On the statistical physics of chains and rods, with application to multi-scale sequence-dependent DNA modelling

THÈSE N^{O} 6977 (2016)

PRÉSENTÉE LE 27 MAI 2016

À LA FACULTÉ DES SCIENCES DE BASE
CHAIRE D'ANALYSE APPLIQUÉE
PROGRAMME DOCTORAL EN MATHÉMATIQUES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Alexandre Emmanuel GRANDCHAMP

acceptée sur proposition du jury:

Prof. A. Abdulle, président du jury
Prof. J. Maddocks, directeur de thèse
Prof. Sir J. M. Ball, rapporteur
Prof. R. Everaers, rapporteur
Prof. H. C. Oettinger, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2016

Acknowledgements

I am grateful to the Prof. J. H. Maddocks who gave me the opportunity to study so many interesting problems in Mechanics and to learn a variety of approaches in Mathematical modelling. The strong interdisciplinary character of his research group offers an ideal atmosphere to examine and to apprehend complex phenomena arising in Nature, such as DNA physics. In particular, the work presented in this thesis has greatly benefited from the tight collaboration with Jarek Glowacki, by the implementation of numerical methods and interactive visualisations. Carine Tschanz has constantly been available to solve and to take care of any administrative issues. I would like to thank Jonatan, Marco, Daiva, Philippe, Ludovica, Nadia, Pauline, Alessandro and Thomas for all the scientific and non-scientific pleasant moments that we have shared. Moreover, I am also grateful to G. Chirikjian, A. Stasiak, F. Benedetti, R. Phillips, T. S. Ratiu, J. Stubbe, S. S. Antman, S. Harvey, F. Gallaire, and P. Monkewitz for the inspiring discussions we had, which have significantly participated to the way I see today's Science.

I wish to acknowledge Prof. A. Abdulle, Prof. H. C. Öttinger, Prof. Sir J. M. Ball and Prof. R. Everaers to have accepted to be jury member for this thesis, for the energy they have invested and for the discerning remarks that they have made.

J'en profite aussi pour remercier mes parents, ma soeur et Myriam pour leur compréhension et leur soutient infaillible durant toutes ces années.

Ce travail a été supporté par le Fonds National Suisse (FNS) de la recherche scientifique. Projets 200021_126666 et 200020_143613.

Résumé

Les mécanismes complexes qui participent au fonctionnement de la cellule sont de mieux en mieux compris depuis une centaine d'années et le rôle central de l'ADN est maintenant établi. La séquence de paires de bases qui caractérise la structure d'un fragment d'ADN ne code pas seulement l'information génétique mais induit aussi localement des propriétés physiques particulières, comme des régions remarquablement courbées ou rigides. On pense par exemple que ces variations dans la constitution de la molécule sont impliquées dans la reconnaissance ADN-protéine et dans le positionnement des nucléosomes. Modéliser comment les propriétés mécaniques de l'ADN dépendent de la séquence constitue donc un progès important dans la compréhension de plusieurs phénomènes biologiques. Cependant, les processus physiques de la cellule interviennent à des échelles très différentes, de quelques paires de bases à plusieurs milliers, et cela rend difficile d'établir un modèle approprié. La modélisation multi-échelle de la mécanique de l'ADN et de sa dépendence en séquence semble donc être une approche intéressante. C'est dans ce contexte que les modèles de bases rigides et de paires de bases rigides ont été proposés. Dans ces modèles gros grains chaque base, ou chaque paire de bases, est décrite par la configuration d'un corps solide, ce qui mène à une représentation de chaîne ou de bichaîne de la molécule d'ADN. Pour chaque séquence une distribution dans l'espace des configurations a été paramétrée, soit à partir de donnés expérimentales soit à partir de trajectoires venant de simulations tout atomes de la dynamique moléculaire.

Les questions importantes que l'on peut étudier avec ces modèles sont par exemple l'influence de la séquence sur la probabilité de contact de deux sites distants le long de la molécule, ou l'espérance de la configuration relative entre ces deux sites. Dans cette thèse, nous proposons d'aborder ces situations physiques du point de vue de la modélisation discrète et continue séparement, puis de discuter en quel sens elles constituent finalement un seul point de vue multi-échelle. Dans la première partie, nous discutons des propriétés méchaniques des chaînes et des bichaînes faites de corps solides hétérogènes, ainsi que de leur analogue continus qui sont les tiges et les bitiges, dans le cadre de la statique classique et dans celui de la physique statistique d'équilibre. Les conditions d'équilibres, les principes variationels et les distributions sur l'espace des configurations sont étudiés pour les chaînes et les tiges seules, puis étendus aux bichaînes et aux bitiges. Nous présentons ensuite des règles de Cauchy-Born déterministe et stochastique permettant d'unifier les représentations discrètes et continues. Dans la seconde partie, nous présentons des applications de la théorie mécanique multi-échelle proposée pour des modèles de chaînes et de tiges qui dépendent de la séquence d'ADN. Nous discutons les

approximations obtenues en utilisant le modèle de bitige pour calculer les configurations de probabilité maximum qui satisfont des conditions de bord données. De manière similaire, nous présentons ensuite le calcul de la matrice de corrélation d'orientations et du vecteur de persistance de Flory pour les chaînes, mais en utilisant le modèle de tige continue, et une méthode d'homogénéisation est proposée. Ces résultats sont pensés comme constituant un important progrès dans la modélisation multi-échelle de la mécanique de l'ADN.

Abstract

The complex mechanisms involved in cellular processes have been increasingly understood this past century and the central role of the DNA molecule has been recognised. The base pair sequence along a DNA fragment is observed not only to encode the genomic information, but also to induce locally very specific physical properties, such as significantly bent or stiff regions. These variations in the molecule constitution are for instance believed to be involved in DNA-protein recognition and in nucleosome positioning. Modelling sequence dependent DNA mechanical properties is consequently an important step towards understanding many biological functions. However, in a cell, vastly different length scales are involved, ranging from a few base pairs to several thousands, which makes the definition of *one* appropriate model difficult. A promising strategy seems to be given by multi-scale modelling of sequence dependent DNA mechanics. In this framework, sequence dependent rigid base and rigid base pair models have been proposed. In these coarse grain models either each base pair or each base is described as a rigid body configuration, which leads to either a chain or a bichain representation of the DNA molecule. A sequence dependent configurational distribution is then parametrized, either from experimental data or directly from atomistic molecular dynamic simulations, to provide an efficient and realistic description at the scale of hundreds of base pairs.

Important questions that can be studied in these models are for instance the influence of the sequence on the probability of contact of two sites, which are distant along the molecule length, or on the expectation of the relative configuration of these two sites. In this thesis, we propose to approach these physical situations both from the discrete and the continuum modelling point of view, and then to discuss in which sense they actually constitute only one multi-scale point of view. In the first part, we discuss mechanical properties of heterogeneous rigid body chains and bichains, as well as continuum rods and birods, in both classical statics and in equilibrium statistical physics. Equilibrium conditions, variational principles and configurational distributions are studied for single chains and rods, and then extended to bichains and birods. We have introduced in particular an original coordinate free Hamiltonian formulation in arc-length of the birod equilibrium conditions, and the notion of the persistence matrix for the configurational moment for chains and rods. We then present deterministic and stochastic exponential Cauchy-Born rules allowing a bridge between the scales of the discrete and continuum representations. In the second part, we present applications of the proposed multi-scale mechanical theory for chains and rods to sequence dependent DNA modelling. We discuss the approximation using the birod model of most probable bichain configurations

satisfying prescribed end conditions. Similarly, we then present a computation of the sequence dependent frame correlation matrix and the Flory persistence vector for chains using a continuum rod model. In addition, a homogenisation method is proposed. These results are believed to constitute a substantial improvement in the multi-scale modelling of DNA mechanics.

Keywords *chain, bichain, rod, birod, equilibrium configuration, configurational distribution, persistence matrix, sequence dependent, DNA mechanics, exponential Cauchy-Born rule.*

Contents

List of Figures

Introduction

Cells have been identified as the elementary blocks of life and significant developments about their complex mechanism have been made. Among the variety of characteristic molecules that they contain, deoxyribonucleic acid, or simply *DNA*, has been observed to play a central role in the functions of regulation and replication, for instance. Informally, DNA has a double stranded structure and each strand is made of bases connected by sugar-phosphate backbones. Each *base* is of four[1] possible compositions, namely adenine (A), thymine (T), guanine (G) and cytosine (C). A standard *complementary pairing* is made between A and T and between G and C with respectively two and three hydrogen bonds. A *base pair* is then the couple made of two complementary bases. Each strand possesses a natural reading direction, called the $5' \to 3'$ direction, and the DNA double helix is such that marching along one strand in this direction is marching along the complementary strand in the opposite direction. In other words, DNA has an *anti-symmetric* structure. The *sequence* is then the base constitution of the fragment, read following the $5' \to 3'$ direction along one of the two strands. The total DNA in a cell is called the *genome*. Its length varies from 10^3 base pairs for some bacterias to 10^{12} base pairs for some plants. The human genome is of the order of 10^9 base pairs. Since the typical distance between two consecutive base pairs is about $3 - 4$ Å, the characteristic length scales of DNA physics go from a few nanometers to several decades of centimetre. An increasing number of full genomes have been sequenced in the past twenty years. Even if not all possible sequences appear in Nature, their variety is gigantic and, consequently, DNA molecules are an archetype of the class of *heterogeneous polymers*. A more precise and more complete discussion on the role of DNA in cells can be found for instance in [Calladine et al., 2004, Phillips et al., 2009].

A fundamental question in the modelling of DNA physics concerns the relation between the sequence and the mechanical properties of a given fragment. The numerous existing sequences together with the laborious character of experiment makes it difficult to arrive at simple and brief conclusions. The sequence dependence of DNA mechanics is nevertheless recognised in many circumstances. For instance, particular fragments have been noted to contain regions that are significantly and intrinsically bent [Marini et al., 1982, Kabsch et al., 1982, Kitchin et al., 1986, Levene et al., 1986, Bednar et al., 1995, Vologodskaia and Vologodskii, 2002] and local DNA stiffness has been observed

[1]The exact number of possible bases is still a topic of research. While the standard DNA bases are made of only four chemical structures, the existence of various base modifications, methylation for instance, is known to be biologically significant.

to vary between difference sequences [Théveny et al., 1988, Olson et al., 1998, Virstedt et al., 2004]. Special regions in the genome are thought to be important in biological processes involved in the cells, see for instance [Hagerman, 1990], and to be related, more particularly, to DNA-protein recognition, see for instance [Matthews, 1988, Koudelka et al., 2006], and to nucleosome positioning in eukaryotic cells, as discussed for instance in [Thåström et al., 1999, Virstedt et al., 2004, Segal and Widom, 2009]. However, most of the existing models cannot accurately encompass the sequence dependent physical behavior of DNA molecules.

Available realistic sequence dependent DNA mechanical models can be divided into two classes: *atomistic models* and *coarse grain models*. Atomistic DNA models aim for a very fine description of the molecule, namely at the level of atoms, see for instance [Lavery and Hartmann, 1994, Lankaš et al., 2000, Lankaš et al., 2003, Lavery et al., 2009, Lavery et al., 2014, Pasi et al., 2014]. The main issues with this approach are first that these models needs a very large set of parameters which are difficult to get since they have to come mainly from *ab initio* principle, such as quantum chemistry, and second that they require intensive computer resources which prevent, nowadays, the study of phenomena involving more than a few decades of base pairs for a few microseconds. Even if the predictions of these models are definitely improving, their accuracy is still questionable. In contrast, coarse grain DNA model describe the molecule as a set of units which contain several atoms, and, keeping the objective of modelling the sequence dependence mechanics of DNA, the two cases of interest here are *rigid base* and *rigid base pair* coarse graining degree, see for instance [Olson et al., 1993, Marky and Olson, 1994, Olson, 1996, Olson et al., 1998, Gonzalez and Maddocks, 2001, Olson et al., 2001, Lankaš et al., 2003, Lankaš et al., 2009, Gonzalez et al., 2013]. Even if these models contain less informations than atomistic models, they are expected to describe in an efficient and appropriate way the sequence dependent features of DNA molecules at the scale of several bases.

Conversely, continuum and homogeneous models have became popular to describe DNA mechanical properties at larger scale. Long macromolecules have been modelled, using homogeneous *Gaussian chains* according to the effective theory of *Kuhn segments*, see for instance [Kuhn, 1934, Flory, 1969]. For shorter fragments, however, this approach fails because the detailed physical constitution of the chain can no longer be ignored, as pointed out in [Bresler and Frenkel, 1943] for instance. Inspired from the work by M. Smoluchowski on stochastic trajectories and originally applied to polymers in [Porod, 1948, Krakty and Porod, 1949], the idea of *Kratky-Porod persistence length* allows the characterisation of some statistical properties of DNA at intermediate scales, say from 10^3 to 10^5 base pairs. This quantity is shown to be related to some bending stiffness parameter in the so-called *Wormlike chain* model which describes DNA as an uniform and intrinsically straight material which only responds to bending deformations, other degrees of freedom like shear, stretch or twist being ignored. Even if this model is quite idealised, it has led to remarkably good results to model average DNA properties. Substantial improvements have been made into the *Helical Wormlike chain* model which additionally allows twist deformations and a helical ground state, see for instance [Bugl,

1969, Yamakawa, 1976, Benham, 1977, Benham, 1979, Yamakawa, 1997]. Despite the propositions of increasingly realistic and sophisticated continuum rod models [Marko and Siggia, 1994, Shi et al., 1996, Manning et al., 1996, Marko, 1997, Becker and Everaers, 2007, Becker, 2007], there exist only a few examples discussing how sequence dependence can explicitly be included in the rod constitutive parameters [Manning et al., 1996, Becker, 2007].

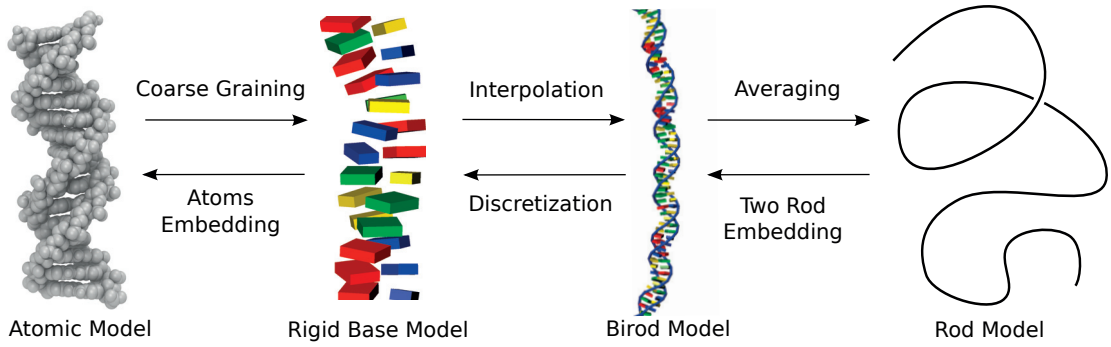


Figure 1: Sketch of multi-scale modelling of DNA molecule.

The development of a mechanical model that has the ability to include the inherent discrete nature of the nucleotide sequence in a DNA fragment and its continuum character at the scale of biological processes in a cell appears, consequently, as a significant step in the description of DNA molecules. Modelling sequence dependence of physical properties is a challenging objective. First, because most of the classical results in polymer mechanics do not apply to heterogenous chains, which implies that new mathematical results have to be stated; and second, because many biological phenomena of interest require hundreds of base pairs to be described and need therefore the ability of predicting large scale behavior as a function of the sequence. These objectives constitute the core of this thesis, and we present here the current state of our progress. More precisely, in part I, we propose a multi-scale mechanical study by considering, on one side, a discrete heterogeneous chain theory and, on the other side, its analogue for continuum rods and, then, a method to bridge these two theories. Then, in part II, these results are illustrated by the computation of sequence dependent most probable configurations respecting end conditions and by the study of the sequence dependent values of the frame correlation matrix and of the Flory persistence vector. A sketch of our multi-scale modelling approach is drawn in figure 1.

Part I, even if it is largely inspired by applications in the context of DNA, can actually be taken as a pure discussion on the mechanical modelling of chains and rods and consequently may be of interest in the context of other microscopic or macroscopic filamentary material. The first chapter is preliminary and comprises the basic, but necessary, mathematical definitions to state in a convenient way the results belonging to this thesis. It should only be read if more technical detail than presented in the following chapters is desired.

The second chapter is devoted to the properties of chains in both classical statics and statistical mechanics. A rigid body chain configuration $\boldsymbol{g} = (\boldsymbol{g}_1, ..., \boldsymbol{g}_N)$ is made of an ordered sequence of N rigid body configurations $\boldsymbol{g}_n = (\boldsymbol{R}_n, \boldsymbol{r}_n) \in \mathrm{SE}(3)$ where the rotation matrix $\boldsymbol{R}_n \in \mathrm{SO}(3)$ denotes the orientation of the n^{th} rigid body and the vector $\boldsymbol{r}_n \in \mathbb{R}^3$ its position. Similarly, a double rigid body chain configuration $(\boldsymbol{g}^+, \boldsymbol{g}^-) = (\boldsymbol{g}_1^+, \boldsymbol{g}_1^-, ..., \boldsymbol{g}_N^+, \boldsymbol{g}_N^-)$ is made of two rigid body chain configurations denoted respectively by $\boldsymbol{g}^+$ and $\boldsymbol{g}^-$. The equilibrium equations are discussed and a variational principle presented for both chains and double chains. Motivated by the rigid base model for DNA molecules in [Gonzalez et al., 2013], the bichain model is introduced. It describes a double chain by a macrostructure configuration $\boldsymbol{g}$, which is a chain configuration designed to characterise the average deformation of the double chain, and a microstructure configuration $\mathcal{P}$, which encodes the local difference between the original double chain configuration and the macrostructure configuration. Based on the analysis of double chains, equilibrium conditions as well as a variational principle are stated for a bichain configuration $(\boldsymbol{g}, \mathcal{P})$. In particular, we show how the problem of computing equilibrium bichain configurations where the end configurations are prescribed can be described using a Lagrange multiplier technique, and, moreover, that the multipliers exactly correspond to the external couple and force needed to respect the desired Dirichlet conditions. The second half of this chapter concerns the study of statistical physics properties of chains and bichains. Following the arguments proposed in [Walter et al., 2010], a rigid body chain in a heat bath, and respectively a bichain, is assumed to be governed by a stationary configurational distribution $d\rho(\boldsymbol{g})$, and respectively, $d\rho(\boldsymbol{g}, \mathcal{P})$, of the Boltzmann form. In the case of a single chain with a local deformation energy, general analytical results on the first moment of the relative rigid body displacement $\langle \boldsymbol{g}_{m,n} \rangle = \langle \boldsymbol{g}_m^{-1} \boldsymbol{g}_n \rangle$ are then developed, where $\langle \, . \, \rangle$ denotes the expectation with respect to the appropriate configurational distribution. In particular, the notion of expected chain configuration is introduced which allows, for chains in a semi-flexible regime, the derivation of an explicit recurrence relation $\langle \boldsymbol{g}_{m,n+1} \rangle = \langle \boldsymbol{g}_{m,n} \rangle \langle a_n \rangle$ based on a first order expansion in local covariance matrices. These analytical developments demonstrate the equal importance of the expected chain configuration $\bar{\boldsymbol{g}}_{m,n}$ and of the local covariance matrices in the behavior of the first moment $\langle \boldsymbol{g}_{m,n} \rangle$ and can be seen as a generalisation, in a mathematical sense, of the work in [Trifonov et al., 1988, Schellman and Harvey, 1995]. They also motivate the definition of the persistence matrix $\langle \boldsymbol{\mathcal{D}}_{m,n} \rangle$ obtained by the factorization $\langle \boldsymbol{g}_{m,n} \rangle = \langle \boldsymbol{\mathcal{D}}_{m,n} \rangle \bar{\boldsymbol{g}}_{m,n}$. The persistence matrix is shown to characterise the memory lost in the entries of the configurational first moment $\langle \boldsymbol{g}_{m,n} \rangle$ and allows a derivation of a closed form expression in a short length expansion.

The third chapter is designed as the continuum analogue of the second, and presents properties of rods in classical statics and equilibrium statistical mechanics. A continuum rod configuration $\boldsymbol{g} = \boldsymbol{g}(s)$ is a continuous curve in SE(3) for $s \in]0, L[$ where L denotes a reference length and $\boldsymbol{g}(s) = (\boldsymbol{R}(s), \boldsymbol{r}(s)) \in \mathrm{SE}(3)$ is the continuum version of the rigid body configuration of the rod cross-section. A double rod configuration is of the form $(\boldsymbol{g}^+, \boldsymbol{g}^-) = (\boldsymbol{g}^+(s), \boldsymbol{g}^-(s))$ where the two rod configuration are denoted respectively by $\boldsymbol{g}^+$ and $\boldsymbol{g}^-$. Following the work in [Moakher and Maddocks, 2005], we then

define the birod configuration, denoted $(\boldsymbol{g}(s), \mathcal{P}(s))$, which comprises a rod macrostructure $\boldsymbol{g}(s)$, representing the double rod configuration in average, and a microstructure $\mathcal{P}(s)$, representing the difference between the actual double rod configuration and the macrostructure, analogously to bichains. Equilibrium conditions and variational principles are presented for rods and birod respectively. In addition, a coordinate independent Hamiltonian formulation of the equilibrium conditions is obtained. Similarly to chains, we present a Lagrange multiplier formulation of the equilibrium birod conditions with prescribed end configurations, and show that the set of multipliers corresponds to external couple and force end loadings. The statistical physics of rods and birods is then studied. The configurational distribution for rods, respectively birods, in the context of equilibrium statistical mechanics, are introduced as path integral densities and we discuss how these infinite dimensional distributions can be interpreted as the appropriate limit of a sequence of chains distributions for the discretised rod configurations. This development is then extended to observables and their expectations and, similarly to chains, we emphasize the special role of the configurational first moment which comprise, now in a continuum version, the frame correlation matrix and the Flory persistence vector. An explicit first order approximation is then obtained for the configurational first moment for single rods which have a local internal energy and are assumed to be in a semi-flexible regime. A similar formula is obtained in [Becker, 2007], but is there based on the analogy with the Brownian dynamics of a rigid body. In analogy to the case of chains, the analytical expression show the equal importance of the expected rod configuration and of the fluctuation matrix, and allow the introduction of a new quantity called the persistence matrix. It is shown to encode the decay rates in the entries of the configurational first moment and yields a closed form expression in a short length expansion.

The fourth chapter discusses how to bridge the scales between chains and rods and the sense in which continuum models can be understood as providing approximations of rigid body chain mechanical properties. We first introduce local interpolation rules for chains and bichains, and discuss separately the notion of deterministic and stochastic exponential Cauchy-Born rules. The term exponential Cauchy-Born rule was introduced originally in [Arroyo and Belytschko, 2004] in the context of continuum modelling of carbon nanotubes from all atom configurations. The notion of consistent discrete and continuum energies and stationary configurations for chains and rods are presented, and we briefly expose why, under specific hypotheses, continuum equilibrium configurations can provide approximations of discrete equilibrium configurations. We then introduce the definitions of consistent discrete and continuum observables and configurational distributions for chains and rods and conclude that in general deterministic exponential Cauchy-Born rules do not provide stochastic exponential Cauchy-Born rules. An explicit deterministic Cauchy-Born rule is proposed for quadratic bichains models which has been designed to apply to the rigid base model in [Gonzalez et al., 2013]. This rule is obtained by matching the first and second order terms in an expansion of the bichain energy which provides consistency only at leading orders. Similarly, we then discuss the construction of an explicit stochastic Cauchy-Born rule. However, this rule is only for single chains and single rods, and in the case that they admit normal configurational distributions and

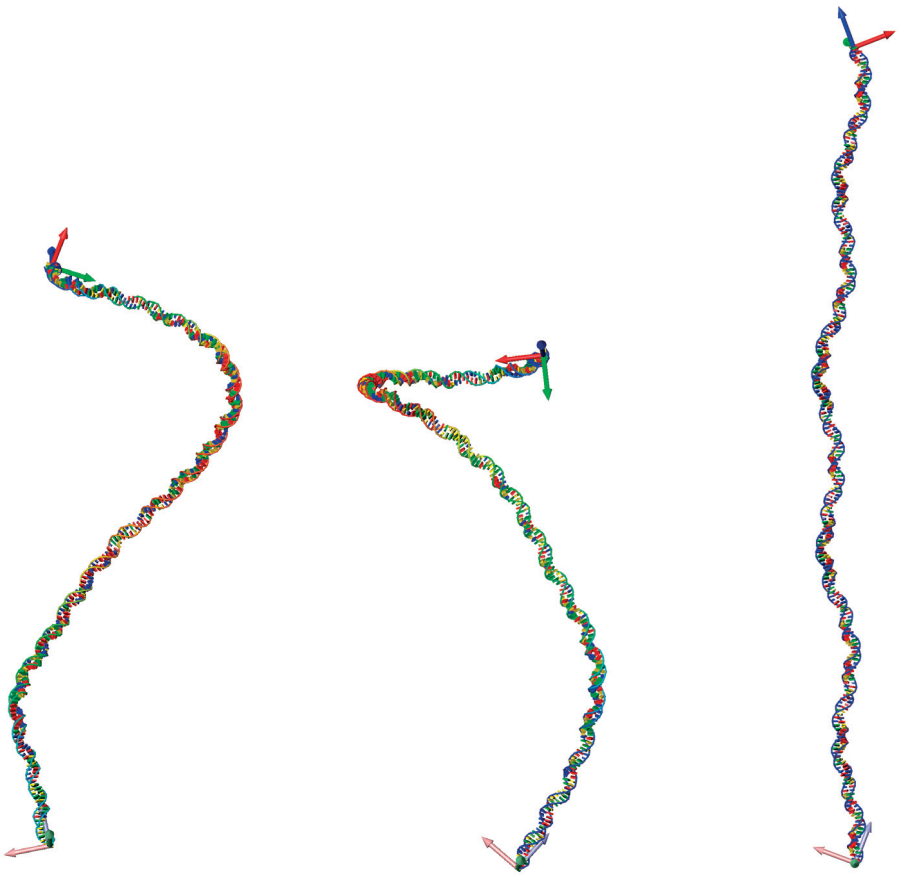
are in a semi-flexible regime. It is deduced by matching the first and second moments of the chain internal coordinates.

Part II is dedicated to the application of chain and rod theories to the multi-scale modelling of sequence dependent DNA mechanics. More precisely, we discuss the computation of the most probable bichain configuration and of the configurational first moment using both a discrete and a continuum approach.

The fifth chapter introduces sequence dependent rigid base pair and rigid base models. They respectively describe as a rigid body configuration either each individual base pair or each individual base in a DNA fragment, which leads to either rigid body chain or rigid body bichain configurations. We propose a sequence dependent continuum rod model, using the stochastic exponential Cauchy-Born rule for chains defined in the previous chapter, and a sequence dependent continuum birod model, using the deterministic exponential Cauchy-Born rule for bichains. In both cases, the associated constitutive coefficients are observed to be discontinuous at each base pair and to have a manifestly heterogeneous character at the scale of hundreds of base pairs. We conclude however that the deterministic exponential Cauchy-Born rule applied to the rigid base model is satisfactory since for instance the continuum birod and the original bichain energy agree to less than 0.5% for randomly sampled bichain configurations.

The sixth chapter discusses the computation of most probable configurations respecting end conditions in the sequence dependent rigid base model and in the sequence dependent birod model. Such configurations are believed to be an interesting first step in the characterisation of the sequence dependent statistical physics properties of a DNA fragment. However, since the second variation still requires further investigation, we restrict the discussion to stationary configurations. In order to demonstrate that the sequence dependent continuum birod model provides accurate approximations of the stationary bichain configurations, we first present numerical solutions of the continuum equilibrium conditions, and then use them in an optimisation method to compute the solution of the discrete equilibrium conditions. We present the Hamiltonian formulation of the birod equilibrium conditions and the sequence dependent coefficients obtained using the deterministic exponential Cauchy-Born rule for bichains. A study of the equilibrium configurations for under and over twisted end conditions, corresponding respectively to applied torque of -300pNÅ and $+300$pNÅ, for the first 240 base pairs of three different sequences denoted by λ , CF and T_L is presented. The first sequence comes from the genome of a virus called *Lambda phage*, which infects the *E. Coli bacterium*, the second is a fragment of the *Crithidia fasciculata* genome, which is a mosquito parasite, and has the particularity to contain a large intrinsic bend [Kitchin et al., 1986], and the last one has been artificially designed to have a pronounced intrinsic super-helical structure which is left-handed. Figure 2 presents the equilibrium birod configurations obtained for these three different sequences, but for the same under twisted load conditions. The colour map along the backbone strands illustrates the distribution of the norm of the local couple around each base pairs. The different mechanical response corresponding to the different sequences is significant. We then compare these solutions with their corresponding bichain numerical solutions, and even if these results are only at a preliminary

stage, we believe them to be quite satisfactory.



(a) λ (b) CF (c) T_L

Figure 2: Under twisted birod configurations for three different sequences, but for the same applied twist loading.

The last chapter discusses the features of the configurational first moment, or equivalently of the frame correlation matrix and of the Flory persistence vector, for the sequence dependent rigid base pair model. These chain observables are important since they can characterise the possible constitutive anisotropy and possible non-trivial intrinsic geometry of a given DNA molecule. We discuss the limitation of the existing models, and in particular of the isotropic Wormlike chain, to describe sequence dependent features. The quality of the discrete and continuum approximation formulas for the configurational first moment of heterogeneous chains and rods, presented in the second and third chapter, are assessed in a numerical comparison with the values obtained by Monte Carlo simulations. However, since these expressions are only valid for local models, we use the sequence dependent rigid base pair and sequence dependent rod model for the discussion. We first concentrate on the chain formulation and the results of the explicit recurrence

relation are compared with a Monte Carlo simulations for the sequences λ , CF and T_L . In each case the differences are hardly perceptible, even after 500 base pairs, and the explicit formulas are consequently thought be very satisfactory. The factorization into the expected chain and the persistence matrix is then illustrated. Remarkably it shows that the fast variation in the configurational first moment reflects the expected chain configuration whereas the slow convergence of the entries is due to the persistence matrix. This factorization can be seen, in a mathematical sense, as a generalisation of the studies in [Trifonov et al., 1988, Schellman and Harvey, 1995] of the effect of localised intrinsic bends on chain statistics, and of the notion of static and dynamic persistence length. We also present the results of the closed form solution given by the short length expansion, that are observed to be valid until a distance of about 135 base pairs. We then apply the exponential stochastic Cauchy-Born rule to discuss the continuum approximations of the configurational first moment of chains. The numerical solutions of the proposed ODE are compared with the results of Monte Carlo simulations obtained for the sequence λ and, as for the discrete approximation, the difference between the values is hardly perceptible. The exponential stochastic Cauchy-Born rule for chains is thus thought to be very satisfactory and to constitute a significant step forwards in the multi-scale modelling of DNA statistical physics properties. However, the strong heterogeneous character of this formulation forces a very dense numerical discretization to solve the ODE and the continuum model appears computationally more expensive than the original base pair model, which is of course not what is desired in the introduction of a continuum model. To overcome this fundamental issue, we finally propose the application of a homogenisation technique for ODE systems which possess two distinct characteristic scales. However, since the coefficients do not exhibit any of the classical hypotheses, such as periodicity or ergodicity, we only obtain a local averaging principle. Even if the optimal size of the window used to average should be further investigated, the choice of about 30Å delivers smooth and slowly varying coefficients which are used to solve the associated homogenised system. Its solution is then compared to the results obtained from the Monte Carlo simulation for the sequence λ and we observe that not only is the difference hardly perceptible, but also the number of numerical steps is now strictly less than the number of base pairs in the original chain. The homogenised sequence dependent ODE constitutes then a large scale description with slowly varying coefficient and is, consequently, less computationally expensive than the original discrete recurrence relation.

Part I

On the Statistical Physics of Chains and Rods

1. Preliminaries

We introduce in this first chapter some basic notations and preliminaries on matrix groups before starting the actual discussion. In particular, we present formally the special Euclidean group SE(3), which describes the set of all possible rigid body displacement in space, and some specific algebraic properties related to infinitesimal variations. Finally, we discuss some notable parametrization of this group in the context of DNA modelling. Since the key expressions are either restated in the main text or explicitly referred to this chapter, we strongly invite the reader to go directly to chapter 2 and only to return to this part if more technical detail is desired.

1.1 Basics on Matrix Groups

Throughout this thesis we use $\mathbb{R}^{m\times n}$ to denote the *set of $m \times n$ matrices with real coefficients*, where m and n are integers. The *product of two matrices* $A \in \mathbb{R}^{k\times m}$ and $B \in \mathbb{R}^{m\times n}$ will simply be denoted by $AB \in \mathbb{R}^{k\times n}$. The *inner product of two matrices* $A, B \in \mathbb{R}^{m\times n}$ is denoted $A : B \in \mathbb{R}$ and is defined as

$$A : B = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}B_{ij} = \mathrm{Tr}\left(A^T B\right) \tag{1.1.1}$$

where A_{ij} and B_{ij} denote the entries of the matrix A and B respectively, A^T the transposed matrix of A and Tr the usual trace of a matrix. In particular, for any matrices $A \in \mathbb{R}^{k\times m}$, $B \in \mathbb{R}^{k\times n}$ and $C \in \mathbb{R}^{n\times m}$ one has the following property

$$A : BC = B^T A : C = AC^T : B. \tag{1.1.2}$$

This inner product induces the matrix norm $|\,.\,|$ defined as

$$|A| = (A : A)^{\frac{1}{2}} \tag{1.1.3}$$

for any matrix $A \in \mathbb{R}^{k\times m}$, which is called the *Fröbenius norm*. In many circumstances, we will also make use of the *spectral matrix norm*, denoted by $|\,.\,|_{\mathrm{Sp}}$ and defined as

$$|A|_{\mathrm{Sp}} = \max\left\{ |\lambda|^{\frac{1}{2}} \,|\, \lambda \in \mathrm{Sp}(A^T A) \right\} \tag{1.1.4}$$

where $\mathrm{Sp}(A^T A)$ denotes the set of eigenvalues of the matrix $A^T A$. In particular the spectral norm has the property that

$$|AB|_{\mathrm{Sp}} \leq |A|_{\mathrm{Sp}} \, |B|_{\mathrm{Sp}} \tag{1.1.5}$$

for any matrices $A \in \mathbb{R}^{k \times m}$, $B \in \mathbb{R}^{k \times n}$. The *matrix commutator* of $A, B \in \mathbb{R}^{n \times n}$ is denoted by $[A, B] \in \mathbb{R}^{n \times n}$ and is defined as

$$[A, B] = AB - BA. \tag{1.1.6}$$

The *space of differentiable curves* on $\mathbb{R}^{m \times n}$ is denoted by $\mathrm{C}^1(]0, L[, \mathbb{R}^{m \times n})$. This space can be defined as $A(s) \in \mathrm{C}^1(]0, L[, \mathbb{R}^{m \times n})$ if and only if all the components of the matrix $A(s)$ are differentiable for all $s \in]0, L[$.

The *Special Orthogonal group* in $\mathbb{R}^{3 \times 3}$, denoted by SO(3), can be defined as

$$\mathrm{SO}(3) = \left\{ \boldsymbol{R} \in \mathbb{R}^{3 \times 3} \,|\, \boldsymbol{R}^T \boldsymbol{R} = I \text{ and } \det \boldsymbol{R} = 1 \right\}. \tag{1.1.7}$$

This group represents, for instance, the set of all possible rotations in the Euclidean space. The tangent space to this group at the identity, which consists in all infinitesimal variations within the group around the identity matrix, is denoted by so(3) and can be shown to be

$$\mathrm{so}(3) = \left\{ [b\times] \in \mathbb{R}^{3 \times 3} \,|\, b \in \mathbb{R}^3 \right\} \tag{1.1.8}$$

which is the vector space of all skew-symmetric matrices in $\mathbb{R}^{3 \times 3}$ and where the operator $[\,.\,\times] : \mathbb{R}^3 \to \mathbb{R}^{3 \times 3}$ is defined by

$$[b\times] = \begin{pmatrix} 0 & -b_3 & b_2 \\ b_3 & 0 & -b_1 \\ -b_2 & b_1 & 0 \end{pmatrix} \text{ for all triples } b = (b_1, b_2, b_3) \in \mathbb{R}^3. \tag{1.1.9}$$

The operator $[\,.\,\times]$ is called the *skew operator* and this notation is meant to emphasize the property

$$[b_1\times]\, b_2 = b_1 \times b_2 \text{ for all } b_1, b_2 \in \mathbb{R}^3 \tag{1.1.10}$$

where $\times$ denotes the usual *vector product* on $\mathbb{R}^3$. Observe that we have the following identity

$$[[b_1\times]\,, [b_2\times]] = [(b_1 \times b_2)\times] \tag{1.1.11}$$

for all $b_1, b_2 \in \mathbb{R}^3$ where the commutator $[\,.\,,\,.\,]$ is defined in (1.1.6). The space so(3) is called the *Lie algebra* of the group SO(3). The operator $[\,.\,\times]$ has a unique adjoint operator, denoted by $\mathrm{Vect}\,(\,.\,) : \mathbb{R}^{3 \times 3} \to \mathbb{R}^3$, with respect to the inner product of matrices defined in (1.1.1). Precisely, one has that

$$B : [b\times] = \mathrm{Vect}\,(B) \cdot b \tag{1.1.12}$$

for any matrix $B \in \mathbb{R}^{3\times 3}$ and for all triples $b \in \mathbb{R}^3$ and where $\cdot$ denotes the usual inner product for vectors in $\mathbb{R}^3$, which leads to the definition

$$\mathrm{Vect}\,(B) = \begin{pmatrix} B_{32} - B_{23} \\ B_{13} - B_{31} \\ B_{21} - B_{12} \end{pmatrix}. \tag{1.1.13}$$

According to the definition (1.1.8), if $\boldsymbol{R}(t) \in \mathrm{SO}(3)$ denotes a differentiable curve on the rotation group, there always exists some triples $\phi^{\boldsymbol{R}}(t) \in \mathbb{R}^3$ and $\boldsymbol{\phi}^{\boldsymbol{R}}(t) \in \mathbb{R}^3$, called *Darboux vector*, such that

$$\frac{d}{dt}\boldsymbol{R}(t) = \left[\boldsymbol{\phi}^{\boldsymbol{R}}(t)\times\right]\boldsymbol{R}(t) = \boldsymbol{R}(t)\left[\phi^{\boldsymbol{R}}(t)\times\right] \tag{1.1.14}$$

and moreover, using the property (1.1.10), these vectors are related by

$$\boldsymbol{R}(t)\phi^{\boldsymbol{R}}(t) = \boldsymbol{\phi}^{\boldsymbol{R}}(t). \tag{1.1.15}$$

Another important Lie group, which will be at the heart of this thesis, is the *Special Euclidean group* in $\mathbb{R}^{4\times 4}$, denoted by SE(3). It can be defined by

$$\mathrm{SE}(3) = \left\{ \boldsymbol{g} \in \mathbb{R}^{4\times 4} \,|\, \boldsymbol{g} = \begin{pmatrix} \boldsymbol{R} & \boldsymbol{r} \\ 0 & 1 \end{pmatrix} \text{ where } \boldsymbol{R} \in \mathrm{SO}(3) \text{ and } \boldsymbol{r} \in \mathbb{R}^3 \right\}. \tag{1.1.16}$$

This group describes, for instance, the set of all possible rigid body configurations in Euclidean space. The matrix $\boldsymbol{R}$ and the vector $\boldsymbol{r}$ are called respectively the rotational and translational part of the element $\boldsymbol{g}$. The product of two elements $\boldsymbol{g}_1, \boldsymbol{g}_2 \in \mathrm{SE}(3)$ is given by the regular matrix product, i.e. by

$$\boldsymbol{g}_1\boldsymbol{g}_2 = \begin{pmatrix} \boldsymbol{R}_1\boldsymbol{R}_2 & \boldsymbol{R}_1\boldsymbol{r}_2 + \boldsymbol{r}_1 \\ 0 & 1 \end{pmatrix} \tag{1.1.17}$$

and the inverse $\boldsymbol{g}^{-1}$ is

$$\boldsymbol{g}^{-1} = \begin{pmatrix} \boldsymbol{R}^T & -\boldsymbol{R}^T\boldsymbol{r} \\ 0 & 1 \end{pmatrix}. \tag{1.1.18}$$

Equivalently, we will also use the shortened notation

$$\boldsymbol{g} = (\boldsymbol{R}, \boldsymbol{r}) \in \mathrm{SE}(3) \tag{1.1.19}$$

and one will have then formally

$$\boldsymbol{g}_1\boldsymbol{g}_2 = (\boldsymbol{R}_1\boldsymbol{R}_2, \boldsymbol{R}_1\boldsymbol{r}_2 + \boldsymbol{r}_1) \text{ and } \boldsymbol{g}^{-1} = (\boldsymbol{R}^T, -\boldsymbol{R}^T\boldsymbol{r}) \tag{1.1.20}$$

according to (1.1.17) and (1.1.18). The tangent space to the special Euclidean group SE(3) at the identity, or its Lie algebra, can be shown to be

$$\mathrm{se}(3) = \left\{ \mathcal{T}\phi \in \mathbb{R}^{4\times 4} \,|\, \phi = (\phi^{\boldsymbol{R}}, \phi^{\boldsymbol{r}}) \text{ with } \phi^{\boldsymbol{R}} \in \mathbb{R}^3 \text{ and } \phi^{\boldsymbol{r}} \in \mathbb{R}^3 \right\} \tag{1.1.21}$$

where the operator $\mathcal{T} : \mathbb{R}^6 \to \mathbb{R}^{4\times 4}$ is

$$\mathcal{T}\phi = \begin{pmatrix} [\phi^{\boldsymbol{R}}\times] & \phi^{\boldsymbol{r}} \\ 0 & 0 \end{pmatrix} \tag{1.1.22}$$

and $[\,.\,\times]$ is defined in (1.1.9). The operator $\mathcal{T}$ is called the *tangent map*. The tangent map has a unique adjoint operator $\mathcal{T}^* : \mathbb{R}^{4\times 4} \to \mathbb{R}^6$ defined by

$$\mathcal{T}b : B = \mathcal{T}^* B \cdot b \tag{1.1.23}$$

for all vectors $b \in \mathbb{R}^6$ and all matrices $B \in \mathbb{R}^{4\times 4}$. According to (1.1.13) one has explicitly

$$\mathcal{T}^* \begin{pmatrix} B_1 & b_2 \\ b_3 & B_4 \end{pmatrix} = \begin{pmatrix} \mathrm{Vect}\,(B_1) \\ b_2 \end{pmatrix} \tag{1.1.24}$$

where $B_1 \in \mathbb{R}^{3\times 3}$, $b_2 \in \mathbb{R}^{3\times 1}$, $b_3 \in \mathbb{R}^{1\times 3}$ and $B_4 \in \mathbb{R}$. In analogy to (1.1.11) one has

$$[\mathcal{T}\phi_1, \mathcal{T}\phi_2] = \mathcal{T} \begin{pmatrix} \phi_1^{\boldsymbol{R}} \times \phi_2^{\boldsymbol{R}} \\ \phi_1^{\boldsymbol{R}} \times \phi_2^{\boldsymbol{r}} + \phi_1^{\boldsymbol{R}} \times \phi_2^{\boldsymbol{r}} \end{pmatrix} \tag{1.1.25}$$

with $\phi_1 = (\phi_1^{\boldsymbol{R}}, \phi_1^{\boldsymbol{r}})$ and $\phi_2 = (\phi_2^{\boldsymbol{R}}, \phi_2^{\boldsymbol{r}})$, and where the commutator $[\,.\,,\,.\,]$ is defined in (1.1.6).

Consequently, using (1.1.20) and (1.1.21), any differentiable curve $\boldsymbol{g}(t) \in \mathrm{SE}(3)$ in the special Euclidean group SE(3) has the property that there exists $\boldsymbol{\phi}(t) = (\boldsymbol{\phi}^{\boldsymbol{R}}(t), \boldsymbol{\phi}^{\boldsymbol{r}}(t)) \in \mathbb{R}^6$ and $\phi(t) = (\phi^{\boldsymbol{R}}(t), \phi^{\boldsymbol{r}}(t)) \in \mathbb{R}^6$ such that

$$\frac{d}{dt}\boldsymbol{g}(t) = \mathcal{T}\boldsymbol{\phi}(t)\boldsymbol{g}(t) = \left(\left[\boldsymbol{\phi}^{\boldsymbol{R}}(t)\times\right]\boldsymbol{R}(t), \boldsymbol{\phi}^{\boldsymbol{R}}(t) \times \boldsymbol{r}(t) + \boldsymbol{\phi}^{\boldsymbol{r}}(t)\right) \tag{1.1.26a}$$

and

$$\frac{d}{dt}\boldsymbol{g}(t) = \boldsymbol{g}(t)\mathcal{T}\phi(t) = \left(\boldsymbol{R}(t)\left[\phi^{\boldsymbol{R}}(t)\times\right], \boldsymbol{R}(t)\phi^{\boldsymbol{r}}(t)\right) \tag{1.1.26b}$$

where we have used the short notation defined in (1.1.19). For a differentiable curve $\boldsymbol{g}(t)$, the vector field $\boldsymbol{\phi}(t)$ which satisfies (1.1.26a) for all t is called the *left infinitesimal generator*, and the analogous $\phi(t)$ in (1.1.26b) is called the *right infinitesimal generator*. The relation between the components of $\boldsymbol{\phi}(t) = (\boldsymbol{\phi}^{\boldsymbol{R}}(t), \boldsymbol{\phi}^{\boldsymbol{r}}(t))$ and $\phi(t) = (\phi^{\boldsymbol{R}}(t), \phi^{\boldsymbol{r}}(t))$ is then

$$\mathcal{T}\boldsymbol{\phi}(t) = \boldsymbol{g}(t)\mathcal{T}\phi(t)\boldsymbol{g}(t)^{-1} \tag{1.1.27}$$

which can be written explicitly as the linear relation

$$\boldsymbol{\phi}(t) = \mathrm{Ad}_{\boldsymbol{g}(t)}\phi(t) \text{ with } \mathrm{Ad}_{\boldsymbol{g}(t)} = \begin{pmatrix} \boldsymbol{R}(t) & 0 \\ [\boldsymbol{r}(t)\times]\boldsymbol{R}(t) & \boldsymbol{R}(t) \end{pmatrix} \tag{1.1.28}$$

for any $\boldsymbol{g}(t) = (\boldsymbol{R}(t), \boldsymbol{r}(t)) \in \mathrm{SE}(3)$. This matrix represents the *operator of adjoint representation* Ad. We observe that (1.1.28) constitutes a slight abuse of notation with

respect to the literature since the general use of this operator is on the Lie algebra and not on one of its representations in coordinates. This operator will be shown to be an important ingredient in many results of this thesis. Note in particular using (1.1.27) one has the following relation

$$\mathrm{Ad}_{\boldsymbol{g}_1\boldsymbol{g}_2} = \mathrm{Ad}_{\boldsymbol{g}_1}\mathrm{Ad}_{\boldsymbol{g}_2} \tag{1.1.29}$$

for any $\boldsymbol{g}_1, \boldsymbol{g}_2 \in \mathrm{SE}(3)$, which implies that

$$\mathrm{Ad}_{\boldsymbol{g}}^{-1} = \mathrm{Ad}_{\boldsymbol{g}^{-1}} = \begin{pmatrix} \boldsymbol{R}^T & 0 \\ -\boldsymbol{R}^T\,[\boldsymbol{r}\times] & \boldsymbol{R}^T \end{pmatrix}. \tag{1.1.30}$$

Another important context where the operator Ad appears is the following. Assume that $\boldsymbol{g}(t) \in \mathrm{SE}(3)$ is a differentiable curve which satisfies $\frac{d}{dt}\boldsymbol{g}(t) = \boldsymbol{g}(t)\mathcal{T}\phi(t)$, as defined in (1.1.26), and that it factorizes as

$$\boldsymbol{g}(t) = \boldsymbol{g}_1(t)\boldsymbol{g}_2(t) \tag{1.1.31a}$$

with

$$\frac{d}{dt}\boldsymbol{g}_1(t) = \boldsymbol{g}_1(t)\mathcal{T}\phi_1(t) \text{ and } \frac{d}{dt}\boldsymbol{g}_2(t) = \boldsymbol{g}_2(t)\mathcal{T}\phi_2(t) \tag{1.1.31b}$$

for $\phi_1(t), \phi_2(t) \in \mathbb{R}^6$. One then has the relation

$$\phi(t) = \mathrm{Ad}_{\boldsymbol{g}_2(t)}^{-1}\phi_1(t) + \phi_2(t) \tag{1.1.32}$$

where the matrix $\mathrm{Ad}_{\boldsymbol{g}_2(t)}^{-1}$ is explicitly defined in (1.1.30). Observe that for a differentiable curve $\boldsymbol{g}(t)$ which satisfies $\frac{d}{dt}\boldsymbol{g}(t) = \boldsymbol{g}(t)\mathcal{T}\phi(t)$ the operator $\mathrm{Ad}_{\boldsymbol{g}(t)}$ satisfies the ODE

$$\frac{d}{dt}\mathrm{Ad}_{\boldsymbol{g}(t)} = \mathrm{Ad}_{\boldsymbol{g}(t)}\mathrm{ad}_{\phi(t)} \tag{1.1.33}$$

where we have used the definition

$$\mathrm{ad}_{\phi(t)} = \begin{pmatrix} \left[\phi^{\boldsymbol{R}}(t)\times\right] & 0 \\ \left[\phi^{\boldsymbol{r}}(t)\times\right] & \left[\phi^{\boldsymbol{R}}(t)\times\right] \end{pmatrix} \text{ where } \phi(t) = (\phi^{\boldsymbol{R}}(t), \phi^{\boldsymbol{r}}(t)) \tag{1.1.34}$$

for all t.

1.2 Calculus on Matrix Groups

The basic definitions on the special orthogonal group SO(3) and on the special Euclidean group SE(3) have been given in section 1.1. In this section we want to present the fundamental expressions of the calculus of variations on the group SE(3). Any property then applies also to SO(3) through the natural embedding $\mathrm{SO}(3) \subset \mathrm{SE}(3)$.

Let $F : \mathrm{SE}(3) \to \mathbb{R}$ be a differentiable function on the special Euclidean group and let $\boldsymbol{g} = (\boldsymbol{R}, \boldsymbol{r}) \in \mathrm{SE}(3)$ be a given rigid body displacement. One can think, for instance, that $F(\boldsymbol{g})$ describes the potential energy between two rigid bodies in Euclidean space as a function of their relative rigid body displacement $\boldsymbol{g}$. A *rigid body perturbation* of the configuration $\boldsymbol{g}$ is denoted by $\delta\boldsymbol{g}$ and is defined as

$$\delta\boldsymbol{g} = \frac{d}{dt}_{|t=0} \boldsymbol{g}(t) = (\mathcal{T}\boldsymbol{\phi})\boldsymbol{g} \tag{1.2.1}$$

for a given $\phi \in \mathbb{R}^6$ according to (1.1.26) and where $\boldsymbol{g}(t)$ stands for a smooth curve in SE(3) with $\boldsymbol{g}(0) = \boldsymbol{g}$. Equivalently, in terms of the rotational part $\boldsymbol{R}$ and the translational part $\boldsymbol{r}$ one has

$$\delta\boldsymbol{R} = \left[\boldsymbol{\phi}^{\boldsymbol{R}}\times\right] \boldsymbol{R} \tag{1.2.2a}$$

and

$$\delta\boldsymbol{r} = \boldsymbol{\phi}^{\boldsymbol{R}} \times \boldsymbol{r} + \boldsymbol{\phi}^{\boldsymbol{r}} \tag{1.2.2b}$$

using (1.1.26a). A rigid body perturbation of the form $\delta\boldsymbol{g} = \boldsymbol{g}\mathcal{T}\phi$ can be used as well, and its explicit expression is given in (1.1.26b). More generally the first variation of the function $F(\boldsymbol{g})$ can be computed, using (1.1.23) and (1.2.1), as

$$\delta F(\boldsymbol{g})\delta\boldsymbol{g} = \frac{d}{dt}_{|t=0} F(\boldsymbol{g}(t)) = \mathcal{T}^* \left(\partial_{\boldsymbol{g}} F(\boldsymbol{g})\boldsymbol{g}^T\right) \cdot \boldsymbol{\phi} \tag{1.2.3}$$

for any rigid body perturbation $\delta\boldsymbol{g}$. In terms of the variations $\delta\boldsymbol{R}$ and $\delta\boldsymbol{r}$ in (1.2.2), one finds

$$\delta F(\boldsymbol{g})\delta\boldsymbol{g} = \left[\mathrm{Vect}\left(\partial_{\boldsymbol{R}} F(\boldsymbol{g})\boldsymbol{R}^T\right) + \boldsymbol{r} \times \partial_{\boldsymbol{r}} F(\boldsymbol{g})\right] \cdot \boldsymbol{\phi}^{\boldsymbol{R}} + \partial_{\boldsymbol{r}} F(\boldsymbol{g}) \cdot \boldsymbol{\phi}^{\boldsymbol{r}} \tag{1.2.4}$$

using the explicit definition of $\mathcal{T}^*$ in (1.1.24).

If a rigid body configuration $\boldsymbol{g}^0 \in \mathrm{SE}(3)$ satisfies

$$\delta F(\boldsymbol{g}^0)\delta\boldsymbol{g} = 0 \text{ for any } \delta\boldsymbol{g} \tag{1.2.5}$$

it is said to be a *stationary configuration* of the function F.

The second order perturbation of the configuration $\boldsymbol{g}$, denoted $\delta^2\boldsymbol{g}$, follows from the definition of the first variation (1.2.1) to be

$$\delta^2\boldsymbol{g} = (\mathcal{T}\boldsymbol{\phi}_1)(\mathcal{T}\boldsymbol{\phi}_2)\boldsymbol{g} \tag{1.2.6}$$

for $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2 \in \mathbb{R}^6$. One obtains then that, at the second order, neighbouring matrices of a given configuration $\bar{\boldsymbol{g}}$ can be expressed as

$$\boldsymbol{g} = \left[\mathrm{Id}_4 + \mathcal{T}\boldsymbol{\phi} + \frac{1}{2}\left(\mathcal{T}\boldsymbol{\phi}\right)^2\right] \bar{\boldsymbol{g}} + \mathrm{O}\left(|\boldsymbol{\phi}|^2\right). \tag{1.2.7}$$

Suppose now that $F : \mathrm{C}^1(]0, L[, \mathrm{SE}(3)) \to \mathbb{R}$ denotes a map defined on the space of differentiable curves on SE(3) and that it is defined as

$$F[\boldsymbol{g}] = \int_0^L \mathcal{L}(\boldsymbol{g}(s), \xi(s); s)ds \tag{1.2.8}$$

where $\xi(s)$ denotes the right generator vector of the curve $\boldsymbol{g}(s)$ as defined in (1.1.26b), i.e.

$$\frac{d}{ds}\boldsymbol{g}(s) = \boldsymbol{g}(s)\mathcal{T}\xi(s). \tag{1.2.9}$$

The mapping F in (1.2.8) said to be a *functional*. The perturbations of a configuration $\boldsymbol{g}(s)$ are then of the form

$$\delta\boldsymbol{g}(s) = (\mathcal{T}\boldsymbol{\phi}(s))\boldsymbol{g}(s) \tag{1.2.10}$$

according to (1.2.1). Using the form of the variations (1.2.9) and (1.2.10) one deduces that the perturbation of the generator vector $\xi(s)$ satisfies

$$\delta\xi(s) = \mathrm{Ad}^{-1}_{\boldsymbol{g}(s)}\left(\frac{d}{ds}\boldsymbol{\phi}(s)\right) \tag{1.2.11}$$

where we have used (1.1.27) and (1.1.28). The perturbations (1.2.10) and (1.2.11) allow the first variation of the functional F in (1.2.8) to be written as

$$\delta F[\boldsymbol{g}]\delta\boldsymbol{g} = \int_0^L \left[\mathcal{T}^*\left(\partial_{\boldsymbol{g}}\mathcal{L}\boldsymbol{g}^T\right) - \frac{d}{ds}\left(\mathrm{Ad}^{-T}_{\boldsymbol{g}(s)}\partial_\xi\mathcal{L}\right)\right]\cdot\boldsymbol{\phi}(s)ds + \mathrm{Ad}^{-T}_{\boldsymbol{g}(s)}\partial_\xi\mathcal{L}\cdot\boldsymbol{\phi}(s)|_0^L \tag{1.2.12}$$

where the explicit dependence of the function $\mathcal{L}$ in the variables $(\boldsymbol{g}, \xi; s)$ has been hidden for compactness. Writing the perturbation $\boldsymbol{\phi}(s) = (\boldsymbol{\phi}^{\boldsymbol{R}}(s), \boldsymbol{\phi}^{\boldsymbol{r}}(s))$, and the generator vector $\xi(s) = (U(s), V(s))$ respectively in terms of their rotational and translational part, as in (1.1.26) leads to

$$\begin{aligned}
&\delta F[\boldsymbol{g}]\delta\boldsymbol{g} = \\
&\int_0^L \left[\mathrm{Vect}\left(\partial_{\boldsymbol{R}}\mathcal{L}(\boldsymbol{g})\boldsymbol{R}^T\right) + \boldsymbol{r}\times\partial_{\boldsymbol{r}}\mathcal{L}(\boldsymbol{g}) - \frac{d}{ds}\left(\boldsymbol{R}(s)\partial_U\mathcal{L} + \boldsymbol{r}(s)\times\boldsymbol{R}(s)\partial_V\mathcal{L}\right)\right]\cdot\boldsymbol{\phi}^{\boldsymbol{R}}(s)ds \\
&+\int_0^L \left[\partial_{\boldsymbol{r}}\mathcal{L}(\boldsymbol{g}) - \frac{d}{ds}\left(\boldsymbol{R}(s)\partial_V\mathcal{L}\right)\right]\cdot\boldsymbol{\phi}^{\boldsymbol{r}}(s)ds \\
&+\left(\boldsymbol{R}(s)\partial_U\mathcal{L} + \boldsymbol{r}(s)\times\boldsymbol{R}(s)\partial_V\mathcal{L}\right)\cdot\boldsymbol{\phi}^{\boldsymbol{R}}(s)|_0^L + \left(\boldsymbol{R}(s)\partial_V\mathcal{L})\right]\cdot\boldsymbol{\phi}^{\boldsymbol{r}}(s)|_0^L.
\end{aligned} \tag{1.2.13}$$

The variation of the functional F in the form of (1.2.12), or (1.2.13), is often called the *Euler-Poincaré variation* in reference to the pioneering works of L. Euler on rigid body dynamics [Euler, 1765] and of H. Poincaré [Poincaré, 1901] on calculus on a smooth group. They are a natural generalisation of the *Euler-Lagrange variations*. Modern

treatment of such matrix calculus can be found in [Marsden and Ratiu, 2002, Ellis et al., 2010, Chirikjian, 2011] for instance. A curve $\boldsymbol{g}^0 \in \mathrm{C}^1(]0, L[, \mathrm{SE}(3))$ is then said to a be *stationary configuration* of the functional F if

$$\delta F[\boldsymbol{g}^0]\delta\boldsymbol{g} = 0 \text{ for any } \delta\boldsymbol{g}(s) \tag{1.2.14}$$

analogously to (1.2.5).

1.3 Coordinate Systems for Matrix Groups

Both the group of special orthogonal matrices $\mathrm{SO}(3) \subset \mathbb{R}^{3\times3}$ and the special euclidean group $\mathrm{SE}(3) \subset \mathbb{R}^{4\times4}$ have the features that their dimension as a smooth manifold, namely three for SO(3) and six for SE(3), is strictly less than the space of matrices where they are naturally embedded, and that they have non-zero Gaussian curvature. In many applications it is then convenient to make use of a *coordinate system* to described these matrix groups. However, there exist many different coordinate system for the rotation group SO(3) and for the group of rigid body motion SE(3), each of which has been used in different physical context and by different scientific communities.

A coordinate system on the matrix group SO(3) is defined as a differentiable mapping of the form

$$u \mapsto Q(u) \in \mathrm{SO}(3) \tag{1.3.1}$$

and similarly a coordinate system on SE(3) is of the form

$$x \mapsto a(x) \in \mathrm{SE}(3). \tag{1.3.2}$$

In the context of DNA modelling, one finds a version of the *Euler angles* for the rotations [Marky and Olson, 1994, Olson et al., 1998, Coleman et al., 2003], a version of the *exponential coordinates* [Becker and Everaers, 2007, Chirikjian, 2011] for the full rigid body motion, the use of the *Cayley vector* and the introduction of a *junction frame* to describe the translation [Lankaš et al., 2009, Walter et al., 2010, Gonzalez et al., 2013, Petkeviciute et al., 2014] or the use of the *unit quaternions* for large rotations [Dichmann et al., 1996, Manning et al., 1996, Manning et al., 1998, Chirikjian, 2010].

In all cases the matrix configuration can be expanded in terms of the internal coordinates, and since it belongs to a matrix group, these expansions can be written in a specific form. For instance in the case of (1.3.2), a second order expansion of the matrix $a(x)$ around the coordinates $\bar{x}$ is

$$a(x) = \left[\mathrm{Id}_4 + a^{(1)}_{\bar{x}}(x-\bar{x}) + \frac{1}{2}a^{(2)}_{\bar{x}}(x-\bar{x})\right]\bar{a} + \mathrm{o}\left(|x-\bar{x}|^2\right) \tag{1.3.3}$$

where $a^{(1)}_{\bar{x}}(x-\bar{x})$ and $a^{(2)}_{\bar{x}}(x-\bar{x})$ stand respectively for the linear and quadratic contribution of $(x-\bar{x})$ in this expansion and where by definition $\bar{a} = a(\bar{x})$. To make clearer the

connexion between the expansion in coordinates x and the one in terms of perturbations of the form obtained in (1.2.7)

$$a = \left[\mathrm{Id}_4 + \mathcal{T}\Theta + \frac{1}{2}(\mathcal{T}\Theta)^2\right] \bar{a} + \mathrm{o}\left(|\Theta|^2\right) \tag{1.3.4}$$

we assume that the vector $\Theta \in \mathbb{R}^6$ can be expressed at the second order as

$$\Theta(x) = \mathbb{L}_{\bar{x}}^{(1)}(x - \bar{x}) + \frac{1}{2}\mathbb{L}_{\bar{x}}^{(2)}(x - \bar{x}) + \mathrm{o}\left(|x - \bar{x}|^2\right) \tag{1.3.5}$$

where $\mathbb{L}_{\bar{x}}^{(1)}(x-\bar{x})$ and $\mathbb{L}_{\bar{x}}^{(2)}(x-\bar{x})$ denote respectively the linear and quadratic contribution of $(x-\bar{x})$. Consequently the relation between the terms in the expansion (1.3.3) and the one in (1.3.5) is

$$\mathcal{T}\mathbb{L}_{\bar{x}}^{(1)}(x - \bar{x}) = a_{\bar{x}}^{(1)}(x - \bar{x}) \tag{1.3.6a}$$

$$\mathcal{T}\mathbb{L}_{\bar{x}}^{(2)}(x - \bar{x}) = a_{\bar{x}}^{(2)}(x - \bar{x}) - \left(a_{\bar{x}}^{(1)}(x - \bar{x})\right)^2 \tag{1.3.6b}$$

in order to match the first order and second order terms in (1.3.4). For a given choice of coordinate system x the problem of computing the terms $a_{\bar{x}}^{(1)}(x - \bar{x})$ and $a_{\bar{x}}^{(2)}(x - \bar{x})$ is then equivalent to the one of computing the terms $\mathbb{L}_{\bar{x}}^{(1)}(x - \bar{x})$ and $\mathbb{L}_{\bar{x}}^{(2)}(x - \bar{x})$. Explicit computations of these terms is done in the section A.1 for the coordinates used in [Gonzalez et al., 2013, Petkeviciute et al., 2014], which are described in (2.1.18).

2. On Single and Double Rigid Body Chains

Let a Cartesian coordinate system
be attached to each bond of the chain
in some arbitrarily specified manner.

P. J. Flory, 1964

This second chapter is dedicated to a discussion of the properties of rigid body chains in classical and equilibrium statistical mechanics. Section 2.1 describes the notions of rigid body chain and bichain configurations and of internal coordinates. Material symmetries are then briefly discussed. The equilibrium conditions subject to end conditions and a related coordinate independent variational principle are then presented in section 2.2 for both the chain and the bichain models. The bichain formulation is original. We then describe how this constrained variational principle can be recast into an unconstrained variational principle using Lagrange multipliers, which in addition can be interpreted as the set of total couple and force. Section 2.3 presents the definition of a configurational distribution for chains and bichains in the context of equilibrium statistical mechanics. Observables and their expectations are then discussed. In particular, we emphasize the special role of the configurational first moment which comprises the frame correlation matrix and the Flory persistence vector. For chains with nearest neighbor internal energy, a detailed analysis of the configurational first moment is then discussed in section 2.4 and an explicit first order approximation is obtained for chains in a semi-flexible regime. These analytical expressions not only show the equal importance of the expected chain configuration and of the fluctuation matrix, but also motivate the definition of the persistence matrix which is shown to characterize the memory lost in the entries of the configurational first moment. We finally obtain a closed form expression in a short length expansion.

2.1 On Chain Configurations

We present in this section the the basic definitions related to rigid body chain models. Rigid body chain configurations for a linear polymer, together with the notion of internal coordinates, are formally introduced. The natural extension of this model to double rigid body chain and bichain descriptions for a double stranded molecule is then discussed in details. In particular, we introduce the notion of macrostructure and microstructure for bichains. We finally examine the transformation rules appearing under different natural material symmetries for chains.

2.1.1 On rigid body chain configurations and internal coordinates

A linear polymer is commonly modelled using a sequence of positions in Euclidean space where two successive elements are connected by a bond, see for instance [Kuhn, 1934, Flory, 1969, Schellman, 1974, Trifonov et al., 1988, Doi and Edwards, 1986, Öttinger, 1996]. The relative orientation of two consecutive bonds can be measured by the introduction of a frame in each unit which constitutes the chain [Eyring, 1932, Flory, 1964, Flory, 1969]. More generally, the sequence of positions and orientations allows to describe the *rigid body displacement* between any element in the chain [Flory, 1973, Maroun and Olson, 1988, Gonzalez and Maddocks, 2001, Becker and Everaers, 2007]. Largely inspired by the existing literature on polymer modelling, we introduce the definition of the *rigid body chain model* for linear polymers where each elementary unit is described as a rigid body configuration, i.e. by a position and an orientation. This model has been for instance extensively studied in [Becker, 2007].

Formally, a *rigid body chain configuration* is made of a sequence of N rigid body configurations $\boldsymbol{g}_1, ..., \boldsymbol{g}_N$ where $\boldsymbol{g}_n = (\boldsymbol{R}_n, \boldsymbol{r}_n) \in \mathrm{SE}(3)$ denotes the *rigid body configuration* of the n^{th} rigid body in the chain, the rotation matrix $\boldsymbol{R}_n \in \mathrm{SO}(3)$ denotes the *rigid body orientation* and the triple $\boldsymbol{r}_n \in \mathbb{R}^3$ the *rigid body position*. The rigid body chain configuration is then denoted as

$$\boldsymbol{g} = (\boldsymbol{g}_1, ..., \boldsymbol{g}_N) \in \mathrm{SE}(3)^N \tag{2.1.1}$$

where $\mathrm{SE}(3)^N$ stands for N copies of the special Euclidean group[1]. A sketch of a rigid body chain configuration is drawn in figure 2.1. In the continuity of several authors [Flory, 1973, Maroun and Olson, 1988, Becker and Everaers, 2007, Chirikjian, 2011] we use the following matrix formulation of the individual rigid body configuration $\boldsymbol{g}_n$

$$\boldsymbol{g}_n = \begin{pmatrix} \boldsymbol{R}_n & \boldsymbol{r}_n \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{4\times 4}. \tag{2.1.2}$$

The relative rotation between the orientation of the n^{th} body and the orientation $(n+1)^{th}$ in the chain is called the *junction rotation* and is denoted by $Q_n \in \mathrm{SO}(3)$, the

[1] See section 1.1 for more detail on the special Euclidean group

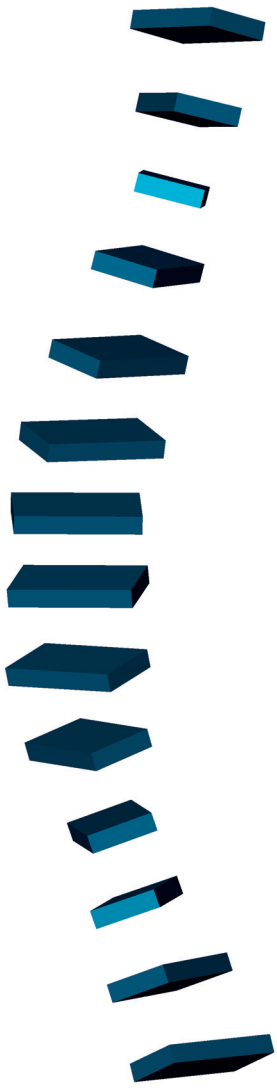


Figure 2.1: Sketch of a rigid body chain configuration

analogous translation is called the *junction translation* and denoted by $q_n \in \mathbb{R}^3$. The rigid body orientations $\boldsymbol{R}_n$ and positions $\boldsymbol{r}_n$ then satisfy the recurrence relation

$$\boldsymbol{R}_{n+1} = \boldsymbol{R}_n Q_n \tag{2.1.3a}$$

$$\boldsymbol{r}_{n+1} = \boldsymbol{r}_n + \boldsymbol{R}_n q_n \tag{2.1.3b}$$

with the first orientation matrix $\boldsymbol{R}_1$ and position vector $\boldsymbol{r}_1$ given. Using the matrix notation (2.1.2), expressions (2.1.3) can be simply summarised in the matrix recurrence relation

$$\boldsymbol{g}_{n+1} = \boldsymbol{g}_n a_n \tag{2.1.4}$$

with the first rigid body configuration $\boldsymbol{g}_1$ prescribed and where the block structure of the *junction displacement* $a_n \in \mathrm{SE}(3)$ is defined as

$$a_n = \begin{pmatrix} Q_n & q_n \\ 0 & 1 \end{pmatrix}. \tag{2.1.5}$$

More generally, the *relative rotation* between the m^{th} body in the chain and n^{th} one is denoted by $\boldsymbol{R}_{m,n} \in \mathrm{SO}(3)$ and the analogous *relative translation* by $\boldsymbol{r}_{m,n} \in \mathbb{R}^3$. They are defined as

$$\boldsymbol{R}_{m,n} = \boldsymbol{R}_m^T \boldsymbol{R}_n \tag{2.1.6a}$$

$$\boldsymbol{r}_{m,n} = \boldsymbol{R}_m^T (\boldsymbol{r}_n - \boldsymbol{r}_m) \tag{2.1.6b}$$

or, in matrix form, the *relative rigid body displacement* $\boldsymbol{g}_{m,n}$ is defined as

$$\boldsymbol{g}_{m,n} = \boldsymbol{g}_m^{-1} \boldsymbol{g}_n \tag{2.1.7}$$

with the block structure

$$\boldsymbol{g}_{m,n} = \begin{pmatrix} \boldsymbol{R}_{m,n} & \boldsymbol{r}_{m,n} \\ 0 & 1 \end{pmatrix} \tag{2.1.8}$$

analogously to (2.1.2). Note that a similar notation has been used previously in [Becker, 2007, Becker and Everaers, 2007]. Similarly to (2.1.4), the *relative rigid displacement* in the chain $\boldsymbol{g}_{m,n}$ can be described through the recurrence relation

$$\boldsymbol{g}_{m,n+1} = \boldsymbol{g}_{m,n} a_n \text{ with } \boldsymbol{g}_{m,m} = \mathrm{Id}_4 \tag{2.1.9}$$

for all $1 \leq m \leq n \leq N$. Observe that definition (2.1.7) implies the *transition identity*

$$\boldsymbol{g}_{m,n} = \boldsymbol{g}_{m,k} \boldsymbol{g}_{k,n} \tag{2.1.10}$$

for all $1 \leq m \leq k \leq n \leq N$.

As described in section 1.3, it is often convenient to describe rigid body displacements in terms of a coordinate system. The junction displacement $a_n \in \mathrm{SE}(3)$ in (2.1.5) can then be parametrized with a minimal set of coordinates $x_n \in \mathbb{R}^6$ as

$$a_n = a(x_n). \tag{2.1.11}$$

The coordinates x_n are called *local internal coordinates* for the junction displacement a_n and it is assumed that the coordinates x_n naturally split into

$$x_n = (u_n, v_n) \tag{2.1.12}$$

where the triple $u_n \in \mathbb{R}^3$ stands for the *local rotational coordinates* and the triple $v_n \in \mathbb{R}^3$ for the *local translational coordinates*. The list

$$x = (x_1, x_2, ..., x_{N-1}) \in \mathbb{R}^{6(N-1)} \tag{2.1.13}$$

are a set of *internal coordinates of the rigid body chain* $\boldsymbol{g}$. Conversely, using the parametrization in (2.1.11) and the recurrence relation in (2.1.4), the set of internal coordinates x defines a rigid body chain configuration through a *chain reconstruction rule*

$$x \mapsto \boldsymbol{g}(x) \tag{2.1.14}$$

or more generally, using the definition (2.1.7), we obtain a mapping of the form

$$x \mapsto \boldsymbol{g}_{m,n}(x) \tag{2.1.15}$$

for any relative rigid body displacement in the chain.

Concerning the choice of parametrization of the junction displacement $a_n = a(x_n)$ in (2.1.11), many versions appear in the literature, such as, in the context of DNA modelling, a version of *Euler angles* [Marky and Olson, 1994, Olson et al., 1998, Coleman et al., 2003], a version of *exponential coordinates* [Becker and Everaers, 2007, Chirikjian, 2011] or the use of the *Cayley vector* of the rotation matrix Q_n and the introduction of a *junction frame* to write the component of the translation q_n [Lankaš et al., 2009, Walter et al., 2010, Gonzalez et al., 2013, Petkeviciute et al., 2014]. This last choice is of particular interest for us, since it satisfies convenient material symmetry relations[2], it allows efficient analytic computations[3] for chain configurational first moment and we possess realistic constitutive parameters in terms of this coordinate system for sequence dependent coarse grain DNA models [Petkeviciute et al., 2014]. Consequently, we need to briefly expose the Cayley parametrization of rotation matrices and the definition of the junction frame as used in [Lankaš et al., 2009, Gonzalez et al., 2013, Petkeviciute

[2]They are described in section 2.1.4.
[3]They are described in section 2.4.

et al., 2014]. Any rotation Q_n, which is not a rotation through an angle of π, can be parametrized by the one-to-one mapping

$$Q_n = Q(u_n) := \left(I + \frac{1}{2}\,[u_n\times]\right)\left(I - \frac{1}{2}\,[u_n\times]\right)^{-1} \tag{2.1.16}$$

where the operator $[\,.\,\times]$ is defined in (1.1.9). The triple u_n in (2.1.16) is called the *Cayley vector* of the rotation matrix Q_n. The junction translation q_n is parametrized as

$$q_n = Q^{\frac{1}{2}}(u_n)v_n \tag{2.1.17}$$

where $Q^{\frac{1}{2}}(u_n)$ denotes the (principal) square root of the rotation matrix $Q(u_n)$ and $v_n \in \mathbb{R}^3$. Observe that the rotation $Q^{\frac{1}{2}}(u_n)$ is the midway rotation between the identity matrix and the rotation matrix $Q(u_n)$. In matrix form, the junction displacement a_n, as defined in (2.1.5) and with the parametrization (2.1.16) and (2.1.17), is then of the form

$$a(x_n) = \begin{pmatrix} Q(u_n) & Q^{\frac{1}{2}}(u_n)v_n \\ 0 & 1 \end{pmatrix} \tag{2.1.18}$$

where $x_n = (u_n, v_n)$.

2.1.2 On rigid body double chain configurations

By observing more closely polymeric structures, there exist some examples, such as DNA molecules for instance, where a finer description can be achieved using multi-stranded rigid body chains. Even if a general theory for such composed object is of interest, we have restrained our study to the case of two interacting rigid body chains since we shall concentrate on the case of DNA modelling.

A *rigid body double chain configuration* $(\boldsymbol{g}^+, \boldsymbol{g}^-)$ can be described as a pair of rigid body chain configurations, denoted respectively $\boldsymbol{g}^+$ and $\boldsymbol{g}^-$, where both chains are exactly made of N rigid bodies. Formally we have

$$(\boldsymbol{g}^+, \boldsymbol{g}^-) = (\boldsymbol{g}_1^+, \boldsymbol{g}_1^-, ..., \boldsymbol{g}_N^+, \boldsymbol{g}_N^-) \in \mathrm{SE}(3)^{2N} \tag{2.1.19}$$

where $\mathrm{SE}(3)^{2N}$ denotes $2N$ copies of the special Euclidean group, as in (2.1.1). A sketch of a rigid body double chain configuration is drawn in figure 2.2. Analogously to the single rigid body chain definition, a rigid body configuration in the double chain $(\boldsymbol{g}^+, \boldsymbol{g}^-)$ is of the form $\boldsymbol{g}_N^\pm = (\boldsymbol{R}_n^\pm, \boldsymbol{r}_n^\pm)$ where $\boldsymbol{R}_n^\pm$ denotes the orientation of the rigid body and $\boldsymbol{r}_n^\pm$ its position. The notation $\boldsymbol{g}^\pm$ is used to describe simultaneously both of the two chains $\boldsymbol{g}^+$ or $\boldsymbol{g}^-$ as in [Moakher and Maddocks, 2005], and each definition related to rigid body chains can be transposed to the chain $\boldsymbol{g}^\pm$ in an unambiguous way. For example, the relative rigid body displacement between the neighbouring rigid bodies $\boldsymbol{g}_n^\pm$ and $\boldsymbol{g}_{n+1}^\pm$ along the same strand is then denoted by $a_n^\pm$. Analogously to the junction displacement $a_n^\pm$, one can introduce the *intra rigid body displacement* $b_n \in \mathrm{SE}(3)$ defined as

$$b_n = (\boldsymbol{g}_n^+)^{-1}\boldsymbol{g}_n^- \tag{2.1.20}$$

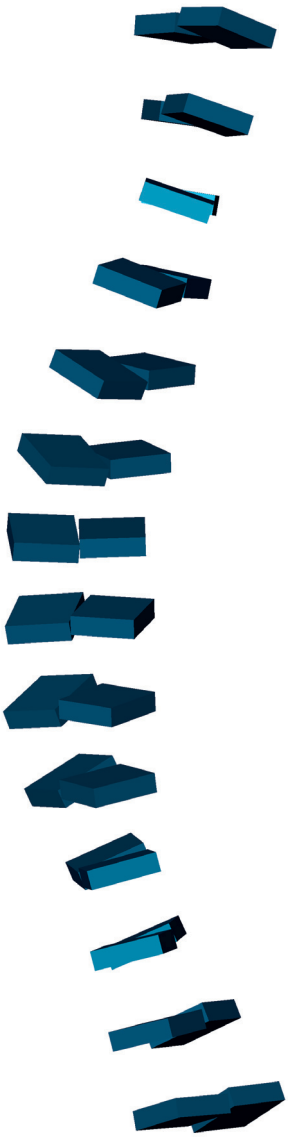


Figure 2.2: Sketch of a rigid body double chain configuration

whose block structure is denoted as

$$b_n = \begin{pmatrix} P_n & p_n \\ 0 & 1 \end{pmatrix}. \tag{2.1.21}$$

The double chain configuration satisfies then the relations

$$\boldsymbol{g}^{\pm}_{n+1} = \boldsymbol{g}^{\pm}_n a^{\pm}_n \tag{2.1.22a}$$

for all $n = 1, ..., N-1$, as discussed in (2.1.4), and

$$\boldsymbol{g}^{-}_n = \boldsymbol{g}^{+}_n b_n \tag{2.1.22b}$$

for all $n = 1, ..., N$.

2.1.3 On rigid body bichain configurations and internal coordinates

The rigid body bichain model has been specifically designed to reflect DNA configuration at the level of individual bases [Lankaš et al., 2009, Gonzalez et al., 2013, Petkeviciute et al., 2014] and can be seen as a natural generalisation to the rigid base pair model for DNA [Olson et al., 1998, Gonzalez and Maddocks, 2001, Becker and Everaers, 2007]. More details about these models are presented in section 5.2. Rigid body bichains are essentially double chains but with a different point of view, but the bichain configurations often constitute a more convenient description.

A *rigid body bichain configuration* $(\boldsymbol{g}, \mathcal{P})$ can be described as a *macrostructure configuration* $\boldsymbol{g}$ together with a *microstructure configuration* $\mathcal{P}$, as in [Moakher and Maddocks, 2005]. The macrostructure configuration is a rigid body chain configuration which is defined to represent the configuration of the double chain in a large scale description, whereas the microstructure configuration encodes the local features of the double chain configuration which allow recovery of each actual chain configuration $\boldsymbol{g}^{\pm}$ from the macrostructure $\boldsymbol{g}$. For DNA models, the macrostructure $\boldsymbol{g}$ is meant to describe the configuration at the level of base pair, whereas the microstructure $\mathcal{P}$ encodes the respective base position *in* the base pairs [Lankaš et al., 2009, Gonzalez et al., 2013, Petkeviciute et al., 2014]. Formally one has then an invertible mapping

$$(\boldsymbol{g}^{+}, \boldsymbol{g}^{-}) \leftrightarrow (\boldsymbol{g}, \mathcal{P}) \tag{2.1.23}$$

between double chain and bichain configurations.

Motivated by previous work on rigid base models for DNA molecules [Lankaš et al., 2009, Gonzalez et al., 2013, Petkeviciute et al., 2014], the chain macrostructure configuration

$$\boldsymbol{g} = (\boldsymbol{g}_1, ..., \boldsymbol{g}_N) \in \mathrm{SE}(3)^N \tag{2.1.24}$$

is defined by $\boldsymbol{g}_n = (\boldsymbol{R}_n, \boldsymbol{r}_n) \in \mathrm{SE}(3)$ where

$$\boldsymbol{R}_n = \boldsymbol{R}^{+}_n \left(\boldsymbol{R}^{+T}_n \boldsymbol{R}^{-}_n \right)^{\frac{1}{2}} \tag{2.1.25a}$$

and

$$\boldsymbol{r}_n = \frac{1}{2}\left(\boldsymbol{r}_n^+ + \boldsymbol{r}_n^-\right) \tag{2.1.25b}$$

for $n = 1, ..., N$. The orientation $\boldsymbol{R}_n$ is the average in the rotation group SO(3) between the orientation $\boldsymbol{R}_n^+$ and $\boldsymbol{R}_n^-$, and the position $\boldsymbol{r}_n$ is the average point in Euclidean space $\mathbb{R}^3$ between the position $\boldsymbol{r}_n^+$ and $\boldsymbol{r}_n^-$. In a complementary way, the microstructure configuration

$$\mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_N) \in \left(\mathrm{SO}(3) \times \mathbb{R}^3\right)^N \tag{2.1.26}$$

is defined by $\mathcal{P}_n = (P_n, \mathsf{w}_n) \in \mathrm{SO}(3) \times \mathbb{R}^3$ where

$$P_n = \boldsymbol{R}_n^{+T} \boldsymbol{R}_n^- \tag{2.1.27a}$$

describes the relative rotation between the rigid body $\boldsymbol{g}_n^+$ and $\boldsymbol{g}_n^-$, and where

$$\mathsf{w}_n = \boldsymbol{R}_n^T \left(\boldsymbol{r}_n^- - \boldsymbol{r}_n^+\right) \tag{2.1.27b}$$

describes the analogous relative translation. According to the definitions in (2.1.25) and in (2.1.27), the double chain configuration can be recovered from its associated bichain configuration by

$$\boldsymbol{R}_n^\pm = \boldsymbol{R}_n P_n^{\mp\frac{1}{2}} \tag{2.1.28a}$$

and

$$\boldsymbol{r}_n^\pm = \boldsymbol{r}_n \mp \frac{1}{2}\boldsymbol{R}_n \mathsf{w}_n \tag{2.1.28b}$$

Concerning the rotational and translational part of the local junction displacements $a_n^\pm = (Q_n^\pm, q_n^\pm)$ we then obtain the relations

$$Q_n^\pm = P_n^{\pm\frac{1}{2}} Q_n P_{n+1}^{\mp\frac{1}{2}} \tag{2.1.29a}$$

and

$$q_n^\pm = P_n^{\pm\frac{1}{2}} \left[q_n \mp \frac{1}{2}\left(Q_n \mathsf{w}_{n+1} - \mathsf{w}_n\right)\right]. \tag{2.1.29b}$$

A bichain configuration can consequently be summarised as the set of local macrostructure and microstructure configurations

$$(\boldsymbol{g}, \mathcal{P}) = (\boldsymbol{g}_1, \mathcal{P}_1, ..., \boldsymbol{g}_N, \mathcal{P}_N) \tag{2.1.30}$$

following the natural order in the chain constitution.

The *bichain internal coordinates* therefore comprise a set of local internal chain coordinate $x = (x_1, ..., x_{N-1}) \in \mathbb{R}^{6(N-1)}$, called *inter coordinates*, for the macrostructure $\boldsymbol{g}$ and a set of local coordinates $y = (y_1, ..., y_N) \in \mathbb{R}^{6N}$, called *intra coordinates*, to parametrize the microstructure $\mathcal{P}$. Precisely, with the application to DNA modelling in mind, we shall follow the definitions of previous works [Lankaš et al., 2009, Gonzalez et al., 2013, Petkeviciute et al., 2014]. The macrostructure configuration $\boldsymbol{g}$ is written in terms of its junction displacements $a_1, ..., a_{N-1}$, similarly to (2.1.4), whereas the microstructure configuration $\mathcal{P}$ is written in terms of the intra displacements $b_1, ..., b_N$, as defined in (2.1.20). The bichain internal coordinates read then

$$(x, y) = (y_1, x_1, y_2, ..., x_{N-1}, y_N) \in \mathbb{R}^{12N-6} \tag{2.1.31}$$

following the ordering in (2.1.30) and are chosen such that

$$\boldsymbol{g}_{n+1} = \boldsymbol{g}_n a(x_n) \text{ and } b_n = b(y_n). \tag{2.1.32}$$

Even if, as in the single rigid body chain case, there exist many bichain internal coordinates (x, y) satisfying (2.1.32), a convenient choice is given by $a(x_n)$ in (2.1.18) and

$$b(y_n) = \begin{pmatrix} P(\eta_n) & P^{\frac{1}{2}}(\eta_n)\mathsf{w}_n \\ 0 & 1 \end{pmatrix} \tag{2.1.33}$$

with $y_n = (\eta_n, \mathsf{w}_n)$ and where η_n denotes the Cayley vector of the rotation matrix P_n, as defined in (2.1.16) and in (2.1.21) respectively [Lankaš et al., 2009, Gonzalez et al., 2013, Petkeviciute et al., 2014]. One obtains then formally the following *bichain reconstruction rule*

$$(x, y) \mapsto (\boldsymbol{g}, \mathcal{P})(x, y). \tag{2.1.34}$$

2.1.4 Material framings and transformation rules

Given a physical fragment, either single or double stranded, it is crucial to investigate whether the notion of configuration and coordinates are ambiguous or not, and, moreover to know what are the transformation rules under different *material framings* of the same physical object.

We first emphasise that both the chain and bichain internal coordinates respect *invariance under overall rigid body motion.* Since the local junction displacements a_n in the rigid body chain have the property that

$$(a_1, ..., a_{N-1}) \mapsto (a_1, ..., a_{N-1}) \text{ when } (\boldsymbol{g}_1, ..., \boldsymbol{g}_N) \mapsto (B\boldsymbol{g}_1, ..., B\boldsymbol{g}_N) \tag{2.1.35}$$

for any $B \in \mathrm{SE}(3)$ according to their definition (2.1.4), then the chain internal coordinates transform as

$$(x_1, ..., x_{N-1}) \mapsto (x_1, ..., x_{N-1}) \text{ when } (\boldsymbol{g}_1, ..., \boldsymbol{g}_N) \mapsto (B\boldsymbol{g}_1, ..., B\boldsymbol{g}_N) \tag{2.1.36}$$

for any $B \in \mathrm{SE}(3)$. Similarly, the local displacements in the macrostructure and the microstructure satisfy

$$\begin{aligned} (\mathcal{P}_1, a_1, ..., a_{N-1}, \mathcal{P}_N) &\mapsto (\mathcal{P}_1, a_1, ..., a_{N-1}, \mathcal{P}_N) \\ \text{when } (\boldsymbol{g}_1^+, \boldsymbol{g}_1^-, ..., \boldsymbol{g}_N^+, \boldsymbol{g}_N^-) &\mapsto (B\boldsymbol{g}_1^+, B\boldsymbol{g}_1^-, ..., B\boldsymbol{g}_N^+, B\boldsymbol{g}_N^-) \end{aligned} \tag{2.1.37}$$

for any $B \in \mathrm{SE}(3)$ which implies

$$\begin{aligned} (y_1, x_1, ..., x_{N-1}, y_N) &\mapsto (y_1, x_1, ..., x_{N-1}, y_N) \\ \text{when } (\boldsymbol{g}_1^+, \boldsymbol{g}_1^-, ..., \boldsymbol{g}_N^+, \boldsymbol{g}_N^-) &\mapsto (B\boldsymbol{g}_1^+, B\boldsymbol{g}_1^-, ..., B\boldsymbol{g}_N^+, B\boldsymbol{g}_N^-) \end{aligned} \tag{2.1.38}$$

for any $B \in \mathrm{SE}(3)$.

For a given physical chain and a definition of its chain configuration $\boldsymbol{g} = (\boldsymbol{g}_1, ..., \boldsymbol{g}_N)$, one can redefine locally the description of orientation and the position of the n^{th} rigid body without changing its physical location. This transformation is called *local change of framing* and can be formally stated as $\boldsymbol{g}_n \mapsto \boldsymbol{g}_n \bar{\boldsymbol{g}}_n^{-1}$ for all $n = 1, ..., N$. For a single chain the local junction displacement then transforms as

$$a_n \mapsto \bar{\boldsymbol{g}}_n \left(a_n \bar{a}_n^{-1}\right) \bar{\boldsymbol{g}}_n^{-1} \text{ when } \boldsymbol{g}_n \mapsto \boldsymbol{g}_n \bar{\boldsymbol{g}}_n^{-1} \tag{2.1.39}$$

where $\bar{\boldsymbol{g}}_{n+1} = \bar{\boldsymbol{g}}_n \bar{a}_n$. For a double chain, if the local change of framing is uniform, i.e. if $\boldsymbol{g}_n^{\pm} \mapsto \boldsymbol{g}_n^{\pm} \bar{\boldsymbol{g}}^{-1}$, then the bichain macrostructure and microstructure transform as

$$\begin{aligned} a_n &\mapsto \bar{\boldsymbol{g}} a_n \bar{\boldsymbol{g}}^{-1} \\ \text{and } (P_n, \mathsf{w}_n) &\mapsto (\bar{\boldsymbol{R}} P_n \bar{\boldsymbol{R}}^T, \bar{\boldsymbol{R}} \mathsf{w}_n - \bar{\boldsymbol{R}}(P_n^{\frac{1}{2}} - P_n^{\frac{T}{2}}) \bar{\boldsymbol{R}}^T \bar{\boldsymbol{r}}) \\ &\text{when } \boldsymbol{g}_n^{\pm} \mapsto \boldsymbol{g}_n^{\pm} \bar{\boldsymbol{g}}_n^{-1} \end{aligned} \tag{2.1.40}$$

where $\mathcal{P}_n = (P_n, \mathsf{w}_n)$ and where $\bar{\boldsymbol{R}} \in \mathrm{SO}(3)$ denotes the rotational part of rigid body displacement $\bar{\boldsymbol{g}}$ and $\bar{\boldsymbol{r}} \in \mathbb{R}^3$ its translational part. A transformation rule can also be explicitly stated for non-uniform local change of framing of double chain configuration, but it is more complicated and not of use in this thesis. For internal bichain coordinates presented in (2.1.18) and (2.1.33), one has then

$$\begin{aligned} (u_n, v_n) &\mapsto (\bar{\boldsymbol{R}} u_n, \bar{\boldsymbol{R}} v_n - \bar{\boldsymbol{R}}(Q_n^{\frac{1}{2}} - Q_n^{\frac{T}{2}}) \bar{\boldsymbol{R}}^T \bar{\boldsymbol{r}}) \\ \text{and } (\eta_n, \mathsf{w}_n) &\mapsto (\bar{\boldsymbol{R}} \eta_n, \bar{\boldsymbol{R}} \mathsf{w}_n - \bar{\boldsymbol{R}}(P_n^{\frac{1}{2}} - P_n^{\frac{T}{2}}) \bar{\boldsymbol{R}}^T \bar{\boldsymbol{r}}) \\ &\text{when } \boldsymbol{g}_n^{\pm} \mapsto \boldsymbol{g}_n^{\pm} \bar{\boldsymbol{g}}_n^{-1}. \end{aligned} \tag{2.1.41}$$

For single chains, another way of changing the material framing is the *downstream indexing* of the chain element which corresponds to switching the role of the two chain ends. The junction displacement a_n transforms then as

$$a_n \mapsto a_{N-n}^{-1} \text{ when } (\boldsymbol{g}_1, ..., \boldsymbol{g}_N) \mapsto (\boldsymbol{g}_N, ..., \boldsymbol{g}_1) \tag{2.1.42a}$$

and the internal coordinates x_n as

$$x_n \mapsto -x_{N-n} \text{ when } (\boldsymbol{g}_1, ..., \boldsymbol{g}_N) \mapsto (\boldsymbol{g}_N, ..., \boldsymbol{g}_1) \tag{2.1.42b}$$

in the case of the choice described in (2.1.18). For a bichain, downstream indexing gives

$$a_n \mapsto a_{N-n}^{-1} \text{ and } b_n \mapsto b_{N-n} \text{ when } (\mathcal{P}_1, \boldsymbol{g}_1, ..., \boldsymbol{g}_N, \mathcal{P}_N) \mapsto (\mathcal{P}_N, \boldsymbol{g}_N, ..., \boldsymbol{g}_1, \mathcal{P}_1) \tag{2.1.43a}$$

and consequently for the internal coordinates

$$x_n \mapsto -x_{N-n} \text{ and } y_n \mapsto y_{N-n} \text{ when } (\mathcal{P}_1, \boldsymbol{g}_1, ..., \boldsymbol{g}_N, \mathcal{P}_N) \mapsto (\mathcal{P}_N, \boldsymbol{g}_N, ..., \boldsymbol{g}_1, \mathcal{P}_1). \tag{2.1.43b}$$

For bichains one can moreover switch the role of the chain denoted by $+$ and $-$, this is called *change of reference strand*. The junction displacement a_n and the intra displacement in the bichain transform then as

$$a_n \mapsto a_n \text{ and } b_n \mapsto b_n^{-1} \text{ when } \boldsymbol{g}^{\pm} \mapsto \boldsymbol{g}^{\mp} \tag{2.1.44a}$$

whereas the microstructure transforms as

$$\mathcal{P}_n = (P_n, \mathsf{w}_n) \mapsto (P_n^T, -\mathsf{w}_n) \text{ when } \boldsymbol{g}^{\pm} \mapsto \boldsymbol{g}^{\mp} \tag{2.1.44b}$$

and then the internal coordinates x_n and y_n transform as

$$x_n \mapsto x_n \text{ and } y_n \mapsto -y_n \text{ when } \boldsymbol{g}^{\pm} \mapsto \boldsymbol{g}^{\pm} \tag{2.1.44c}$$

in the case of the choice described in (2.1.33).

However, in DNA modelling, the natural change of material framing which arises is not change of reference strand, but the *reading symmetry*. This symmetry is due to the fact that a DNA molecule is an anti-symmetric object in the sense that reading the strand $\boldsymbol{g}^+$ upstream corresponds to read the strand $\boldsymbol{g}^-$ downstream according to its chemical structure. The reading symmetry is then the composition of the local change of framing (with $\bar{\boldsymbol{r}} = 0$) (2.1.40), the downstream indexing (2.1.43) and the change of reference strand (2.1.44). Formally, this change of framing involves the transformation $(\boldsymbol{g}^+, \boldsymbol{g}^-) \mapsto (\boldsymbol{g}^{+\prime}, \boldsymbol{g}^{-\prime})$ where by definition

$$\boldsymbol{g}^{\pm\prime} = (\boldsymbol{g}_N^{\mp}, \boldsymbol{g}_{N-1}^{\mp}, ..., \boldsymbol{g}_1^{\mp}). \tag{2.1.45}$$

As shown in [Gonzalez et al., 2013], the bichain internal coordinates (x, y) defined respectively in (2.1.18) and in (2.1.33) then transform under the reading symmetry as

$$(x, y) \mapsto (\mathbb{E}_2 y_N, \mathbb{E}_2 x_N, ..., \mathbb{E}_2 x_1, \mathbb{E}_2 y_1) \text{ when } (\boldsymbol{g}^+, \boldsymbol{g}^-) \mapsto (\boldsymbol{g}^{+\prime}, \boldsymbol{g}^{-\prime}) \tag{2.1.46}$$

where $\mathbb{E}_2$ stands for the diagonal matrix $\mathrm{diag}(-1, 1, 1, -1, 1, 1)$.

2.2 On Equilibrium Conditions and Variational Principles

This section discusses the notions of couple and force, as well as constitutive relations, for chains and the definition of equilibrium conditions. Starting with a single chain, we show how equilibrium configurations can be seen as stationary configurations of a internal energy with respect to end conditions, with or without the introduction of internal coordinates. The argument is then extended to the cases of double chains and bichains. The precise relations between the constitutive relations in the double chain and bichain descriptions are presented and lead to the definition of the bichain equilibrium conditions. It is then shown how the end conditions can explicitly be incorporated in these variational principles with a Lagrange multiplier formulation. Moreover, the Lagrange multipliers are explicitly recognised as the couple and force needed to obtain the desired end conditions. Finally, we present a very short discussion on the relation between local bichain and non-local chain theory.

2.2.1 On rigid body chain equilibrium conditions and constitutive relations

The physical configuration of a rigid body chain is modified by the application of local couples and forces. More precisely, each rigid body $\boldsymbol{g}_n$ in the chain feels the action of *internal couples and forces* coming from the internal deformation of the entire chain, as well as the action of *external couples and forces* coming from the interaction of the chain with an external field. A rigid body chain is said to be in an *equilibrium configuration* if the total couple and force on each rigid body which constitutes it vanish. Formally, if we denote by $\mathbf{m}_n \in \mathbb{R}^3$ the total internal couple around the point $\boldsymbol{r}_n$ acting on the n^{th} rigid body in the chain from the $(n-1)^{th}$, and by $\mathbf{n}_n \in \mathbb{R}^3$ the analogous total force, the *chain equilibrium conditions* can be stated as

$$-\mathbf{m}_{n+1} + \mathbf{m}_n - \boldsymbol{r}_{n+1} \times \mathbf{n}_{n+1} + \boldsymbol{r}_n \times \mathbf{n}_n + \mathbf{c}_n + \boldsymbol{r}_n \times \mathbf{f}_n = 0 \tag{2.2.1a}$$

$$-\mathbf{n}_{n+1} + \mathbf{n}_n + \mathbf{f}_n = 0 \tag{2.2.1b}$$

for $n = 2, ..., N-1$ and where $\mathbf{c}_n \in \mathbb{R}^3$ and $\mathbf{f}_n \in \mathbb{R}^3$ denote respectively the total external couple (around the point $\boldsymbol{r}_n$) and the total external force on the n^{th} rigid body. Conditions (2.2.1) are, for instance, deduced from the balance of a single rigid body in an external field . The chain is said to be an *isolated rigid body chain* if there is no external contributions to the total couples and forces acting on the chain, i.e. if $\mathbf{c}_n = 0$ and $\mathbf{f}_n = 0$. The chain equilibrium conditions in (2.2.1) only provide conditions for rigid bodies in the interior of the chain and one needs therefore to introduce *chain end conditions*. In this thesis, the conditions of interest are

i. *Dirichlet conditions*: the orientation and the position of each end of the chain are prescribed, i.e.

$$\boldsymbol{g}_1 \text{ and } \boldsymbol{g}_N \text{ are prescribed.} \tag{2.2.2a}$$

ii. *Unstressed conditions*: the orientation and the position are free at each end of the chain, i.e.

$$-\mathbf{m}_2 + \mathbf{c}_1 = 0,\ -\mathbf{n}_2 + \mathbf{f}_1 = 0 \text{ and } \mathbf{m}_N + \mathbf{c}_N = 0,\ \mathbf{n}_{N-1} + \mathbf{f}_N = 0 \tag{2.2.2b}$$

For compactness we will denote generically by Γ the subset of chain configurations which satisfies one of the end conditions i. or ii., i.e.

$$\Gamma = \left\{ \boldsymbol{g} \in \mathrm{SE}(3)^N \,|\, \boldsymbol{g} \text{ satisfies i. or ii. } \right\} \tag{2.2.3}$$

The key relations to translate the chain balance laws (2.2.1) into a statement about the actual chain configuration are the *chain constitutive relations*. Formally, the constitutive relations for a rigid body chain can be stated as a mapping $\boldsymbol{g} \mapsto (\mathbf{m}, \mathbf{n}, \mathbf{c}, \mathbf{f})$ where $\mathbf{m} = (\mathbf{m}_1, ..., \mathbf{m}_{N-1})(\boldsymbol{g})$, etc. In particular, we can define the notion of *local chain constitutive relations* which means that the local internal couple and force in the chain are only functions of the local chain deformation, i.e.

$$\begin{pmatrix} \mathbf{m}_n \\ \mathbf{n}_n \end{pmatrix} = \boldsymbol{\Psi}_n^{(\mathbf{m},\mathbf{n})}(a_n) \text{ for all } n = 2, ..., N. \tag{2.2.4}$$

The constitutive relations reflect informations about the physical constitution of the chain. The construction of realistic constitutive relations for a given material is generally far from being an easy task, see for instance [Antman, 1995], and it is still at the heart of much research in material science. The *chain equilibrium conditions* then comprise the conditions (2.2.1) on a set of configurations (2.2.3) together with a given set of constitutive relations, of the form (2.2.4) for instance.

2.2.2 On rigid body chain internal energy and variational principle

The *chain internal energy* E describes the chain stored potential energy arising exclusively from internal deformations of the rigid body chain. The internal energy is said to be a *local energy*, or a *nearest-neighbour energy*, if the function $E : \mathrm{SE}(3)^N \to \mathbb{R}$ is of the form

$$E(\boldsymbol{g}) = \sum_{n=1}^{N-1} w_n(a_n) \tag{2.2.5}$$

where the chain configuration $\boldsymbol{g} = (\boldsymbol{g}_1, ..., \boldsymbol{g}_N)$ is defined in (2.1.1) and $a_n \in \mathrm{SE}(3)$ denotes the local junction displacement as defined in (2.1.4). Observe that the energy E is internal in the sense that it is invariant through global rigid body motion of the chain, as defined in (2.1.35), formally one has

$$E(B\boldsymbol{g}_1, ..., B\boldsymbol{g}_N) = E(\boldsymbol{g}_1, ..., \boldsymbol{g}_N) \tag{2.2.6}$$

for any $B \in \mathrm{SE}(3)$. We should emphasise that the nearest-neighbour dependence assumption of the chain internal energy is often violated for realistic descriptions of macromolecules and one needs to add longer range dependence in the chain [Flory, 1969, Lankaš

et al., 2009, Gonzalez et al., 2013]. However, the local energy expression in (2.2.5) is sufficiently general to embrace most of the commonly used chain models, for instance in DNA modelling [Olson et al., 1998, Becker and Everaers, 2007, Fathizadeh et al., 2013] or in macroscopic rigid body systems [Schiehlen et al., 2006, Chirikjian, 2011], and it is thought to be the necessary starting point to understand mechanical and statistical properties of rigid body chains subject to more sophisticated non-local internal energies.

One of the major interest of the chain internal energy E is that it provides a *variational principle for rigid body chains* in the sense that stationary configurations of the function E on the set of configurations Γ, as defined respectively in (1.2.5) and in (2.2.3), satisfies balance laws of the form (2.2.1) for an isolated chain together with constitutive relations of the form (2.2.4). We note that this result can be generalised to include the presence of an external energy. However, we have to stress that not all constitutive relations of the form (2.2.4) admit a variational principle, the class of chains which do admit an internal energy is referred to as *hyper-elastic chains*.

More precisely, suppose that Γ denotes the set of chain configuration where the first and last rigid body configuration in the chain are prescribed and suppose that the chain configuration $\boldsymbol{g}^0 = (\boldsymbol{g}_1^0, ..., \boldsymbol{g}_N^0) \in \Gamma$ makes the internal energy E stationary on the set Γ. Then by definition[4] it satisfies

$$\delta E(\boldsymbol{g}^0)\delta\boldsymbol{g} = 0 \tag{2.2.7a}$$

for all perturbation of the form $\delta\boldsymbol{g} = (\delta\boldsymbol{g}_1, ..., \delta\boldsymbol{g}_N)$ where

$$\delta\boldsymbol{g}_n = (\mathcal{T}\boldsymbol{\phi}_n)\boldsymbol{g}_n^0 \text{ for } n = 2, ..., N-1 \tag{2.2.7b}$$

and

$$\delta\boldsymbol{g}_1 = 0 \text{ and } \delta\boldsymbol{g}_N = 0 \tag{2.2.7c}$$

according to the specific form of the perturbation of rigid body configurations detailed in (1.2.1) and the specific definition of the set of configurations Γ. Furthermore, at any given chain configurations $\boldsymbol{g} \in \Gamma$ any perturbation $\delta\boldsymbol{g}$ is of the form (2.2.7b) and (2.2.7c) and consequently the perturbation of the local junction a_n displacement reads $\delta a_1 = (\boldsymbol{g}_1)^{-1}(\mathcal{T}\boldsymbol{\phi}_2)\boldsymbol{g}_2$,

$$\delta a_n = \boldsymbol{g}_n^{-1}(\mathcal{T}\boldsymbol{\phi}_{n+1} - \mathcal{T}\boldsymbol{\phi}_n)\boldsymbol{g}_{n+1} \tag{2.2.8}$$

for $n = 2, ..., N-2$ and $\delta a_{N-1} = -(\boldsymbol{g}_{N-1})^{-1}(\mathcal{T}\boldsymbol{\phi}_{N-1})\boldsymbol{g}_N$. Then, using the form of the internal energy E in (2.2.5), its first variation can be written as

$$\delta E(\boldsymbol{g})\delta\boldsymbol{g} = -\sum_{n=2}^{N-1} (\mathbf{m}_{n+1} - \mathbf{m}_n + \boldsymbol{r}_{n+1} \times \mathbf{n}_{n+1} - \boldsymbol{r}_n \times \mathbf{n}_n) \cdot \boldsymbol{\phi}_n^{\boldsymbol{R}} + (\mathbf{n}_{n+1} - \mathbf{n}_n) \cdot \boldsymbol{\phi}_n^{\boldsymbol{r}} \tag{2.2.9}$$

[4] See section 1.2 for more details on variations of functions on a matrix group.

for $\boldsymbol{\phi}_n = (\boldsymbol{\phi}_n^{\boldsymbol{R}}, \boldsymbol{\phi}_n^{\boldsymbol{r}}) \in \mathbb{R}^6$, where we have defined for $n = 1, ..., N-1$

$$\mathbf{m}_{n+1} = \boldsymbol{R}_n \mathrm{Vect}\left(\partial_{Q_n} w_n(a_n) Q_n^T\right) \tag{2.2.10a}$$

$$\mathbf{n}_{n+1} = \boldsymbol{R}_n \partial_{q_n} w_n(a_n) \tag{2.2.10b}$$

and where we recall that $Q_n \in \mathrm{SO}(3)$ and $q_n \in \mathbb{R}^3$ denote respectively the rotational and translational sub-block of the matrix a_n as defined in (2.1.5). When the configuration set Γ denotes the conditions ii. in (2.2.3), a similar argument shows that the variational principle for the chain equilibrium conditions in (2.2.9) is also obtained. For completeness, we observe that a similar result for the variation in 2.2.9 can be found in [Coleman et al., 2003]. The expression (2.2.9) is called *discrete Euler-Poincaré variation for chains*[5].

Consequently, if the chain constitutive relations in (2.2.4) admit a variational representation of the form (2.2.10) for all chain configurations $\boldsymbol{g} \in \Gamma$, then the stationary conditions for the internal energy E on the set of configurations Γ in (2.2.7) is literally an expression of the chain equilibrium conditions (2.2.1) for isolated chains. Equations (2.2.9) imply that the six quantities

$$\mathbf{m}_n + \boldsymbol{r}_n \times \mathbf{n}_n \in \mathbb{R}^3 \text{ and } \mathbf{n}_n \in \mathbb{R}^3 \tag{2.2.11}$$

are conserved along stationary configuration of the internal energy E, i.e. along equilibrium chain configuration. The existence of conserved quantities in (2.2.11) is a manifestation of the famous *Noether's Theorem*, since it can be shown that they reflect that the internal chain energy E defined in (2.2.5) is invariant through a global rigid body motion of the chain, as presented in (2.2.6).

An alternative expression of the stationary conditions is given by introducing the triples $\mathsf{m}_{n+1} \in \mathbb{R}^3$ and $\mathsf{n}_{n+1} \in \mathbb{R}^3$ as the respective components of the couple $\mathbf{m}_{n+1}$ and the force $\mathbf{n}_{n+1}$ in the director frame $\boldsymbol{R}_n$ i.e.

$$\mathsf{m}_{n+1} = \mathrm{Vect}\left(\partial_{Q_n} w_n(a_n) Q_n^T\right) \tag{2.2.12a}$$

$$\mathsf{n}_{n+1} = \partial_{q_n} w_n(a_n) \tag{2.2.12b}$$

according to the definition (2.2.10). Using then the identity

$$\begin{pmatrix} \mathbf{m}_n + \boldsymbol{r}_n \times \mathbf{n}_n \\ \mathbf{n}_n \end{pmatrix} = \mathrm{Ad}_{\boldsymbol{g}_{n-1}}^{-T} \begin{pmatrix} \mathsf{m}_n + q_{n-1} \times \mathsf{n}_n \\ \mathsf{n}_n \end{pmatrix} \tag{2.2.13}$$

where the operator $\mathrm{Ad}_{\boldsymbol{g}_n}$ is defined as

$$\mathrm{Ad}_{\boldsymbol{g}_n} = \begin{pmatrix} \boldsymbol{R}_n & 0 \\ [\boldsymbol{r}_n \times] \boldsymbol{R}_n & \boldsymbol{R}_n \end{pmatrix} \tag{2.2.14}$$

and as discussed in more details in (1.1.28), the stationary condition (2.2.7) reads

$$\begin{pmatrix} \mathsf{m}_{n+1} + q_n \times \mathsf{n}_{n+1} \\ \mathsf{n}_{n+1} \end{pmatrix} = \mathrm{Ad}_{a_{n-1}}^{T} \begin{pmatrix} \mathsf{m}_n + q_{n-1} \times \mathsf{n}_n \\ \mathsf{n}_n \end{pmatrix} \tag{2.2.15a}$$

[5] See section 1.2 for more detail on Euler-Poincaré variations.

for $n = 2, ..., N-1$

$$\boldsymbol{g}_{n+1} = \boldsymbol{g}_n a_n \tag{2.2.15b}$$

for $n = 1, ..., N$ and with

$$\boldsymbol{g} \in \Gamma \tag{2.2.15c}$$

subject to the constitutive relations (2.2.12), and for a choice of the configuration set Γ in (2.2.3).

However, realistic internal chain energyies as in [Olson et al., 1998, Becker and Everaers, 2007, Gonzalez et al., 2013] for instance, are given in terms of the internal chain coordinates x of the form (2.1.12), i.e. as

$$E(\boldsymbol{g}(x)) = \sum_{n=1}^{N-1} w_n^x(x_n) \tag{2.2.16}$$

where one has to have for consistency

$$w_n^x(x_n) = w_n(a(x_n)) \tag{2.2.17}$$

for all local internal chain coordinates x_n and for all $n = 1, ..., N-1$, according to the reconstruction rule (2.1.14). The first variation of the local internal chain energy E with respect to the internal coordinates x_n is then obtained using the identity

$$\delta a(x_n) = \left(\mathcal{T}\mathbb{L}_{x_n}^{(1)}\delta x_n\right) a(x_n) \tag{2.2.18}$$

where the explicit expression of the linear operator $\mathbb{L}_{x_n}^{(1)}$ depends on the choice of the internal chain coordinates[6] x_n. The stationary conditions are then of the form (2.2.15) but the constitutive relations (2.2.12) read

$$\begin{pmatrix} \mathsf{m}_{n+1} \\ \mathsf{n}_{n+1} \end{pmatrix} = \begin{pmatrix} \mathrm{Id}_3 & -[q_n\times] \\ 0_3 & \mathrm{Id}_3 \end{pmatrix} \mathbb{L}_{x_n}^{(1)^{-T}} \partial_{x_n} w_n^x(x_n) \tag{2.2.19}$$

for $n = 1, ..., N-1$. For the internal chain coordinates x_n used in [Lankaš et al., 2009, Gonzalez et al., 2013, Petkeviciute et al., 2014], the linear operator $\mathbb{L}_{x_n}^{(1)}$ is explicitly given by

$$\mathbb{L}_{x_n}^{(1)} = \begin{pmatrix} \mathbb{P}_1(u_n) & 0 \\ Q^{\frac{1}{2}}(u_n)\,[v_n\times]\,\mathbb{P}_2(u_n) & Q^{\frac{1}{2}}(u_n) \end{pmatrix} \tag{2.2.20}$$

for $x_n = (u_n, v_n)$ as defined in (2.1.18), and where we have used

$$\mathbb{P}_1(u_n) = \frac{1}{1 + \left(\frac{|u_n|}{2}\right)^2} \left(\mathrm{Id}_3 + \frac{1}{2}\,[u_n\times]\right) \tag{2.2.21a}$$

$$\mathbb{P}_2(u_n) = \left(\mathrm{Id}_3 + Q^{\frac{1}{2}}(u_n)\right)^{-1} \mathbb{P}_1(u_n). \tag{2.2.21b}$$

A more detailed discussion of the expressions (2.2.21) is presented in appendix A.1.

[6] See section 1.3 for more detail on matrix expansion with respect to coordinate systems.

2.2.3 On rigid body double chain equilibrium conditions and constitutive relations

We next aim to define equilibrium conditions for rigid body bichain configurations, in an analogous way to rigid body chains. However since the bichain configurations $(\boldsymbol{g}, \mathcal{P})$ are an alternative expression of double chain configurations $(\boldsymbol{g}^+, \boldsymbol{g}^-)$ it is of interest to first present the notion of a double chain equilibrium configuration and, then, discuss how it can be translated into an expression for bichains.

The system made of two interacting rigid body chains is with no doubt more complex than one single chain. A way to deduce the expression of the double chain balance laws is to see each strand as a single rigid body chain in an external field. Formally, if for each rigid body $\boldsymbol{g}_n^\pm$ we denote by $\mathbf{m}_n^\pm \in \mathbb{R}^3$ the total internal couple around the point $\boldsymbol{r}_n^\pm$ and by $\mathbf{n}_n^\pm \in \mathbb{R}^3$ the analogous total force, similarly to the single rigid body chain case (2.2.1), the *double chain equilibrium conditions* can be stated as

$$-\mathbf{m}_{n+1}^\pm + \mathbf{m}_n^\pm - \boldsymbol{r}_{n+1}^\pm \times \mathbf{n}_{n+1}^\pm + \boldsymbol{r}_n^\pm \times \mathbf{n}_n^\pm + \mathbf{c}_n^\pm + \boldsymbol{r}_n^\pm \times \mathbf{f}_n^\pm = 0 \tag{2.2.22a}$$

$$-\mathbf{n}_{n+1}^\pm + \mathbf{n}_n^\pm + \mathbf{f}_n^\pm = 0 \tag{2.2.22b}$$

for $n = 2, ..., N-1$ and where $\mathbf{c}_n^\pm \in \mathbb{R}^3$ and $\mathbf{f}_n^\pm \in \mathbb{R}^3$ denote respectively the total external couple (around the point $\boldsymbol{r}_n^\pm$) and the total external force on the n^{th} rigid body of strand $\pm$. Similarly to single chain, a particular case is the *isolated double rigid body chain* which means that there is no external contribution to the total couples and forces acting on the double chain. A similar expression of the balance laws (2.2.22) can be found in [Moakher and Maddocks, 2005] for the system of two continuum rods.

In order to make use of the single chain case, the *double chain end conditions* are assumed to be of the form

$$\Gamma = \left\{ (\boldsymbol{g}^+, \boldsymbol{g}^-) \in \mathrm{SE}(3)^{2N} \,|\, \boldsymbol{g}^+ \text{ and } \boldsymbol{g}^- \text{ satisfies i. or ii. } \right\} \tag{2.2.23}$$

where the conditions i. and ii. are defined in (2.2.3). In particular, we do not allow that $\boldsymbol{g}^+$ satisfies i. and $\boldsymbol{g}^-$ satisfies ii. . Of course, in the context of DNA modelling, it is difficult to ensure that these conditions can represent physical phenomena, but they serve nevertheless as an instructive starting point for the discussion. In what follows more sophisticated double chain end conditions can be used, but we will not describe that for the clarity of the discussion.

The *double chain constitutive relations* are then formally a mapping $(\boldsymbol{g}^+, \boldsymbol{g}^-) \mapsto (\mathbf{m}^+, \mathbf{m}^-, \mathbf{n}^+, \mathbf{n}^-, \mathbf{c}^+, \mathbf{c}^-, \mathbf{f}^+, \mathbf{f}^-)$ where $\mathbf{m}^\pm = (\mathbf{m}_1^\pm, ..., \mathbf{m}_{N-1}^\pm)(\boldsymbol{g}^+, \boldsymbol{g}^-)$, etc. As in the case of a single chain, we can define the notion of *nearest neighbour double chain constitutive relations*, which means that the local couple and force in the double chain are only functions of the nearest neighbour deformation, i.e.

$$\begin{pmatrix} \mathbf{m}_n^\pm \\ \mathbf{n}_n^\pm \end{pmatrix} = \boldsymbol{\Psi}_n^{(\mathbf{m},\mathbf{n})}(b_{n-1}, a_{n-1}^+, a_{n-1}^-, b_n) \tag{2.2.24a}$$

for $n = 2, ..., N$ and

$$\begin{pmatrix} \mathbf{c}_n^{\pm} \\ \mathbf{f}_n^{\pm} \end{pmatrix} = \boldsymbol{\Psi}_n^{(\mathbf{c},\mathbf{f})}(b_{n-1}, a_{n-1}^+, a_{n-1}^-, b_n, a_n^+, a_n^-, b_{n+1}) \tag{2.2.24b}$$

for $n = 2, ..., N-1$ and

$$\begin{pmatrix} \mathbf{c}_1^{\pm} \\ \mathbf{f}_1^{\pm} \end{pmatrix} = \boldsymbol{\Psi}_1^{(\mathbf{c},\mathbf{f})}(b_1, a_1^+, a_1^-, b_2) \text{ and } \begin{pmatrix} \mathbf{c}_N^{\pm} \\ \mathbf{f}_N^{\pm} \end{pmatrix} = \boldsymbol{\Psi}_N^{(\mathbf{c},\mathbf{f})}(b_{N-1}, a_{N-1}^+, a_{N-1}^-, b_N). \tag{2.2.24c}$$

Observe that the arguments $(b_{n-1}, a_{n-1}^+, a_{n-1}^-, b_n)$ are not independent, but we use this formulation for convenience. The constitutive relations reflect the physical constitution of the double chain and their explicit formulation for a given material require very subtle studies. The *rigid body double chain equilibrium conditions* then comprise the conditions (2.2.22) on a configuration set Γ together with a given set of constitutive relations, of the form (2.2.24) for instance.

2.2.4 On rigid body double chain internal energy and variational principles

The *double chain internal energy* is denoted by $E^{||}$ and describes the stored potential energy associated to internal deformations of the rigid body double chain. A *nearest-neighbor energy*, or a *local energy*, for the double chain is then defined as $E^{||} : \mathrm{SE}(3)^{2N} \to \mathbb{R}$ and is assumed to be of the form

$$E^{||}(\boldsymbol{g}^+, \boldsymbol{g}^-) = \sum_{n=1}^{N-1} w_n^{||}(b_n, a_n^+, a_n^-, b_{n+1}) \tag{2.2.25}$$

where $a_n^{\pm} = (\boldsymbol{g}_n^{\pm})^{-1}\boldsymbol{g}_{n+1}^{\pm} \in \mathrm{SE}(3)$ is the local junction displacement along each strand $\pm$ respectively and where $b_n \in \mathrm{SE}(3)$ denotes the rigid body displacement from the rigid body $\boldsymbol{g}_n^+$ to $\boldsymbol{g}_n^-$ on the complementary strand, as defined in (2.1.22). In particular, the displacements $a_n^{\pm}$, b_n and b_{n+1} are not independent one from another, but nevertheless this formulation will be shown to be appropriate. Observe that, similarly to the single chain internal energy in (2.2.6), the double chain energy $E^{||}$ is invariant through a global rigid body motion of the double chain, i.e.

$$E^{||}(B\boldsymbol{g}_1^+, B\boldsymbol{g}_1^-, ..., B\boldsymbol{g}_N^+, B\boldsymbol{g}_N^-) = E^{||}(\boldsymbol{g}_1^+, \boldsymbol{g}_1^-, ..., \boldsymbol{g}_N^+, \boldsymbol{g}_N^-) \tag{2.2.26}$$

for any $B \in \mathrm{SE}(3)$. It is worth to emphasize that when the internal double chain energy $E^{||}$ is assumed to be local, the two individual chains $\boldsymbol{g}^{\pm}$ are governed by an *effective non-local internal energy*, as mentioned for instance in [Lankaš et al., 2009, Gonzalez et al., 2013]. That is partially why such a double chain energy, even if they only involve nearest neighbours, provides a potentially more realistic description of double stranded macromolecules which are known to have long range couplings than single chain local energies. The internal energy $E^{||}$ allows to state a *variational principle for rigid body*

double chains, and if the double chain constitutive relations (2.2.24) admit a variational form it will be called a *hyper-elastic double chain*.

More precisely, suppose that Γ denote the set of double chain configuration where each end of both chains is prescribed, which corresponds to the case i. in (2.2.23). Using conditions of the form (2.2.7b) and (2.2.7c) for the perturbation of the double chain configuration $(\boldsymbol{g}^+, \boldsymbol{g}^-)$ and using moreover that $\delta b_1 = 0$

$$\delta b_n = (\boldsymbol{g}_n^+)^{-1}(\mathcal{T}\boldsymbol{\phi}_n^- - \mathcal{T}\boldsymbol{\phi}_n^+)\boldsymbol{g}_n^- \tag{2.2.27}$$

for $n = 2, ..., N-1$ and $\delta b_N = 0$ implies that if a double chain configuration $(\boldsymbol{g}^+, \boldsymbol{g}^-)$ makes the internal energy $E^{||}$ in (2.2.25) stationary, then it satisfies the double chain equilibrium conditions (2.2.22) for

$$\mathbf{m}_{n+1}^{\pm} = \boldsymbol{R}_n^{\pm}\mathrm{Vect}\left(\partial_{Q_n^{\pm}} w_n^{||} Q_n^{\pm T}\right) \tag{2.2.28a}$$

$$\mathbf{n}_{n+1}^{\pm} = \boldsymbol{R}_n^{\pm}\partial_{q_n^{\pm}} w_n^{||} \tag{2.2.28b}$$

for $n = 1, ..., N-1$ and

$$\mathbf{c}_n^- = \boldsymbol{R}_n^+\mathrm{Vect}\left(\partial_{P_n}\left(w_n^{||} + w_{n-1}^{||}\right) P_n^T\right) \tag{2.2.28c}$$

$$\mathbf{f}_n^- = \boldsymbol{R}_n^+\partial_{p_n}\left(w_n^{||} + w_{n-1}^{||}\right) \tag{2.2.28d}$$

$$\mathbf{c}_n^+ = -\mathbf{c}_n^- - (\boldsymbol{r}_n^- - \boldsymbol{r}_n^+) \times \mathbf{f}_n^- \tag{2.2.28e}$$

$$\mathbf{f}_n^+ = -\mathbf{f}_n^- \tag{2.2.28f}$$

for $n = 2, ..., N-1$ and where the argument $(b_n, a_n^+, a_n^-, b_{n+1})$ of the internal energy $w_n^{||}$ has been hidden for compactness. Observe that the apparent asymmetry between the definitions of the couple $\mathbf{c}_n^-$ and $\mathbf{c}_n^+$ and the force $\mathbf{f}_n^-$ and $\mathbf{f}_n^+$ is explained by the fact the internal energy $E^{||}$ is not invariant through the transformation $\boldsymbol{g}^+ \mapsto \boldsymbol{g}^-$ presented in (2.1.44). The transformation rule on the expressions (2.2.28) when changing the reference strand is obtained using $\delta b_n^{-1} = -b_n^{-1}\delta b_n b_n^{-1}$.

As for rigid body chains, the variational principle for the double chain balance laws in (2.2.22) together with the constitutive relations (2.2.28) can also be obtained when the condition set Γ denotes the condition ii. in (2.2.23). However, the following additional relations have to be satisfied

$$\begin{aligned} \mathbf{c}_1^- &= \boldsymbol{R}_1^+\mathrm{Vect}\left(\partial_{P_1} w_1^{||} P_1^T\right) \\ \mathbf{c}_N^- &= \boldsymbol{R}_N^+\mathrm{Vect}\left(\partial_{P_N} w_{N-1}^{||} P_N^T\right) \end{aligned} \tag{2.2.29a}$$

$$\begin{aligned} \mathbf{f}_1^- &= \boldsymbol{R}_1^+\partial_{p_1} w_1^{||} \\ \mathbf{f}_N^- &= \boldsymbol{R}_N^+\partial_{p_N} w_{N-1}^{||} \end{aligned} \tag{2.2.29b}$$

with $\mathbf{c}_n^+ = -\mathbf{c}_n^- + (\boldsymbol{r}_n^- - \boldsymbol{r}_n^+) \times \mathbf{f}_n^-$ and $\mathbf{f}_n^+ = -\mathbf{f}_n^-$ for $n = 1$ and $n = N$.

Consequently, if the chain constitutive relations in (2.2.24) admit a variational representation of the form (2.2.28) for all the double chain configurations $(\boldsymbol{g}^+, \boldsymbol{g}^-)$, then the stationary conditions for the internal energy $E^{||}$ on the set of configuration Γ is an expression of the chain balance laws (2.2.22) for isolated double chains. This variational principle is then called the *discrete Euler-Poincaré variation for double chains* [7].

In their respective director frames $\boldsymbol{R}_n^{\pm}$ the constitutive relations (2.2.28) then reads

$$\mathsf{m}_{n+1}^{\pm} = \mathrm{Vect}\left(\partial_{Q_n^{\pm}} w_n^{||} Q_n^{\pm T}\right) \tag{2.2.30a}$$

$$\mathsf{n}_{n+1}^{\pm} = \partial_{q_n^{\pm}} w_n^{||} \tag{2.2.30b}$$

for $n = 1, ..., N-1$ and

$$\mathsf{c}_n^- = P_n^T \mathrm{Vect}\left(\partial_{P_n}\left(w_n^{||} + w_{n-1}^{||}\right) P_n^T\right) \tag{2.2.30c}$$

$$\mathsf{f}_n^- = P_n^T \partial_{p_n}\left(w_n^{||} + w_{n-1}^{||}\right) \tag{2.2.30d}$$

$$\mathsf{c}_n^+ = -P_n \mathsf{c}_n^- + p_n \times P_n \mathsf{f}_n^- \tag{2.2.30e}$$

$$\mathsf{f}_n^+ = -P_n \mathsf{f}_n^- \tag{2.2.30f}$$

for $n = 2, ..., N-1$ and consequently, the double chain equilibrium conditions can be summarised as

$$\begin{pmatrix} \mathsf{m}_{n+1}^{\pm} \\ \mathsf{n}_{n+1}^{\pm} \end{pmatrix} = \mathrm{Ad}_{a_n^{\pm}}^T \begin{pmatrix} \mathsf{m}_n^{\pm} + \mathsf{c}_n^{\pm} \\ \mathsf{n}_n^{\pm} + \mathsf{f}_n^{\pm} \end{pmatrix} \tag{2.2.31a}$$

$$\boldsymbol{g}_{n+1}^{\pm} = \boldsymbol{g}_n^{\pm} a_n^{\pm} \tag{2.2.31b}$$

for $n = 1, ..., N-1$ and with

$$(\boldsymbol{g}^+, \boldsymbol{g}^-) \in \Gamma \tag{2.2.31c}$$

subject to the constitutive relations (2.2.30), and for a choice of the configuration set Γ in (2.2.23), analogously to the single chain in (2.2.15). We note that the operator $\mathrm{Ad}_{\boldsymbol{g}_n}$ is defined in (2.2.14).

2.2.5 On rigid body bichain variational principles and constitutive relations

In contrast, the *bichain internal energy* is denoted by $E : \mathrm{SE}(3)^{2N} \to \mathbb{R}$, as in the single rigid body chain case (2.2.5), and is defined as

$$E(\boldsymbol{g}, \mathcal{P}) = \sum_{n=1}^{N-1} w_n(\mathcal{P}_n, a_n, \mathcal{P}_{n+1}) \tag{2.2.32}$$

[7] See section 1.2 for more detail

where $a_n \in \mathrm{SE}(3)$ and $\mathcal{P}_n \in \mathrm{SE}(3)$ are the local rigid body displacement in the macrostructure and in the microstructure respectively, and where the energy terms $w_n(\mathcal{P}_n, a_n, \mathcal{P}_{n+1})$ are defined such that the total bichain internal energy satisfies

$$E(\boldsymbol{g}, \mathcal{P}) = E^{||}(\boldsymbol{g}^+, \boldsymbol{g}^-) \tag{2.2.33}$$

for the corresponding double chain configuration, as described in section 2.1.2. It is then equivalent to study either double chain configurations $(\boldsymbol{g}^+, \boldsymbol{g}^-)$ which make the internal energy $E^{||}$ stationary or bichain configurations $(\boldsymbol{g}, \mathcal{P})$ which make E stationary. In particular, the bichain energy is internal in the sense that it is invariant with respect to any overall rigid body motion of the bichain configuration, i.e.

$$E(\mathcal{P}_1, B\boldsymbol{g}_1, ..., B\boldsymbol{g}_N, \mathcal{P}_N) = E(\mathcal{P}_1, \boldsymbol{g}_1, ..., \boldsymbol{g}_N, \mathcal{P}_N) \tag{2.2.34}$$

for any $B \in \mathrm{SE}(3)$. The apparent asymmetry between the transformation of the macrostructure and microstructure under global rigid body motions of the bichain configuration is explained by the fact that the microstructure is itself invariant under this transformation, as discussed in (2.1.37).

In the case of the Dirichlet conditions for double chains in (2.2.23) one also gets Dirichlet conditions for the bichain macrostructure and microstructure $(\boldsymbol{g}, \mathcal{P})$, and the first variation of the internal energy E can then be shown to be

$$-\mathbf{m}_{n+1} + \mathbf{m}_n - \boldsymbol{r}_{n+1} \times \mathbf{n}_{n+1} + \boldsymbol{r}_n \times \mathbf{n}_n = 0 \tag{2.2.35a}$$

$$-\mathbf{n}_{n+1} + \mathbf{n}_n = 0 \tag{2.2.35b}$$

$$\mathsf{C}_n^{\mathcal{P}} = 0 \tag{2.2.35c}$$

$$\mathsf{F}_n^{\mathcal{P}} = 0 \tag{2.2.35d}$$

for $n = 2, ..., N-1$ and with the relations

$$\mathbf{m}_{n+1} = \boldsymbol{R}_n \mathrm{Vect}\left(\partial_{Q_n} w_n Q_n^T\right) \tag{2.2.36a}$$

$$\mathbf{n}_{n+1} = \boldsymbol{R}_n \partial_{q_n} w_n \tag{2.2.36b}$$

for $n = 1, ..., N-1$ and

$$\mathsf{C}_n^{\mathcal{P}} = \mathrm{Vect}\left(P_n^{\frac{T}{2}} \partial_{P_n} \left(w_n + w_{n-1}\right) P_n^{\frac{T}{2}}\right) \tag{2.2.36c}$$

$$\mathsf{F}_n^{\mathcal{P}} = \partial_{\mathsf{w}_n} \left(w_n + w_{n-1}\right) \tag{2.2.36d}$$

for $n = 2, ..., N-1$ and where the argument $(\mathcal{P}_n, a_n, \mathcal{P}_{n+1})$ of the internal energy w_n has been suppressed for compactness of the notation. We remark that we have used the following non-standard perturbation of the rotation matrix P_n

$$\delta P_n = P_n^{\frac{1}{2}} \left[\phi_n \times\right] P_n^{\frac{1}{2}} \tag{2.2.37}$$

for a triple $\phi_n \in \mathbb{R}^3$. This choice has been made in order to respect the underlying symmetry between the strand + and −.

Equations (2.2.35a) and (2.2.35b) correspond to the single chain balance laws, as presented in (2.2.1), and capture the *macrostructure equilibrium conditions*. The triple $\mathbf{m}_n \in \mathbb{R}^3$ denotes the internal couple around the point $\boldsymbol{r}_n$ acting on the rigid bodies $\boldsymbol{g}_n^+$ and $\boldsymbol{g}_n^-$ in the double chain from the pair indexed by $(n-1)$, and $\mathbf{n}_n \in \mathbb{R}^3$ the analogous force. In contrast, equations (2.2.35c) and (2.2.35d) define the *microstructure equilibrium conditions*, where $\mathsf{C}_n^{\mathcal{P}}$ and $\mathsf{F}_n^{\mathcal{P}}$ denote respectively the couple and the force *in* the microstructure. The expressions in (2.2.35) are then called the *bichain equilibrium conditions*. Note that the expressions (2.2.35) constitute a discrete version of analogous formulation in [Moakher and Maddocks, 2005] for the system of two continuum rods.

Using the double chain constitutive relations (2.2.30) together with the identities (2.1.29), the relations between the double chain couples and forces in (2.2.28) and the bichain ones in (2.2.36) can be explicitly written as

$$\mathbf{m}_n = \mathbf{m}_n^+ + \mathbf{m}_n^- + \frac{1}{2}(\boldsymbol{r}_n^- - \boldsymbol{r}_n^+) \times (\mathbf{n}_n^- - \mathbf{n}_n^+) \tag{2.2.38a}$$

$$\mathbf{n}_n = \mathbf{n}_n^+ + \mathbf{n}_n^- \tag{2.2.38b}$$

and

$$\begin{aligned} \mathsf{C}_n^{\mathcal{P}} = {} & M_n^- \left[-\mathsf{m}_{n+1}^- - q_n^- \times \mathsf{n}_{n+1}^- + \mathsf{c}_n^-\right] - M_n^+ (Q_{n-1}^-)^T \mathsf{m}_n^- \\ & - M_n^+ \left[-\mathsf{m}_{n+1}^+ - q_n^+ \times \mathsf{n}_{n+1}^+ + \mathsf{c}_n^+\right] + M_n^- (Q_{n-1}^+)^T \mathsf{m}_n^+ \end{aligned} \tag{2.2.38c}$$

$$\mathsf{F}_n^{\mathcal{P}} = \frac{1}{2} P_n^+ \left(-\mathsf{n}_{n+1}^- + \mathsf{n}_n^- + \mathsf{f}_n^-\right) - \frac{1}{2} P_n^- \left(-\mathsf{n}_{n+1}^+ + \mathsf{n}_n^+ + \mathsf{f}_n^+\right) \tag{2.2.38d}$$

for $n = 2, ..., N-1$. Note that we used the short notation $P_n^+ = P_n^{\frac{1}{2}}$ and $P_n^- = P_n^{\frac{T}{2}}$ so that the matrices $M_n^{\pm}$ are defined by

$$M_n^{\pm} = \left(\mathrm{Id}_3 + P_n^{\pm\frac{1}{2}}\right)^{-1}. \tag{2.2.39}$$

The presence of the matrices $M_n^{\pm}$ in (2.2.38) is explained by the fact that one needs to differentiate the matrices $P_n^{\pm\frac{1}{2}}$, according to the expression of the double chain local displacement in (2.1.29). More detail about expressions (2.2.38) can be found in appendix A.2.

It is interesting to observe that the relations in (2.2.38) are sums and weighted differences of the expressions in (2.2.22). In particular, the variables $\mathbf{m}_n$ and $\mathbf{n}_n$ are respectively interpreted as the total couple (around the point $\boldsymbol{r}_n$) and force acting on the rigid bodies $\boldsymbol{g}_n^+$ and $\boldsymbol{g}_n^-$. The double chain equilibrium conditions and the bichain equilibrium equations are thus equivalent.

More generally, using the identities in (2.2.38), the double chain end conditions (2.2.23) can be written in terms of the bichain configuration $(\boldsymbol{g}, \mathcal{P})$ and induces a formulation of the *bichain end conditions*. They are defined as

i′. *Dirichlet conditions*: The macrostructure and the microstructure are prescribed at both ends, i.e.

$$(\boldsymbol{g}_1, \mathcal{P}_1) \text{ and } (\boldsymbol{g}_N, \mathcal{P}_N) \text{ are prescribed.} \tag{2.2.40a}$$

ii′. *Unstressed conditions*: The macrostructure and the microstructure are free at both ends, i.e.

$$\begin{aligned} -\mathbf{m}_2 = 0,\ -\mathbf{n}_2 = 0,\ \mathsf{C}_1^{\mathcal{P}} = 0,\ \mathsf{F}_1^{\mathcal{P}} = 0 \\ \text{and } -\mathbf{m}_N = 0,\ -\mathbf{n}_N = 0,\ \mathsf{C}_N^{\mathcal{P}} = 0,\ \mathsf{F}_N^{\mathcal{P}} = 0 \end{aligned} \tag{2.2.40b}$$

We remark that the extra constitutive relations

$$\begin{aligned} \mathsf{C}_1^{\mathcal{P}} &= \text{Vect}\left(P_1^{\frac{T}{2}} \partial_{P_1} w_1 P_1^{\frac{T}{2}}\right) \\ \mathsf{C}_N^{\mathcal{P}} &= \text{Vect}\left(P_N^{\frac{T}{2}} \partial_{P_N} w_{N-1} P_N^{\frac{T}{2}}\right) \end{aligned} \tag{2.2.41a}$$

$$\begin{aligned} \mathsf{F}_1^{\mathcal{P}} &= \partial_{\mathsf{w}_1} w_1 \\ \mathsf{F}_N^{\mathcal{P}} &= \partial_{\mathsf{w}_N} w_{N-1}. \end{aligned} \tag{2.2.41b}$$

are obtained in the cases ii′.

As in the single rigid body chain case, we generically denote Γ to be the subset of all bichain configurations satisfying one end condition i′. or ii′., i.e.

$$\Gamma = \left\{(\boldsymbol{g}, \mathcal{P}) \in \text{SE}(3)^{2N} \,|\, (\boldsymbol{g}, \mathcal{P}) \text{ satisfies i′. or ii′. } \right\} \tag{2.2.42}$$

The variational principle for the bichain balance laws in (2.2.35) together with the constitutive relations (2.2.36) is also valid when the condition set Γ denotes the condition ii'. in (2.2.42). Consequently, if the bichain constitutive relations admit a variational representation of the form (2.2.36) for all bichain configurations $(\boldsymbol{g}, \mathcal{P})$, then the stationary conditions for the internal energy E on the set of configuration Γ is an expression of the bichain equilibrium conditions (2.2.35) for isolated bichains. This expression is then referred as *discrete Euler-Poincaré variation for bichains*[8]. As a corollary of the bichain balance laws (2.2.35), the six following quantities

$$\mathbf{m}_n + \boldsymbol{r}_n \times \mathbf{n}_n \in \mathbb{R}^3 \text{ and } \mathbf{n}_n \in \mathbb{R}^3 \tag{2.2.43}$$

are conserved along equilibrium bichain configuration. Similarly to the single rigid body chain case in (2.2.11), *Noether's Theorem* reveals the connexion between the conserved quantities in (2.2.43) and the invariance of the internal double chain internal energy $E^{||}$ under global rigid body motions of the system in (2.2.26), or equivalently in terms of the bichain internal energy E in (2.2.34).

[8]See section 1.2 for more detail

In the director frame of the macrostructure $\boldsymbol{R}_n$, the constitutive relations (2.2.36a) and (2.2.36b) become

$$\mathsf{m}_{n+1} = \mathrm{Vect}\left(\partial_{Q_n} w_n Q_n^T\right) \tag{2.2.44a}$$

$$\mathsf{n}_{n+1} = \partial_{q_n} w_n \tag{2.2.44b}$$

for $n = 1, ..., N-1$ and consequently, the bichain equilibrium conditions can be summarized as

$$\begin{pmatrix} \mathsf{m}_{n+1} + q_n \times \mathsf{n}_{n+1} \\ \mathsf{n}_{n+1} \end{pmatrix} = \mathrm{Ad}^T_{a_n} \begin{pmatrix} \mathsf{m}_n + q_{n-1} \times \mathsf{n}_n \\ \mathsf{n}_n \end{pmatrix} \\ \begin{pmatrix} \mathsf{C}_n^{\mathcal{P}} \\ \mathsf{F}_n^{\mathcal{P}} \end{pmatrix} = 0 \tag{2.2.45a}$$

for $n = 2, ..., N-1$

$$\boldsymbol{g}_{n+1} = \boldsymbol{g}_n a_n \tag{2.2.45b}$$

for $n = 1, ..., N-1$ and with

$$(\boldsymbol{g}, \mathcal{P}) \in \Gamma \tag{2.2.45c}$$

subject to the constitutive relations in (2.2.36) and in (2.2.44), and for a choice of the configuration set Γ in (2.2.42). It has to be noticed that the equilibrium conditions for a double chain in (2.2.31) can be written as a system which contains the equilibrium conditions for the macrostructure, which are of the form of those for a single rigid body chain in (2.2.15), and equilibrium conditions for the microstructure, and that the two subsystems are coupled through the bichain constitutive relations which reflect the original double chain structure. Similar observations are made in [Moakher and Maddocks, 2005] for the system of two continuum rods.

The main application of bichain theory that we will discuss is in the context of DNA modelling, where a realistic function E can be built in terms of internal bichain coordinates [Gonzalez et al., 2013, Petkeviciute et al., 2014]. It remains then to translate the bichain constitutive relations to an internal energy of the form

$$E(\boldsymbol{g}(x), \mathcal{P}(y)) = \sum_{n=1}^{N-1} w_n^{(x,y)}(y_n, x_n, y_{n+1}) \tag{2.2.46}$$

with

$$w_n^{(x,y)}(y_n, x_n, y_{n+1}) = w_n(\mathcal{P}(y_n), a(x_n), \mathcal{P}(y_{n+1})) \tag{2.2.47}$$

for all local internal bichain coordinates and for all $n = 1, ..., N-1$, according to the bichain reconstruction rule (2.1.34).

The first variation of the local internal chain energy E with respect to the internal coordinates x_n and y_n is then obtained using the identities

$$\begin{aligned} \delta a(x_n) &= \left(\mathcal{T}\mathbb{L}_{x_n}^{(1)}\delta x_n\right) a(x_n) \\ \delta\mathcal{P}(y_n) &= \left(P_n^{\frac{1}{2}}\left[\mathbb{B}_{\eta_n}\delta\eta_n\times\right]P_n^{\frac{1}{2}}, \delta\mathsf{w}_n\right) \end{aligned} \tag{2.2.48}$$

where the explicit expression of the linear operators $\mathbb{L}_{x_n}^{(1)}$ and $\mathbb{B}_{\eta_n}$ depends on the choice of the internal bichain coordinates[9] (x, y). The equilibrium bichain conditions are then of the form (2.2.45), but the constitutive relations in (2.2.36) and in (2.2.44) read

$$\begin{pmatrix}\mathsf{m}_{n+1}\\ \mathsf{n}_{n+1}\end{pmatrix} = \begin{pmatrix}\mathrm{Id}_3 & -[q_n\times]\\ 0_3 & \mathrm{Id}_3\end{pmatrix}\mathbb{L}_{x_n}^{(1)^{-T}}\partial_{x_n}w_n^{(x,y)} \tag{2.2.49a}$$

for $n = 1, ..., N-1$ and

$$\begin{pmatrix}\mathsf{C}_n^{\mathcal{P}}\\ \mathsf{F}_n^{\mathcal{P}}\end{pmatrix} = \begin{pmatrix}\mathbb{B}_{\eta_n}^{-T} & 0_3\\ 0_3 & \mathrm{Id}_3\end{pmatrix}\partial_{y_n}\left(w_n^{(x,y)} + w_{n+1}^{(x,y)}\right) \tag{2.2.49b}$$

for $n = 1, ..., N$. For the internal coordinates used in [Lankaš et al., 2009, Gonzalez et al., 2013, Petkeviciute et al., 2014], an explicit form of the operator $\mathbb{L}_{x_n}^{(1)}$ is given in (2.2.20) and the operator $\mathbb{B}_{\eta_n}$ can be shown to be

$$\mathbb{B}_{\eta_n} = P^{\frac{T}{2}}(\eta_n)\mathbb{P}_1(\eta_n) \tag{2.2.50}$$

where the matrix $\mathbb{P}_1(\eta_n)$ is defined in (2.2.21a). More detail about explicit formulation of the linear operators $\mathbb{L}_{x_n}^{(1)}$ and $\mathbb{B}_{\eta_n}$ can be found in appendix A.1.

2.2.6 Formulation of bichain equilibrium conditions through Lagrange multipliers

An alternative formulation of the bichain equilibrium conditions on a configuration set Γ is obtained using the *Lagrange multipliers method* regarding the condition (2.2.45c) as a set of constraints on the internal bichain displacements. Precisely, in the case of the Dirichlet conditions described in (2.2.40a), a bichain equilibrium configuration is characterised by the fact that the set of internal junction displacements $(a_1, ..., a_{N-1})$ and of microstructure configurations $(\mathcal{P}_1, ..., \mathcal{P}_N)$ make the bichain internal energy E, as defined in (2.2.32), stationary under the constraints that

$$\boldsymbol{g}_1 = \mathrm{Id}_4,\ \boldsymbol{g}_N = \prod_{n=1}^{N-1} a_n = {\boldsymbol{g}_N}^* \tag{2.2.51a}$$

$$\mathcal{P}_1 = {\mathcal{P}_1}^* \text{ and } \mathcal{P}_N = {\mathcal{P}_N}^*. \tag{2.2.51b}$$

[9] See section 1.3 for more detail on matrix expansion with respect to coordinate systems.

Observe that the Dirichlet conditions can always be assumed of the form (2.2.51a) since the energy E is assumed to respect the invariance through global rigid body motion of the bichain configuration $(\boldsymbol{g}, \mathcal{P})$, as discussed in (2.2.34).

The set of *admissible perturbations* induced by the constraints in (2.2.51) can be shown to be

$$\left\{ (\phi, \Theta) \,|\, \sum_{n=1}^{N-1} \mathrm{Ad}_{\boldsymbol{g}_n} \Theta_n = 0,\, \phi_1 = 0,\, \phi_N = 0,\, \delta \mathsf{w}_1 = 0 \text{ and } \delta \mathsf{w}_N = 0 \right\} \tag{2.2.52a}$$

with

$$(\phi, \Theta) = (\phi_1, \Theta_1, ..., \Theta_{N-1}, \phi_N) \in \mathbb{R}^{12N-6} \tag{2.2.52b}$$

using the specific form of the perturbation $\delta a_n = (\mathcal{T}\Theta_n)\, a_n$ and $\delta P_n = P_n^{\frac{1}{2}} \left[\phi_n \times\right] P_n^{\frac{1}{2}}$ for all n, similarly to (2.2.7b) and (2.2.37), and the definition of the operator $\mathrm{Ad}_{\boldsymbol{g}_n}$ in (2.2.14).

Consequently, the bichain configurations $(\boldsymbol{g}, \mathcal{P})$ which make stationary the internal energy $E(\boldsymbol{g}, \mathcal{P})$ in (2.2.32) with the respect to any admissible perturbation (ϕ, Θ) in 2.2.52 have to satisfy for some real triples $\boldsymbol{\lambda^R}$, $\boldsymbol{\lambda^r}$, λ_1^P, λ_N^P, $\lambda_1^{\mathsf{w}}, \lambda_N^{\mathsf{w}}$ the conditions

$$\begin{pmatrix} \mathsf{m}_{n+1} + q_n \times \mathsf{n}_{n+1} \\ \mathsf{n}_{n+1} \end{pmatrix} - \mathrm{Ad}_{\boldsymbol{g}_n}^T \begin{pmatrix} \boldsymbol{\lambda^R} \\ \boldsymbol{\lambda^r} \end{pmatrix} = 0 \tag{2.2.53a}$$

$$\begin{pmatrix} \mathsf{C}_n^{\mathcal{P}} \\ \mathsf{F}_n^{\mathcal{P}} \end{pmatrix} = 0 \tag{2.2.53b}$$

for $n = 2, ..., N-1$ and

$$\begin{pmatrix} \mathsf{C}_1^{\mathcal{P}} \\ \mathsf{F}_1^{\mathcal{P}} \end{pmatrix} = \begin{pmatrix} \lambda_1^P \\ \lambda_1^{\mathsf{w}} \end{pmatrix} \text{ and } \begin{pmatrix} \mathsf{C}_N^{\mathcal{P}} \\ \mathsf{F}_N^{\mathcal{P}} \end{pmatrix} = \begin{pmatrix} \lambda_N^P \\ \lambda_N^{\mathsf{w}} \end{pmatrix} \tag{2.2.53c}$$

according to the variational constitutive relations in (2.2.36) and in (2.2.44). The set

$$\boldsymbol{\lambda} = (\boldsymbol{\lambda^R}, \boldsymbol{\lambda^r}, \lambda_1^P, \lambda_N^P, \lambda_1^{\mathsf{w}}, \lambda_N^{\mathsf{w}}) \in \mathbb{R}^{18} \tag{2.2.54}$$

is called the set of *bichain Lagrange multipliers*.

Interestingly, it can be shown that introducing the *Lagrangian function* $\mathcal{L} : \mathrm{SE}(3)^{2N} \times \mathbb{R}^{18} \to \mathbb{R}$ as

$$\mathcal{L}(\boldsymbol{g}, \mathcal{P}, \boldsymbol{\lambda}) = E(\boldsymbol{g}, \mathcal{P}) - \boldsymbol{\lambda} \Phi(\mathcal{P}_1, \boldsymbol{g}_N, \mathcal{P}_N; {\mathcal{P}_1}^*, {\boldsymbol{g}_N}^*, {\mathcal{P}_N}^*) \tag{2.2.55}$$

where the function Φ is defined as

$$\Phi(\mathcal{P}_1, \boldsymbol{g}_N, \mathcal{P}_N; {\mathcal{P}_1}^*, {\boldsymbol{g}_N}^*, {\mathcal{P}_N}^*) = \begin{pmatrix} \frac{1}{2}\mathrm{Vect}\left(\boldsymbol{R}_N \left({\boldsymbol{R}_N}^*\right)^T\right) \\ \boldsymbol{r}_N - \boldsymbol{R}_N \left({\boldsymbol{R}_N}^*\right)^T {\boldsymbol{r}_N}^* \\ \frac{1}{2}\mathrm{Vect}\left(P_1 \left({P_1}^*\right)^T\right) \\ \frac{1}{2}\mathrm{Vect}\left(P_N \left({P_N}^*\right)^T\right) \\ \mathsf{w}_1 - {\mathsf{w}_1}^* \\ \mathsf{w}_N - {\mathsf{w}_N}^* \end{pmatrix} \tag{2.2.56}$$

then the stationary conditions under constraints (2.2.53) exactly correspond to the regular stationary conditions for the function $\mathcal{L}$ over the set of internal junction displacement and microstructure configuration $(\mathcal{P}_1, a_1, ..., a_{N-1}, \mathcal{P}_N)$ and over the Lagrange multipliers $\boldsymbol{\lambda}$, i.e.

$$\delta\mathcal{L}(\boldsymbol{g}(a), \mathcal{P}, \boldsymbol{\lambda})(\delta a, \delta\mathcal{P}, \delta\boldsymbol{\lambda}) = 0 \tag{2.2.57}$$

for any perturbations $(\delta a, \delta\mathcal{P}, \delta\boldsymbol{\lambda})$ and where the macrostructure configuration $\boldsymbol{g}$ is seen as a function of the set local junction displacement $a = (a_1, ..., a_{N-1})$ according to the definition (2.1.4).

We should also observe that, using the identity (2.2.13), the conditions in (2.2.53a) can be equivalently expressed as

$$\begin{pmatrix} \mathbf{m}_{n+1} + \boldsymbol{r}_n \times \mathbf{n}_{n+1} \\ \mathbf{n}_{n+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\lambda}^{\boldsymbol{R}} \\ \boldsymbol{\lambda}^{\boldsymbol{r}} \end{pmatrix}, \tag{2.2.58}$$

which, together with (2.2.53c), allows to interpret the Lagrange multipliers $\boldsymbol{\lambda} = (\boldsymbol{\lambda}^{\boldsymbol{R}}, \boldsymbol{\lambda}^{\boldsymbol{r}}, \lambda_1^P, \lambda_N^P, \lambda_1^{\mathsf{w}}, \lambda_N^{\mathsf{w}})$ as the set of couple and force which are needed to enforce the constraints in (2.2.51), and that moreover the triples $\boldsymbol{\lambda}^{\boldsymbol{R}}$ and $\boldsymbol{\lambda}^{\boldsymbol{r}}$ exactly correspond, in the case of internal bichain energy, to the conserved quantities described in (2.2.43). This remark shows that the expressions in (2.2.53) can be used to express the equilibrium conditions where the Lagrange multipliers, i.e. the couple and force, are prescribed by removing the constraints in (2.2.51). An application of the system (2.2.53) in the context of the rigid base DNA model is presented in section 6.3.

2.2.7 On bichain models for single chains

We conclude this part on the mathematical formulation of chain and bichain equilibrium conditions by some remarks on the relations between the rigid body chain model and and the bichain model.

As discussed in (2.2.45), the bichain configuration contains naturally a macrostructure $\boldsymbol{g}$ which is a single rigid body chain configuration, and moreover part of the bichain equilibrium conditions are literally the equilibrium conditions for a single chain. It is then the bichain constitutive relations which make the study of bichain equilibrium configuration very different from the one single chain equilibrium configuration. A natural question is then whether the mechanical behaviour of the bichain macrostructure can be understood as a single chain behaviour.

A trivial example is when the bichain internal energy (2.2.32) decouples into the contributions from the macrostructure and the microstructure deformations, i.e. when E is of the form

$$E(\boldsymbol{g}, \mathcal{P}) = \sum_{n=1}^{N-1} w_n^a(a_n) + w_n^{\mathcal{P}}(\mathcal{P}_n, \mathcal{P}_{n+1}). \tag{2.2.59}$$

In such cases, the energy E is said to be a *separable internal bichain energy.* This decoupling implies then that the constitutive relations (2.2.36) have the property that

the macrostructure deformations are independent from the microstructure ones, and then that any bichain equilibrium configuration $(\boldsymbol{g}, \mathcal{P})$ governed by this internal energy is indistinguishable from a single chain equilibrium configuration $\boldsymbol{g}$ governed by the internal energy $E(\boldsymbol{g}) = \sum_{n=1}^{N-1} w_n^a(a_n)$. The bichain theory with nearest-neighbor internal energy contains then strictly the single chain theory with local internal energy.

Conversely, another extreme case is given by a *constrained microstructure*, and it corresponds to constraining the microstructure $\mathcal{P}$ to a given (possibly stressed) configuration. Even if a rigorous discussion of this problem is left as a future development, we believe that all these local constraints imply that the single chain behaviour of the bichain macrostructure cannot be described by an internal nearest neighbor chain energy.

Consequently, the bichain model can be seen as an intermediate model between local and fully non-local chain models.

2.3 On Stationary Configurational Distributions and Chain Moments

This section discusses an approach to the stochastic behaviour of rigid body chains in the context of equilibrium statistical mechanics. Based on the existence of a variational principle for the equilibrium conditions, we present a stationary configurational distribution for single chains, and respectively for bichains, in a coordinate independent way. The definition of chain observables and their first moment is then introduced, and the particular role of configurational chain first moment, as well as the coordinate first and second moments, is emphasised.

2.3.1 On stationary configurational distributions for isolated chains and bichains

A rigid body chain is said to be an *isolated rigid body chain in a stochastic bath* if the only couples and forces which act on it arise from the internal deformations of the material and from the interactions with the surrounding fluid, which are modelled as linear damping and Brownian forcing. This definition is a natural extension of the case of a single rigid body [Steele, 1963a, Steele, 1963b, Brenner, 1964, Brenner, 1965, Hubbard, 1972, Doi and Edwards, 1986, Walter et al., 2010]

The dynamic of an isolated chain in a stochastic bath, is then, in the context of *equilibrium statistical mechanics*, assumed to be be governed by a *stationary chain configurational distribution* $d\rho(\boldsymbol{g})$. This assumption is made by many authors, and is rigorously discussed in [Walter et al., 2010] for instance. We denote by k_B the *Boltzmann constant* and T the *kinetic temperature of the bath*. Moreover, for an isolated chain with free ends, the configurational distribution can be written as a *Boltzmann distribution* in terms of the internal chain energy E and a choice of internal chain coordinates x. Precisely, this distribution of the from

$$d\rho(\boldsymbol{g})(x) = \frac{1}{Z_\Gamma} \exp\left\{-\frac{1}{k_B T} E(\boldsymbol{g}(x))\right\} J(x)dx \text{ for } \boldsymbol{g} \in \Gamma \tag{2.3.1a}$$

with the *normalisation constant*

$$Z_\Gamma = \int_\Gamma \exp\left\{-\frac{1}{k_B T} E(\boldsymbol{g}(x))\right\} J(x)dx \tag{2.3.1b}$$

where

$$J(x) = \prod_{n=1}^{N-1} J_n(x_n) \tag{2.3.1c}$$

is the *Jacobian factor* coming from the integration over the rotation group [Eyring, 1932, Gonzalez and Maddocks, 2001, Becker and Everaers, 2007, Lankaš et al., 2009, Walter et al., 2010, Chirikjian, 2011, Gonzalez et al., 2013], where the configuration space Γ is

defined in (2.2.3) using the unstressed end conditions, described in (2.2.2b), and where dx is a compact notation for $dx_1...dx_{N-1}$.

We should emphasize that the form of the configurational distribution in (2.3.1a) is the same for local and non-local chain internal energies and, moreover, that this expression is not strictly a Boltzmann distribution because of the presence of the Jacobian correction $J(x)$ but is its exact analogue on a non-flat configuration space. The Jacobian factor ensures that the value of the probability density $d\rho(\boldsymbol{g})(x)$ for a given chain configuration $\boldsymbol{g}$ transforms consistently in any choice of the internal chain coordinates x. This is why, one can write formally the rigid body chain configurational distribution in (2.3.1a) in the following compact form

$$d\rho(\boldsymbol{g}) = \frac{1}{Z_\Gamma} \exp\left\{-\frac{1}{k_B T} E(\boldsymbol{g}(a))\right\} da \text{ for } \boldsymbol{g} \in \Gamma \tag{2.3.2a}$$

with the *normalisation constant*

$$Z_\Gamma = \int_\Gamma \exp\left\{-\frac{1}{k_B T} E(\boldsymbol{g}(a))\right\} da \tag{2.3.2b}$$

and with the identification

$$da = J(x)dx \tag{2.3.2c}$$

where $a = (a_1, ..., a_{N-1})$ is the set of local junction displacements in the chain, as discussed in detail by [Chirikjian, 2011].

The functional form of the local Jacobian $J_n(x_n)$ depends on the choice of internal chain coordinates x_n. For instance, in terms of the coordinates used in [Lankaš et al., 2009, Gonzalez et al., 2013, Petkeviciute et al., 2014], one has explicitly

$$J_n(x_n) = \frac{1}{\left(1 + \frac{|u_n|}{4}\right)^2} \text{ with } x_n = (u_n, v_n). \tag{2.3.3}$$

Analogous expressions for some other choice of coordinates are systematically presented in [Chirikjian, 2011]. If the internal energy chain E is assumed to be local, as in (2.2.5), the distribution $d\rho(\boldsymbol{g})(x)$ factorizes into

$$d\rho(\boldsymbol{g})(x) = \prod_{n=1}^{N} d\rho_n(a(x_n)) \tag{2.3.4a}$$

with

$$d\rho_n(a(x_n)) = \frac{1}{Z_n} \exp\left\{-\frac{1}{k_B T} w_n(a(x_n))\right\} J_n(x_n) dx_n \tag{2.3.4b}$$

where Z_n is the local normalising constant and where the chain energy E expressed in internal coordinates x is presented in (2.2.16). Equivalently, one can write

$$d\rho(\boldsymbol{g}) = \prod_{n=1}^{N} d\rho_n(a_n) \tag{2.3.5a}$$

with

$$d\rho_n(a_n) = \frac{1}{Z_n} \exp\left\{-\frac{1}{k_B T} w_n(a_n)\right\} da_n. \tag{2.3.5b}$$

We remark that, in the DNA modelling literature, approximation of the distribution in (2.3.4) have been built for different choice of the internal coordinates [Marky and Olson, 1994, Olson et al., 1998, Coleman et al., 2003, Gonzalez and Maddocks, 2001, Becker and Everaers, 2007, Lankaš et al., 2009].

We define an *isolated rigid body double chain in a stochastic bath* similarly to single chains and we also assume the existence of a *stationary bichain configurational distribution* $d\rho(\boldsymbol{g}, \mathcal{P})$. By adapting the development presented in [Walter et al., 2010] for an isolated bichain with free ends, we obtain a configurational distribution in terms of the internal bichain energy E and some internal bichain coordinates (x, y) of the form

$$d\rho(\boldsymbol{g}, \mathcal{P})(x, y) = \frac{1}{Z_\Gamma} \exp\left\{-\frac{1}{k_B T} E(\boldsymbol{g}(x), \mathcal{P}(y))\right\} J(x, y) dx dy \text{ for } (\boldsymbol{g}, \mathcal{P}) \in \Gamma \tag{2.3.6a}$$

with the *normalization constant*

$$Z_\Gamma = \int_\Gamma \exp\left\{-\frac{1}{k_B T} E(\boldsymbol{g}(x), \mathcal{P}(y))\right\} J(x, y) dx dy \tag{2.3.6b}$$

where

$$J(x, y) = \prod_{n=1}^{N-1} J_n(x_n) \prod_{m=1}^{N-1} J_m(y_m) \tag{2.3.6c}$$

is the *Jacobian factor* associated to bichain coordinates [Lankaš et al., 2009, Gonzalez et al., 2013] and where the configuration set Γ is defined in (2.2.42) using the end conditions described in (2.2.40b). Since the value of the bichain configurational density $d\rho(\boldsymbol{g}, \mathcal{P})$ for a given bichain configuration transforms consistently in any internal coordinates (x, y), one can write the expression in (2.3.6) as

$$d\rho(\boldsymbol{g}, \mathcal{P}) = \frac{1}{Z_\Gamma} \exp\left\{-\frac{1}{k_B T} E(\boldsymbol{g}(a), \mathcal{P})\right\} da d\mathcal{P} \text{ for } (\boldsymbol{g}, \mathcal{P}) \in \Gamma \tag{2.3.7a}$$

with the *normalization constant*

$$Z_\Gamma = \int_\Gamma \exp\left\{-\frac{1}{k_B T} E(\boldsymbol{g}(a), \mathcal{P})\right\} da d\mathcal{P} \tag{2.3.7b}$$

where

$$da d\mathcal{P} = J(x, y) dx dy. \tag{2.3.7c}$$

For the choice of the internal bichain coordinates described in (2.1.33), one has $J_n(x_n)$ as in (2.3.3) and similarly

$$J_m(y_m) = \frac{1}{\left(1 + \frac{|\eta_m|}{4}\right)^2} \text{ with } y_m = (\eta_m, \mathsf{w}_m). \tag{2.3.8}$$

Using the arguments presented in the section 2.2.7, the single chain configurational distribution $d\rho(\boldsymbol{g})$ in (2.3.1) can be seen as a marginal of the bichain configurational distribution $d\rho(\boldsymbol{g}, \mathcal{P})$ in (2.3.6). Even for a nearest neighbor bichain internal energy of the form (2.2.32), the marginal distribution for the macrostructure is in general very different from the one for local single chains presented in (2.3.4). In contrast, this marginal distribution exactly factorizes into (2.3.4) when the internal bichain energy decouples into independent macrostructure and microstructure configurations.

2.3.2 Observables and expectations for chains and bichains

A *rigid body chain observable* $\mathcal{A}$ is formally a function of the chain configuration $\mathcal{A}(\boldsymbol{g}_1, ..., \boldsymbol{g}_N) \in \mathbb{R}$, which will simply be denoted by $\mathcal{A}(\boldsymbol{g}) \in \mathbb{R}$. The most common observables $\mathcal{A}(\boldsymbol{g})$ in polymer physics include, for instance, the projection of some given bond in the chain on the first one, or of the end-to-end vector [Flory, 1969, Schellman, 1974, Doi and Edwards, 1986, Maroun and Olson, 1988, Yamakawa, 1997]. The chain configurational distribution $d\rho(\boldsymbol{g})$ allows the definition of the *expectation of a chain observable* $\mathcal{A}(\boldsymbol{g})$ as

$$\langle \mathcal{A}(\boldsymbol{g}(x)) \rangle = \int_\Gamma \mathcal{A}(\boldsymbol{g}(x)) d\rho(\boldsymbol{g})(x) \tag{2.3.9}$$

where x denotes a set of chain internal coordinates of the chain configuration $\boldsymbol{g}$, as defined in (2.1.14), and Γ denotes the set of chain configuration where the ends are free in (2.2.3), and where $d\rho(\boldsymbol{g})(x)$ denotes the stationary chain configurational distribution written in terms of the coordinates x, as in (2.3.1). As discussed in section 2.3.1, the distribution $d\rho(\boldsymbol{g})(x)$ transforms consistently in any choice of internal coordinates x, and one can then write the expectation in (2.3.9) as

$$\langle \mathcal{A}(\boldsymbol{g}) \rangle = \int_\Gamma \mathcal{A}(\boldsymbol{g}) d\rho(\boldsymbol{g}). \tag{2.3.10}$$

The symbol $\langle \, . \, \rangle$ stands therefore for the expectation over the configuration ensemble Γ, and the value of $\langle \mathcal{A}(\boldsymbol{g}) \rangle$ is independent of the choice of internal chain coordinates used to describe the Boltzmann distribution and the observable $\mathcal{A}(\boldsymbol{g}(x))$.

Similarly, a *rigid body bichain observable* is also denoted by $\mathcal{A}$ but it is a function of the bichain configuration $\mathcal{A}(\mathcal{P}_1, \boldsymbol{g}_1, ..., \boldsymbol{g}_N, \mathcal{P}_N) \in \mathbb{R}$, or simply $\mathcal{A}(\boldsymbol{g}, \mathcal{P}) \in \mathbb{R}$. Using the bichain configurational distribution in (2.3.6) allows one to define the *expectation of a bichain observable* as

$$\langle \mathcal{A}(\boldsymbol{g}(x), \mathcal{P}(y)) \rangle = \int_\Gamma \mathcal{A}(\boldsymbol{g}(x), \mathcal{P}(y)) d\rho(\boldsymbol{g}, \mathcal{P})(x, y) \tag{2.3.11}$$

where (x, y) denotes a set of bichain internal coordinates of the bichain configuration $(\boldsymbol{g}, \mathcal{P})$, as defined in (2.1.34) and where Γ denotes the set of bichain configurations with free ends in (2.2.42). Since the distribution $d\rho(\boldsymbol{g}, \mathcal{P})(x, y)$ transforms consistently in any

choice of internal coordinates (x, y), one can then write the expectation in (2.3.11) as

$$\langle \mathcal{A}(\boldsymbol{g}, \mathcal{P}) \rangle = \int_{\Gamma} \mathcal{A}(\boldsymbol{g}, \mathcal{P}) d\rho(\boldsymbol{g}, \mathcal{P}). \tag{2.3.12}$$

As for the single chains in (2.3.10), one deduces that the value of the expectation $\langle \mathcal{A}(\boldsymbol{g}, \mathcal{P}) \rangle$ is independent of the choice of internal bichain coordinates.

A chain observable $\mathcal{A}(\boldsymbol{g})$ is said to be an *internal chain observable* if it is invariant under any overall rigid body motion of the chain, i.e. if

$$\mathcal{A}(B\boldsymbol{g}_1, ..., B\boldsymbol{g}_N) = \mathcal{A}(\boldsymbol{g}_1, ..., \boldsymbol{g}_N) \tag{2.3.13}$$

for any $B \in \mathrm{SE}(3)$. Similarly, an *internal bichain observable* $\mathcal{A}(\boldsymbol{g}, \mathcal{P})$ has the property

$$\mathcal{A}(\mathcal{P}_1, B\boldsymbol{g}_1, ..., B\boldsymbol{g}_N, \mathcal{P}_N) = \mathcal{A}(\mathcal{P}_1, \boldsymbol{g}_1, ..., \boldsymbol{g}_N, \mathcal{P}_N) \tag{2.3.14}$$

for any $B \in \mathrm{SE}(3)$. In the context of isolated chain and isolated bichains, internal observables are of interest since they are the only observables whose expectation is well defined. In other words, if an observable $\mathcal{A}$ is not internal then the computation of its expectation, in (2.3.10) and (2.3.12) respectively, requires information which is not described by a configurational distribution that is only a function of the internal energy E.

Among the variety of internal observables of chain and bichain configurations, the relative rigid body motion $\boldsymbol{g}_{m,n}$ between the m^{th} element and the n^{th} element in the chain, as defined in (2.1.7), or respectively in the bichain macrostructure, is of particular interest. The *first moment of the relative rigid body motion*, or simply the *configurational first moment*, is defined as a general chain observable, i.e. as

$$\langle \boldsymbol{g}_{m,n} \rangle = \int_{\Gamma} \boldsymbol{g}_{m,n} d\rho(\boldsymbol{g}) \tag{2.3.15}$$

according to the definitions (2.3.10), and as

$$\langle \boldsymbol{g}_{m,n} \rangle = \int_{\Gamma} \boldsymbol{g}_{m,n} d\rho(\boldsymbol{g}, \mathcal{P}) \tag{2.3.16}$$

for bichains according to (2.3.12). In both cases, the block structure of $\langle \boldsymbol{g}_{m,n} \rangle$ can be written in terms of the relative rotation $\boldsymbol{R}_{m,n}$ and relative translation $\boldsymbol{r}_{m,n}$, defined in (2.1.6), as

$$\langle \boldsymbol{g}_{m,n} \rangle = \begin{pmatrix} \langle \boldsymbol{R}_{m,n} \rangle & \langle \boldsymbol{r}_{m,n} \rangle \\ 0 & 1 \end{pmatrix}. \tag{2.3.17}$$

Moments of a chain configuration provide condensed information about their statistical properties and offer a useful way to compare and characterise different chains, however they are probably not sufficient to deduce the detailed mechanical parameters of the chain [Flory, 1969, Yamakawa, 1997]. The study of the configurational first moment

$\langle \boldsymbol{g}_{m,n} \rangle$ is of importance because it encodes many other significant observables in polymer physics. More details are presented in section 2.4.

Another set of observables of interest are the moments of the chain, or respectively bichain, internal coordinates. For a bichain configurational distribution and for some internal coordinates (x, y) we will use the specific notation $\hat{x}_n$ and $\hat{y}_n$ for the *first moment of the internal coordinates* x_n and y_n respectively. We denote the *centered second moment of the internal coordinates* by the matrix c. The triples $\hat{x}_n$ and $\hat{y}_n$ are also referred to as the *local bichain coordinate expectations* and the matrix c as the *bichain covariance matrix*. Using the ordering of the internal coordinates in (2.1.31), the coordinate expectation can be summarized as

$$(\hat{x}, \hat{y}) = (\hat{y}_1, \hat{x}_1, ..., \hat{x}_{N-1}, \hat{y}_N) \tag{2.3.18a}$$

where

$$\hat{x}_n = \langle x_n \rangle \text{ and } \hat{y}_n = \langle y_n \rangle , \tag{2.3.18b}$$

and where the expectation $\langle \, . \, \rangle$ is understood in the sense of (2.3.12). The covariance matrix then has the block structure

$$c = \begin{pmatrix} c^{yy}_{1,1} & c^{yx}_{1,1} & ... & c^{yx}_{1,N} & c^{yy}_{1,N} \\ c^{xy}_{1,1} & c^{xx}_{1,1} & ... & ... & c^{xy}_{1,N} \\ ... & ... & ... & ... & ... \\ c^{xy}_{N-1,1} & ... & ... & c^{xx}_{N-1,N-1} & c^{yx}_{N,N-1} \\ c^{yy}_{N,1} & c^{xy}_{N,1} & ... & c^{xy}_{N,N-1} & c^{yy}_{N,N} \end{pmatrix} \tag{2.3.19a}$$

where the blocks are defined as

$$\begin{aligned} c^{yy}_{m,n} = \langle (y_m - \hat{y}_m) \otimes (y_n - \hat{y}_n) \rangle \, , \, c^{xx}_{m,n} = \langle (x_m - \hat{x}_m) \otimes (x_n - \hat{x}_n) \rangle \\ c^{yx}_{m,n} = \langle (y_m - \hat{y}_m) \otimes (x_n - \hat{x}_n) \rangle \text{ and } c^{xy}_{m,n} = (c^{yx}_{m,n})^T . \end{aligned} \tag{2.3.19b}$$

Furthermore, if the internal coordinates are of the form $x_n = (u_n, v_n) \in \mathbb{R}^6$, as are the ones described in (2.1.18), and $y_n = (\eta_n, \mathsf{w}_n) \in \mathbb{R}^6$, as are the ones described in (2.1.33), then the expectations in (2.3.18b) naturally decompose into

$$\hat{x}_n = \begin{pmatrix} \hat{u}_n \\ \hat{v}_n \end{pmatrix} \text{ and } \hat{y}_n = \begin{pmatrix} \hat{\eta}_n \\ \hat{\mathsf{w}}_n \end{pmatrix} \tag{2.3.20}$$

and consequently the blocks in (2.3.19b) decompose into 3-by-3 sub-blocks denoted

$$c^{yy}_{m,n} = \begin{pmatrix} c^{\eta\eta}_{m,n} & c^{\eta\mathsf{w}}_{m,n} \\ c^{\mathsf{w}\eta}_{m,n} & c^{\mathsf{ww}}_{m,n} \end{pmatrix} , \, c^{xx}_{m,n} = \begin{pmatrix} c^{uu}_{m,n} & c^{uv}_{m,n} \\ c^{vu}_{m,n} & c^{vv}_{m,n} \end{pmatrix} \text{ and } c^{yx}_{m,n} = \begin{pmatrix} c^{\eta u}_{m,n} & c^{\eta v}_{m,n} \\ c^{\mathsf{w}u}_{m,n} & c^{\mathsf{w}v}_{m,n} \end{pmatrix} \tag{2.3.21}$$

where we have used the same notation as in (2.3.19b).

In the case of a single chain, the first and second moment of internal coordinates x are defined similarly to (2.3.18) and (2.3.19). Moreover, if the chain has a local internal

energy, with the concomitant factorisation of the configurational distribution $d\rho(\boldsymbol{g})$ in the form of (2.3.5), then the local internal chain coordinates x_n are statistically independent and the covariance matrix c^{xx} is 6-by-6 block diagonal since

$$\langle (x_n - \hat{x}_n) \otimes (x_m - \hat{x}_m) \rangle = 0 \tag{2.3.22}$$

for any $m \neq n$.

We would like to stress that the above definition of the first and second moments of internal chain, and bichain, coordinates show that different choices of coordinates lead to different values of the expectation $(\hat{x}, \hat{y})$ and of the covariance matrix c, despite that underlying statistical physics is not different. In particular the knowledge of the configurational distribution in any set of coordinates allows to compute, in principle, the first and second moments in any other coordinates.

In the context of DNA modelling, the sequence dependence of the values for the first and second moments of the internal coordinates, in (2.3.18) and (2.3.19), have been particularly studied and have been used to build sequence dependent normal distributions for chain configurations [Marky and Olson, 1994, Olson et al., 1998, Coleman et al., 2003, Gonzalez and Maddocks, 2001, Becker and Everaers, 2007, Lankaš et al., 2009] and bichain configurations [Lankaš et al., 2009, Gonzalez et al., 2013, Petkeviciute et al., 2014]. More details about these distributions are presented in section 5.2.

2.4 The Persistence Matrix for a Heterogenous Rigid Body Chain

A detailed analysis of the configurational chain first moment is presented in this section. We discuss more precisely how it encodes other frequently adopted polymer observables and in particular the frame correlation matrix and the Flory persistence vector. The notion of the expected chain configuration and of the fluctuation matrix allows us to state a formal definition of a semi-flexible regime for chains and to derive a first order expansion for the configurational first moment. This analytical expression emphasises the equal role of the expected chain and the fluctuation matrix and leads to a general description of the configurational first moment behaviour for non degenerate chains. In particular, the frame correlation matrix is shown to be generally asymptotically vanishing as a function of the molecular distance, but in a non-monotonic way, and that the components of the Flory persistence vector converge to finite values. Moreover, the explicit formula motivates the factorisation of the configurational first moment into the expected chain and the persistence matrix, whose blocks characterise the memory lost as well as the limiting values of the Flory persistence vector. Furthermore, the formula obtained can be also expanded in terms of the molecular length and lead to the discussion of short and long length approximations. The short length expansion delivers explicit closed form solutions.

2.4.1 Recurrence relation for the chain configurational first moment

As defined in section 2.3.2, the *first moment of the relative rigid body motion* from the m^{th} and n^{th} rigid body in the chain, or just the *configurational first moment*, is denoted by $\langle \boldsymbol{g}_{m,n} \rangle = (\langle \boldsymbol{R}_{m,n} \rangle, \langle \boldsymbol{r}_{m,n} \rangle)$. It is of particular interest since it encodes some observables which are frequently used to characterise polymer statistical physics properties, such as:

(a) The *Kratky-Porod persistence length*, which is the component of the vector $\langle \boldsymbol{r}_{m,\infty} \rangle$ in the direction of the first translation vector. It was proposed, following the work of M. Smoluchowski on stochastic trajectories, as a characteristic value for a chain [Porod, 1948, Krakty and Porod, 1949, Flory, 1969, Doi and Edwards, 1986].

(b) The *Flory persistence vector* which is the triple $\langle \boldsymbol{r}_{m,n} \rangle$. It generalises the Kratky-Porod persistence length [Flory, 1973, Schellman and Harvey, 1995].

(c) The *tangent-tangent correlation* which is defined as $t_m \cdot \langle \boldsymbol{R}_{m,n} \rangle \, t_n$ where $t_m \in \mathbb{R}^3$ and $t_n \in \mathbb{R}^3$ describe a given and deterministic direction in the frames $\boldsymbol{R}_m$ and $\boldsymbol{R}_n$ respectively. In some very specific cases this quantity can be simply related to the Kratky-Porod persistence length and to some bending parameters of the chain [Landau and Lifshitz, 1959, Schellman, 1974, Doi and Edwards, 1986, Yamakawa, 1997]. We note, moreover, that tangent-tangent correlation is in general very different from the *chord-chord correlation* $\langle (\boldsymbol{r}_{m,m+1} - \boldsymbol{r}_{m,m}) \cdot (\boldsymbol{r}_{m,n+1} - \boldsymbol{r}_{m,n}) \rangle$ because of the possible presence of extension and shear deformations.

(d) The *frame correlation matrix* which is the matrix $\langle \boldsymbol{R}_{m,n} \rangle$ that contains all the cross correlations between the director frames indexed by m and n. It generalises the tangent-tangent correlation [Maroun and Olson, 1988, Panyukov and Rabin, 2000].

(e) The *mean squared end-to-end distance* is strongly related to configurational first moments, even if in general it cannot be directly deduced from the entries in $\langle \boldsymbol{g}_{m,n} \rangle$ [Flory, 1969, Marky and Olson, 1994].

In order to study some properties of the first moment $\langle \boldsymbol{g}_{m,n} \rangle$, we concentrate for the rest of this section on the case of rigid body chains with local internal energy E, as defined in (2.2.5). According to this hypothesis, the configurational distribution $d\rho(\boldsymbol{g})$ factorizes into local distributions over junction displacements a_n, as presented in (2.3.4). Consequently, the recurrence relation for the relative rigid body displacement $\boldsymbol{g}_{m,n}$ in (2.1.9) leads to the following recurrence relation for the first moment

$$\langle \boldsymbol{g}_{m,n+1} \rangle = \langle \boldsymbol{g}_{m,n} \rangle \langle a_n \rangle \tag{2.4.1}$$

where

$$\langle a_n \rangle = \int_{\mathbb{R}^6} a(x_n) d\rho_n(a(x_n)) \tag{2.4.2}$$

in terms of the chain internal coordinates (2.1.11), or equivalently as

$$\langle a_n \rangle = \int_{\mathrm{SE}(3)} a(x_n) d\rho_n(a_n) \tag{2.4.3}$$

according to the distribution in (2.3.5). More generally, the relation (2.4.1) gives

$$\langle \boldsymbol{g}_{m,n} \rangle = \langle \boldsymbol{g}_{m,k} \rangle \langle \boldsymbol{g}_{k,n} \rangle \tag{2.4.4}$$

for all $1 \leq m \leq k \leq n \leq N$ according to the transition identity (2.1.10). In terms of the moments $\langle \boldsymbol{R}_{m,n} \rangle$ and $\langle \boldsymbol{r}_{m,n} \rangle$ the recurrence relation (2.4.1) reads

$$\langle \boldsymbol{R}_{m,n+1} \rangle = \langle \boldsymbol{R}_{m,n} \rangle \langle Q_n \rangle \tag{2.4.5a}$$

$$\langle \boldsymbol{r}_{m,n+1} \rangle = \langle \boldsymbol{r}_{m,n} \rangle + \langle \boldsymbol{R}_{m,n} \rangle \langle q_n \rangle \tag{2.4.5b}$$

for $1 \leq m \leq n \leq N$. The definition of the moment $\langle \boldsymbol{R}_{m,n} \rangle$ is originally found in [Eyring, 1932] and is extensively discussed in [Flory, 1969] as the *matrix multiplication method* [Flory and Jernigan, 1965, Flory and Miller, 1966]. This method is then generalised to the matrix $\langle \boldsymbol{g}_{m,n} \rangle$ in [Flory, 1973].

Consequently, one of the key steps in the understanding of the observables listed in (a) to (e) above in the case of rigid body chains with local energy is the explicit knowledge of the first moment of the junction displacement $\langle a_n \rangle$ in (2.4.2) [Flory, 1969, Flory, 1973, Maroun and Olson, 1988, Marky and Olson, 1994]. It would be even more valuable to be able to write the matrix $\langle a_n \rangle$ directly as a function of the Boltzmann distribution

$d\rho_n(a_n)$ for each n, or equivalently as a function of the local potential energy w_n, as defined in (2.2.5), for each n [Eyring, 1932, Schellman, 1980, Schellman and Harvey, 1995]. Exact analytical values of the entries of the matrix $\langle a_n \rangle$ are, for most of the realistic Boltzmann distribution $d\rho_n(a(x_n))$, still challenging to obtain. We will however show that an approximation can be provided for general local chain configurational distributions in the *semi-flexible regime*. We observe that the matrix multiplication method in (2.4.1) is no longer valid for *non-local chain energies* [Eyring, 1932, Flory, 1969], or even for general nearest neighbor bichain distribution $d\rho(\boldsymbol{g}, \mathcal{P})$, which emphasises the difficulty of studying analytically the configurational first moment $\langle \boldsymbol{g}_{m,n} \rangle$ for realistic chain models.

2.4.2 An explicit expansion for semi-flexible chains

In order to approximate the junction displacement expectation $\langle a_n \rangle$ in (2.4.2), it is convenient to first introduce the *expected junction* $\bar{a}_n \in \mathrm{SE}(3)$ which has the following block structure

$$\bar{a}_n = \begin{pmatrix} \bar{Q}_n & \bar{q}_n \\ 0 & 1 \end{pmatrix} \tag{2.4.6}$$

analogous to (2.1.5). The expected junction is defined as follows. We factorize the junction displacement a_n in the form

$$a_n = \alpha_n \bar{a}_n \tag{2.4.7}$$

and we expand the matrix α_n, according to (1.3.4), as

$$\alpha_n = \mathrm{Id}_4 + \mathcal{T}\Theta_n + \frac{1}{2}(\mathcal{T}\Theta_n)^2 + \mathrm{o}\left(|\Theta_n|^2\right) \tag{2.4.8}$$

where we recall that the operator $\mathcal{T}$ stands for the *tangent map* defined as

$$\mathcal{T}\Theta_n = \begin{pmatrix} [\Theta_n^{\boldsymbol{R}}\times] & \Theta_n^{\boldsymbol{r}} \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{4\times 4} \text{ for } \Theta_n = (\Theta_n^{\boldsymbol{R}}, \Theta_n^{\boldsymbol{r}}) \in \mathbb{R}^6 \tag{2.4.9}$$

with the superscripts $^{\boldsymbol{R}}$ and $^{\boldsymbol{r}}$ denoting the rotational and translational parts of the vector Θ_n respectively, as in (1.1.22). The matrix a_n is then described by the deformation α_n from a given junction configuration $\bar{a}_n$ and, moreover, the variable Θ_n can be interpreted as a parametrization of linear perturbations around $\bar{a}_n$. The expected junction $\bar{a}_n$ is then defined as the junction displacement in (2.4.7) such that

$$\langle \Theta_n \rangle = 0 \tag{2.4.10}$$

where the expectation is taken in the sense of (2.3.10). The expected junction can then be seen as a junction displacement such that the expectation of the linear perturbation vanishes. The expected junction always exists, and it is moreover unique if the distribution of

the a_n is sufficiently concentrated [Kendall, 1990, Becker and Everaers, 2007, Chirikjian, 2011]. For completeness, we should note that the expression of the junction in (2.4.7) is the one used in [Becker and Everaers, 2007, Becker, 2007] to parametrize the junction displacement in a rigid body chain, and that the expansion (2.4.8) can be interpreted as higher order terms for the exponential coordinates of the deformation matrix α_n.

The *expected chain* $\bar{\boldsymbol{g}}$ is the rigid body chain configuration generated by the set of expected junction $\bar{a}_1, ..., \bar{a}_{N-1}$. More generally, the *expected chain relative rigid body displacement* is defined, in analogy to (2.1.7), by

$$\bar{\boldsymbol{g}}_{m,n+1} = \bar{\boldsymbol{g}}_{m,n}\bar{a}_n \text{ with } \bar{\boldsymbol{g}}_{m,m} = \mathrm{Id}_4. \tag{2.4.11}$$

The block structure of $\bar{\boldsymbol{g}}_{m,n}$ is denoted as

$$\bar{\boldsymbol{g}}_{m,n} = \begin{pmatrix} \bar{\boldsymbol{R}}_{m,n} & \bar{\boldsymbol{r}}_{m,n} \\ 0 & 1 \end{pmatrix} \tag{2.4.12}$$

where $\bar{\boldsymbol{R}}_{m,n}$ and $\bar{\boldsymbol{r}}_{m,n}$ denote respectively the *expected chain relative rotation* and the *expected chain relative translation* from the m^{th} to the n^{th} element in the chain, as in (2.1.7).

It is crucial to observe that for a given choice of rigid body chain coordinates x, one does not in general have that $a(\hat{x}_n) = \bar{a}_n$ where $\hat{x}_n$ denotes the expectation of the coordinates x_n, as defined in (2.3.18b). To compute the expected junction $\bar{a}_n$ from a distribution $d\rho_n(a(x_n))$ is then generally non-trivial and a priori requires a numerical approximation. However, in the case of a sufficiently concentrated normal distribution $d\rho_n(a(x_n))$ in (2.4.2), an analytical approximation can be derived. This construction is presented in appendix A.5.

We now need to define formally the notion of the *semi-flexible regime* for rigid body chains. Observing that the expansion (2.4.8) allows the expectation of the junction deformation α_n to be written as

$$\langle \alpha_n \rangle = \mathrm{Id}_4 + \frac{1}{2}\left\langle (\mathcal{T}\Theta_n)^2 \right\rangle + \left\langle \mathrm{o}\left(|\Theta_n|^2\right) \right\rangle \tag{2.4.13}$$

we say that the corresponding chain is in the semi-flexible regime if there exists a small parameter $\epsilon > 0$ such that

$$\left| \left\langle (\mathcal{T}\Theta_n)^2 \right\rangle \right| = \mathrm{O}(\epsilon) \text{ and } \left\langle \mathrm{o}\left(|\Theta_n|^2\right) \right\rangle = \mathrm{o}(\epsilon) \tag{2.4.14}$$

for all $n = 1, .., N-1$ in the expression (2.4.13). This definition implies that the deformation matrix α_n is on average close to the identity and moreover that its expectation can be well approximated by truncation at the first order. According to the definition (2.4.9), one finds the identity

$$(\mathcal{T}\Theta_n)^2 = \begin{pmatrix} \left[\Theta_n^{\boldsymbol{R}}\times\right]^2 & \Theta_n^{\boldsymbol{R}} \times \Theta_n^{\boldsymbol{r}} \\ 0 & 0 \end{pmatrix}. \tag{2.4.15}$$

Defining the *local chain fluctuation matrix* $\mathcal{C}_n$ together with its block structure

$$\mathcal{C}_n = \begin{pmatrix} \mathcal{C}_n^{\boldsymbol{RR}} & \mathcal{C}_n^{\boldsymbol{Rr}} \\ \mathcal{C}_n^{\boldsymbol{Rr}T} & \mathcal{C}_n^{\boldsymbol{rr}} \end{pmatrix} \tag{2.4.16a}$$

where

$$\mathcal{C}_n^{\boldsymbol{RR}} = \left\langle \Theta_n^{\boldsymbol{R}} \otimes \Theta_n^{\boldsymbol{R}} \right\rangle \, , \, \mathcal{C}_n^{\boldsymbol{Rr}} = \left\langle \Theta_n^{\boldsymbol{R}} \otimes \Theta_n^{\boldsymbol{r}} \right\rangle \text{ and } \mathcal{C}_n^{\boldsymbol{rr}} = \langle \Theta_n^{\boldsymbol{r}} \otimes \Theta_n^{\boldsymbol{r}} \rangle \tag{2.4.16b}$$

leads then to the expression

$$\left\langle (\mathcal{T}\Theta_n)^2 \right\rangle = - \begin{pmatrix} \mathrm{Tr}\left(\mathcal{C}_n^{\boldsymbol{RR}}\right) \mathrm{Id}_3 - \mathcal{C}_n^{\boldsymbol{RR}} & \mathrm{Vect}\left(\mathcal{C}_n^{\boldsymbol{Rr}}\right) \\ 0 & 0 \end{pmatrix} \tag{2.4.17}$$

where the operator $\mathrm{Vect}\,(\,.\,)$ is defined in (1.1.13). We note that to deduce (2.4.17) we have used the following identities for any triples $y, y_1, y_2 \in \mathbb{R}^3$

$$[y\times]^2 = y \otimes y - (y \cdot y)\mathrm{Id}_3 \tag{2.4.18}$$

and

$$y_1 \times y_2 = \mathrm{Vect}\,(y_2 \otimes y_1)\,. \tag{2.4.19}$$

In the definition (2.4.14) one can then replace the first estimation by the following ones in terms of the blocks of the fluctuation matrices

$$\left|\mathcal{C}_n^{\boldsymbol{RR}}\right|_{\mathrm{Sp}} = \mathrm{O}\,(\epsilon) \text{ and } \left|\mathrm{Vect}\left(\mathcal{C}_n^{\boldsymbol{Rr}}\right)\right| = \mathrm{O}\,(\epsilon) \tag{2.4.20}$$

for all $n = 1, ..., N$ where $\left|\mathcal{C}_n^{\boldsymbol{RR}}\right|_{\mathrm{Sp}}$ denotes the largest eigenvalue of the matrix $\mathcal{C}_n^{\boldsymbol{RR}}$, as presented in (1.1.4).

Under the hypothesis of the rigid body chain being in the semi-flexible regime, one then obtains the following asymptotic expansion of the deformation matrix expectation

$$\langle \alpha_n \rangle = \mathrm{Id}_4 + \frac{1}{2}\left\langle (\mathcal{T}\Theta_n)^2 \right\rangle + \mathrm{o}\,(\epsilon) \tag{2.4.21}$$

and consequently, using the factorisation (2.4.7) together with the expansion (2.4.21) and the identity (2.4.17), the first moment of the junction displacement $\langle a_n \rangle$ reads

$$\langle a_n \rangle = \left(\mathrm{Id}_4 - \frac{1}{2}\Lambda_n\right) \bar{a}_n + \mathrm{o}\,(\epsilon) \tag{2.4.22}$$

where we have defined for compactness

$$\Lambda_n = \begin{pmatrix} \mathrm{Tr}\left(\mathcal{C}_n^{\boldsymbol{RR}}\right) \mathrm{Id}_3 - \mathcal{C}_n^{\boldsymbol{RR}} & \mathrm{Vect}\left(\mathcal{C}_n^{\boldsymbol{Rr}}\right) \\ 0 & 0 \end{pmatrix}. \tag{2.4.23}$$

We denote the sub-blocks of Λ_n by respectively

$$\mathrm{Tr}\left(\mathcal{C}_n^{\boldsymbol{RR}}\right)\mathrm{Id}_3 - \mathcal{C}_n^{\boldsymbol{RR}} = \Lambda_n^{\boldsymbol{RR}} \text{ and } \mathrm{Vect}\left(\mathcal{C}_n^{\boldsymbol{Rr}}\right) = \Lambda_n^{\boldsymbol{Rr}} \tag{2.4.24}$$

and observe that the assumption (2.4.14), or equivalently (2.4.20), implies that

$$\left|\Lambda_n^{\boldsymbol{RR}}\right|_{\mathrm{Sp}} = \mathrm{O}\left(\epsilon\right) \text{ and } \left|\Lambda_n^{\boldsymbol{Rr}}\right| = \mathrm{O}\left(\epsilon\right). \tag{2.4.25}$$

Expansion (2.4.22) then provides a way to factorize the first moment of the junction displacement $\langle a_n \rangle$ into the expected junction $\bar{a}_n$ and a first order perturbation. The configurational first moment $\langle \boldsymbol{g}_{m,n} \rangle$, as described in (2.4.1), can then be explicitly written using (2.4.22) as the recurrence relation

$$\langle \boldsymbol{g}_{m,n+1} \rangle = \langle \boldsymbol{g}_{m,n} \rangle \left(\mathrm{Id}_4 - \frac{1}{2}\Lambda_n\right)\bar{a}_n + \mathrm{o}\left(\epsilon\right) \text{ with } \boldsymbol{g}_{m,m} = \mathrm{Id}_4 \tag{2.4.26}$$

or, equivalently in terms of $\langle \boldsymbol{R}_{m,n} \rangle$ and $\langle \boldsymbol{r}_{m,n} \rangle$, as

$$\langle \boldsymbol{R}_{m,n+1} \rangle = \langle \boldsymbol{R}_{m,n} \rangle \left(\mathrm{Id}_3 - \frac{1}{2}\Lambda_n^{\boldsymbol{RR}}\right)\bar{Q}_n + \mathrm{o}\left(\epsilon\right) \text{ with } \boldsymbol{R}_{m,m} = \mathrm{Id}_3 \tag{2.4.27a}$$

$$\langle \boldsymbol{r}_{m,n+1} \rangle = \langle \boldsymbol{r}_{m,n} \rangle + \langle \boldsymbol{R}_{m,n} \rangle \left[\left(\mathrm{Id}_3 - \frac{1}{2}\Lambda_n^{\boldsymbol{RR}}\right)\bar{q}_n - \frac{1}{2}\Lambda_n^{\boldsymbol{Rr}}\right] + \mathrm{o}\left(\epsilon\right) \text{ with } \boldsymbol{r}_{m,m} = 0 \tag{2.4.27b}$$

where $\bar{Q}_n$ and $\bar{q}_n$ are defined in (2.4.6), and $\Lambda_n^{\boldsymbol{RR}}$ and $\Lambda_n^{\boldsymbol{Rr}}$ are defined in (2.4.24).

These expressions generalise the work in [Schellman and Harvey, 1995] using an isotropic *wedge model* on the *static and dynamic contributions*, as originally proposed in [Trifonov et al., 1988], to the average chain configuration. For completeness, we note that in the context of continuum chains similar expressions to (2.4.26) have been deduced, in the constant coefficient and diagonal case [Panyukov and Rabin, 2000] and in the general case [Becker, 2007]. One should emphasize that these recurrence relations stress the equal importance of the expected chain configuration $\bar{\boldsymbol{g}}_n$ and of the local covariance matrices $\mathcal{C}_n$ on the chain first moment $\langle \boldsymbol{g}_{m,n} \rangle$, and consequently on the Flory persistence vector and on the frame correlation matrix for instance. Moreover the characteristic value of the parameter ϵ can be chosen as

$$\epsilon = \max_{1 \leq n \leq N-1} \left\{ \frac{4\left|\mathcal{C}_n^{\boldsymbol{RR}}\right|_{\mathrm{Sp}}}{\pi^2}, \frac{2\left|\mathrm{Vect}\left(\mathcal{C}_n^{\boldsymbol{Rr}}\right)\right|}{\pi\left|\bar{q}\right|} \right\} \tag{2.4.28}$$

where $\frac{\pi}{2}$ and $|\bar{q}| = \max_{1 \leq n \leq N-1} |\bar{q}_n|$ have been chosen as characteristic scales for the rotations and translations respectively. Applications of the expressions in (2.4.26), or equivalently in (2.4.27), are presented in section 7.2.1 in the context of coarse grained DNA modelling. By comparing their predictions with the results given by a full Monte Carlo simulation, we conclude that they provide a remarkably good analytical approximation for semi-flexible chains.

2.4.3 General behaviour of semi-flexible chain first moment

In contrast to what is widely assumed, the expression (2.4.26) shows that the entries in the matrix $\langle \boldsymbol{R}_{m,n} \rangle$, as well as the components of the vector $\langle \boldsymbol{r}_{m,n} \rangle$, do *not* exhibit in general a monotone behaviour for increasing values of n, mainly because the expected junction $\bar{a}_n$ is in general far from the identity Id_4, as for example in coarse grain models of DNA molecule where the intrinsic helicity, as well as local intrinsic bends, can be quite large. In particular the tangent-tangent correlation $t_m \cdot \langle \boldsymbol{R}_{m,n} \rangle \, t_n$ should not in general be expected to be exponentially decaying and can even become negative. Moreover, according the recurrence relation (2.4.26), this quantity is not only a function of the bending parameters of the chain, but also of the expected chain configuration. This fact has already be observed in the case of simple chain models in [Schellman and Harvey, 1995, Yamakawa, 1997, Panyukov and Rabin, 2000, Vaillant, 2001, Liu et al., 2011], and in the general case in [Becker, 2007]. Similarly, the third component of the vector $\langle \boldsymbol{r}_{m,\infty} \rangle$ is only a partial measure of the chain persistence, since its first and second component do not vanish in general, and can also be negative, as noted in [Flory, 1973, Maroun and Olson, 1988, Schellman and Harvey, 1995, Yamakawa, 1997, Becker, 2007].

In contrast, the relation (2.4.27a) implies that the spectral norm of the matrix $\langle \boldsymbol{R}_{m,n} \rangle$, as defined in (1.1.4), is, at leading order, decaying for increasing values of n since

$$|\langle \boldsymbol{R}_{m,n+1} \rangle|_{\mathrm{Sp}} \leq |\langle \boldsymbol{R}_{m,n} \rangle|_{\mathrm{Sp}} \left| \left(\mathrm{Id}_3 - \frac{1}{2} \Lambda_n^{\boldsymbol{RR}} \right) \right|_{\mathrm{Sp}} + \mathrm{o}\,(\epsilon) \leq |\langle \boldsymbol{R}_{m,n} \rangle|_{\mathrm{Sp}} + \mathrm{o}\,(\epsilon) \tag{2.4.29}$$

and the block $\Lambda_n^{\boldsymbol{RR}} = \mathrm{Tr}\left(\mathcal{C}_n^{\boldsymbol{RR}}\right) \mathrm{Id}_3 - \mathcal{C}_n^{\boldsymbol{RR}}$, as defined in (2.4.24), is always positive definite. More generally, one has then

$$|\langle \boldsymbol{R}_{m,n} \rangle|_{\mathrm{Sp}} \leq (1 - \lambda^{\min})^{n-m} + \mathrm{o}\left(\frac{1 - (1 - \lambda^{\min})^{n-m}}{\lambda^{\min}} \epsilon \right) \tag{2.4.30}$$

where $\lambda^{\min} = \min_{m \leq k \leq n} \lambda_k^{\min}$ and $\lambda_k^{\min}$ denotes the minimum eigenvalue of $\frac{1}{2}\Lambda_k^{\boldsymbol{RR}}$. By definition of the matrix $\Lambda_k^{\boldsymbol{RR}}$ the value of $\lambda^{\min}$ can be shown to be

$$\lambda^{\min} = \min_{m \leq k \leq n} \frac{1}{2} (\gamma_{1,k} + \gamma_{2,k}) \tag{2.4.31}$$

where $\gamma_{1,k}$ and $\gamma_{2,k}$ denote the two smallest eigenvalues of $\mathcal{C}_k^{\boldsymbol{RR}}$. We observe in particular that the value of $\lambda^{\min}$ vanishes if and only if at least one of the matrices $\mathcal{C}_k^{\boldsymbol{RR}}$ has exactly two zero eigenvalues. A rigid body chain with $\lambda^{\min} = 0$ is called a *degenerate chain*. Expression (2.4.30) allows us to deduce that if the chain is non degenerate, then the spectral norm $|\langle \boldsymbol{R}_{m,n} \rangle|_{\mathrm{Sp}}$ is, at leading order, strictly decaying and all frame cross-correlations are asymptotically vanishing, i.e.

$$\langle \boldsymbol{R}_{m,n} \rangle \to 0 \text{ for } n - m >> 1. \tag{2.4.32}$$

Using the recurrence relation (2.4.27b) together with the expression (2.4.30) we obtain, at leading order, the estimate

$$|\langle \boldsymbol{r}_{m,n+1} \rangle - \langle \boldsymbol{r}_{m,n} \rangle| \leq (1 - \lambda^{\min})^{n-m} \left[(1 - \lambda^{\min}) \, |\bar{q}| + \frac{1}{2} \left| \Lambda^{\boldsymbol{Rr}} \right| \right] \tag{2.4.33}$$

where we have used the notation $|\bar{q}| = \max_{m\leq k\leq n} |\bar{q}_k|$ and $\left|\Lambda^{\boldsymbol{Rr}}\right| = \max_{m\leq k\leq n} \left|\Lambda_k^{\boldsymbol{Rr}}\right|$. For non-degenerate chains, the relation (2.4.33) gives then

$$|\langle \boldsymbol{r}_{m,n}\rangle| \leq \frac{1-(1-\lambda^{\min})^{n-m}}{\lambda^{\min}} \left[(1-\lambda^{\min})\,|\bar{q}| + \frac{1}{2}\left|\Lambda^{\boldsymbol{Rr}}\right|\right] \tag{2.4.34}$$

and, consequently, the three components of the relative translation vector $\langle \boldsymbol{r}_{m,n}\rangle$ converge to finite values $\langle \boldsymbol{r}_{m,\infty}\rangle$. Precisely we have that

$$\langle \boldsymbol{r}_{m,n}\rangle \to \langle \boldsymbol{r}_{m,\infty}\rangle \text{ for } n-m >> 1 \tag{2.4.35}$$

with

$$|\langle \boldsymbol{r}_{m,\infty}\rangle| \leq \frac{1}{\lambda^{\min}} \left[(1-\lambda^{\min})\,|\bar{q}| + \left|\Lambda^{\boldsymbol{Rr}}\right|\right]. \tag{2.4.36}$$

Statements (2.4.32) and (2.4.35) constitute general convergence results for non-degenerate semi-flexible chains. They show that the larger the value of $m-n$ is, the less *correlated* is the relative orientation $\boldsymbol{R}_n$ from $\boldsymbol{R}_m$, and moreover that the limiting value of the Flory persistence vector is always finite. Even if these facts seem very natural, to our knowledge no precise and general argument has previously been presented in the literature.

Finer statements about the behaviour of the first moment of the relative rigid body motion $\langle \boldsymbol{g}_{m,n}\rangle$ requires more detailed information about the structure of the chain, i.e. about the expected chain $\bar{\boldsymbol{g}}_{m,n}$ and the values of the fluctuation matrix $\mathcal{C}_n$. Numerical studies of the behaviour of both the entries in $\langle \boldsymbol{R}_{m,n}\rangle$ and in $\langle \boldsymbol{r}_{m,n}\rangle$ has been performed, for instance in [Maroun and Olson, 1988, Marky and Olson, 1994, Liu et al., 2011] in the context of DNA modelling, by using Monte Carlo methods to approximate the first moment of the junction displacement $\langle a_n\rangle$.

In the specific case of a homogeneous chain, i.e. $\langle a_n\rangle = \langle a\rangle$ for all $1 \leq n \leq N$, analytical formulas can obtained for the matrix $\langle \boldsymbol{R}_{m,n}\rangle$ and for the component of the vector $\langle \boldsymbol{r}_{m,n}\rangle$ as a function of $m-n$ and of the entries of the matrix $\langle a\rangle$. This allows the computation of the triple $\langle \boldsymbol{r}_{m,\infty}\rangle$ [Flory, 1969]. The analytical approximation presented in (2.4.22) leads then to the following expression for homogeneous semi-flexible chains

$$\langle \boldsymbol{R}_{m,n}\rangle = \langle Q\rangle^{n-m} \tag{2.4.37a}$$

$$\langle \boldsymbol{r}_{m,n}\rangle = \left(\mathrm{Id}_3 - \langle Q\rangle^{m-n}\right)\left(\mathrm{Id}_3 - \langle Q\rangle\right)^{-1}\langle q\rangle \tag{2.4.37b}$$

where

$$\langle Q\rangle = \left(\mathrm{Id}_3 - \frac{1}{2}\Lambda^{\boldsymbol{RR}}\right)\bar{Q} \tag{2.4.38a}$$

$$\langle q\rangle = \left(\mathrm{Id}_3 - \frac{1}{2}\Lambda^{\boldsymbol{RR}}\right)\bar{q} - \frac{1}{2}\Lambda^{\boldsymbol{Rr}} \tag{2.4.38b}$$

The limiting values of the Flory persistence vector is then,

$$\langle \boldsymbol{r}_{m,\infty}\rangle = \left(\mathrm{Id}_3 - \langle Q\rangle\right)^{-1}\langle q\rangle. \tag{2.4.39}$$

2.4.4 Global factorization and the persistence matrix

Even if the use of the recurrence relation (2.4.26) allows the deduction of a few general statements about the behaviour of the first moment $\langle \boldsymbol{g}_{m,n} \rangle$ for any $n \geq m$, a finer understanding of the detailed features in the entries of $\langle \boldsymbol{g}_{m,n} \rangle$ is still missing. Some interesting insights are obtained from the factorization of the expected chain configuration in the configurational chain first moment.

In analogy to (2.4.7) we can express the relative rigid body displacement $\boldsymbol{g}_{m,n}$ in the form

$$\boldsymbol{g}_{m,n} = \boldsymbol{\mathcal{D}}_{m,n} \bar{\boldsymbol{g}}_{m,n} \tag{2.4.40}$$

where $\bar{\boldsymbol{g}}_{m,n}$ is presented in (2.4.11) and where $\boldsymbol{\mathcal{D}}_{m,n}$ will be called the *deformation matrix*. Consequently the first moment $\langle \boldsymbol{g}_{m,n} \rangle$ can then be expressed as

$$\langle \boldsymbol{g}_{m,n} \rangle = \langle \boldsymbol{\mathcal{D}}_{m,n} \rangle \, \bar{\boldsymbol{g}}_{m,n}. \tag{2.4.41}$$

Writing the block structure of $\langle \boldsymbol{\mathcal{D}}_{m,n} \rangle$ as

$$\langle \boldsymbol{\mathcal{D}}_{m,n} \rangle = \begin{pmatrix} \langle \boldsymbol{\Delta}_{m,n} \rangle & \langle \boldsymbol{\delta}_{m,n} \rangle \\ 0 & 1 \end{pmatrix} \tag{2.4.42}$$

gives the following expressions of the first moment of the relative rotation $\langle \boldsymbol{R}_{m,n} \rangle$ and the relative translation $\langle \boldsymbol{r}_{m,n} \rangle$

$$\langle \boldsymbol{R}_{m,n} \rangle = \langle \boldsymbol{\Delta}_{m,n} \rangle \, \bar{\boldsymbol{R}}_{m,n} \tag{2.4.43a}$$

$$\langle \boldsymbol{r}_{m,n} \rangle = \langle \boldsymbol{\delta}_{m,n} \rangle + \langle \boldsymbol{\Delta}_{m,n} \rangle \, \bar{\boldsymbol{r}}_{m,n} \tag{2.4.43b}$$

where we have used the block structure of $\bar{\boldsymbol{g}}_{m,n}$ defined in (2.4.12). Expression (2.4.41), or equivalently (2.4.43), decomposes the first moment of the relative rigid body displacement $\langle \boldsymbol{g}_{m,n} \rangle$ into the expected chain displacement $\bar{\boldsymbol{g}}_{m,n}$ and the first moment of the deformation matrix $\langle \boldsymbol{\mathcal{D}}_{m,n} \rangle$.

Using the explicit approximation of the junction expectation $\langle a_n \rangle$ in (2.4.22) and the recurrence relation for $\bar{\boldsymbol{g}}_{m,n}$ in (2.4.11), the sequence of matrices $\langle \boldsymbol{\mathcal{D}}_{m,n} \rangle$ can be expressed, up to $\mathrm{o}(\epsilon)$, through their own recurrence relation for $n \geq m$ by

$$\langle \boldsymbol{\mathcal{D}}_{m,n+1} \rangle = \langle \boldsymbol{\mathcal{D}}_{m,n} \rangle \left(\mathrm{Id}_4 - \frac{1}{2} \bar{\boldsymbol{g}}_{m,n} \Lambda_n \bar{\boldsymbol{g}}_{m,n}^{-1} \right) \text{ with } \boldsymbol{\mathcal{D}}_{m,m} = \mathrm{Id}_4 \tag{2.4.44}$$

or equivalently for the blocks in (2.4.42) by

$$\langle \boldsymbol{\Delta}_{m,n+1} \rangle = \langle \boldsymbol{\Delta}_{m,n} \rangle \left(\mathrm{Id}_3 - \frac{1}{2} \bar{\boldsymbol{R}}_{m,n} \Lambda_n^{\boldsymbol{RR}} \bar{\boldsymbol{R}}_{m,n}^T \right) \tag{2.4.45a}$$

$$\langle \boldsymbol{\delta}_{m,n+1} \rangle = \langle \boldsymbol{\delta}_{m,n} \rangle + \langle \boldsymbol{\Delta}_{m,n} \rangle \left[\frac{1}{2} \bar{\boldsymbol{R}}_{m,n} \Lambda_n^{\boldsymbol{RR}} \bar{\boldsymbol{R}}_{m,n}^T \bar{\boldsymbol{r}}_{m,n} - \frac{1}{2} \bar{\boldsymbol{R}}_{m,n} \Lambda_n^{\boldsymbol{Rr}} \right]. \tag{2.4.45b}$$

In contrast to the first moment $\langle \boldsymbol{R}_{m,n} \rangle$, the matrix $\langle \boldsymbol{\Delta}_{m,n} \rangle$ is a product of strictly positive definite matrices which leads to the estimates

$$(1-\lambda^{\max})\,|\langle \boldsymbol{\Delta}_{m,n}\rangle|_{\mathrm{Sp}} \leq |\langle \boldsymbol{\Delta}_{m,n+1}\rangle|_{\mathrm{Sp}} \leq (1-\lambda^{\min})\,|\langle \boldsymbol{\Delta}_{m,n}\rangle|_{\mathrm{Sp}} \tag{2.4.46}$$

where $\lambda^{\max} = \max_{m\leq k\leq n} \lambda_k^{\max}$ and $\lambda_k^{\max}$ denotes the maximum eigenvalue of $\frac{1}{2}\Lambda_k^{\boldsymbol{RR}}$, which can be computed analogously to $\lambda^{\min}$ in (2.4.31) using the two largest eigenvalues of the matrix $\mathcal{C}_k^{\boldsymbol{RR}}$. One then has

$$(1-\lambda^{\max})^{n-m} \leq |\langle \boldsymbol{\Delta}_{m,n}\rangle|_{\mathrm{Sp}} \leq (1-\lambda^{\min})^{n-m} \tag{2.4.47}$$

and consequently, for non-degenerate chains, the spectral norm of the matrix $\langle \boldsymbol{\Delta}_{m,n} \rangle$ is strictly decaying and is asymptotically vanishing, i.e.

$$|\langle \boldsymbol{\Delta}_{m,n+1}\rangle|_{\mathrm{Sp}} < |\langle \boldsymbol{\Delta}_{m,n}\rangle|_{\mathrm{Sp}} \text{ and } \langle \boldsymbol{\Delta}_{m,n}\rangle \to 0 \text{ for } n-m >> 1. \tag{2.4.48}$$

We observe moreover that the variations in the entries of the matrix $\langle \boldsymbol{\Delta}_{m,n} \rangle$ as a function of the index n have to be small since

$$|\langle \boldsymbol{\Delta}_{m,n+1}\rangle - \langle \boldsymbol{\Delta}_{m,n}\rangle|_{\mathrm{Sp}} \leq \frac{1}{2}\,|\langle \boldsymbol{\Delta}_{m,n}\rangle|_{\mathrm{Sp}}\,\left|\Lambda_n^{\boldsymbol{RR}}\right|_{\mathrm{Sp}} = \mathrm{O}\left((1-\lambda^{\min})^{n-m}\epsilon\right) \tag{2.4.49}$$

according to the hypothesis (2.4.25) and the estimate (2.4.47), whereas, in general, the variations in the entries of the relative rotation first moment only satisfy $|\langle \boldsymbol{R}_{m,n+1}\rangle - \langle \boldsymbol{R}_{m,n}\rangle| = \mathrm{O}\,(1)$. Consequently, for non-degenerate semi-flexible chains, the first moment of the relative rotation $\langle \boldsymbol{R}_{m,n} \rangle$ can always be factorized into a fast and non-decaying rotation matrix $\bar{\boldsymbol{R}}_{m,n}$ and another slowly varying and strictly decaying matrix $\langle \boldsymbol{\Delta}_{m,n} \rangle$.

Concerning the triple $\langle \boldsymbol{\delta}_{m,n} \rangle$, the definition (2.4.45b), using the estimation (2.4.47), gives for $n > m$

$$|\langle \boldsymbol{\delta}_{m,n}\rangle| \leq \sum_{k=m}^{n-1} (1-\lambda^{\min})^{k-m}\left[\lambda^{\max}\left|\bar{\boldsymbol{r}}_{m,k}\right| + \frac{1}{2}\left|\Lambda^{\boldsymbol{RR}}\right|\right] \tag{2.4.50}$$

which implies

$$\begin{aligned} |\langle \boldsymbol{\delta}_{m,n}\rangle| \leq \frac{1-(1-\lambda^{\min})^{n-m}}{\lambda^{\min}}\left[\frac{\lambda^{\max}}{\lambda^{\min}}(1-\lambda^{\min})\,|\bar{q}| + \left|\Lambda^{\boldsymbol{Rr}}\right|\right] \\ -\frac{\lambda^{\max}}{\lambda^{\min}}\,|\bar{q}|\,(n-m)(1-\lambda^{\min})^{n-m} \end{aligned} \tag{2.4.51}$$

where $|\bar{q}|$ is defined in (2.4.33) and where we have used $\left|\bar{\boldsymbol{r}}_{m,n}\right| \leq |\bar{q}|\,(n-m)$ for $n > m$. The limiting value $|\langle \boldsymbol{\delta}_{m,\infty}\rangle|$ is then finite and admits the upper bound

$$|\langle \boldsymbol{\delta}_{m,\infty}\rangle| \leq \frac{1}{\lambda^{\min}}\left(\frac{\lambda^{\max}}{\lambda^{\min}}(1-\lambda^{\min})\,|\bar{q}| + \left|\Lambda^{\boldsymbol{Rr}}\right|\right). \tag{2.4.52}$$

We observe moreover that

$$\langle \boldsymbol{\delta}_{m,\infty} \rangle = \langle \boldsymbol{r}_{m,\infty} \rangle \tag{2.4.53}$$

according to (2.4.43) and that we have the estimate

$$\begin{aligned} |\langle \boldsymbol{\delta}_{m,n+1} \rangle - \langle \boldsymbol{\delta}_{m,n} \rangle| &\leq \frac{1}{2} |\langle \boldsymbol{\Delta}_{m,n} \rangle|_{\mathrm{Sp}} \left(\left| \Lambda_n^{\boldsymbol{RR}} \right|_{\mathrm{Sp}} \left| \bar{\boldsymbol{r}}_{m,n} \right| + \left| \Lambda_n^{\boldsymbol{Rr}} \right| \right) \\ &= \mathrm{O}\left(\epsilon (1 - \lambda^{\min})^{n-m} \left(\left| \bar{\boldsymbol{r}}_{m,n} \right| + 1 \right) \right) \end{aligned} \tag{2.4.54}$$

using (2.4.45b). The first moment of the translation vector $\langle \boldsymbol{r}_{m,n} \rangle$ of non-degenerate semi-flexible chains can therefore be decomposed into the non-converging vector $\bar{\boldsymbol{r}}_{m,n}$ damped by the matrix $\langle \boldsymbol{\Delta}_{m,n} \rangle$ and by a convergent part $\langle \boldsymbol{\delta}_{m,n} \rangle$. Moreover, the triple $\langle \boldsymbol{\delta}_{m,n} \rangle$ is made up of a damped sum of the expected translation vector $\bar{\boldsymbol{r}}_{m,n}$ according to its definition (2.4.45b), which grows slower than $\bar{\boldsymbol{r}}_{m,n}$ and which has the same limiting values as the vector $\langle \boldsymbol{r}_{m,n} \rangle$.

We should notice that the estimation proposed in (2.4.47) and (2.4.51) are very coarse in general, and that they are used to depict qualitatively the behaviour of the first moment $\langle \boldsymbol{\Delta}_{m,n} \rangle$ and $\langle \boldsymbol{\delta}_{m,n} \rangle$ and then of $\langle \boldsymbol{R}_{m,n} \rangle$ and $\langle \boldsymbol{r}_{m,n} \rangle$. More precise estimation could be established when more information is known about the chain constitution.

The matrix $\langle \boldsymbol{\mathcal{D}}_{m,n} \rangle$ can be observed to have the interesting properties that the block $\langle \boldsymbol{\Delta}_{m,n} \rangle$ encodes how much of the expected chain configuration $\bar{\boldsymbol{g}}_{m,n}$ has persisted in the value of configurational first moment $\langle \boldsymbol{g}_{m,n} \rangle$, by definition in (2.4.43), and that the limiting value of $\langle \boldsymbol{\delta}_{m,n} \rangle$ behaves like the Flory persistence vector $\langle \boldsymbol{r}_{m,n} \rangle$, as shown in (2.4.53). We will consequently call the matrix $\langle \boldsymbol{\mathcal{D}}_{m,n} \rangle$ the *persistence matrix*. The factorization formula in (2.4.43) can be seen as a more sophisticated manifestation of the ideas proposed originally in [Trifonov et al., 1988], and then discussed in details in [Schellman and Harvey, 1995], which is that the behaviour of the configurational first moment $\langle \boldsymbol{g}_{m,n} \rangle$ is governed both by a deterministic, or *static*, part and by a stochastic, or *dynamic*, part, and it therefore generalises, in a mathematical sense, the expressions proposed in [Trifonov et al., 1988, Schellman and Harvey, 1995] in the context of DNA modelling for the static and dynamic *persistence lengths*. Application of this factorization of the configurational first moment into the expected chain configuration and the persistence matrix are presented in section 7.2.2 for a rigid base pair model of DNA.

2.4.5 Short and long length behaviour approximation

The approximate recurrence relation for the relative rigid body displacement first moment $\langle \boldsymbol{g}_{m,n} \rangle$, obtained in (2.4.26) for a chain in the semi-flexible regime, can be further investigated using the persistence matrix factorization in (2.4.41) by studying short and long length behaviours.

Under the semi-flexible regime hypothesis asserted in (2.4.25), it can be shown that the leading order term of the persistence matrix $\langle \boldsymbol{\mathcal{D}}_{m,n} \rangle$ of the recurrence relation (2.4.44)

for n sufficiently close to m is

$$\langle \boldsymbol{\mathcal{D}}_{m,n} \rangle = \exp\left\{ -\frac{1}{2} \sum_{k=m}^{n-1} \bar{\boldsymbol{g}}_{m,k} \Lambda_k \bar{\boldsymbol{g}}_{m,k}^{-1} \right\} + \mathrm{o}\left((n-m)\epsilon\right) \tag{2.4.55}$$

using the *Baker-Campbell-Hausdorff formula* for the exponential form of a product of non-commutating matrices, see for instance [Chirikjian, 2011], and combining with the fact that for any matrix $\in \mathbb{R}^{4\times 4}$

$$\ln\left(\mathrm{Id}_4 + A\right) = A + \mathrm{o}\left(|A|_{\mathrm{Sp}}\right). \tag{2.4.56}$$

More precisely, defining the matrices

$$\boldsymbol{\Omega}_{m,n} = \frac{1}{2} \sum_{k=m}^{n-1} \bar{\boldsymbol{R}}_{m,k} \Lambda_k^{\boldsymbol{RR}} \bar{\boldsymbol{R}}_{m,k}^T \tag{2.4.57a}$$

and the triples

$$\boldsymbol{\omega}_{m,n} = \left(\sum_{k=m}^{n-1} \bar{\boldsymbol{R}}_{m,k} \Lambda_k^{\boldsymbol{RR}} \bar{\boldsymbol{R}}_{m,k}^T \right)^{-1} \left[\sum_{k=m}^{n-1} \bar{\boldsymbol{R}}_{m,k} \Lambda_k^{\boldsymbol{RR}} \bar{\boldsymbol{R}}_{m,k}^T \bar{\boldsymbol{r}}_{m,k} - \bar{\boldsymbol{R}}_{m,k} \Lambda_k^{\boldsymbol{Rr}} \right] \tag{2.4.57b}$$

for each $n \geq m$, the expression in (2.4.55) leads to

$$\langle \boldsymbol{\Delta}_{m,n} \rangle = \exp\left\{ -\boldsymbol{\Omega}_{m,n} \right\} + \mathrm{o}\left((n-m)\epsilon\right) \tag{2.4.58a}$$

$$\langle \boldsymbol{\delta}_{m,n} \rangle = \left[\mathrm{Id}_3 - \exp\left\{ -\boldsymbol{\Omega}_{m,n} \right\} \right] \boldsymbol{\omega}_{m,n} + \mathrm{o}\left((n-m)\epsilon\right) \tag{2.4.58b}$$

as approximations of the blocks of the matrix $\langle \boldsymbol{\mathcal{D}}_{m,n} \rangle$ in (2.4.42). Note that, to deduce (2.4.58), we have used the identity

$$\exp\left\{ \begin{pmatrix} B & b \\ 0 & 0 \end{pmatrix} \right\} = \begin{pmatrix} \exp B & \left(\exp B - \mathrm{Id}_3\right) B^{-1} b \\ 0 & 1 \end{pmatrix} \tag{2.4.59}$$

for any invertible matrix $B \in \mathbb{R}^{3\times 3}$ in (2.4.55). Consequently, using the short length behaviour expressions (2.4.58) together with the factorization in (2.4.43) we deduce that the first moment of the relative rotation $\boldsymbol{R}_{m,n}$ and of the relative translation $\boldsymbol{r}_{m,n}$ can be approximated as

$$\langle \boldsymbol{R}_{m,n} \rangle = \exp\left\{ -\boldsymbol{\Omega}_{m,n} \right\} \bar{\boldsymbol{R}}_{m,n} + \mathrm{o}\left((n-m)\epsilon\right) \tag{2.4.60a}$$

$$\langle \boldsymbol{r}_{m,n} \rangle = \left[\mathrm{Id}_3 - \exp\left\{ -\boldsymbol{\Omega}_{m,n} \right\} \right] \boldsymbol{\omega}_{m,n} + \exp\left\{ -\boldsymbol{\Omega}_{m,n} \right\} \bar{\boldsymbol{r}}_{m,n} + \mathrm{o}\left((n-m)\epsilon\right) \tag{2.4.60b}$$

for n sufficiently close to m. The formulas in (2.4.60) constitute the *short length asymptotic behaviour* for rigid body chains.

We deduce that the direct dependence of the expected chain configuration $\bar{\boldsymbol{g}}_{m,n}$ in the entries of the first moment $\langle \boldsymbol{g}_{m,n} \rangle$ is, at the first order, exponentially damped at a rate governed by the symmetric and positive definite matrix

$$\boldsymbol{\Omega}_{m,n} = \frac{1}{2} \sum_{k=m}^{n-1} \mathrm{Tr}\left(\mathcal{C}_k^{\boldsymbol{RR}}\right) \mathrm{Id}_3 - \bar{\boldsymbol{R}}_{m,k} \mathcal{C}_k^{\boldsymbol{RR}} \bar{\boldsymbol{R}}_{m,k}^T \tag{2.4.61}$$

according to the expression (2.4.60) and the definition of $\Lambda_k^{\boldsymbol{RR}}$ in (2.4.24). Interestingly, for n sufficiently close to m, the persistence vector $\langle \boldsymbol{r}_{m,n} \rangle$ is expressed as a convex sum of the vector $\boldsymbol{\omega}_{m,n}$ and $\bar{\boldsymbol{r}}_{m,n}$, and, in addition, the vector $\boldsymbol{\omega}_{m,n}$ is made up of a weighted average of the translation vector $\bar{\boldsymbol{r}}_{m,n}$. In particular one has the following estimates for non-degenerate chains

$$(n-m)\lambda^{\min} \leq |\boldsymbol{\Omega}_{m,n}|_{\mathrm{Sp}} \leq (n-m)\lambda^{\max} \tag{2.4.62a}$$

$$|\boldsymbol{\omega}_{m,n}| \leq \frac{\lambda^{\max}}{\lambda^{\min}} \frac{1}{n-m} \sum_{k=m}^{n-1} \left|\bar{\boldsymbol{r}}_{m,k}\right| + \frac{1}{2} \frac{\left|\Lambda^{\boldsymbol{Rr}}\right|}{\lambda^{\min}} \tag{2.4.62b}$$

analogously to (2.4.47) and (2.4.50). Since the coefficients in $\boldsymbol{\Omega}_{m,n}$ and $\boldsymbol{\omega}_{m,n}$ are strongly dependent on the detailed constitution of the chain, more information is needed on the expected chain $\bar{\boldsymbol{g}}_{m,n}$ and on the fluctuation matrices $\mathcal{C}_n$ if one wants to describe more precisely their behaviour. One can however use relations (2.4.62) in (2.4.60) to deduce that for non-degenerate semi-flexible chains and for n sufficiently close to m, the first moments $\langle \boldsymbol{R}_{m,n} \rangle$ and $\langle \boldsymbol{r}_{m,n} \rangle$ satisfy

$$\exp\left\{-(n-m)\lambda^{\max}\right\} \leq |\langle \boldsymbol{R}_{m,n} \rangle|_{\mathrm{Sp}} \leq \exp\left\{-(n-m)\lambda^{\min}\right\} \tag{2.4.63a}$$

$$\begin{aligned} |\langle \boldsymbol{r}_{m,n} \rangle| \leq (1 - \exp\left\{-(n-m)\lambda^{\max}\right\}) \left[\frac{\lambda^{\max}}{\lambda^{\min}} \frac{1}{n-m} \sum_{k=m}^{n-1} \left|\bar{\boldsymbol{r}}_{m,k}\right| + \frac{1}{2} \frac{\left|\Lambda^{\boldsymbol{Rr}}\right|}{\lambda^{\min}} \right] \\ + \exp\left\{-(n-m)\lambda^{\min}\right\} \left|\bar{\boldsymbol{r}}_{m,n}\right| \end{aligned} \tag{2.4.63b}$$

In contrast, the long length behaviour, i.e. when the index difference $n-m$ is large, is governed by the asymptotic values given by $|m-n| \to \infty$. For non-degenerate semi-flexible chains, the solution of the recurrence relation (2.4.27) can be approximated by the constant values

$$\langle \boldsymbol{R}_{m,n} \rangle = 0 + \mathrm{O}\left(\frac{(1-\lambda^{\min})^{n-m}}{\lambda^{\min}} \right) \tag{2.4.64a}$$

$$\langle \boldsymbol{r}_{m,n} \rangle = \langle \boldsymbol{r}_{m,\infty} \rangle + \mathrm{O}\left(\frac{(1-\lambda^{\min})^{n-m}}{\lambda^{\min}} \right) \tag{2.4.64b}$$

according to (2.4.27), (2.4.30) and (2.4.34). A more detailed analysis is required however to estimate the first correction to describe the behaviour of the first moments $\langle \boldsymbol{R}_{m,n} \rangle$ and $\langle \boldsymbol{r}_{m,n} \rangle$ close to their limiting values and is left as an open topic. The quality of the short and long length approximations in (2.4.60) and (2.4.64) are discussed in section 7.2.3 in the context of a rigid base pair model of DNA.

3. On Single and Double Continuum Rods

L'ensemble continu à une dimension de tels trièdres (...)
sera ce que nous appellerons une ligne déformable.

E. & F. Cosserat, 1909.

This third chapter has been built as the continuum analogue of chapter 2 for chains and presents rod properties in classical statics and equilibrium statistical mechanics. Section 3.1 describes the continuum rod configuration and the associated infinitesimal generator vector. These notions are then extendedd to double rod, and then birod, configurations and to internal coordinates. The equilibrium conditions subject to end conditions are discussed and the related variational principles are presented in section 3.2. The birod formulation is a generalised version of the original one in [Moakher and Maddocks, 2005]. In particular, in section 3.2.8, an original coordinate independent Hamiltonian structure is obtained for the birod equilibrium conditions. We then present how an unconstrained variational principle using Lagrange multipliers, which are interpreted as the set of total couple and force, can be obtained. Section 3.3 presents the definition of a configurational distribution for rods, and respectively birods, in the context of equilibrium statistical mechanics as a path integral density. We discuss how these infinite dimensional distributions can be interpreted as the appropriate limit of a sequence of chain distributions for discretised rod configurations. This development is then extended to observables and their expectations. Similarly to chains, we emphasize the special role of the configurational first moment which comprise, here in a continuum version, the frame correlation matrix and the Flory persistence vector. An explicit first order approximation is obtained in section 3.4 for the configurational first moment for single rods which have a local internal energy and are assumed to be in a semi-flexible regime. In analogy to the case of chains, the analytical expressions shows the equal importance of the expected rod configuration and of the fluctuation matrix. They also lead to the definition of the persistence matrix which is shown to characterise the decay rates in the entries of the configurational first moment and allows the derivation of a closed form expression in a short length expansion.

3.1 On Rod Configurations

This section discuss the notion of configurations and generator vectors for rods. We then extend these definitions to two rod and birod systems, and present a set of internal variables to describe birod configurations as a macrostructure and microstructure. Finally, some specific transformations rules arising from natural material symmetries are explicitly derived.

3.1.1 On continuum rod configurations and infinitesimal generators

Many continuum models have been used to study and to predict polymer statistical physics properties. The most famous is doubtless the *Wormlike chain* which appears in the pioneering work [Krakty and Porod, 1949] and then in [Landau and Lifshitz, 1959, Doi and Edwards, 1986] for instance. This model has been extended to the *Helical Wormlike chain* in [Bugl, 1969, Yamakawa, 1976, Yamakawa, 1997] for instance, and then has taken more and more sophisticated forms [Benham, 1979, Marko and Siggia, 1994, R.S. Manning, 1996, Becker and Everaers, 2007]. In the context of classical mechanics, all of these models belong to a family of continuum rod models which can be described as special *Cosserat rod models* [Cosserat and Cosserat, 1909].

A *continuum rod configuration* is a continuum path $\boldsymbol{g}(s) = (\boldsymbol{R}(s), \boldsymbol{r}(s)) \in \mathrm{SE}(3)$ in the special Euclidean group[1]. The rotation matrices $\boldsymbol{R}(s) \in \mathrm{SO}(3)$ describe the orientation of the *rod cross sections* and the triple $\boldsymbol{r}(s) \in \mathbb{R}^3$ the position of its *centreline*. Typically, the continuum parameter s is chosen to represent unstressed arc length of the rod, i.e. $s \in]0, L[$ where L is the total length of the rod in its unstressed state. Formally, a rod configuration is then denoted by

$$\boldsymbol{g} = \boldsymbol{g}(s) \in \mathrm{SE}(3)^{]0,L[} \tag{3.1.1}$$

where $\mathrm{SE}(3)^{]0,L[}$ denotes the space of continuous curve in the special Euclidean group.

In analogy to the rigid body chain configurations defined (2.1.2), we use the following matrix formulation

$$\boldsymbol{g}(s) = \begin{pmatrix} \boldsymbol{R}(s) & \boldsymbol{r}(s) \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{4\times 4} \tag{3.1.2}$$

as in [Becker and Everaers, 2007, Chirikjian, 2011] for instance. Note that, by convention, we use *rod* as the name for the continuum representation of a polymer in contrast to *chain* for the discrete one. xThe infinitesimal variations of a continuum rod configuration at a value $s \in]0, L[$ can be expressed as

$$\frac{d}{ds}\boldsymbol{R}(s) = \boldsymbol{R}(s)\,[U(s)\times] \tag{3.1.3a}$$

$$\frac{d}{ds}\boldsymbol{r}(s) = \boldsymbol{R}(s)V(s) \tag{3.1.3b}$$

[1] See section 1.1 for more detail on the special Euclidean group

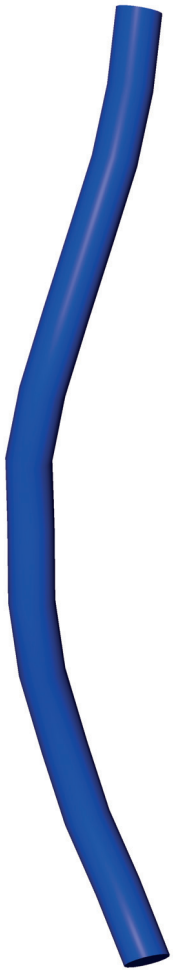


Figure 3.1: Sketch of a continuum rod configuration

for all $s \in]0, L[$. The triple $U(s) \in \mathbb{R}^3$ denotes the three *infinitesimal rotational degrees of freedom* at s and is often referred as the *Darboux vector*[Darboux, 1894], and the triple $V(s) \in \mathbb{R}^3$ denotes the analogous three *infinitesimal translational degrees of freedom*. More details about the standard expressions (3.1.3) are given in section 1.2. In terms of the rod configuration in (3.1.2), expressions (3.1.3) can be summarized as

$$\frac{d}{ds}\boldsymbol{g}(s) = \boldsymbol{g}(s)\mathcal{T}\xi(s) \tag{3.1.4}$$

where

$$\xi(s) = (U(s), V(s)) \in \mathbb{R}^6 \tag{3.1.5}$$

is the *rod infinitesimal generator vector*[2] and where the *tangent map*[3] $\mathcal{T}$ is defined as

$$\mathcal{T}\xi(s) = \begin{pmatrix} [U(s)\times] & V(s) \\ 0 & 0 \end{pmatrix} \text{ for } \xi(s) = (U(s), V(s))\,. \tag{3.1.6}$$

The *skew operator* $[\,.\,\times]$ is defined in (1.1.9). We emphasize that the evolution equation (3.1.4) is the continuum analogue of the discrete one in (2.1.4) for rigid body chains.

The *relative rigid body displacement* between two cross sections located at s' and s is denoted $\boldsymbol{g}(s', s) \in \mathrm{SE}(3)$ and defined by

$$\boldsymbol{g}(s', s) = \boldsymbol{g}^{-1}(s')\boldsymbol{g}(s). \tag{3.1.7}$$

The matrix $\boldsymbol{g}(s', s)$ has the following block structure

$$\boldsymbol{g}(s', s) = \begin{pmatrix} \boldsymbol{R}(s', s) & \boldsymbol{r}(s', s) \\ 0 & 1 \end{pmatrix} \tag{3.1.8}$$

with

$$\boldsymbol{R}(s', s) = \boldsymbol{R}^T(s')\boldsymbol{R}(s) \tag{3.1.9a}$$

$$\boldsymbol{r}(s', s) = \boldsymbol{R}^T(s')[\boldsymbol{r}(s) - \boldsymbol{r}(s')]. \tag{3.1.9b}$$

Analogously to (3.1.4), the relative rigid body motion $\boldsymbol{g}(s', s)$ is the solution of

$$\partial_s \boldsymbol{g}(s', s) = \boldsymbol{g}(s', s)\mathcal{T}\xi(s) \text{ with } \boldsymbol{g}(s', s') = \mathrm{Id}_4 \tag{3.1.10}$$

for $s' \leq s < L$, which satisfies the following *transition identity*

$$\boldsymbol{g}(s'', s) = \boldsymbol{g}(s'', s')\boldsymbol{g}(s', s) \tag{3.1.11}$$

[2]The vector $\xi(s)$ is precisely a *right infinitesimal generator*, in contrast to a *left infinitesimal generator*, we will not insist on this difference however. See section 1.2 for more details.

[3]See sections 1.1 and 1.2 for more details on variations on matrix groups.

for all $0 < s'' \leq s' \leq s < L$. The relation (3.1.10) leads then formally to the mapping

$$\xi(s) \mapsto \boldsymbol{g}[\xi] \tag{3.1.12}$$

or more generally to

$$\xi(s) \mapsto \boldsymbol{g}[\xi](s', s). \tag{3.1.13}$$

In contrast to the analogous relation for a rigid body chain in (2.1.14), we here use square brackets $[\,.\,]$ to emphasize the fact that relations (3.1.12) and (3.1.13) are mappings between functions and may involve the solution of an ODE system. Nevertheless, the expression (3.1.12) has to be understood as the continuum version of the one in (2.1.14).

3.1.2 On continuum double rod configurations

As already noted in the context of chain model, several examples of macromolecule exhibit an interesting multi-stranded structure which is only poorly captured by single rod models. This statement is motivated the study of a system made of two rods in [Moakher and Maddocks, 2005], especially designed to produce DNA models.

A *continuum double rod configuration* $(\boldsymbol{g}^+, \boldsymbol{g}^-)$ can be described as a pair of continuum rod, denoted respectively $\boldsymbol{g}^+$ and $\boldsymbol{g}^-$, which are assumed to be parametrized by a common parameter $s \in]0, L[$. Formally we have

$$(\boldsymbol{g}^+, \boldsymbol{g}^-) = (\boldsymbol{g}^+(s), \boldsymbol{g}^-(s)) \in \left(\mathrm{SE}(3)^{]0,L[}\right)^2 \tag{3.1.14}$$

where $\left(\mathrm{SE}(3)^{]0,L[}\right)^2$ denotes two copies of the set of continuous paths in the special Euclidean group, as in (3.1.1). Note that we use the same notation for double rod configurations and double chain configurations, as in (3.1.14), since their use will be unambiguous.

Each continuum rod in the double rod configuration $(\boldsymbol{g}^+, \boldsymbol{g}^-)$ is of the form $\boldsymbol{g}^\pm(s) = (\boldsymbol{R}^\pm(s), \boldsymbol{r}^\pm(s))$ where $\boldsymbol{R}^\pm(s)$ denotes the rod orientation and $\boldsymbol{r}^\pm(s)$ its position. The notation $\boldsymbol{g}^\pm$ is used to described one of the two rods $\boldsymbol{g}^+$ or $\boldsymbol{g}^-$ respectively, as in [Moakher and Maddocks, 2005], and each definition related to rods can be translated to the rods $\boldsymbol{g}^\pm$ in an unambiguous way. For example, the infinitesimal generator of the rod $\boldsymbol{g}^\pm(s)$ is then denoted by $\xi^\pm(s)$.

Analogously to the double chain description in (2.1.20), we introduce the *intra rigid body displacement* $b(s)$ defined as

$$b(s) = \left(\boldsymbol{g}^+(s)\right)^{-1} \boldsymbol{g}^-(s) \tag{3.1.15}$$

whose block structure is denoted as

$$b(s) = \begin{pmatrix} P(s) & p(s) \\ 0 & 1 \end{pmatrix}. \tag{3.1.16}$$

Figure 3.2: Sketch of a continuum double rod configuration

Consequently, the double rod configuration satisfies the relations

$$\frac{d}{ds}\boldsymbol{g}^{\pm}(s) = \boldsymbol{g}^{\pm}(s)\mathcal{T}\xi^{\pm}(s) \tag{3.1.17a}$$

for all $s \in]0, L[$ and

$$\boldsymbol{g}^{-}(s) = \boldsymbol{g}^{+}(s)b(s) \tag{3.1.17b}$$

for all $s \in]0, L[$.

3.1.3 On continuum birod configurations and internal variables

The birod model was originally presented in [Moakher and Maddocks, 2005] and was developed to describe double stranded DNA molecules as a continuum. The model is a natural extension of single rods and describes the mechanics of two rods with a local interaction.

A *continuum birod configuration* $(\boldsymbol{g}, \mathcal{P})$ can be described as a *macrostructure configuration* $\boldsymbol{g}(s)$ together with a *microstructure configuration* $\mathcal{P}(s)$. The macrostructure configuration is a continuum rod configuration and is aimed to represent on average the configuration of the double rod, whereas the microstructure configuration encodes the relative rotation and the relative translation and allows the reconstruction of the actual double rod configurations $\boldsymbol{g}^{\pm}$ from the macrostructure $\boldsymbol{g}$. The continuum birod configuration is then essentially a double rod configuration, but described in a different way. Formally one has then a one-to-one mapping

$$(\boldsymbol{g}^{+}, \boldsymbol{g}^{-}) \leftrightarrow (\boldsymbol{g}, \mathcal{P}) \tag{3.1.18}$$

between double rod configurations and birod configurations.

In [Moakher and Maddocks, 2005], the microstructure is defined as a function of the half rotation and half translation between the two rods at a given common parameter s. However, since we will use the birod model to produce a large scale continuum DNA model it will be convenient if the microstructure is parametrized consistently with the discrete models [Lankaš et al., 2009, Gonzalez et al., 2013, Petkeviciute et al., 2014], i.e as a function of the full rotation and translation.

Consequently, the rod macrostructure configuration $\boldsymbol{g}$ is defined as $\boldsymbol{g}(s) = (\boldsymbol{R}(s), \boldsymbol{r}(s)) \in \mathrm{SE}(3)$ where

$$\boldsymbol{R}(s) = \boldsymbol{R}^{+}(s)\left[\boldsymbol{R}^{+}(s)^{T}\boldsymbol{R}^{-}(s)\right]^{\frac{1}{2}} \tag{3.1.19a}$$

and

$$\boldsymbol{r}(s) = \frac{1}{2}\left[\boldsymbol{r}^{+}(s) + \boldsymbol{r}^{-}(s)\right] \tag{3.1.19b}$$

for $s \in]0, L[$. The orientation $\boldsymbol{R}(s)$ is the average element in the rotation group between the orientation $\boldsymbol{R}^{+}(s)$ and $\boldsymbol{R}^{-}(s)$, and the position $\boldsymbol{r}(s)$ is the average point in the

Euclidean space between the position $\boldsymbol{r}^+(s)$ and $\boldsymbol{r}^-(s)$. The microstructure configuration $\mathcal{P}$ is then defined as $\mathcal{P}(s) = (P(s), \mathsf{w}(s)) \in \mathrm{SO}(3) \times \mathbb{R}^3$ where

$$P(s) = \boldsymbol{R}^+(s)^T \boldsymbol{R}^-(s) \tag{3.1.20a}$$

describes the relative rotation between the cross-sections $\boldsymbol{g}^+(s)$ and $\boldsymbol{g}^-(s)$, and where

$$\mathsf{w}(s) = \boldsymbol{R}(s)^T \left[\boldsymbol{r}^-(s) - \boldsymbol{r}^+(s)\right] \tag{3.1.20b}$$

describes the analogous relative translation. Observe that the definitions (3.1.19) and (3.1.20) are precisely the continuum analogues of the discrete expressions in (2.1.25) and (2.1.27).

According to the definition in (3.1.19) and in (3.1.20), the double rod configuration can be recovered from its associated birod configuration by

$$\boldsymbol{R}^\pm(s) = \boldsymbol{R}(s) P^{\mp\frac{1}{2}}(s) \tag{3.1.21a}$$

and

$$\boldsymbol{r}^\pm(s) = \boldsymbol{r}(s) \mp \frac{1}{2}\boldsymbol{R}(s)\mathsf{w}(s). \tag{3.1.21b}$$

The infinitesimal generator $\xi^\pm(s) = (U^\pm(s), V^\pm(s))$ and each continuum rod $\boldsymbol{g}^\pm(s)$ is then shown to satisfy

$$U^\pm(s) = P^{\pm\frac{1}{2}}(s) U(s) \mp M^\mp(s) U_P(s) \tag{3.1.22a}$$

and

$$V^\pm(s) = P^{\pm\frac{1}{2}}(s) \left[V(s) \mp \frac{1}{2} U(s) \times \mathsf{w}(s) \mp \frac{1}{2} V_\mathsf{w}\right]. \tag{3.1.22b}$$

where we have used

$$\frac{d}{ds}\boldsymbol{g}(s) = \boldsymbol{g}(s)\mathcal{T}\xi(s) \tag{3.1.23}$$

where $\xi(s) = (U(s), V(s)) \in \mathbb{R}^6$ stands for the *macrostructure generator vector* similarly to the rod case (3.1.4), and the expressions for the microstructure

$$\frac{d}{ds}P(s) = P^{\frac{1}{2}}(s) \left[U_P(s)\times\right] P^{\frac{1}{2}}(s) \tag{3.1.24a}$$

$$\frac{d}{ds}\mathsf{w}(s) = V_\mathsf{w}(s) \tag{3.1.24b}$$

and the matrices

$$M^\pm(s) = \left(\mathrm{Id}_3 + P^{\pm\frac{1}{2}}(s)\right)^{-1}. \tag{3.1.25}$$

similarly to (2.2.39). The variation in (3.1.24a) is not standard and is introduced to force a simple symmetry relation when interchanging the role of the rods denoted by + and −. A similar expression has been used in the case of bichains in (2.2.37). The expressions in (3.1.24) motivate the definition of the *microstructure generator vector* $\xi^{\mathcal{P}}(s) \in \mathbb{R}^6$ as

$$\xi^{\mathcal{P}}(s) = (U_P(s), V_{\mathsf{w}}(s)) \tag{3.1.26}$$

and the associated tangent map $\mathcal{T}^{\mathcal{P}}$ defined by the relation

$$\frac{d}{ds}\mathcal{P}(s) = \mathcal{T}^{\mathcal{P}}\xi^{\mathcal{P}}(s) \tag{3.1.27}$$

according to (3.1.24). The microstructure generator vector $\xi^{\mathcal{P}}$ has the property of satisfying a simple transformation rule when the role of the strands $\boldsymbol{g}^+$ and $\boldsymbol{g}^-$ are exchanged, as described in more details in section 3.1.4. It constitutes a *centered generator vector* because it is neither a left nor a right generator vector.

We should stress that the continuum microstructure $\mathcal{P}(s)$ can then be recovered from the vector $\xi^{\mathcal{P}}(s)$ using the equations in (3.1.27), whereas the continuum macrostructure $\boldsymbol{g}(s)$ is recovered for the vector $\xi(s)$ using (3.1.4). Observe that again the expressions in (3.1.22) are continuum versions of the analogous discrete ones in (2.1.29).

The *birod internal variables* then comprise a one parameter family of infinitesimal rod generators $\xi(s)$, $s \in]0, L[$, for the macrostructure $\boldsymbol{g}$ and a set of local coordinates $y(s)$, $s \in]0, L[$, called *intra coordinates*, to parametrize the microstructure $\mathcal{P}$. Following the choices for the bichain model made in [Lankaš et al., 2009, Gonzalez et al., 2013, Petkeviciute et al., 2014], the intra displacements $b(s)$, as defined in (3.1.15), are parametrized in terms of the birod internal coordinates $y(s)$ according to

$$b(y(s)) = \begin{pmatrix} P(\eta(s)) & P^{\frac{1}{2}}(\eta(s))\mathsf{w}(s) \\ 0 & 1 \end{pmatrix} \tag{3.1.28}$$

where $y(s) = (\eta(s), \mathsf{w}(s))$ and where $\eta(s)$ denotes the Cayley vector of the rotation matrix $P(s)$ as defined in (2.1.16), and in (3.1.16), similarly to (2.1.33).

3.1.4 Material framings and transformation rules

As discussed in the analogous section 2.1.4, it is important to present the transformation rules related to different *material framings* in the description of continuum rod and birod configurations.

For rods, the infinitesimal generator $\xi(s)$ defined in (3.1.4) respect *invariance under overall rigid body motion*, i.e.

$$\xi(s) \mapsto \xi(s) \text{ when } \boldsymbol{g} \mapsto B\boldsymbol{g} \tag{3.1.29}$$

for any $B \in \mathrm{SE}(3)$. Similarly, for a birod configuration the generator vector of the macrostructure and the local displacements in the microstructure satisfy

$$\xi(s) \mapsto \xi(s) \text{ and } \mathcal{P}(s) \mapsto \mathcal{P}(s) \text{ when } (\boldsymbol{g}^+, \boldsymbol{g}^-) \mapsto (B\boldsymbol{g}^+, B\boldsymbol{g}^-) \tag{3.1.30a}$$

for any $B \in \mathrm{SE}(3)$ according to its definition in (3.1.20), or in terms of the birod internal variables

$$\xi(s) \mapsto \xi(s) \text{ and } y(s) \mapsto y(s) \text{ when } (\boldsymbol{g}^+, \boldsymbol{g}^-) \mapsto (B\boldsymbol{g}^+, B\boldsymbol{g}^-) \tag{3.1.30b}$$

for any $B \in \mathrm{SE}(3)$.

For a given rod configuration $\boldsymbol{g}(s)$ one can redefine locally the description of orientation and the position of the cross-section at s. This can be formally stated as $\boldsymbol{g}(s) \mapsto \boldsymbol{g}(s)\bar{\boldsymbol{g}}(s)^{-1}$ for all $s \in]0, L[$ and is called a *local change of framing*. For a single rod one has then that the infinitesimal rod generator vector transforms as

$$\xi(s) \mapsto \mathrm{Ad}_{\bar{\boldsymbol{g}}(s)} \left(\xi(s) - \bar{\xi}(s)\right) \text{ when } \boldsymbol{g}(s) \mapsto \boldsymbol{g}(s)\bar{\boldsymbol{g}}(s)^{-1} \tag{3.1.31}$$

where $\frac{d}{ds}\bar{\boldsymbol{g}}(s) = \bar{\boldsymbol{g}}(s)\mathcal{T}\bar{\xi}(s)$ and where we have used the result in (2.2.14). For a double rod, if the local change of framing is uniform, i.e. if $\boldsymbol{g}^{\pm}(s) \mapsto \boldsymbol{g}^{\pm}(s)\bar{\boldsymbol{g}}^{-1}$, then the birod macrostructure and microstructure transform as

$$\begin{aligned} \xi(s) &\mapsto \mathrm{Ad}_{\bar{\boldsymbol{g}}(s)}\xi(s) \\ \text{and } (P(s), \mathsf{w}(s)) &\mapsto (\bar{\boldsymbol{R}}P(s)\bar{\boldsymbol{R}}^T, \bar{\boldsymbol{R}}\mathsf{w}(s) - \bar{\boldsymbol{R}}(P^{\frac{1}{2}}(s) - P^{\frac{T}{2}}(s))\bar{\boldsymbol{R}}^T\bar{\boldsymbol{r}}) \\ &\text{when } \boldsymbol{g}^{\pm}(s) \mapsto \boldsymbol{g}^{\pm}(s)\bar{\boldsymbol{g}}^{-1} \end{aligned} \tag{3.1.32}$$

where $\mathcal{P}(s) = (P(s), \mathsf{w}(s))$ and where $\bar{\boldsymbol{R}} \in \mathrm{SO}(3)$ denotes the rotational part of rigid body displacement $\bar{\boldsymbol{g}}$ and $\bar{\boldsymbol{r}} \in \mathbb{R}^3$ its translational part. A transformation rule can also be explicitly stated for non-uniform local change of framing of double rod configurations, but it is more complicated and will not be used in this thesis. For the internal birod coordinates presented in (3.1.28), one has then

$$\begin{aligned} (\eta(s), \mathsf{w}(s)) \mapsto (\bar{\boldsymbol{R}}\eta(s), \bar{\boldsymbol{R}}\mathsf{w}(s) - \bar{\boldsymbol{R}}(P^{\frac{1}{2}}(s) - P^{\frac{T}{2}}(s))\bar{\boldsymbol{R}}^T\bar{\boldsymbol{r}}) \\ \text{when } \boldsymbol{g}^{\pm}(s) \mapsto \boldsymbol{g}^{\pm}(s)\bar{\boldsymbol{g}}^{-1}. \end{aligned} \tag{3.1.33}$$

Similarly to chains, one can also change the framing by *downstream indexing* which corresponds to starting the parametrization of the rod from the other end. The rod infinitesimal generator $\xi(s)$ then transforms as

$$\xi(s) \mapsto -\xi(L - s) \text{ when } \boldsymbol{g}(s) \mapsto \boldsymbol{g}(L - s) \tag{3.1.34}$$

for all $s \in]0, L[$. For a birod configuration, downstream indexing leads to

$$\begin{aligned} \xi(s) \mapsto -\xi(L - s) \text{ and } \mathcal{P}(s) \mapsto \mathcal{P}(L - s) \\ \text{when } (\boldsymbol{g}(s), \mathcal{P}(s)) \mapsto (\boldsymbol{g}(L - s), \mathcal{P}(L - s)) \end{aligned} \tag{3.1.35a}$$

and consequently the internal coordinates become

$$\begin{aligned} \xi(s) \mapsto -\xi(L - s) \text{ and } y(s) \mapsto y(L - s) \\ \text{when } (\boldsymbol{g}(s), \mathcal{P}(s)) \mapsto (\boldsymbol{g}(L - s), \mathcal{P}(L - s)) \end{aligned} \tag{3.1.35b}$$

As in bichains, one can also switch the role of the rods denoted by + and −, this is called *change of reference strand*. The infinitesimal rod generator $\xi(s)$ and the intra displacement in the microstructure of the birod then transform as

$$\xi(s) \mapsto \xi(s) \text{ and } b(s) \mapsto b(s)^{-1} \text{ when } \boldsymbol{g}^{\pm} \mapsto \boldsymbol{g}^{\mp} \tag{3.1.36a}$$

whereas the microstructure transforms as

$$\begin{aligned} \mathcal{P}(s) = (P(s), \mathsf{w}(s)) \mapsto (P(s)^T, -\mathsf{w}(s)) \text{ and } \xi^{\mathcal{P}}(s) \mapsto -\xi^{\mathcal{P}}(s) \\ \text{when } \boldsymbol{g}^{\pm} \mapsto \boldsymbol{g}^{\mp} \end{aligned} \tag{3.1.36b}$$

and the internal coordinates y transform as

$$y(s) \mapsto -y(s) \text{ when } \boldsymbol{g}^{\pm} \mapsto \boldsymbol{g}^{\mp} \tag{3.1.36c}$$

in the case of the choice described in (3.1.28).

In DNA modeling, the natural symmetry is not a simple change of reference strand, but the *reading symmetry*. This symmetry, as described in the analogous section 2.1.4, reflects the fact that DNA molecules are anti-symmetric objects since reading the strand $\boldsymbol{g}^+$ upstream corresponds to reading the strand $\boldsymbol{g}^-$ downstream according to its chemical structure. The reading symmetry is then the composition of the local change of framing (3.1.32), the downstream indexing (3.1.35) and the double rod symmetry (3.1.36). Formally, this symmetry involves the transformation $(\boldsymbol{g}^+, \boldsymbol{g}^-) \mapsto (\boldsymbol{g}^{+\prime}, \boldsymbol{g}^{-\prime})$ where by definition

$$\boldsymbol{g}^{\pm\prime}(s) = \boldsymbol{g}^{\mp}(L - s). \tag{3.1.37}$$

One obtains then that the birod variables, similarly to the bichain internal coordinates, transforms under the reading symmetry, as

$$\begin{aligned} \xi(s) \mapsto \mathbb{E}_2 \xi(L - s) \text{ and } y(s) \mapsto \mathbb{E}_2 y(L - s) \\ \text{when } (\boldsymbol{g}^+, \boldsymbol{g}^-) \mapsto (\boldsymbol{g}^{+\prime}, \boldsymbol{g}^{-\prime}) \end{aligned} \tag{3.1.38}$$

and where $\mathbb{E}_2$ stands for the diagonal matrix $\mathrm{diag}(-1, 1, 1, -1, 1, 1)$.

3.2 On Equilibrium Conditions and Variational Principles

We discuss in this section the standard definitions of couple and force, as well as equilibrium conditions for continuum rods. We describe also the notion of rod internal energy and show that it provides a coordinate independent variational principle for the rod equilibrium and, moreover, leads to a Hamiltonian formulation. These results are then extended to double rods and birods. In particular, the birod equilibrium conditions also admit a Hamiltonian structure and are shown to comprise the rod equilibrium conditions for the macrostructure configuration which is coupled to the description of the microstructure equilibrium.

3.2.1 On continuum rod equilibrium conditions and constitutive relations

Similarly to chains, continuum rod configurations adapt to the application of local couples and forces. One can distinguish between the action of *internal couples and forces* coming from the internal deformation of the rod and the action of *external couples and forces* coming from the interaction of the rod with an external field. A continuum rod is said to be in an *equilibrium configuration* if the total couple and force densities acting on each cross section vanish. Using a similar notation than the one we have used for rigid body chains, we denote by $\mathbf{m}(s) \in \mathbb{R}^3$ the total internal couple around the point $\boldsymbol{r}(s)$ acting on the cross section located at s from the material part of the rod parametrized by $s \in]0, L[$, and by $\mathbf{n}(s) \in \mathbb{R}^3$ the analogous total force, the *rod equilibrium conditions* can be stated as

$$-\frac{d}{ds}\left[\mathbf{m}(s) + \boldsymbol{r}(s) \times \mathbf{n}(s)\right] + \mathbf{c}(s) + \boldsymbol{r}(s) \times \mathbf{f}(s) = 0 \tag{3.2.1a}$$

$$-\frac{d}{ds}\mathbf{n}(s) + \mathbf{f}(s) = 0 \tag{3.2.1b}$$

for $s \in]0, L[$ and where $\mathbf{c}(s) \in \mathbb{R}^3$ and $\mathbf{f}(s) \in \mathbb{R}^3$ denote respectively the total external couple density (around the point $\boldsymbol{r}(s)$) and the total external force density at s. Conditions (3.2.1) can be deduced for instance from the equilibrium conditions in three dimensional continuum elasticity [Love, 1927, Antman, 1995]. We observe that the rod balance laws can be interpreted as a continuum version of the chain balance laws in (2.2.1). The rod is said to be an *isolated continuum rod* if there is no external contributions to the total couples and forces acting on the chain, i.e. if $\mathbf{c}(s) = 0$ and $\mathbf{f}(s) = 0$.

As for rigid body chains, one needs to introduce *rod end conditions*. We are interested in the following ones

i. *Dirichlet conditions*: the orientation and the position of each end of the rod are prescribed, i.e.

$$\boldsymbol{g}(0) \text{ and } \boldsymbol{g}(L) \text{ are prescribed.} \tag{3.2.2a}$$

ii. *Unstressed conditions*: the orientation and the position are free at each end of the rod, i.e.

$$\mathbf{m}(0) = 0,\ \mathbf{n}(0) = 0 \text{ and } \mathbf{m}(L) = 0,\ \mathbf{n}(L) = 0 \tag{3.2.2b}$$

For compactness we will denote generically by Γ the subset of rod configurations which satisfies one of the end conditions i. or ii., i.e.

$$\Gamma = \left\{ \boldsymbol{g} \in \mathrm{SE}(3)^N \,|\, \boldsymbol{g} \text{ satisfies i. or ii. } \right\} \tag{3.2.3}$$

We observe that the rod end conditions in (3.2.2) are the continuum version of the ones in (2.2.2).

The *rod constitutive relations* are formally a mapping of the form $\boldsymbol{g} \mapsto (\mathbf{m}, \mathbf{n}, \mathbf{c}, \mathbf{f})$. In particular, we can define the notion of *local rod constitutive relations* which means that the local internal couple and force in the chain are only functions of the local rod deformation, i.e.

$$\begin{pmatrix} \mathbf{m}(s) \\ \mathbf{n}(s) \end{pmatrix} = \boldsymbol{\Psi}^{(\mathbf{m},\mathbf{n})}(\xi(s); s) \text{ for all } s \in]0, L[. \tag{3.2.4}$$

The constitutive relations contain information about the physical constitution of the rod. The construction of realistic constitutive relations for a given material is far from being straightforward and it is still at the heart of research in material science, see [Antman, 1995] for instance. The balance law conditions (3.2.1) on a set of configuration (3.2.3) together with a given set of constitutive relations, of the form (3.2.4) for instance, are called the *rod equilibrium conditions*.

3.2.2 On continuum rod internal energy and variational principle

The *rod internal energy* is denoted by E, as for rigid body chains. It describes the stored potential energy arising exclusively from internal deformations of the continuum rod. The internal energy is said to be a *local energy*, if the functional $E : \mathrm{SE}(3)^{]0,L[} \to \mathbb{R}$ is of the form

$$E[\boldsymbol{g}] = \int_0^L \mathsf{W}(\xi(s); s) ds \tag{3.2.5}$$

where the rod configuration $\boldsymbol{g}(s)$ is defined in (3.1.1) and where $\xi(s) \in \mathbb{R}^6$ denotes the rod infinitesimal generator as defined in (3.1.4). This energy formulation for rods appears originally in the work of E. and F. Cosserat [Cosserat and Cosserat, 1909] which is why such a rod theory is called a *Cosserat rod theory*. The energy E is said to be internal because it is invariant through global rigid body motion of the rod, as defined in (3.1.29), formally one has

$$E[B\boldsymbol{g}] = E[\boldsymbol{g}] \tag{3.2.6}$$

for any $B \in \mathrm{SE}(3)$.

It is not clear that the locality assumption of the internal rod energy is very accurate in general, since this hypothesis has been show to break for analogous rigid body chain models of macromolecules [Flory, 1969, Lankaš et al., 2009, Gonzalez et al., 2013]. However, the local energy expression in (3.2.5) is sufficiently general to embrace most of the commonly used continuum rod models, for instance in polymer and DNA modelling [Landau and Lifshitz, 1959, Bugl, 1969, Yamakawa, 1976, Marko and Siggia, 1994, R.S. Manning, 1996, Yamakawa, 1997].

The rod internal energy E allows the statement of a *variational principle for continuum rods* in the sense that stationary configurations of the functional E on the set of configurations Γ, as defined respectively in (1.2.14) and in (3.2.3), satisfy equilibrium conditions of the form (3.2.1) for an isolated rod together with constitutive relations of the form (3.2.4). Note that it is straightforward to generalise the result in the presence of an external potential. However, we have to stress that not all constitutive relations of the form (3.2.4) admit a variational principle, the class of rods which admit an internal energy is called *hyper-elastic rods*.

More precisely, suppose that Γ denotes the set of continuum rod configurations where the cross-sections at $s = 0$ and $s = L$ in the rod are prescribed and suppose that the rod configuration $\boldsymbol{g}^0(s) \in \Gamma$ makes the internal energy E stationary on the set Γ. Then by definition[4] it satisfies

$$\delta E[\boldsymbol{g}^0]\delta\boldsymbol{g} = 0 \tag{3.2.7a}$$

for all perturbations $\delta\boldsymbol{g}(s)$ where

$$\delta\boldsymbol{g}(s) = (\mathcal{T}\boldsymbol{\phi}(s))\boldsymbol{g}^0(s) \text{ for } s \in]0, L[\tag{3.2.7b}$$

and

$$\delta\boldsymbol{g}(0) = 0 \text{ and } \delta\boldsymbol{g}(L) = 0 \tag{3.2.7c}$$

according to the specific form of the perturbation of continuum rod configurations in (1.2.10) and the specific definition of the set of configuration Γ. Furthermore, at any smooth enough configuration $\boldsymbol{g} \in \Gamma$ any perturbation $\delta\boldsymbol{g}$ is of the form (3.2.7b) and (3.2.7c) and the perturbation of the rod generator vector $\xi(s)$ reads

$$\delta\xi(s) = \mathrm{Ad}^{-1}_{\boldsymbol{g}(s)}\left(\frac{d}{ds}\boldsymbol{\phi}(s)\right) \tag{3.2.8}$$

for $s \in]0, L[$, where the matrix $\mathrm{Ad}_{\boldsymbol{g}(s)}$ is defined as

$$\mathrm{Ad}_{\boldsymbol{g}(s)} = \begin{pmatrix} \boldsymbol{R}(s) & 0 \\ [\boldsymbol{r}(s)\times]\,\boldsymbol{R}(s) & \boldsymbol{R}(s) \end{pmatrix}. \tag{3.2.9}$$

[4] See section 1.2 for more detail on variations of functions on a matrix group.

We observe that the expression (3.2.8) is discussed in more detail in (1.2.11) and that it can be interpreted as the continuum version of the variation in (2.2.8). Using then the form of the internal energy E in (3.2.5), its first variation can be written as

$$\delta E[\boldsymbol{g}]\delta\boldsymbol{g} = -\int_0^L \left\{ \frac{d}{ds}\left[\mathbf{m}(s) + \boldsymbol{r}(s)\times\mathbf{n}(s)\right]\cdot\boldsymbol{\phi}^{\boldsymbol{R}}(s) + \frac{d}{ds}\mathbf{n}(s)\cdot\boldsymbol{\phi}^{\boldsymbol{r}}(s)\right\} ds \tag{3.2.10}$$

for $\boldsymbol{\phi}(s) = (\boldsymbol{\phi}^{\boldsymbol{R}}(s), \boldsymbol{\phi}^{\boldsymbol{r}}(s)) \in \mathbb{R}^6$, where we have defined for $s \in]0, L[$

$$\mathbf{m}(s) = \boldsymbol{R}(s)\partial_U \mathsf{W}(\xi; s) \tag{3.2.11a}$$

$$\mathbf{n}(s) = \boldsymbol{R}(s)\partial_V \mathsf{W}(\xi; s) \tag{3.2.11b}$$

and where we recall that $U(s)$ and $V(s)$ denote respectively the rotational and translational part of the rod generator vector $\xi(s)$ as defined in (3.1.5). When the boundary condition set Γ denotes case ii. in (3.2.3), a similar argument shows that the variational principle for the chain balance laws in (3.2.10) is also obtained. The variational principle (3.2.10) is called the *continuum Euler-Poincaré variation for rods* [5]. Observe that the expressions in (3.2.10) are a continuum version of (2.2.9). More details about variational principles for rods can be found in [Antman, 1995, Dichmann et al., 1996, Chouaïeb, 2003, Cotta-Ramusino, 2008, Ellis et al., 2010] for instance.

Consequently, if the rod constitutive relations in (3.2.4) admit a variational representation of the form (3.2.11) for all rod configurations $\boldsymbol{g} \in \Gamma$, then the stationarity conditions for the internal energy E on the set of configurations Γ in (3.2.7) is literally an expression of the rod equilibrium conditions (3.2.1) for isolated chains. Equations (3.2.10) imply that the six quantities

$$\mathbf{m}(s) + \boldsymbol{r}(s)\times\mathbf{n}(s) \in \mathbb{R}^3 \text{ and } \mathbf{n}(s) \in \mathbb{R}^3 \tag{3.2.12}$$

are conserved along equilibrium rod configuration, or equivalently, on stationary configuration of the internal energy E. The existence of conserved quantities in (3.2.12) is again the manifestation of the famous *Noether's Theorem*, since it expresses the invariance of the internal rod energy E defined in (3.2.5) under a global rigid body motion of the rod, as presented in (3.2.6).

An alternative expression of the stationary conditions is given by introducing the triples $\mathsf{m}(s) \in \mathbb{R}^3$ and $\mathsf{n}(s) \in \mathbb{R}^3$ as the respective components of the couple $\mathbf{m}(s)$ and the force $\mathbf{n}(s)$ in the director frame $\boldsymbol{R}(s)$ i.e.

$$\mathsf{m}(s) = \partial_U \mathsf{W}(\xi(s); s) \tag{3.2.13a}$$

$$\mathsf{n}(s) = \partial_V \mathsf{W}(\xi(s); s) \tag{3.2.13b}$$

according to the definition (3.2.11). Using then the identities

$$\begin{pmatrix} \mathbf{m}(s) + \boldsymbol{r}(s)\times\mathbf{n}(s) \\ \mathbf{n}(s) \end{pmatrix} = \mathrm{Ad}^{-T}_{\boldsymbol{g}(s)} \begin{pmatrix} \mathsf{m}(s) \\ \mathsf{n}(s) \end{pmatrix} \tag{3.2.14}$$

[5] See section 1.2 for more details on variation on matrix groups

and

$$\frac{d}{ds}\mathrm{Ad}_{\boldsymbol{g}(s)} = \mathrm{Ad}_{\boldsymbol{g}(s)}\mathrm{ad}_{\xi(s)} \tag{3.2.15}$$

where the operator $\mathrm{Ad}_{\boldsymbol{g}(s)}$ is defined in (3.2.9) and the operator $\mathrm{ad}_{\xi(s)}$ by

$$\mathrm{ad}_{\xi(s)} = \begin{pmatrix} [U(s)\times] & 0 \\ [V(s)\times] & [U(s)\times] \end{pmatrix} \tag{3.2.16}$$

for $\xi(s) = (U(s), V(s))$, and as discussed in more detail in (1.1.34), the stationary condition (3.2.7) reads

$$\frac{d}{ds}\begin{pmatrix} \mathsf{m}(s) \\ \mathsf{n}(s) \end{pmatrix} = \mathrm{ad}^T_{\xi(s)}\begin{pmatrix} \mathsf{m}(s) \\ \mathsf{n}(s) \end{pmatrix} \tag{3.2.17a}$$

$$\frac{d}{ds}\boldsymbol{g}(s) = \boldsymbol{g}(s)\mathcal{T}\xi(s) \tag{3.2.17b}$$

for $s \in]0, L[$ and with

$$\boldsymbol{g} \in \Gamma \tag{3.2.17c}$$

subject to the constitutive relations (3.2.13), and for a choice of the configuration set Γ in (3.2.3). Observe that the conditions in (3.2.17) can be seen as a continuum description of the analogous chain conditions (2.2.15).

3.2.3 Hamiltonian formulation of rod equilibrium conditions

An interesting property of the rod equilibrium conditions in (3.2.17) is that they can be shown to admit a Hamiltonian structure. For compactness, we introduce the notation

$$\zeta(s) = \begin{pmatrix} \mathsf{m}(s) \\ \mathsf{n}(s) \end{pmatrix} \in \mathbb{R}^6 \tag{3.2.18}$$

which allow the constitutive relation (3.2.13) to be written as

$$\zeta(s) = \partial_\xi \mathsf{W}(\xi(s); s). \tag{3.2.19}$$

Using the famous *Legendre transform* of the function W, defined as

$$\mathsf{H}(\zeta; s) = \max_{\zeta\in\mathbb{R}^6}\left\{\xi\cdot\zeta - \mathsf{W}(\xi; s)\right\}, \tag{3.2.20}$$

one can write the inverse of the mapping in (3.2.19) as

$$\xi(s) = \partial_\zeta \mathsf{H}(\zeta(s); s) \tag{3.2.21}$$

for any function W strictly convex in ξ. The function H is here called the *rod Hamiltonian*. The equilibrium conditions (3.2.17) then become for $s \in]0, L[$

$$\frac{d}{ds}\zeta(s) = \mathrm{ad}^{\dagger}_{\zeta(s)}\partial_{\zeta}\mathsf{H}(\zeta(s); s) \tag{3.2.22a}$$

$$\frac{d}{ds}\boldsymbol{g}(s) = \boldsymbol{g}(s)\mathcal{T}\partial_{\zeta}\mathsf{H}(\zeta(s); s) \tag{3.2.22b}$$

$$\text{with } \boldsymbol{g} \in \Gamma \tag{3.2.22c}$$

where we have defined

$$\mathrm{ad}^{\dagger}_{\zeta(s)} = \begin{pmatrix} [\mathsf{m}(s)\times] & [\mathsf{n}(s)\times] \\ [\mathsf{n}(s)\times] & 0 \end{pmatrix} \text{ for all } \zeta(s) = (\mathsf{m}(s), \mathsf{n}(s)) \in \mathbb{R}^6, \tag{3.2.23}$$

and which satisfies $\mathrm{ad}^{T}_{\phi_1}\phi_2 = \mathrm{ad}^{\dagger}_{\phi_2}\phi_1$ for any triples $\phi_1, \phi_2 \in \mathbb{R}^6$, according to the definition of the operator ad in (1.1.34). Moreover, defining the *rod Hamiltonian variable*

$$\mathsf{z}(s) = (\zeta(s), \boldsymbol{g}(s)) \in \mathbb{R}^6 \times \mathrm{SE}(3) \tag{3.2.24}$$

one can write the rod equilibrium equations in an even more compact formulation as

$$\frac{d}{ds}\mathsf{z}(s) = \mathsf{J}\left(\mathsf{z}(s)\right)\partial_{\mathsf{z}}\mathsf{H}(\mathsf{z}(s); s) \tag{3.2.25a}$$

with

$$\mathsf{J}\left(\mathsf{z}(s)\right) = \begin{pmatrix} \mathrm{ad}^{\dagger}_{\zeta(s)} & -\mathcal{T}^{*}\boldsymbol{g}(s)^{T} \\ \boldsymbol{g}(s)\mathcal{T} & 0 \end{pmatrix} \tag{3.2.25b}$$

with $\boldsymbol{g} \in \Gamma$, and where the operators ad and $\mathcal{T}^{*}$ are defined in (3.2.16) and (1.1.23) respectively. Equations (3.2.25) constitute the *Hamiltonian form of the rod equilibrium equations*. We have to stress that the locality assumption on the internal rod energy E in (3.2.5) is key here to obtain the system (3.2.25). More detailed discussion of the rod Hamiltonian properties can be found for instance in [Dichmann et al., 1996, R.S. Manning, 1996, Kehrbaum, 1997, Rey, 2002, Chouaïeb, 2003, Ellis et al., 2010].

3.2.4 Continuum double rod equilibrium conditions and constitutive relations

The goal of this section is to define *continuum birod equilibrium configurations* in an analogous way to continuum rods. However since the birod configurations $(\boldsymbol{g}, \mathcal{P})$ are an alternative expression of double rod configurations $(\boldsymbol{g}^{+}, \boldsymbol{g}^{-})$ it is of interest to first present the notion of *continuum double rod equilibrium configurations* and then discuss how they can be translated into an expression for birods. This strategy follows the pioneering work of [Moakher and Maddocks, 2005] in birod theory.

As for rigid body double chains, a way to deduce the expression of the balance laws is to see each strand of the double rod as a single continuum rod in an external field. Formally, if for each cross section whose configuration is given by $\boldsymbol{g}^{\pm}(s)$ we denote by $\mathbf{m}^{\pm}(s) \in \mathbb{R}^3$ the total internal couple around the point $\boldsymbol{r}^{\pm}(s)$ and by $\mathbf{n}^{\pm}(s) \in \mathbb{R}^3$ the analogous total force, similarly to the single continuum rod case (3.2.1), the *double rod balance laws* can be stated as

$$-\frac{d}{ds}\left[\mathbf{m}^{\pm}(s) + \boldsymbol{r}^{\pm}(s) \times \mathbf{n}^{\pm}(s)\right] + \mathbf{c}^{\pm}(s) + \boldsymbol{r}^{\pm}(s) \times \mathbf{f}^{\pm}(s) = 0 \tag{3.2.26a}$$

$$-\frac{d}{ds}\mathbf{n}^{\pm}(s) + \mathbf{f}^{\pm}(s) = 0 \tag{3.2.26b}$$

for $s \in]0, L[$ and where $\mathbf{c}^{\pm}(s) \in \mathbb{R}^3$ and $\mathbf{f}^{\pm}(s) \in \mathbb{R}^3$ denote respectively the total external couple (around the point $\boldsymbol{r}^{\pm}(s)$) density and the total external force density at s on the strand $\pm$. A double rod is said to be an *isolated double rod* if there is no external contribution to the total couples and forces acting on the double chain. A similar expression of the balance laws (3.2.26) can be found in [Moakher and Maddocks, 2005].

Based on the double chain end conditions (2.2.23), the *double rod end conditions* can be stated to be of the form

$$\Gamma = \left\{ (\boldsymbol{g}^{+}, \boldsymbol{g}^{-}) \in \left(\mathrm{SE}(3)^{]0,L[}\right)^2 \mid \boldsymbol{g}^{+} \text{ and } \boldsymbol{g}^{-} \text{ satisfy i. or ii.} \right\} \tag{3.2.27}$$

where the conditions i. and ii. refer to the single rod case in (3.2.3). In particular, we recall that we do not allow that $\boldsymbol{g}^{+}$ satisfies i. and $\boldsymbol{g}^{-}$ satisfies ii. . The accuracy of these end conditions in the context of DNA modelling has of course to be discussed, but we believe that they serve as instructive starting points.

The *double rod constitutive relations* are then formally a mapping $(\boldsymbol{g}^{+}, \boldsymbol{g}^{-}) \mapsto (\mathbf{m}^{+}, \mathbf{m}^{-}, \mathbf{n}^{+}, \mathbf{n}^{-}, \mathbf{c}^{+}, \mathbf{c}^{-}, \mathbf{f}^{+}, \mathbf{f}^{-})$. In the case of an isolated double rod, we can define the notion of *local double rod constitutive relations*, which means that the local couple and force in the double rod are only functions of the local deformations, i.e.

$$\begin{pmatrix} \mathbf{m}^{\pm}(s) \\ \mathbf{n}^{\pm}(s) \end{pmatrix} = \boldsymbol{\Psi}^{(\mathbf{m},\mathbf{n})}(b(s), \xi^{+}(s), \xi^{-}(s); s) \tag{3.2.28a}$$

and

$$\begin{pmatrix} \mathbf{c}^{\pm}(s) \\ \mathbf{f}^{\pm}(s) \end{pmatrix} = \boldsymbol{\Psi}^{(\mathbf{c},\mathbf{f})}(b(s), \xi^{+}(s), \xi^{-}(s); s) \tag{3.2.28b}$$

for $s \in]0, L[$.

As in the single rod case (3.2.4), the constitutive relations reflect the physical constitution of the double rod and their explicit determination for a given material requires subtle studies. The *double rod equilibrium conditions* then comprise the conditions (3.2.26) on a configuration set Γ together with a given set of constitutive relations, of the form (3.2.28) for instance.

3.2.5 On continuum double rod internal energy and variational principle

The *double rod internal energy* is denoted by $E^{||}$, as in rigid body double chain theory. It describes the stored potential energy associated to internal deformations of the continuum double rod system. A *local energy*, for the double rod is then defined as $E^{||} : \left(\mathrm{SE}(3)^{]0,L[}\right)^2 \to \mathbb{R}$ of the form

$$E^{||}[\boldsymbol{g}^+, \boldsymbol{g}^-] = \int_0^L \mathsf{W}^{||}(b(s), \xi^+(s), \xi^-(s); s)ds \tag{3.2.29}$$

where $\xi^{\pm}(s) \in \mathbb{R}^6$ is the rod generator vector along each strand $\pm$ respectively and $b(s) \in \mathrm{SE}(3)$ denotes the rigid body displacement from the cross section whose configuration is described by $\boldsymbol{g}^+(s)$ to $\boldsymbol{g}^-(s)$ on the complementary strand, as defined in (3.1.17). Observe that, similarly to the single rod internal energy in (3.2.6), the double rod energy $E^{||}$ is invariant through a global rigid body motion of the double rod, i.e.

$$E^{||}[B\boldsymbol{g}^+, B\boldsymbol{g}^-] = E^{||}[\boldsymbol{g}^+, \boldsymbol{g}^-] \tag{3.2.30}$$

for any $B \in \mathrm{SE}(3)$. It is worth to emphasize that when the internal double rod energy $E^{||}$ is assumed to be local, the two individual rods $\boldsymbol{g}^{\pm}$ are governed by an effective nonlocal energy, similarly to double chains. The internal energy $E^{||}$ allows the statement of a *variational principle for continuum double rods*, and if the double rod constitutive relations (3.2.28) admit a potential energy it will be called a *hyper-elastic double rod*.

More precisely, suppose that Γ denote the set of double rod configurations i. in (3.2.27). Using conditions of the form (3.2.7b) and (3.2.7c) for the perturbation of the double rod configuration $(\boldsymbol{g}^+, \boldsymbol{g}^-)$, and using moreover that $\delta b(0) = 0$

$$\delta b(s) = (\boldsymbol{g}^+(s))^{-1}(\mathcal{T}\boldsymbol{\phi}^-(s) - \mathcal{T}\boldsymbol{\phi}^+(s))\boldsymbol{g}^-(s) \tag{3.2.31}$$

for $s \in]0, L[$ and $\delta b(L) = 0$, implies that a double rod configuration $(\boldsymbol{g}^+, \boldsymbol{g}^-)$ makes the internal energy $E^{||}$ in (3.2.29) stationary if it satisfies the double chain balance laws (3.2.26) for

$$\mathsf{m}^{\pm}(s) = \boldsymbol{R}^{\pm}(s)\partial_{U^{\pm}}\mathsf{W}^{||} \tag{3.2.32a}$$

$$\mathsf{n}^{\pm}(s) = \boldsymbol{R}^{\pm}(s)\partial_{V^{\pm}}\mathsf{W}^{||} \tag{3.2.32b}$$

for $s \in]0, L[$ and

$$\mathsf{c}^-(s) = \boldsymbol{R}^+(s)\mathrm{Vect}\left(\partial_P \mathsf{W}^{||} P(s)^T\right) \tag{3.2.32c}$$

$$\mathsf{f}^-(s) = \boldsymbol{R}^+(s)\partial_p \mathsf{W}^{||} \tag{3.2.32d}$$

$$\mathsf{c}^+(s) = -\mathsf{c}^-(s) - (\boldsymbol{r}^-(s) - \boldsymbol{r}^+(s)) \times \mathsf{f}^-(s) \tag{3.2.32e}$$

$$\mathbf{f}^+(s) = -\mathbf{f}^-(s) \tag{3.2.32f}$$

for $s \in]0, L[$ and where the argument $(b(s), \xi^+(s), \xi^-(s); s)$ of the internal energy $\mathsf{W}^{||}$ has been hidden for compactness. Observe that the apparent asymmetry between the definitions of the couple $\mathbf{c}^-(s)$ and $\mathbf{c}^+(s)$ and between the force $\mathbf{f}^-(s)$ and $\mathbf{f}^+(s)$ is explained by the fact the internal energy $E^{||}$ is not invariant through the transformation $\boldsymbol{g}^+ \mapsto \boldsymbol{g}^-$ presented in (3.1.36). The transformation rule is then obtained using $\delta b(s)^{-1} = -b(s)^{-1}\delta b(s) b(s)^{-1}$.

As for a single continuum rod, the variational principle for the double rod balance laws in (3.2.26) together with the constitutive relations (3.2.32) can also be obtained when the set Γ denotes the condition ii. (3.2.27). Consequently, if the double rod constitutive relations in (3.2.28) admit a variational representation of the form (3.2.32) for all configurations $(\boldsymbol{g}^+, \boldsymbol{g}^-)$, then the stationary conditions for the internal energy $E^{||}$ on the set of configurations Γ is an expression of the equilibrium conditions (3.2.26) for isolated double rods. This expression is then called a *continuum Euler-Poincaré variation for double rods*.[6]

In their respective director frames $\boldsymbol{R}^\pm(s)$ the constitutive relations (3.2.32) reads then

$$\mathsf{m}^\pm(s) = \partial_{U^\pm}\mathsf{W}^{||} \tag{3.2.33a}$$

$$\mathsf{n}^\pm(s) = \partial_{V^\pm}\mathsf{W}^{||} \tag{3.2.33b}$$

for $s \in]0, L[$ and

$$\mathsf{c}^-(s) = P(s)^T \mathrm{Vect}\left(\partial_P \mathsf{W}^{||} P(s)^T\right) \tag{3.2.33c}$$

$$\mathsf{f}^-(s) = P(s)^T \partial_{p(s)}\mathsf{W}^{||} \tag{3.2.33d}$$

$$\mathsf{c}^+(s) = -P(s)\mathsf{c}^-(s) + p(s) \times P(s)\mathsf{f}^-(s) \tag{3.2.33e}$$

$$\mathsf{f}^+(s) = -P(s)\mathsf{f}^-(s) \tag{3.2.33f}$$

for $s \in]0, L[$ and consequently, the double rod equilibrium conditions can be summarized as

$$\frac{d}{ds}\begin{pmatrix}\mathsf{m}^\pm(s)\\ \mathsf{n}^\pm(s)\end{pmatrix} = \mathrm{ad}^T_{\xi^\pm(s)}\begin{pmatrix}\mathsf{m}^\pm(s)\\ \mathsf{n}^\pm(s)\end{pmatrix} + \begin{pmatrix}\mathsf{c}^\pm(s)\\ \mathsf{f}^\pm(s)\end{pmatrix} \tag{3.2.34a}$$

$$\boldsymbol{g}^\pm(s) = \boldsymbol{g}^\pm(s)\mathcal{T}\xi^\pm(s) \tag{3.2.34b}$$

for $s \in]0, L[$ and with

$$(\boldsymbol{g}^+, \boldsymbol{g}^-) \in \Gamma \tag{3.2.34c}$$

subject to the constitutive relations (3.2.33), and for a choice of the configuration set Γ in (3.2.27), where we recall that the operator $\mathrm{ad}_{\boldsymbol{g}}$ is defined in (3.2.16). Observe that the evolution equations in (3.2.34) can be seen as a continuum version of the ones in (2.2.31).

[6] See section 1.2 for more detail

3.2.6 Hamiltonian formulation of double rod equilibrium conditions

Since the double rod equilibrium conditions in (3.2.34) is made of two copies of the rod equilibrium conditions in (3.2.17) with couples and forces which act between the two rods, a similar Hamiltonian formulation to the one for a single rod (3.2.25) can be deduced.

Introducing the notation

$$\zeta^{\pm}(s) = \begin{pmatrix} \mathsf{m}^{\pm}(s) \\ \mathsf{n}^{\pm}(s) \end{pmatrix} \tag{3.2.35}$$

and

$$\mathcal{F}^{\pm}(s) = \begin{pmatrix} \mathsf{c}^{\pm}(s) \\ \mathsf{f}^{\pm}(s) \end{pmatrix} \tag{3.2.36}$$

the constitutive relation in (3.2.33) can be written as

$$\zeta^{\pm}(s) = \partial_{\xi^{\pm}} \mathsf{W}^{||} \tag{3.2.37a}$$

$$\mathcal{F}^{\pm}(s) = \mathcal{T}^{*}\left(\boldsymbol{g}^{\pm}(s)\right)^{T} \partial_{\boldsymbol{g}^{\pm}} \mathsf{W}^{||} \tag{3.2.37b}$$

where the arguments $(b(s), \xi^{+}(s), \xi^{-}(s); s)$ of the function $\mathsf{W}^{||}$ have been suppressed for compactness and where the operator $\mathcal{T}^{*}$ is defined in (1.1.23). More detail about the expressions in (3.2.37) can be found in section 1.2.

Similarly to (3.2.20) we introduce the *Legendre transform* $\mathsf{H}^{||}$ of the function $\mathsf{W}^{||}$ which is defined as

$$\mathsf{H}^{||}(\zeta^{+}, \boldsymbol{g}^{+}, \zeta^{-}, \boldsymbol{g}^{-}; s) = \max_{(\zeta^{+}, \zeta^{-}) \in \mathbb{R}^{6}} \left\{ \xi^{+} \cdot \zeta^{+} + \xi^{-} \cdot \zeta^{-} - \mathsf{W}^{||}\left((\boldsymbol{g}^{+})^{-1}\boldsymbol{g}^{-}, \xi^{+}, \xi^{-}; s\right) \right\} \tag{3.2.38}$$

which has the property that

$$\xi^{\pm}(s) = \partial_{\zeta^{\pm}} \mathsf{H}^{||} \tag{3.2.39a}$$

$$\mathcal{F}^{\pm}(s) = -\mathcal{T}^{*}\left(\boldsymbol{g}^{\pm}(s)\right)^{T} \partial_{\boldsymbol{g}^{\pm}} \mathsf{H}^{||} \tag{3.2.39b}$$

for any function W strictly convex in ξ according to the relations in (3.2.37) and where the arguments $(\zeta^{+}, \boldsymbol{g}^{+}, \zeta^{-}, \boldsymbol{g}^{-}; s)$ of the function $\mathsf{H}^{||}$ have been suppressed for compactness. The function H is here called the *double rod Hamiltonian*. Using the notation

$$\mathsf{z}^{\pm}(s) = \left(\zeta^{\pm}(s), \boldsymbol{g}^{\pm}(s)\right) \tag{3.2.40}$$

for the *double rod Hamiltonian variables*, similarly to the rod case in (3.2.24), the equilibrium conditions (3.2.34) become for $s \in]0, L[$

$$\frac{d}{ds}\begin{pmatrix} \mathsf{z}^{+}(s) \\ \mathsf{z}^{-}(s) \end{pmatrix} = \begin{pmatrix} \mathsf{J}\left(\mathsf{z}^{+}(s)\right) & 0 \\ 0 & \mathsf{J}\left(\mathsf{z}^{-}(s)\right) \end{pmatrix} \begin{pmatrix} \partial_{\mathsf{z}^{+}} \mathsf{H}^{||}(\mathsf{z}^{+}(s), \mathsf{z}^{-}(s); s) \\ \partial_{\mathsf{z}^{-}} \mathsf{H}^{||}(\mathsf{z}^{+}(s), \mathsf{z}^{-}(s); s) \end{pmatrix} \tag{3.2.41}$$

with $(\boldsymbol{g}^{+}, \boldsymbol{g}^{-}) \in \Gamma$. Note that the operator $\mathsf{J}(\,.\,)$ is defined in (3.2.25b). The equations (3.2.41) constitute the *Hamiltonian form of the double rod equilibrium equations*.

3.2.7 On continuum birod variational principle and constitutive relations

In contrast to double rods, the *birod internal energy* is denoted by $E : \left(\mathrm{SE}(3)^{]0,L[}\right)^2 \to \mathbb{R}$, as in the single rod case (3.2.5), and is defined as

$$E[\boldsymbol{g}, \mathcal{P}] = \int_0^L \mathsf{W}\left(\mathcal{P}(s), \xi^{\mathcal{P}}(s), \xi(s); s\right) ds \tag{3.2.42}$$

where $\xi^{\mathcal{P}}(s)$ and $\xi(s)$ denote the generator vector of the microstructure and macrostructure respectively, $\mathcal{P}(s)$ is the microstructure configuration, and where the energy terms $\mathsf{W}\left(\mathcal{P}(s), \xi^{\mathcal{P}}(s), \xi(s); s\right)$ are defined such that the total birod internal energy satisfies

$$E[\boldsymbol{g}, \mathcal{P}] = E^{||}[\boldsymbol{g}^+, \boldsymbol{g}^-] \tag{3.2.43}$$

for the corresponding double rod configuration, as described in (3.1.18). In particular the terms $\xi^{\mathcal{P}}(s)$ and $\xi(s)$ appear naturally from the relations (3.1.22). It is then equivalent to study double rod configurations $(\boldsymbol{g}^+, \boldsymbol{g}^-)$ which make the internal energy $E^{||}$ stationary and birod configurations $(\boldsymbol{g}, \mathcal{P})$ which make E stationary. In particular, the birod energy is internal in the sense that it satisfies invariance through global rigid body motion of the birod configuration, i.e.

$$E[B\boldsymbol{g}, \mathcal{P}] = E[\boldsymbol{g}, \mathcal{P}] \tag{3.2.44}$$

for any $B \in \mathrm{SE}(3)$.

More precisely, in the case of the Dirichlet conditions for double rods in (3.2.27) one also gets Dirichlet conditions for the birod macrostructure and microstructure $(\boldsymbol{g}, \mathcal{P})$. The first variation of the internal energy E can then be shown to be

$$-\frac{d}{ds}\left[\mathbf{m}(s) + \boldsymbol{r}(s) \times \mathbf{n}(s)\right] = 0 \tag{3.2.45a}$$

$$-\frac{d}{ds}\mathbf{n}(s) = 0 \tag{3.2.45b}$$

$$-\frac{d}{ds}\mathsf{m}^{\mathcal{P}}(s) + \frac{1}{2}\left[\left(M^-(s) - M^-(s)\right)\mathsf{m}^{\mathcal{P}}(s)\times\right] U_P(s) + \mathsf{c}^{\mathcal{P}}(s) = 0 \tag{3.2.45c}$$

$$-\frac{d}{ds}\mathsf{n}^{\mathcal{P}}(s) + \mathsf{f}^{\mathcal{P}}(s) = 0 \tag{3.2.45d}$$

for $s \in]0, L[$ with the relations

$$\mathbf{m}(s) = \boldsymbol{R}(s)\partial_U \mathsf{W} \tag{3.2.46a}$$

$$\mathbf{n}(s) = \boldsymbol{R}(s)\partial_V \mathsf{W} \tag{3.2.46b}$$

$$\mathsf{m}^{\mathcal{P}}(s) = \partial_{U_P} \mathsf{W} \tag{3.2.46c}$$

$$\mathsf{n}^{\mathcal{P}}(s) = \partial_{V_{\mathsf{w}}}\mathsf{W} \tag{3.2.46d}$$

$$\mathsf{c}^{\mathcal{P}}(s) = \mathrm{Vect}\left(P^{\frac{T}{2}}(s)\partial_P \mathsf{W} P^{\frac{T}{2}}(s)\right) \tag{3.2.46e}$$

$$\mathsf{f}^{\mathcal{P}}(s) = \partial_{\mathsf{w}}\mathsf{W} \tag{3.2.46f}$$

for $s \in]0, L[$ and where the argument $\big(\mathcal{P}(s), \xi^{\mathcal{P}}(s), \xi(s); s\big)$ of the internal energy W has been suppressed for compactness. Note that we have used the following symmetric perturbation of the microstructure $\mathcal{P}(s)$

$$\delta\mathcal{P}(s) = \mathcal{T}^{\mathcal{P}}\phi(s) \tag{3.2.47}$$

for some differentiable function $\phi(s) \in \mathbb{R}^6$ and where the operator $\mathcal{T}^{\mathcal{P}}$ is defined in (3.1.27). The perturbation for the generator vector $\xi^{\mathcal{P}}(s) = (U_P(s), V_{\mathsf{w}}(s))$ of the microstructure then satisfy

$$\delta U_P(s) = \frac{d}{ds}\phi^{\boldsymbol{R}}(s) + \left[M^-(s)\phi^{\boldsymbol{R}}(s)\times\right] U_P(s) + \left[M^+(s)U_P(s)\times\right]\phi^{\boldsymbol{R}}(s) \tag{3.2.48a}$$

$$\delta V_{\mathsf{w}}(s) = \frac{d}{ds}\phi^{\boldsymbol{r}}(s) \tag{3.2.48b}$$

where $\phi(s) = (\phi^{\boldsymbol{R}}(s), \phi^{\boldsymbol{r}}(s))$ and the matrices $M^{\pm}(s)$ are defined in (3.1.25).

The equations (3.2.45a) and (3.2.45b) correspond to the single rod balance laws, as presented in (3.2.1), and describe the *macrostructure balance laws*. The triple $\mathbf{m}(s) \in \mathbb{R}^3$ denotes the internal couple around the point $\boldsymbol{r}(s)$ acting on the cross sections $\boldsymbol{g}^+(s)$ and $\boldsymbol{g}^-(s)$ in the double rod from the material parametrized by $]0, s[$, and $\mathbf{n}(s) \in \mathbb{R}^3$ the analogous force. In contrast, the equations (3.2.45c) and (3.2.45d) define the *microstructure balance laws*, where $\mathsf{c}^{\mathcal{P}}(s)$ and $\mathsf{f}^{\mathcal{P}}(s)$ denote respectively the couple and the force *in* the microstructure. The expressions in (3.2.45) are called *birod equilibrium conditions*. We should emphasize that the form of the equations (3.2.45) are different from the one appearing in [Moakher and Maddocks, 2005], because the internal energy is here expressed in a coordinate free formulation.

Using the double rod constitutive relations (3.2.32) and (3.2.33) together with the identities (3.1.22), the relations between the double rod couples and forces (3.2.32) and the birod ones (3.2.46) can be explicitly written as

$$\mathbf{m}(s) = \mathbf{m}^+(s) + \mathbf{m}^-(s) + \left[\boldsymbol{r}^-(s) - \boldsymbol{r}^+(s)\right] \times \boldsymbol{R}(s)\mathsf{n}^{\mathcal{P}}(s) \tag{3.2.49a}$$

$$\mathbf{n}(s) = \mathbf{n}^+(s) + \mathbf{n}^-(s) \tag{3.2.49b}$$

$$\mathsf{m}^{\mathcal{P}}(s) = M^-(s)\mathsf{m}^-(s) - M^+(s)\mathsf{m}^+(s) \tag{3.2.49c}$$

$$\mathsf{n}^{\mathcal{P}}(s) = \frac{1}{2}\left[P^+(s)\mathsf{n}^-(s) - P^-(s)\mathsf{n}^+(s)\right] \tag{3.2.49d}$$

and

$$\begin{aligned}\mathsf{c}^{\mathcal{P}} &= M^-(s)\left[\mathsf{m}^-(s)\times U^-(s)+\mathsf{n}^-(s)\times V^-(s)+\mathsf{c}^-(s)\right]\\ &\quad -M^+(s)\left[\mathsf{m}^+(s)\times U^+(s)+\mathsf{n}^+(s)\times V^+(s)+\mathsf{c}^+(s)\right]\\ &\quad +\frac{d}{ds}\left(M^+(s)\right)P^+(s)\mathsf{m}^-(s)-\frac{d}{ds}\left(M^-(s)\right)P^-(s)\mathsf{m}^+(s)\end{aligned} \tag{3.2.49e}$$

$$\mathsf{f}^{\mathcal{P}}(s)=\frac{1}{2}\left[P^+(s)\mathsf{f}^-(s)-P^-(s)\mathsf{f}^+(s)\right]-U(s)\times\mathsf{n}^{\mathcal{P}}(s) \tag{3.2.49f}$$

for $s\in]0,L[$ and where the presence of the matrices $M^{\pm}(s)$ is explained by the fact that one needs to differentiate the matrix $P^{\frac{1}{2}}(s)$, according to the expression of the double rod generator vector (3.1.22). More details about the identities in (3.2.49) can be found in appendix A.3.

It is interesting to observe that the relations in (3.2.49) are sums and weighted differences of the expressions in (3.2.26). In particular, the variables $\mathbf{m}(s)$ and $\mathbf{n}(s)$ are respectively interpreted as the total couple (around the point $\boldsymbol{r}(s)$) and force acting on the rigid bodies $\boldsymbol{g}^+(s)$ and $\boldsymbol{g}^-(s)$. The double rod equilibrium conditions and the birod equilibrium equations under Dirichlet conditions are then equivalent.

More generally, using the identities in (3.2.49), the double rod end conditions (3.2.27) written in terms of the birod configuration $(\boldsymbol{g},\mathcal{P})$ comprise the *birod end conditions* and are consequently defined as

i′. *Dirichlet conditions*: The macrostructure and the microstructure are prescribed at both ends, i.e.

$$(\boldsymbol{g}(0),\mathcal{P}(0)) \text{ and } (\boldsymbol{g}(L),\mathcal{P}(L)) \text{ are prescribed.} \tag{3.2.50a}$$

ii′. *Unstressed conditions*: The macrostructure and the microstructure are free at both ends, i.e.

$$\begin{aligned}-\mathbf{m}(0)=0,\ -\mathbf{n}(0)=0,\ \mathsf{m}^{\mathcal{P}}(0)=0,\ \mathsf{n}^{\mathcal{P}}(0)=0\\ \text{and }\ -\mathbf{m}(L)=0,\ -\mathbf{n}(L)=0,\ \mathsf{m}^{\mathcal{P}}(L)=0,\ \mathsf{n}^{\mathcal{P}}(L)=0\end{aligned} \tag{3.2.50b}$$

As in the single rigid rod case, we will denote generically by Γ the subset of birod configurations which satisfies one of the end conditions i′. or ii′., i.e.

$$\Gamma=\left\{(\boldsymbol{g},\mathcal{P})\in\left(\mathrm{SE}(3)^{]0,L[}\right)^2 \mid (\boldsymbol{g},\mathcal{P}) \text{ satisfies i′. or ii′.}\right\} \tag{3.2.51}$$

The variational principle for the double chain balance laws in (3.2.45) together with the constitutive relations (3.2.46) is also valid when the condition set Γ denotes the condition ii′. in (3.2.51).

Consequently, if the birod constitutive relations admit a variational representation of the form (3.2.46) for all birod configurations $(\boldsymbol{g},\mathcal{P})$, then the stationary conditions for the internal energy E on the set of configurations Γ is an expression of the birod

equilibrium conditions (3.2.45) for an isolated birod. This expression is then called the *continuum Euler-Poincaré variation for birods.*

As a corollary of the birod balance laws (3.2.45), the six quantities

$$\mathbf{m}(s) + \boldsymbol{r}(s) \times \mathbf{n}(s) \in \mathbb{R}^3 \text{ and } \mathbf{n}(s) \in \mathbb{R}^3 \tag{3.2.52}$$

are conserved along equilibrium birod configuration, or equivalently, along stationary configuration of the internal energy E. Similarly to the single continuum rod case in (3.2.12), the *Noether's Theorem* associates the conserved quantities in (3.2.52) to the invariance of internal double rod internal energy $E^{||}$ under global rigid body motions of the system in (3.2.30), or equivalently in terms of the birod internal energy E in (3.2.44).

In the director frame of the macrostructure $\boldsymbol{R}(s)$, the constitutive relations (3.2.46a) and (3.2.46b) become

$$\mathsf{m}(s) = \partial_U \mathsf{W} \tag{3.2.53a}$$

$$\mathsf{n}(s) = \partial_V \mathsf{W} \tag{3.2.53b}$$

for $s \in]0, L[$ where the argument $\big(\mathcal{P}(s), \xi^{\mathcal{P}}(s), \xi(s); s\big)$ of the internal energy W has been hidden for compactness. Consequently, the birod equilibrium conditions can be summarized as

$$\begin{aligned} \frac{d}{ds}\begin{pmatrix}\mathsf{m}(s)\\ \mathsf{n}(s)\end{pmatrix} &= \mathrm{ad}^T_{\xi(s)}\begin{pmatrix}\mathsf{m}(s)\\ \mathsf{n}(s)\end{pmatrix} \\ \frac{d}{ds}\begin{pmatrix}\mathsf{m}^{\mathcal{P}}(s)\\ \mathsf{n}^{\mathcal{P}}(s)\end{pmatrix} &= \begin{pmatrix}\frac{1}{2}\left[(M^-(s) - M^+(s))\,\mathsf{m}^{\mathcal{P}}(s)\times\right] U_P(s) + \mathsf{c}^{\mathcal{P}}(s)\\ \mathsf{f}^{\mathcal{P}}(s)\end{pmatrix} \end{aligned} \tag{3.2.54a}$$

$$\boldsymbol{g}(s) = \boldsymbol{g}(s)\mathcal{T}\xi(s) \tag{3.2.54b}$$

for $s \in]0, L[$ and with

$$(\boldsymbol{g}, \mathcal{P}) \in \Gamma \tag{3.2.54c}$$

subject to the constitutive relations (3.2.46), and for a choice of the configuration set Γ in (3.2.51). It is interesting to observe that the equilibrium conditions for a birod in (3.2.54) can be written as a system which contains the equilibrium conditions for the macrostructure, which are of the form of those for a single continuum rod (3.2.17), along with equilibrium conditions for the microstructure, and that two subsystems are coupled through the birod constitutive relations which reflect the original double rod structure, as already noticed in [Moakher and Maddocks, 2005].

The main application of the birod theory that we will discuss is in the context of DNA modelling, where a realistic function E will not be given in terms of the rotation $P(s)$ used to describe the microstructure, but in terms of a parametrization vector of this rotation matrix. It this then useful to describe the birod internal energy in terms of the generator vector of macrostructure $\xi(s)$ and a set of internal coordinates $y(s)$ for the microstructure.

If one uses the continuum analogue of internal coordinate in (2.1.33) to define

$$y(s) = (\eta(s), \mathsf{w}(s)) \tag{3.2.55}$$

we obtain that the generator vector of the microstructure $\xi^{\mathcal{P}}(s) = (U_P(s), V_{\mathsf{w}}(s))$ is expressed by

$$U_P(s) = \mathbb{B}_{\eta(s)} U_\eta(s) \text{ with } U_\eta(s) = \frac{d}{ds}\eta(s) \text{ and } V_{\mathsf{w}}(s) = \frac{d}{ds}\mathsf{w}(s) \tag{3.2.56}$$

according to the expression in (2.2.48) and where the matrix $\mathbb{B}_{\eta(s)}$ is defined in (2.2.50). The internal birod energy then reads as

$$E\left[\boldsymbol{g}, \mathcal{P}(y)\right] = \int_0^L \mathrm{W}\left(y(s), \frac{d}{ds}y(s), \xi(s); s\right) ds \tag{3.2.57}$$

with

$$\mathrm{W}\left(y(s), \frac{d}{ds}y(s), \xi(s); s\right) = \mathsf{W}\left(\mathcal{P}(s), \xi^{\mathcal{P}}(s), \xi(s)\right) \tag{3.2.58}$$

for all birod configuration and for all $s \in]0, L[$, according to the identities in (3.2.56). The stationary conditions for the birod internal energy become

$$-\frac{d}{ds}\left[\mathbf{m}(s) + \boldsymbol{r}(s) \times \mathbf{n}(s)\right] = 0 \tag{3.2.59a}$$

$$-\frac{d}{ds}\mathbf{n}(s) = 0 \tag{3.2.59b}$$

$$-\frac{d}{ds}\mathrm{m}^{\mathcal{P}}(s) + \mathrm{c}^{\mathcal{P}}(s) = 0 \tag{3.2.59c}$$

$$-\frac{d}{ds}\mathsf{n}^{\mathcal{P}}(s) + \mathsf{f}^{\mathcal{P}}(s) = 0 \tag{3.2.59d}$$

for $s \in]0, L[$ with the relations

$$\mathbf{m}(s) = \boldsymbol{R}(s)\partial_U \mathrm{W} \tag{3.2.60a}$$

$$\mathbf{n}(s) = \boldsymbol{R}(s)\partial_V \mathrm{W} \tag{3.2.60b}$$

$$\mathrm{m}^{\mathcal{P}}(s) = \partial_{U_\eta}\mathrm{W} \tag{3.2.60c}$$

$$\mathsf{n}^{\mathcal{P}}(s) = \partial_{V_{\mathsf{w}}}\mathrm{W} \tag{3.2.60d}$$

$$\mathrm{c}^{\mathcal{P}}(s) = \partial_\eta \mathrm{W} \tag{3.2.60e}$$

$$\mathsf{f}^{\mathcal{P}}(s) = \partial_{\mathsf{w}}\mathrm{W} \tag{3.2.60f}$$

for $s \in]0, L[$ and where the argument $\left(y(s), \frac{d}{ds}y(s), \xi(s); s\right)$ of the internal energy W has been suppressed for compactness. The conditions (3.2.59) together with the relations

(3.2.60) are called the *equilibrium birod conditions in internal coordinates*. It is then equivalent to solve equilibrium conditions (3.2.45) or (3.2.59) with the corresponding constitutive relations (3.2.46) or (3.2.60). The main difference between these two formulations is the expression of the moment balance in the microstructure in (3.2.45c) and (3.2.59c). The relation between the variables $\mathsf{m}^{\mathcal{P}}(s)$ and $\mathrm{m}^{\mathcal{P}}(s)$ is deduced, using the expression in (3.2.56), to be

$$\mathsf{m}^{\mathcal{P}}(s) = \mathbb{B}^{-T}_{\eta(s)} \mathrm{m}^{\mathcal{P}}(s) \tag{3.2.61}$$

whereas $\mathsf{c}^{\mathcal{P}}(s)$ and $\mathrm{c}^{\mathcal{P}}(s)$ satisfy

$$\mathsf{c}^{\mathcal{P}}(s) = \mathbb{B}^{-T}_{\eta(s)} \mathrm{c}^{\mathcal{P}}(s) \tag{3.2.62}$$

similarly to (2.2.49b).

3.2.8 Hamiltonian formulation of birod equilibrium conditions

An interesting property of the birod equilibrium conditions in (3.2.45) and (3.2.59) is that they can be shown to admit a Hamiltonian structure which is similar to the one deduced for rods in section 3.2.3.

We use the compact notation $\zeta(s) = (\mathsf{m}(s), \mathsf{n}(s)) \in \mathbb{R}^6$, presented in (3.2.18), and similarly we introduce

$$\zeta^{\mathcal{P}}(s) = \begin{pmatrix} \mathsf{m}^{\mathcal{P}}(s) \\ \mathsf{n}^{\mathcal{P}}(s) \end{pmatrix} \in \mathbb{R}^6 \tag{3.2.63}$$

and

$$\mathcal{F}^{\mathcal{P}}(s) = \begin{pmatrix} \mathsf{c}^{\mathcal{P}}(s) \\ \mathsf{f}^{\mathcal{P}}(s) \end{pmatrix} \in \mathbb{R}^6 \tag{3.2.64}$$

which allow the constitutive relation (3.2.46) and (3.2.53) to be written as

$$\zeta(s) = \partial_{\xi} \mathsf{W} \tag{3.2.65a}$$

$$\zeta^{\mathcal{P}}(s) = \partial_{\xi^{\mathcal{P}}} \mathsf{W} \tag{3.2.65b}$$

$$\mathcal{F}^{\mathcal{P}}(s) = \left(\mathcal{T}^{\mathcal{P}}\right)^{*} \partial_{\mathcal{P}} \mathsf{W} \tag{3.2.65c}$$

where the arguments $(\mathcal{P}(s), \xi^{\mathcal{P}}(s), \xi(s); s)$ of the function W have been suppressed for compactness and where we recall that the operator $\mathcal{T}^{\mathcal{P}}$ associated to perturbation of the microstructure is defined in (3.1.27).

Similarly to the rod case in (3.2.20), we denote by H the *Legendre transform* of the function W, but it is here defined as

$$\mathsf{H}(\mathcal{P}, \zeta^{\mathcal{P}}, \zeta; s) = \max_{(\zeta, \zeta^{\mathcal{P}}) \in \mathbb{R}^{12}} \left\{ \xi \cdot \zeta + \xi^{\mathcal{P}} \cdot \zeta^{\mathcal{P}} - \mathsf{W}(\mathcal{P}, \xi^{\mathcal{P}}, \xi; s) \right\}. \tag{3.2.66}$$

The function H is then called the *birod Hamiltonian*. According to (3.2.65) it satisfies

$$\xi(s) = \partial_{\zeta}\mathsf{H} \tag{3.2.67a}$$

$$\xi^{\mathcal{P}}(s) = \partial_{\zeta^{\mathcal{P}}}\mathsf{H} \tag{3.2.67b}$$

$$\mathcal{F}^{\mathcal{P}}(s) = -\left(\mathcal{T}^{\mathcal{P}}\right)^{*}\partial_{\mathcal{P}}\mathsf{H} \tag{3.2.67c}$$

for any function W strictly convex in the variables ξ and $\xi^{\mathcal{P}}$, and where the arguments $(\mathcal{P}(s), \zeta^{\mathcal{P}}(s), \zeta(s); s)$ of the function H have been suppressed for compactness. The *birod Hamiltonian variables* can be described as

$$\mathsf{z}(s) = (\mathsf{z}_{\boldsymbol{g}}(s), \mathsf{z}_{\mathcal{P}}(s)) \tag{3.2.68a}$$

where

$$\mathsf{z}_{\boldsymbol{g}}(s) = (\zeta(s), \boldsymbol{g}(s)) \in \mathbb{R}^6 \times \mathrm{SE}(3) \tag{3.2.68b}$$

are the conjugate variables associated to the macrostructure $\boldsymbol{g}$ and

$$\mathsf{z}_{\mathcal{P}}(s) = (\zeta^{\mathcal{P}}(s), \mathcal{P}(s)) \in \mathbb{R}^6 \times \mathrm{SO}(3) \times \mathbb{R}^3 \tag{3.2.68c}$$

to the microstructure $\mathcal{P}$. The equilibrium conditions (3.2.54) can then be written for $s \in]0, L[$ as

$$\frac{d}{ds}\mathsf{z}_{\boldsymbol{g}}(s) = \mathsf{J}_{\boldsymbol{g}}\left(\mathsf{z}_{\boldsymbol{g}}(s)\right)\partial_{\mathsf{z}_{\boldsymbol{g}}}\mathsf{H}(\mathsf{z}(s); s) \tag{3.2.69a}$$

$$\frac{d}{ds}\mathsf{z}_{\mathcal{P}}(s) = \mathsf{J}_{\mathcal{P}}\left(\mathsf{z}_{\mathcal{P}}(s)\right)\partial_{\mathsf{z}_{\mathcal{P}}}\mathsf{H}(\mathsf{z}(s); s) \tag{3.2.69b}$$

$$\text{with } (\boldsymbol{g}, \mathcal{P}) \in \Gamma \tag{3.2.69c}$$

where the operator $\mathsf{J}_{\boldsymbol{g}}(\,.\,)$ is the one appearing in the rod Hamiltonian system in (3.2.25) and the operator $\mathsf{J}_{\mathcal{P}}(\,.\,)$ is defined as

$$\mathsf{J}_{\mathcal{P}}\left(\mathsf{z}_{\mathcal{P}}\right) = \begin{pmatrix} \begin{pmatrix} \frac{1}{2}\left[(M^{-} - M^{+})\,\mathsf{m}^{\mathcal{P}}\times\right] & 0 \\ 0 & 0 \end{pmatrix} & -\left(\mathcal{T}^{\mathcal{P}}\right)^{*} \\ \mathcal{T}^{\mathcal{P}} & 0 \end{pmatrix}. \tag{3.2.70}$$

Consequently, the *Hamiltonian form of the birod equilibrium conditions* can be written as

$$\frac{d}{ds}\mathsf{z}(s) = \mathsf{J}\left(\mathsf{z}(s)\right)\partial_{\mathsf{z}}\mathsf{H}(\mathsf{z}(s); s) \tag{3.2.71a}$$

with

$$\mathsf{J}\left(\mathsf{z}(s)\right) = \begin{pmatrix} \mathsf{J}_{\boldsymbol{g}}\left(\mathsf{z}_{\boldsymbol{g}}(s)\right) & 0 \\ 0 & \mathsf{J}_{\mathcal{P}}\left(\mathsf{z}_{\mathcal{P}}(s)\right) \end{pmatrix} \tag{3.2.71b}$$

$$(\boldsymbol{g}, \mathcal{P}) \in \Gamma. \tag{3.2.71c}$$

In the case where the birod internal energy E is given in terms of internal birod variables, as in (3.2.57), then the birod equilibrium conditions take the form presented in (3.2.59) according to the constitutive relations in (3.2.60). If one denotes, analogously to (3.2.18), (3.2.63) and (3.2.64), by

$$\zeta(s) = (\mathsf{m}(s), \mathsf{n}(s)),\ \zeta_y^{\mathcal{P}}(s) = (\mathrm{m}^{\mathcal{P}}(s), \mathsf{n}^{\mathcal{P}}(s)) \text{ and } \mathcal{F}_y^{\mathcal{P}}(s) = (\mathrm{c}^{\mathcal{P}}(s), \mathsf{f}^{\mathcal{P}}(s)) \tag{3.2.72}$$

where in particular $\mathrm{m}^{\mathcal{P}}(s)$ and $\mathrm{c}^{\mathcal{P}}(s)$ are the variables appearing in (3.2.60), then the *birod Hamiltonian in internal coordinates*, denoted by H, is defined as

$$\mathrm{H}(y, \zeta_y^{\mathcal{P}}, \zeta; s) = \max_{(\zeta, \zeta_y^{\mathcal{P}}) \in \mathbb{R}^{12}} \left\{ \xi \cdot \zeta + \xi_y^{\mathcal{P}} \cdot \zeta_y^{\mathcal{P}} - \mathrm{W}(y, \xi_y^{\mathcal{P}}, \xi; s) \right\} \tag{3.2.73}$$

where we have used the notation $\xi_y^{\mathcal{P}}(s) = \frac{d}{ds} y(s)$. Defining the variable

$$\mathrm{z}(s) = (\mathsf{z}_{\boldsymbol{g}}(s), \mathrm{z}_y(s)) \tag{3.2.74}$$

where $\mathsf{z}_{\boldsymbol{g}}(s)$ is the set of Hamiltonian variables associated to the macrostructure in (3.2.68b) and where

$$\mathrm{z}_y(s) = (\zeta_y^{\mathcal{P}}(s), y(s)) \in \mathbb{R}^{12} \tag{3.2.75}$$

are the conjugate variable related to the microstructure, the birod equilibrium equations in terms of the internal coordinates $y(s)$ (3.2.59) can be written as the Hamiltonian system

$$\frac{d}{ds}\mathrm{z}(s) = J\left(\mathrm{z}(s)\right) \partial_{\mathrm{z}} \mathrm{H}(\mathrm{z}(s); s) \tag{3.2.76a}$$

with

$$J\left(\mathrm{z}(s)\right) = \begin{pmatrix} \mathsf{J}_{\boldsymbol{g}}\left(\mathsf{z}_{\boldsymbol{g}}(s)\right) & 0 \\ 0 & \begin{pmatrix} 0 & -\mathrm{Id}_6 \\ \mathrm{Id}_6 & 0 \end{pmatrix} \end{pmatrix} \tag{3.2.76b}$$

with $(\boldsymbol{g}, y) \in \Gamma$, and where $\mathsf{J}_{\boldsymbol{g}}(\,.\,)$ is defined in (3.2.25). The expressions (3.2.76) are the *Hamiltonian form of birod the equilibrium conditions in internal coordinates*. An application of this Hamiltonian system is presented in section 6.2.

It is worthwhile to observe that the birod Hamiltonian function H and the double rod Hamiltonian function $\mathsf{H}^{||}$, defined in (3.2.38), take the same value for corresponding descriptions of equilibrium birod configurations, i.e.

$$\mathsf{H}(\mathsf{z}_{\boldsymbol{g}}(s), \mathsf{z}_{\mathcal{P}}(s); s) = \mathsf{H}^{||}(\mathsf{z}^{+}(s), \mathsf{z}^{-}(s); s) \tag{3.2.77}$$

for all equilibrium birod configuration, as can be proven using the relations between the birod and double rod constitutive relations in (2.2.38). The transformation rule

between the birod and the double rod Hamiltonian variables are summarised in table 3.1. Similarly the functions H and H also have the same value for corresponding descriptions of equilibrium birod configurations, i.e.

$$\mathsf{H}(\mathsf{z}_{\boldsymbol{g}}(s), \mathsf{z}_{\mathcal{P}}(s); s) = \mathrm{H}(\mathsf{z}_{\boldsymbol{g}}(s), \mathrm{z}_y(s); s) \tag{3.2.78}$$

as can be proven using the relations (3.2.61) and (3.2.62).

Table 3.1: Relations between birod and double rod Hamiltonian variables

Birod	Double Rod
$\boldsymbol{R}$	$\boldsymbol{R}^+ \left[\boldsymbol{R}^{+T}\boldsymbol{R}^-\right]^{\frac{1}{2}}$
$\boldsymbol{r}$	$\frac{1}{2}\left(\boldsymbol{r}^+ + \boldsymbol{r}^-\right)$
P	$\boldsymbol{R}^{+T}\boldsymbol{R}^-$
w	$\boldsymbol{R}^T\left[\boldsymbol{r}^- - \boldsymbol{r}^+\right]$
m	$P^{\frac{T}{2}}\mathsf{m}^+ + P^{\frac{1}{2}}\mathsf{m}^- + \mathsf{w}\times\mathsf{n}^{\mathcal{P}}$
n	$P^{\frac{1}{2}}\mathsf{n}^- + P^{\frac{T}{2}}\mathsf{n}^+$
$\mathsf{m}^{\mathcal{P}}$	$M^-\mathsf{m}^- - M^+\mathsf{m}^+$
$\mathsf{n}^{\mathcal{P}}$	$\frac{1}{2}\left[P^{\frac{1}{2}}\mathsf{n}^- - P^{\frac{T}{2}}\mathsf{n}^+\right]$

Birod	Double Rod
$\boldsymbol{R}P^{\frac{T}{2}}$	$\boldsymbol{R}^+$
$\boldsymbol{R}P^{\frac{1}{2}}$	$\boldsymbol{R}^-$
$\boldsymbol{r} - \frac{1}{2}\boldsymbol{R}\mathsf{w}$	$\boldsymbol{r}^+$
$\boldsymbol{r} + \frac{1}{2}\boldsymbol{R}\mathsf{w}$	$\boldsymbol{r}^-$
$M^-\left[\mathsf{m} - \mathsf{w}\times\mathsf{n}^{\mathcal{P}}\right] - P^{\frac{1}{2}}\mathsf{m}^{\mathcal{P}}$	m^+
$M^+\left[\mathsf{m} - \mathsf{w}\times\mathsf{n}^{\mathcal{P}}\right] + P^{\frac{T}{2}}\mathsf{m}^{\mathcal{P}}$	m^-
$P^{\frac{1}{2}}\left[\frac{1}{2}\mathsf{n} - \mathsf{n}^{\mathcal{P}}\right]$	n^+
$P^{\frac{T}{2}}\left[\frac{1}{2}\mathsf{n} + \mathsf{n}^{\mathcal{P}}\right]$	n^-

3.2.9 Formulation of the birod equilibrium conditions through Lagrange Multipliers

Similarly to the bichain case, the birod equilibrium conditions (3.2.54) can alternatively be described through the *Lagrange multiplier method*. For instance, assuming the Dirichlet conditions described in (3.2.50a) of the form

$$\boldsymbol{g}(0) = \mathrm{Id}_4,\ \boldsymbol{g}(L) = {\boldsymbol{g}_L}^* \tag{3.2.79a}$$

$$\mathcal{P}(0) = {\mathcal{P}_0}^* \text{ and } \mathcal{P}(L) = {\mathcal{P}_L}^* \tag{3.2.79b}$$

we define the *Lagrangian function* $\mathcal{L} : \left(\mathrm{SE}(3)^{]0,L[}\right)^2 \times \mathbb{R}^{18} \to \mathbb{R}$ as

$$\mathcal{L}[\boldsymbol{g}, \mathcal{P}, \boldsymbol{\lambda}] = E[\boldsymbol{g}, \mathcal{P}] - \boldsymbol{\lambda}\Phi\left(\mathcal{P}(0), \boldsymbol{g}(L), \mathcal{P}(L); {\mathcal{P}_0}^*, {\boldsymbol{g}_L}^*, {\mathcal{P}_L}^*\right) \tag{3.2.80}$$

where the function Φ is defined as

$$\Phi\left(\mathcal{P}(0), \boldsymbol{g}(L), \mathcal{P}(L); {\mathcal{P}_0}^*, {\boldsymbol{g}_L}^*, {\mathcal{P}_L}^*\right) = \begin{pmatrix} \frac{1}{2}\mathrm{Vect}\left(\boldsymbol{R}(L)\left({\boldsymbol{R}_L}^*\right)^T\right) \\ \boldsymbol{r}(L) - \boldsymbol{R}(L)\left({\boldsymbol{R}_L}^*\right)^T {\boldsymbol{r}_L}^* \\ \frac{1}{2}\mathrm{Vect}\left(P(0)\left({P_0}^*\right)^T\right) \\ \frac{1}{2}\mathrm{Vect}\left(P(L)\left({P_L}^*\right)^T\right) \\ \mathsf{w}(0) - {\mathsf{w}_0}^* \\ \mathsf{w}(L) - {\mathsf{w}_L}^* \end{pmatrix} \tag{3.2.81}$$

and where $E[\boldsymbol{g}, \mathcal{P}]$ denotes the internal bichain energy, defined in (3.2.42). The set

$$\boldsymbol{\lambda} = (\boldsymbol{\lambda}^{\boldsymbol{R}}, \boldsymbol{\lambda}^{\boldsymbol{r}}, \lambda_0^P, \lambda_L^P, \lambda_0^{\mathsf{w}}, \lambda_L^{\mathsf{w}}) \in \mathbb{R}^{18} \tag{3.2.82}$$

is called the set of *birod Lagrange multipliers*. Note that the expression in (3.2.80) is strongly inspired from the analogous one for bichains in (2.2.55).

The stationary conditions for the functional $\mathcal{L}$ over the set of birod configuration $(\boldsymbol{g}, \mathcal{P})$ and over the Lagrange multipliers $\boldsymbol{\lambda}$ can be shown to be the birod equilibrium conditions in (3.2.45) with the extra boundary conditions

$$\begin{pmatrix} \mathsf{m}(L) + \boldsymbol{r}(L) \times \mathsf{n}(L) \\ \mathsf{n}(L) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\lambda}^{\boldsymbol{R}} \\ \boldsymbol{\lambda}^{\boldsymbol{r}} \end{pmatrix} \tag{3.2.83a}$$

$$\begin{pmatrix} \mathsf{m}^{\mathcal{P}}(0) \\ \mathsf{n}^{\mathcal{P}}(0) \end{pmatrix} = \begin{pmatrix} \lambda_0^P \\ \lambda_0^{\mathsf{w}} \end{pmatrix} \text{ and } \begin{pmatrix} \mathsf{m}^{\mathcal{P}}(L) \\ \mathsf{n}^{\mathcal{P}}(L) \end{pmatrix} = \begin{pmatrix} \lambda_L^P \\ \lambda_L^{\mathsf{w}} \end{pmatrix} \tag{3.2.83b}$$

together with the requirement that the array of constraints in (3.2.81) vanishes.

Consequently, the Lagrange multipliers $\boldsymbol{\lambda} = (\boldsymbol{\lambda}^{\boldsymbol{R}}, \boldsymbol{\lambda}^{\boldsymbol{r}}, \lambda_1^P, \lambda_N^P, \lambda_1^{\mathsf{w}}, \lambda_N^{\mathsf{w}})$ can be interpreted as the set of couple and force which are needed to enforce the constraints in (3.2.79) and moreover the triples $\boldsymbol{\lambda}^{\boldsymbol{R}}$ and $\boldsymbol{\lambda}^{\boldsymbol{r}}$ exactly correspond, in the case of internal birod energy, to the conserved quantities described in (3.2.52).

3.2.10 On birod models for single rods

We conclude this part on the variational principle for rods with some remarks on the relations between the rod model and and the birod model.

Similarly to rigid body bichains, the birod configuration is made up of a rod configuration to describe their macrostructure and the birod equilibrium conditions contain the rod equilibrium conditions to express the macrostructure balance laws, either in the Euler-Poincaré form (3.2.54) or in the Hamiltonian form (3.2.69). However, the mechanical behaviour of a birod macrostructure is in principle very different from the one of a single rod whose equilibrium conditions are governed by the Hamiltonian system (3.2.25) due to the fact that the respective constitutive relations can be very different.

A particular case of the birod internal energy (3.2.42) is when it decouples into separate contributions from the macrostructure and the microstructure deformations,

i.e. when E is of the form

$$E[\boldsymbol{g}, \mathcal{P}] = \int_0^L \mathsf{W}^{\xi}(\xi(s); s) + \mathsf{W}^{\mathcal{P}}(\mathcal{P}(s), \xi^{\mathcal{P}}(s); s) ds. \tag{3.2.84}$$

In such cases, the energy E is said to be a *separable internal birod energy*. The birod Hamiltonian function is then of the form

$$\mathsf{H}(\mathsf{z}_{\boldsymbol{g}}(s), \mathsf{z}_{\mathcal{P}}(s); s) = \mathsf{H}^{\mathsf{z}_{\boldsymbol{g}}}(\mathsf{z}_{\boldsymbol{g}}(s); s) + \mathsf{H}^{\mathsf{z}_{\mathcal{P}}}(\mathsf{z}_{\mathcal{P}}(s); s) \tag{3.2.85}$$

according to the Legendre transform definition in (3.2.66). Consequently, equilibrium equations for the macrostructure in (3.2.69) are independent from those for the microstructure, and then both are only, and exactly, governed by their respective Hamiltonian function denoted by $\mathsf{H}^{\mathsf{z}_{\boldsymbol{g}}}$ and $\mathsf{H}^{\mathsf{z}_{\mathcal{P}}}$ in (3.2.85).

Conversely, one can study the case of *constrained microstructure* which corresponds to prescribing the microstructure $\mathcal{P}$ to a given (possibly stressed) configuration. The rigorous analysis of this problem is left as an open topic, but we believe that the induced rod mechanical behaviour of the macrostructure will not be described by a single rod Hamiltonian of the form (3.2.25) since a non-local character arises from this constraint. Similarly to the rigid body bichain, we conclude that the birod model can be seen as an intermediate model between local rods and non-local rods.

3.3 On Stationary Configurational Distributions and Rod Moments

This section presents the definition of stationary configurational distributions for continuum rods and birods in a stochastic environment. In particular, these infinite dimensional distributions are understood as the appropriate limit of finite dimensional chains and bichain distributions. We then discuss the notion of moments of rod and birod observables and emphasize the interest of the rod configurational first moment.

3.3.1 Stationary configurational distributions for isolated rods and birods

Similarly to rigid body chains, we say that a rod is said to be an *isolated continuum rod in a stochastic bath* if the only external couples and forces which act on it are due to the interactions with the surrounding fluid. This definition is appearing for instance in previous studies in continuum rod statistical mechanics and it is often assumed, moreover, that in such conditions the continuum rod admits a *stationary rod configurational distribution* $d\rho[\boldsymbol{g}]$ [Landau and Lifshitz, 1959, Bugl, 1969, Yamakawa, 1976, Doi and Edwards, 1986, Marko and Siggia, 1994, Manning et al., 1996, Yamakawa, 1997, Becker, 2007, Cotta-Ramusino, 2008]. If, moreover, the ends of the rod are free we define the rod configurational distribution to be expressed as a function of the rod generator vector ξ as the *path integral density*

$$d\rho\left[\boldsymbol{g}[\xi]\right] \sim \exp\left\{-\frac{1}{k_B T}E[\boldsymbol{g}]\right\} D[\xi] \text{ for } \boldsymbol{g} \in \Gamma \tag{3.3.1}$$

where E denotes the internal energy of the rod, defined in (3.2.5), k_B the *Boltzmann constant*, T the *kinetic temperature of the bath* and Γ the set of continuum rod configurations with free ends, similarly to the rigid body chain (2.3.1). The distribution $d\rho\left[\boldsymbol{g}[\xi]\right]$ is called the *Boltzmann distribution for a Cosserat rod*.

We should stress that the rod distribution $d\rho[\boldsymbol{g}]$ is a probability distribution on a space of rod configurations and, in contrast to rigid body chain, is therefore a distribution on an infinite dimensional space. It can then be understood through the following construction. For a rod configuration $\boldsymbol{g}(s)$ with $s \in]0, L[$, introduce a mesh containing d nodes

$$0 < {s_1}^{(d)} < {s_2}^{(d)} < ... < {s_d}^{(d)} < L \tag{3.3.2}$$

and suppose that we can observe the rigid body chain configuration defined by the *discretised rod configuration* $\boldsymbol{g}^{(d)} = ({\boldsymbol{g}_1}^{(d)}, ..., {\boldsymbol{g}_d}^{(d)})$ where ${\boldsymbol{g}_i}^{(d)} \in \mathrm{SE}(3)$ is given by

$${\boldsymbol{g}_i}^{(d)} = \boldsymbol{g}({s_i}^{(d)}) \text{ for } i = 1, ..., d. \tag{3.3.3}$$

The *discretised junction displacement* between two consecutive elements in the discretised rod is denoted by ${a_i}^{(d)} \in \mathrm{SE}(3)$ and is defined similarly to the rigid body chain case (2.1.4) as

$${a_i}^{(d)} = ({\boldsymbol{g}_i}^{(d)})^{-1} {\boldsymbol{g}_{i+1}}^{(d)} \tag{3.3.4}$$

which leads to the following set of internal chain coordinates of the discretised rod

$$x^{(d)} = ({x_1}^{(d)}, ..., {x_{d-1}}^{(d)}) \tag{3.3.5}$$

where

$${a_i}^{(d)} = a({x_i}^{(d)}) \text{ for } i = 1, ..., d \tag{3.3.6}$$

similarly to (2.1.11).

Given the observed discretised rod configuration (3.3.23) we introduce a smooth *rod interpolation* of the chain configuration $\boldsymbol{g}^{(d)}$ as

$$\boldsymbol{g}^{(d)}(s) = \boldsymbol{g}^{(d)}(x^{(d)}; s) \text{ for } s \in]0, L[\tag{3.3.7}$$

which satisfies

$$\boldsymbol{g}^{(d)}(x^{(d)}; {s_i}^{(d)}) = {\boldsymbol{g}_i}^{(d)} \tag{3.3.8}$$

for all nodes s_i in (3.3.2). This allows the definition, for a given mesh in (3.3.2) and for a given interpolation rule in (3.3.7), of the *discretised space of rod configurations* $\Gamma^{(d)}$ as the set of interpolations of discretised rod configurations $\boldsymbol{g}^{(d)}$, i.e. as

$$\Gamma^{(d)} = \left\{ \boldsymbol{g}^{(d)}(s) \,|\, \boldsymbol{g} \in \Gamma \right\} \tag{3.3.9}$$

where Γ denotes the set of rod configurations with free ends in (3.3.1). Similarly to (3.1.4), we define the *rod interpolation generator vector* $\xi^{(d)}(s)$ by the relation

$$\frac{d}{ds}\boldsymbol{g}^{(d)}(s) = \boldsymbol{g}^{(d)}(s)\mathcal{T}\xi^{(d)}(s) \tag{3.3.10}$$

for all $s \in]0, L[$. Formally, one then has a map

$$\xi^{(d)}(s) = \xi^{(d)}(x^{(d)}; s) \text{ for } s \in]0, L[\tag{3.3.11}$$

analogously to (3.3.7). The definition (3.3.10) together with (3.3.11) imply that both the interpolation generator vector $\xi^{(d)}(s)$ and the interpolation of the discretised rod configuration $\boldsymbol{g}^{(d)}(s)$ converge asymptotically to the continuum rod generator vector $\xi(s)$ and to the configuration $\boldsymbol{g}(s)$ when the number of nodes d in the interval $]0, L[$ increases, i.e.

$$\lim_{d\to\infty} \xi^{(d)}(s) = \xi(s) \tag{3.3.12a}$$

$$\lim_{d\to\infty} \boldsymbol{g}^{(d)}(s) = \boldsymbol{g}(s). \tag{3.3.12b}$$

for smooth rod configurations $\boldsymbol{g}(s)$. In other terms we have formally that

$$\lim_{d\to\infty} \Gamma^{(d)} = \Gamma. \tag{3.3.13}$$

Moreover, using the definition of ${a_i}^{(d)}$ in (3.3.4) and $\xi^{(d)}(s)$ in (3.3.10), one can show that for ${s_i}^{(d)} < s < {s_{i+1}}^{(d)}$

$$\lim_{d\to\infty} \frac{1}{{h_i}^{(d)}} \left({a_i}^{(d)} - \mathrm{Id}_4 \right) = \mathcal{T}\xi(s) \tag{3.3.14}$$

where we have defined

$${h_i}^{(d)} = {s_{i+1}}^{(d)} - {s_i}^{(d)}. \tag{3.3.15}$$

Consequently $\boldsymbol{g}^{(d)}(s)$ can be thought of as a rigid body chain approximation of the original rod configuration $\boldsymbol{g}(s)$. Observe that this limiting process does not depend on the mesh nodes s_i chosen in (3.3.2) and that the construction (3.3.7) has strong similarity to the *geodesic finite element method* for rods proposed in [Sander, 2010] in the context of classical mechanics.

The configurational distribution for a Cosserat rod $d\rho[\boldsymbol{g}]$ in (3.3.1) is then defined using the strong similarity to rigid body chain stochastic calculus. Let $\mathcal{A}[\boldsymbol{g}] \in \mathbb{R}$ be a smooth functional on the continuum rod configurations set Γ. This functional can be thought of as the end-to-end vector of the rod for instance. For clarity, we assume that the internal chain coordinates $x^{(d)} = ({x_1}^{(d)}, ..., {x_{d-1}}^{(d)}) \in \mathbb{R}^{6(d-1)}$ in (3.3.5) are chosen according to the definitions (2.1.18). The distribution $d\rho\,[\boldsymbol{g}[\xi]]$ is then defined as the limit

$$\int_\Gamma \mathcal{A}[\boldsymbol{g}] d\rho\,[\boldsymbol{g}[\xi]] = \lim_{d\to\infty} \int_{\Gamma^{(d)}} \mathcal{A}[\boldsymbol{g}^{(d)}] d\rho^{(d)}(\boldsymbol{g}^{(d)})(x^{(d)}) \tag{3.3.16a}$$

for any smooth functional $\mathcal{A}[\boldsymbol{g}] \in \mathbb{R}$ where

$$d\rho^{(d)}(\boldsymbol{g}^{(d)})(x^{(d)}) = \frac{1}{Z^{(d)}} \exp\left\{ -\frac{1}{k_B T} E[\boldsymbol{g}^{(d)}] \right\} J(x^{(d)}) dx^{(d)} \tag{3.3.16b}$$

with

$$Z^{(d)} = \int_{\Gamma^{(d)}} \exp\left\{ -\frac{1}{k_B T} E[\boldsymbol{g}^{(d)}] \right\} J(x^{(d)}) dx^{(d)}. \tag{3.3.16c}$$

and where we have used with the short notation $dx^{(d)}$ for $d{x_1}^{(d)}...d{x_{d-1}}^{(d)}$, $J(x^{(d)})$ is the explicit Jacobian factor in (2.3.2c) and in (2.3.3) related to the definition of local coordinates $x^{(d)}$ in (3.3.5), the finite dimensional set $\Gamma^{(d)}$ is defined in (3.3.9) and $Z^{(d)}$ is the normalising constant for each fixed d. Note that in (3.3.16) the interpolation $\boldsymbol{g}^{(d)}(s)$ is seen as an explicit function of the local chain coordinates through the relation (3.3.7). The distribution $d\rho^{(d)}(\boldsymbol{g}^{(d)})(x^{(d)})$ is called a *finite dimensional rod marginal distribution* since it corresponds to the observation of only a finite number of rod cross section configurations ${\boldsymbol{g}_i}^{(d)}$. In other words, the rod configurational distribution $d\rho\,[\boldsymbol{g}[\xi]]$ is defined as the appropriate limit of the rigid body chain configurational distribution $d\rho^{(d)}(\boldsymbol{g}^{(d)})(x^{(d)})$ in (3.3.16). This construction can be referred to as *time slicing*, by analogy to the literature on imaginary path integrals.

The assumptions of the functional form of the finite dimensional distribution $d\rho^{(d)}(\boldsymbol{g}^{(d)})(x^{(d)})$ in (3.3.16) have to be discussed. A complete argument would require a rigorous mathematical proof, however we have decided to give here only some informal justification. As presented in (2.3.1), this form of the stationary distribution is correct for isolated rigid body chains when the fluid interactions are modelled in the framework of *Langevin stochastic dynamics*, as discussed in detail in [Walter et al., 2010]. Assuming the analogous stochastic dynamics for a continuum rod, we claim that any discretised rod configuration $\boldsymbol{g}^{(d)}$ satisfies a stationary distribution of the form that is precisely like the chain configurational distribution but with the discretised energy. Consequently, the presence of the Jacobian factor ensures that the finite dimensional rod marginal distribution $d\rho^{(d)}(\boldsymbol{g}^{(d)})(x^{(d)})$ transforms consistently in any internal discretised chain coordinates $x^{(d)}$, and this implies that the limiting distribution $d\rho\,[\boldsymbol{g}[\xi]]$ is independent of the internal coordinates chosen to describe the discretised configurations.

We observe, moreover, that we can always choose a chain interpolation $\boldsymbol{g}^{(d)}(s)$ in (3.3.7) such that the associated generator vector $\xi^{(d)}(s)$ in (3.3.11) is only a function of the local chain coordinates ${x_i}^{(d)}$ in the interval ${s_i}^{(d)} < s < {s_{i+1}}^{(d)}$ for all $i = 1, ..., d-1$, i.e.

$$\xi^{(d)}(s) = \xi^{(d)}({x_i}^{(d)}; s) \text{ for } {s_i}^{(d)} < s < {s_{i+1}}^{(d)}. \tag{3.3.17}$$

A rod interpolation $\boldsymbol{g}^{(d)}(s)$ whose generator vector satisfies (3.3.17) is said to be a *local rod interpolation*. This specific choice, together with the form of the Cosserat rod energy E in (3.2.5) and the Jacobian definition in (2.3.2c), allows the factorization of the distribution $d\rho^{(d)}(\boldsymbol{g}^{(d)})(x^{(d)})$ as

$$d\rho^{(d)}(\boldsymbol{g}^{(d)})(x^{(d)}) = \prod_{i=1}^{d-1} d{\rho_i}^{(d)}(a({x_i}^{(d)})) \tag{3.3.18a}$$

and

$$d{\rho_i}^{(d)}(a({x_i}^{(d)})) = \frac{1}{{Z_i}^{(d)}} \exp\left\{-\frac{1}{k_B T} {w_i}^{(d)}(a({x_i}^{(d)}))\right\} J_i({x_i}^{(d)}) d{x_i}^{(d)} \tag{3.3.18b}$$

where

$${w_i}^{(d)}(a({x_i}^{(d)})) = \int_{s_i}^{s+1} \mathrm{W}(\xi({x_i}^{(d)}; s); s) ds \tag{3.3.18c}$$

and where ${Z_i}^{(d)}$ denotes the local normalization constant. Expressions (3.3.18) make even stronger the relation with the stochastic calculus for local rigid body chains described in (2.3.1).

An interesting fact however, which contrasts with the calculus for rigid body chains, is that the Jacobian factor can be neglected for continuum rod configurational distributions. This fact has been noticed in [Becker, 2007] when using exponential coordinates to described discretised chain configuration. In terms of the internal coordinates $x^{(d)}$

used in the definition (3.3.16), we observe that, according to the limit in (3.3.14), for all $1 \leq i \leq d$

$$x_i{}^{(d)} = \mathrm{O}\left(h_i{}^{(d)}\right) \tag{3.3.19}$$

where $h_i{}^{(d)} = s_{i+1}{}^{(d)} - s_i{}^{(d)}$ which implies that

$$\lim_{d\to\infty} \sum_{i=1}^{d-1} \ln\left\{J_i(x_i{}^{(d)})\right\} = 0 \tag{3.3.20a}$$

since

$$J_i(x_i{}^{(d)}) = 1 + \mathrm{O}\left(h_i{}^{(d)2}\right) \tag{3.3.20b}$$

according to the definition of the Jacobian factor J_i in (2.3.3). Consequently, one can replace the definition of the finite dimensional rod marginal in (3.3.16a) of the local normalising constant (3.3.16c) by

$$d\rho^{(d)}(\boldsymbol{g}^{(d)})(x^{(d)}) = \frac{1}{Z^{(d)}} \exp\left\{-\frac{1}{k_B T} E[\boldsymbol{g}^{(d)}]\right\} dx^{(d)} \tag{3.3.21a}$$

and

$$Z^{(d)} = \int_{\Gamma^{(d)}} \exp\left\{-\frac{1}{k_B T} E[\boldsymbol{g}^{(d)}]\right\} dx^{(d)}. \tag{3.3.21b}$$

without changing the limiting behaviour, or in other words the Jacobian factor correction vanishes in the limit $d \to \infty$.

We define the notion of an *isolated continuum birod in a stochastic bath* in a similar way to the single rod case and so assume the existence of a *stationary birod configurational distribution* $d\rho[\boldsymbol{g}, \mathcal{P}]$. In the case where both ends of the birod configuration are free, the birod distribution is defined in terms of the macrostructure generator ξ and of the microstructure configuration $\mathcal{P}$

$$d\rho\,[\boldsymbol{g}[\xi], \mathcal{P}] \sim \exp\left\{-\frac{1}{k_B T} E[\boldsymbol{g}, \mathcal{P}]\right\} D[\xi, \mathcal{P}] \text{ for } (\boldsymbol{g}, \mathcal{P}) \in \Gamma \tag{3.3.22}$$

in analogy to the configurational rod distribution in (3.3.1), but where E now denotes the continuum birod internal energy and where Γ denotes the set of continuum birod configurations with free ends. The distribution $d\rho\,[\boldsymbol{g}[\xi], \mathcal{P}]$ is then also understood as a limit of finite dimensional marginal distributions.

For a given birod configuration $(\boldsymbol{g}, \mathcal{P})$ and a given mesh of the form (3.3.2), we define the *discretised birod configuration* $(\boldsymbol{g}^{(d)}, \mathcal{P}^{(d)}) = (\boldsymbol{g}_1{}^{(d)}, \mathcal{P}_1{}^{(d)}, ..., \boldsymbol{g}_d{}^{(d)}, \mathcal{P}_d{}^{(d)})$ where

$$\boldsymbol{g}_i{}^{(d)} = \boldsymbol{g}(s_i{}^{(d)}) \text{ and } \mathcal{P}_i{}^{(d)} = \mathcal{P}(s_i{}^{(d)}) \text{ for } i = 1, ..., d \tag{3.3.23}$$

and where $(\boldsymbol{g}^{(d)}, \mathcal{P}^{(d)})$ can be seen as a bichain configuration of the form (2.1.30). Consequently, the local rigid body displacements ${a_i}^{(d)}, {b_i}^{(d)} \in \mathrm{SE}(3)$ in the discretised birod configuration are defined as

$$ {a_i}^{(d)} = ({\boldsymbol{g}_i}^{(d)})^{-1} {\boldsymbol{g}_{i+1}}^{(d)} \text{ and } {b_i}^{(d)} = ({\boldsymbol{g}_i^+}^{(d)})^{-1} {\boldsymbol{g}_i^-}^{(d)} \tag{3.3.24} $$

and this leads to the set of internal bichain coordinates

$$ (x^{(d)}, y^{(d)}) = ({y_1}^{(d)}, {x_1}^{(d)}, ..., {x_{d-1}}^{(d)}, {y_d}^{(d)}) \tag{3.3.25} $$

where

$$ {a_i}^{(d)} = a({x_i}^{(d)}) \text{ and } {b_i}^{(d)} = b({y_i}^{(d)}) \tag{3.3.26} $$

similarly to (2.1.32).

Similarly to the single rod case, we introduce a smooth *birod interpolation* of the bichain configuration $(\boldsymbol{g}^{(d)}, \mathcal{P}^{(d)})$ as

$$ \boldsymbol{g}^{(d)}(s) = \boldsymbol{g}^{(d)}(x^{(d)}; s) \text{ and } \mathcal{P}^{(d)}(s) = \mathcal{P}^{(d)}(y^{(d)}; s) \text{ for } s \in]0, L[\tag{3.3.27} $$

which satisfy

$$ \boldsymbol{g}^{(d)}(x^{(d)}; {s_i}^{(d)}) = {\boldsymbol{g}_i}^{(d)} \text{ and } \mathcal{P}^{(d)}(y^{(d)}; {s_i}^{(d)}) = {\mathcal{P}_i}^{(d)} \tag{3.3.28} $$

for all nodes s_i in (3.3.2), and then the *discretised space of birod configurations* $\Gamma^{(d)}$ as the set of interpolations of discretised birod configurations $\boldsymbol{g}^{(d)}$ for a given mesh in (3.3.2) and for any given interpolation rule of the form (3.3.27), i.e. as

$$ \Gamma^{(d)} = \left\{ (\boldsymbol{g}^{(d)}(s), \mathcal{P}^{(d)}(s)) \,|\, (\boldsymbol{g}, \mathcal{P}) \in \Gamma \right\} \tag{3.3.29} $$

where Γ denotes the set of birod configuration with free ends in (3.3.22).

The *birod interpolation generator vectors* of the macrostructure $\xi^{(d)}(s)$ and the microstructure $\xi^{\mathcal{P}^{(d)}}(s)$ are then defined by the relations

$$ \frac{d}{ds}\boldsymbol{g}^{(d)}(s) = \boldsymbol{g}^{(d)}(s)\mathcal{T}\xi^{(d)}(s) \text{ and } \frac{d}{ds}\mathcal{P}^{(d)}(s) = \mathcal{T}^{\mathcal{P}}\xi^{\mathcal{P}^{(d)}}(s) \tag{3.3.30} $$

for all $s \in]0, L[$ according to their definition in (3.1.23) and (3.1.27) respectively and one then has maps of the form

$$ \xi^{(d)}(s) = \xi^{(d)}(x^{(d)}; s) \text{ and } \xi^{\mathcal{P}^{(d)}}(s) = \xi^{\mathcal{P}^{(d)}}(y^{(d)}; s) \text{ for } s \in]0, L[\tag{3.3.31} $$

analogously to (3.3.27).

The definitions (3.3.30) together with (3.3.31) imply that the birod interpolation generator vectors $\xi^{(d)}(s)$ and $\xi^{\mathcal{P}^{(d)}}(s)$, and the interpolation of the discretised birod configuration $(\boldsymbol{g}^{(d)}(s), \mathcal{P}^{(d)}(s))$ converge asymptotically to the continuum generator vectors

$\xi(s)$ and $\xi^{\mathcal{P}}(s)$ and to the birod configuration $(\boldsymbol{g}(s), \mathcal{P}(s))$ when the number of nodes d in the interval $]0, L[$ increases for smooth birod configuration $(\boldsymbol{g}(s), \mathcal{P}(s))$.

The configurational distribution for a birod $d\rho[\boldsymbol{g}, \mathcal{P}]$ in (3.3.22) is then defined similarly to the single rod distribution. Let $\mathcal{A}[\boldsymbol{g}, \mathcal{P}] \in \mathbb{R}$ be a smooth functional on the continuum birod configuration set Γ, and define the internal coordinates of the discretised birod configuration $(x^{(d)}, y^{(d)})$ to be the bichain coordinates defined in (2.1.18) and (2.1.33) respectively. The distribution $d\rho\,[\boldsymbol{g}[\xi], \mathcal{P}]$ is then defined as the limit

$$\int_{\Gamma} \mathcal{A}[\boldsymbol{g}, \mathcal{P}] d\rho\,[\boldsymbol{g}[\xi], \mathcal{P}] = \lim_{d\to\infty} \int_{\Gamma^{(d)}} \mathcal{A}[\boldsymbol{g}^{(d)}, \mathcal{P}^{(d)}] d\rho^{(d)}(\boldsymbol{g}^{(d)}, \mathcal{P}^{(d)})(x^{(d)}, y^{(d)}) \tag{3.3.32a}$$

for any smooth functional $\mathcal{A}[\boldsymbol{g}, \mathcal{P}] \in \mathbb{R}$ where

$$d\rho^{(d)}(\boldsymbol{g}^{(d)}, \mathcal{P}^{(d)})(x^{(d)}, y^{(d)}) = \frac{1}{Z^{(d)}} \exp\left\{-\frac{1}{k_B T} E[\boldsymbol{g}^{(d)}, \mathcal{P}^{(d)}]\right\} J(x^{(d)}, y^{(d)}) dx^{(d)} dy^{(d)} \tag{3.3.32b}$$

with

$$Z^{(d)} = \int_{\Gamma^{(d)}} \exp\left\{-\frac{1}{k_B T} E[\boldsymbol{g}^{(d)}, \mathcal{P}^{(d)}]\right\} J(x^{(d)}, y^{(d)}) dx^{(d)} dy^{(d)}. \tag{3.3.32c}$$

and where we have used with the short notation $dx^{(d)}dy^{(d)}$ for ${dy_1}^{(d)}{dx_1}^{(d)}...{dx_{d-1}}^{(d)}{dy_d}^{(d)}$, $J(x^{(d)}, y^{(d)})$ is the explicit Jacobian factor in (2.3.6c), (2.3.3) and (2.3.8) related to the definition of local coordinates $(x^{(d)}, y^{(d)})$ in (3.3.25), the finite dimensional set $\Gamma^{(d)}$ is defined in (3.3.29) and $Z^{(d)}$ is the normalising constant for each fixed d. The distribution $d\rho^{(d)}(\boldsymbol{g}^{(d)}, \mathcal{P}^{(d)})(x^{(d)}, y^{(d)})$ is called a *finite dimensional birod marginal distribution* and is deduced similarly to the single rod case.

Conversely, we should stress the asymmetry between the roles played by the internal coordinates $x^{(d)}$ for the discretised macrostructure and by the internal coordinates $y^{(d)}$ for the discretised microstructure. In particular one has

$$\lim_{d\to\infty} {x_i}^{(d)} = 0 \text{ whereas } \lim_{d\to\infty} {y_i}^{(d)} = y(s) \text{ with } \mathcal{P}(s) = \mathcal{P}(y(s)) \tag{3.3.33}$$

which implies that the contribution of the Jacobian factor associated with the macrostructure coordinates $x^{(d)}$ can be neglected, similarly to the single rod case in (3.3.21), but not the one coming from the microstructure coordinates $y^{(d)}$.

As discussed in section 3.2.7, and more precisely in (3.2.57), a realistic birod internal energy will be given in terms of the microstructure internal coordinates. The birod configurational distribution in terms of the internal coordinates $y(s)$ is denoted by $d\rho\,[\boldsymbol{g}[\xi], \mathcal{P}(y)]$ and is defined by

$$d\rho\,[\boldsymbol{g}[\xi], \mathcal{P}(y)] \sim \exp\left\{-\frac{1}{k_B T} E[\boldsymbol{g}, \mathcal{P}(y)]\right\} D[\xi, y] \text{ for } (\boldsymbol{g}, \mathcal{P}(y)) \in \Gamma \tag{3.3.34}$$

according to the birod configurational distribution in (3.3.22).

Using the arguments presented in section 3.2.10, one can see the single rod configurational distribution $d\rho[\boldsymbol{g}]$ in (3.3.1) as a marginal of the birod configurational distribution $d\rho[\boldsymbol{g}, \mathcal{P}]$ in (3.3.22). Even for a local birod internal energy of the form (3.2.42), the marginal distribution for the macrostructure is in general very different from the one for local single rod presented in (3.3.1), but, in contrast, this marginal distribution exactly factorizes into (3.3.1) when the internal birod energy decouples into the macrostructure and the microstructure configurations.

3.3.2 Observables and expectations of rods and birods

A *continuum rod observable* $\mathcal{A}$ is formally a functional on a set of rod configurations and it is denoted by $\mathcal{A}[\boldsymbol{g}] \in \mathbb{R}$. In contrast to the rigid body chain case, we use here square brackets $[\,.\,]$ to emphasize that the evaluation of $\mathcal{A}[\boldsymbol{g}]$ can involve the evaluation of derivatives of the configuration $\boldsymbol{g}(s)$ for instance. The most common observables $\mathcal{A}[\boldsymbol{g}]$ in polymer physics, include the projection of some given direction at arc-length $s \in]0, L[$ in the rod on some other direction at $s = 0$, or of the end-to-end vector for instance [Krakty and Porod, 1949, Landau and Lifshitz, 1959, Doi and Edwards, 1986, Yamakawa, 1997, Panyukov and Rabin, 2000].

The *expectation of a rod observable* $\mathcal{A}[\boldsymbol{g}] \in \mathbb{R}$ with respect to the continuum rod configurational distribution $d\rho[\boldsymbol{g}]$, as presented in (3.3.1), is defined as

$$\langle \mathcal{A}[\boldsymbol{g}] \rangle = \int_{\Gamma} \mathcal{A}[\boldsymbol{g}] d\rho\,[\boldsymbol{g}[\xi]] \tag{3.3.35}$$

according to the expression in (3.3.16a). Practically, we introduce a mesh on the interval $]0, L[$, as described in (3.3.2), and a set of interpolated rod configuration $\Gamma^{(d)}$ presented in (3.3.9) to define the *finite dimensional observable approximation* $\mathcal{A}[\boldsymbol{g}^{(d)}]$ of the observable $\mathcal{A}[\boldsymbol{g}]$ which has the property that

$$\lim_{d\to\infty} \mathcal{A}[\boldsymbol{g}^{(d)}] = \mathcal{A}[\boldsymbol{g}] \tag{3.3.36}$$

according to (3.3.12). Then, the *finite dimensional expectation* $\langle \mathcal{A}[\boldsymbol{g}] \rangle^{(d)}$ of the finite dimensional approximation of the observable $\mathcal{A}[\boldsymbol{g}^{(d)}]$ is given by

$$\langle \mathcal{A}[\boldsymbol{g}] \rangle^{(d)} = \int_{\Gamma^{(d)}} \mathcal{A}[\boldsymbol{g}^{(d)}] d\rho^{(d)}(\boldsymbol{g}^{(d)})(x^{(d)}) \tag{3.3.37}$$

where $d\rho^{(d)}(\boldsymbol{g}^{(d)})(x^{(d)})$ denotes the finite dimensional rod marginal distribution defined in (3.3.16b). Using the definition in (3.3.16a), the expectation in (3.3.35) is then understood as the limit $d \to \infty$ of the finite dimensional expectation, i.e.

$$\langle \mathcal{A}[\boldsymbol{g}] \rangle = \lim_{d\to\infty} \langle \mathcal{A}[\boldsymbol{g}] \rangle^{(d)}. \tag{3.3.38}$$

As discussed in section 3.3.1, the limiting value in (3.3.38), when it exists, is independent of the discretization and of the internal discretised rod coordinates $x^{(d)}$ chosen to

evaluate the finite dimensional expectation $\langle \mathcal{A}[\boldsymbol{g}] \rangle^{(d)}$. The symbol $\langle\, . \,\rangle$ stands therefore for the expectation over an ensemble of continuum rod configurations and is a function of only the rod internal energy E. A related discussion can be found in [Becker, 2007].

In the case where the observable $\mathcal{A}[\boldsymbol{g}]$ is a function of the arc-length s, we define its expectation $\langle \mathcal{A}[\boldsymbol{g}] \rangle$ to be the function such that

$$\int_0^L \langle \mathcal{A}[\boldsymbol{g}] \rangle \cdot b(s) ds = \left\langle \int_0^L \mathcal{A}[\boldsymbol{g}] \cdot b(s) ds \right\rangle \tag{3.3.39}$$

for any non-stochastic and smooth function $b(s)$ and where $\langle\, . \,\rangle$ is in the sense of (3.3.35), or equivalently (3.3.38).

A *continuum birod observable* is also denoted by $\mathcal{A}$ but is a functional of birod configurations $\mathcal{A}[\boldsymbol{g}, \mathcal{P}] \in \mathbb{R}$. Similarly to the single rod case, we define the *expectation of a birod observable* as

$$\langle \mathcal{A}[\boldsymbol{g}, \mathcal{P}] \rangle = \int_\Gamma \mathcal{A}[\boldsymbol{g}, \mathcal{P}] d\rho[\boldsymbol{g}, \mathcal{P}] \tag{3.3.40}$$

according to the definition of the continuum birod configurational distribution $d\rho[\boldsymbol{g}, \mathcal{P}]$ in (3.3.22). Defining the *finite dimensional expectation* of the birod observable $\langle \mathcal{A}[\boldsymbol{g}, \mathcal{P}] \rangle^{(d)}$ in a similar way to the single rod case (3.3.37), we also understand (3.3.40) as

$$\langle \mathcal{A}[\boldsymbol{g}, \mathcal{P}] \rangle = \lim_{d \to \infty} \langle \mathcal{A}[\boldsymbol{g}, \mathcal{P}] \rangle^{(d)} \tag{3.3.41}$$

and conclude that the value of the birod expectation $\langle \mathcal{A}[\boldsymbol{g}, \mathcal{P}] \rangle$, when it exists, is independent of the choices made to describe the discretised birod configurations. Similarly to rods, if the observable $\mathcal{A}[\boldsymbol{g}, \mathcal{P}]$ is a function of the parameter s, its expectation is defined by

$$\int_0^L \langle \mathcal{A}[\boldsymbol{g}, \mathcal{P}] \rangle \cdot b(s) ds = \left\langle \int_0^L \mathcal{A}[\boldsymbol{g}, \mathcal{P}] \cdot b(s) ds \right\rangle \tag{3.3.42}$$

for any non-stochastic and smooth function $b(s)$.

A rod observable $\mathcal{A}[\boldsymbol{g}]$ is said to be an *internal rod observable* if it is invariant through any overall rigid body motion of the chain, i.e. if

$$\mathcal{A}[B\boldsymbol{g}] = \mathcal{A}[\boldsymbol{g}] \tag{3.3.43}$$

for any $B \in \mathrm{SE}(3)$. Similarly, an *internal birod observable* $\mathcal{A}[\boldsymbol{g}, \mathcal{P}]$ has the property

$$\mathcal{A}[B\boldsymbol{g}, \mathcal{P}] = \mathcal{A}[\boldsymbol{g}, \mathcal{P}] \tag{3.3.44}$$

for any $B \in \mathrm{SE}(3)$. Internal observables are the only observables whose expectation is well defined for isolated rods and birods.

As in the case of chains, we are particularly interested in the specific internal observable of the relative rigid body motion $\boldsymbol{g}(s', s)$ between the cross section located at s'

and the one located at s in the rod configuration, as defined in (3.1.7), or respectively in the birod macrostructure. The *first moment of the relative rigid body motion*, or *configurational first moment*, is defined as for a general rod observable, i.e. as

$$\big\langle \boldsymbol{g}(s',s)\big\rangle = \int_\Gamma \boldsymbol{g}(s',s)d\rho[\boldsymbol{g}] \tag{3.3.45}$$

according to the definitions (3.3.35), and as

$$\big\langle \boldsymbol{g}(s',s)\big\rangle = \int_\Gamma \boldsymbol{g}(s',s)d\rho[\boldsymbol{g},\mathcal{P}] \tag{3.3.46}$$

for birods according to (3.3.40). In both cases, the block structure of $\langle \boldsymbol{g}(s',s)\rangle$ can be written in terms of the relative rotation $\boldsymbol{R}(s',s)$ and relative translation $\boldsymbol{r}(s',s)$, defined in (3.1.9), as

$$\big\langle \boldsymbol{g}(s',s)\big\rangle = \begin{pmatrix} \langle \boldsymbol{R}(s',s)\rangle & \langle \boldsymbol{r}(s',s)\rangle \\ 0 & 1 \end{pmatrix}. \tag{3.3.47}$$

The importance of the study of the first moment $\langle \boldsymbol{g}(s',s)\rangle$ is due to the fact that it encodes, similarly to chains in section 2.4, many other important observables in polymer physics, as discussed more precisely in section 3.4.

An other set of internal observables is the first moment of the birod internal variables $\xi(s)$ and $y(s)$, as defined in (3.2.55). The *first moment of the macrostructure generator vector* $\langle \xi(s)\rangle \in \mathbb{R}^6$ is denoted by $\bar{\xi}(s) \in \mathbb{R}^6$ and defined as

$$\left\langle \int_0^L \xi(s)\cdot b(s)ds \right\rangle = \int_0^L \bar{\xi}(s)\cdot b(s)ds \tag{3.3.48}$$

for any deterministic function $b(s) \in \mathbb{R}^6$, according (3.3.39). The *first moment of the birod microstructure internal variables* $\langle y(s)\rangle$ is denoted by $\hat{y}(s)$ and defined analogously to (3.3.48).

Interestingly, if one assumes a birod configurational distribution of the form (3.3.34) with an internal energy of the form (3.2.57), then the cross-correlations between the birod variables $y(s'), \frac{d}{ds}y(s')$ and $\xi(s')$ and $y(s), \frac{d}{ds}y(s)$ and $\xi(s)$ are in general non-vanishing for $s' \neq s$. For instance, the *second moment of the macrostructure generator vector*, denoted by $\mathcal{C}^{\xi\xi}(s',s) \in \mathbb{R}^{6\times 6}$, is defined as

$$\left\langle \int_0^L \int_0^L \big(\xi(s') - \bar{\xi}(s')\big) \otimes \big(\xi(s) - \bar{\xi}(s)\big) : B(s',s)ds'ds \right\rangle = \int_0^L \int_0^L \mathcal{C}^{\xi\xi}(s',s)\cdot B(s',s)ds'ds \tag{3.3.49}$$

for any deterministic function $B(s',s) \in \mathbb{R}^{6\times 6}$. All the other cross correlations are defined similarly.

Moreover, if $\langle\,.\,\rangle$ in (3.3.49) stands for the expectation with respect to a rod configurational distribution governed by a local internal energy of the form in (3.2.5), or

with respect to a birod distribution governed by a separable internal energy of the form (3.2.84), the generator vectors $\xi(s')$ and $\xi(s)$ have the property of being independent for $s' \neq s$. Precisely, one obtains

$$\begin{aligned}
&\int_0^L \int_0^L \mathcal{C}^{\xi\xi}(s',s) : B(s',s)ds'ds \\
&= \left\langle \int_0^L \int_0^L \big(\xi(s') - \bar{\xi}(s')\big) \otimes \big(\xi(s) - \bar{\xi}(s)\big) : B(s',s)ds'ds \right\rangle \\
&= \lim_{d\to\infty} \sum_{i=1}^{d-1} \sum_{j=1}^{d-1} \int_{{s_i}^{(d)}}^{{s_{i+1}}^{(d)}} \int_{{s_j}^{(d)}}^{{s_{j+1}}^{(d)}} {h_i}^{(d)} \left\langle \Big(\xi(s')^{(d)} - \bar{\xi}(s')\Big) \otimes \Big(\xi(s)^{(d)} - \bar{\xi}(s)\Big) \right\rangle^{(d)} \frac{\delta_{ij}}{{h_i}^{(d)}} : B(s',s)ds'ds \\
&= \int_0^L \mathcal{C}(s)\delta_s(s') : B(s',s)ds'ds
\end{aligned} \tag{3.3.50}$$

for any deterministic matrix function $B(s',s)$, where the matrix $\mathcal{C}(s)$ is defined as

$$\mathcal{C}(s) = \lim_{d\to\infty} {h_i}^{(d)} \left\langle \Big(\xi(s)^{(d)} - \bar{\xi}(s)\Big) \otimes \Big(\xi(s)^{(d)} - \bar{\xi}(s)\Big) \right\rangle^{(d)} \text{ for } {s_i}^{(d)} < s < {s_{i+1}}^{(d)} \tag{3.3.51}$$

and δ_s denotes the famous *Dirac function* on the interval $]0, L[$. Consequently the matrix $\mathcal{C}^{\xi\xi}(s',s)$ is of the form

$$\mathcal{C}^{\xi\xi}(s',s) = \mathcal{C}(s)\delta_s(s'). \tag{3.3.52}$$

3.4 The Persistence Matrix for Heterogenous Continuum rods

This section presents detailed results on the configurational first moment for continuum rods. Its relation to other important observables in polymer physics is briefly discussed and we propose to study the behaviour of the frame correlation matrix and of the Flory persistence vector, as in the chain case. Introducing the definitions of the expected rod configuration and of the fluctuation matrix leads to the notion of a semi-flexible regime for rods and to an explicit ODE governing, at first order, the rod configurational first moment. As in the case of chains, the equal importance of expected rod configuration and of the fluctuation matrix is observed. For non-degenerate rods, the frame correlation matrix is shown to have, in general, a non-monotonic decay and the Flory persistence vector to be convergent. The definition of the persistence matrix for rods allows a factorization of the configurational first moment and emphasises, as for semi-flexible chains, the contribution of the possibly rapidly varying expected configuration. A short length expansion is then obtained which yields very explicit expressions for configurational first moment.

3.4.1 Ordinary differential equation for the rod first moment

The *first moment of the relative rigid body motion* between the cross section located at s' and the one located at s in the rod configuration, or *configurational first moment*, is denoted by $\langle \boldsymbol{g}(s', s)\rangle = (\langle \boldsymbol{R}(s', s)\rangle, \langle \boldsymbol{r}(s', s)\rangle)$ as defined in section 3.3.2. Similarly to the discussion made for chains in section 2.4.1, this matrix is of interest because of its relation to popular observables in polymer physics, but here in their continuum versions.

(a) The *Kratky-Porod persistence length*, which is the component of the vector $\langle \boldsymbol{r}(s', \infty)\rangle$ in the direction of the rod tangent vector located at the arc length s' and has been proposed, following the work of M. Smoluchowski on stochastic trajectories, as a characteristic value for a rod [Porod, 1948, Krakty and Porod, 1949, Landau and Lifshitz, 1959, Yamakawa, 1976, Rivetti et al., 1998, Panyukov and Rabin, 2000, Becker and Everaers, 2007]

(b) The *Flory persistence vector* which is the vector $\langle \boldsymbol{r}(s', s)\rangle$, analogously to chains, and generalises the Kratky-Porod persistence length [Yamakawa, 1997, Panyukov and Rabin, 2000, Becker, 2007]

(c) The *tangent-tangent correlation* which is defined as $t \cdot \langle \boldsymbol{R}(s', s)\rangle\, t$ and where $t \in \mathbb{R}^3$ describes a given and deterministic direction in the frame $\boldsymbol{R}(s')$ and $\boldsymbol{R}(s)$ respectively. In some very specific cases this quantity is related to the Kratky-Porod persistence length and to some bending parameters of the rod [Landau and Lifshitz, 1959, Schellman, 1974, Doi and Edwards, 1986, Yamakawa, 1997].

(d) The *frame correlation matrix* $\langle \boldsymbol{R}(s', s)\rangle$ contains all the cross correlations between the director frames at the arc-length s' and s. It generalises, analogously to chains,

the tangent-tangent correlation [Maroun and Olson, 1988, Panyukov and Rabin, 2000, Vaillant, 2001, Becker and Everaers, 2007, Liu et al., 2011].

(e) The *mean squared end-to-end distance* is strongly related, but cannot be directly deduced in general, from the matrix $\langle \boldsymbol{R}(s',s)\rangle$ [Landau and Lifshitz, 1959, Yamakawa, 1976, Panyukov and Rabin, 2000, Becker, 2007].

Similarly to section 2.4 for the case of chains, we aim to study more precisely some general properties of the rod first moment $\langle \boldsymbol{g}(s',s)\rangle$ and we will focus the discussion on a single rod admitting a local internal energy $E[\boldsymbol{g}]$, in the sense of (3.2.5). Since the associated configurational distribution has therefore the property that each finite dimensional distribution is of the particular form (3.3.18), the following ODE for the first moment $\langle \boldsymbol{g}(s',s)\rangle$ is implied

$$\partial_s \left\langle \boldsymbol{g}(s',s)\right\rangle = \left\langle \boldsymbol{g}(s',s)\right\rangle \mathcal{A}(s) \tag{3.4.1a}$$

where

$$\mathcal{A}(s) = \lim_{d\to\infty} \frac{1}{{h_i}^{(d)}} \left(\left\langle {a_i}^{(d)} \right\rangle^{(d)} - \mathrm{Id}_4 \right) \text{ for } {s_i}^{(d)} < s < {s_{i+1}}^{(d)} \tag{3.4.1b}$$

provided that this limit exists. In more details, using the definitions (3.1.10) and (3.3.38), and the property (3.3.14), the coefficient matrix $\mathcal{A}(s)$ has to be understood as satisfying

$$\begin{aligned} \int_0^L \left\langle \partial_s \boldsymbol{g}(s',s)\right\rangle : B(s)ds &= \left\langle \int_0^L \partial_s \boldsymbol{g}(s',s) : B(s)ds \right\rangle \\ = \lim_{d\to\infty} \sum_{i=1}^{d-1} \int_{{s_i}^{(d)}}^{{s_{i+1}}^{(d)}} \boldsymbol{g}^{(d)}(s',s) \frac{1}{{h_i}^{(d)}} \left(\left\langle {a_i}^{(d)} \right\rangle^{(d)} - \mathrm{Id}_4 \right) &: B(s)ds \\ &= \int_0^L \boldsymbol{g}(s',s)\mathcal{A}(s) : B(s)ds. \end{aligned} \tag{3.4.2}$$

for any deterministic and smooth matrix function $B(s) \in \mathbb{R}^{4\times 4}$, where ${a_i}^{(d)}$ denotes the junction displacement of a discretised rod configuration as defined in (3.3.4). A similar approach has been used in [Panyukov and Rabin, 2000] for a specific case of the rod energy $E[\boldsymbol{g}]$. The precise entries of the matrix $\mathcal{A}(s)$ depend on the details of the rod constitution. We will show however, that some general behaviour of the first moment $\langle \boldsymbol{g}(s',s)\rangle$ can be stated, similarly to the rigid body chain case (2.4.1). More generally, the relation in (3.4.1) gives

$$\left\langle \boldsymbol{g}(s'',s)\right\rangle = \left\langle \boldsymbol{g}(s'',s')\right\rangle \left\langle \boldsymbol{g}(s',s)\right\rangle \tag{3.4.3}$$

for all $0 < s'' < s' < s < L$, analogously to (3.1.11).

In terms of the moment $\langle \boldsymbol{R}(s',s)\rangle$ and $\langle \boldsymbol{r}(s',s)\rangle$ the ODE (3.4.1) then reads

$$\partial_s \left\langle \boldsymbol{R}(s',s)\right\rangle = \left\langle \boldsymbol{R}(s',s)\right\rangle \mathcal{A}^{\boldsymbol{R}}(s) \tag{3.4.4a}$$

$$\partial_s \left\langle \boldsymbol{r}(s',s) \right\rangle = \left\langle \boldsymbol{R}(s',s) \right\rangle \mathcal{A}^{\boldsymbol{r}}(s) \tag{3.4.4b}$$

where $\mathcal{A}^{\boldsymbol{R}}(s) \in \mathbb{R}^{3\times 3}$ and $\mathcal{A}^{\boldsymbol{r}}(s) \in \mathbb{R}^3$ denote the corresponding blocks of the matrix $\mathcal{A}(s)$, i.e.

$$\mathcal{A}(s) = \begin{pmatrix} \mathcal{A}^{\boldsymbol{R}}(s) & \mathcal{A}^{\boldsymbol{r}}(s) \\ 0 & 0 \end{pmatrix}. \tag{3.4.5}$$

for $0 < s' < s < L$. Observe that the expressions in (3.4.1) and (3.4.22) are the continuum analogue of the matrix multiplication method in (2.4.1) and (2.4.5), as originally proposed in [Flory and Jernigan, 1965, Flory and Miller, 1966]. They are also a generalisation of the similar expression appearing in [Panyukov and Rabin, 2000].

Similarly to chains, we conclude that the key step for understanding the features of the observables listed in (a) to (e) above in the case of rods with a local internal energy is the explicit knowledge of the coefficient matrix $\mathcal{A}(s)$ in (3.4.1b) [Panyukov and Rabin, 2000, Becker and Everaers, 2007]. It would then be valuable to know the relation between the rod configurational distribution in (3.2.5) and the entries in $\mathcal{A}(s)$. However an exact and analytic computation for most realistic rod distributions is not available. We will however show in section 3.4.2 that an approximation can be provided for general local rod configurational distributions in the *semi-flexible regime*. Observe that the argument to deduce the ODE (3.4.1a) fails for a *non-local rod energy*, or even for a general local birod distribution $d\rho[\boldsymbol{g}, \mathcal{P}]$, which emphasises the difficulty of studying the moment $\langle \boldsymbol{g}(s',s) \rangle$ for non idealised rod model.

3.4.2 Explicit expansion for semi-flexible rods

Analogously to the expansion made for rigid body chain first moment in section 2.4.2, we need to introduce the particular rod configuration $\bar{\boldsymbol{g}}(s)$, called the *expected rod configuration*. Its definition is strongly related to that of the expected chain configuration $\bar{\boldsymbol{g}}_n$.

Introduce a mesh of the interval $]0, L[$ in the form (3.3.2) and for any continuous rod configuration $\boldsymbol{g}(s)$ denote by $\boldsymbol{g}^{(d)}(s)$ its interpolation on this mesh. Similarly to (2.4.7), the chain junction displacement ${a_i}^{(d)}$, as defined in (3.3.4), can be factorized in the form

$${a_i}^{(d)} = {\alpha_i}^{(d)} \bar{a}_i^{(d)} \tag{3.4.6a}$$

where

$${\alpha_i}^{(d)} = \exp\left\{ \mathcal{T} {\Theta_i}^{(d)} \right\} \tag{3.4.6b}$$

for some ${\Theta_i}^{(d)} \in \mathbb{R}^6$ which is assumed to have the property that

$$\left\langle {\Theta_i}^{(d)} \right\rangle^{(d)} = 0 \tag{3.4.6c}$$

where $\langle\,.\,\rangle^{(d)}$ denotes the finite dimensional expectation with respect to the finite dimensional rod marginal distribution $d\rho^{(d)}(\boldsymbol{g}^{(d)})(x^{(d)})$ defined in (3.3.18). We recall that the existence of $\bar{a}_i^{(d)}$ is guaranteed if the configurational distribution $d\rho^{(d)}(\boldsymbol{g}^{(d)})(x^{(d)})$ is sufficiently peaked [Kendall, 1990, Becker and Everaers, 2007, Chirikjian, 2011].

Using the property (3.3.14), the *expected generator vector* is defined as the triple $\bar{\xi}(s) \in \mathbb{R}^3$ which satisfies

$$\lim_{d\to\infty} \frac{1}{{h_i}^{(d)}} \left(\bar{a}_i^{(d)} - \mathrm{Id}_4\right) = \mathcal{T}\bar{\xi}(s) \text{ for } {s_i}^{(d)} < s < {s_{i+1}}^{(d)} \tag{3.4.7}$$

and the *expected rod* configuration $\bar{\boldsymbol{g}}(s)$ is defined as the solution of

$$\frac{d}{ds}\bar{\boldsymbol{g}}(s) = \bar{\boldsymbol{g}}(s)\mathcal{T}\bar{\xi}(s) \text{ for } s \in]0, L[\tag{3.4.8}$$

with $\bar{\boldsymbol{g}}(0) = \mathrm{Id}_4$, according to the definition (3.1.4). Observe that if ${a_i}^{(d)}(s)$ and $\bar{a}_i^{(d)}(s)$ denote the local interpolations from Id_4 to ${a_i}^{(d)}$ and $\bar{a}_i^{(d)}$ respectively, and if $\xi^{(d)}(s)$ and $\bar{\xi}^{(d)}(s)$ denote their associated interpolation generator vectors, then the definitions in (3.4.6a) and (3.4.6b), lead to the relation

$$\frac{1}{{h_i}^{(d)}}\mathcal{T}{\Theta_i}^{(d)} = \bar{a}_i^{(d)}(s)\mathcal{T}(\xi^{(d)} - \bar{\xi}^{(d)})(s)(\bar{a}_i^{(d)}(s))^{-1} \tag{3.4.9}$$

or equivalently to

$$\frac{1}{{h_i}^{(d)}}{\Theta_i}^{(d)} = \mathrm{Ad}_{\bar{a}_i^{(d)}(s)}\left(\xi^{(d)}(s) - \bar{\xi}^{(d)}(s)\right) \tag{3.4.10}$$

where the matrix $\mathrm{Ad}_{\bar{a}_i^{(d)}(s)}$ is defined using (1.1.28). Interestingly, in contrast to the rigid body chain case in (2.4.22), the expected generator vector $\bar{\xi}(s)$ is strongly related to the first moment of the generator vector $\xi(s)$. Precisely, for a given mesh of the interval $]0, L[$ of the form in (3.3.2), if we take the expectation on both sides of (3.4.10) and if we use the definitions (3.4.6c) and (3.3.39), we conclude that

$$\langle\xi(s)\rangle = \lim_{d\to\infty}\left\langle\xi^{(d)}(s)\right\rangle^{(d)} = \bar{\xi}(s) \tag{3.4.11}$$

which justifies the notation of the rod generator vector first moment in (3.3.48). More generally, we can define the *expected rod relative rigid body displacement* $\bar{\boldsymbol{g}}(s', s)$ by

$$\partial_s\bar{\boldsymbol{g}}(s', s) = \bar{\boldsymbol{g}}(s', s)\mathcal{T}\bar{\xi}(s) \text{ with } \bar{\boldsymbol{g}}(s', s') = \mathrm{Id}_4 \tag{3.4.12}$$

whose block structure is denoted

$$\bar{\boldsymbol{g}}(s', s) = \begin{pmatrix} \bar{\boldsymbol{R}}(s', s) & \bar{\boldsymbol{r}}(s', s) \\ 0 & 1 \end{pmatrix} \tag{3.4.13}$$

where $\bar{\boldsymbol{R}}(s',s)$ and $\bar{\boldsymbol{r}}(s',s)$ denote respectively the *expected rod relative rotation* and the *expected rod relative translation*.

The *local rod fluctuation matrix* $\mathcal{C}(s)$ is defined as the limit

$$\mathcal{C}(s) = \lim_{d\to\infty} \frac{1}{{h_i}^{(d)}} \left\langle {\Theta_i}^{(d)} \otimes {\Theta_i}^{(d)} \right\rangle^{(d)} \text{ for } {s_i}^{(d)} < s < {s_{i+1}}^{(d)} \tag{3.4.14}$$

when this limit exists in the sense of (3.3.39). Using the natural splitting of the vector ${\Theta_i}^{(d)} = (\Theta_i^{\boldsymbol{R}(d)}, \Theta_i^{\boldsymbol{r}(d)})$, as in (3.1.6) for instance, its block structure can be written as

$$\mathcal{C}(s) = \begin{pmatrix} \mathcal{C}^{\boldsymbol{RR}}(s) & \mathcal{C}^{\boldsymbol{Rr}}(s) \\ \mathcal{C}^{\boldsymbol{Rr}}(s)^T & \mathcal{C}^{\boldsymbol{rr}}(s) \end{pmatrix} \tag{3.4.15a}$$

with

$$\begin{aligned} \mathcal{C}^{\boldsymbol{RR}}(s) &= \lim_{d\to\infty} \frac{1}{{h_i}^{(d)}} \left\langle \Theta_i^{\boldsymbol{R}(d)} \otimes \Theta_i^{\boldsymbol{R}(d)} \right\rangle^{(d)} \\ \mathcal{C}^{\boldsymbol{Rr}}(s) &= \lim_{d\to\infty} \frac{1}{{h_i}^{(d)}} \left\langle \Theta_i^{\boldsymbol{R}(d)} \otimes \Theta_i^{\boldsymbol{r}(d)} \right\rangle^{(d)} \\ \mathcal{C}^{\boldsymbol{rr}}(s) &= \lim_{d\to\infty} \frac{1}{{h_i}^{(d)}} \left\langle \Theta_i^{\boldsymbol{r}(d)} \otimes \Theta_i^{\boldsymbol{r}(d)} \right\rangle^{(d)} \end{aligned} \tag{3.4.15b}$$

where the equality is in the sense of (3.3.39). The fluctuation matrix $\mathcal{C}(s)$ is strongly related to the second moment of the generator vector $\xi(s)$. Precisely, for a configurational distribution of the form (3.3.18) the centered second moment of $\xi(s)$ can be shown, using the identity (3.4.10) and the expression in (3.3.52), to satisfy

$$\left\langle \left(\langle \xi(s) \rangle - \bar{\xi}(s) \right) \otimes \left(\langle \xi(s) \rangle - \bar{\xi}(s) \right) \right\rangle = \mathcal{C}(s) \delta_s(s'). \tag{3.4.16}$$

Persevering with the similarity with the expansion made in the case of chains in section 2.4.2, we introduce a formal definition of the idea of *semi-flexible regime* for continuum rods. According to (3.4.6) the junction first moment of the discretised rod configuration $\left\langle {a_i}^{(d)} \right\rangle^{(d)}$ can be expanded as

$$\left\langle {a_i}^{(d)} \right\rangle^{(d)} = \left\langle {\alpha_i}^{(d)} \right\rangle^{(d)} \bar{a}_i^{(d)} \tag{3.4.17a}$$

with

$$\left\langle {\alpha_i}^{(d)} \right\rangle^{(d)} = \mathrm{Id}_4 + \left\langle \left(\mathcal{T} {\Theta_i}^{(d)} \right)^2 \right\rangle^{(d)} + \left\langle \mathrm{o}\left(\left| {\Theta_i}^{(d)} \right|^2 \right) \right\rangle^{(d)}. \tag{3.4.17b}$$

A continuum rod is then said to be in the semi-flexible regime if there exists a small parameter $\epsilon > 0$ such that

$$\lim_{d\to\infty} \frac{1}{{h_i}^{(d)}} \left| \left\langle \left(\mathcal{T} {\Theta_i}^{(d)} \right)^2 \right\rangle^{(d)} \right| = \mathrm{O}(\epsilon) \text{ and } \lim_{d\to\infty} \frac{1}{{h_i}^{(d)}} \left\langle \mathrm{o}\left(\left| {\Theta_i}^{(d)} \right|^2 \right) \right\rangle^{(d)} = \mathrm{o}(\epsilon). \tag{3.4.18}$$

Consequently, defining the matrix

$$\mathcal{L}(s) = \begin{pmatrix} \mathcal{L}^{\boldsymbol{RR}}(s) & \mathcal{L}^{\boldsymbol{Rr}}(s) \\ 0 & 0 \end{pmatrix} = \lim_{d\to\infty} \frac{-1}{{h_i}^{(d)}} \left\langle \left(\mathcal{T}{\Theta_i}^{(d)} \right)^2 \right\rangle^{(d)} \tag{3.4.19a}$$

whose block structure is given, according to the identity in (2.4.15) and to the definition of the rod fluctuation matrix in (3.4.15), by

$$\mathcal{L}^{\boldsymbol{RR}}(s) = \mathrm{Tr}\left(\mathcal{C}^{\boldsymbol{RR}}(s)\right) \mathrm{Id}_3 - \mathcal{C}^{\boldsymbol{RR}}(s) \text{ and } \mathcal{L}^{\boldsymbol{Rr}}(s) = \mathrm{Vect}\left(\mathcal{C}^{\boldsymbol{Rr}}(s)\right) \tag{3.4.19b}$$

where the operator Vect (.) is defined in (1.1.13), shows that the coefficient matrix $\mathcal{A}(s)$, as defined in (3.4.1b), can be approximated to first order by

$$\mathcal{A}(s) = \mathcal{T}\bar{\xi}(s) - \frac{1}{2}\mathcal{L}(s) + \mathrm{o}\,(\epsilon) \tag{3.4.20}$$

for a semi-flexible continuum rod, according to (3.4.7) and (3.4.17). The rigid body displacement first moment $\langle \boldsymbol{g}(s', s)\rangle$ therefore satisfies

$$\partial_s \left\langle \boldsymbol{g}(s', s)\right\rangle = \left\langle \boldsymbol{g}(s', s)\right\rangle \left(\mathcal{T}\bar{\xi}(s) - \frac{1}{2}\mathcal{L}(s) \right) + \mathrm{o}\,(\epsilon) \text{ with } \left\langle \boldsymbol{g}(s', s')\right\rangle = \mathrm{Id}_4. \tag{3.4.21}$$

In terms of the relative rotation $\boldsymbol{R}(s', s)$ and translation $\boldsymbol{r}(s', s)$, as defined in (3.1.9), one has then

$$\partial_s \left\langle \boldsymbol{R}(s', s)\right\rangle = \left\langle \boldsymbol{R}(s', s)\right\rangle \left(\left[\bar{U}(s)\times\right] - \frac{1}{2}\mathcal{L}^{\boldsymbol{RR}}(s) \right) + \mathrm{o}\,(\epsilon) \text{ with } \boldsymbol{R}(s', s') = \mathrm{Id}_3 \tag{3.4.22a}$$

$$\partial_s \left\langle \boldsymbol{r}(s', s)\right\rangle = \left\langle \boldsymbol{R}(s', s)\right\rangle \left(\bar{V}(s) - \frac{1}{2}\mathcal{L}^{\boldsymbol{Rr}}(s) \right) + \mathrm{o}\,(\epsilon) \text{ with } \boldsymbol{r}(s', s') = 0. \tag{3.4.22b}$$

These expressions generalise the ones in [Yamakawa, 1997, Panyukov and Rabin, 2000] and a similar formulation can be found in [Becker, 2007] where they are deduced in the context of Brownian dynamics on Lie groups. Relations (3.4.22) emphasize the equal importance of the expected rod configuration $\bar{\boldsymbol{g}}(s', s)$ and of the fluctuation matrix $\mathcal{C}(s)$ in the behaviour of the rod first moment $\langle \boldsymbol{g}(s', s)\rangle$, similarly to the rigid body chain case in (2.4.26). We observe that in the semi-flexible definition (3.4.18) the first estimate can then be replaced by

$$\left|\mathcal{C}^{\boldsymbol{RR}}(s)\right|_{\mathrm{Sp}} = \mathrm{O}\,(\epsilon) \text{ and } \left|\mathrm{Vect}\left(\mathcal{C}^{\boldsymbol{Rr}}(s)\right)\right| = \mathrm{O}\,(\epsilon) \tag{3.4.23}$$

for all $s \in]0, L[$ and where we recall that $|\,.\,|_{\mathrm{Sp}}$ denotes the spectral norm, as defined in (1.1.4), and, moreover, the value of the parameter ϵ can be estimated as

$$\epsilon = \max_{0\le s\le L} \left\{ \frac{2\left|\mathcal{C}^{\boldsymbol{RR}}(s)\right|_{\mathrm{Sp}}}{\pi}, \frac{2\left|\mathrm{Vect}\left(\mathcal{C}^{\boldsymbol{Rr}}(s)\right)\right|}{\pi} \right\} \tag{3.4.24}$$

where $\frac{\pi}{2}$ and 1 have been chosen as characteristic scales for angles and angular speed respectively. We note that (3.4.20) implies that the first moment $\langle \boldsymbol{g}(s', s) \rangle$ does not belong the group of rigid motions, as expected from the definition of the expectation $\langle \, . \, \rangle$. Moreover, we want to stress the similarity between expressions (3.4.11) and (3.4.16) and the one which appears in the theory of rigid body stochastic dynamics. In particular, the matrix $\mathcal{C}(s)$ plays the role of diffusion tensor [Steele, 1963a, Steele, 1963b, Brenner, 1964, Brenner, 1965] and $\bar{\xi}(s)$ the role of intrinsic velocity [Zhou and Chirikjian, 2003, Becker, 2007, Long et al., 2012]. It is finally interesting to observe that the relations (3.4.22) constitute literally the continuum version of the ones deduces for semi-flexible rigid body chains in (2.4.27).

3.4.3 General behavior of semi-flexible rod first moment

The expressions (3.4.22) show that the general behaviour of the first moment of a continuum rod configuration can be very diverse and strongly dependent on the detailed constitution of the rod. In particular, there is no reason to expect any monotone behaviour in the entries of $\langle \boldsymbol{R}(s', s) \rangle$ nor $\langle \boldsymbol{r}(s', s) \rangle$, specifically because the expected generator vector $\bar{\xi}(s)$ is generally relatively large, as in DNA models for instance. However, and similarly to the rigid body chain case discussed in section 2.4.3, one can show some general features for semi-flexible continuum rods.

Concerning the first moment of the relative rotation matrix $\boldsymbol{R}(s', s)$, the relation (3.4.22a) implies, at leading order, that

$$\partial_s \left| \left\langle \boldsymbol{R}(s', s) \right\rangle \right|_{\mathrm{Sp}}^2 \leq - \left| \mathcal{L}^{\boldsymbol{RR}}(s) \right|_{\mathrm{Sp}} \left| \left\langle \boldsymbol{R}(s', s) \right\rangle \right|_{\mathrm{Sp}}^2 + \mathrm{o}\left(\epsilon^2\right) \tag{3.4.25}$$

for $s' \leq s$ which then leads, using the *Grönwall Inequality*, to

$$\left| \left\langle \boldsymbol{R}(s', s) \right\rangle \right|_{\mathrm{Sp}} \leq \exp\left\{ -(s - s')\lambda^{\min} \right\} + \mathrm{o}\left(\frac{\exp\left\{ -(s - s')\lambda^{\min} \right\}}{\lambda^{\min}} \epsilon \right) \tag{3.4.26}$$

where $\lambda^{\min} = \min_{s' < t < s} \lambda^{\min}(t)$ and where $\lambda^{\min}(t)$ stands for the minimum eigenvalue of the matrix $\frac{1}{2}\mathcal{L}^{\boldsymbol{RR}}(t)$. Using the definition (3.4.19b), the value of $\lambda^{\min}$ can be written as

$$\lambda^{\min} = \min_{s' < t < s} \frac{1}{2}(\gamma_1(t) + \gamma_2(t)) \tag{3.4.27}$$

where $\gamma_1(t)$ and $\gamma_2(t)$ denote the two smallest eigenvalues of $\mathcal{C}^{\boldsymbol{RR}}(t)$. As in the rigid body chain case, the value of $\lambda^{\min}$ vanishes if at least one of the matrices $\mathcal{C}^{\boldsymbol{RR}}(t)$ has two zero eigenvalues. If $\lambda^{\min} = 0$ the rod is referred to as a *degenerate rod*. Consequently, if the rod is non degenerate expression (3.4.26) shows that all frame cross-correlations, at leading order, are asymptotically vanishing, i.e.

$$\left\langle \boldsymbol{R}(s', s) \right\rangle \to 0 \text{ for } s' - s >> 1. \tag{3.4.28}$$

For the first moment of the translation vector $\langle \boldsymbol{r}(s',s)\rangle$, the relations (3.4.22) and (3.4.26) lead, at leading order, to

$$\left|\partial_s \left\langle \boldsymbol{r}(s',s)\right\rangle\right| \leq \exp\left\{-(s-s')\lambda^{\min}\right\} \left|\bar{V} - \frac{1}{2}\mathcal{L}^{\boldsymbol{Rr}}\right| \tag{3.4.29}$$

and

$$\left|\left\langle \boldsymbol{r}(s',s)\right\rangle\right| \leq \frac{1}{\lambda^{\min}}\left(1 - \exp\left\{-(s-s')\lambda^{\min}\right\}\right) \left|\bar{V} - \frac{1}{2}\mathcal{L}^{\boldsymbol{Rr}}\right| \tag{3.4.30}$$

where $\left|\bar{V} - \frac{1}{2}\mathcal{L}^{\boldsymbol{Rr}}\right| = \min_{s'<t<s}\left|\bar{V}(t) - \frac{1}{2}\mathcal{L}^{\boldsymbol{Rr}}(t)\right|$. For a non-degenerate rod, the first moment $\langle \boldsymbol{r}(s',s)\rangle$ is then bounded and convergent to a limiting value $\langle \boldsymbol{r}(s',\infty)\rangle$ which satisfies

$$\left\langle \boldsymbol{r}(s',\infty)\right\rangle \leq \frac{1}{\lambda^{\min}} \left|\bar{V} - \frac{1}{2}\mathcal{L}^{\boldsymbol{Rr}}\right|. \tag{3.4.31}$$

Statements (3.4.28) and (3.4.31) constitute general convergence results for non-degenerate semi-flexible rods. They show that the larger the arc-length distance $s'-s$, the more *uncorrelated* is the relative orientation $\boldsymbol{R}(s)$ from $\boldsymbol{R}(s')$ and, moreover, that the limiting value of the Flory persistence vector is always finite. More informations about the specific physical constitution of the rod has to be known in order to obtain finer analytical estimates.

Some particular cases of constitutively homogeneous rods are discussed in [Yamakawa, 1997, Panyukov and Rabin, 2000] and more generally in [Becker and Everaers, 2007]. In particular, a closed form expression can be written for the first moments $\langle \boldsymbol{R}(s',s)\rangle$ and $\langle \boldsymbol{r}(s',s)\rangle$ in (3.4.22), i.e.

$$\left\langle \boldsymbol{R}(s',s)\right\rangle = \exp\left\{(s-s')\mathcal{A}^{\boldsymbol{R}}\right\} \tag{3.4.32a}$$

$$\left\langle \boldsymbol{r}(s',s)\right\rangle = \left(\exp\left\{(s-s')\mathcal{A}^{\boldsymbol{R}}\right\} - \mathrm{Id}_3\right)\left(\mathcal{A}^{\boldsymbol{R}}\right)^{-1}\mathcal{A}^{\boldsymbol{r}} \tag{3.4.32b}$$

where

$$\mathcal{A}^{\boldsymbol{R}} = \left[\bar{U}\times\right] - \frac{1}{2}\mathcal{L}^{\boldsymbol{RR}} \tag{3.4.33a}$$

$$\mathcal{A}^{\boldsymbol{r}} = \bar{V} - \frac{1}{2}\mathcal{L}^{\boldsymbol{Rr}} \tag{3.4.33b}$$

The limiting values of the Flory persistence vector is then,

$$\left\langle \boldsymbol{r}(s',\infty)\right\rangle = -\left(\mathcal{A}^{\boldsymbol{R}}\right)^{-1}\mathcal{A}^{\boldsymbol{r}}. \tag{3.4.34}$$

We emphasize the analogy between these expressions and the ones obtained in (2.4.37) and (2.4.39) for the rigid body chain first moments.

3.4.4 Global factorization and the persistence matrix

The explicit ODE in (3.4.21) allows the study of the configurational first moment $\boldsymbol{g}(s',s)$ behaviour using numerical methods. However, clear and general relations between the values of the expected rod generator vector $\bar{\xi}(s)$, the fluctuation matrix $\mathcal{C}(s)$ and the behaviour of the entries in $\boldsymbol{g}(s',s)$ are still missing. Using an analogous approach to the one made for chains in section 2.4.4 provides key features of the statistical physics of continuum rods.

In analogy to the rigid body chain case, we define the *deformation matrix* $\boldsymbol{\mathcal{D}}(s',s)$ through the factorization

$$\boldsymbol{g}(s',s) = \boldsymbol{\mathcal{D}}(s',s)\bar{\boldsymbol{g}}(s',s) \tag{3.4.35}$$

where $\bar{\boldsymbol{g}}(s',s)$ is given in (3.4.12). The first moment $\langle \boldsymbol{g}(s',s)\rangle$ then becomes

$$\big\langle \boldsymbol{g}(s',s)\big\rangle = \big\langle \boldsymbol{\mathcal{D}}(s',s)\big\rangle\, \bar{\boldsymbol{g}}(s',s) \tag{3.4.36}$$

or equivalently in terms of the relative rotation and translation

$$\big\langle \boldsymbol{R}(s',s)\big\rangle = \big\langle \boldsymbol{\Delta}(s',s)\big\rangle\, \bar{\boldsymbol{R}}(s',s) \tag{3.4.37a}$$

$$\big\langle \boldsymbol{r}(s',s)\big\rangle = \big\langle \boldsymbol{\delta}(s',s)\big\rangle + \big\langle \boldsymbol{\Delta}(s',s)\big\rangle\, \bar{\boldsymbol{r}}(s',s) \tag{3.4.37b}$$

where we have defined the block structure of $\langle \boldsymbol{\mathcal{D}}(s',s)\rangle$ as

$$\big\langle \boldsymbol{\mathcal{D}}(s',s)\big\rangle = \begin{pmatrix} \langle \boldsymbol{\Delta}(s',s)\rangle & \langle \boldsymbol{\delta}(s',s)\rangle \\ 0 & 1 \end{pmatrix} \tag{3.4.38}$$

and used the block structure of $\bar{\boldsymbol{g}}(s',s)$ in (3.4.13).

Given the ODE deduced for the first moment $\langle \boldsymbol{g}(s',s)\rangle$ in (3.4.21), together with the one for the expected rod configuration $\bar{\boldsymbol{g}}(s',s)$ in (3.4.8), one can then show that the deformation matrix first moment $\langle \boldsymbol{\mathcal{D}}(s',s)\rangle$ is governed by the relation

$$\partial_s \big\langle \boldsymbol{\mathcal{D}}(s',s)\big\rangle = \big\langle \boldsymbol{\mathcal{D}}(s',s)\big\rangle \left\{ -\frac{1}{2}\bar{\boldsymbol{g}}(s',s)\mathcal{L}(s)\bar{\boldsymbol{g}}^{-1}(s',s) \right\} \text{ with } \big\langle \boldsymbol{\mathcal{D}}(s',s')\big\rangle = \mathrm{Id}_4 \tag{3.4.39}$$

or equivalently

$$\partial_s \big\langle \boldsymbol{\Delta}(s',s)\big\rangle = \big\langle \boldsymbol{\Delta}(s',s)\big\rangle \left\{ -\frac{1}{2}\bar{\boldsymbol{R}}(s',s)\mathcal{L}^{\boldsymbol{RR}}(s)\bar{\boldsymbol{R}}^{T}(s',s) \right\} \tag{3.4.40a}$$

$$\partial_s \big\langle \boldsymbol{\delta}(s',s)\big\rangle = \big\langle \boldsymbol{\Delta}(s',s)\big\rangle \left\{ \frac{1}{2}\bar{\boldsymbol{R}}(s',s)\mathcal{L}^{\boldsymbol{RR}}(s)\bar{\boldsymbol{R}}^{T}(s',s)\bar{\boldsymbol{r}}(s',s) - \frac{1}{2}\bar{\boldsymbol{R}}(s',s)\mathcal{L}^{\boldsymbol{Rr}}(s) \right\} \tag{3.4.40b}$$

where $\mathcal{L}^{\boldsymbol{RR}}(s)$ and $\mathcal{L}^{\boldsymbol{Rr}}(s)$ denote the blocks of $\mathcal{L}(s)$ defined in (3.4.19b). These relations imply that the entries in the block $\langle \boldsymbol{\Delta}(s',s)\rangle$ are asymptotically vanishing for non degenerate rods since

$$\exp\left\{-(s-s')\lambda^{\max}\right\} \leq \big|\big\langle \boldsymbol{\Delta}(s',s)\big\rangle\big|_{\mathrm{Sp}} \leq \exp\left\{-(s-s')\lambda^{\min}\right\} \tag{3.4.41}$$

according to the Grönwall Inequality, where $\lambda^{\max} = \max_{s' \leq t \leq s} \lambda^{\max}(t)$ with $\lambda^{\max}(t)$ is the maximum eigenvalue of $\frac{1}{2}\mathcal{L}^{\boldsymbol{RR}}(t)$ and is obtained similarly to $\lambda^{\min}$ in (3.4.27). Moreover, in contrast to the first moment of the relative rotation $\langle \boldsymbol{R}(s', s)\rangle$, the variation has to be small since

$$\left|\partial_s \left\langle \boldsymbol{\Delta}(s', s)\right\rangle\right|_{\mathrm{Sp}} = \mathrm{O}\left(\exp\left\{-(s - s')\lambda^{\min}\right\}\epsilon\right) \tag{3.4.42}$$

according to the semi-flexible hypothesis (3.4.23). Consequently, analogously to rigid body chains, the relative rotation first moment $\langle \boldsymbol{R}(s', s)\rangle$ can be factorized, for non-degenerate semi-flexible rods, into a fast and non-decaying rotation matrix $\bar{\boldsymbol{R}}(s', s)$ and another slowly varying and strictly decaying matrix $\langle \boldsymbol{\Delta}(s', s)\rangle$.

The estimate in (3.4.41) and the expression (3.4.40b) show that the translation part $\langle \boldsymbol{\delta}(s', s)\rangle$ also has a slow variation since

$$\left|\partial_s \left\langle \boldsymbol{\delta}(s', s)\right\rangle\right| = \mathrm{O}\left(\epsilon \exp\left\{-(s - s')\lambda^{\min}\right\}\left(\left|\bar{\boldsymbol{r}}(s', s)\right| + 1\right)\right) \tag{3.4.43}$$

and is bounded using the estimate

$$\begin{aligned}\left|\left\langle \boldsymbol{\delta}(s', s)\right\rangle\right| \leq \frac{1}{\lambda^{\min}}\left(1 - \exp\left\{-(s - s')\lambda^{\min}\right\}\right)\left(\frac{\lambda^{\max}}{\lambda^{\min}}\left|\bar{V}\right| + \left|\mathcal{L}^{\boldsymbol{Rr}}\right|\right) \\ -\frac{\lambda^{\max}}{\lambda^{\min}}\left|\bar{V}\right|\left(1 - \exp\left\{-(s - s')\lambda^{\min}\right\}\right)(s - s').\end{aligned} \tag{3.4.44}$$

Moreover using the definition (3.4.37b) one has that

$$\left\langle \boldsymbol{\delta}(s', \infty)\right\rangle = \left\langle \boldsymbol{r}(s', \infty)\right\rangle. \tag{3.4.45}$$

We again emphasize the similarity between the estimates (3.4.41) and (3.4.44) and the corresponding ones for the rigid body chain case in (2.4.47) and in (2.4.51).

The first moment of the translation vector $\langle \boldsymbol{r}(s', s)\rangle$ can then be decomposed into the non-converging triple $\bar{\boldsymbol{r}}(s', s)$ damped by the matrix $\langle \boldsymbol{\Delta}(s', s)\rangle$ and by a convergent part $\langle \boldsymbol{\delta}(s', s)\rangle$. Moreover, the triple $\langle \boldsymbol{\delta}(s', s)\rangle$ is made by a damped integral of the expected translation vector $\bar{\boldsymbol{r}}(s', s)$ according to its definition (3.4.40b), and has the same limiting values as the vector $\langle \boldsymbol{r}(s', s)\rangle$.

Similarly to the rigid body chain case, the matrix $\langle \boldsymbol{\mathcal{D}}(s', s)\rangle$ will be called the *persistence matrix*. The block $\langle \boldsymbol{\Delta}(s', s)\rangle$ encodes how much of the expected rod configuration $\bar{\boldsymbol{g}}(s', s)$ persists in the value of the first moment of the configurational first moment $\langle \boldsymbol{g}(s', s)\rangle$ according to (3.4.37), and the limiting value of $\langle \boldsymbol{\delta}(s', s)\rangle$ behaves like the Flory persistence vector $\langle \boldsymbol{r}(s', s)\rangle$, as shown in (3.4.45). The proposed factorization can be seen as a generalisation of the results in [Trifonov et al., 1988, Schellman and Harvey, 1995], but in a continuum version, as a way to deconvolve the effect of local intrinsic bent and thermal fluctuation on the behaviour of the first moment $\langle \boldsymbol{g}(s', s)\rangle$ for molecules.

3.4.5 Short and long length behavior approximation

Similarly to the discussion made for chains in section 2.4.5, we describe here how analytic solutions can be established in the case of continuum rods in the semi-flexible regime using the persistence matrix factorization in (3.4.36) and an asymptotic development in the rod reference length.

For s sufficiently close to s', and using the semi-flexible hypothesis in (3.4.23) the leading order term of the persistence matrix $\langle \boldsymbol{\mathcal{D}}(s',s)\rangle$ solution of (3.4.39) can be shown to be

$$\left\langle \boldsymbol{\mathcal{D}}(s',s)\right\rangle = \exp\left\{-\frac{1}{2}\int_{s'}^{s} \bar{\boldsymbol{g}}(s',t)\mathcal{L}(t)\bar{\boldsymbol{g}}(s',t)^{-1}dt\right\} + \mathrm{o}\left((s-s')\epsilon\right) \tag{3.4.46}$$

using the *Magnus expansion* for linear ODE systems.

A more explicit formulation is obtained by defining the matrices

$$\boldsymbol{\Omega}(s',s) = \frac{1}{2}\int_{s'}^{s} \bar{\boldsymbol{R}}(s',t)\mathcal{L}^{\boldsymbol{RR}}(t)\bar{\boldsymbol{R}}(s',t)^{T}dt \tag{3.4.47a}$$

and the triples

$$\boldsymbol{\omega}(s',s) = \left(\int_{s'}^{s} \bar{\boldsymbol{R}}(s',t)\mathcal{L}^{\boldsymbol{RR}}(t)\bar{\boldsymbol{R}}(s',t)^{T}dt\right)^{-1}\left[\int_{s'}^{s} \bar{\boldsymbol{R}}(s',t)\mathcal{L}^{\boldsymbol{RR}}(t)\bar{\boldsymbol{R}}(s',t)^{T}\bar{\boldsymbol{r}}(s',t) - \bar{\boldsymbol{R}}(s',t)\mathcal{L}^{\boldsymbol{Rr}}(t)dt\right] \tag{3.4.47b}$$

for each $s \geq s'$. The blocks of the persistence matrix in (3.4.38) then read, according to expression (3.4.46) and to the identity in (2.4.59),

$$\left\langle \boldsymbol{\Delta}(s',s)\right\rangle = \exp\left\{-\boldsymbol{\Omega}(s',s)\right\} + \mathrm{o}\left((s-s')\epsilon\right) \tag{3.4.48a}$$

$$\left\langle \boldsymbol{\delta}(s',s)\right\rangle = \left[\mathrm{Id}_3 - \exp\left\{-\boldsymbol{\Omega}(s',s)\right\}\right]\boldsymbol{\omega}(s',s) + \mathrm{o}\left((s-s')\epsilon\right) \tag{3.4.48b}$$

Consequently, using the factorization in (3.4.37), the first moment of the relative rotation $\boldsymbol{R}(s',s)$ and of the relative translation $\boldsymbol{r}(s',s)$ can respectively be approximated as

$$\left\langle \boldsymbol{R}(s',s)\right\rangle = \exp\left\{-\boldsymbol{\Omega}(s',s)\right\}\bar{\boldsymbol{R}}(s',s) + \mathrm{o}\left((s-s')\epsilon\right) \tag{3.4.49a}$$

$$\left\langle \boldsymbol{r}(s,s')\right\rangle = \left[\mathrm{Id}_3 - \exp\left\{-\boldsymbol{\Omega}(s',s)\right\}\right]\boldsymbol{\omega}(s',s) + \exp\left\{-\boldsymbol{\Omega}(s',s)\right\}\bar{\boldsymbol{r}}(s',s) + \mathrm{o}\left((s'-s)\epsilon\right). \tag{3.4.49b}$$

for s sufficiently close to s'. The formulas in (3.4.49) comprises the *short length asymptotic behaviour* for continuum rods. At first order, the direct dependence of the expected rod configuration $\bar{\boldsymbol{g}}(s',s)$ on the behaviour of first moment $\langle \boldsymbol{g}_{m,n}\rangle$ is then observed to be exponentially damped at a rate controlled by the symmetric and positive definite matrix

$$\boldsymbol{\Omega}(s',s) = \frac{1}{2}\int_{s'}^{s} \mathrm{Tr}\left(\mathcal{C}^{\boldsymbol{RR}}(t)\right)\mathrm{Id}_3 - \bar{\boldsymbol{R}}(s',t)\mathcal{C}^{\boldsymbol{RR}}(t)\bar{\boldsymbol{R}}(s',t)^{T}dt \tag{3.4.50}$$

according to the definition of $\mathcal{L}^{\boldsymbol{RR}}(s)$ in (3.4.19b).

Moreover, similarly to the case of rigid body chains, the persistence vector $\langle \boldsymbol{r}(s',s)\rangle$ is observed to be made of a convex sum of the vector $\boldsymbol{\omega}(s',s)$ and $\bar{\boldsymbol{r}}(s',s)$, and the vector $\boldsymbol{\omega}(s',s)$ of a weighted average of the translation vector $\bar{\boldsymbol{r}}(s',s)$. In particular the following estimates can be deduced for non-degenerate rods

$$(s-s')\lambda^{\min} \leq \big|\boldsymbol{\Omega}(s',s)\big|_{\mathrm{Sp}} \leq (s-s')\lambda^{\max} \tag{3.4.51a}$$

$$\big|\boldsymbol{\omega}(s',s)\big| \leq \frac{\lambda^{\max}}{\lambda^{\min}} \frac{1}{s-s'} \int_{s'}^{s} \big|\bar{\boldsymbol{r}}(s',t)\big|\, dt + \frac{1}{2}\frac{\big|\mathcal{L}^{\boldsymbol{Rr}}\big|}{\lambda^{\min}}. \tag{3.4.51b}$$

and they lead, for the first moments $\langle \boldsymbol{R}(s',s)\rangle$ and $\langle \boldsymbol{r}(s',s)\rangle$, to

$$\exp\big\{-(s-s')\lambda^{\max}\big\} \leq \big|\big\langle \boldsymbol{R}(s',s)\big\rangle\big|_{\mathrm{Sp}} \leq \exp\big\{-(s-s')\lambda^{\min}\big\} \tag{3.4.52a}$$

$$\begin{aligned}\big|\big\langle \boldsymbol{r}(s',s)\big\rangle\big| \leq (1-\exp\big\{-(s-s')\lambda^{\max}\big\}) \left[\frac{\lambda^{\max}}{\lambda^{\min}} \frac{1}{s-s'} \int_{s'}^{s} \big|\bar{\boldsymbol{r}}(s',t)\big|\, dt + \frac{1}{2}\frac{\big|\mathcal{L}^{\boldsymbol{Rr}}\big|}{\lambda^{\min}}\right] \\ + \exp\big\{-(s-s')\lambda^{\min}\big\}\, \big|\bar{\boldsymbol{r}}(s',s)\big|\end{aligned} \tag{3.4.52b}$$

for s sufficiently close to s'.

In the opposite case, when the arc-length difference $s-s'$ is large, the behaviour of the first moment $\langle \boldsymbol{g}(s',s)\rangle$ is dominated by the asymptotic values given by $|s-s'| \to \infty$. Using the ODE system in (3.4.22), together with the estimates (3.4.26) and (3.4.30), we conclude that for non-degenerate semi-flexible rods

$$\big\langle \boldsymbol{R}(s',s)\big\rangle = 0 + \mathrm{O}\left(\frac{\exp\big\{-(s-s')\lambda^{\min}\big\}}{\lambda^{\min}}\right) \tag{3.4.53a}$$

$$\big\langle \boldsymbol{r}(s',s)\big\rangle = \big\langle \boldsymbol{r}(s',\infty)\big\rangle + \mathrm{O}\left(\frac{\exp\big\{-(s-s')\lambda^{\min}\big\}}{\lambda^{\min}}\right) \tag{3.4.53b}$$

for $s >> s'$. The associated first order correction requires a more detailed analysis and is left as an open topic.

4. Bridging the Scales from Chain to Rod Mechanics

Mathematically, there is some subtlety in going from discrete n to continuous n, but for the present purpose it is sufficient to understand that eqn (...) is a formal rewriting of eqn (...). In this book we use the discrete n and continuous n interchangeably.

M. Doi & S. F Edwards, 1986.

This fourth chapter presents both general principles and explicit methods to bridge the scales from chain to rod mechanics. In section 4.1 we discuss how the original Cauchy-Born rule, introduced to provide continuum modelling of crystal deformations, can be adapted to the context of chains and rods. Inspired by the work of [Arroyo and Belytschko, 2004] on the application of such rules for configuration space with an intrinsically non-flat geometry, we also refer to exponential Cauchy-Born rules, even if our context is different. Local interpolation rules for chains and bichains are presented and lead to the definition of consistent discrete and continuum energies and stationary configurations. We briefly expose why, under specific hypotheses, continuum equilibrium configurations can provide approximations of the discrete equilibrium configurations. The discussion is then extended to the notion of consistent observable and configurational distributions, and it is noticed that in general the definition of a deterministic Cauchy-Born rule does not imply a stochastic Cauchy-Born rule. Section 4.2 presents an explicit definition of a deterministic Cauchy-Born rule between bichains and birods with quadratic energy. This rule is obtained by matching the first and second order terms in an expansion of the bichain energy and provides then only consistency at leading orders. Similarly, we discuss in section 4.3 the construction of an explicit stochastic Cauchy-Born rule. However, this rule is only for a single chain and a single rod which admit normal configurational distributions and which are in a semi-flexible regime. We deduce the rule by matching the first and second moments of the chain internal coordinates.

4.1 On Exponential Cauchy-Born Rules for Rigid Body Chains and Bichains

We discuss in this section general principles in the scale bridging between chain and rod mechanics, and more precisely, in which sense a continuum rod model can be thought of as an approximation of a chain model. Originally the *Cauchy-Born rule* was used to approximate a potential energy for continuum deformable bodies configurations given the analogous all atom energy and configurations, see for instance [E and Ming, 2006, Tadmor and Miller, 2011]. For rodlike objects similar Cauchy-Born rules have been proposed in the context of carbon nanotube modelling [Arroyo and Belytschko, 2004, Yang and E, 2006] or precisely in the context of Cosserat rods [Yang and E, 2006, Hakobyan et al., 2012, Kumar et al., 2015]. In particular in [Arroyo and Belytschko, 2004] they define the notion of *exponential Cauchy-Born rule* which is the direct analogue of the classical Cauchy-Born rule but for a configuration space with an intrinsic non-flat geometry. In the context of continuum rod approximation of rigid body chain mechanics, these Cauchy-Born rules have to be modified since the configuration space of the original discrete system is different than the one appearing in atomic representations. In other words, we have not to work with groups of atoms but directly with rigid bodies. The constructions are also referred to as exponential Cauchy-Born rules, even if they are different from the one in [Arroyo and Belytschko, 2004], since they significantly use the underlying geometric structure of chain and rod configurations. Moreover, we need to discuss separately a rule based on the chain internal energy, which we call a *deterministic exponential Cauchy-Born rule*, and a rule based on the chain configurational distribution, which we call a *stochastic exponential Cauchy-Born rule*. In order to provide a precise discussion, we define chain, and respectively bichain, local interpolation rules. The notion of consistent discrete and continuum energies is then introduced which leads naturally to notion of consistent discrete and continuum stationary configurations. We briefly discuss why, under specific hypotheses, continuum equilibrium configurations can be expected to be close to the equilibrium configurations of the original problem. Similarly, we present the definition of consistent discrete and continuum observables and stationary configurational distributions. We finally conclude that in general deterministic exponential Cauchy-Born rules do not lead to stochastic exponential Cauchy-Born rules.

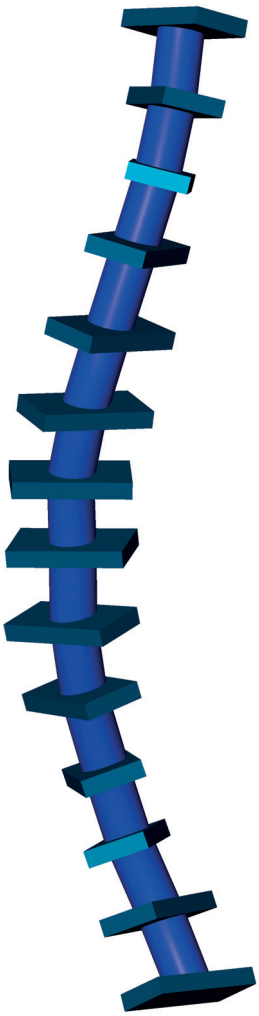


Figure 4.1: Sketch of a chain interpolation

4.1.1 Interpolation rules for chains and bichains

In section (3.3.1), we discussed how the discretisation of a set of continuum rod configurations can naturally be interpreted as a set of rigid body chain configurations, and similarly for discretised birods and bichains. The inverse procedure, based on a given chain or bichain interpolation rule, allows consideration of a set of rigid body chain, or rigid body bichains configurations, as a particular subset of continuum rods, or respectively birod, configurations.

Precisely, for a chain configuration $\boldsymbol{g} = (\boldsymbol{g}_1, ..., \boldsymbol{g}_N)$, as defined in (2.1.1), we introduce a reference length L and a set of N nodes, denoted respectively by ${s_n}^{(N)}$, such that

$$0 = {s_1}^{(N)} < {s_2}^{(N)} < ... < {s_N}^{(N)} = L. \tag{4.1.1}$$

We can then introduce a continuous *local rod interpolation*, denoted by $\boldsymbol{g}^{(N)}(s)$ which is of the form

$$\boldsymbol{g}^{(N)}(s) = \boldsymbol{g}^{(N)}(a; s) \text{ for } s \in]0, L[\tag{4.1.2}$$

where $a = (a_1, ..., a_{N-1})$ denote the set of local junction displacement in the chain associated to the chain configuration $\boldsymbol{g}$, as defined in (2.1.4), and which have the properties that

$$\boldsymbol{g}^{(N)}(a; {s_n}^{(N)}) = \boldsymbol{g}_n \tag{4.1.3}$$

and

$$\boldsymbol{g}^{(N)}(s) = \boldsymbol{g}^{(N)}(a_n; s) \text{ for } {s_n}^{(N)} < s < {s_{n+1}}^{(N)} \tag{4.1.4}$$

for all $n = 1, ..., N$. Note that we have used a similar notation to (3.3.7) and (3.3.8). The *rod interpolation generator vector* $\xi^{(N)}(s)$ is then defined by the relation

$$\frac{d}{ds}\boldsymbol{g}^{(N)}(s) = \boldsymbol{g}^{(N)}(s)\mathcal{T}\xi^{(N)}(s) \tag{4.1.5}$$

and we observe that, according to the assumption (4.1.4), the dependence of the generator vector $\xi^{(N)}(s)$ on the chain configuration $\boldsymbol{g}$ is local in the sense that

$$\xi^{(N)}(s) = \xi^{(N)}(a_n; s) \text{ for } {s_n}^{(N)} < s < {s_{n+1}}^{(N)} \tag{4.1.6}$$

for all $n = 1, ..., N$. Consequently, if we denote by $\Gamma^{(N)}$ the set of rigid body chain configurations satisfying one of the conditions in (2.2.3) and by Γ the set of continuous rod configurations satisfying the analogous condition in (3.2.3), one has formally the embedding

$$\Gamma^{(N)} \subset \Gamma \tag{4.1.7}$$

using the interpolation rule (4.1.2). A sketch of an interpolated rod configuration is drawn in figure 4.1.

For a bichain configuration $(\boldsymbol{g}, \mathcal{P}) = (\boldsymbol{g}_1, \mathcal{P}_1, ..., \boldsymbol{g}_N, \mathcal{P}_N)$ and a mesh of the form (4.1.1) the *local birod interpolation* $(\boldsymbol{g}^{(N)}(s), \mathcal{P}^{(N)}(s))$ is defined similarly to that for single chains, i.e. with

$$\boldsymbol{g}^{(N)}(s) = \boldsymbol{g}^{(N)}(a; s) \text{ and } \mathcal{P}^{(N)}(s) = \mathcal{P}^{(N)}(\mathcal{P}; s) \text{ for } s \in]0, L[\tag{4.1.8}$$

where $\mathcal{P} = (\mathcal{P}_1, ..., \mathcal{P}_N)$ denotes the bichain microstructure configuration, and with the properties that

$$\boldsymbol{g}^{(N)}(a; {s_n}^{(N)}) = \boldsymbol{g}_n \text{ and } \mathcal{P}^{(N)}(\mathcal{P}; {s_n}^{(N)}) = \mathcal{P}_n \tag{4.1.9}$$

and

$$\boldsymbol{g}^{(N)}(s) = \boldsymbol{g}^{(N)}(a_n; s) \text{ and } \mathcal{P}^{(N)}(s) = \mathcal{P}^{(N)}(\mathcal{P}_n, \mathcal{P}_{n+1}; s) \text{ for } {s_n}^{(N)} < s < {s_{n+1}}^{(N)} \tag{4.1.10}$$

for all $n = 1, .., N$. The associated macrostructure and microstructure *birod interpolations generator vectors* are then of the form

$$\frac{d}{ds}\boldsymbol{g}^{(N)}(s) = \boldsymbol{g}^{(N)}(s)\mathcal{T}\xi^{(N)}(s) \text{ and } \frac{d}{ds}\mathcal{P}^{(N)}(s) = \mathcal{T}^{\mathcal{P}}\xi^{\mathcal{P}^{(N)}}(s) \tag{4.1.11}$$

for all $s \in]0, L[$ and with

$$\xi^{(N)}(s) = \xi^{(N)}(a_n; s) \text{ and } \xi^{\mathcal{P}^{(N)}}(s) = \xi^{\mathcal{P}^{(N)}}(\mathcal{P}_n, \mathcal{P}_{n+1}; s) \text{ for } {s_n}^{(N)} < s < {s_{n+1}}^{(N)}. \tag{4.1.12}$$

Denoting, similarly as for single chain configurations, by $\Gamma^{(N)}$ the set of rigid body bichain configurations satisfying one of the conditions in (2.2.42) and by Γ the set of continuous birod configuration satisfying the analogous condition in (3.2.51), one also obtains a formal embedding of the form (4.1.7). A sketch of an interpolated rod configuration is drawn in figure 4.2.

4.1.2 On the consistency of energies and stationary configurations

The interpolations rules presented in the previous section allow a representation of rigid body chains and bichains as well defined continuum objects and this idea leads naturally to the notion of *consistent energy*.

For a given set of interpolated chain configuration $\Gamma^{(N)}$ and a set of continuum configurations Γ satisfying (4.1.7), a continuum rod energy $E^{(N)}$ is said to be consistent with the chain energy E if

$$E^{(N)}[\boldsymbol{g}^{(N)}] = E(\boldsymbol{g}) \text{ for all } \boldsymbol{g}^{(N)} \in \Gamma^{(N)}. \tag{4.1.13}$$

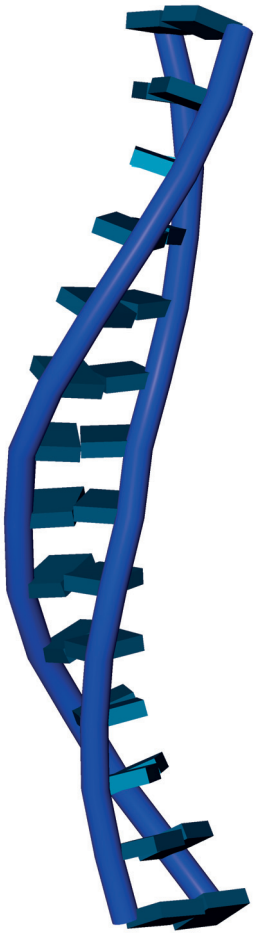


Figure 4.2: Sketch of a bichain interpolation

In the case of chain and rod internal energies defined respectively in (2.2.5) and (3.2.5), the internal energies are consistent if

$$\int_{s_n}^{s_{n+1}} \mathsf{W}(\xi^{(N)}(s); s) = w_n(a_n) \text{ for all } \boldsymbol{g}^{(N)} \in \Gamma^{(N)} \tag{4.1.14}$$

for all nodes s_n in the interpolation mesh (3.3.2), according to the assumption (4.1.6). The explicit construction of a consistent continuum rod energy $E^{(N)}[\boldsymbol{g}^{(N)}]$ for a given interpolation rule of the form in (4.1.6) and given chain energy $E(\boldsymbol{g})$ is said to be a *deterministic exponential Cauchy-Born rule for chains*.

Similarly, a bichain energy E and a continuum birod energy $E^{(N)}$ are then said to be consistent energies if

$$E^{(N)}[\boldsymbol{g}^{(N)}, \mathcal{P}^{(N)}] = E(\boldsymbol{g}, \mathcal{P}) \text{ for all } (\boldsymbol{g}^{(N)}, \mathcal{P}^{(N)}) \in \Gamma^{(N)} \tag{4.1.15}$$

where $\Gamma^{(N)}$ denotes the set of interpolated bichain configurations according to the rules in (4.1.8). When both the bichain and the birod energy are local, as defined respectively in (2.2.32) and (3.2.42), then the internal energies are consistent if

$$\int_{s_n}^{s_{n+1}} \mathsf{W}(\mathcal{P}^{(N)}(s), \xi^{\mathcal{P}^{(N)}}(s), \xi^{(N)}(s); s) = w_n(\mathcal{P}_n, a_n, \mathcal{P}_{n+1}) \text{ for all } (\boldsymbol{g}^{(N)}, \mathcal{P}^{(N)}) \in \Gamma^{(N)} \tag{4.1.16}$$

and for all nodes s_n in the interpolation mesh (3.3.2), according to the assumptions (4.1.10) and (4.1.12). The construction of a consistent continuum birod energy $E^{(N)}[\boldsymbol{g}^{(N)}, \mathcal{P}^{(N)}]$ for a given bichain energy $E(\boldsymbol{g}, \mathcal{P})$ is called a *deterministic exponential Cauchy-Born rule for bichains*.

A fundamental question arising from the definition of discrete and continuum consistent energies, either of the form (4.1.13) or (4.1.15), is whether their respective stationary configuration, in the sense of (2.2.7) and (3.2.7) for instance, or even minimum energy configuration have any chance to be consistent, in the sense that one is the exact interpolation of the other. Of course, the answer is in general no, since the definition of consistency only involves information on the subset of interpolations, a detailed analysis is required in particular cases to prove if that happens or not, see for instance the inspiring papers [Friesecke and Theil, 2002, E and Ming, 2006] in the context of the Cauchy-Born rule for crystals.

However, the formulations of the bichain and birod equilibrium conditions in terms of Lagrange multipliers, presented respectively in section 2.2.6 and 3.2.9, allows us to state an estimate of the difference between bichain and birod stationary configurations, which, even if it has to be regarded as very coarse, is instructive.

In the case of bichains, we assume that there exists a unique configuration denoted by $(\hat{\boldsymbol{g}}, \hat{\mathcal{P}})$ which makes the internal energy E stationary with respect to unstressed boundary conditions presented in (2.2.40b). The configuration $(\hat{\boldsymbol{g}}, \hat{\mathcal{P}})$ is then called the *bichain ground state configuration*. Using the expressions in (2.2.53) we deduce that consequently the configuration $(\hat{\boldsymbol{g}}, \hat{\mathcal{P}})$ makes the Lagrangian function $\mathcal{L}$, defined in (2.2.55), stationary with the Lagrange multipliers all vanishing identically, i.e. $\boldsymbol{\lambda} = 0$, for the Dirichlet

conditions $(\boldsymbol{g}_1{}^*, \mathcal{P}_1{}^*) = (\hat{\boldsymbol{g}}_1, \hat{\mathcal{P}}_1)$ and $(\boldsymbol{g}_N{}^*, \mathcal{P}_N{}^*) = (\hat{\boldsymbol{g}}_N, \hat{\mathcal{P}}_N)$. Moreover, we shall assume that there exists a value $\lambda_c > 0$ such that for all $\boldsymbol{\lambda}$ with $|\boldsymbol{\lambda}| < \lambda_c$ one can find Dirichlet conditions $(\boldsymbol{g}_1{}^*, \mathcal{P}_1{}^*)(\boldsymbol{\lambda})$ and $(\boldsymbol{g}_N{}^*, \mathcal{P}_N{}^*)(\boldsymbol{\lambda})$ such that there exists a unique bichain configuration solution of the stationary conditions (2.2.35) or, equivalently, of (2.2.53). Formally, one obtains a mapping

$$\boldsymbol{\lambda} \mapsto (\boldsymbol{g}, \mathcal{P})(\boldsymbol{\lambda}) \text{ for all } \boldsymbol{\lambda} \text{ with } |\boldsymbol{\lambda}| < \lambda_c \tag{4.1.17a}$$

with

$$(\boldsymbol{g}, \mathcal{P})(0) = (\hat{\boldsymbol{g}}, \hat{\mathcal{P}}) \tag{4.1.17b}$$

which satisfies by definition

$$\delta\mathcal{L}\left((\boldsymbol{g}, \mathcal{P})(\boldsymbol{\lambda}), \boldsymbol{\lambda}\right)(\delta\boldsymbol{g}, \delta\mathcal{P}, \delta\boldsymbol{\lambda}) = 0 \text{ for all } \boldsymbol{\lambda} \text{ with } |\boldsymbol{\lambda}| < \lambda_c \tag{4.1.17c}$$

and for any perturbation $(\delta\boldsymbol{g}, \delta\mathcal{P}, \delta\boldsymbol{\lambda})$.

Similarly, if there exists a unique birod configuration which makes stationary the internal birod energy with respect to unstressed boundary conditions presented in (2.2.40b), it is called the *birod ground state configuration*. For a birod internal energy $E^{(N)}$ consistent with the bichain energy E in the sense of (4.1.15), one can show that the stationary birod configuration $(\boldsymbol{g}, \mathcal{P})$ respecting the Dirichlet conditions $(\boldsymbol{g}(0), \mathcal{P}(0)) = (\boldsymbol{g}_1{}^*, \mathcal{P}_1{}^*)$ and $(\boldsymbol{g}(L), \mathcal{P}(L)) = (\boldsymbol{g}_N{}^*, \mathcal{P}_N{}^*)$ are unconstrained stationary birod configurations of a functional of the form

$$\mathcal{L}^{(N)}[\boldsymbol{g}, \mathcal{P}, \boldsymbol{\lambda}] = E^{(N)}[\boldsymbol{g}, \mathcal{P}] - \boldsymbol{\lambda} \cdot \Phi\left(\mathcal{P}(0), \boldsymbol{g}(L), \mathcal{P}(L); \mathcal{P}_1{}^*, \boldsymbol{g}_N{}^*, \mathcal{P}_N{}^*\right) \tag{4.1.18}$$

according to (3.2.80). Assuming that the continuum energy $E^{(N)}$ admits an unstressed configuration and, moreover, that it is given by the interpolation in the sense of (4.1.9) of the bichain unstressed configuration in (4.1.17b), i.e. by $(\hat{\boldsymbol{g}}^{(N)}, \hat{\mathcal{P}}^{(N)})$, then it makes the functional $\mathcal{L}^{(N)}$ defined in (4.1.18) stationary with the Lagrange multipliers all identically vanishing, i.e. $\boldsymbol{\lambda} = 0$.

Furthermore, analogously to (4.1.17), we make the hypothesis that for all $\boldsymbol{\lambda}$ with $|\boldsymbol{\lambda}| < \lambda_c$ there exist Dirichlet conditions $(\boldsymbol{g}_0{}^*, \mathcal{P}_0{}^*)(\boldsymbol{\lambda})$ and $(\boldsymbol{g}_L{}^*, \mathcal{P}_L{}^*)(\boldsymbol{\lambda})$ such the stationary conditions (3.2.45) admit a unique birod configuration solution. Formally, this can be expressed by the mapping

$$\boldsymbol{\lambda} \mapsto (\boldsymbol{g}, \mathcal{P})[\boldsymbol{\lambda}] \text{ for all } \boldsymbol{\lambda} \text{ with } |\boldsymbol{\lambda}| < \lambda_c \tag{4.1.19a}$$

with

$$(\boldsymbol{g}, \mathcal{P})[0] = (\hat{\boldsymbol{g}}, \hat{\mathcal{P}}) \tag{4.1.19b}$$

and which satisfies by definition

$$\delta\mathcal{L}^{(N)}\left[(\boldsymbol{g}, \mathcal{P})[\boldsymbol{\lambda}], \boldsymbol{\lambda}\right](\delta\boldsymbol{g}, \delta\mathcal{P}, \delta\boldsymbol{\lambda}) = 0 \text{ for all } \boldsymbol{\lambda} \text{ with } |\boldsymbol{\lambda}| < \lambda_c \tag{4.1.19c}$$

and for any perturbation $(\delta \boldsymbol{g}, \delta \mathcal{P}, \delta \boldsymbol{\lambda})$.

In such circumstances, for given Dirichlet conditions

$$(\boldsymbol{g}_1{}^*, \mathcal{P}_1{}^*) \text{ and } (\boldsymbol{g}_N{}^*, \mathcal{P}_N{}^*) \tag{4.1.20}$$

and for a given bichain energy E and an associated consistent birod energy $E^{(N)}$ satisfying (4.1.17) and (4.1.19) respectively, we denote by $(\boldsymbol{g}, \mathcal{P})(\boldsymbol{\lambda}_1)$ and $(\boldsymbol{g}, \mathcal{P})[\boldsymbol{\lambda}_2]$ the bichain and birod configuration which satisfies (4.1.17c) and (4.1.19c) respectively. Then, using expansions around $\boldsymbol{\lambda} = 0$, one deduces the following estimates

$$\left| E\left((\boldsymbol{g}, \mathcal{P})(\boldsymbol{\lambda}_1), \boldsymbol{\lambda}_1\right) - E^{(N)}\left[(\boldsymbol{g}, \mathcal{P})[\boldsymbol{\lambda}_2], \boldsymbol{\lambda}_2\right] \right| = \mathrm{O}\left(\lambda_c^2\right) \tag{4.1.21a}$$

$$\left| (\boldsymbol{g}, \mathcal{P})(\boldsymbol{\lambda}_1)) - (\boldsymbol{g}, \mathcal{P})[\boldsymbol{\lambda}_2] \right| = \mathrm{O}\left(\lambda_c\right) \tag{4.1.21b}$$

if the expressions in (4.1.17a) and (4.1.19a) are smooth enough. The optimality of these approximations as well as the statement of finer estimates are left as open topics.

4.1.3 On the consistency of observables and configurational distributions

Another natural question which appears in the definition of continuum models as a representation of discrete systems is comparing their statistical physics properties. To assess this point we need to introduce the notion of *consistent observable* and the notion of *consistent configurational distribution*.

Let the rigid body chain configuration set $\Gamma^{(N)}$ and a set of continuous rod interpolations Γ satisfying (4.1.7), where both stand for unstressed configurations, as defined in (2.2.2b) and (3.2.2b). We say then that the continuum rod observable $\mathcal{A}^{(N)}$ is consistent with the chain observable $\mathcal{A}$ if

$$\mathcal{A}^{(N)}[\boldsymbol{g}^{(N)}] = \mathcal{A}(\boldsymbol{g}) \text{ for all } \boldsymbol{g}^{(N)} \in \Gamma^{(N)}. \tag{4.1.22}$$

Similarly, a birod observable $\mathcal{A}^{(N)}$ and a bichain observable $\mathcal{A}$ are consistent if

$$\mathcal{A}^{(N)}[\boldsymbol{g}^{(N)}, \mathcal{P}^{(N)}] = \mathcal{A}(\boldsymbol{g}, \mathcal{P}) \text{ for all } (\boldsymbol{g}^{(N)}, \mathcal{P}^{(N)}) \in \Gamma^{(N)} \tag{4.1.23}$$

where the set $\Gamma^{(N)}$ denotes unstressed configurations, as defined in (2.2.40b) and (3.2.50b). We note that the definition of consistent observables (4.1.22) is analogous to the one made for consistent internal energies in (4.1.13) and (4.1.15).

A chain configurational distribution $d\rho(\boldsymbol{g})$ and a rod configurational distribution $d\rho^{(N)}[\boldsymbol{g}]$ are said to be consistent if the expectation value of any consistent observable $\mathcal{A}$ and $\mathcal{A}^{(N)}$ coincides, i.e. if

$$\left\langle \mathcal{A}^{(N)}[\boldsymbol{g}^{(N)}] \right\rangle^{(N)} = \langle \mathcal{A}(\boldsymbol{g}) \rangle \tag{4.1.24}$$

where $\langle\,.\,\rangle^{(N)}$ and $\langle\,.\,\rangle$ stand for the expectation with respect to the rod distribution $d\rho^{(N)}[\boldsymbol{g}]$ and the chain distribution $d\rho(\boldsymbol{g})$, as defined in (2.3.10) and in (3.3.35) respectively. The construction of a consistent continuum rod configuration distribution $d\rho^{(N)}[\boldsymbol{g}]$ for a given chain interpolation rule and a given chain distribution $d\rho(\boldsymbol{g})$ is said to be a *stochastic exponential Cauchy-Born rule for chains*. Similarly, a bichain distribution $d\rho(\boldsymbol{g},\mathcal{P})$ and a birod distribution $d\rho^{(N)}[\boldsymbol{g},\mathcal{P}]$ are said to be consistent if

$$\left\langle \mathcal{A}^{(N)}[\boldsymbol{g}^{(N)},\mathcal{P}^{(N)}]\right\rangle^{(N)} = \langle \mathcal{A}(\boldsymbol{g},\mathcal{P})\rangle \tag{4.1.25}$$

for any consistent bichain observables $\mathcal{A}$ and birod observables $\mathcal{A}^{(N)}$. The construction of a birod distribution $d\rho^{(N)}[\boldsymbol{g},\mathcal{P}]$ satisfying (4.1.25) for a given bichain distribution and a given interpolation rule is called a *stochastic exponential Cauchy-Born rule for bichains*.

We would like to observe, without providing any rigorous examples, that in general deterministic Cauchy-Born rules, in the sense of (4.1.13) or (4.1.15), do not provide stochastic Cauchy-Born rules, in the sense of (4.1.24) or (4.1.25). This statement is motivated by the idea that deterministic Cauchy-Born rules are consistent on the set of chain, or respectively bichain, interpolations and this condition is not strong enough to imply the consistency of the associated configurational distributions. The required mathematical analysis goes beyond the present thesis and is left as an open topic.

4.2 A Deterministic Exponential Cauchy-Born Rule for Quadratic Bichain Internal energies

The general idea of deterministic Cauchy-Born rules in the context of chains has been discussed in section (4.1.2). We now would like to concentrate on the particular cases of quadratic internal energies. As anticipated, we are mainly interested in the application of chain and rod theory in the context of sequence dependent DNA modelling. In particular, a *quadratic internal bichain energy* of the form

$$E(\boldsymbol{g}(x),\mathcal{P}(y)) = \sum_{n=1}^{N-1} \frac{1}{2} \begin{pmatrix} y_n - \hat{y}_n^{\alpha\beta} \\ x_n - \hat{x}_n^{\alpha\beta} \\ y_{n+1} - \hat{y}_{n+1}^{\alpha\beta} \end{pmatrix} \cdot K_n^{\alpha\beta} \begin{pmatrix} y_n - \hat{y}_n^{\alpha\beta} \\ x_n - \hat{x}_n^{\alpha\beta} \\ y_{n+1} - \hat{y}_{n+1}^{\alpha\beta} \end{pmatrix} + \sum_{n=1}^{N} \frac{1}{2} (y_n - \hat{y}_n^{\alpha}) \cdot K_n^{\alpha} (y_n - \hat{y}_n^{\alpha}) \tag{4.2.1}$$

has been proposed in [Petkeviciute, 2012, Gonzalez et al., 2013, Petkeviciute et al., 2014] where the bichain internal coordinates (x, y) are defined in (2.1.31) and where the subscripts $\alpha\beta$ and α refer to the physical constitution of the bichain. This model corresponds to a nearest neighbor quadratic rigid base energy. Precisely we have $\hat{x}_n^{\alpha\beta}, \hat{y}_n^{\alpha\beta} \in \mathbb{R}^6$, $K_n^{\alpha\beta} \in \mathbb{R}^{18\times 18}$ and $K_n^{\alpha} \in \mathbb{R}^{6\times 6}$. This quadratic form can then be written as

$$E(\boldsymbol{g}(x),\mathcal{P}(y)) = \frac{1}{2} \begin{pmatrix} y_1 - \hat{y}_1 \\ x_1 - \hat{x}_1 \\ ... \\ x_{N-1} - \hat{x}_{N-1} \\ y_N - \hat{y}_N \end{pmatrix} \cdot K \begin{pmatrix} y_1 - \hat{y}_1 \\ x_1 - \hat{x}_1 \\ ... \\ x_{N-1} - \hat{x}_{N-1} \\ y_N - \hat{y}_N \end{pmatrix} \tag{4.2.2}$$

where $K \in \mathbb{R}^{(12N-6)\times(12N-6)}$ is called the bichain stiffness matrix and $(\hat{x}, \hat{y}) \in \mathbb{R}^{12N-6}$ the bichain ground state internal coordinates. We underline in particular that the entries $(\hat{x}_n, \hat{y}_n)$ do not correspond to $(\hat{x}_n^{\alpha\beta}, \hat{y}_n^{\alpha\beta})$, since one needs to *complete the square* to obtain (4.2.2) from (4.2.1), which is a non-local computation. More details on the rigid base model are provided in section 5.2.2. We observe that the bichain configuration $(\hat{\boldsymbol{g}}, \hat{\mathcal{P}})$, defined by the internal coordinates $(\hat{x}, \hat{y})$, is a called ground state since it is the global minimiser of the function $E(\boldsymbol{g}(x), \mathcal{P}(y))$. The construction of a deterministic Cauchy-Born rule for the rigid base model is of interest since it allows the approximation of stationary configurations of the energy $E(\boldsymbol{g}(x), \mathcal{P}(y))$ for prescribed bichain end conditions, such as in (4.1.20) for instance, using a continuum birod model. Numerical examples are discussed in the context of sequence dependent DNA modelling in chapter 6.

According to the discussion made in section (4.1.2), the strategy is to define a *quadratic internal birod energy* $\hat{E}^{(N)}[\boldsymbol{g}, \mathcal{P}(y)]$ which has the property to be consistent for a given interpolation rule, in the sense of (4.1.15), to the quadratic bichain energy

$E(\boldsymbol{g}(x), \mathcal{P}(y))$ in (4.2.2). Precisely, we assume a birod energy of the form

$$\begin{aligned}\hat{E}^{(N)}[\boldsymbol{g}, \mathcal{P}(y)] = \int_0^L \frac{1}{2} \begin{pmatrix} y(s) - \hat{y}(s) \\ \frac{d}{ds} y(s) - \frac{d}{ds} \hat{y}(s) \\ \xi(s) - \hat{\xi}(s) \end{pmatrix} \cdot \mathcal{K}(s) \begin{pmatrix} y(s) - \hat{y}(s) \\ \frac{d}{ds} y(s) - \frac{d}{ds} \hat{y}(s) \\ \xi(s) - \hat{\xi}(s) \end{pmatrix} ds \\ + \frac{1}{2} \left(y(0) - \hat{y}(0)\right) \cdot \mathcal{K}_0 \left(y(0) - \hat{y}(0)\right) + \frac{1}{2} \left(y(L) - \hat{y}(L)\right) \cdot \mathcal{K}_L \left(y(L) - \hat{y}(L)\right)\end{aligned} \tag{4.2.3}$$

where the macrostructure generator vector $\xi(s)$ and the microstructure internal coordinates $y(s)$ are presented in (3.2.57), $\mathcal{K}(s) \in \mathbb{R}^{18\times 18}$ is the interior birod stiffness matrix, $\mathcal{K}_0 \in \mathbb{R}^{6\times 6}$ and $\mathcal{K}_L \in \mathbb{R}^{6\times 6}$ are boundary intra stiffness matrices, and $(\hat{\xi}(s), \hat{y}(s))$ the birod ground state internal variables, similarly to bichains.

Before presenting how to construct a birod internal energy of the form (4.2.3) given a quadratic bichain energy of the form (4.2.1), we need to remark that one cannot ask simultaneously to define an exact consistent energy and that both discrete and continuum energy formulations are quadratic. The reason is that the dependence of the birod macrostructure generator $\xi(s)$ vector on the bichain macrostructure coordinates x is non-linear. However, to pursue the goal of parametrizing a continuum birod model with a realistic bichain model we will conserve the bichain coordinates used in [Gonzalez et al., 2013], and, to achieve all the future necessary analytical computations, we will also conserve the quadratic birod energy formulation (4.2.3). The compromise is then to only look for an approximate consistency of the discrete and continuum energies. More precisely we will only ask for consistency at the leading orders through quadratic terms.

In order to show how such a continuum energy can be defined, we first need to introduce a specific birod interpolation rule for bichain configurations. Let $\hat{\boldsymbol{g}} = (\hat{\boldsymbol{g}}_1, ..., \hat{\boldsymbol{g}}_N)$ be the macrostructure ground state associated to the internal coordinates $(\hat{x}_1, ..., \hat{x}_{N-1})$ in (4.2.2), the *reference length* L is defined as the sum of all the *junction lengths* ${h_n}^{(N)}$ which are chosen to be exactly the lengths of each local helical interpolation between the rigid body configuration $\hat{\boldsymbol{g}}_n$ and $\hat{\boldsymbol{g}}_{n+1}$. The interpolation of the macrostructure configuration $\hat{\boldsymbol{g}}$ is then defined as a *piecewise helical interpolation* $\hat{\boldsymbol{g}}^{(N)}(s)$. Formally, the reference length L together with the nodes ${s_n}^{(N)} \in]0, L[$, and the interpolation $\hat{\boldsymbol{g}}^{(N)}(s)$ are defined by the relations

$${s_1}^{(N)} = 0,\ {s_{n+1}}^{(N)} = {s_n}^{(N)} + {h_n}^{(N)} \text{ and } {s_N}^{(N)} = L \text{ for all } n = 1, ..., N-1 \tag{4.2.4a}$$

$$\frac{d}{ds}\hat{\boldsymbol{g}}^{(N)}(s) = \hat{\boldsymbol{g}}^{(N)}(s)\mathcal{T}\hat{\xi}^{(N)}(s) \text{ for all } s \in]0, L[\tag{4.2.4b}$$

$$\hat{\boldsymbol{g}}^{(N)}({s_n}^{(N)}) = \hat{\boldsymbol{g}}_n \text{ for all } n = 1, ..., N \tag{4.2.4c}$$

$$\mathcal{T}\hat{\xi}^{(N)}(s) = \frac{1}{{h_n}^{(N)}} \ln\left\{a(\hat{x}_n)\right\} \text{ for } {s_n}^{(N)} < s < {s_{n+1}}^{(N)} \tag{4.2.4d}$$

$$\left|\hat{V}^{(N)}(s)\right| = 1 \text{ for all } s \in]0, L[\tag{4.2.4e}$$

where we recall that $a(x_n) = \boldsymbol{g}_n^{-1}\boldsymbol{g}_{n+1}$ denotes the local junction displacement and $\hat{V}^{(N)}(s) \in \mathbb{R}^3$ the translational part of the macrostructure generator vector $\hat{\xi}^{(N)}(s)$, as

presented in (3.1.5). Observe that a piecewise helical interpolation of chain configuration is characterised by the fact the generator vector $\hat{\xi}^{(N)}(s)$ is piece wise constant on the interval $]0, L[$. A tempting choice would be to define the interpolation of any bichain macrostructure configuration $\boldsymbol{g}$ as piecewise helices, similarly to the ground state macrostructure configuration $\hat{\boldsymbol{g}}$. However this choice is not the most convenient to use and, in contrast, we propose to make piecewise helical interpolations of the deformations from the ground state instead of the actual configurations. This leads to the definition of the interpolation $\boldsymbol{g}^{(N)}(s)$ of the bichain macrostructure configuration $\boldsymbol{g} = (\boldsymbol{g}_1, ..., \boldsymbol{g}_N)$ by

$$\frac{d}{ds}\boldsymbol{g}^{(N)}(s) = \boldsymbol{g}^{(N)}(s)\mathcal{T}\xi^{(N)}(s) \text{ for all } s \in]0, L[\tag{4.2.5a}$$

with

$$\mathcal{T}\xi^{(N)}(s) = \mathcal{T}\hat{\xi}^{(N)}(s) + \frac{1}{{h_n}^{(N)}} \left[\hat{\boldsymbol{g}}_n^{-1}\hat{\boldsymbol{g}}^{(N)}(s)\right]^{-1} \ln\left\{a(x_n)a^{-1}(\hat{x}_n)\right\} \left[\hat{\boldsymbol{g}}_n^{-1}\hat{\boldsymbol{g}}^{(N)}(s)\right] \tag{4.2.5b}$$

for ${s_n}^{(N)} < s < {s_{n+1}}^{(N)}$, where we recall that ${h_n}^{(N)} = {s_{n+1}}^{(N)} - {s_n}^{(N)}$ denotes the local junction length. The main interest in this choice of interpolation is that it directly allows to deduce the following expansion for $s \in]{s_n}^{(N)}, {s_{n+1}}^{(N)}[$

$$\xi^{(N)}(s) = \hat{\xi}^{(N)}(s) + \frac{1}{{h_n}^{(N)}} \mathrm{Ad}^{-1}_{\hat{\boldsymbol{g}}_n^{-1}\hat{\boldsymbol{g}}^{(N)}(s)} \mathbb{L}^{(1)}_{\hat{x}_n} (x_n - \hat{x}_n) + \mathrm{O}\left(|x_n - \hat{x}_n|^2\right) \tag{4.2.6}$$

according to the identity[1] in (2.2.18), and where the operator Ad and the matrix $\mathbb{L}^{(1)}_{\hat{x}_n}$ are defined in (2.2.14) and in (2.2.18) respectively. Note that the matrix $\mathbb{L}^{(1)}_{\hat{x}_n}$ is explicitly given in (2.2.20) for the internal coordinates x_n used in [Petkeviciute, 2012, Gonzalez et al., 2013, Petkeviciute et al., 2014].

In contrast, it is possible to use a *piecewise linear interpolation* for internal bichain microstructure coordinates $y = (y_1, ..., y_N)$ since the internal variables for the birod microstructure, defined in (3.1.28), are chosen consistently with the bichain internal variables defined in (2.1.33). Precisely, we define the interpolation $y^{(N)}(s)$ using the rule

$$y^{(N)}(s) = y_n + \frac{s - {s_n}^{(N)}}{{h_n}^{(N)}}(y_{n+1} - y_n) \text{ for } s \in]{s_n}^{(N)}, {s_{n+1}}^{(N)}[\tag{4.2.7}$$

which gives

$$\begin{pmatrix} y^{(N)}(s) \\ \frac{d}{ds}y^{(N)}(s) \end{pmatrix} = \begin{pmatrix} (1 - \frac{s-{s_n}^{(N)}}{{h_n}^{(N)}})\mathrm{Id}_6 & \frac{s-{s_n}^{(N)}}{{h_n}^{(N)}}\mathrm{Id}_6 \\ -\frac{1}{{h_n}^{(N)}}\mathrm{Id}_6 & \frac{1}{{h_n}^{(N)}}\mathrm{Id}_6 \end{pmatrix} \begin{pmatrix} y_n \\ y_{n+1} \end{pmatrix} \text{ for } s \in]{s_n}^{(N)}, {s_{n+1}}^{(N)}[. \tag{4.2.8}$$

The interpolated microstructure variables $\hat{y}^{(N)}(s)$ are defined using the rule in (4.2.7).

[1] See section 1.3 for more detail on matrix expansion with respect to coordinate systems.

We then have the following expansion of the birod variables around the bichain ground state $(\hat{x},\hat{y})$ for $s\in]{s_n}^{(N)},{s_{n+1}}^{(N)}[$

$$\begin{pmatrix} y^{(N)}(s)-\hat{y}^{(N)}(s)\\ \frac{d}{ds}y^{(N)}(s)-\frac{d}{ds}\hat{y}^{(N)}(s)\\ \xi^{(N)}(s)-\hat{\xi}^{(N)}(s)\end{pmatrix}=\mathbf{L}_{\hat{x}_n}(s)\begin{pmatrix} y_n-\hat{y}_n\\ x_n-\hat{x}_n\\ y_{n+1}-\hat{y}_{n+1}\end{pmatrix}+\mathrm{O}\left(|x_n-\hat{x}_n|^2\right) \tag{4.2.9}$$

with

$$\mathbf{L}_{\hat{x}_n}(s)=\begin{pmatrix}(1-\frac{s-{s_n}^{(N)}}{{h_n}^{(N)}})\mathrm{Id}_6 & 0 & \frac{s-{s_n}^{(N)}}{{h_n}^{(N)}}\mathrm{Id}_6\\ -\frac{1}{{h_n}^{(N)}}\mathrm{Id}_6 & 0 & \frac{1}{{h_n}^{(N)}}\mathrm{Id}_6\\ 0 & \frac{1}{{h_n}^{(N)}}\mathrm{Ad}^{-1}_{\hat{\boldsymbol{g}}_n^{-1}\hat{\boldsymbol{g}}^{(N)}(s)}\mathbb{L}^{(1)}_{\hat{x}_n} & 0\end{pmatrix} \tag{4.2.10}$$

according to (4.2.6) and (4.2.8). Consequently, defining the birod ground state internal variables $(\hat{\xi}(s),\hat{y}(s))$ in (4.2.3) to be the interpolation of the bichain ground state using the rules proposed in (4.2.5) and in (4.2.7), i.e. by

$$\hat{\xi}(s)=\hat{\xi}^{(N)}(s)\text{ and }\hat{y}(s)=\hat{y}^{(N)}(s) \tag{4.2.11}$$

and using the local expansion in (4.2.10), we have that the quadratic bichain and birod energy satisfy for all the bichain interpolated configurations $(\boldsymbol{g}^{(N)}(s),y^{(N)}(s))$

$$E(\boldsymbol{g}(x),\mathcal{P}(y))=\hat{E}^{(N)}[\boldsymbol{g}^{(N)},\mathcal{P}(y^{(N)})]+\mathrm{O}\left(N\,|x-\hat{x}|^3\right) \tag{4.2.12}$$

where $|x-\hat{x}|=\max_{1\leq n\leq N-1}|x_n-\hat{x}_n|$, if

$$\begin{aligned}&\sum_{n=1}^{N-1}\frac{1}{2}\begin{pmatrix} y_n-\hat{y}_n^{\alpha\beta}\\ x_n-\hat{x}_n^{\alpha\beta}\\ y_{n+1}-\hat{y}_{n+1}^{\alpha\beta}\end{pmatrix}\cdot K_n^{\alpha\beta}\begin{pmatrix} y_n-\hat{y}_n^{\alpha\beta}\\ x_n-\hat{x}_n^{\alpha\beta}\\ y_{n+1}-\hat{y}_{n+1}^{\alpha\beta}\end{pmatrix}+\sum_{n=1}^{N}\frac{1}{2}(y_n-\hat{y}_n^{\alpha})\cdot K_n^{\alpha}(y_n-\hat{y}_n^{\alpha})\\ &=\sum_{n=1}^{N-1}\frac{1}{2}\begin{pmatrix} y_n-\hat{y}_n\\ x_n-\hat{x}_n\\ y_{n+1}-\hat{y}_{n+1}\end{pmatrix}\cdot\left[\int_{{s_n}^{(N)}}^{{s_{n+1}}^{(N)}}\mathbf{L}^T_{\hat{x}_n}(s)\mathcal{K}(s)\mathbf{L}_{\hat{x}_n}(s)ds\right]\begin{pmatrix} y_n-\hat{y}_n\\ x_n-\hat{x}_n\\ y_{n+1}-\hat{y}_{n+1}\end{pmatrix}\\ &\qquad+\frac{1}{2}\left(y_1-\hat{y}_1\right)\cdot\mathcal{K}_0\left(y_1-\hat{y}_1\right)+\frac{1}{2}\left(y_N-\hat{y}_N\right)\cdot\mathcal{K}_L\left(y_N-\hat{y}_N\right)\end{aligned} \tag{4.2.13}$$

for all internal bichain coordinates (x,y). Comparing the pure quadratic terms we deduce that one possible definition of the birod stiffness matrix $\mathcal{K}(s)$, and of the two boundary intra stiffness matrices $\mathcal{K}_0$ and $\mathcal{K}_L$ is then

$$\mathcal{K}(s)=\frac{1}{{h_n}^{(N)}}\mathbf{L}^{-T}_{\hat{x}_n}(s)\left[K_n^{\alpha\beta}+\frac{1}{2}\begin{pmatrix}K_n^{\alpha}&0&0\\0&0&0\\0&0&K_{n+1}^{\beta}\end{pmatrix}\right]\mathbf{L}^{-1}_{\hat{x}_n}(s) \tag{4.2.14a}$$

for $s \in]{s_n}^{(N)}, {s_{n+1}}^{(N)}[$ and

$$\mathcal{K}_0 = \frac{1}{2} K_1^\alpha \text{ and } \mathcal{K}_L = \frac{1}{2} K_N^\alpha \tag{4.2.14b}$$

in order to ensure the relation in (4.2.12). We should remark that the weight $\frac{1}{2}$ in the expression (4.2.14a) splits the local contribution of each quadratic form associated to the matrices K_n^α in (4.2.13) into equal contributions to the birod energy in the interval $]{s_{n-1}}^{(N)}, {s_n}^{(N)}[$ and $]{s_n}^{(N)}, {s_{n+1}}^{(N)}[$ respectively. Any unequal splitting also provides a consistent quadratic birod model of the form (4.2.12), but none of them respect a simple transformation rule under the reading symmetry presented in (3.1.38). Moreover we observe that any matrix of the form $\mathcal{K}(s) + \mathcal{K}'(s)$ also satisfy the condition in (4.2.13) if $\int_{{s_n}^{(N)}}^{{s_{n+1}}^{(N)}} \mathbf{L}_{\hat{x}_n}^T(s) \mathcal{K}'(s) \mathbf{L}_{\hat{x}_n}(s) ds = 0$. The complete description of the freedom in the definition of the continuum birod parameters need require more investigations.

The *deterministic exponential Cauchy-Born rule for bichains* is then defined by the relations in (4.2.11) and in (4.2.14) and has been deduced by matching the quadratic terms in an expansion of the bichain energy around its ground state. Numerical applications of this construction are presented in chapter 6 in the context of sequence dependent DNA modelling.

Finally, we would like to stress that the above result admits as a particular case a *deterministic exponential Cauchy-Born rule for chains* which appears naturally when the macrostructure and microstructure internal coordinates x and y are not coupled by the quadratic form in (4.2.2). For an internal chain energy of the form

$$E(\boldsymbol{g}) = \sum_{n=1}^{N-1} \frac{1}{2} (x_n - \hat{x}_n) \cdot K_n (x_n - \hat{x}_n) \tag{4.2.15}$$

one obtains a rod internal energy of the form

$$\hat{E}^{(N)}[\boldsymbol{g}] = \int_0^L \frac{1}{2} \left(\xi(s) - \hat{\xi}^{(N)}(s) \right) \cdot \mathcal{K}(s) \left(\xi(s) - \hat{\xi}^{(N)}(s) \right) ds \tag{4.2.16}$$

where the reference length L and the generator vector $\hat{\xi}^{(N)}(s)$ are defined in (4.2.4) and where

$$\mathcal{K}(s) = {h_n}^{(N)} \left(\mathrm{Ad}^{-1}_{\hat{\boldsymbol{g}}_n^{-1} \hat{\boldsymbol{g}}^{(N)}(s)} \mathbb{L}^{(1)}_{\hat{x}_n} \right)^{-T} K_n \left(\mathrm{Ad}^{-1}_{\hat{\boldsymbol{g}}_n^{-1} \hat{\boldsymbol{g}}^{(N)}(s)} \mathbb{L}^{(1)}_{\hat{x}_n} \right)^{-1} \tag{4.2.17}$$

for $s \in]s_n, s_{n+1}[$, according to the expression in (4.2.14).

4.3 A Stochastic Exponential Cauchy-Born Rule for Normal Configurational Chain Distributions

As discussed in section (4.1.3), stochastic Cauchy-Born rules are designed to deliver continuum models which respect the statistical properties of a given discrete stochastic physical model and deterministic Cauchy-Born rules do not always possess this feature. In contrast to the discussion made for quadratic bichain energies, we present here an explicit construction for normal configurational distributions. However, the definition of a proper stochastic Cauchy-Born rule for bichain configurational distributions is still under investigation and we shall, therefore, restrict the discussion here to the case of a single chain in a semi-flexible regime.

Having still DNA modelling in mind, we consider a *chain normal configurational distribution*

$$d\rho(\boldsymbol{g})(x) = \prod_{n=1}^{N} \sqrt{\det\left(\frac{1}{2\pi}K_n\right)} \exp\left\{-\frac{1}{2}(x_n - \hat{x}_n)\cdot K_n(x_n - \hat{x}_n)\right\} dx_n \tag{4.3.1}$$

where the matrix $K_n \in \mathbb{R}^{6\times 6}$ and the vector $\hat{x}_n \in \mathbb{R}^6$ appear in (4.2.15). This model refers typically to quadratic rigid base pair models in the context of DNA modelling, as presented in section 5.2.1. The goal here is to use a stochastic Cauchy-Born rule to design a normal distribution for continuum rod configurations consistent with the bichain distribution $d\rho(\boldsymbol{g})(x)$, in the sense of (4.1.24). Numerical examples of this construction are presented in chapter 7. They illustrate its application to the study of the configurational first moment in the context of DNA modelling. We say that $d\rho^{(N)}\,[\boldsymbol{g}[\xi]]$ is a *rod normal configurational distribution* if it is of the form

$$d\rho^{(N)}\,[\boldsymbol{g}[\xi]] \sim \exp\left\{-\bar{E}^{(N)}[\boldsymbol{g}]\right\} D[\xi] \tag{4.3.2a}$$

with

$$\bar{E}^{(N)}[\boldsymbol{g}] = \int_0^L \frac{1}{2}\left(\xi(s) - \bar{\xi}(s)\right)\cdot \bar{\mathcal{K}}(s)\left(\xi(s) - \bar{\xi}(s)\right) ds. \tag{4.3.2b}$$

We recall that the rod configurational distributions of the form in (4.3.2a) are discussed in (3.3.1). Similarly to finite dimensional normal distribution, a rod normal distributions is fully determined by prescribing the first and second moment of the variable $\xi(s)$ in the sense of (3.3.39). Using an argument very similar to the one used in (3.3.50), it can be shown that we have the the following identities

$$\langle \xi(s)\rangle^{(N)} = \bar{\xi}(s) \tag{4.3.3a}$$

$$\left\langle \left(\xi(s) - \bar{\xi}(s)\right)\otimes\left(\xi(s') - \bar{\xi}(s')\right)\right\rangle^{(N)} = \bar{\mathcal{K}}^{-1}(s)\delta_s(s') \tag{4.3.3b}$$

for $s \in]0, L[$, where $\langle\,.\,\rangle^{(N)}$ stands for the expectation with respect to the distribution $d\rho^{(N)}\,[\boldsymbol{g}[\xi]]$ in (4.3.2) and where $\delta_s(s')$ denotes the *Dirac function* as in (3.3.52).

As in the case of the deterministic Cauchy-Born rule, we have to emphasise that it is incompatible to look for a rod configurational distribution which is consistent with the chain distribution in (4.3.1), in the sense of (4.1.24), and which is normal, i.e. of the form (4.3.2). The cause is again that the mapping between the internal chain coordinates x and the rod generator vector $\xi(s)$ in non-linear. Since we want to keep both the chain stochastic model in (4.3.1), because it allows realistic sequence dependent modelling for DNA molecules, and the rod distribution (4.3.2b), because it allows analytical predictions, we need to make a compromise here as well. It is given by constructing a stochastic Cauchy-Born rule which only respects the moments to leading order through the second moment.

In order to construct the desired rule, we need to assume that the chain configurational distribution is concentrated enough such that it admits a unique expected chain configuration $\bar{\boldsymbol{g}} = (\bar{\boldsymbol{g}}_1, ..., \bar{\boldsymbol{g}}_N)$, as defined in (2.4.12), and we denote $\bar{\boldsymbol{g}}^{(N)}(s)$ its *piecewise helical interpolation*. We recall that an explicit approximation of the expected configuration is derived in appendix A.5 for chains in a semi-flexible regime and governed by a normal distribution. In a similar way to the deterministic Cauchy-Born rule in (4.2.4), we define formally the reference length L, the nodes ${s_n}^{(N)} \in]0, L[$ and the interpolation $\bar{\boldsymbol{g}}^{(N)}(s)$ by the relations

$$ {s_1}^{(N)} = 0,\ {s_{n+1}}^{(N)} = {s_n}^{(N)} + {h_n}^{(N)} \text{ and } {s_N}^{(N)} = L \text{ for all } n = 1, ..., N-1 \tag{4.3.4a} $$

$$ \frac{d}{ds}\bar{\boldsymbol{g}}^{(N)}(s) = \bar{\boldsymbol{g}}^{(N)}(s)\mathcal{T}\bar{\xi}^{(N)}(s) \text{ for all } s \in]0, L[\tag{4.3.4b} $$

$$ \bar{\boldsymbol{g}}^{(N)}({s_n}^{(N)}) = \bar{\boldsymbol{g}}_n \text{ for all } n = 1, ..., N \tag{4.3.4c} $$

$$ \mathcal{T}\bar{\xi}^{(N)}(s) = \frac{1}{{h_n}^{(N)}} \ln\{\bar{a}_n\} \text{ for } {s_n}^{(N)} < s < {s_{n+1}}^{(N)} \tag{4.3.4d} $$

$$ \left|\bar{V}^{(N)}(s)\right| = 1 \text{ for all } s \in]0, L[\tag{4.3.4e} $$

where $\bar{a}_n = \bar{\boldsymbol{g}}_n^{-1}\bar{\boldsymbol{g}}_{n+1}$ and where $\bar{V}^{(N)}(s) \in \mathbb{R}^3$ denotes the translational part of the rod generator vector $\bar{\xi}^{(N)}(s)$, as presented in (3.1.5). The interpolation $\boldsymbol{g}^{(N)}(s)$ of the chain configuration $\boldsymbol{g} = (\boldsymbol{g}_1, ..., \boldsymbol{g}_N)$ is then defined as

$$ \frac{d}{ds}\boldsymbol{g}^{(N)}(s) = \boldsymbol{g}^{(N)}(s)\mathcal{T}\xi^{(N)}(s) \text{ for all } s \in]0, L[\tag{4.3.5a} $$

with

$$ \mathcal{T}\xi^{(N)}(s) = \mathcal{T}\bar{\xi}^{(N)}(s) + \frac{1}{{h_n}^{(N)}} \left[\bar{\boldsymbol{g}}_n^{-1}\bar{\boldsymbol{g}}^{(N)}(s)\right]^{-1} \ln\left\{a(x_n)\bar{a}_n^{-1}\right\} \left[\bar{\boldsymbol{g}}_n^{-1}\bar{\boldsymbol{g}}^{(N)}(s)\right] \tag{4.3.5b} $$

for ${s_n}^{(N)} < s < {s_{n+1}}^{(N)}$ and where we recall that ${h_n}^{(N)} = {s_{n+1}}^{(N)} - {s_n}^{(N)}$ denotes the local junction length, similarly to (4.2.5).

Given the described interpolation rules and according to the definition in (4.1.24), a necessary condition for the chain and rod configurational distribution $d\rho(\boldsymbol{g})(x)$ and $d\rho^{(N)}[\boldsymbol{g}[\xi]]$ to be consistent is that the first moments satisfy

$$\left\langle \int_{s_n{}^{(N)}}^{s_{n+1}{}^{(N)}} \xi^{(N)}(s)\cdot b(s)ds \right\rangle^{(N)} = \left\langle \int_{s_n{}^{(N)}}^{s_{n+1}{}^{(N)}} \xi^{(N)}(x_n;s)\cdot b(s)ds \right\rangle \tag{4.3.6}$$

for any deterministic function $b(s)\in\mathbb{R}^6$ and for all $s_n{}^{(N)}$ in the interpolation mesh in (4.3.4a), and where $\langle\,.\,\rangle^{(N)}$ and $\langle\,.\,\rangle$ stand for the expectation with respect to the distribution $d\rho(\boldsymbol{g})(x)$ and $d\rho^{(N)}[\boldsymbol{g}[\xi]]$ respectively. The notation $\xi^{(N)}(x_n;s)(s)$ on the right hand side of (4.3.6) is meant to emphasise that the rod generator vector $\xi^{(N)}(s)$ is seen as an explicit function of the internal chain coordinates x_n according to the interpolation rule (4.3.5b). Using the specific form of the distributions in (4.3.1) and in (4.3.2) leads then to the equality

$$\bar{\xi}(s) = \bar{\xi}^{(N)}(s) \tag{4.3.7}$$

where we have used the identity in (4.3.3a) and, according to its definition in (2.4.10), that the expected chain junction displacement $\bar{a}_n$ has the property that $\left\langle \ln\left\{a(x_n)\bar{a}_n^{-1}\right\}\right\rangle = 0$.

A second other necessary conditions for the chain and rod distribution to be consistent is then that the second moments satisfy

$$\begin{aligned}&\left\langle \int_{s_n{}^{(N)}}^{s_{n+1}{}^{(N)}} \int_{s_m{}^{(N)}}^{s_{m+1}{}^{(N)}} \left(\xi^{(N)}(s)-\bar{\xi}^{(N)}(s)\right)\otimes\left(\xi^{(N)}(s')-\bar{\xi}^{(N)}(s')\right) : B(s',s)ds'ds \right\rangle^{(N)}\\ &= \left\langle \int_{s_n{}^{(N)}}^{s_{n+1}{}^{(N)}} \int_{s_m{}^{(N)}}^{s_{m+1}{}^{(N)}} \left(\xi^{(N)}(x_n;s)-\bar{\xi}^{(N)}(s)\right)\otimes\left(\xi^{(N)}(x_m;s')-\bar{\xi}^{(N)}(s')\right) : B(s',s)ds'ds \right\rangle\end{aligned} \tag{4.3.8}$$

for any deterministic function $B(s',s)\in\mathbb{R}^{6\times 6}$ and for all $s_n{}^{(N)}$ in the interpolation mesh (4.3.4a), and where $\langle\,.\,\rangle^{(N)}$ and $\langle\,.\,\rangle$ stand for expectation with respect to the distribution $d\rho(\boldsymbol{g})(x)$ and $d\rho^{(N)}[\boldsymbol{g}[\xi]]$ respectively.

Denoting by $\bar{x}_n\in\mathbb{R}^6$ the local chain coordinates which satisfy $\bar{a}_n = a(\bar{x}_n)$ in (4.3.5b) and using the expansion of the generator vector $\xi^{(N)}(x_n;s)$ around $\bar{x}_n$ in (4.2.6), gives then for $s\in]s_n{}^{(N)}, s_{n+1}{}^{(N)}[$

$$\begin{aligned}\bar{\mathcal{K}}^{-1}(s) = \frac{1}{h_n{}^{(N)}}\left(\mathrm{Ad}^{-1}_{\bar{\boldsymbol{g}}_n^{-1}\bar{\boldsymbol{g}}^{(N)}(s)}\mathbb{L}^{(1)}_{\bar{x}_n}\right)\langle(x_n-\bar{x}_n)\otimes(x_n-\bar{x}_n)\rangle\left(\mathrm{Ad}^{-1}_{\bar{\boldsymbol{g}}_n^{-1}\bar{\boldsymbol{g}}^{(N)}(s)}\mathbb{L}^{(1)}_{\bar{x}_n}\right)^T\\ +\left\langle \mathrm{O}\left(|x_n-\bar{x}_n|^3\right)\right\rangle\end{aligned} \tag{4.3.9}$$

according to the identity (4.3.3b). Moreover, using that the chain is in a semi-flexible regime, as defined in (2.4.14), leads to

$$\bar{\mathcal{K}}^{-1}(s) = \frac{1}{h_n{}^{(N)}}\left(\mathrm{Ad}^{-1}_{\bar{\boldsymbol{g}}_n^{-1}\bar{\boldsymbol{g}}^{(N)}(s)}\mathbb{L}^{(1)}_{\bar{x}_n}\right)K_n^{-1}\left(\mathrm{Ad}^{-1}_{\bar{\boldsymbol{g}}_n^{-1}\bar{\boldsymbol{g}}^{(N)}(s)}\mathbb{L}^{(1)}_{\bar{x}_n}\right)^T + \mathrm{o}(\epsilon) \tag{4.3.10}$$

since $\bar{x}_n = \hat{x}_n + \mathrm{O}(\epsilon)$ as presented in appendix A.5.

Consequently, for a given normal chain configuration $d\rho(\boldsymbol{g})(x)$ in (4.3.1) and for a given interpolation, we can define a normal rod configurational distribution $d\rho^{(N)}[\boldsymbol{g}[\xi]]$ using the expression (4.3.7) and (4.3.10) with the property that the first and second moment are consistent at the leading order. This construction constitutes a *stochastic Cauchy-Born rule for chains*. A method to parametrize a stochastic continuum rod model from quadratic rigid body chains is also discussed in [Becker, 2007]. It is however based on analogy with the Brownian dynamics of a rigid body and is greatly simplified by the fact that he uses an appropriate set of exponential coordinates to describe the discrete chain configuration.

We should observe that the quadratic form obtained for continuum rods in the case of the deterministic Cauchy-Born rule and of the stochastic Cauchy-Born rule are different. More precisely, in the deterministic case we obtain the ground state interpolation $\hat{\xi}^{(N)}$ and, in contrast, the expected chain interpolation $\bar{\xi}^{(N)}$ for the stochastic case. This distinction can be intuitively explained by the fact that the expected chain configuration represents in some sense the mean of the configurational distribution whereas the ground state configuration is its peak, and in general they simply do not coincide. Similarly, we obtain the matrix $\mathcal{K}(s)$ for the deterministic case, which corresponds to an approximation of the chain internal energy landscape around the ground state configuration and in the stochastic case we obtain the matrix $\bar{\mathcal{K}}(s)$ which is interpreted as an approximation of the centered second moment around the expected chain. An explicit example of these differences appears in appendix A.5.

Part II

On the Multi-scale and Sequence Dependent Statistical Physics of B-DNA

5. On the Multi-scale Modelling of Sequence Dependent DNA Mechanics

No single model explains the entire body of experimental data. [...] An improved explanation of solution properties should include sequence-dependent variations of local flexibility as well as differences in local chain structure.

R. C. Maroun & W. K. Olson, 1998.

This fifth chapter discusses discrete and continuum sequence dependent mechanical models for DNA molecules. We recall briefly in section 5.1 experimental evidence for the importance of understanding sequence dependence in DNA physical properties. We then present some historically important models, most of which cannot describe properly the sequence dependence. In section 5.2, we describe the sequence dependent rigid base pair and rigid base models, which are both coarse grain models but at different levels, which allow a DNA molecule to be described either as a single chain or as a bichain configuration. Three different sequences, denoted by λ , CF and T_L respectively, that will be repeatedly used as examples in the two following chapters are introduced. Sequence dependent rod and birod models are proposed in section 5.3 using the exponential Cauchy-Born rules presented in chapter 4. The associated constitutive coefficients are observed to be discontinuous at each base pair and to have a manifestly heterogeneous character at the scale of hundreds of base pairs. We conclude however that the deterministic exponential Cauchy-Born rule applied to the rigid base model is satisfactory since for instance the continuum birod and the original bichain energy agree to less than 0.5% for random bichain configurations. These rules relating discrete and continuum models are consequently believed to constitute significant progress towards a multi-scale approach of sequence dependent DNA mechanics.

5.1 Modelling Sequence Dependent DNA Mechanics

As has been discussed in the introduction, the objective of understanding the physical properties of DNA has become increasingly important since the second half of the last century, after a succession of discoveries on the crucial role of this macromolecule in cells of living organism. In particular, a fundamental question is the relation between the DNA sequence of a fragment ant its mechanical properties. On the experimental side, it has been proposed that particular fragments contain regions that are intrinsically bent [Marini et al., 1982, Kabsch et al., 1982, Kitchin et al., 1986, Levene et al., 1986, Bednar et al., 1995, Vologodskaia and Vologodskii, 2002] and that the local DNA stiffness may substantially vary with the sequence [Théveny et al., 1988, Olson et al., 1998, Virstedt et al., 2004]. These special regions are thought to be important in biological processes involved in the cells, see for instance [Hagerman, 1990], and more particularly to be related to DNA-protein recognition, see for instance [Matthews, 1988, Koudelka et al., 2006], and to nucleosome positioning in eukaryotic cells, as discussed for instance in [Thåström et al., 1999, Virstedt et al., 2004, Segal and Widom, 2009, Meyer and Everaers, 2015]. Originally proposed in [Porod, 1948, Krakty and Porod, 1949], the idea of *Kratky-Porod persistence length* allows the characterisation of some statistical properties[1] of DNA at intermediate scales (say from 10^3 to 10^5 base pairs). This quantity is shown to be related to a bending stiffness parameter in the so-called *Wormlike chain* model, either in its continuum formulation, see for instance [Landau and Lifshitz, 1959, Doi and Edwards, 1986], or in its discrete version [Schellman, 1974], which in both cases describe DNA as a uniform and intrinsically straight material which only responds to bending deformation. Even if this model is quite idealised, it has led to remarkably good results in average DNA properties modelling. In order to account for local intrinsic bends in the DNA structure, which is *not* possible in the Wormlike chain models, the idea of a set *static wedges* has been proposed in discrete representations of DNA configurations, see [Trifonov and Sussman, 1980, Trifonov et al., 1988, Crothers et al., 1990, Schellman and Harvey, 1995] for instance. For continuum formulations, substantial improvements have been made in the *Helical Wormlike chain* model, see [Bugl, 1969, Yamakawa, 1976, Benham, 1977, Benham, 1979, Yamakawa, 1997] for instance, which have taken more and more sophisticated forms [Marko and Siggia, 1994, Shi et al., 1996, Manning et al., 1996, Marko, 1997, Becker, 2007]. Available realistic sequence dependent DNA mechanical models can be divided into two classes: *atomistic models* and *coarse grain models*. Atomistic DNA models aim for a very fine description of molecule, namely at the level of atoms, see for instance [Lavery and Hartmann, 1994, Lankaš et al., 2000, Lankaš et al., 2003, Lavery et al., 2009, Lavery et al., 2014, Pasi et al., 2014]. The main issues with this approach are first that these models need a very large set parameters which are difficult to get since they have to come from *ab initio* principle, such as quantum chemistry, and second that they requires intensive computer resources. Even if the predictions of these models are definitely improving, their accuracy is still questionable. Coarse grain DNA models describe the molecule as a set of units which contain several atoms, and,

[1] See sections 2.3.2 and 3.3.2 for a more detailed discussion on chain and rod observables.

keeping the objective of modelling the sequence dependence mechanics of DNA, the two cases of interest here are *rigid base* and *rigid base pair* coarse graining descriptions, see for instance [Olson et al., 1993, Marky and Olson, 1994, Olson, 1996, Olson et al., 1998, Gonzalez and Maddocks, 2001, Olson et al., 2001, Lankaš et al., 2003, Lankaš et al., 2009, Becker and Everaers, 2009, Gonzalez et al., 2013]. Even if these models contain less information than atomistic models, they are expected to provide an appropriate description of sequence dependent mechanical features of DNA molecules at the scale of several bases, and in a efficient way since the number of degree of freedom together with the number of parameters have substantially decreased. The two models are described in section 5.2.

Modelling the sequence dependent statistical physics of DNA molecule is a challenging objective. First, because most of the classical results in polymer mechanics do not apply to heterogenous chains, and, second, most of the biological phenomena of interest require at least some hundreds of base pairs to be described and need therefore an understanding of the sequence dependent large scale behaviour. We believe however that the definition of the chain models in 5.2 and their associated rod ones in 5.3 constitute a significant progress towards a multi-scale approach of sequence dependent DNA mechanics.

5.2 On Rigid Base and Rigid Base Pair DNA Models

We present in this section two discrete models used to study the sequence dependence of DNA statistical physics. The rigid base pair model describes DNA molecules as a single rigid body chain, where each rigid body characterises an individual base pair configuration. The configuration is made of the product of local and sequence dependent normal distributions. There exist several different versions of this model, based on different data and different internal chain coordinates. In contrast, the rigid base model describes each individual base as a rigid body and consequently each DNA fragment as a bichain configuration. The configurational distribution is also normal but, even though it has a very specific nearest neighbour structure, it no longer factorizes into local distributions. We introduce in particular three chosen sequences, denoted by λ , CF and T_L respectively, that will be used as illustrations in the following applications.

5.2.1 Sequence Dependent Rigid Base Pair Model

In order to describe the sequence dependent mechanical properties of DNA in term of a coarse grain model, a first approach is to represent each base pair as a single rigid body. A rigid base pair model describes a DNA fragment as a *rigid body chain configuration* $\boldsymbol{g} = (\boldsymbol{g}_1, ..., \boldsymbol{g}_N)$, as defined in (2.1.1), where N denotes the number of pairs of nucleotides and where $\boldsymbol{g}_n \in \mathrm{SE}(3)$ denotes the rigid body configuration of the n^{th} rigid body in the chain, see for instance [Olson, 1996, Olson et al., 1998, Gonzalez and Maddocks, 2001, Olson et al., 2001, Lankaš et al., 2003]. The junction displacement $a_n = \boldsymbol{g}_n^{-1}\boldsymbol{g}_{n+1}$ is the rigid body displacement between two consecutive elements in the chain and the local internal coordinates $x_n \in \mathbb{R}^6$ parametrize the junction displacement in the form $a_n = a(x_n)$. The

internal chain coordinates are then denoted by $x = (x_1, .., x_{N-1}) \in \mathbb{R}^{6(N-1)}$. Several different choices of internal chain coordinates have been introduced in in the literature [Olson et al., 1998, Becker and Everaers, 2007, Gonzalez et al., 2013]. More detail about rigid body chains configurations and the internal chain coordinates is discussed in section 2.1.1.

Following the pioneering work of Wilma K. Olson, see for instance [Olson et al., 1998], the *sequence dependent rigid base pair model* configurational distribution is assumed to be of the factorized form in (2.3.4) for any DNA fragment. More details on stationary configurational distribution for rigid body chains are discussed in section 2.3.1. The sequence variation in the molecule constitution is then encoded through a set of *sequence dependent chain local configurational distribution* $d\rho_n(a_n)$ defined as

$$d\rho_n(a_n) = \sqrt{\det\left(\frac{1}{2\pi}K_n^{xx}\right)}\exp\left\{-\frac{1}{2}(x_n - \hat{x}_n)\cdot K_n^{xx}(x_n - \hat{x}_n)\right\}dx_n \tag{5.2.1}$$

where $\hat{x}_n$ are called the *ground state internal coordinates* and the matrix $K_n^{xx} \in \mathbb{R}^{6\times 6}$ the *local stiffness matrix* in analogy to linear elasticity theory. Practically, given a set of data from DNA fragments with known sequences, one first converts them into a set of rigid body configurations $\boldsymbol{g}$ in order to extract statistics for the local internal coordinates x_n. Then, the expectation $\hat{x}_n$ and the matrix K_n^{xx} can be estimated, according to the relations in (2.3.20) and in (2.3.21), using the empirical first and second moments of the internal coordinate x_n on the set of data, see for instance [Olson et al., 1998, Becker and Everaers, 2007, Gonzalez and Maddocks, 2001, Lankaš et al., 2009, Becker and Everaers, 2009]. Observe that neither the bath temperature k_BT nor the Jacobian factor $J_n(x_n)$, as defined in (2.3.1), appear in the distribution (5.2.1) since it is seen as the best normal approximation in the internal chain coordinates x. Figure 5.1 a sketch of the coarse graining process between an all atom and a rigid base, or rigid base pair, representation.

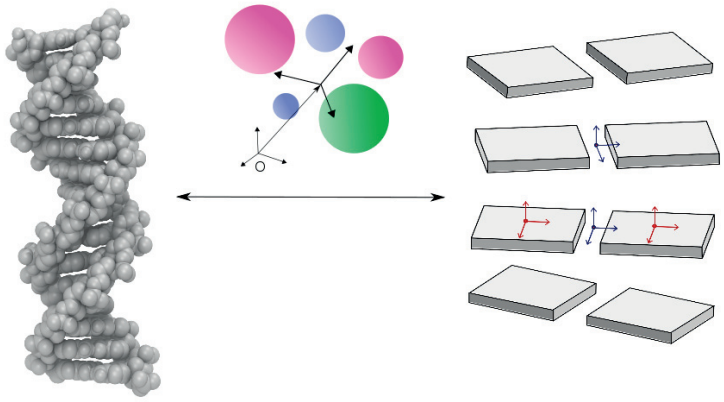


Figure 5.1: Sketch of the coarse graining process between an all atom and a rigid base, or rigid base pair, representation.

The main difference between the models in [Olson et al., 1998, Becker and Everaers, 2007, Lankaš et al., 2009] are the definition of the internal coordinate x and the data ensemble as well as the method used to obtain the distribution in (5.2.1). In particular, the parameter estimation in [Olson et al., 1998] is based on crystal structure data, whereas [Lankaš et al., 2009] is based on molecular dynamic simulations.

A crucial point made in [Lankaš et al., 2009] is that the locality assumption in the functional form of the configurational distribution in (5.2.1) can be assessed by evaluating the cross covariances matrices $c^{xx}_{m,n}$ for $m \neq n$, since they should vanish according to the expression in (2.3.22). Interestingly enough, based on time series trajectories coming from all atom molecular dynamic simulations, it has been emphasised that the assumption that the configurational distribution for the rigid base pair model factorizes into local distribution is not very accurate since the cross covariances $c^{xx}_{m,n}$ can be quite large even for substantial differences of index $|m-n|$ [Lankaš et al., 2009, Gonzalez et al., 2013]. This observation motivated the introduction of the rigid base model.

5.2.2 Sequence Dependent Rigid Base Model

A slightly less coarse grained description of a DNA fragment is to represent the configuration of each individual base as a rigid body configuration, instead of the complete base pair, or, using the definition of section 2.1.2, as a *rigid body double chain configuration* $(\boldsymbol{g}^+, \boldsymbol{g}^-) = (\boldsymbol{g}^+_1, \boldsymbol{g}^-_1, ..., \boldsymbol{g}^+_N, \boldsymbol{g}^-_N)$ where N denotes the number of pair of nucleotides and where $\boldsymbol{g}^\pm_n$ denotes the rigid body configuration of the n^{th} base on the strand referred to as $+$ or $-$ respectively. However, following prior work on DNA modelling [Olson et al., 2001, Lankaš et al., 2009, Gonzalez et al., 2013], it is more convenient to describe a DNA molecule coarse grained at the level of bases as a *rigid body bichain configuration* $(\boldsymbol{g}, \mathcal{P}) = (\boldsymbol{g}_1, \mathcal{P}_1, ..., \boldsymbol{g}_N, \mathcal{P}_N)$, which is called the *rigid base model*. The rigid body chain configuration $\boldsymbol{g}$ defines the *bichain macrostructure* and represents the set of rigid base pair configurations, whereas $\mathcal{P}$ defines the *bichain microstructure* and each element $\mathcal{P}_n$ describes the relative rigid body motion between the two bases in the same base pair indexed by n. The detailed transformation rule between double chain configurations $(\boldsymbol{g}^+, \boldsymbol{g}^-)$ and the bichain configurations $(\boldsymbol{g}, \mathcal{P})$ is presented in (2.1.25) and in (2.1.27). The bichain internal coordinates $(x, y) = (y_1, x_1, y_2, ..., x_{N-1}, y_N) \in \mathbb{R}^{12N-6}$ are defined such that $a_n = a(x_n)$ and $\mathcal{P}_n = \mathcal{P}(y_n)$, where a_n denotes the local junction displacement. The choice proposed in [Lankaš et al., 2009, Gonzalez et al., 2013] is observed to be particularly convenient and is presented in more detail in (2.1.18) and in (2.1.33). More detail on the bichain configuration and the bichain internal coordinates is discussed in section 2.1.3.

As a generalisation of the sequence dependent rigid base pair model presented in section 5.2.1, and as discussed in detail in [Lankaš et al., 2009, Gonzalez et al., 2013], the *sequence dependent rigid base model* comprises a set of bichain coordinates (x, y) together with a *sequence dependent configurational distribution* $d\rho(\boldsymbol{g}, \mathcal{P})(x, y)$ of the form

$$d\rho(\boldsymbol{g}, \mathcal{P})(x, y) = \sqrt{\det\left(\frac{1}{2\pi}K\right)} \exp\left\{-\frac{1}{2}\begin{pmatrix} x-\hat{x} \\ y-\hat{y}\end{pmatrix} \cdot K \begin{pmatrix} x-\hat{x} \\ y-\hat{y}\end{pmatrix}\right\} dxdy \qquad (5.2.2)$$

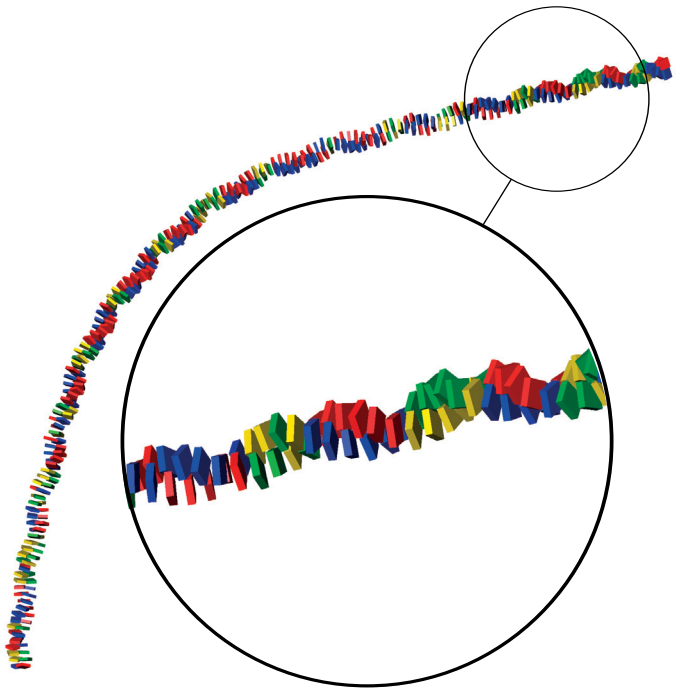


Figure 5.2: Example of a bichain ground state configuration for a DNA fragment made of 150 selected base pairs in the sequence *CF* .

where $(\hat{x}, \hat{y})$ are called the *bichain ground state internal coordinates* and the matrix $K \in \mathbb{R}^{(12N-6)\times(12N-6)}$ the *bichain stiffness matrix*, similarly to (5.2.1), and where the ordering of the coordinate vector (x, y) and of the entries in the matrix K is defined in (2.1.31). More detail about bichain configurational distribution can be found in section 2.3.1. Note that, as for the rigid base pair configurational distribution, the temperature $k_B T$ and the Jacobian factor $J(x, y)$, defined in (2.3.6), do not explicitly appear since the distribution (5.2.2) is a normal approximation in the bichain coordinates (x, y). Let us define $E(\boldsymbol{g}(x), \mathcal{P}(y))$ to be the *quadratic internal bichain energy* in the internal bichain coordinates (x, y)

$$E(\boldsymbol{g}(x), \mathcal{P}(y)) = \frac{1}{2} \begin{pmatrix} y_1 - \hat{y}_1 \\ x_1 - \hat{x}_1 \\ \dots \\ x_{N-1} - \hat{x}_{N-1} \\ y_N - \hat{y}_N \end{pmatrix} \cdot K \begin{pmatrix} y_1 - \hat{y}_1 \\ x_1 - \hat{x}_1 \\ \dots \\ x_{N-1} - \hat{x}_{N-1} \\ y_N - \hat{y}_N \end{pmatrix} \tag{5.2.3}$$

where the internal coordinates $(\hat{x}, \hat{y})$ as well as the matrix K are the ones in the distribution (5.2.2). We should stress that, according to the results on stationary bichain configurational distribution in section 2.3.1, the function $E(\boldsymbol{g}(x), \mathcal{P}(y))$ is not rigorously the bichain energy appearing in (2.3.6) because of the necessary presence of the associated Jacobian factor. We shall however make this abuse of nomenclature for convenience

of the discussion. The bichain configuration corresponding to the ground state internal coordinates $(\hat{x}, \hat{y})$ is then called the *bichain ground state configuration* and is denoted by $(\hat{\boldsymbol{g}}, \hat{\mathcal{P}})$. It is formally defined as

$$\hat{\boldsymbol{g}}_{n+1} = \hat{\boldsymbol{g}}_n a(\hat{x}_n) \text{ and } \hat{\mathcal{P}}_n = \mathcal{P}(\hat{y}_n) \tag{5.2.4}$$

for all $n = 1, ..., N-1$, where the first rigid body configuration $\hat{\boldsymbol{g}}_1$ is given. An example of a three dimensional bichain ground state configuration is drawn in figure 5.2.

In the sequence dependent rigid base model developed in [Petkeviciute, 2012, Gonzalez et al., 2013, Petkeviciute et al., 2014], one of the key hypotheses is that the quadratic bichain internal energy is local, in the sense of (2.2.46), and that it can be written as a sum of local nearest neighbor contributions, i.e. of the form

$$E(\boldsymbol{g}(x), \mathcal{P}(y)) = \sum_{n=1}^{N-1} \frac{1}{2} \begin{pmatrix} y_n - \hat{y}_n^{\alpha\beta} \\ x_n - \hat{x}_n^{\alpha\beta} \\ y_{n+1} - \hat{y}_{n+1}^{\alpha\beta} \end{pmatrix} \cdot K_n^{\alpha\beta} \begin{pmatrix} y_n - \hat{y}_n^{\alpha\beta} \\ x_n - \hat{x}_n^{\alpha\beta} \\ y_{n+1} - \hat{y}_{n+1}^{\alpha\beta} \end{pmatrix} + \sum_{n=1}^{N} \frac{1}{2} (y_n - \hat{y}_n^{\alpha}) \cdot K_n^{\alpha} (y_n - \hat{y}_n^{\alpha}) \tag{5.2.5}$$

where the subscripts $\alpha\beta$ and α refer respectively to the base pair sequence of the macrostructure junction between the base pairs indexed by n and $n+1$, and to the base composition of the microstructure indexed by n. This decomposition allows the deconvolution of the sequence dependent parameters appearing in the configurational distribution $d\rho(\boldsymbol{g}, \mathcal{P})(x, y)$. In particular the bichain stiffness matrix K in (5.2.2) has then the very specific *nearest neighbor pattern* drawn in figure 5.3. One of the important properties of this model is that the ground state internal coordinates $(\hat{x}, \hat{y})$ are predicted to have a non local dependence in sequence even if the constitutive parameters are locally defined, as discussed in [Gonzalez et al., 2013].

With a parameter set estimated from the all-atom simulations, reported in [Lavery et al., 2009, Pasi et al., 2014], of approximately forty B-DNA fragments each of eighteen base pairs length in explicit solvent, and using the decomposition (5.2.5), a very fast numerical procedure allows to approximate the sequence dependent configurational distribution $d\rho(\boldsymbol{g}, \mathcal{P})(x, y)$ in (5.2.2) for *any* B-DNA sequence. This procedure is automated using the freely available package *cgDNA* described in [Petkeviciute et al., 2014]. Moreover, the obtained distributions are shown to be in a very good agreement with the original atomic simulations after the coarse grained procedure, and with independent simulations as well [Petkeviciute, 2012, Gonzalez et al., 2013, Petkeviciute et al., 2014], which attests that the rigid base model has to be preferred to the rigid base pair model to study DNA statistical physics properties at relatively short scales.

To illustrate the results of this thesis on the application of chain and rod theory to the modelling of DNA statistical physics properties, we have chosen three sequences that we shall repeatedly use in most of the following examples. The first one is denoted by λ and is a fragment of the virus genome *Lambda phage* which infects the *E. Coli bacterium*. This virus has been extensively studied and the relatively short sequence λ that we have chosen serves as a reference sequence since it has a roughly intrinsically straight ground

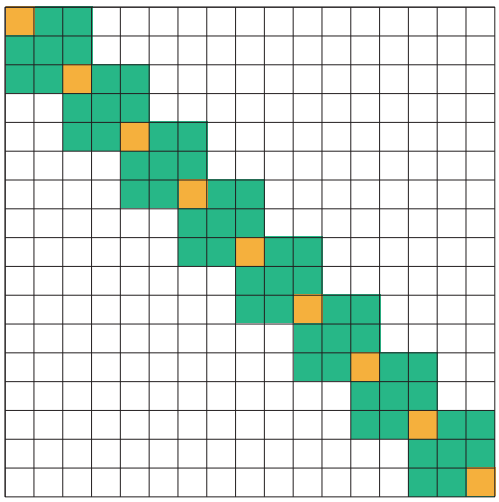


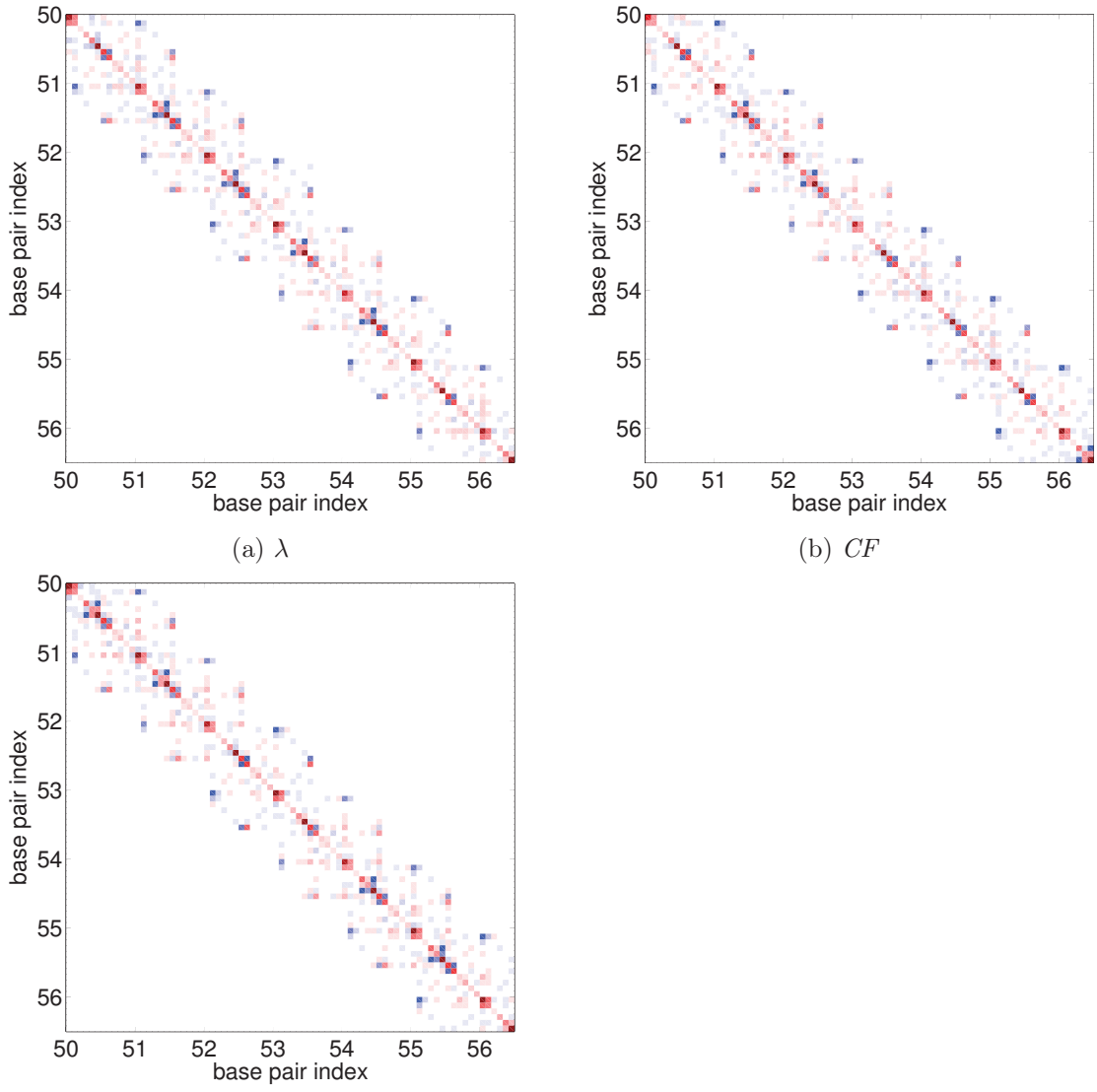
Figure 5.3: Sketch of a nearest neighbor bichain stiffness matrix: the green blocks represents the position of the blocks $K_n^{\alpha\beta}$, the orange ones are the superposition of the blocks $K_n^{\alpha\beta}$, K_n^{α} and $K_{n+1}^{\alpha\beta}$ and the white ones have zero as entries. The size of mesh indicates 6×6 blocks.

state, as drawn in figure 5.4a, and no exceptional region expected. The second one is a section of the genome of the *Crithidia fasciculata*, which is a mosquito parasite, and is here denoted by *CF* . It has also been well studied because it has the particularity to contain a large intrinsic bend as proposed in [Kitchin et al., 1986] and as it is observed in its cgDNA ground state shown in figure 5.4b. Finally, the third sequence has been artificially designed to have an pronounced intrinsic super-helical structure which is left-handed, as drawn in 5.4c. It is denoted by T_L and is made of repeats of the sequence $(\mathrm{CATGGTTAAC})_p$. Such helical repeats have been used in [Dubochet et al., 1994] for instance and a review on their role in genomics can be found in [Padeken et al., 2015]. Entries in the associated stiffness matrices given in the cgDNA model are respectively shown in figures 5.5a, 5.5b, 5.5c. The precise sequences λ and *CF* here are listed in appendix A.6.

Figure 5.6 shows the full nearest neighbor stiffness matrix for the sequence λ as computed in the cgDNA model. Even if at the scale of a few base pairs this matrix is locally dense, as shown in figure 5.5a, the bandwidth looks tiny at the scale of a few hundred base pairs.

(a) λ (b) CF (c) T_L

Figure 5.4: Bichain ground state configuration for three different sequences as predicted by the cgDNA model. The figures show base pairs 1 to 240.



(a) λ

(b) CF

(c) T_L

Figure 5.5: Zoom on the bichain stiffness matrices predicted by the cgDNA model for three different sequences. The plots shows the entries from the 50^{th} to 56^{th} base pairs

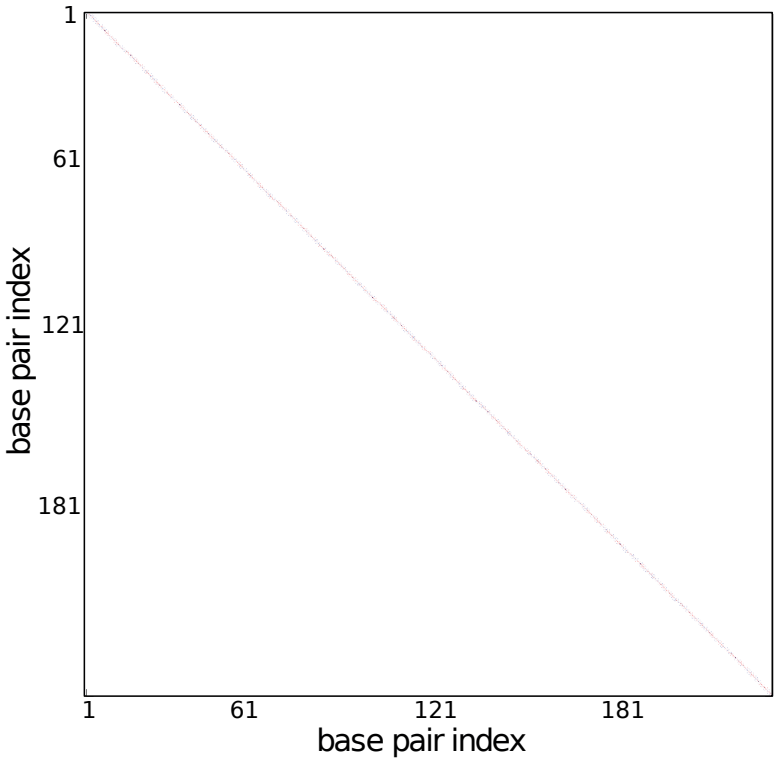


Figure 5.6: Bichain stiffness matrix for the sequence λ (240 first base pairs).

5.3 On Rod and Birod DNA Models

This section presents the continuum analogue of the rigid base pair and rigid base model, which are the sequence dependent rod and birod models respectively. There exist many continuum rod models that have been proposed to model the large scale behaviour of DNA molecules. However all of them are homogeneous, with only a few exceptions. We first propose a sequence dependent quadratic rod model based on the stochastic exponential Cauchy-Born rule discussed in section 4.3. It is interesting to note that the obtained continuum coefficients are only continuous between two consecutive base pairs and that their heterogeneous character is manifest at the scale of hundreds of base pairs. Then, a sequence dependent quadratic birod model built using the deterministic exponential Cauchy-Born rule discussed in section 4.2 is studied. The stiffness coefficient are also noticed to be discontinuous at each base pair position and have an obvious heterogeneous character. However, this deterministic Cauchy-Born rule is concluded to be quite satisfactory by comparing numerically the energy value for the rigid base and for the associated birod energies. In particular an error of less than 0.5% is observed for 2500 random bichain configurations sampled from its corresponding configurational distribution.

5.3.1 Sequence Dependent Rod Model

As originally proposed in [Porod, 1948, Krakty and Porod, 1949], even if the natural characterisation of polymers is through discrete models, an intermediate scale description should be made in terms of a continuum. In particular, a quadratic continuum rod energy has been proposed in [Landau and Lifshitz, 1959, Bugl, 1969] to model the behaviour of long linear polymers, which has lead to the idea of the *Helical Wormlike chain* [Yamakawa, 1976, Yamakawa, 1997] which uses bending and twisting stiffness parameters and an intrinsically helical geometry. Specifically in the context of DNA, equilibrium conditions, with idealised linear constitutive relations are introduced in [Benham, 1977, Benham, 1979] and special solutions are studied. The energy functional is then generalised in [Marko and Siggia, 1994, R.S. Manning, 1996, Shi et al., 1996]. As mentioned in section 3.1.1, all these models belong to a general class of hyper-elastic rod models called *Cosserat rod models* [Cosserat and Cosserat, 1909].

Concerning models with explicit sequence dependence, a rod model similar to the one in [Benham, 1977, Benham, 1979] is parametrized in [R.S. Manning, 1996] to study sequence dependence in the intrinsic shape, although the stiffness parameters are sequence independent however. They also introduce a Hamiltonian form of the equilibrium conditions using a quaternion formulation, which is used to compute equilibrium configurations of closed loops by applying a numerical continuation method provided by the package *AUTO*, see for instance [Doedel et al., 2009]. In [Becker, 2007] a general quadratic continuum rod model of the form 4.3.2 is introduced in the context of DNA modelling and then, in [Becker, 2007, Becker and Everaers, 2007], they discuss how, based on a local sequence dependent rigid base pair model expressed through *exponential coordinates*, an effective homogeneous wormlike chain can be deduced though some averaging processes. They also emphasize the obvious limitation of such models for relatively short DNA fragments.

Based on the theory of Cosserat rods presented in section 3.2.2, the continuum analog of the rigid base pair model in (5.2.1) is given by a *sequence dependent rod local configurational distribution* $d\rho\,[\boldsymbol{g}[\xi]]$

$$d\rho\,[\boldsymbol{g}[\xi]] \sim \exp\left\{-\int_0^L \frac{1}{2}\left(\xi(s)-\bar{\xi}(s)\right)\cdot\bar{\mathcal{K}}(s)\left(\xi(s)-\bar{\xi}(s)\right)ds\right\}D[\xi] \tag{5.3.1}$$

where $\xi(s)\in\mathbb{R}^6$ is the generator vector of the rod configuration $\boldsymbol{g}(s)\in\mathrm{SE}(3)$, as defined in (3.1.4), where $\bar{\xi}(s)\in\mathbb{R}^6$ denotes the generator vector of the *expected rod configuration* $\bar{\boldsymbol{g}}(s)\in\mathrm{SE}(3)$, and the matrix $\bar{\mathcal{K}}(s)\in\mathbb{R}^{6\times6}$ the *rod local stiffness matrix* analogously to the rigid base pair case in section 5.2.1. The length L denotes a reference length, typically chosen as the length of the expected rod configuration. The non-uniformity of the constitutive parameters $\bar{\xi}(s)$ and $\bar{\mathcal{K}}(s)$ is supposed to encode the sequence dependence.

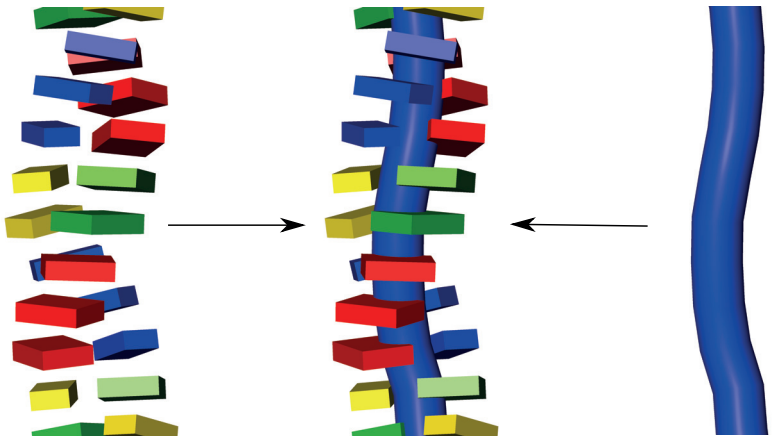


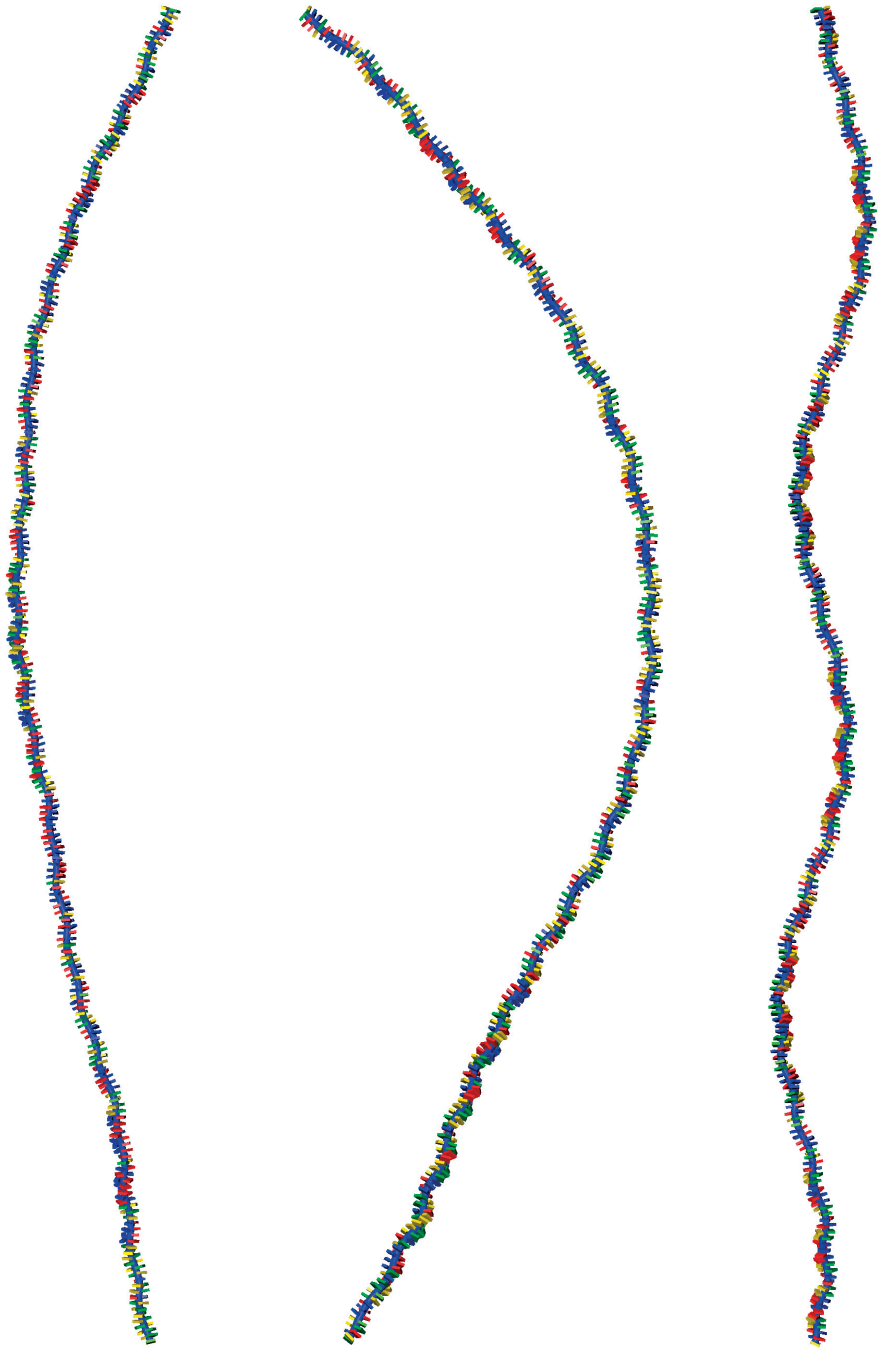
Figure 5.7: Sketch of the expected chain configuration interpolation

The rod configurational distribution $d\rho\,[\boldsymbol{g}[\xi]]$ in (5.3.1) can be defined for instance directly from the rigid base pair distribution $d\rho(\boldsymbol{g})(x)$ in (5.2.1) using the exponential stochastic Cauchy-Born rule for chains proposed in section 4.3. A sketch of the interpolated rigid base pair configuration is drawn in figure 5.7, it illustrates the scale bridging between a rigid base pair and a continuum rod description. The expected rod configurations $\bar{\boldsymbol{g}}(s)$, deduced using the rule in (4.3.7), are presented in figure 5.8 for the sequence λ , CF and T_L . It is interesting to note that for these three sequences it is difficult to see the difference between the base pair positions in the expected configuration and in the ground state configuration drawn in 5.4. However, according to the discussion made in A.5.3, each junction displacement differs slightly and a deviation may appear for long enough fragments.

For these three sequences the coefficient in the matrix $\bar{\mathcal{K}}(s)$ obtained using (4.3.10) are illustrated in figures 5.9, 5.10a and 5.10b respectively. In each figure, the top left plot shows the entries in the symmetric block $\mathcal{K}^{\boldsymbol{RR}}(s) \in \mathbb{R}^{6\times 6}$, the top right the entries of $\mathcal{K}^{\boldsymbol{Rr}}(s) \in \mathbb{R}^{6\times 6}$ and the bottom left the symmetric block $\mathcal{K}^{\boldsymbol{rr}}(s) \in \mathbb{R}^{6\times 6}$ where by definition

$$\mathcal{K}(s) = \begin{pmatrix} \mathcal{K}^{\boldsymbol{RR}}(s) & \mathcal{K}^{\boldsymbol{Rr}}(s) \\ (\mathcal{K}^{\boldsymbol{Rr}})^T(s) & \mathcal{K}^{\boldsymbol{rr}}(s) \end{pmatrix}. \tag{5.3.2}$$

It is interesting to observe that the exponential stochastic Cauchy-Born rule delivers coefficients which reflect the DNA sequence, and that they are not globally continuous. We note that the discontinuity is exactly at the base pair level. In contrast, at the scale of a few hundred base pairs, the entries of the matrix $\bar{\mathcal{K}}(s)$ take a more heterogeneous character. Figure 5.11 shows the coefficient associated to the first 240 base pairs of the sequence λ . We note that a possible way to parametrize the rod distribution in (5.3.1), using a given sequence dependent rigid base pair model of the form (5.2.1), in terms of appropriate exponential coordinates, is mentioned in [Becker, 2007].



(a) λ (b) CF (c) T_L

Figure 5.8: Rod expected configuration for three different sequences of length 240 base pairs, as used in the stochastic Cauchy-Born rule for chains.

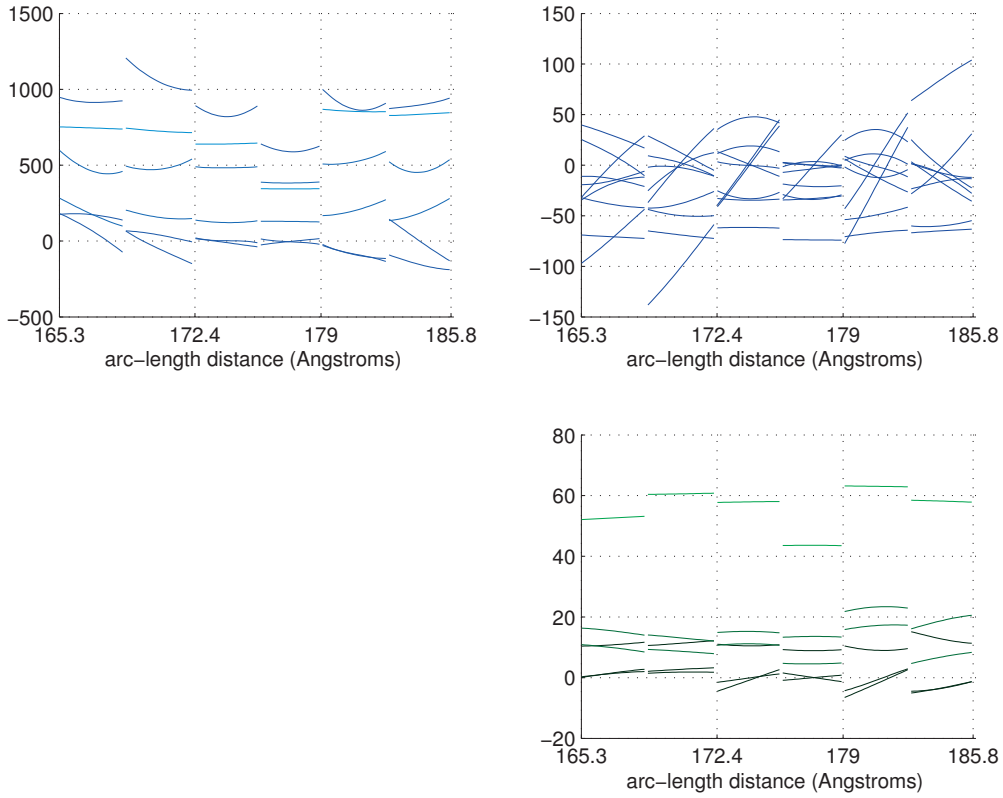
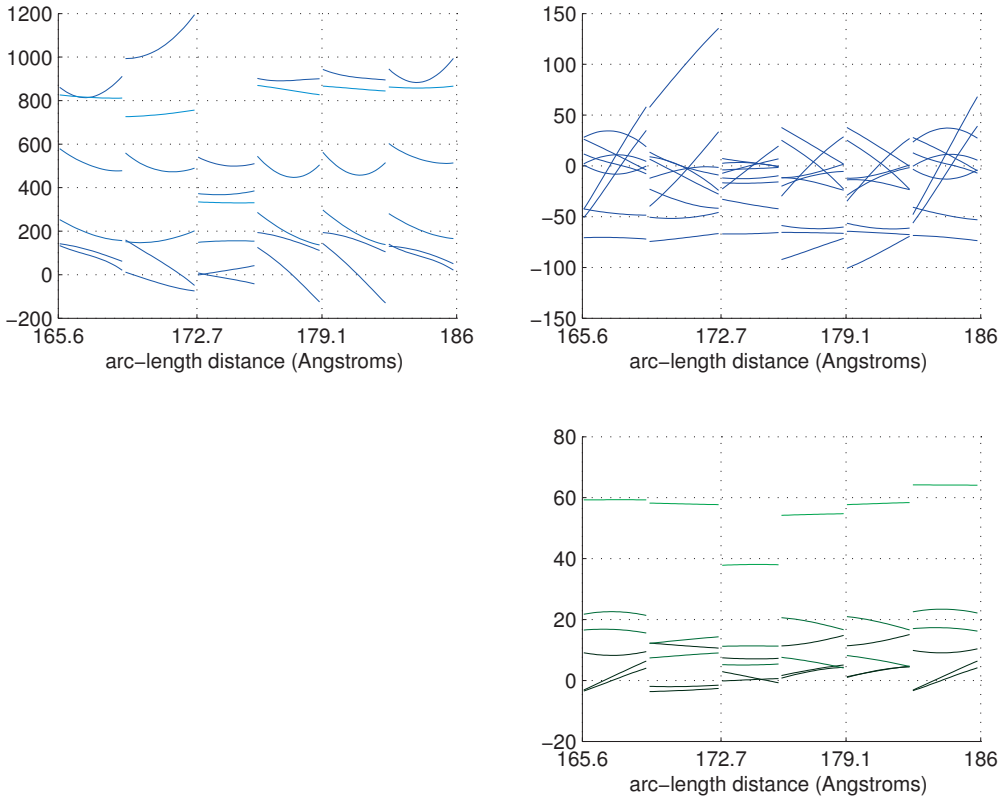
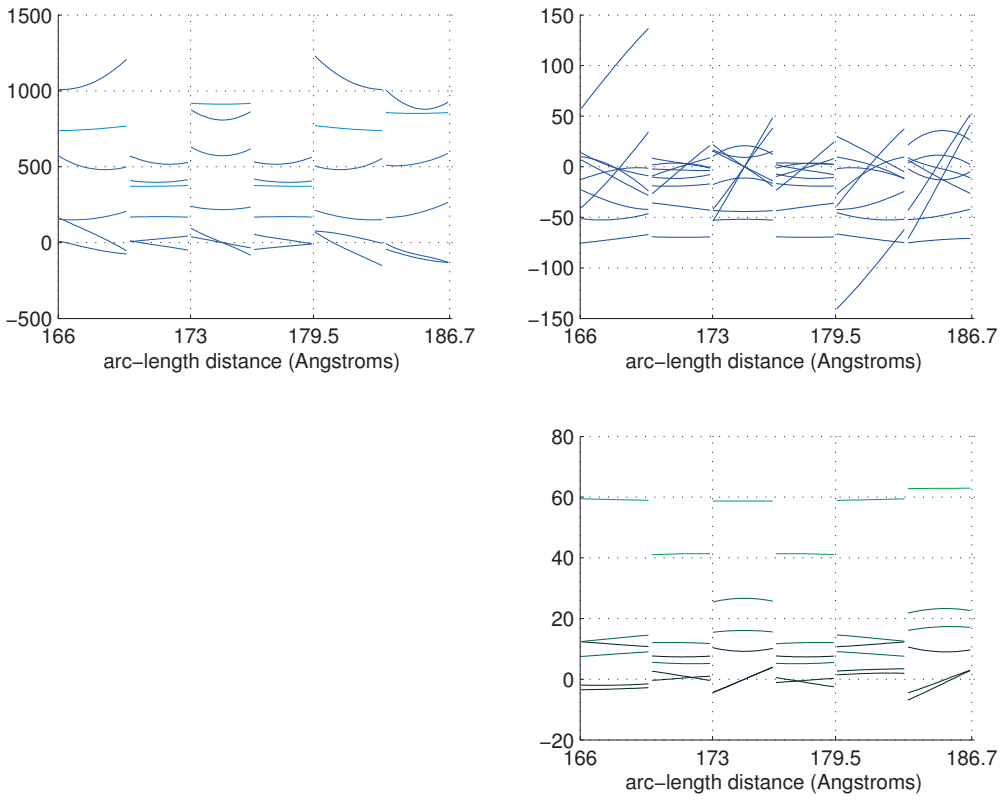


Figure 5.9: Zoom on the entries of the rod stiffness matrices predicted by the stochastic Cauchy-Born rule for the sequence λ . The plots shows the entries from the 50^{th} to 56^{th} base pair. A more detailed description is provided in the text.

However, as reported in section 5.2.1, the local rigid base pair model does not provide a very accurate description of coarse grained DNA statistical physics, whereas the local rigid base model, described in section 5.2.2, is shown to be able to reproduce a much more rich and reliable behaviour [Lankaš et al., 2009, Gonzalez et al., 2013]. This statement is then contradictory with a local rod model of the form in (5.3.1), or even more generally in (3.2.5), principally because of the non-local sparsity pattern involved in the definition of the nearest neighbor rigid base energy, as shown in figure 5.3. There are two main solutions to face this issue and persist in continuum modelling of DNA molecules. The first possibility is going towards non-local rod models. We believe however that this route is not promising to allow a better understanding of coarse grained DNA fragments since non-local theories in continuum mechanics are themselves still at an early stage of development. The other possibility is to abandon a rod theory in favour of a birod theory.



(a) *CF*



(b) T_L

Figure 5.10: Zoom on the entries of the rod stiffness matrices predicted by the exponential stochastic Cauchy-Born rule for the sequence *CF* and T_L . See caption of figure 5.9.

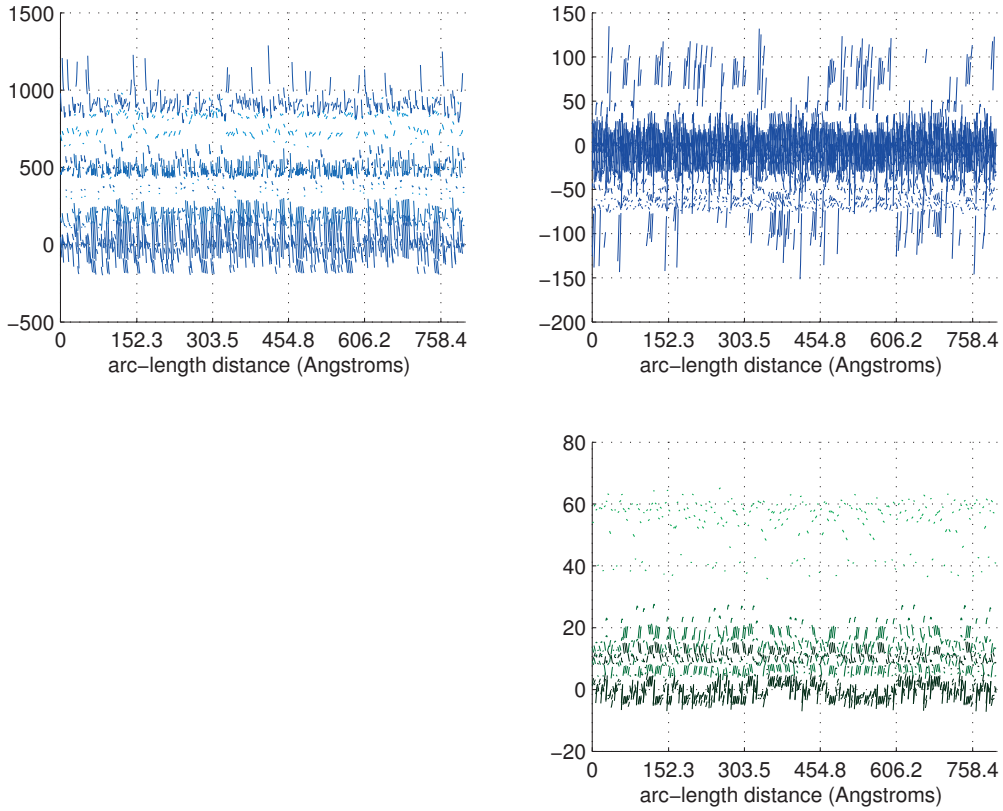


Figure 5.11: Entries of the rod stiffness matrices predicted by the exponential stochastic Cauchy-Born rule for the sequence λ . The plots shows the entries from the 1^{st} to 240^{th} base pair. A more detailed description is provided in the text.

5.3.2 Sequence Dependent Birod Model

The deterministic continuum birod model was first introduced in [Moakher and Maddocks, 2005] in order to deliver a mechanical theory for two rods interacting together, and especially for coarse grained DNA molecules, even if no realistic constitutive relations were available at that time. In order to make this connection, the birod theory has been adapted as described in sections 3.1.3 and 3.2.7. However, the theory of stochastic birod models is still under investigation, and we will only discuss the deterministic formulation in this thesis.

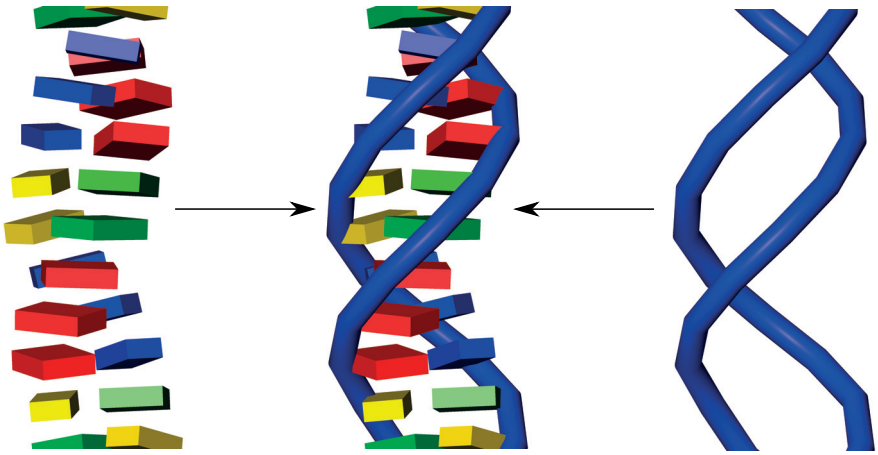


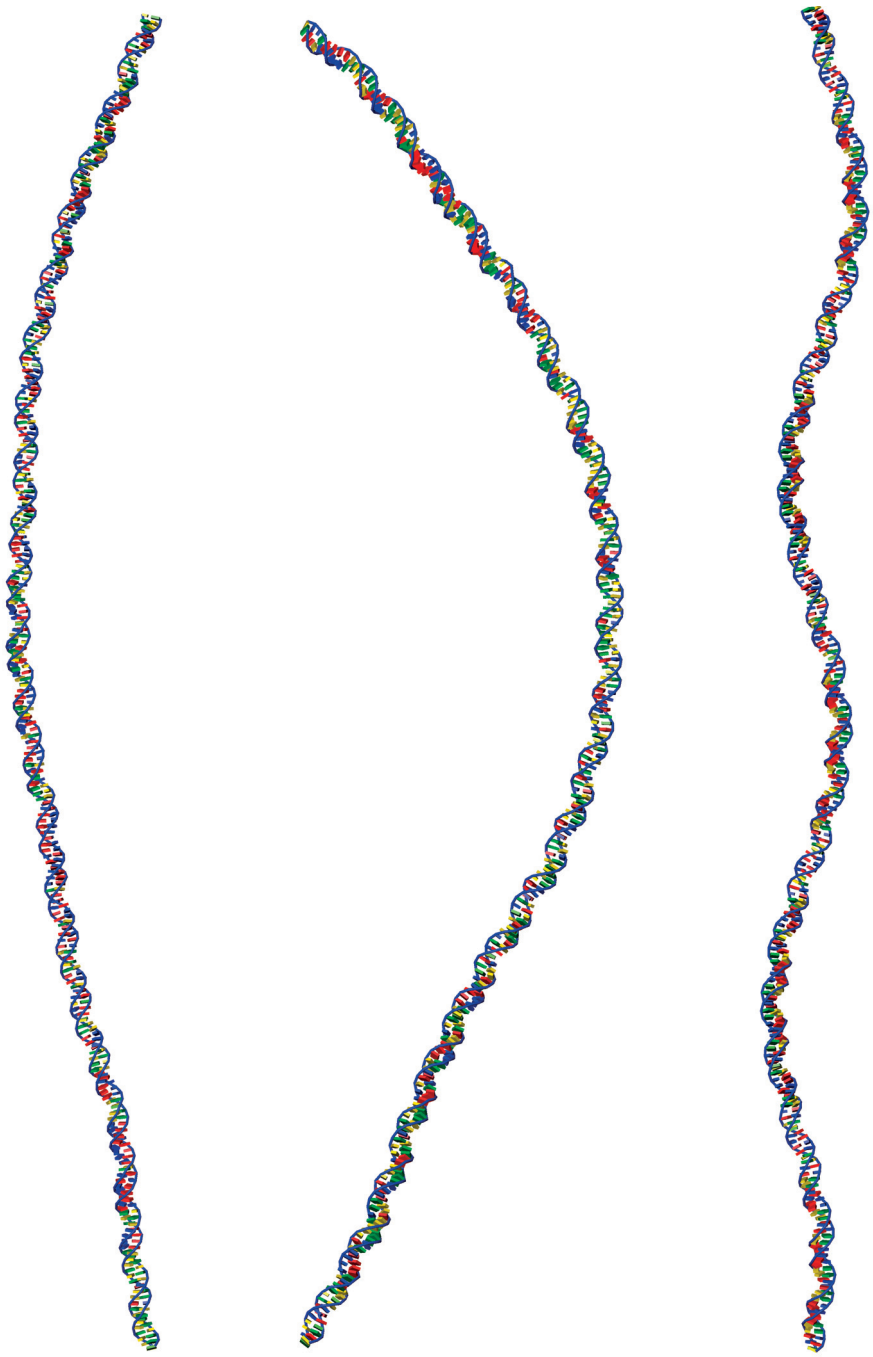
Figure 5.12: Sketch of the bichain ground state configuration interpolation

Motivated by the discussion made in section 4.2 on the continuum formulation of a quadratic bichain internal energy of the form (5.2.3), we define the *sequence dependent quadratic birod energy* in terms of the internal birod variables (ξ, y) to be of the form

$$\begin{aligned}\hat{E}[\boldsymbol{g}, \mathcal{P}(y)] = \int_0^L \frac{1}{2} \begin{pmatrix} y(s) - \hat{y}(s) \\ \frac{d}{ds} y(s) - \frac{d}{ds} \hat{y}(s) \\ \xi(s) - \hat{\xi}(s) \end{pmatrix} \cdot \mathcal{K}(s) \begin{pmatrix} y(s) - \hat{y}(s) \\ \frac{d}{ds} y(s) - \frac{d}{ds} \hat{y}(s) \\ \xi(s) - \hat{\xi}(s) \end{pmatrix} ds \\ + \frac{1}{2} \left(y(0) - \hat{y}(0)\right) \cdot \mathcal{K}_0 \left(y(0) - \hat{y}(0)\right) + \frac{1}{2} \left(y(L) - \hat{y}(L)\right) \cdot \mathcal{K}_L \left(y(L) - \hat{y}(L)\right)\end{aligned} \tag{5.3.3}$$

where $\xi(s)$ denotes the generator vector of the birod macrostructure and $y(s)$ the internal coordinates for the microstructure, as presented respectively in (3.1.23) and (3.1.28) and where the notion of birod internal energy has been discussed in (3.2.57). Similarly to bichains, the matrix $\mathcal{K}(s) \in \mathbb{R}^{18\times 18}$ is called the *birod stiffness matrix* and the configuration associated with the variables $(\hat{\xi}(s), \hat{y}(s))$ the *birod ground state configuration* $(\hat{\boldsymbol{g}}(s), \hat{\mathcal{P}}(s))$.

In order to define a quadratic birod energy in (5.3.3) consistent, in the sense of (4.2.12), with the sequence dependent bichain energy in (5.2.3) we propose to use the deterministic exponential Cauchy-Born rule described in section 4.2. Precisely, the birod



(a) λ (b) CF (c) T_L

Figure 5.13: Birod ground state configuration for three different sequences of length 240 base pairs, as used in the deterministic exponential Cauchy-Born rule for bichains.

ground state is obtained using the expression in (4.2.11) and the birod stiffness matrix using (4.2.14). Figures 5.13a, 5.13b and 5.13c show the ground state configuration for the sequence λ , CF and T_L . We note that they look very similar to the bichain ones in 5.4a, 5.4b and 5.4c since they are birod interpolations.

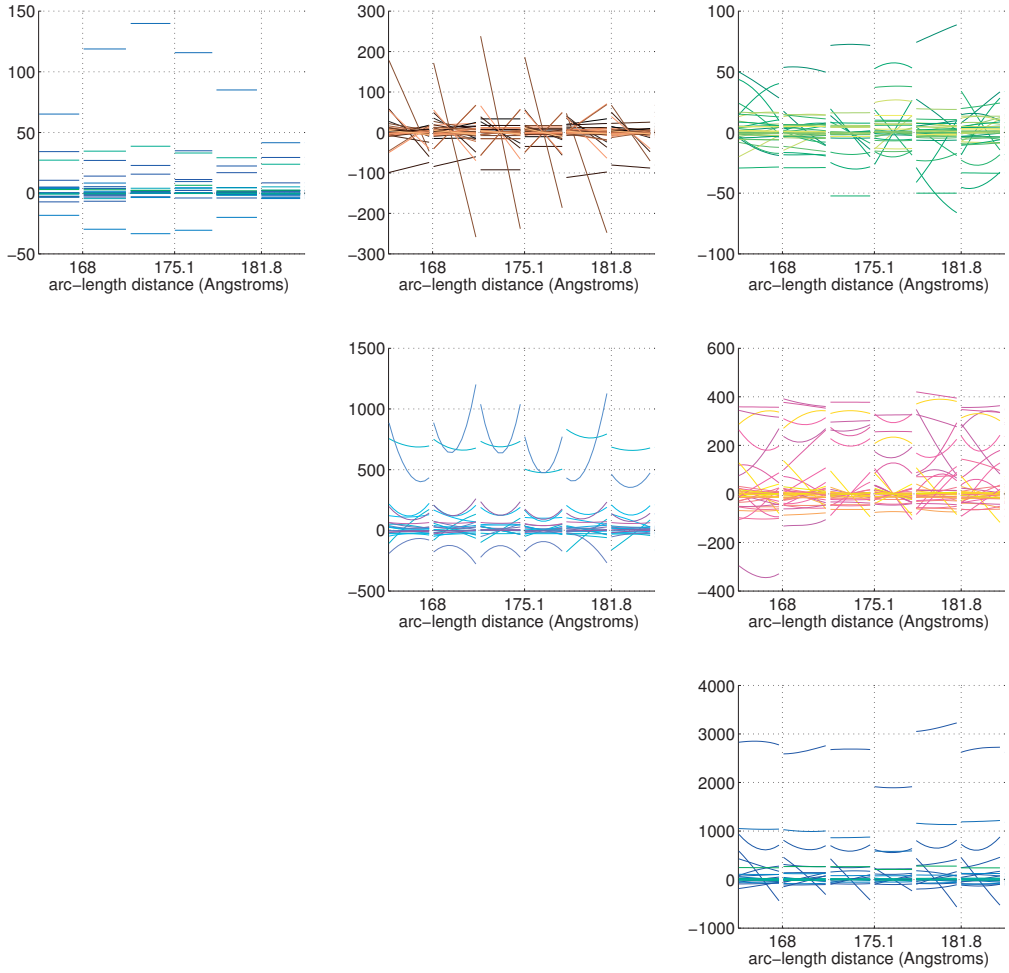


Figure 5.14: Zoom on the entries of the birod stiffness matrices predicted by the deterministic exponential Cauchy-Born rule for the sequence λ . The plots shows the entries from the 50^{th} to 56^{th} base pair. More details are discussed in the text.

The stiffness matrix is illustrated for the sequence λ in figure 5.14 on the scale of a few base pairs and in figure 5.15 on the scale of hundreds of base pairs. Similarly to the coefficients shown in figures 5.9 and 5.11 obtained using a stochastic Cauchy-Born rule for chains, the entries in the stiffness matrix are observed to be only piece wise continuous between each base pair and a strong heterogenous character appears at larger scales. Each subfigure in 5.14 and 5.15 shows the entries in a 6×6 sub-block of the

matrix $\mathcal{K}(s)$. This 6×6 sub-block structure is the one corresponding to the natural decomposition into the variables $y(s) \in \mathbb{R}^6$, $\frac{d}{ds}y(s) \in \mathbb{R}^6$ and $\xi(s) \in \mathbb{R}^6$ in 5.3.3.

We should remark that the energy formulation in (5.3.3) differs slightly from the more general one in (3.2.57) because of the presence of the two *pointwise energy terms* associated respectively to the *pointwise stiffness matrices* $\mathcal{K}_0$ and $\mathcal{K}_L$ at the two ends. These terms have been added to make the construction of a consistent continuum energy for the nearest-neighbor bichain energy in (5.2.5) more symmetric. They also emphasize the particular mechanical behaviour of the last and first microstructure configuration in a bichain as has been observed in DNA modelling [Gonzalez et al., 2013].

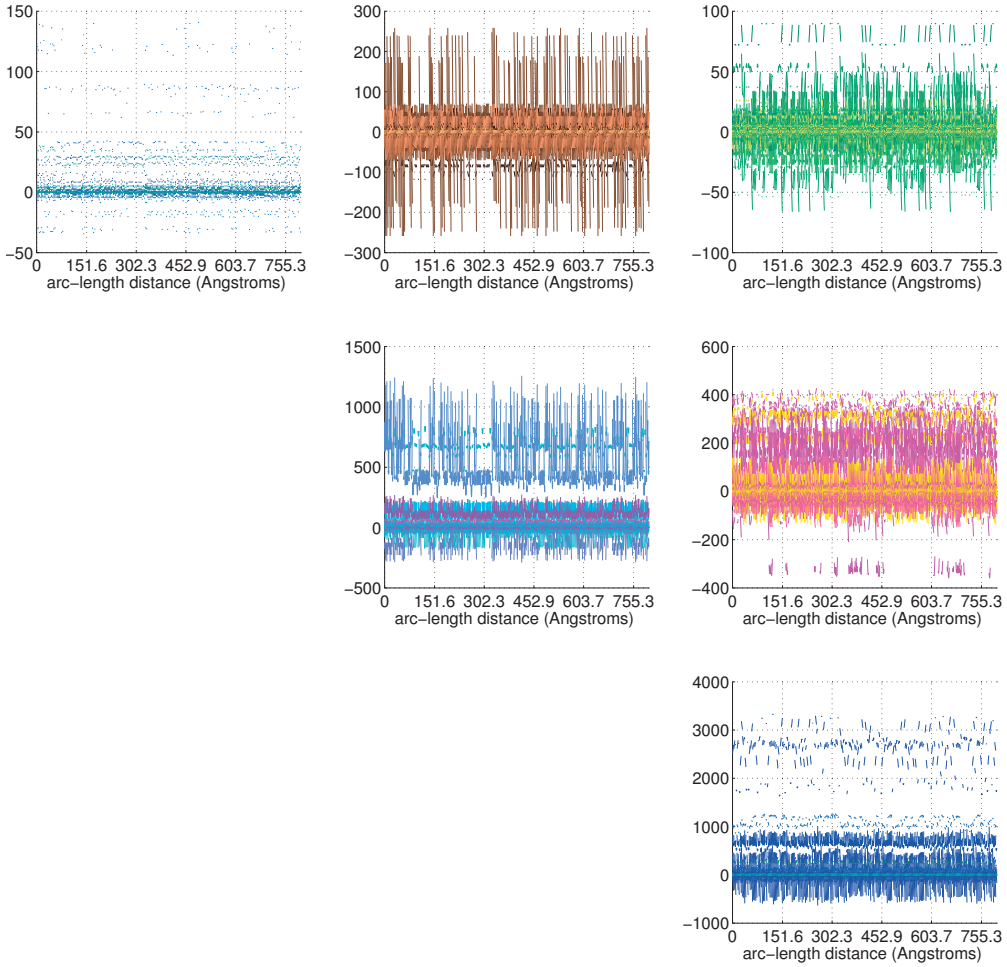


Figure 5.15: Entries of the birod stiffness matrix predicted by the deterministic exponential Cauchy-Born rule for the sequence λ . The plots shows the entries from the 1^{st} to 240^{th} base pair. More details are discussed in the text.

Another point to make concerns the assumption made in the formulation (5.3.3)

that the shift in the quadratic contribution of the derivative $\frac{d}{ds}y(s)$ is the derivative of the shift in the contribution of the variable $y(s)$. In general, this assumption is very restrictive since otherwise one cannot ensure that the energy density can be pointwise minimised. However, in the quadratic case it can be shown that one can always transform the energy into this particular form, with $(\hat{\xi}(s), \hat{y}(s))$ being the birod variables associated to the absolute minimiser of E, since it can be proven that any other equivalent quadratic energy differs from this one by a *null Lagrangian*. More details about this transformation can be found in appendix A.4.

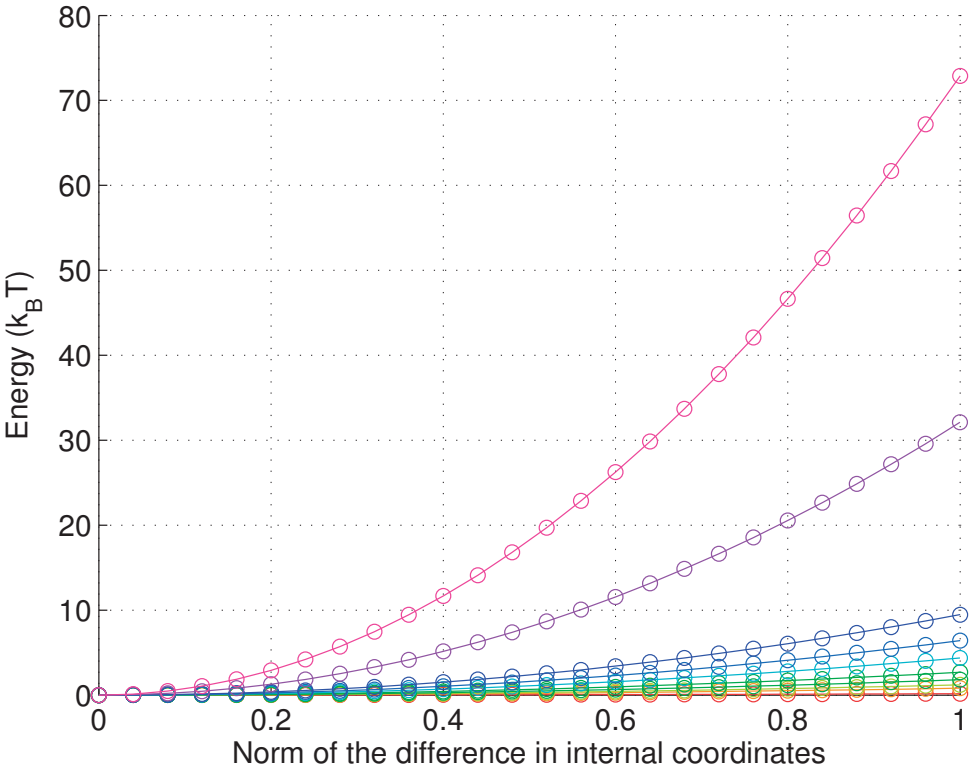


Figure 5.16: Energy Comparison between the deterministic exponential Cauchy-Born rule (−) and the bichain energy (∘) for the sequence λ . More details are discussed in the text.

The main value of the local birod energy in (5.3.3) is that, in contrast to the local rod energy, it can be seen as a continuum version of the rigid base energy in (5.2.3) and therefore has some chance to provide a realistic continuum description of DNA physics. In order to quantify the quality of the deterministic exponential Cauchy-Born rule used to define the ground state configuration $(\hat{\boldsymbol{g}}(s), \hat{\mathcal{P}}(s))$, drawn in figure 5.13a for instance, and the entries in the stiffness matrix, drawn in figure 5.15 for instance, appearing in the birod energy (5.3.3), we have computed the energy E given by the original discrete model of

[Gonzalez et al., 2013, Petkeviciute et al., 2014] in (5.2.5) for a family of deformed bichain configuration and the analogous energy given by the continuum model in (5.3.3). More precisely, we have compared these two energy definitions for the sequence λ and for ten sets of deformations which are built by marching along the eigenvectors of the bichain stiffness matrix K in (5.2.3), chosen to correspond to the ten deciles of its eigenvalue spectrum. Figure 5.16 shows the values obtained for the discrete, using the symbol $\circ$, and continuum description, using a continuous line, as a function of the norm of the difference in internal coordinates $|(x, y) - (\hat{x}, \hat{y})|$. It is not easy to define a way to measure whether a deformation between a given configuration and the bichain ground state is large or not, but energies of order of $70\, k_B T$ are thought to be large for a DNA fragment.

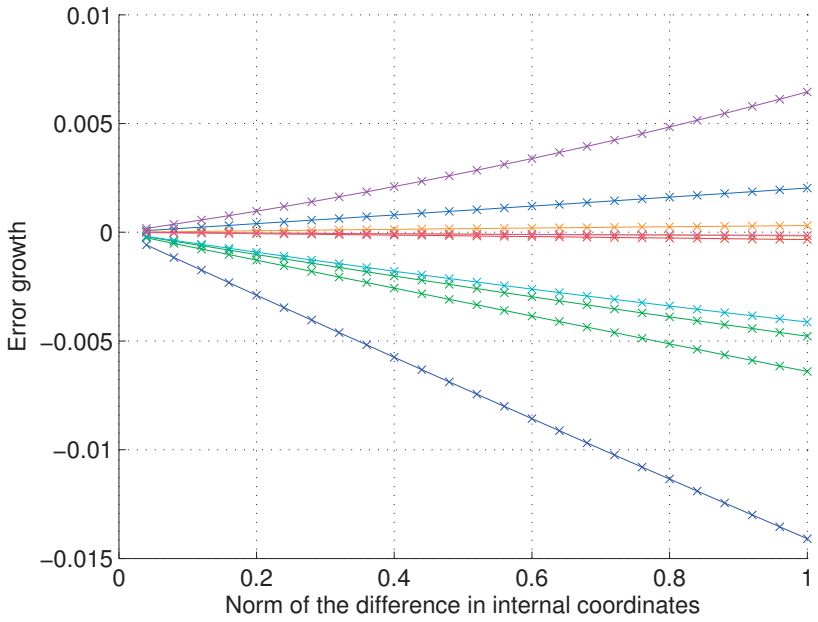


Figure 5.17: Estimation of the truncation error for the deterministic exponential Cauchy-Born rule for the sequence λ . More details are discussed in the text.

As presented in section 4.2, the deterministic exponential Cauchy-Born rule has been built by keeping the quadratic terms in an expansion of the birod continuum energy in terms of the bichain internal coordinates. The numerical evaluation of the discrete and continuum energy then allows estimation of the truncation error between these two definitions. According to the expression (4.2.12), we expect the difference to grow like $\mathrm{O}\left(N\,|(x, y) - (\hat{x}, \hat{y})|^3\right)$, or perhaps more intelligibly, that

$$\frac{E(\boldsymbol{g}(x), \mathcal{P}(y)) - \hat{E}[\boldsymbol{g}^{(N)}, \mathcal{P}(y^{(N)})]}{|(x, y) - (\hat{x}, \hat{y})|^2} = \mathrm{O}\left(N\,|x - \hat{x}|\right). \tag{5.3.4}$$

Figure 5.17 shows the value of the left hand side of (5.3.4) for the ten sets of deformations used in figure 5.16 as a function on the norm $|(x, y) - (\hat{x}, \hat{y})|$. Finally, to provide a more global picture of the accuracy of the proposed deterministic exponential Cauchy-Born rule for bichains, we have computed the relative error between discrete and continuum definitions for 2500 random configurations sampled from the distribution corresponding to the sequence λ in (5.2.2). Figure 5.18 shows a histogram of the relative errors that are obtained. Even if it is quite satisfactory that they are all less than 0.5% in absolute value, a finer understanding of this deterministic exponential Cauchy-Born would be of interest.

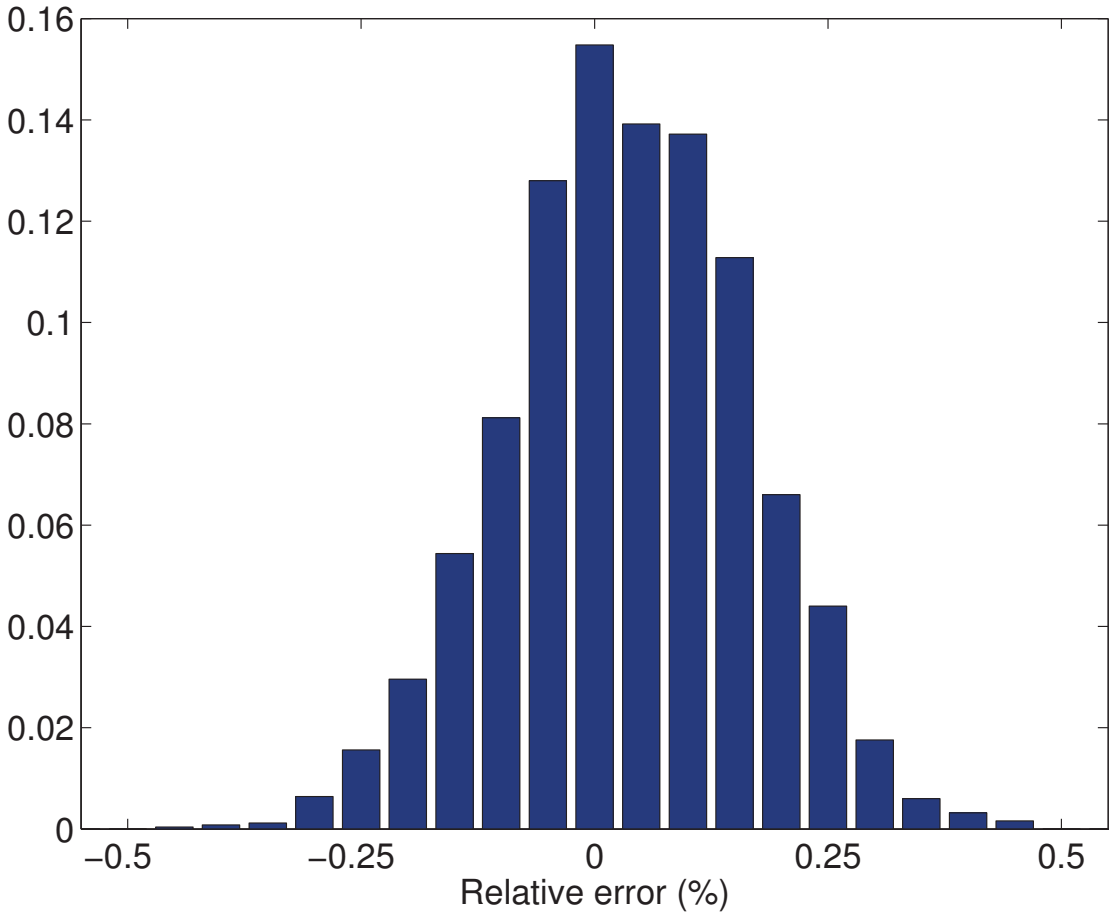


Figure 5.18: Histogram of relative errors associated to the deterministic exponential Cauchy-Born rule for the 2500 random configurations of the sequence λ . More details are discussed in the text.

6. On the Multi-scale Modelling of Most Probable DNA Configurations

The calculations of energy and molecular dynamics go beyond the present work and will hopefully lead to an evaluation of the potential function for local conformational distortions of DNA so that the isotropic model can be supplanted.

J. A. Schellman & S. C. Harvey, 1995.

This sixth chapter discusses the computation of most probable configurations respecting end conditions in the sequence dependent rigid base model. In section 6.1, we present why such configurations are believed to be an interesting first step in the characterization of the sequence dependent statistical physics properties of a DNA fragment. However, since the second variation still requires further investigations, we shall restrict the discussion to stationary configurations. In order to demonstrate that the sequence dependent continuum birod model provides accurate approximations of the stationary bichain configurations, we first present the numerical solutions of the continuum equilibrium conditions and then use them in an optimization method to compute the solution of the discrete equilibrium conditions. In section 6.2, we present the Hamiltonian formulation of the birod equilibrium conditions and discuss the sequence dependent coefficients obtained by the deterministic exponential Cauchy-Born rule for bichains. We then illustrate the equilibrium configurations for the first 240 base pairs of the sequences λ , CF and T_L for under and over twisted end conditions, corresponding respectively to -300pNÅ and $+300$pNÅ, computed using the software bBDNA. The different mechanical response corresponding to different sequences is significant and, even if a more detailed analysis is required, the effect of the ground state geometry appear to be important in each case. The results are then compared to their discrete analogues in section 6.3. We present the sequence dependent bichain equilibrium conditions numerically obtained for the rigid base model. Although the comparison of the discrete and continuum solution is satisfactory, these results are only preliminary and a finer analysis of the ability of the continuum solutions to reproduce bichain stationary configurations remains to be investigated.

6.1 Computing Most Probable DNA Configurations

Understanding general properties of the bichain configurational distribution in (5.2.2) as a function of the DNA sequence is a difficult task because the configuration space is large and the relation between the bichain internal coordinates (x, y) and the associated bichain configuration $(\boldsymbol{g}, \mathcal{P})$ is non-linear. In particular, in the case of lengths of biological interest i.e. a few hundreds of base pairs, the dimension of the configuration space is of order 10^3.

A first approach of interest to characterise potential physical differences between different sequences, or to recognise special sequences with special mechanical properties, is to study the *most probable configuration* achieved subject to prescribed end conditions. This approach has been mentioned previously in the context of DNA in [Olson, 1996, Becker, 2007] and is applied using numerical methods to idealised DNA models in [Manning et al., 1996, Zhang and Crothers, 2003, Cotta-Ramusino and Maddocks, 2010] in order to study *cyclisation j-factor* values for instance. We propose in particular to discuss the case where the first and last rigid body configuration in the bichain macrostructure $\boldsymbol{g}$ are prescribed, but where the end microstructure configurations $\mathcal{P}_1$ and $\mathcal{P}_N$ are free. This set of boundary conditions is thought to provide a reasonable compromise between the detailed end values needed to describe mathematically a deformed DNA configuration and the lack of physical knowledge on what are interesting values for the microstructure end conditions. According to the specific functional form of the bichain configurational distribution in (5.2.2), we observe that the most probable configurations respecting given boundary conditions are necessarily minimisers of the function $E(\boldsymbol{g}(x), \mathcal{P}(y))$ in (5.2.3) and, consequently, they have to solve stationary conditions of the form (2.2.45). The computation of such stationary configurations for a realistic sequence dependent model, such as in [Gonzalez et al., 2013, Petkeviciute et al., 2014], has to help in the understanding of the DNA molecule physical properties and, even if not all stationary configurations are minimisers, constitutes a necessary first step. The second variation of the bichain energy is not discussed in this thesis and is left as a future work.

As originally proposed in [Manning et al., 1996], an efficient way to compute numerically stationary configurations of realistic DNA fragments, whose length goes naturally from of a few hundred to several thousand base pairs, is to work with a model where the numerical degrees of freedom can be chosen to be considerably less than the physical ones which typically grow linearly with the number of base pairs. This requirement is at the heart of the objective to design accurate continuum description of DNA molecules. Continuing this work, we propose to solve continuum birod stationary configurations using the Hamiltonian formulation in (3.2.76) parametrized using the sequence dependent coefficients, as described in (5.3.3), coming from the deterministic Cauchy-Born rule proposed in section 4.2. These results are then compared with solutions of the original discrete system by using a root solving method which takes as initial guess the continuum birod configuration sampled at the base positions.

6.2 A Birod Model Formulation

As anticipated, we are interested in the use of a continuum birod model to approximate sequence dependent stationary bichain configurations. More precisely, we are interested in the computation of stationary birod configurations $(\boldsymbol{g}, \mathcal{P}(y))$ with prescribed end macrostructure configurations $\boldsymbol{g}(0) = \mathrm{Id}_4$ and $\boldsymbol{g}(L) = \boldsymbol{g}_N^*$ but with free end microstructure configurations. Formally, these conditions can be stated as

$$\delta E[\boldsymbol{g}, \mathcal{P}(y)]\delta(\boldsymbol{g}, y) = 0 \tag{6.2.1a}$$

for any smooth birod perturbations $\delta(\boldsymbol{g}, y)(s)$ satisfying

$$\delta\boldsymbol{g}(0) = 0 \text{ and } \delta\boldsymbol{g}(L) = 0. \tag{6.2.1b}$$

In section 4.1.2, we discussed briefly why for a consistent birod internal energy, in the sense of (4.1.15), and for Dirichlet conditions sufficiently close to the ground state end configuration, the continuum stationary configuration is expected to provide an approximation to the desired bichain configuration. A similar argument can be made for the mixed end conditions used in this case. We recall that the quality of the consistency between the discrete and continuum energy definition in the context of DNA modelling is illustrated in figure 5.18.

As discussed in more detail in sections 3.2.7 and 3.2.8, the birod configuration $(\boldsymbol{g}, \mathcal{P}(y))$ making stationary an internal energy of the form (5.3.3) with respect to some prescribed end configurations, can be written in a Hamiltonian form. More precisely, we first introduce the compact notation

$$\mathrm{z}(s) = (\zeta(s), \boldsymbol{g}(s), \zeta_y^{\mathcal{P}}(s), y(s)) \tag{6.2.2}$$

where $\zeta(s) = (\mathsf{m}(s), \mathsf{n}(s)) \in \mathbb{R}^6$ stands for the local torque and stress acting on the birod macrostructure and is the conjugate variable to the macrostructure configuration $\boldsymbol{g}(s) = (\boldsymbol{R}(s), \boldsymbol{r}(s)) \in \mathrm{SE}(3)$, and similarly $\zeta_y^{\mathcal{P}}(s) = (\mathrm{m}^{\mathcal{P}}(s), \mathsf{n}^{\mathcal{P}}(s)) \in \mathbb{R}^6$ is the conjugate to the birod microstructure coordinates $y(s) = (\eta(s), \mathsf{w}(s)) \in \mathbb{R}^6$. It yields the following *sequence dependent birod Hamiltonian system*

$$\frac{d}{ds}\mathrm{z}(s) = J\left(\mathrm{z}(s)\right)\partial_{\mathrm{z}}\mathrm{H}(\mathrm{z}(s); s) \tag{6.2.3a}$$

$$\begin{aligned} \boldsymbol{g}(0) = \mathrm{Id}_4,\ \boldsymbol{g}(L) = {\boldsymbol{g}_N}^* \\ \mathcal{K}_0(y(0) - \hat{y}(0)) - \zeta_y^{\mathcal{P}}(0) = 0 \text{ and } \mathcal{K}_L(y(L) - \hat{y}(L)) + \zeta_y^{\mathcal{P}}(L) = 0 \end{aligned} \tag{6.2.3b}$$

with

$$J\left(\mathrm{z}(s)\right) = \begin{pmatrix} \mathsf{J}_{\boldsymbol{g}}\left(\zeta(s), \boldsymbol{g}(s)\right) & 0 \\ 0 & \begin{pmatrix} 0 & -\mathrm{Id}_6 \\ \mathrm{Id}_6 & 0 \end{pmatrix} \end{pmatrix} \tag{6.2.3c}$$

where the operator $\mathsf{J}_{\boldsymbol{g}}(\,.\,)$ is defined in (3.2.25), the boundary conditions (6.2.3b) have been chosen consistently with (6.2.1b) and the matrices $\mathcal{K}_0$ and $\mathcal{K}_L$ come from the point wise energy terms in (5.3.3). The function H denotes the *sequence dependent birod Hamiltonian*. Moreover, since the internal energy in (5.3.3) is quadratic, and according to the definition of the *Legendre transform* in (3.2.73), the birod Hamiltonian can easily be shown to be of the form

$$\mathrm{H}(y,\zeta_y^{\mathcal{P}},\zeta;s)=\frac{1}{2}\begin{pmatrix} y-\hat{y}(s)\\ \zeta_y^{\mathcal{P}}\\ \zeta\end{pmatrix}\cdot\mathcal{H}(s)\begin{pmatrix} y-\hat{y}(s)\\ \zeta_y^{\mathcal{P}}\\ \zeta\end{pmatrix}+\begin{pmatrix}\zeta_y^{\mathcal{P}}\\ \zeta\end{pmatrix}\cdot\begin{pmatrix}\frac{d}{ds}\hat{y}(s)\\ \hat{\xi}(s)\end{pmatrix} \tag{6.2.4a}$$

where the matrix $\mathcal{H}(s)\in\mathbb{R}^{18\times 18}$ is explicitly given as a function of the birod stiffness matrix $\mathcal{K}(s)$ in (5.3.3) as

$$\mathcal{H}=\begin{pmatrix}\mathcal{K}_2\mathcal{K}_3^{-1}\mathcal{K}_2^T-\mathcal{K}_1 & -\mathcal{K}_2\mathcal{K}_3^{-1}\\ -\mathcal{K}_3^{-1}\mathcal{K}_2^T & \mathcal{K}_3^{-1}\end{pmatrix} \tag{6.2.4b}$$

where we have introduced the block structure

$$\mathcal{K}=\begin{pmatrix}\mathcal{K}_1 & \mathcal{K}_2\\ \mathcal{K}_2^T & \mathcal{K}_3\end{pmatrix} \tag{6.2.5}$$

with $\mathcal{K}_1\in\mathbb{R}^{6\times 6}$, $\mathcal{K}_2\in\mathbb{R}^{6\times 12}$ and $\mathcal{K}_3\in\mathbb{R}^{12\times 12}$. Consequently, a sequence dependent birod Hamiltonian, together with its symplectic structure in (6.2.3a) can be analytically deduced from the sequence dependent rigid base model in [Gonzalez et al., 2013, Petkeviciute et al., 2014] using the determinist Cauchy-Born rule presented in section 4.2. We recall that sequence dependent birod ground state configurations $(\hat{\boldsymbol{g}}(s),\mathcal{P}(\hat{y}(s)))$ and the corresponding birod stiffness matrix $\mathcal{K}(s)$ are illustrated for the sequence λ in figure 5.13 and 5.15 respectively. The entries in the matrix $\mathcal{H}(s)$, as defined in (6.2.4b), are shown in figure 6.1 using an analogous 6×6 block decomposition to figure 5.15 but with respect to the variables y, $\zeta_y^{\mathcal{P}}$ and ζ.

The implementation of an efficient numerical method to solve the Hamiltonian system in (6.2.3), as well as a general discussion of some solution sets for different given DNA sequences, requires a further significant investment of effort and constitutes a large part of the doctoral thesis[1] of *J. Glowacki*. Based on the work of [Manning et al., 1996], an equivalent system to (6.2.3) involving a *quaternion parametrization* of the rotation matrix $\boldsymbol{R}(s)$ has been derived and a *continuation method* is used to produce sets of stationary sequence dependent birod configurations. Moreover, significant effort has been invested in the design of interactive software tools allowing useful representations of these solutions sets. The resulting software *bBDNA* allows the computation and the visualisation of solutions to the equilibrium birod conditions for several end conditions.[2] Figure 6.2 shows solutions computed for the first 240 base pairs of the sequence λ using this software and

[1] in preparation

[2] This software should be available for free downloading very soon. The interested reader should contact his author Jaroslaw Glowacki for more information.

the Hamiltonian matrix illustrated in figure 6.1. The middle configuration is obtained by the application of a pure extension load on the ground state configuration and correspond to a force of 0.75 picoNewton (pN). We have then applied pure twist loads corresponding respectively to -300pNÅ and $+300$pNÅ. The left configuration is then under twisted, with respect to intrinsic right handedness of B-DNA fragments, and the right one is over twisted. The colour map along the strands shows the norm of the local total couple $|\mathsf{m}|\,(s)$ around the birod macrostructure position $\boldsymbol{r}(s)$. Figures 6.3 and 6.4 show analogous configurations (same loading values) for the sequence CF and T_L respectively. The sequences λ , CF and T_L were detailed in section 5.2.2.

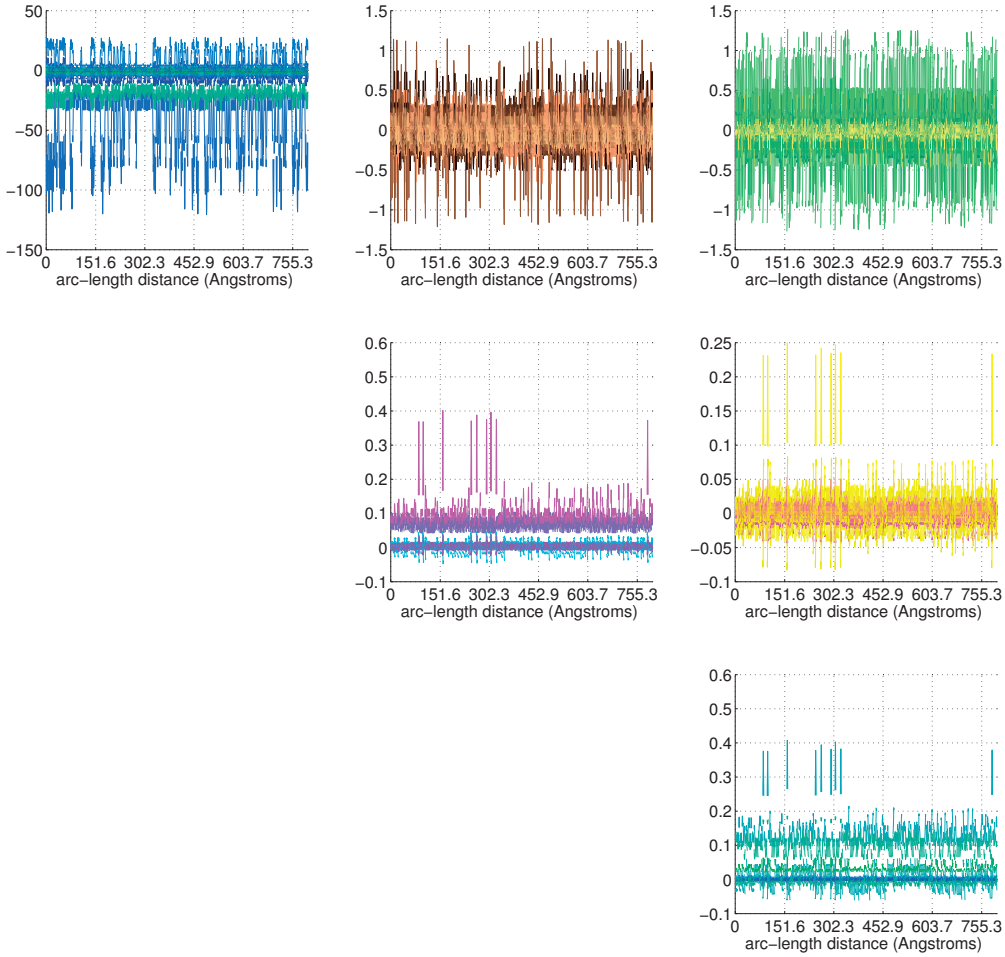


Figure 6.1: Entries of the birod Hamiltonian matrix predicted by the deterministic Cauchy-Born rule for the sequence λ . The plots shows the entries from the 1^{st} to 240^{th} base pair. More detail is discussed in the text.

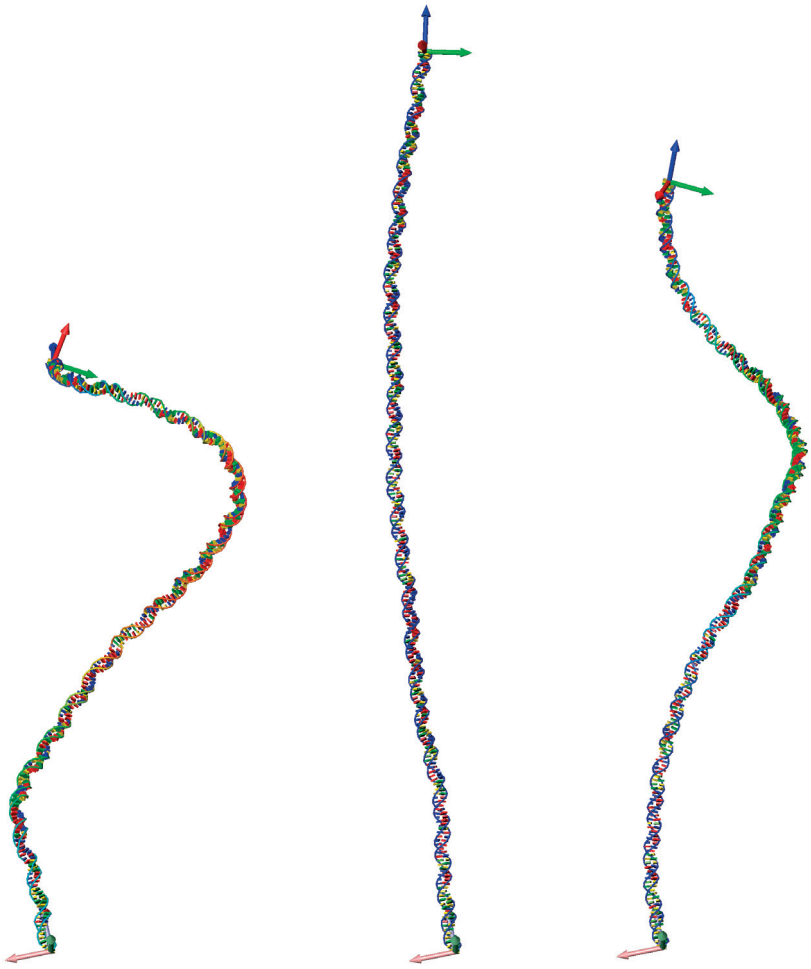


Figure 6.2: Deformed continuum birod configurations for the sequence λ . The left configuration has been under twisted, the middle one is slightly pulled and the right one has been over twisted. The colour map along the strands shows the norm of the local total couple around the base pair position. More detail is discussed in the text.

In all three cases, the difference in the mechanical response between under and over twisting is evident. We observe, according to the colour map along the strands, that the local couple seems to be reach higher values for the under twisted λ fragment, in figure 6.2, whereas the converse situation appears for the T_L fragment, in figure 6.4. This could be explained by the fact that the λ unstressed configuration, drawn in figure 5.13a, is roughly straight and it is then its intrinsic right handedness which is deformed. In contrast the T_L unstressed configuration has been designed to have a pronounced left handed super-helix, as drawn in figure 5.13c, and it consequently seems that it is its super-

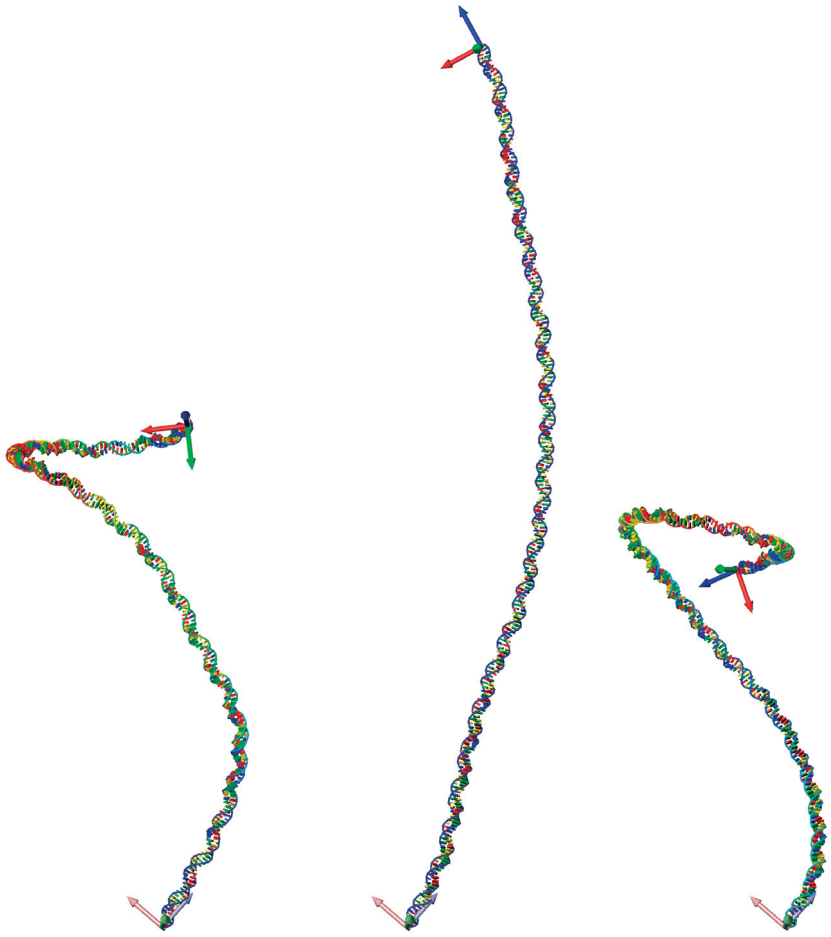


Figure 6.3: Deformed continuum birod configurations for the sequence *CF* . The left configuration has been under twisted, the middle one is slightly pulled and the right one has been over twisted. The colour map along the strands shows the norm of the local total couple around the base pair position. More detail is discussed in the text.

helical structure which reacts the most to the applied deformation. The sequence *CF* , even if it not completely clear from the figure 6.3, has very interesting three dimensional structures when it has been under and over twisted. They are probably due the interplay between the large intrinsic bends in its unstressed configuration, drawn in figure 5.13b, and the applied external couples. A more detailed analysis on the sequence dependent mechanical response is not pursued here and is left for a future work. We merely remark that these examples illustrate that different sequences respond in markedly different ways.

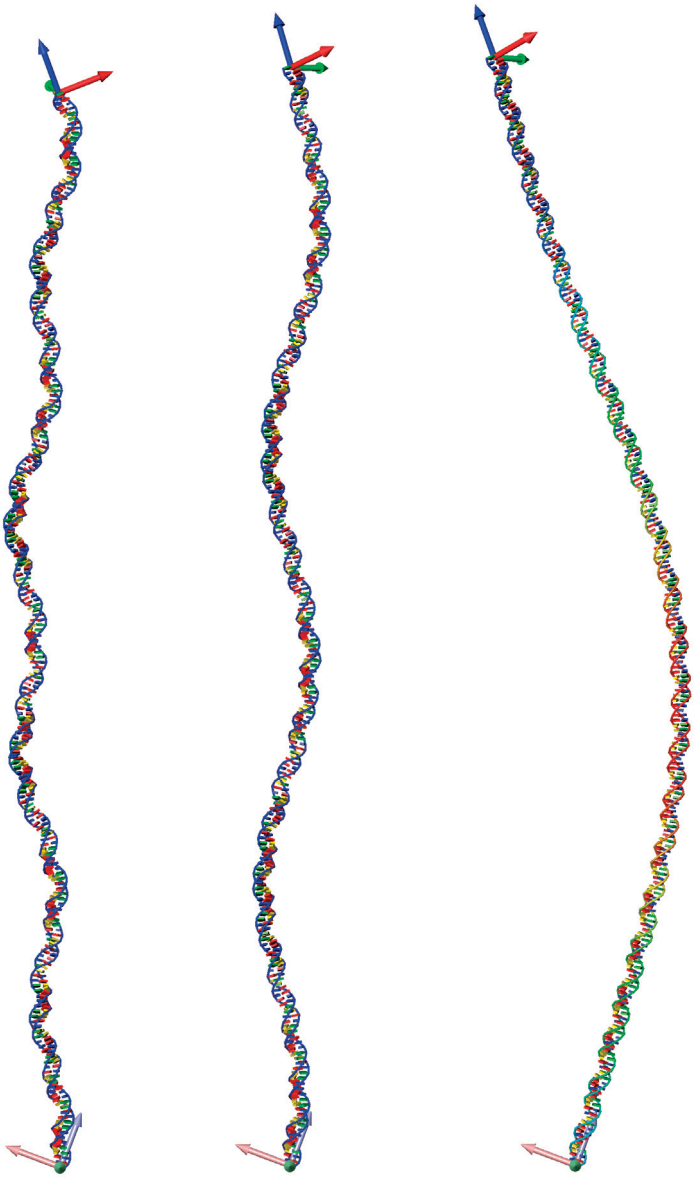


Figure 6.4: Deformed continuum birod configurations for the sequence T_L . The left configuration has been under twisted, the middle one is slightly pulled and the right one has been over twisted. The colour map along the strands shows the norm of the local total couple around the base pair position. More detail is discussed in the text.

6.3 A Rigid Base Model Formulation

We would like now to assess whether the sequence dependent birod equilibrium configurations, computed using the software bBDNA, provide accurate approximations of the analogous bichain equilibrium configurations for the rigid base model. One approach is to solve numerically the bichain sequence dependent equilibrium conditions, and to compare the resulting solution with the corresponding birod configuration evaluated at base pair positions.

A discussion on this problem has been made in section 2.2.6 and stationary conditions have been deduced for a general energy bichain E. However, in the specific case of the quadratic energy appearing in (5.2.3) and for the specific end conditions used in the continuum computation, i.e. prescribed end macrostructure configurations to be Id_4 and ${\boldsymbol{g}_N}^*$ respectively and free microstructure configurations, we obtain explicitly

$$K\begin{pmatrix} y_1-\hat{y}_1 \\ x_1-\hat{x}_1 \\ y_2-\hat{y}_2 \\ x_2-\hat{x}_2 \\ \cdots \\ y_{N-1}-\hat{y}_{N-1} \\ x_{N-1}-\hat{x}_{N-1} \\ y_N-\hat{y}_N \end{pmatrix} - \begin{pmatrix} 0 \\ {\mathbb{L}^{(1)}_{x_1}}^T \mathrm{Ad}^T_{\boldsymbol{g}_1}\begin{pmatrix}\boldsymbol{\lambda^R}\\ \boldsymbol{\lambda^r}\end{pmatrix} \\ 0 \\ {\mathbb{L}^{(1)}_{x_2}}^T \mathrm{Ad}^T_{\boldsymbol{g}_2}\begin{pmatrix}\boldsymbol{\lambda^R}\\ \boldsymbol{\lambda^r}\end{pmatrix} \\ \cdots \\ 0 \\ {\mathbb{L}^{(1)}_{x_{N-1}}}^T \mathrm{Ad}^T_{\boldsymbol{g}_{N-1}}\begin{pmatrix}\boldsymbol{\lambda^R}\\ \boldsymbol{\lambda^r}\end{pmatrix} \\ 0 \end{pmatrix} = 0 \tag{6.3.1a}$$

$$\boldsymbol{g}_1 = \mathrm{Id}_4,\ \boldsymbol{g}_N = \prod_{n=1}^{N-1} a(x_n) = {\boldsymbol{g}_N}^*. \tag{6.3.1b}$$

where we have used the expressions in (2.2.51) and (2.2.53) together with the identities in (2.2.49) and where the matrix $\mathbb{L}^{(1)}_{x_n}$ is explicitly given in (2.2.20) for the bichain coordinates used in [Gonzalez et al., 2013, Petkeviciute et al., 2014]. The system (6.3.1) is then made up of $12N$ equations and unknowns, which are the $12N-6$ bichain internal coordinates (x, y) plus 6 Lagrange multipliers $(\boldsymbol{\lambda^R}, \boldsymbol{\lambda^r})$. Note that the Lagrange multipliers can be interpreted as the couple and force needed to enforce the Dirichlet conditions according to (2.2.58), and that all the non-linear dependencies are in the matrices $\mathrm{Ad}_{\boldsymbol{g}_n}$ which come from the geometric constraints, induced by the Dirichlet conditions, on the internal coordinates.

Numerical solutions of these bichain equilibrium conditions have been computed by *A. Patelli* using the function *fsolve* provided by the software *matlab*. This method is based on a Newton algorithm and therefore requires an initial approximation of internal coordinates and Lagrange multipliers. According to the discussion made in section 4.1.2 about the consistency of discrete and continuum stationary configurations, natural initial

approximations are given by the evaluation of the associated birod continuum configuration at base pair positions and by the associated external couple and force, as described in (3.2.83a). Figure 6.5 shows the bichain internal coordinates for the under twisted sequence dependent birod solution, using the symbol $-$, appearing in figure 6.2 and the corresponding solution of the bichain system (6.3.1), using the symbol $\circ$. The first two plots show respectively the components of the intra bichain coordinates $y_n = (\eta_n, \mathsf{w}_n)$ as a function of the base pair distance, and the two others shows respectively the entry of the inter bichain coordinates $x_n = (u_n, v_n)$. The analogous bichain internal coordinates for the over twisted configuration are shown in figure 6.6. We also present in figures 6.7 and 6.8 the corresponding plots for the under twisted configuration of the sequence CF and T_L respectively[3]. We note that the characteristic periodicity in figure 6.8 reflects the 10 base pairs periodicity of the sequence T_L . Their associated three dimensional birod configurations are shown in figure 6.3 and 6.4 respectively. Even if these results are satisfactory, they are, unfortunately, only preliminary. A finer analysis of the ability of the continuum solutions to reproduce bichain stationary configurations has to be investigated.

An important remark is the following. To solve the Hamiltonian system in 6.2.3 a very dense mesh is needed because of the strong heterogeneous character of the coefficients matrix illustrated in figure 6.1. More precisely, the number of nodes used in the numerical discretization is considerably higher than the number of base pairs. Consequently, even if these continuum computations are very fast and moreover provide an initial configuration for the discrete system, the interest of the continuum model is then questionable. We believe however the definition of this sequence dependent continuum model to be valuable since it provides the necessary first step in the definition of rigorous large scale DNA models, and, consequently, in the construction of the multi-scale modelling of DNA statistical physics. We are confident that homogenisation techniques, such as the one presented in section 7.3.2, will allow this issue to be overcame and to define an sequence dependent continuum model with more slowly varying coefficients.

[3]The analogous figures for the over twisted configurations are omitted for compactness.

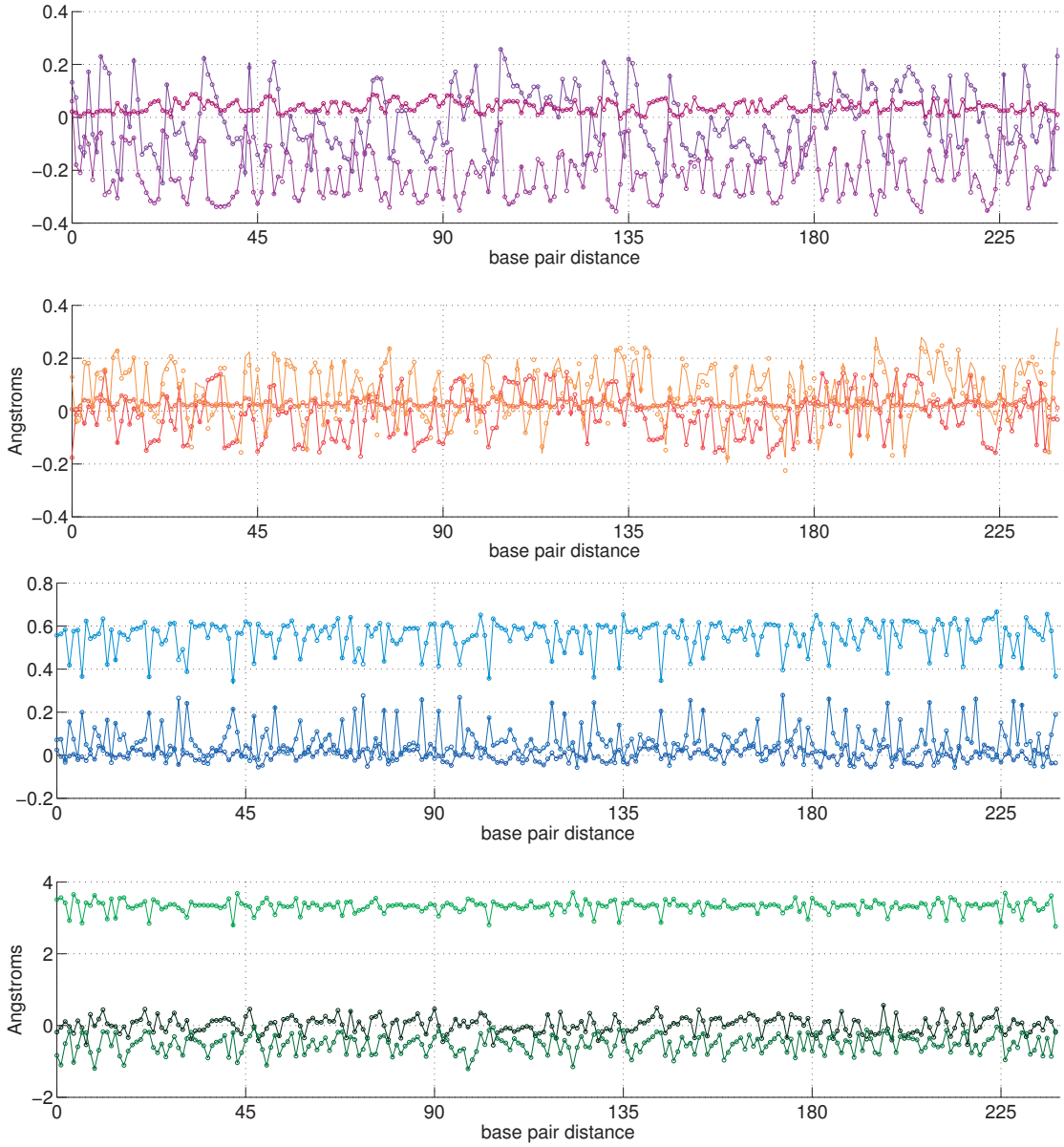


Figure 6.5: Comparison of the bichain internal coordinates of an under twisted configuration for the sequence λ computed with the discrete ($\circ$) and continuum (-) equilibrium conditions. Each plot shows from top to bottom the entries of the triple η, w, u and v as a function of base pair distance.

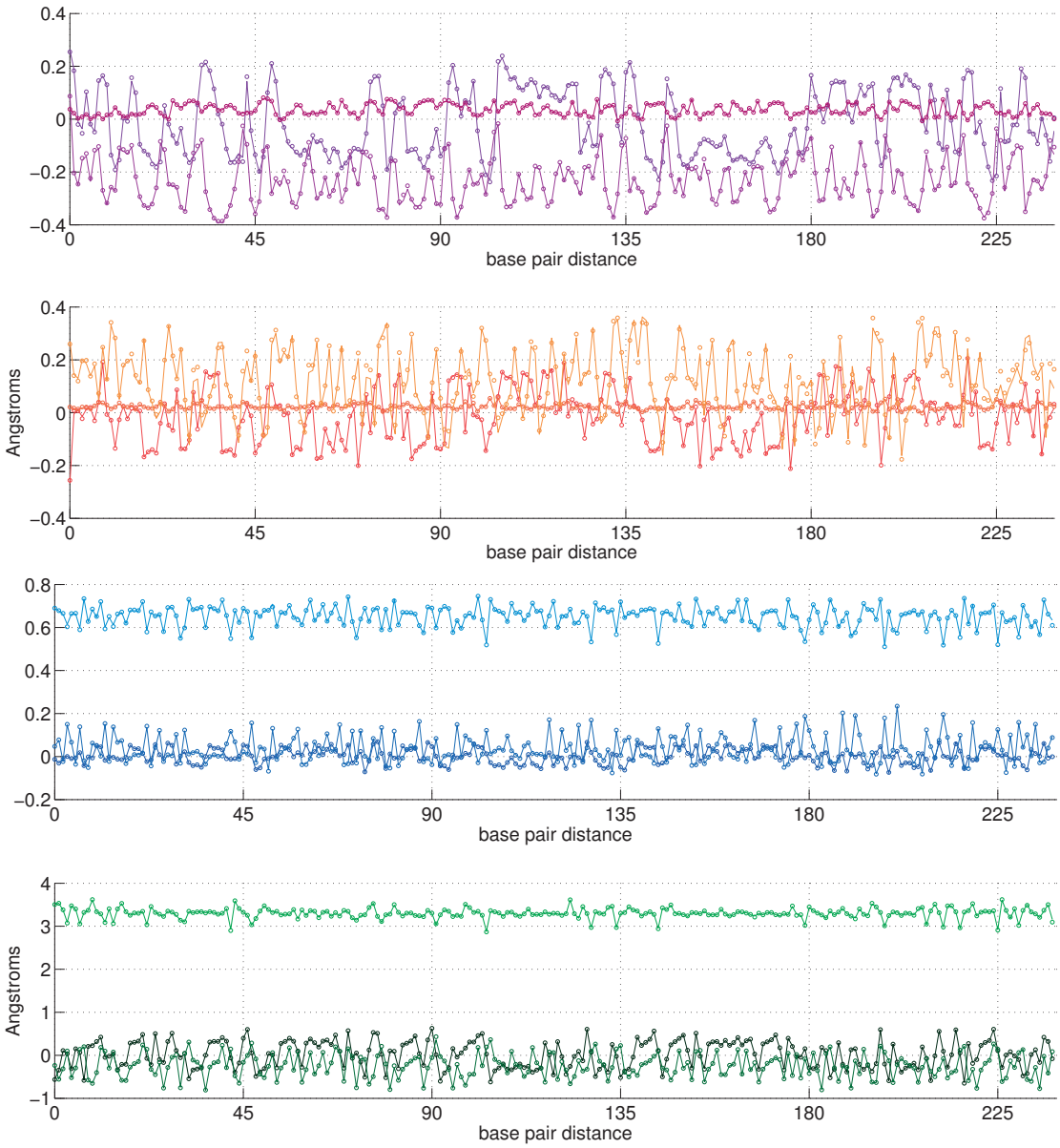


Figure 6.6: Comparison of the bichain internal coordinates of an over twisted configuration for the sequence λ computed with the discrete ($\circ$) and continuum (-) equilibrium conditions. Each plot shows from top to bottom the entries of the triple η, w, u and v as a function of base pair distance.

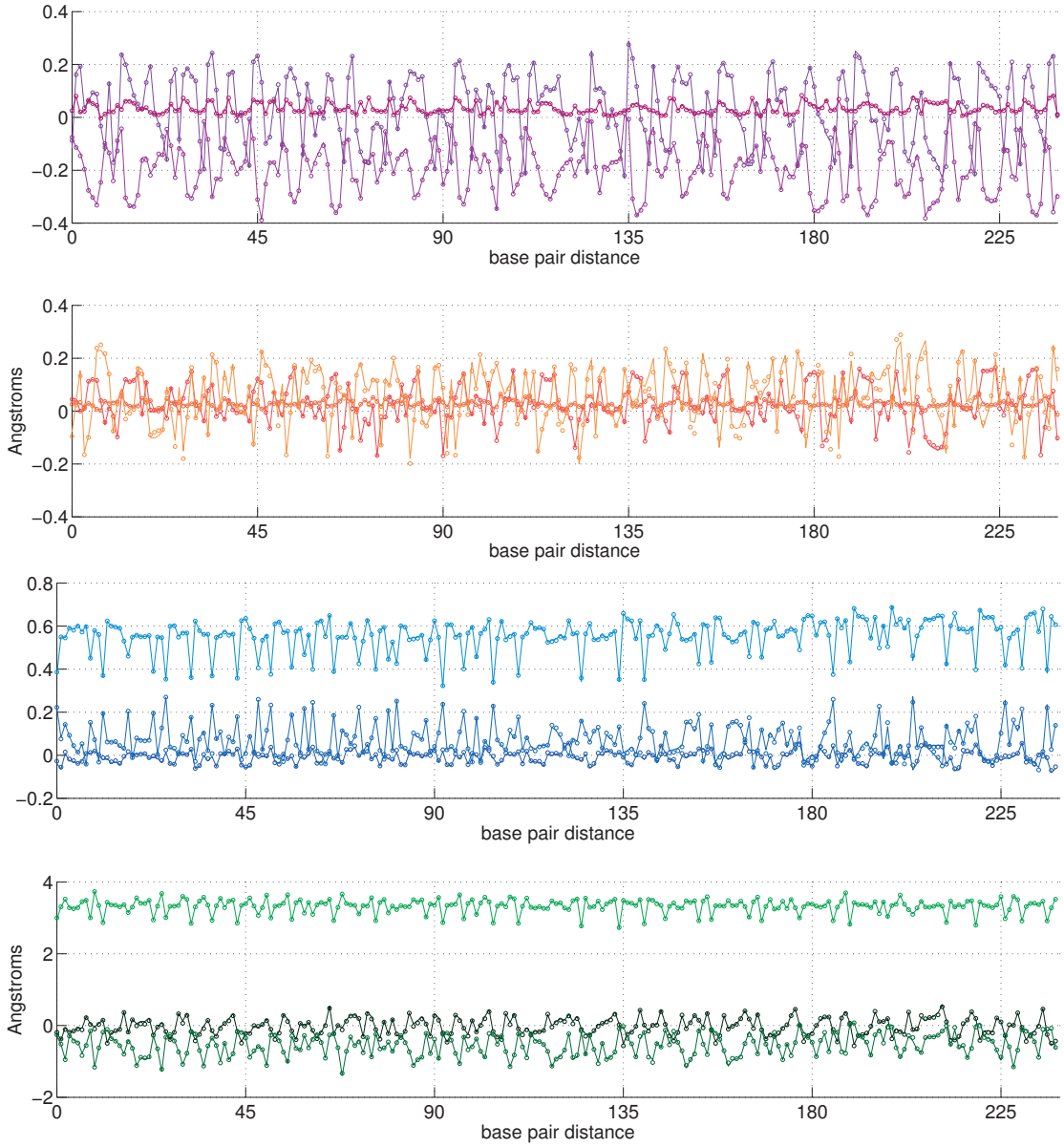


Figure 6.7: Comparison of the bichain internal coordinates of an under twisted configuration for the sequence *CF* computed with the discrete (∘) and continuum (-) equilibrium conditions. Each plot shows from top to bottom the entries of the triple η, w, u and v as a function of base pair distance.

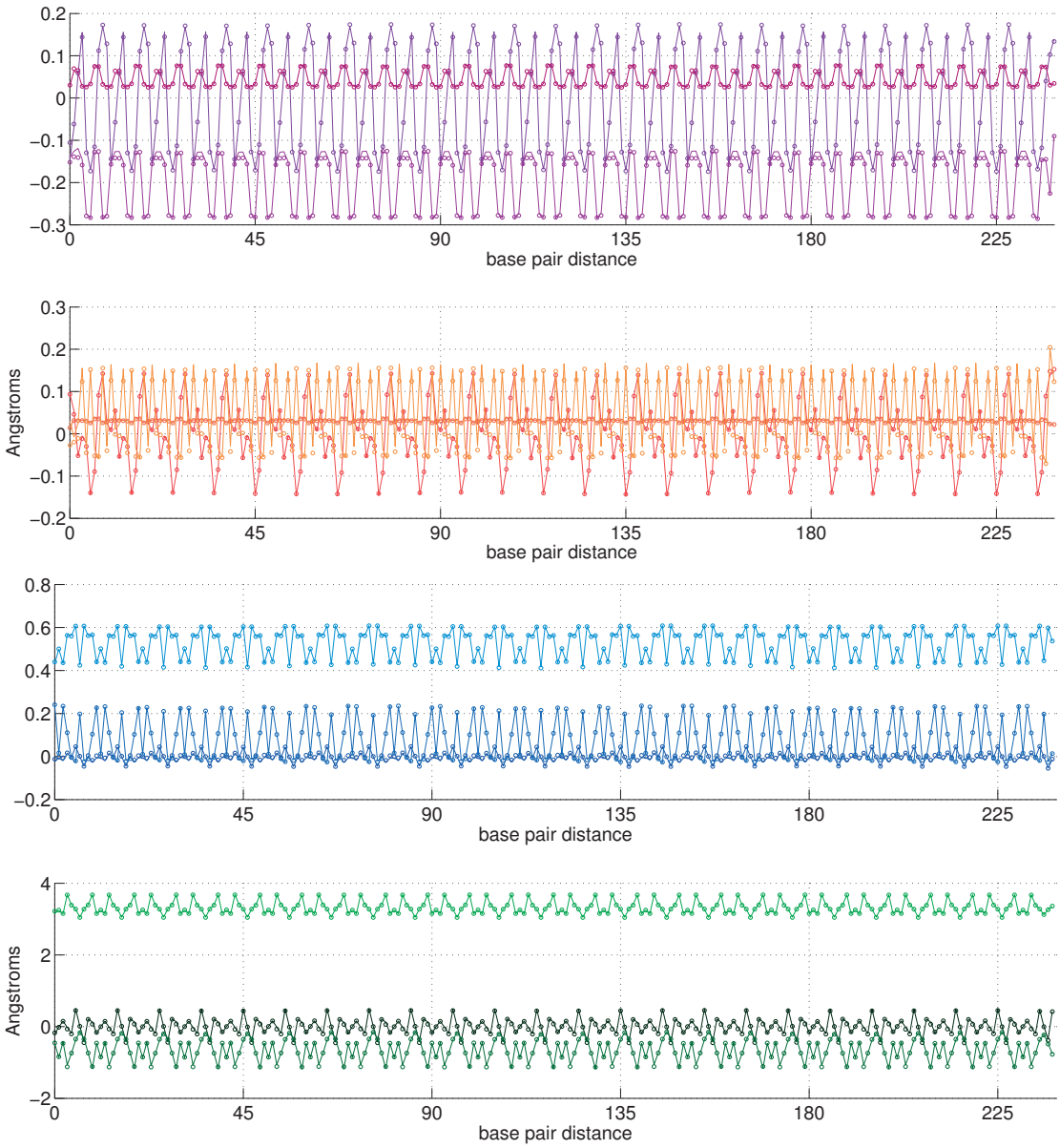


Figure 6.8: Comparison of the bichain internal coordinate of an under twisted configuration for the sequence T_L computed with the discrete (∘) and continuum (-) equilibrium conditions. Each plot shows from top to bottom the entries of the triple η, w, u and v as a function of base pair distance.

7. On the Multi-scale Modelling of DNA configurational first moment

The local structure of the DNA double helix was assumed to be that of idealised B-DNA, with the helix axis following a straight line. It is now known that this latter assumption is an over simplification.

E. N. Trifonov, R. K.-Z. Tan & S. C. Harvey, 1988.

This last chapter discusses the features of the configurational first moment for the sequence dependent rigid base pair model as well as its approximation in a continuum description. In section 7.1 we recall the importance of this chain observable since it comprises the frame correlation matrix and the Flory persistence vector which are both of interest to characterise the possible constitutive anisotropy and possible non-trivial intrinsic geometry of the DNA molecule. We discuss the limitation of the existing models, in particular of the isotropic Wormlike chain, to describe sequence dependent properties. The value of the explicit approximations for the configurational first moment of heterogeneous chains and rods is then emphasised and, to quantify their accuracy, we compare numerically their predictions with results obtained by a Monte Carlo simulation. However, these results are only valid for local chains and local rod, so that we only use the rigid base pair model for the discussion. In section 7.2, we concentrate on the chain formulation and the results of the recurrence relation for the configurational first moment deduced in section 2.4 are compared with Monte Carlo simulations for the sequences λ , CF and T_L . In each case the differences are hardly perceptible, even after 500 base pairs, and the explicit formula are consequently very satisfactory. The factorization into the expected chain and the persistence matrix is then illustrated. Remarkably, it shows that the fast variations in the configurational first moment reflect the expected chain configuration whereas the slow convergence of the entries is due to the persistence matrix. This factorization is then, in a mathematical sense, a generalisation of the studies in [Trifonov et al., 1988, Schellman and Harvey, 1995] of the effect of localised intrinsic bends in chain statistics. We present also the results of the short length expansion, which is a closed form expression and which is observed to be valid until a distance of about 135 base pairs. In section 7.3 we apply the exponential stochastic Cauchy-Born rule to discuss the continuum approximations of the configurational first moment of chains. We describe the sequence dependent continuum rod parameters, obtained from the rigid

base pair model, and the ODE governing the rod configurational first moment, deduced in section 3.4. The numerical solutions of this ODE are compared with the results of Monte Carlo simulations obtained for the sequence λ and, as for the discrete approximation, the difference between the values is hardly perceptible. The exponential stochastic Cauchy-Born rule for chains, together with the proposed ODE, are then thought to be very satisfactory and to constitute a significant step towards the multi-scale modelling of DNA statistical physics properties. However, the strong heterogeneous character of this formulation forces the number of mesh point of the numerical discretization used to solve the ODE, to be significantly larger than the actual of number of base pairs which lead to a pragmatic objection to the use of the continuum model. In order to overcome this fundamental issue, we propose the application of a homogenisation technique for ODE systems which possess two distinct characteristic scales. Since the coefficients do not exhibit any of the classical hypotheses, such as periodicity or ergodicity, we only obtain a local averaging principle. Even if the optimal size of the window used to average should be investigated more, the choice of about 30Å is observed to deliver very smooth and slowly varying coefficients. The solution of the homogenised system is then compared to the Monte Carlo simulation for the sequence λ . Not only is the difference is hardly perceptible but also the number of numerical mesh points are now significantly less than the number of base pairs of the original chain. The results, even if they require a more detailed analysis, are believed to be encouraging in the multi-scale modelling of DNA statistical physics and in the definition of sequence dependent large scale models.

7.1 On the configurational first moment in DNA mechanics

In both discrete and continuum modelling of DNA molecules, the first moment of some configurational observables has been intensively studied and has been used to describe their general statistical properties and to discuss the existence of exceptional sequences. A brief presentation of these observables has been made in section 2.4 for the discrete approach and in section 3.4 for the continuum one. In the following discussion we will focus on the *Flory persistence vector* and on the *frame correlation matrix*.

The Flory persistence vector, which is denoted $\langle \boldsymbol{r}_{m,n} \rangle \in \mathbb{R}^3$ or $\langle \boldsymbol{r}(s',s) \rangle \in \mathbb{R}^3$, is defined as the expectation of the relative position of two sites in the molecule, indexed by m and n in the discrete setting and by s' and s in the continuum one. It generalises the *Kratky-Porod persistence length* [Porod, 1948, Krakty and Porod, 1949, Flory, 1973, Schellman and Harvey, 1995]. Analogously, the frame correlation matrix is denoted as $\langle \boldsymbol{R}_{m,n} \rangle \in \mathbb{R}^{3\times3}$ or $\langle \boldsymbol{R}(s',s) \rangle \in \mathbb{R}^{3\times3}$ and is defined as the expectation of the relative rotation between two elements in the molecule which generalises the *tangent-tangent correlation* [Landau and Lifshitz, 1959, Schellman, 1974, Maroun and Olson, 1988, Panyukov and Rabin, 2000]. Using the definitions presented in sections 2.4 and 3.4, the study of the Flory persistence vector and of the frame cross-correlations is then encompassed in the analysis of the first moment of the relative rigid body displacement, or *configurational first moment*, denoted by $\langle \boldsymbol{g}_{m,n} \rangle \in \mathbb{R}^{4\times4}$ or $\langle \boldsymbol{g}(s',s) \rangle \in \mathbb{R}^{4\times4}$. This point of view has been used already in [Becker, 2007].

The main interest of these two observables is that they encode the *possible constitutive anisotropy* as well as the *possible non-trivial intrinsic geometry* of the molecule. A large confusion is, however, still observed today about the notion of persistence length and tangent-tangent correlation and it is important for us to make here a very clear point. The most popular model used to describe the discrete or continuum DNA statistical physics is with no doubt the *Wormlike chain* model. This model is defined by a quadratic energy density with only one bending stiffness parameter and where the minimum energy configuration is straight. In such circumstances, it can be shown that the Kratky-Porod persistence length and the tangent-tangent correlation are simple functions of the bending parameter [Landau and Lifshitz, 1959, Doi and Edwards, 1986]. The application of this model on a large ensemble of long (a few kilo base pairs) DNA molecules to discuss large scale effective properties can be justified arguing that the fine details of each fragment constitution is, somehow, averaged. The confusion starts when the Wormlike chain is applied to an ensemble of reasonably short fragments and of similar constitutions because, in this situation, both the hypotheses of uniform stiffness distribution and of intrinsically straight ground state are strongly violated. For instance, it has been observed experimentally that some regions of specific sequences are intrinsically bent [Marini et al., 1982, Kabsch et al., 1982, Kitchin et al., 1986, Levene et al., 1986, Bednar et al., 1995, Vologodskaia and Vologodskii, 2002]. In such cases, since the Wormlike chain model no longer applies, the values of the Kratky-Porod persistence length and the tangent-tangent correlation can be very different, as it is shown in the *Helical hinge* model [Schellman, 1980] or in the *Helical Wormlike chain* model

for instance [Yamakawa, 1997]. In order to take into account the possible existence of local bends in the minimum energy shape of the molecule, the *wedge model* has been presented [Trifonov and Sussman, 1980] which allows the introduction of the notion of *static persistence length* and *dynamic persistence length* [Trifonov et al., 1988] which are designed to reflect the contributions of the intrinsic geometry of the molecule and of the thermal fluctuation on the total statistical persistence length. A finer mathematical analysis on the influence of localised bends in the ground state configuration as well as the sequence dependent molecule composition has been done in [Schellman and Harvey, 1995], they show how the tangent-tangent correlation can be factorized into a static part and a dynamic part under specific material symmetries.

Nowadays, we believe that realistic DNA models, at the scale of hundreds of base pairs, are given by a *sequence dependent rigid base model* and, to some extent, by a *sequence dependent rigid base pair model*. These models are described in sections 5.2.2 and 5.2.1 respectively. The Flory persistence vector $\langle \boldsymbol{r}_{m,n} \rangle$ and the frame correlation matrix $\langle \boldsymbol{R}_{m,n} \rangle$ have already been studied in [Maroun and Olson, 1988] using Monte Carlo simulations on a sequence dependent rigid base pair model. Using a DNA sequence made of periodic repeats of A_5G_5, which has the property to have a superhelical structure in its ground state, they show that the three components of the persistence vector reflect the superhelical periodicity and converge to different and non-zero values, and that diagonal entries in the frame correlation matrix also indicates the intrinsic superhelical structure and are asymptotically vanishing.

In sections 2.4 and 3.4 it has been shown how in both rigid body chain and continuum rod models explicit analytical expansions for the entries in the Flory persistence vector and in the frame correlation matrix can be built in the *semi-flexible regime* and, moreover, how this leads to the definition of the persistence matrix. In order to assess the accuracy of these expansions in the case of DNA modelling we would like to compare them to the results obtained using Monte Carlo simulations for realistic sequence dependent parameters. However, one of the key simplifying assumptions needed to apply the theory that we have developed is that the first moment $\langle \boldsymbol{g}_{m,n} \rangle$ satisfies (2.4.1) in the discrete setting, and that $\langle \boldsymbol{g}(s', s) \rangle$ satisfies (3.4.1a) in the continuum formulation, and this is not verified for the rigid base model in [Gonzalez et al., 2013, Petkeviciute et al., 2014], as presented in (5.2.5) because of its 18×18 overlapping blocks structure depicted in figure 5.3. For illustration purpose, we will present a best-fit sequence dependent rigid base pair model derived from the rigid base model and use it to demonstrate the quality of the expansions which have been deduced. We should emphasize that even if a rigid base pair model does not reproduce all the complexity of the sequence dependent DNA statistical physics as well as the rigid base model, the rigorous conclusions that it implies allow an original and valuable viewpoint on the rich spectrum of behaviours that a semi-flexible heterogeneous polymer can have.

7.2 A Rigid Base Pair Formulation

At the level of base pairs, a DNA molecule can be described as a rigid body chain configuration $\boldsymbol{g} = (\boldsymbol{g}_1, ..., \boldsymbol{g}_N)$ where N denotes the number of base pairs. As discussed in more detail in section 5.2.1, we assume the existence of a stationary configurational distribution $d\rho(\boldsymbol{g})$ of the form

$$d\rho(\boldsymbol{g}) = \prod_{1 \leq n \leq N} d\rho_n(a_n) \tag{7.2.1a}$$

with

$$d\rho_n(a_n) = \sqrt{\det\left(\frac{1}{2\pi} K_n^{xx}\right)} \exp\left\{-\frac{1}{2}(x_n - \hat{x}_n) \cdot K_n^{xx}(x_n - \hat{x}_n)\right\} dx_n \tag{7.2.1b}$$

where $x_n \in \mathbb{R}^6$ stands for local internal chain coordinates which parametrize the local junction displacement in the form $a_n = a(x_n)$ according to the definition in (2.1.18). In order to define a realistic sequence dependent rigid base pair model of the form in (7.2.1), we use the parameter set and the method presented in [Gonzalez et al., 2013, Petkeviciute et al., 2014] to build a rigid base model of the form in (5.2.3) and we define, for each $n = 1, ..., N$, the triple $\hat{x}_n$ to be the corresponding macrostructure internal coordinates in the first moment $(\hat{x}, \hat{y})$, as in (2.3.18b), and the matrix K_n^{xx} by

$$K_n^{xx} = \left(c_{n,n}^{xx}\right)^{-1} \tag{7.2.2}$$

where $c_{n,n}^{xx}$ denotes the local covariance matrices of the macrostructure internal coordinates as defined in (2.3.19b). This construction can be motivated by a *maximum entropy principle* subject to a given block diagonal sparsity pattern on the sequence dependent rigid base distribution in [Gonzalez et al., 2013, Petkeviciute et al., 2014].

Consequently, the analytic relation for the relative rigid body displacement $\boldsymbol{g}_{m,n} = \boldsymbol{g}_m^{-1}\boldsymbol{g}_n$ for a chain in a semi-flexible regime, presented in section (2.4.2), can be stated as

$$\langle \boldsymbol{g}_{m,n+1} \rangle = \langle \boldsymbol{g}_{m,n} \rangle \left(\mathrm{Id}_4 - \frac{1}{2}\Lambda_n\right) \bar{a}_n + \mathrm{o}(\epsilon) \text{ with } \boldsymbol{g}_{m,m} = \mathrm{Id}_4 \tag{7.2.3a}$$

with

$$\Lambda_n = \begin{pmatrix} \mathrm{Tr}\left(\mathcal{C}_n^{\boldsymbol{RR}}\right) \mathrm{Id}_3 - \mathcal{C}_n^{\boldsymbol{RR}} & \mathrm{Vect}\left(\mathcal{C}_n^{\boldsymbol{Rr}}\right) \\ 0 & 0 \end{pmatrix} \tag{7.2.3b}$$

and where the matrix $\bar{a}_n$ denotes the expected junction displacement, defined in (2.4.6), and $\mathcal{C}_n^{\boldsymbol{RR}}$ and $\mathcal{C}_n^{\boldsymbol{Rr}}$ denote blocks of the local fluctuation matrix $\mathcal{C}_n$, defined in (2.4.16). For a given distribution of the form in (7.2.1) no analytical expression for $\bar{a}_n$ and $\mathcal{C}_n$ is available and numerical methods have been proposed [Becker, 2007, Becker and Everaers, 2007, Long et al., 2012]. However, using again the semi-flexible hypothesis we have

deduced explicit first order approximations for these matrices as a function of the triple $\hat{x}_n$ and the matrix K_n^{xx}. For compactness these expansions are presented in the appendix A.5. We should also observe that we have

$$\epsilon \sim 0.01 \tag{7.2.4}$$

for the parameters in [Gonzalez et al., 2013, Petkeviciute et al., 2014] according to the expression in (2.4.28).

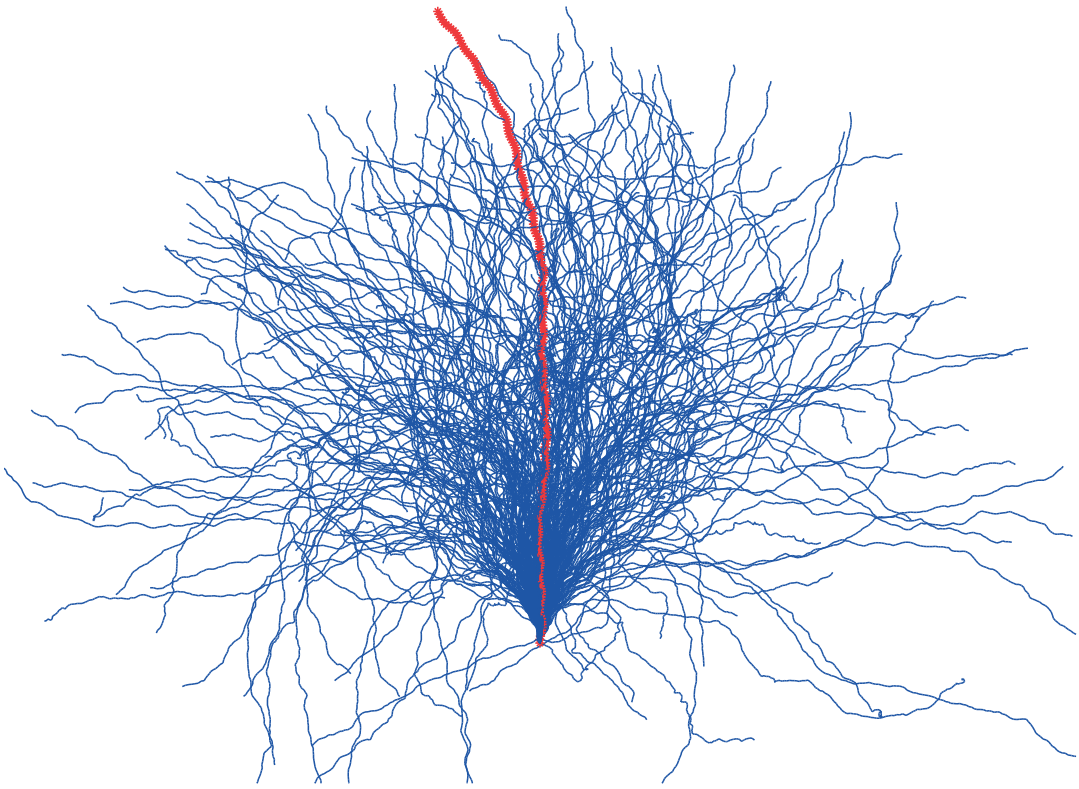


Figure 7.1: Sketch of three dimensional configurations for 250 Monte Carlo samples (in blue) and for the expected chain configuration (in red) for the sequence λ (first 220 base pairs).

7.2.1 Comparison between the analytic recurrence relations and Monte Carlo simulations

We would like to assess the quality of the proposed recurrence relation for the first moment $\langle \boldsymbol{g}_{m,n} \rangle = (\langle \boldsymbol{R}_{m,n} \rangle, \langle \boldsymbol{r}_{m,n} \rangle)$ in (7.2.3) in the case of the rigid base pair coarse grain DNA distribution $d\rho(\boldsymbol{g})$ in (7.2.1). In order to have quantitative estimates we propose to compare numerically the results obtained using the analytic formulas with the analogous ones obtained using a *Monte Carlo method*. More precisely, on one hand, the recurrence relation (7.2.3) can be explicitly evaluated for a given DNA sequence, and, on the other hand, the normal configurational distribution $d\rho(\boldsymbol{g})$ can be densely sampled in order to approximate the entries of $\langle \boldsymbol{g}_{m,n} \rangle$ by an arithmetic average. Figures 7.1 depict in blue the three dimensional shape of 250 Monte Carlo samples and in red the expected chain configuration $\bar{\boldsymbol{g}}_{1,n}$.

Figures 7.2 shows the first column of the frame correlation matrix $\langle \boldsymbol{R}_{1,n} \rangle$ as a function of the base pair index n for the sequence[1] λ . The first line shows the values of the three components $[\langle \boldsymbol{R}_{1,n} \rangle]_{11}$, $[\langle \boldsymbol{R}_{1,n} \rangle]_{21}$ and $[\langle \boldsymbol{R}_{1,n} \rangle]_{31}$ at the scale of 50 base pairs. The second and third line show the same components but at the scale of 150 and 450 base pairs respectively. The analogous entries of second and third columns of the frame correlation matrix $\langle \boldsymbol{R}_{1,n} \rangle$ are drawn in figures 7.3 and 7.4. In all of these plots the results of the Monte Carlo simulation is denoted by the symbol $\circ$ and the results of the recurrence relation in (7.2.3) by $*$ and connected by dashed lines. Similarly, figure 7.5 shows the components of the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$ using the same illustration choices.

The difference between the Monte Carlo simulations and the prediction of the recurrence relation is hardly perceptible. Relative errors are reported in figure 7.6 for the frame correlation matrix $\langle \boldsymbol{R}_{1,n} \rangle$ and for the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$ respectively. The difference for the relative rotation first moment $\langle \boldsymbol{R}_{1,n} \rangle$ reaches approximatively 8% around $n = 500$, however the actual values of the entries are around 10^{-2} which explains why this error is not apparent in figures 7.2, 7.3 and 7.4. In contrast, even if the values of the component in $\langle \boldsymbol{r}_{1,n} \rangle$ can become relatively large, the error stays less than 1%. Consequently, we believe the recurrence relation (7.2.3) to be a very accurate approximation of the chain first moment $\langle \boldsymbol{g}_{1,n} \rangle$. As discussed in section 2.4.3, we observe that the entries in the frame correlation matrix in $\langle \boldsymbol{R}_{1,n} \rangle$ are all asymptotically vanishing but in a non-monotonic fashion for a sufficiently long base pair distance, and that the component of the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$ are similarly converging to finite values. Moreover the frame correlation matrix demonstrates oscillations with a relatively small period of about 11 base pairs, which corresponds to the intrinsic helical twist in B-DNA.

[1] More detail about this sequence is discussed in section 5.2.2.

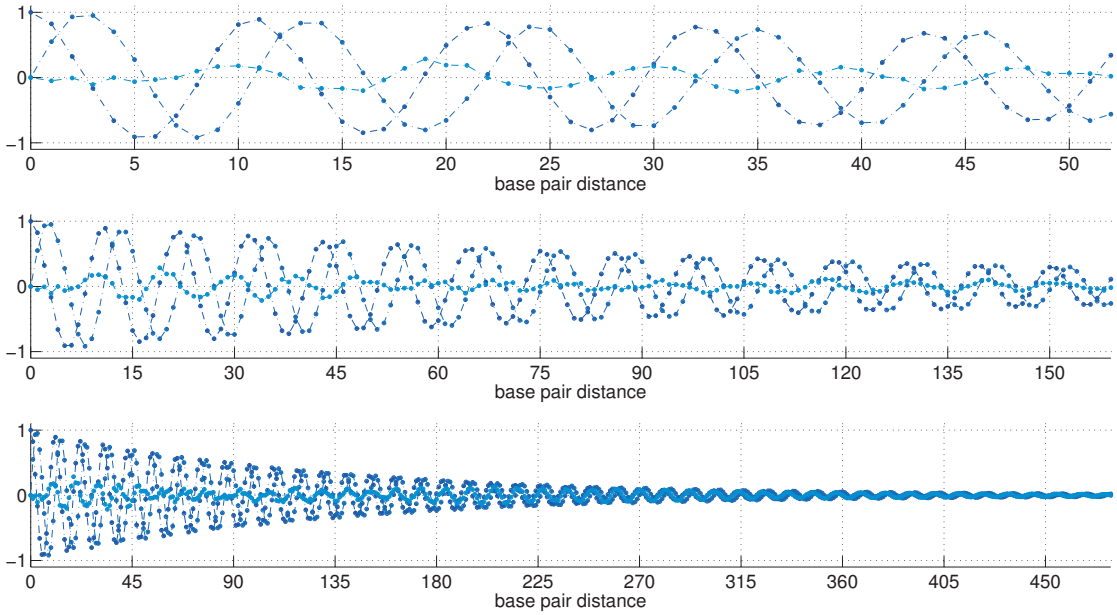


Figure 7.2: First column of the frame correlation matrix $\langle \boldsymbol{R}_{1,n} \rangle$ for the sequence λ and for three different molecule lengths. Monte Carlo simulations ($\circ$) versus the analytical predictions ($*$). More detail is given in the text.

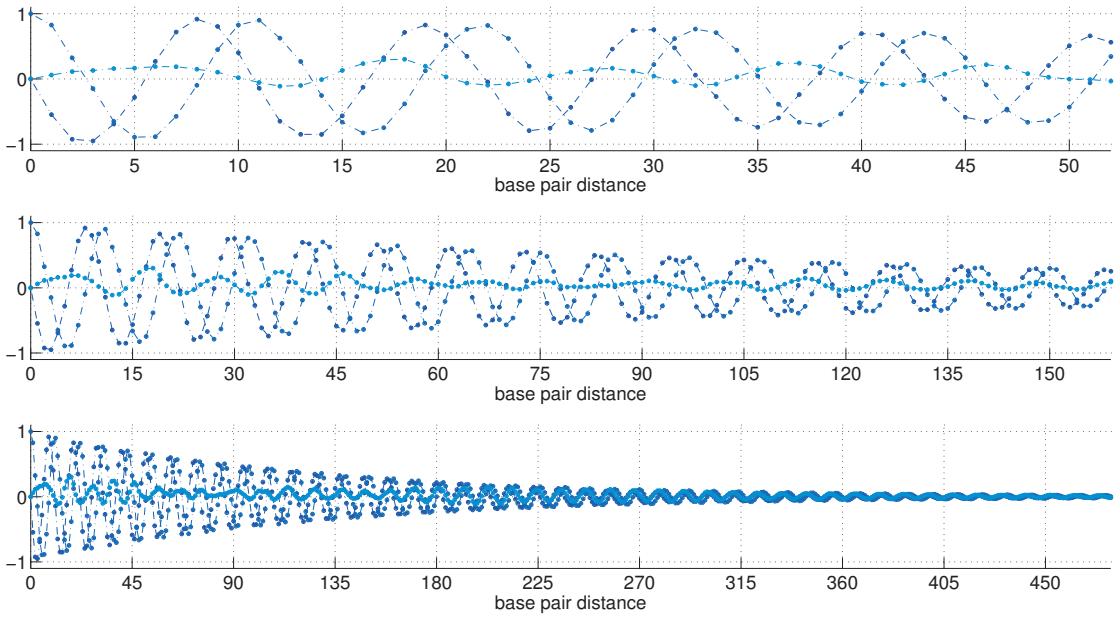


Figure 7.3: Second column of the frame correlation matrix $\langle \boldsymbol{R}_{1,n} \rangle$ for the sequence λ and for three different molecule lengths. Monte Carlo simulations ($\circ$) versus the analytical predictions ($*$). More detail is given in the text.

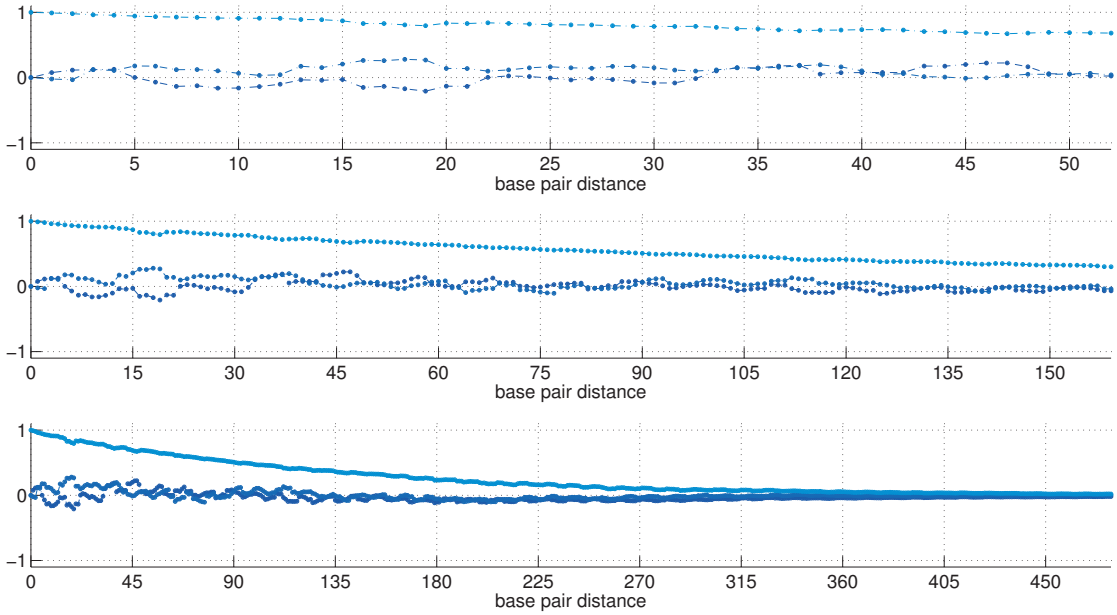


Figure 7.4: Third column of the frame correlation matrix $\langle \boldsymbol{R}_{1,n} \rangle$ for the sequence λ and for three different molecule lengths. Monte Carlo simulations ($\circ$) versus the analytical predictions ($*$). More detail is given in the text.

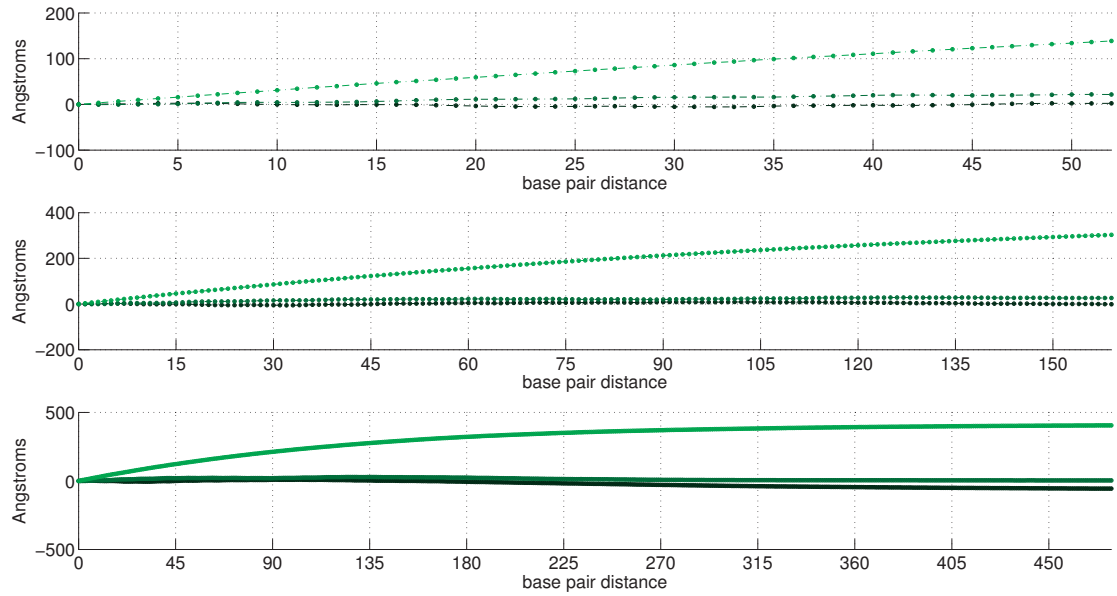


Figure 7.5: Components of the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$ for the sequence λ and for three different molecule lengths. Monte Carlo simulations ($\circ$) versus the analytical predictions ($*$). More detail is given in the text.

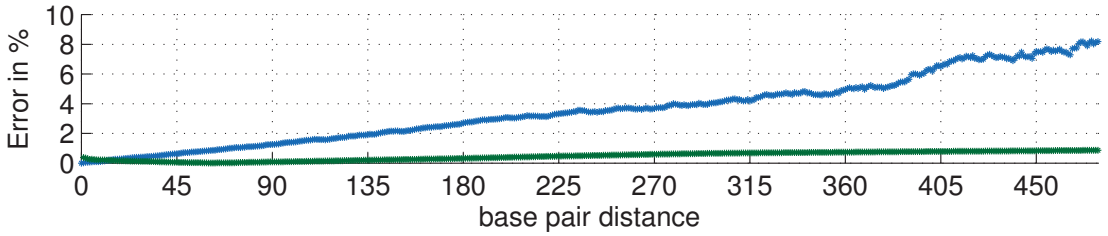
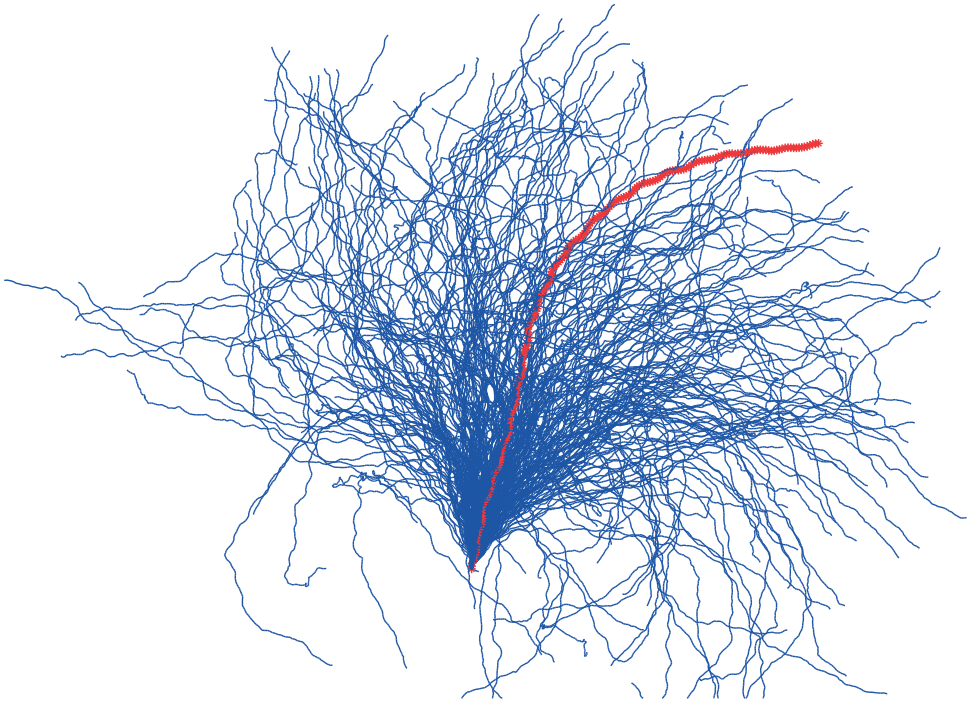


Figure 7.6: Relative errors between Monte Carlo simulations and analytical predictions for the frame correlation matrix $\langle \boldsymbol{R}_{1,n} \rangle$ (blue) and for the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$ (green) associated to the sequence λ as a function of the molecule length. More detail is given in the text.

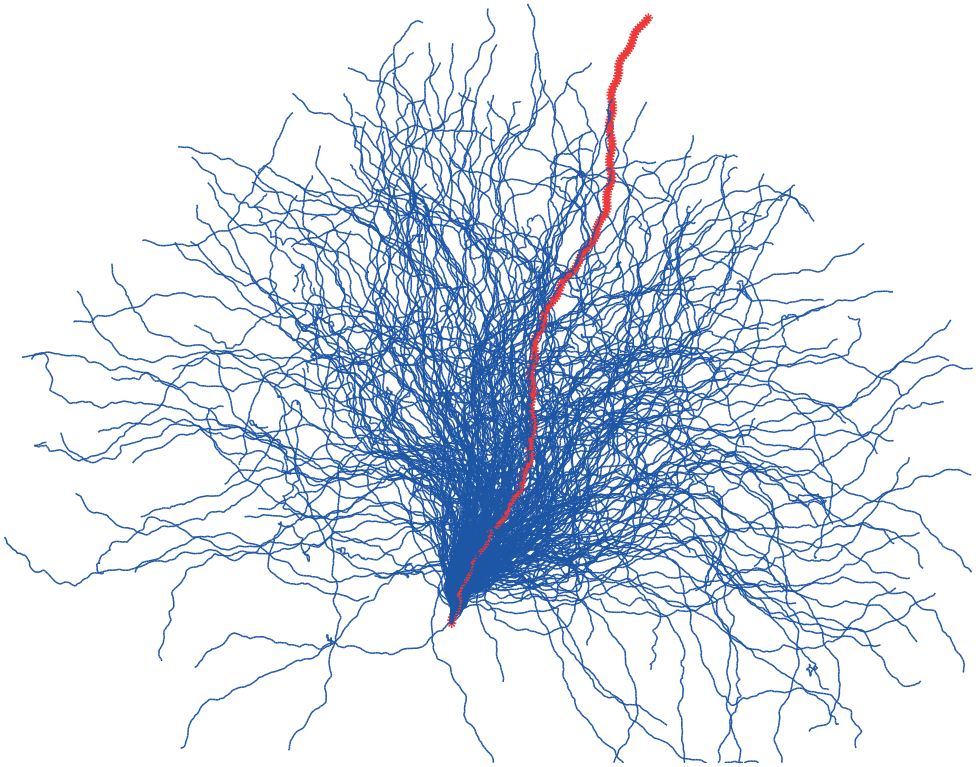
Interestingly, for different sequences we actually observe different behaviours of the configurational first moment $\langle \boldsymbol{g}_{1,n} \rangle$. The figures 7.7a and 7.7b depict in blue the three dimensional shape of 250 Monte Carlo samples and in red the expected chain configuration for the sequences CF and T_L respectively. We recall the sequences λ , CF and T_L are discussed in section 5.2.2. It is interesting to observe the qualitative organisation of the Monte Carlo samples and how this correlates with the expected chain configuration. More precisely, the detailed decay of the frame correlation matrix $\langle \boldsymbol{R}_{1,n} \rangle$ is for instance very different for the sequence λ , CF and T_L . Figure 7.8 shows the third column of the frame correlation matrix $\langle \boldsymbol{R}_{1,n} \rangle$ for the fragments CF and T_L . We note that the component $[\langle \boldsymbol{R}_{1,n} \rangle]_{33}$ is decaying faster for the sequence CF and even becomes slightly negative around $n = 225$, which is intuitively explained by the presence of a large intrinsic bend in this fragment as drawn in figure 5.4b. In contrast, the analogous component for the fragment T_L exhibits oscillations with two distinct periods, the shorter one is about 11 base pairs, and corresponds to the intrinsic B-DNA helicity, and the longer one is about 90 base pairs, and corresponds to the super-helical structure that this particular sequence has in its ground state configuration as shown in figure 5.4c. Concerning the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$, even if the detailed behaviour differs for each distinct sequences λ , CF and T_L , we mainly would like to underline that the limiting values are different, as presented in table 7.1. The ordering of the values for the components $\langle \boldsymbol{r}_{3,500} \rangle$ can be intuitively understood from the respective ground state configurations in 5.4, but it is difficult to give to the rest of the table such a naive interpretation. The different limits in these Flory persistence vectors are of importance because they supports the idea that realistic sequence dependent DNA models have to allow physical features that the wormlike chain, for instance, cannot describe, such as anisotropy of the ground state.

Table 7.1: Limiting values of the Flory Persistence vector

	λ	CF	T_L
$\langle \boldsymbol{r}_{1,500} \rangle$ [Å]	-55.71	116.25	168.69
$\langle \boldsymbol{r}_{2,500} \rangle$ [Å]	5.72	98.03	-28.97
$\langle \boldsymbol{r}_{3,500} \rangle$ [Å]	403.05	338.62	362.68

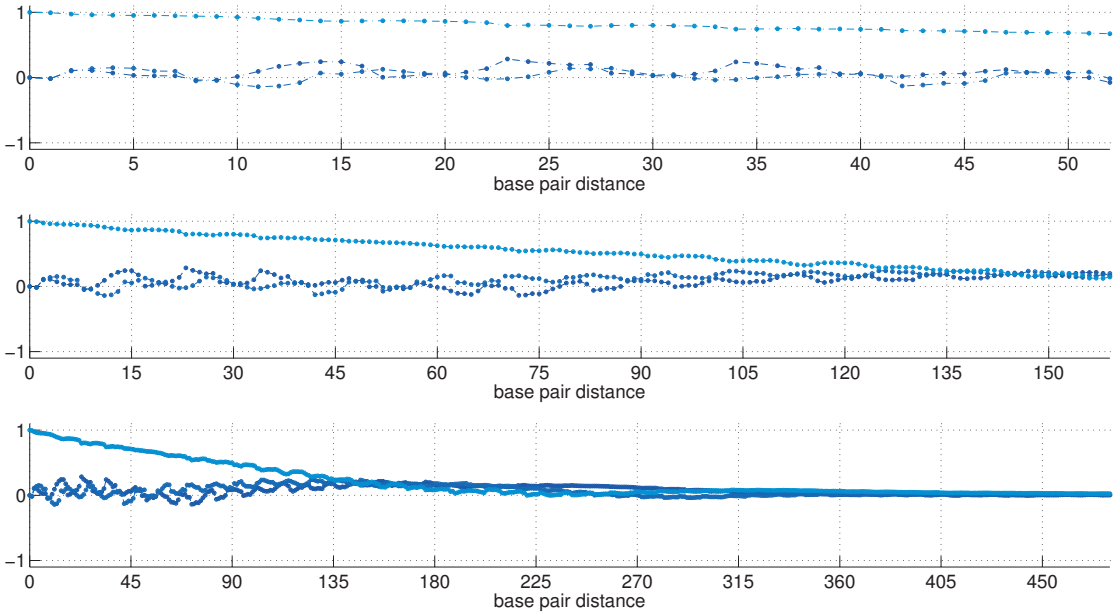


(a) *CF*

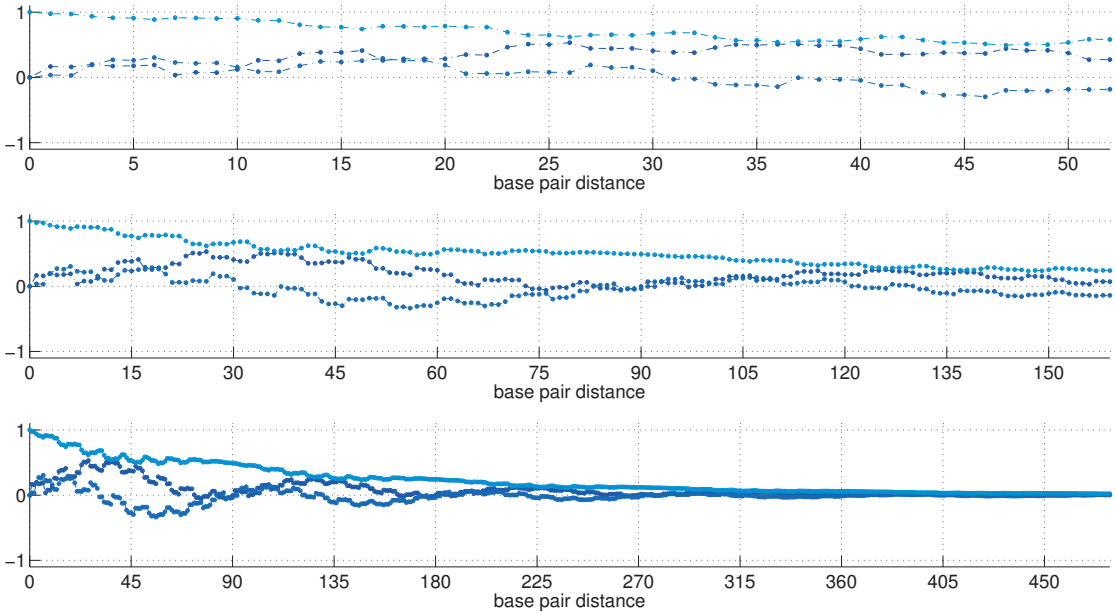


(b) T_L

Figure 7.7: Sketch of three dimensional configuration of 250 Monte Carlo samples (in blue) and of the expected chain configuration (in red) for the sequences *CF* (a) and T_L (b) (first 220 base pairs)

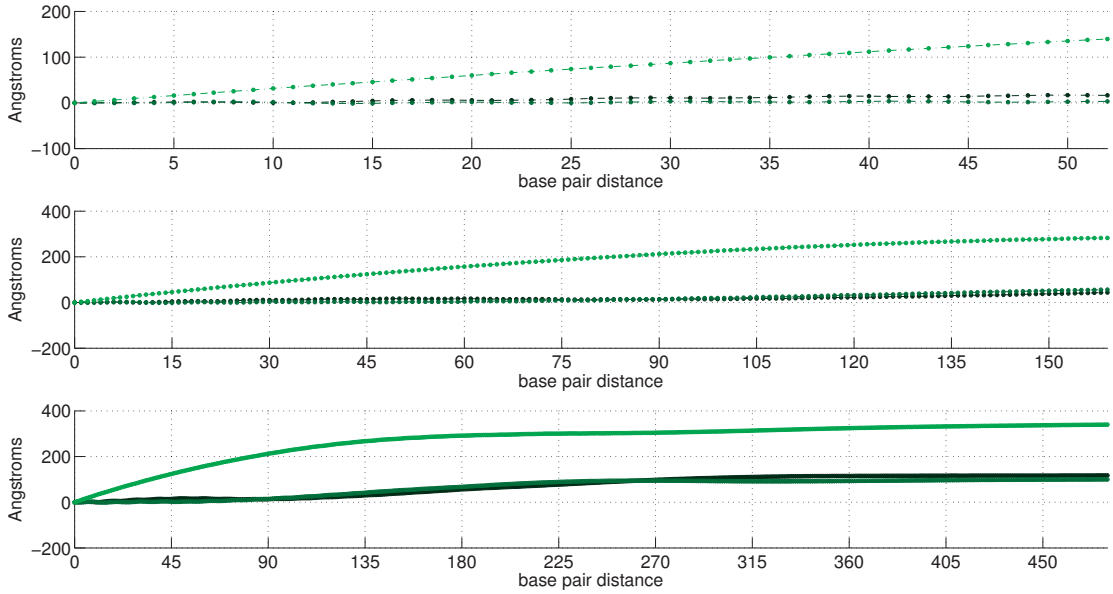


(a) *CF*

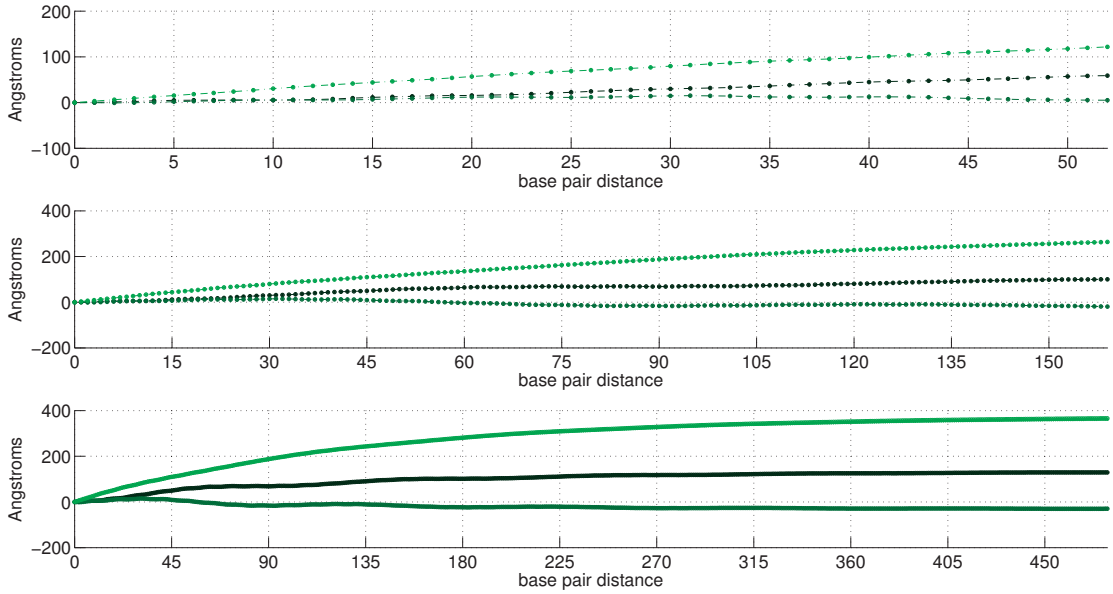


(b) T_L

Figure 7.8: Third column of the frame correlation matrix $\langle \boldsymbol{R}_{1,n} \rangle$ for the sequences *CF* (a) and T_L (b) and for three different molecule lengths. Monte Carlo simulations ($\circ$) versus the analytical predictions ($*$). More detail is given in the text.



(a) *CF*



(b) T_L

Figure 7.9: Components of the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$ for the sequences *CF* (a) and T_L (b) and for three different molecule lengths. Monte Carlo simulations ($\circ$) versus the analytical predictions ($*$). More detail is given in the text.

7.2.2 Factorization into the persistence matrix and the expected chain configuration

As presented in section 2.4.4, the factorization of the configurational first moment $\langle \boldsymbol{g}_{1,n} \rangle = (\langle \boldsymbol{R}_{1,n} \rangle, \langle \boldsymbol{r}_{1,n} \rangle)$ into the expected chain configuration $\bar{\boldsymbol{g}}_{1,n} = (\bar{\boldsymbol{R}}_{1,n}, \bar{\boldsymbol{r}}_{1,n})$ and the persistence matrix $\langle \boldsymbol{\mathcal{D}}_{1,n} \rangle = (\langle \boldsymbol{\Delta}_{1,n} \rangle, \langle \boldsymbol{\delta}_{1,n} \rangle)$ allows a decomposition of the frame correlation and of the Flory persistence vector in an interesting way. Precisely, for chains in a semi-flexible regime, we have shown that they can be written as

$$\langle \boldsymbol{R}_{1,n} \rangle = \langle \boldsymbol{\Delta}_{1,n} \rangle \, \bar{\boldsymbol{R}}_{1,n} \tag{7.2.5a}$$

$$\langle \boldsymbol{r}_{1,n} \rangle = \langle \boldsymbol{\delta}_{1,n} \rangle + \langle \boldsymbol{\Delta}_{1,n} \rangle \, \bar{\boldsymbol{r}}_{1,n} \tag{7.2.5b}$$

with the property that the entries in the matrix $\langle \boldsymbol{\Delta}_{1,n} \rangle$ are slowly decaying as a function of n and the components of $\langle \boldsymbol{\delta}_{1,n} \rangle$ are converging to the same limiting values as the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$.

The factorization of the matrix $\langle \boldsymbol{R}_{1,n} \rangle$ in (7.2.5a) is illustrated in figure 7.10 for the sequence λ . The three plots in the top line show respectively the three components of the three columns of the matrix $\langle \boldsymbol{R}_{1,n} \rangle$ a as function of the base pair index n, from left to right. This first line is essentially a summary of plots 7.2, 7.3 and 7.4. The second and the bottom lines present the components of the three columns of the rotation $\bar{\boldsymbol{R}}_{1,n}$ and of the matrix $\langle \boldsymbol{\Delta}_{1,n} \rangle$ respectively, following the same ordering as for $\langle \boldsymbol{R}_{1,n} \rangle$. Remarkably, it shows that all the fast variations in the frame correlation $\langle \boldsymbol{R}_{1,n} \rangle$ are encoded in the expected orientation configuration $\bar{\boldsymbol{R}}_{1,n}$, drawn in red in figure 7.10, whereas the slow decay is entirely due to the entries of the matrix $\langle \boldsymbol{\Delta}_{1,n} \rangle$. Moreover, this matrix is almost diagonal since the diagonal entries exhibit a kind of polynomial decay and the others stay close to zero as a function of n.

The expression for the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$ in (7.2.5b) suggests, as discussed in section 2.4.4, that it can be seen as being made up of the damped part $\langle \boldsymbol{\Delta}_{1,n} \rangle \, \bar{\boldsymbol{r}}_{1,n}$ and the part $\langle \boldsymbol{\delta}_{1,n} \rangle$. Figure 7.11 illustrates this decomposition for the sequence λ . The three first plots show the three components of $\langle \boldsymbol{r}_{1,n} \rangle$ in green. The components of the triples $\langle \boldsymbol{\Delta}_{1,n} \rangle \, \bar{\boldsymbol{r}}_{1,n}$ and $\langle \boldsymbol{\delta}_{1,n} \rangle$ are drawn in orange and red respectively. The bottom line represents the relative contribution in euclidean norm of $\langle \boldsymbol{\Delta}_{1,n} \rangle \, \bar{\boldsymbol{r}}_{1,n}$, drawn in orange, and $\langle \boldsymbol{\delta}_{1,n} \rangle$, drawn in red, of the triple $\langle \boldsymbol{r}_{1,n} \rangle$. We observe that for a length shorter than roughly 45 base pairs the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$ is strongly dominated by the expected chain position $\bar{\boldsymbol{r}}_{1,n}$, whereas the triple $\langle \boldsymbol{\delta}_{1,n} \rangle$ is the only remaining contribution for lengths significantly larger than roughly 400 base pairs.

The analogous factorization of the configurational first moment into the expected chain configuration and the persistence matrix of the form $\langle \boldsymbol{g}_{1,n} \rangle = \langle \bar{\boldsymbol{g}}_{1,n} \rangle \, \langle \boldsymbol{\mathcal{D}}_{1,n} \rangle$, or equivalently in (7.2.5), is presented in figure 7.12a and 7.12b and for the sequences CF and T_L respectively.

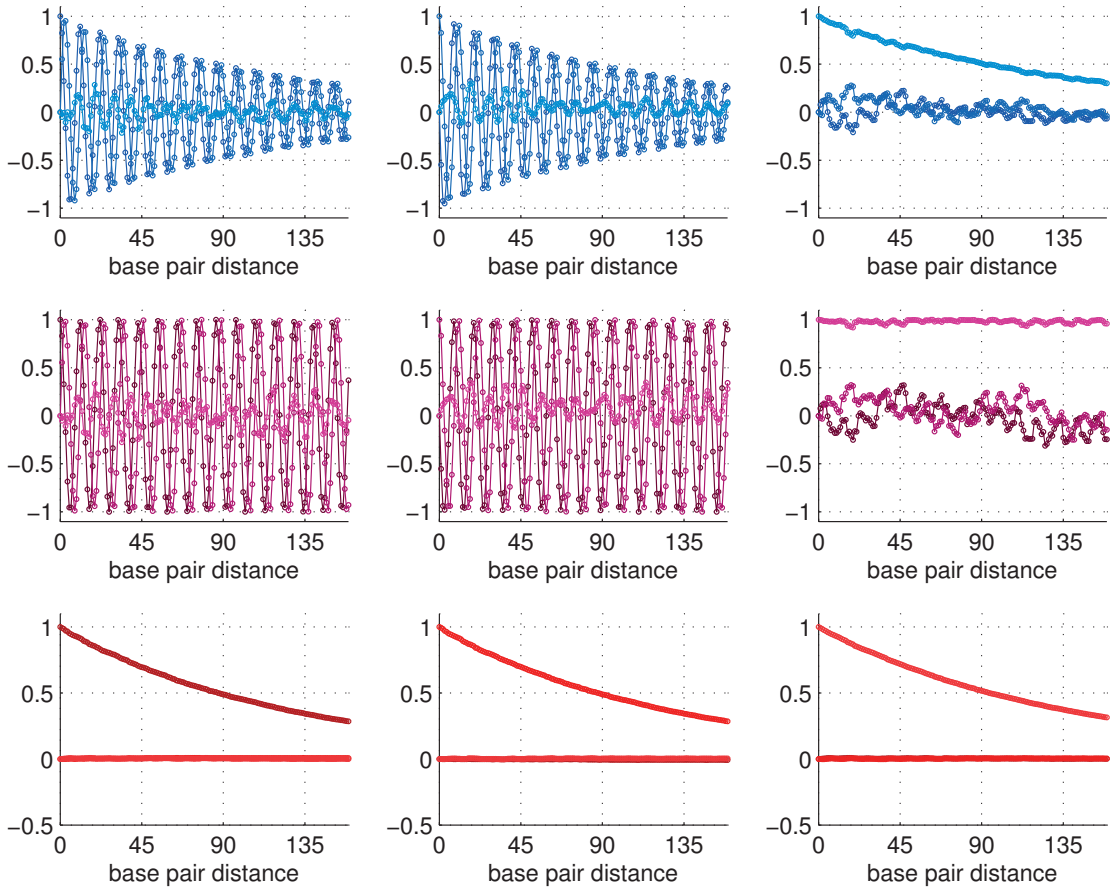


Figure 7.10: The three plots in the top line show respectively the components of the three columns of the matrix $\langle \boldsymbol{R}_{1,n} \rangle$ (in blue) for the sequence λ . The second and bottom lines present the matrix factors $\bar{\boldsymbol{R}}_{1,n}$ (in purple) and $\langle \boldsymbol{\Delta}_{1,n} \rangle$ (in red) respectively using the same ordering.

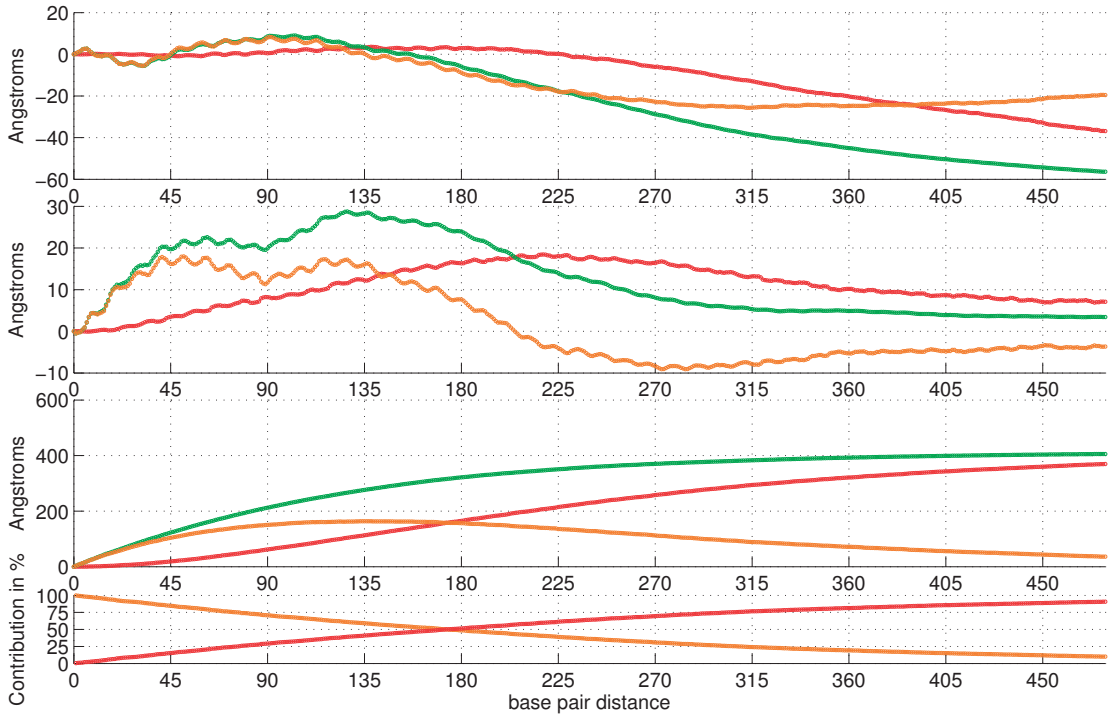
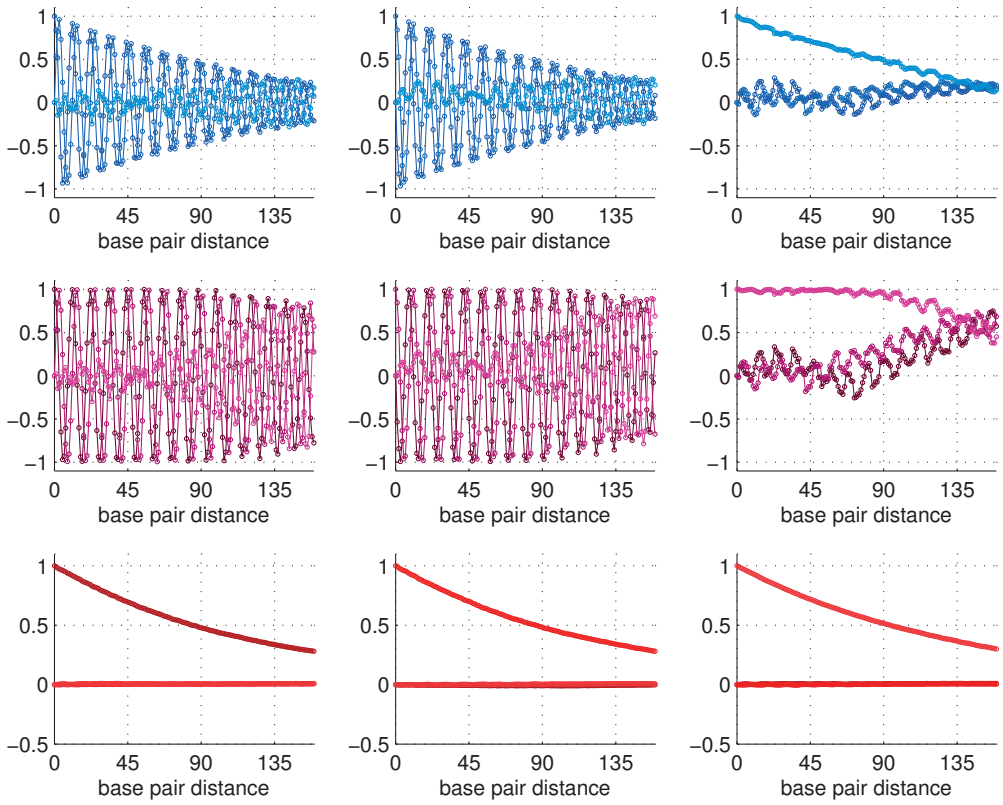
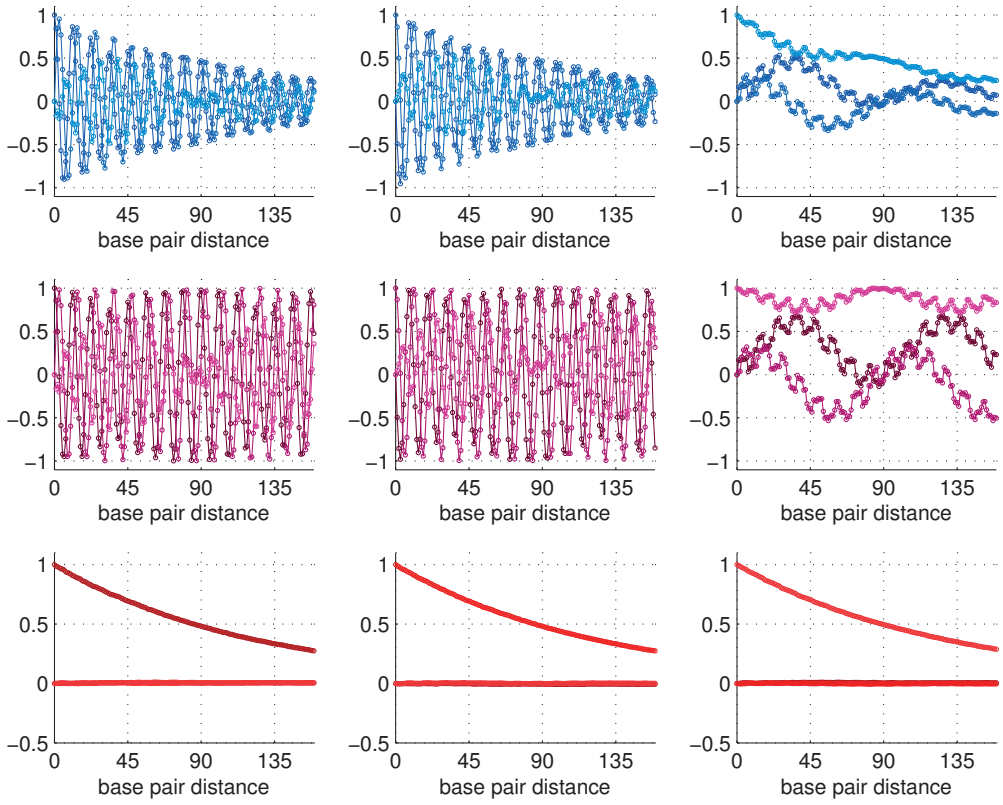


Figure 7.11: The three first plots show the three components of the triples $\langle \boldsymbol{r}_{1,n} \rangle$ (in green), $\langle \boldsymbol{\Delta}_{1,n} \rangle \bar{\boldsymbol{r}}_{1,n}$ (in orange) and $\langle \boldsymbol{\delta}_{1,n} \rangle$ (in red) for the sequence λ . The bottom panel represents the relative contribution in euclidean norm of $\langle \boldsymbol{\Delta}_{1,n} \rangle \bar{\boldsymbol{r}}_{1,n}$ (in orange) and $\langle \boldsymbol{\delta}_{1,n} \rangle$ (in red) of the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$.

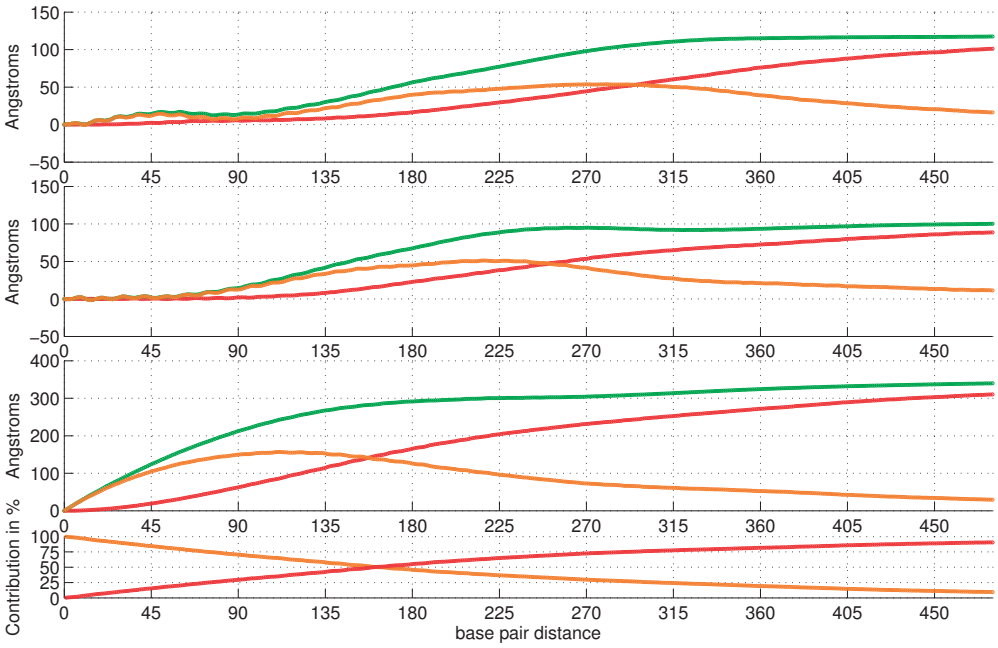


(a) *CF*

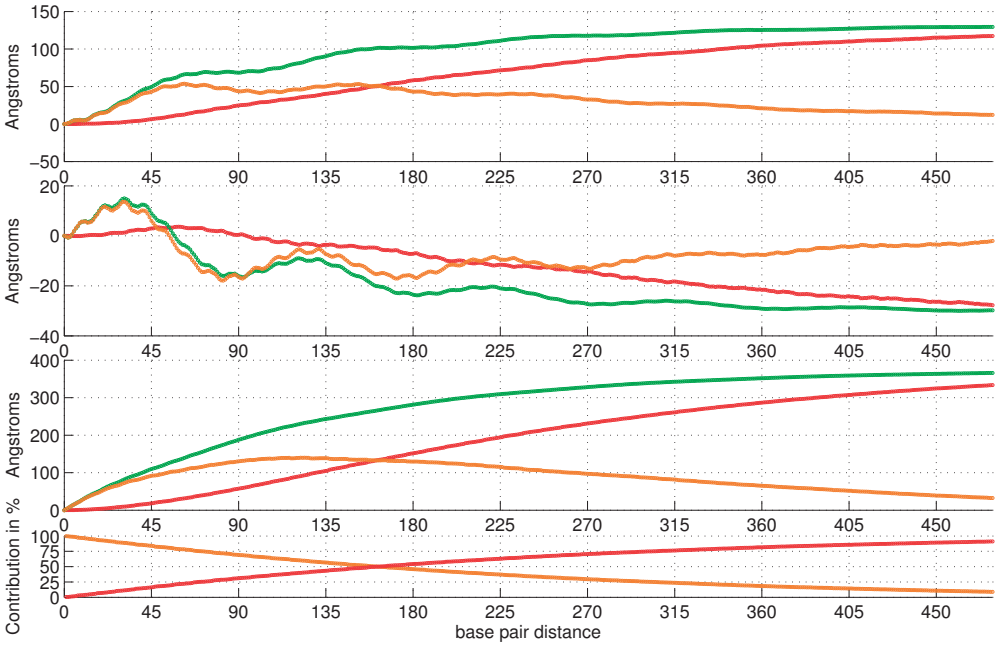


(b) T_L

Figure 7.12: In each subfigure, the three plots in the top line show respectively the components of the three columns of the matrix $\langle \boldsymbol{R}_{1,n} \rangle$ (in blue) for the sequences *CF* (a) and T_L (b). The second and bottom lines represent the matrix $\bar{\boldsymbol{R}}_{1,n}$ (in purple) and $\langle \boldsymbol{\Delta}_{1,n} \rangle$ (in red) respectively using the same ordering.



(a) CF



(b) T_L

Figure 7.13: In each subfigure, the three first plots show the three components of the triples $\langle \boldsymbol{r}_{1,n} \rangle$ (in green), $\langle \boldsymbol{\Delta}_{1,n} \rangle \bar{\boldsymbol{r}}_{1,n}$ (in orange) and $\langle \boldsymbol{\delta}_{1,n} \rangle$ (in red) for the sequences CF (a) and T_L (b) respectively. The bottom panels represent the relative contribution in euclidean norm of $\langle \boldsymbol{\Delta}_{1,n} \rangle \bar{\boldsymbol{r}}_{1,n}$ (in orange) and $\langle \boldsymbol{\delta}_{1,n} \rangle$ (in red) in the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$.

The proposed factorization of the configurational first moment $\langle \boldsymbol{g}_{1,n} \rangle$ has interesting properties when applied to the sequence dependent rigid base pair model described in section 7.2.1. We observe that most of the characteristic and fast variations appearing in component of the first moment $\langle \boldsymbol{R}_{1,n} \rangle$ and $\langle \boldsymbol{r}_{1,n} \rangle$ can be mainly attributed to the analogous variations in the expected chain configuration $\bar{\boldsymbol{g}}_{1,n} = (\bar{\boldsymbol{R}}_{1,n}, \bar{\boldsymbol{r}}_{1,n})$, such as the B-DNA intrinsic twist or the super-helical structure in the case of the sequence T_L illustrated in figures 7.12b and 7.13b.

Table 7.2: Characteristic inverse decays (in base pairs) of the matrix $\langle \boldsymbol{\Delta}_{1,n} \rangle$

	λ	CF	T_L
$\vartheta_1^{(bp)}$	126.93	129.69	124.19
$\vartheta_2^{(bp)}$	124.38	126.84	122.10
$\vartheta_3^{(bp)}$	131.05	130.61	128.39

The behaviour of the entries in the block $\langle \boldsymbol{\Delta}_{1,n} \rangle$ of the persistence matrix $\langle \boldsymbol{\mathcal{D}}_{1,n} \rangle$ looks similar for the three sequences λ , CF and T_L presented in figures 7.10, 7.12a and 7.12b. In order to quantify how close these decays are, we simply propose to fit these matrices by matrices of the form

$$\exp\left\{-n \ \mathrm{diag}\left(\frac{1}{\vartheta_1^{(bp)}}, \frac{1}{\vartheta_2^{(bp)}}, \frac{1}{\vartheta_3^{(bp)}}\right)\right\}. \tag{7.2.6}$$

The values of the constants $\vartheta_1^{(bp)}$,$\vartheta_2^{(bp)}$ and $\vartheta_3^{(bp)}$ are thought to be characteristic inverse decays since they roughly quantify the memory lost of the expected chain configuration $\bar{\boldsymbol{g}}_{1,n}$ in the configurational first moment $\langle \boldsymbol{g}_{1,n} \rangle$ according to the factorization (7.2.5). These values (measured in base pairs) are reported in table 7.2.

We conclude this section with a short digression. We recall that, as discussed briefly in section 2.4.4, the factorization of the configurational first moment into the persistence matrix $\langle \boldsymbol{\mathcal{D}}_{1,n} \rangle$ and the expected chain configuration $\bar{\boldsymbol{g}}_{1,n}$ can be seen as a generalisation of the expressions proposed in [Trifonov et al., 1988, Schellman and Harvey, 1995]. In these studies, based on (very) simple polymer models they discuss how the presence of localised static curvature decreases the apparent inverse decay rate of the tangent-tangent correlation. Moreover, in the DNA literature the value of 150 base pairs is the accepted characteristic value for the decay of the entries $[\langle \boldsymbol{R}_{1,n} \rangle]_{33}$. Therefore, even if the rigid base pair model proposed in 7.2.1 is itself an approximation of the more accurate rigid base model in [Gonzalez et al., 2013, Petkeviciute et al., 2014], which is itself parametrized from molecular dynamic, we are satisfied to observe that the proposed analytical theory delivers results, which are consistent with the DNA literature, but considerably more detailed.

7.2.3 On the short length approximation

In section 2.4.5 we have shown that very explicit approximation formulas can be stated for sufficiently short semi-flexible rigid body chains. In particular we have deduced the following asymptotic behaviour

$$\langle \boldsymbol{R}_{1,n} \rangle = \exp\{-\boldsymbol{\Omega}_{1,n}\}\, \bar{\boldsymbol{R}}_{1,n} \tag{7.2.7a}$$

$$\langle \boldsymbol{r}_{1,n} \rangle = [\mathrm{Id}_3 - \exp\{-\boldsymbol{\Omega}_{1,n}\}]\, \boldsymbol{\omega}_{1,n} + \exp\{-\boldsymbol{\Omega}_{1,n}\}\, \bar{\boldsymbol{r}}_{1,n} \tag{7.2.7b}$$

for a molecular length n small compared to $\frac{1}{\epsilon}$ and with the definitions

$$\boldsymbol{\Omega}_{1,n} = \frac{1}{2} \sum_{k=1}^{n-1} \bar{\boldsymbol{R}}_{1,k} \Lambda_k^{\boldsymbol{RR}} \bar{\boldsymbol{R}}_{1,k}^T \tag{7.2.8a}$$

$$\boldsymbol{\omega}_{1,n} = \left(\sum_{k=1}^{n-1} \bar{\boldsymbol{R}}_{1,k} \Lambda_k^{\boldsymbol{RR}} \bar{\boldsymbol{R}}_{1,k}^T \right)^{-1} \left[\sum_{k=1}^{n-1} \bar{\boldsymbol{R}}_{1,k} \Lambda_k^{\boldsymbol{RR}} \bar{\boldsymbol{R}}_{1,k}^T \bar{\boldsymbol{r}}_{1,k} - \bar{\boldsymbol{R}}_{1,k} \Lambda_k^{\boldsymbol{Rr}} \right] \tag{7.2.8b}$$

and where we recall that by convention

$$\Lambda_k^{\boldsymbol{RR}} = \mathrm{Tr}\left(\mathcal{C}_k^{\boldsymbol{RR}}\right) \mathrm{Id}_3 - \mathcal{C}_k^{\boldsymbol{RR}} \text{ and } \Lambda_k^{\boldsymbol{Rr}} = \mathrm{Vect}\left(\mathcal{C}_k^{\boldsymbol{Rr}}\right) \tag{7.2.9}$$

with $\mathcal{C}_k^{\boldsymbol{RR}}$ and $\mathcal{C}_k^{\boldsymbol{Rr}}$ denoting blocks of the local fluctuation matrix $\mathcal{C}_k$, defined in (2.4.16), and that $\epsilon \sim 0.01$ for the rigid base pair model parametrised as proposed in section 7.2.1.

The main interest of the factorization (7.2.7) with respect to the more general one proposed in (7.2.5), is that it is now in closed form, even if that is only the leading order terms in an expansion in length. In other words the expression (7.2.7) literally allows to read off the values of the entries of the configurational first moment $\langle \boldsymbol{g}_{1,n} \rangle = (\langle \boldsymbol{R}_{1,n} \rangle, \langle \boldsymbol{r}_{1,n} \rangle)$ as a function of the set of fluctuation matrices $\mathcal{C}_n$ and the expected chain rigid body configurations $\bar{\boldsymbol{g}}_{1,n} = (\bar{\boldsymbol{R}}_{1,n}, \bar{\boldsymbol{r}}_{1,n})$. The accuracy of the expansion for the frame correlation matrix in (7.2.7a) can be appreciated from figure 7.14. The components of the three columns of the matrix $\langle \boldsymbol{R}_{1,n} \rangle$ computed using a Monte Carlo simulation are drawn in blue in the three first plots. The analogous results obtained using the expansion are shown in red. Similarly, figure 7.15 illustrates the short length expansion for the Flory persistence vector in (7.2.7b). The three plots show in green the three components of $\langle \boldsymbol{r}_{1,n} \rangle$ computed using Monte Carlo sampling. The results of the expansion are drawn in red. The grey horizontal lines illustrate the limiting values of the triple $\langle \boldsymbol{r}_{1,n} \rangle$ appearing in table 7.1. It is interesting to observe that the short length expansion gives good results for the frame correlation matrix for the entire molecular length, whereas it seems to be acceptable until only roughly $n = 135$ for the Flory persistence vector. The relative errors between the results obtained using the expansion and the Monte Carlo simulation are reported in figure 7.16 as a function of the base pair length. The analogous short length expansions for the sequences CF and T_L are presented in figures 7.17 and 7.18.

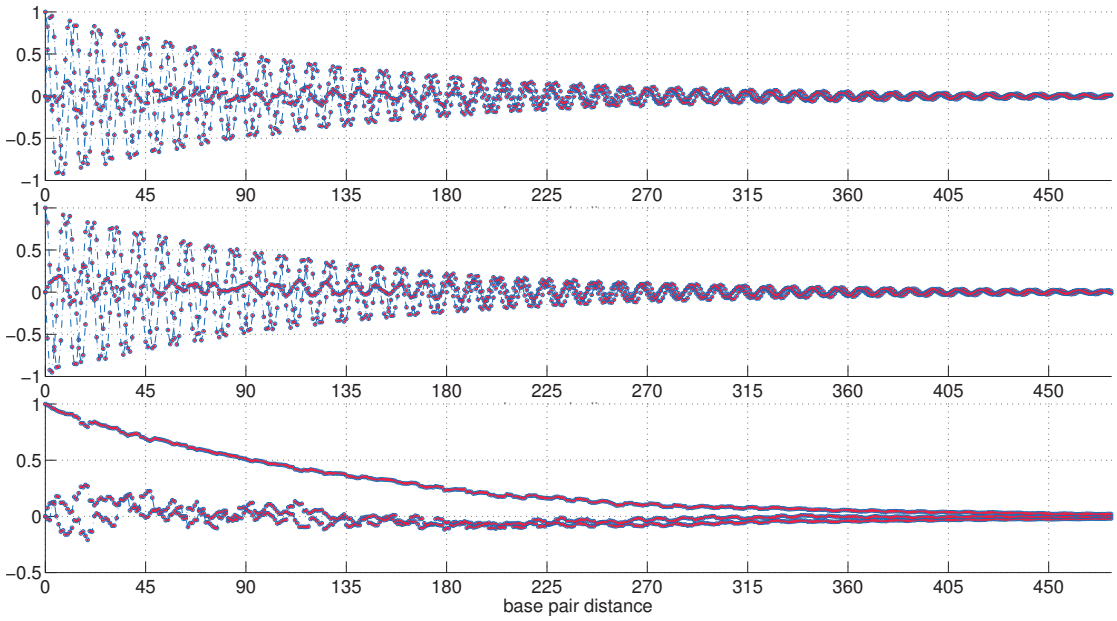


Figure 7.14: The three plots show respectively the components of the three columns of the frame correlation matrix $\langle \boldsymbol{R}_{1,n} \rangle$ computed using a Monte Carlo simulation (in blue) and the short length expansion (in red) for the sequence λ .

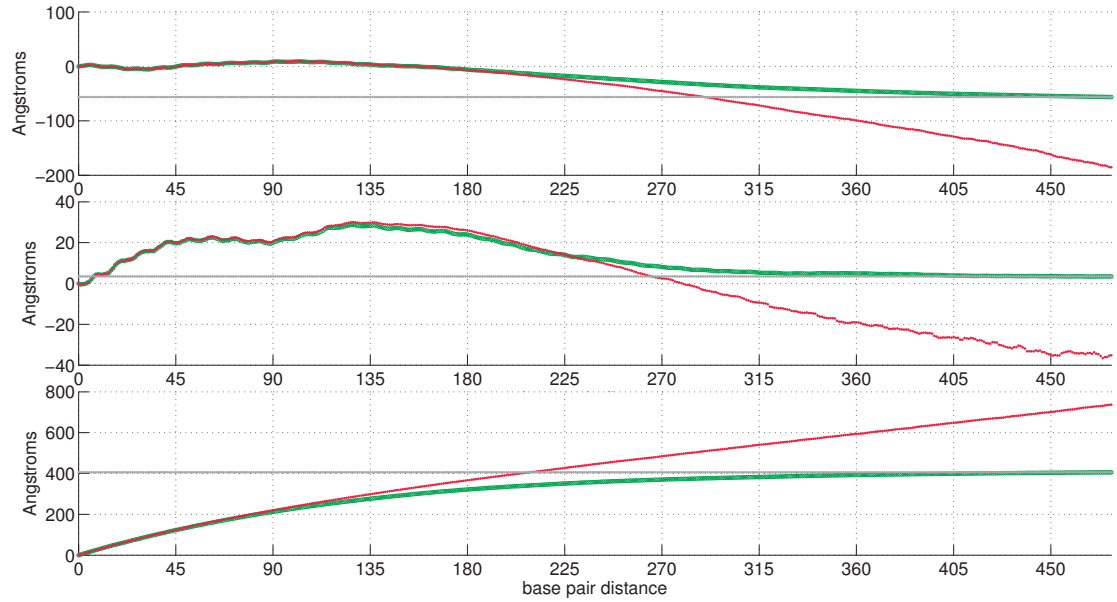


Figure 7.15: The three plots show respectively the components of the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$ computed using a Monte Carlo simulation (in green) and the short length expansion (in red) for the sequence λ . The grey line representing the limiting values of the components of $\langle \boldsymbol{r}_{1,n} \rangle$.

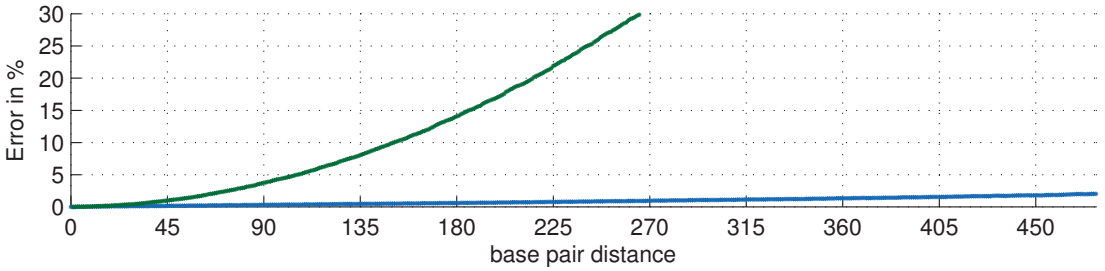


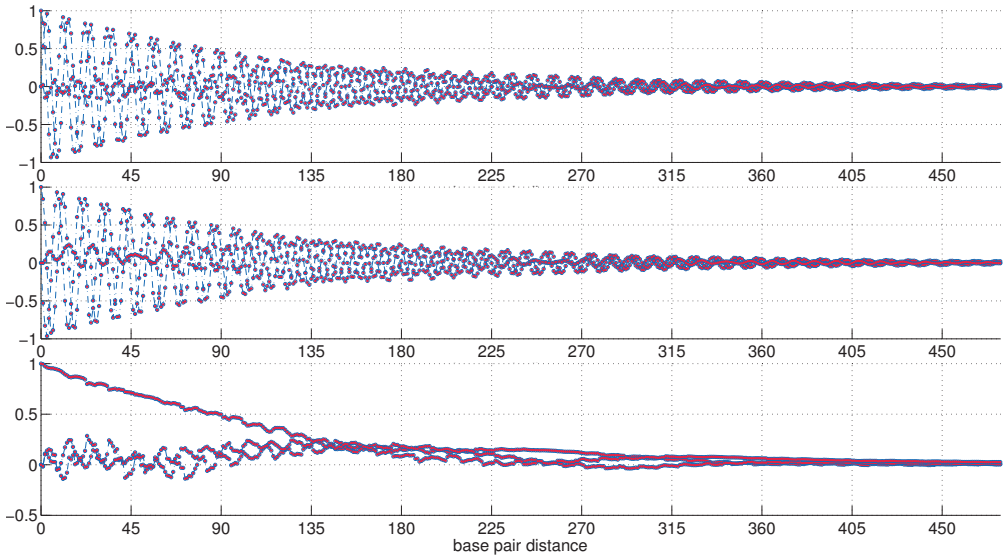
Figure 7.16: Relative errors between the short length expansion and a Monte Carlo simulation for the frame correlation matrix (in blue) and the Flory persistence vector (in green) for the sequence λ .

The quality of the short length expansion formulas in (7.2.7) is observed to be similar for the three sequences λ , CF and T_L . The frame correlation matrix $\langle \boldsymbol{R}_{1,n} \rangle$ can reasonably be approximated irrespective of the base pair distance, whereas the length of roughly 150 base pairs seems to be the threshold of applicability of the short length approximation for the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$.

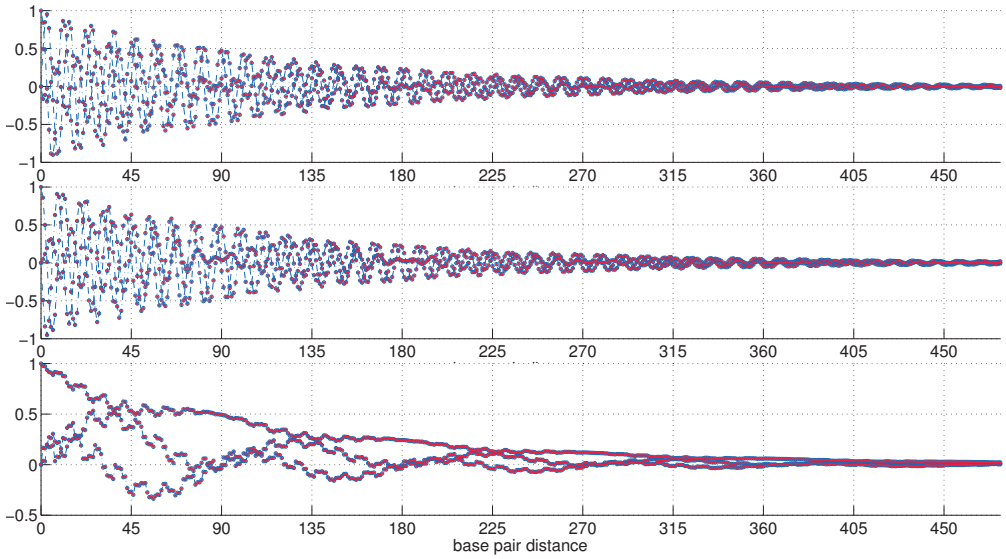
Table 7.3: Characteristic decays (in base pairs) of the matrix $\exp\{-\boldsymbol{\Omega}_{1,n}\}$

	λ	CF	T_L
$\vartheta_1^{(bp)}$	127.62	130.38	124.82
$\vartheta_2^{(bp)}$	124.90	127.37	122.68
$\vartheta_3^{(bp)}$	131.56	131.09	128.95

We conclude this section by noting that the entries of the matrix $\exp\{-\boldsymbol{\Omega}_{1,n}\}$ in (7.2.7a) are very close to the ones in the matrix $\langle \boldsymbol{\Delta}_{1,n} \rangle$ in (7.2.7a), as one can deduce by comparing figures 7.10 and 7.14 for instance. To make the comparison more quantitative we report in table 7.3 the characteristic inverse decay values $\vartheta_1^{(bp)}$, $\vartheta_2^{(bp)}$ and $\vartheta_3^{(bp)}$ (measured in base pairs) obtained by a linear fit to the matrix $\boldsymbol{\Omega}_{1,n}$. These values are observed to be close but slightly above the analogous ones in table 7.2.

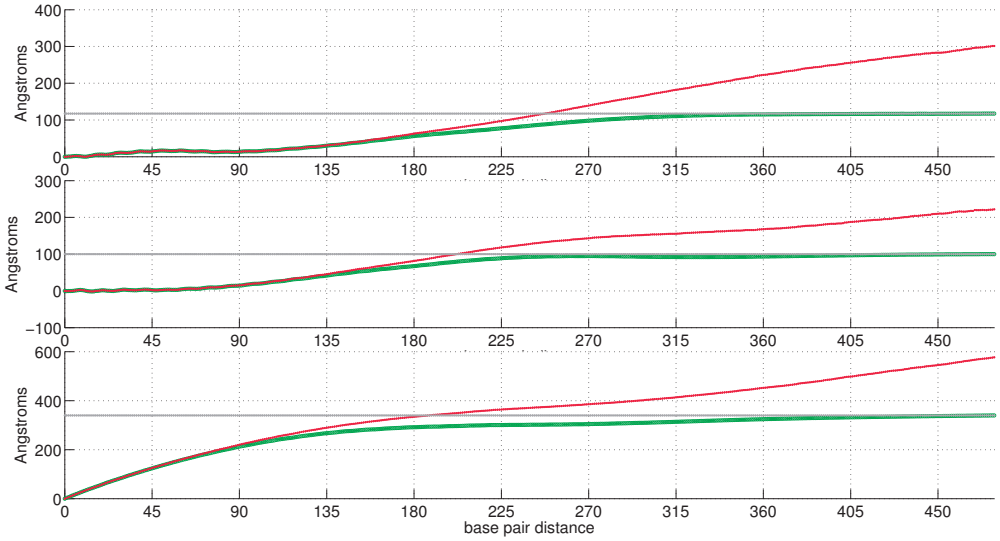


(a) *CF*

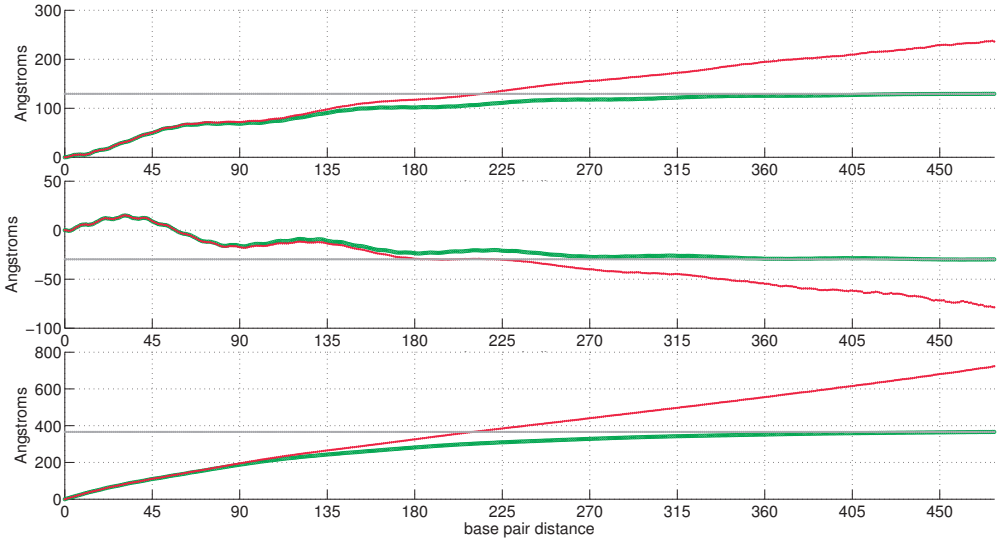


(b) T_L

Figure 7.17: In each subfigure, the three plots show respectively the components of the three columns of the frame correlation matrix $\langle \boldsymbol{R}_{1,n} \rangle$ computed using a Monte Carlo simulation (in blue) and the short length expansion (in red) for the sequence *CF* and T_L respectively.



(a) CF



(b) T_L

Figure 7.18: In each subfigure, the three plots show respectively the components of the Flory persistence vector $\langle \boldsymbol{r}_{1,n} \rangle$ computed using the Monte Carlo simulation (in green) and using the short length expansion (in red) for the sequence CF and T_L respectively. The grey line representing the limiting values of the components of $\langle \boldsymbol{r}_{1,n} \rangle$.

7.3 A Rod Formulation

As discussed in section 5.3.1, a continuum description of the rigid base pair model in (7.2.1) is given by a rod stationary configurational distribution of the form

$$d\rho\left[\boldsymbol{g}[\xi]\right] \sim \exp\left\{-\int_0^L \frac{1}{2}\left(\xi(s)-\bar{\xi}(s)\right)\cdot\bar{\mathcal{K}}(s)\left(\xi(s)-\bar{\xi}(s)\right)ds\right\}D[\xi] \tag{7.3.1}$$

where $\xi(s)\in\mathbb{R}^6$ is the generator vector associated with the continuum rod configuration $\boldsymbol{g}(s)\in \mathrm{SE}(3)$.

The explicit approximation for the configurational first moment $\langle \boldsymbol{g}(s',s)\rangle = \langle \boldsymbol{g}(s')^{-1}\boldsymbol{g}(s)\rangle$ proposed in section 3.4.2 then reads

$$\partial_s\left\langle \boldsymbol{g}(s',s)\right\rangle = \left\langle \boldsymbol{g}(s',s)\right\rangle\left(\mathcal{T}\bar{\xi}(s)-\frac{1}{2}\mathcal{L}(s)\right)+\mathrm{o}\left(\epsilon\right) \text{ with } \left\langle \boldsymbol{g}(s',s')\right\rangle = \mathrm{Id}_4. \tag{7.3.2a}$$

where

$$\mathcal{L}(s)=\begin{pmatrix}\mathrm{Tr}\left(\mathcal{C}^{\boldsymbol{RR}}(s)\right)\mathrm{Id}_3-\mathcal{C}^{\boldsymbol{RR}}(s) & \mathrm{Vect}\left(\mathcal{C}^{\boldsymbol{Rr}}(s)\right)\\ 0 & 0\end{pmatrix}. \tag{7.3.2b}$$

The vector $\bar{\xi}(s)\in\mathbb{R}^6$ denotes the generator vector of the expected rod configuration, as defined in (3.4.12), and appears explicitly in the distribution $d\rho^{(N)}\left[\boldsymbol{g}[\xi]\right]$ in (7.3.1). The matrices $\mathcal{C}^{\boldsymbol{RR}}(s)$ and $\mathcal{C}^{\boldsymbol{Rr}}(s)$ stand for blocks of the rod fluctuation matrix $\mathcal{C}(s)$, as defined in (3.4.15), and we have the relation $\mathcal{C}(s)=\bar{\mathcal{K}}^{-1}(s)$ according to (4.3.3b).

In order to define a rod configurational distribution $d\rho\left[\boldsymbol{g}[\xi]\right]$ of the form in (7.3.1) consistent with the sequence dependent rigid base pair model discussed in (7.2.1) we use the stochastic Cauchy-Born rule for chains presented in section 4.3 to have

$$\bar{\xi}(s)=\bar{\xi}^{(N)}(s) \tag{7.3.3a}$$

$$\mathcal{C}(s)=\frac{1}{{h_n}^{(N)}}\left(\mathrm{Ad}^{-1}_{\bar{\boldsymbol{g}}_n^{-1}\bar{\boldsymbol{g}}^{(N)}(s)}\mathbb{L}^{(1)}_{\bar{x}_n}\right)K_n^{-1}\left(\mathrm{Ad}^{-1}_{\bar{\boldsymbol{g}}_n^{-1}\bar{\boldsymbol{g}}^{(N)}(s)}\mathbb{L}^{(1)}_{\bar{x}_n}\right)^T+\mathrm{o}\left(\epsilon\right) \tag{7.3.3b}$$

where $\bar{\boldsymbol{g}}_n$ and $\bar{x}_n$ stand for the rigid body chain expected configuration and its local internal coordinates, as discussed in (4.3.9), and where similarly $\bar{\boldsymbol{g}}^{(N)}(s)$ and $\bar{\xi}^{(N)}(s)$ denote its interpolated configuration and its related generator vector, as defined in 4.3.4. The entries of the matrix $\bar{\mathcal{K}}(s)$ are drawn for a few base pairs for the sequences λ , CF and T_L in figures 5.9, 5.10a and 5.10b respectively, and in figure 5.11 for a few hundreds of base pairs. We observe that if the original rigid body chain is in a semi-flexible regime, in the sense of (2.4.20), then the induced continuum rod is also in a semi-flexible regime in the sense of (3.4.23), according to the definition (7.3.3b).

7.3.1 Comparison between the analytical ODE and Monte Carlo simulations

Similarly to the numerical comparisons made in section 7.2.1, we shall discuss the accuracy of the sequence dependent ODE proposed in (7.3.2). The strategy is simply to quantify the difference between a numerical solution $\langle \boldsymbol{g}(0,s)\rangle = (\langle \boldsymbol{R}(0,s)\rangle, \langle \boldsymbol{r}(0,s)\rangle)$, obtained using a standard ODE solver in *matlab*, with the associated Monte Carlo simulations used in section 7.2.1.

Analogously to figure 7.2, figure 7.20 shows the three components of the first column of the continuum frame correlation matrix $\langle \boldsymbol{R}(0,s)\rangle$ for the sequence λ , using a continuum line, and the same components of the matrix $\langle \boldsymbol{R}_{1,n}\rangle$ obtained from Monte Carlo sampling for the same sequence, using the symbol $\circ$, at the scale of 50, 150 and 450 base pairs. Similarly figures 7.21 and 7.22 present the comparison for the second and third columns respectively.

The components of the Flory persistent vector $\langle \boldsymbol{r}(0,s)\rangle$ for the sequence λ , as predicted by the differential equation (7.3.2), together with the analogous ones $\langle \boldsymbol{r}_{1,n}\rangle$, computed using the Monte Carlo simulation, are drawn in figure 7.23.

The relative error between the values of the continuum frame correlation matrix $\langle \boldsymbol{R}(0,s)\rangle$ evaluated at base pair position, obtained by solving the ODE in (7.3.2), and its discrete version $\langle \boldsymbol{R}_{1,n}\rangle$, obtained from the Monte Carlo simulation, and similarly relative error between the continuum Flory persistence vector $\langle \boldsymbol{r}(0,s)\rangle$ and its discrete analog $\langle \boldsymbol{r}_{1,n}\rangle$, are presented in figure 7.19 for the sequence λ . These plots look very similar to the ones shown in figure 7.6 obtained by comparing the rigid base pair description analytic relation with the Monte Carlo simulation. This is because the continuum solution and the discrete one are very close one to the other. We conclude that the large relative errors observed in the entries of the frame correlation matrix are due to the relatively small size of their absolute values, and that the Flory persistence vector, which has relatively large components, is approximated to within 1%. Consequently, the solution $\langle \boldsymbol{g}(0,s)\rangle$ of the sequence dependent ODE (7.3.2) is thought to give an accurate description of the configurational first moment $\langle \boldsymbol{g}_{1,n}\rangle$. The exponential stochastic Cauchy-Born rule for chains, together with the proposed ODE, are then consequently considered to be very satisfactory.

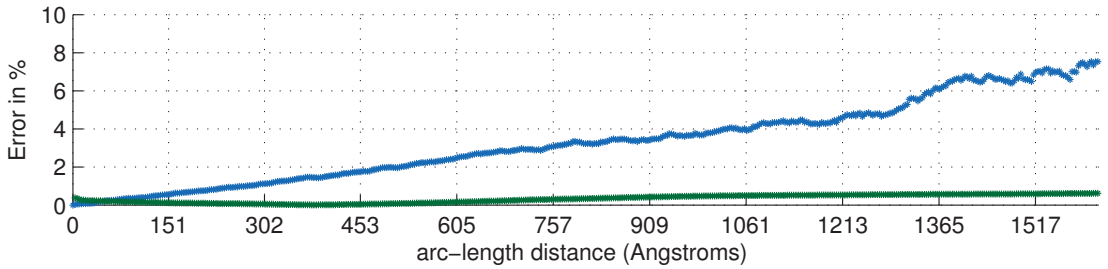


Figure 7.19: Relative errors between Monte Carlo simulations and analytic predictions for the frame correlation matrix $\langle \boldsymbol{R}(0,s)\rangle$ (blue) and for Flory persistence vector $\langle \boldsymbol{r}(0,s)\rangle$ (green) associated to the sequence λ as a function of the molecule length.

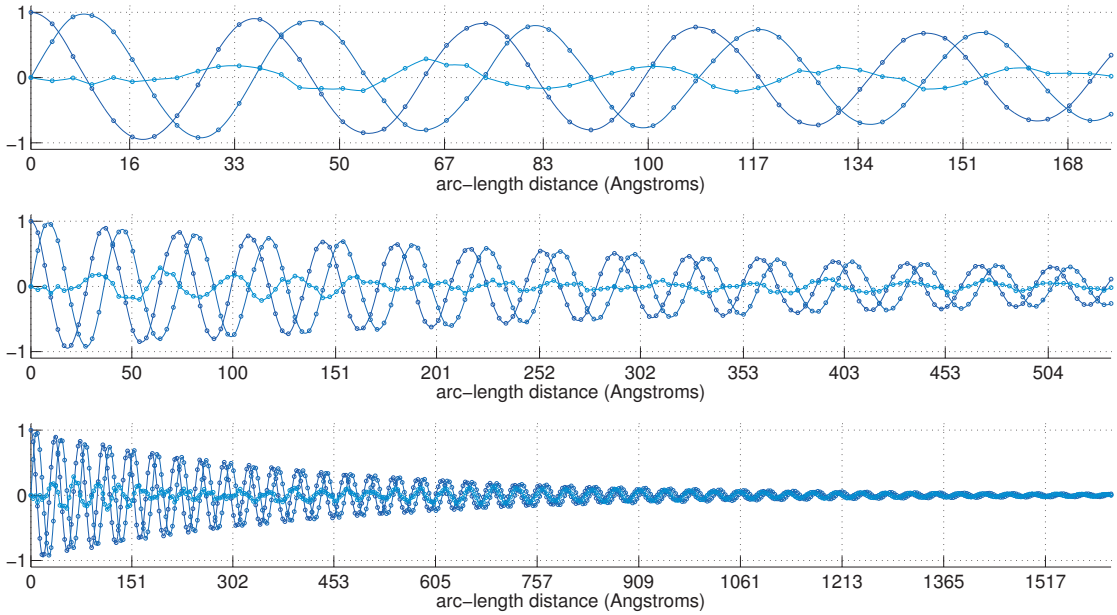


Figure 7.20: First column of the frame correlation matrix $\langle \boldsymbol{R}(0,s) \rangle$ for the sequence λ and for three different molecule lengths. Monte Carlo simulation ($\circ$) versus analytic predictions ($-$).

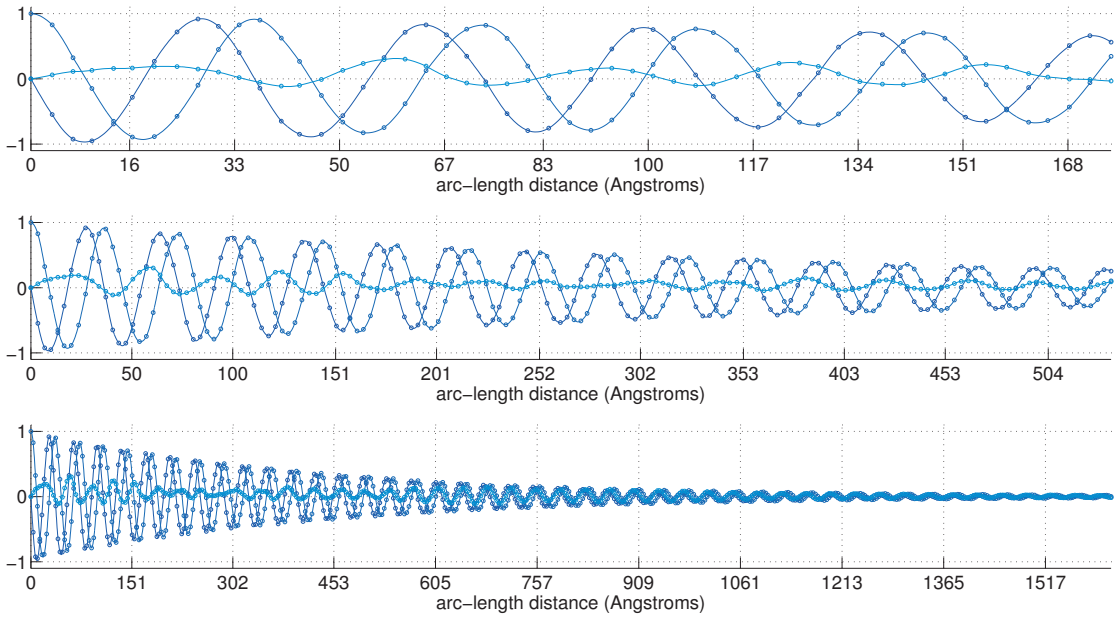


Figure 7.21: Second column of the frame correlation matrix $\langle \boldsymbol{R}(0,s) \rangle$ for the sequence λ and for three different molecule lengths. Monte Carlo simulation ($\circ$) versus analytic predictions ($-$).

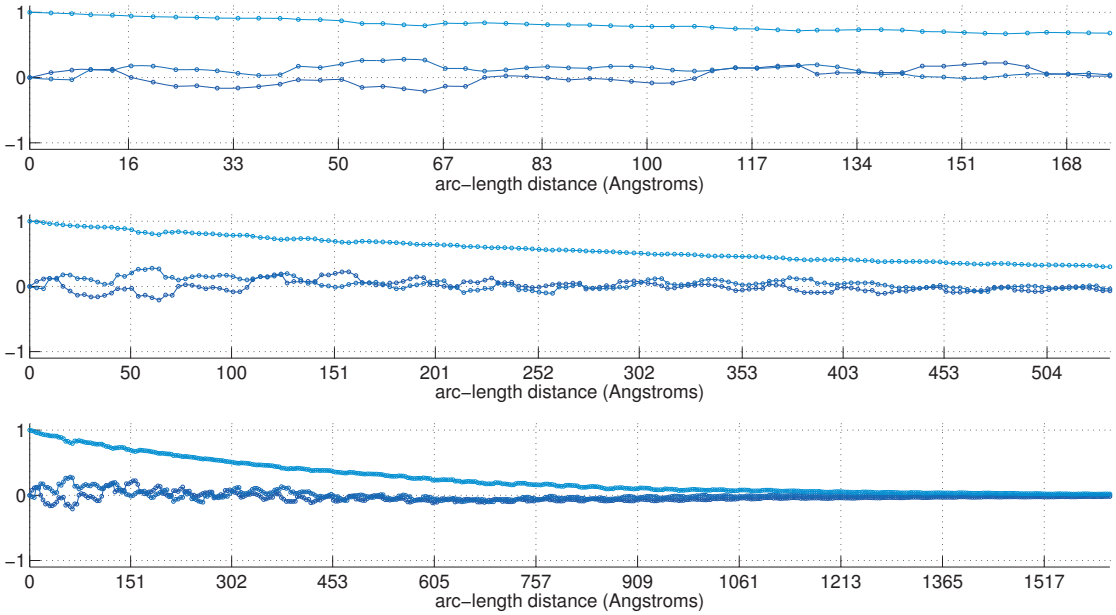


Figure 7.22: Third column of the frame correlation matrix $\langle \boldsymbol{R}(0,s) \rangle$ for the sequence λ and for three different molecule lengths. Monte Carlo simulation ($\circ$) versus analytic predictions ($-$).

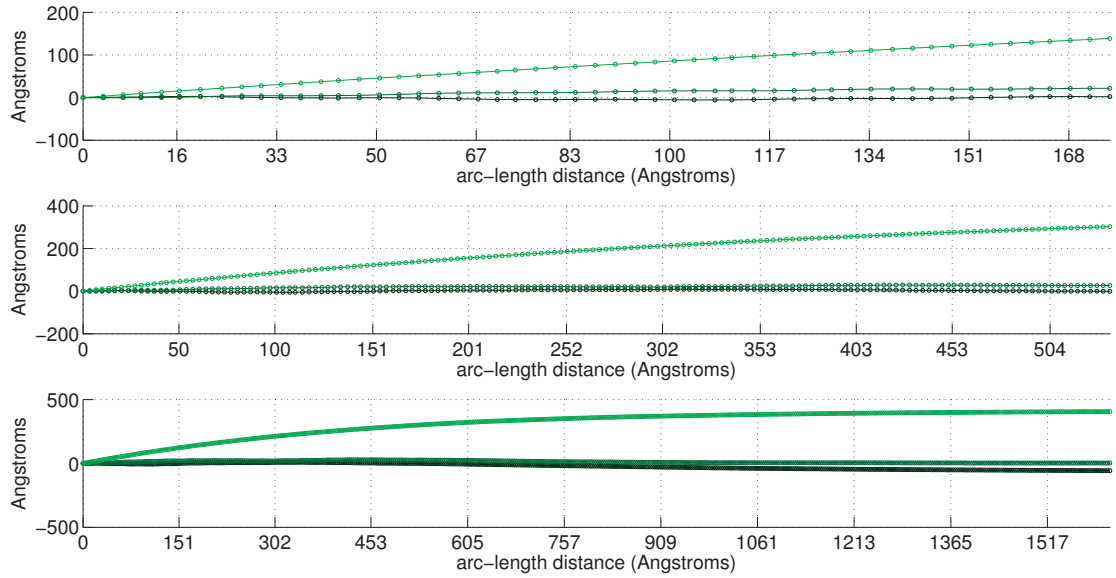


Figure 7.23: Component of the Flory persistence vector $\langle \boldsymbol{r}(0,s) \rangle$ for the sequence λ and for three different molecule lengths. Monte Carlo simulation ($\circ$) versus analytic predictions ($-$).

The other results presented in section 3.4, such as the factorization using the persistence matrix $\langle \boldsymbol{\mathcal{D}}(0,s)\rangle$ or the short length expansion, are not illustrated for reasons of compactness of the discussion. The reader should however be assured that the results are analogous to the ones shown in 7.10, 7.11, 7.14 and 7.15, since the configurational first moment $\langle \boldsymbol{g}(0,s)\rangle$ has been shown to be very similar to $\langle \boldsymbol{g}_{1,n}\rangle$ and, by definition, the expected rod configuration $\bar{\boldsymbol{g}}(0,s)$ is an interpolation of the expected chain configuration $\bar{\boldsymbol{g}}_{1,n}$. We do report in table 7.4 the continuum analog of the inverse characteristic decays $\vartheta_1^{(\text{Å})}$, $\vartheta_2^{(\text{Å})}$ and $\vartheta_3^{(\text{Å})}$ measured in Ångström for the matrix $\langle \boldsymbol{\Delta}(0,s)\rangle$ defined by the fit using the expression

$$\exp\left\{-s\ \text{diag}\left(\frac{1}{\vartheta_1^{(\text{Å})}},\frac{1}{\vartheta_2^{(\text{Å})}},\frac{1}{\vartheta_3^{(\text{Å})}}\right)\right\}. \tag{7.3.4}$$

These values are slightly lower than the accepted value of 450Å for the characteristic inverse decay usually found in the DNA literature. A discussion similar to the one made for the analogous table 7.2 can be made.

Table 7.4: Characteristic inverse decays (in Ångström) of the matrix $\langle \boldsymbol{\Delta}(0,s)\rangle$

	λ	CF	T_L
$\vartheta_1^{(\text{Å})}$	423.01	436.85	416.64
$\vartheta_2^{(\text{Å})}$	413.54	426.65	415.08
$\vartheta_3^{(\text{Å})}$	438.52	439.34	432.57

From the point of view of both analytical and numerical understanding of the behaviour of the frame correlation $\langle \boldsymbol{R}_{1,n}\rangle$ and of the Flory persistence vector $\langle \boldsymbol{r}_{1,n}\rangle$, it is not clear that the continuum description formulated in (7.3.2) is of interest with respect to its discrete analog in (7.2.3). The main reason is that the coefficients in the matrix $\mathcal{K}(s)$, or equivalently in the fluctuation matrix $\mathcal{C}(s)$, are discontinuous and highly varying, as drawn in figure 5.11 for instance, and consequently any accurate numerical implementation of such a continuum description has to be at least as intensive as the original discrete system. However, the continuum formulation in (7.3.2) allows the possibility of using homogenisation techniques coming from ODE analysis and, therefore, the chance of delivering large scale behaviour of the configurational first moment $\langle \boldsymbol{g}_{1,n}\rangle$ deduced from the large scale solution $\langle \boldsymbol{g}(0,s)\rangle$.

7.3.2 A large scale homogenization approach

As discussed in section 5.1, understanding the sequence dependent statistical physics properties of DNA molecules involves the analysis of several phenomena taking place at several length scales, ranging from the detailed local sequence of bases to the mechanical properties of sites made of hundreds of base pairs. A valuable theory would then be capable of describing and studying both short and long scales. However, short scale theories quickly become intensive and confused at larger scale and, in contrast, large scale theories only gives inaccurate and coarse results at shorter scales. Interestingly, the continuum description of the configurational first moment $\boldsymbol{g}_{1,n}$ through the sequence dependent ODE (7.3.2) allows us to make a first step in the construction of a multi-scale theory of DNA molecule statistical physics. More precisely, it allows to state a *local averaging principle* for the expressions governing the behaviour of the rod persistence matrix $\langle \boldsymbol{\mathcal{D}}(0,s) \rangle$, and consequently for the first moment $\langle \boldsymbol{g}(0,s) \rangle$ through an analytic transformation.

We first have to discuss the notion of *scale separation* for an ODE system. Let

$$\begin{aligned} \frac{d}{ds} x_\epsilon(s) &= \epsilon f(x_\epsilon(s), y_\epsilon(s)) \\ \frac{d}{ds} y_\epsilon(s) &= g(x_\epsilon(s), y_\epsilon(s)) \end{aligned} \tag{7.3.5}$$

for $s \in]0, L[$ and assume $f, g = \mathrm{O}(1)$. For $0 < \epsilon << 1$, the scale separation arises from the fact that the variable y_ϵ exhibits strong variations with respect to the variable x_ϵ at the length scale $s = \mathrm{O}(1)$. The variable x_ϵ is then called a *slow variable* while y_ϵ is called a *fast variable*, see for instance the discussion in [Hartmann, 2007, Freidlin and Wentzell, 2012, Abdulle et al., 2012]. As discussed in more detail in appendix A.7, the behaviour of the variable $x_\epsilon(s)$ can be approximated by the solution $x^{(H)}(s)$ of the following ODE

$$\frac{d}{ds} x^{(H)}(s) = f_{\Delta s}^{(H)} \left(x^{(H)}(s) \right) \tag{7.3.6a}$$

where by definition

$$f_{\Delta s}^{(H)} \left(x(s) \right) = \frac{1}{2\Delta s} \int_{s-\Delta s}^{s+\Delta s} f(x(s), y_0(s'; s)) ds' \tag{7.3.6b}$$

for a given value of Δs and for all function $x(s)$ and where the variable $y_0(s; s_0)$ is defined by

$$\frac{d}{ds} y_0(s; s_0) = g(x^{(H)}(s_0), y(s; s_0)) \text{ with } y_0(s_0; s_0) = y(s_0). \tag{7.3.6c}$$

The system (7.3.6a) constitutes an averaging principle, or *homogenised system*, for the ODE in (7.3.5) and has the property that its solutions $x^{(H)}(s)$ and $y_0(s; s_0)$ provide, in some sense, approximations of the original solution $x_\epsilon(s)$ and $y_\epsilon(s)$. The precise sense of the convergence, as well as necessary hypotheses on the function f and g, require a

discussion which goes beyond this thesis, see for instance [Freidlin and Wentzell, 2012, Abdulle et al., 2012]. Importantly, if Δs is chosen such that $\epsilon \Delta s = \mathrm{o}(1)$, we can expect that the homogenised solution $x^{(H)}(s)$ of the system (7.3.6a) provides an approximation of order $\mathrm{O}(\epsilon \Delta s)$ of the variable $x_\epsilon(s)$ not only at the scale of $s = \mathrm{O}(1)$ but also for $s = \mathrm{O}\left(\frac{1}{\epsilon}\right)$, and, moreover, the averaging procedure in (7.3.6b) delivers a function $f_{\Delta s}^{(H)}$ which is better behaved than f. We note that this approach is more direct when the function g does not depend explicitly on the variable $x_\epsilon(s)$ in (7.3.5), or even if the variable $y_\epsilon(s)$ is explicitly known.

In the case of the configurational first moment $\langle \boldsymbol{g}(0,s) \rangle$ for a rod in a semi-flexible regime the governing ODE is given in (7.3.2). We note that in general the averaging principle cannot be directly applied to this ODE system since, even if the semi-flexible assumption implies $\mathcal{L}(s) = \mathrm{O}(\epsilon)$ according to (3.4.19b) and (3.4.23), the expected chain generator vector $\bar{\xi}(s)$ is not always small, i.e. $\bar{\xi}(s) = O(1)$, as it is precisely the case for B-DNA for instance. However, using the factorization $\langle \boldsymbol{g}(0,s) \rangle = \langle \boldsymbol{\mathcal{D}}(0,s) \rangle \, \bar{\boldsymbol{g}}(0,s)$ with the persistence matrix $\langle \boldsymbol{\mathcal{D}}(0,s) \rangle$ and the expected rod configuration $\bar{\boldsymbol{g}}(0,s)$ presented in section 3.4.4, along with the fact that the variable $\bar{\boldsymbol{g}}(0,s)$ is an analytic interpolation by definition in (7.3.3a), it is equivalent to study the configurational first moment $\langle \boldsymbol{g}(0,s) \rangle$ or the persistent matrix $\langle \boldsymbol{\mathcal{D}}(0,s) \rangle$. More precisely, the ODE system governing the persistence matrix in (3.4.39) can be written in the form

$$\partial_s \langle \boldsymbol{\mathcal{D}}(0,s) \rangle = \langle \boldsymbol{\mathcal{D}}(0,s) \rangle \left\{ -\frac{1}{2} \bar{\boldsymbol{g}}(0,s) \mathcal{L}(s) \bar{\boldsymbol{g}}^{-1}(0,s) \right\} \tag{7.3.7a}$$

$$\partial_s \left\{ \bar{\boldsymbol{g}}(0,s) \mathcal{L}(s) \bar{\boldsymbol{g}}^{-1}(0,s) \right\} = \left\{ \bar{\boldsymbol{g}}(0,s) \mathcal{L}(s) \bar{\boldsymbol{g}}^{-1}(0,s) \right\} \mathcal{B}(s) \tag{7.3.7b}$$

with

$$\bar{\boldsymbol{g}}(0,s) \mathcal{L}(s) \bar{\boldsymbol{g}}^{-1}(0,s) = \mathrm{O}(\epsilon) \tag{7.3.7c}$$

$$\mathcal{B}(s) = \mathrm{O}(1) \,. \tag{7.3.7d}$$

according to the semi-flexible hypothesis for rods. We note that for rod theories obtained from the stochastic Cauchy-Born rule proposed in section 4.3, the hypothesis of the semi-flexible regime is verified if the underlying chain has this property according to the definition (7.3.3b). The second ODE (7.3.7b) is somehow artificial since the entries in the matrix $\bar{\boldsymbol{g}}(0,s) \mathcal{L}(s) \bar{\boldsymbol{g}}^{-1}(0,s)$ are explicitly known as a function of s. It emphasises however the scale separation observed between the characteristic size of the entries in this coefficient matrix, i.e. the characteristic size of the variation of the entries of the persistence matrix $\langle \boldsymbol{\mathcal{D}}(0,s) \rangle$, and the characteristic size of their own variation, denoted by the matrix $\mathcal{B}(s) \in \mathbb{R}^{4 \times 4}$, which is of order $\mathrm{O}(1)$.

Consequently, the configurational first moment $\langle \boldsymbol{g}(0,s) \rangle$ can be approximated by

$$\langle \boldsymbol{g}(0,s) \rangle \simeq \langle \boldsymbol{\mathcal{D}}(0,s) \rangle^{(H)} \, \bar{\boldsymbol{g}}(0,s) \tag{7.3.8a}$$

where the matrix $\langle \boldsymbol{\mathcal{D}}(0,s)\rangle^{(H)}$ is defined as the solution of the homogenised ODE

$$\partial_s \langle \boldsymbol{\mathcal{D}}(0,s)\rangle^{(H)} = \langle \boldsymbol{\mathcal{D}}(0,s)\rangle^{(H)} \left\{ -\frac{1}{2}\bar{\boldsymbol{g}}(0,s)\mathcal{L}(s)\bar{\boldsymbol{g}}^{-1}(0,s) \right\}_{\Delta s}^{(H)} \tag{7.3.8b}$$

with

$$\left\{ -\frac{1}{2}\bar{\boldsymbol{g}}(0,s)\mathcal{L}(s)\bar{\boldsymbol{g}}^{-1}(0,s) \right\}_{\Delta s}^{(H)} = \frac{1}{2\Delta s}\int_{s-\Delta s}^{s+\Delta s} \left\{ -\frac{1}{2}\bar{\boldsymbol{g}}(0,s')\mathcal{L}(s')\bar{\boldsymbol{g}}^{-1}(0,s') \right\} ds' \tag{7.3.8c}$$

for a fixed value of Δs, according to the averaging principle in (7.3.6).

Equivalently, in terms of the frame correlation matrix $\langle \boldsymbol{R}(0,s)\rangle$ and the Flory persistence vector $\langle \boldsymbol{r}(0,s)\rangle$ we have

$$\langle \boldsymbol{R}(0,s)\rangle \simeq \langle \boldsymbol{\Delta}(0,s)\rangle^{(H)}\, \bar{\boldsymbol{R}}(0,s) \tag{7.3.9a}$$

$$\langle \boldsymbol{r}(0,s)\rangle \simeq \langle \boldsymbol{\Delta}(0,s)\rangle^{(H)}\, \bar{\boldsymbol{r}}(0,s) + \langle \boldsymbol{\delta}(0,s)\rangle^{(H)} \tag{7.3.9b}$$

where the matrix $\langle \boldsymbol{\Delta}(0,s)\rangle^{(H)}$ and the triple $\langle \boldsymbol{\delta}(0,s)\rangle^{(H)}$ are defined as solutions of

$$\partial_s \langle \boldsymbol{\Delta}(0,s)\rangle^{(H)} = \langle \boldsymbol{\Delta}(0,s)\rangle^{(H)} \left\{ -\frac{1}{2}\bar{\boldsymbol{R}}(0,s)\mathcal{L}^{\boldsymbol{RR}}(s)\bar{\boldsymbol{R}}^{-1}(0,s) \right\}_{\Delta s}^{(H)} \tag{7.3.9c}$$

$$\partial_s \langle \boldsymbol{\delta}(0,s)\rangle^{(H)} = \langle \boldsymbol{\Delta}(0,s)\rangle^{(H)} \left\{ \frac{1}{2}\bar{\boldsymbol{R}}(0,s)\mathcal{L}^{\boldsymbol{RR}}(s)\bar{\boldsymbol{R}}^{T}(0,s)\bar{\boldsymbol{r}}(0,s) - \frac{1}{2}\bar{\boldsymbol{R}}(0,s)\mathcal{L}^{\boldsymbol{Rr}}(s) \right\}_{\Delta s}^{(H)} \tag{7.3.9d}$$

where we have used the notation $\mathcal{L}^{\boldsymbol{RR}}(s) = \mathrm{Tr}\left(\mathcal{C}^{\boldsymbol{RR}}(s)\right)\mathrm{Id}_3 - \mathcal{C}^{\boldsymbol{RR}}(s)$ and $\mathcal{L}^{\boldsymbol{Rr}}(s) = \mathcal{C}^{\boldsymbol{Rr}}(s)$, as in (3.4.40), and where $\{\,.\,\}_{\Delta s}^{(H)}$ has to be understood in the same way as in (7.3.8c).

For the case of the stochastic Cauchy-Born rule presented in (7.3.3b), the optimal choice of the value Δs requires further investigations but for illustration of the proposed averaging principle we use $\Delta s = 17$Å, i.e. approximately 5 base pairs. Figure 7.24 shows the coefficient matrices in (7.3.9c) in their original and in their homogenised version for the sequence denoted by λ . Similarly, figure 7.25 shows the coefficients with and without homogenisation for the ODE (7.3.9d). The original discontinuous coefficient are shown in blue and green respectively, similarly to the ones of the rod stiffness matrix $\bar{\mathcal{K}}(s)$ shown in figures 5.9 and 5.11 for instance, and the homogenised coefficient are shown in both figure in orange. Interestingly, the averaging principle for the choice $\Delta s = 17$Å delivers very slowly varying coefficients. For the rotation part $\langle \boldsymbol{\Delta}(0,s)\rangle$, in figure 7.24, they are observed to stay at a constant magnitude with some local and sequence dependent variations, and where the three curves with ordinate values between -3 and -2 correspond to the diagonal entries in the homogenised coefficient matrix. In contrast, the observed growth in figure 7.25 is explained by the terms $\bar{\boldsymbol{r}}(0,s)$ in the corresponding expression (7.3.9d).

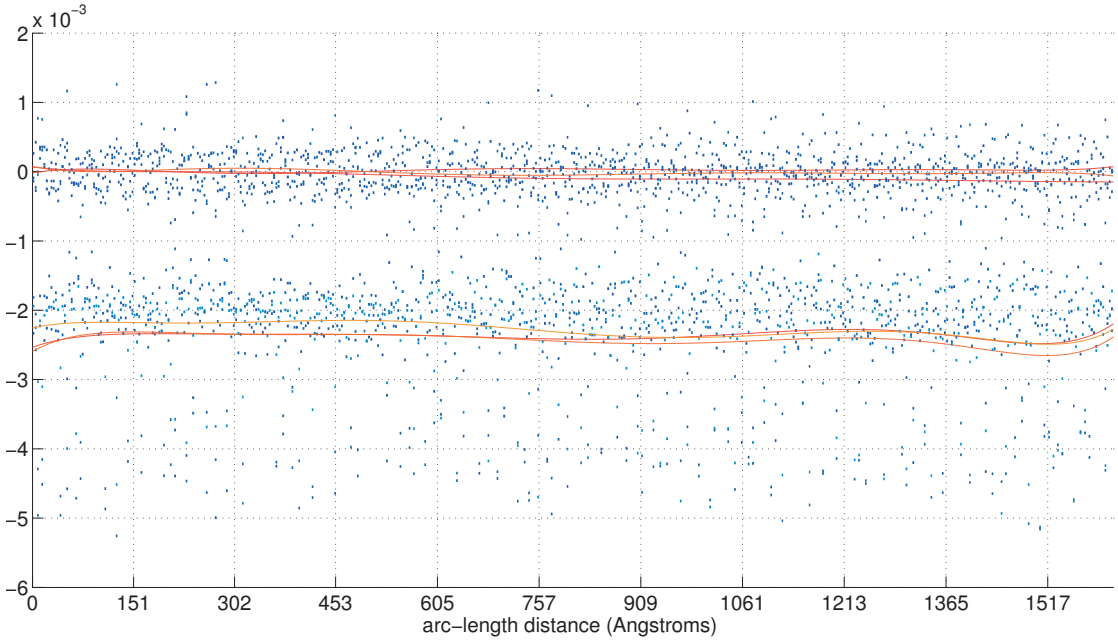


Figure 7.24: Original (blue) and homogenised (orange) coefficient in (7.3.9c) for the sequence λ .

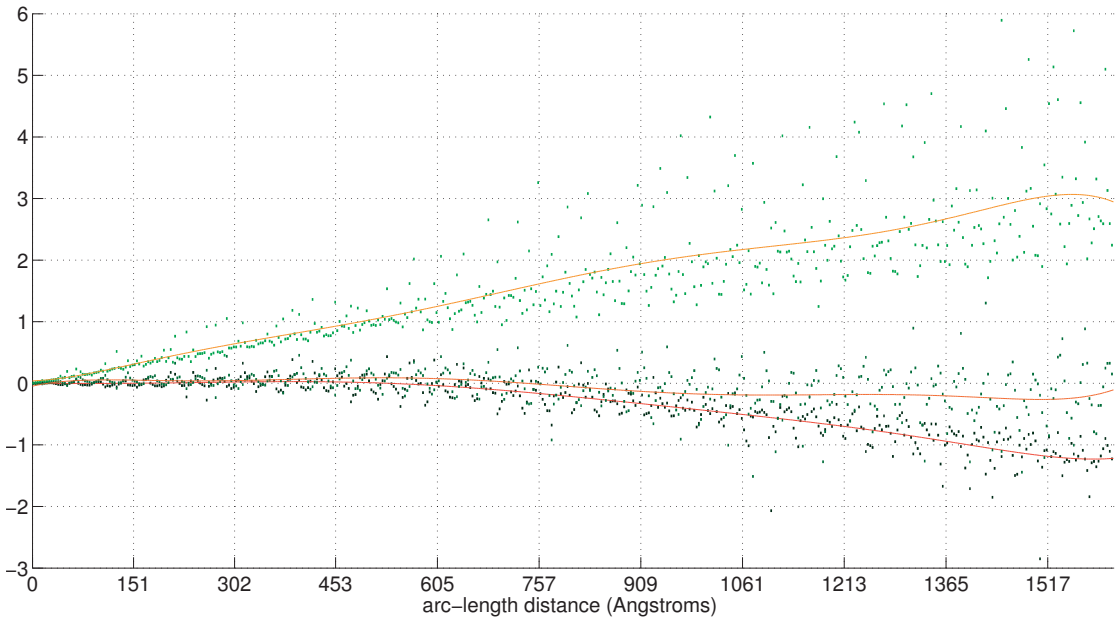


Figure 7.25: Original (green) and homogenised (orange) coefficient in (7.3.9d) for the sequence λ .

In order to quantify the difference between the computation of the configurational first moment $\langle \boldsymbol{g}(0,s)\rangle$ using the original ODE in (7.3.2) and using the homogenised version of the persistence matrix together with the relation (7.3.8a), we propose to use the same numerical method[2] as the one used to obtain the solutions shown in figure 7.20, 7.21,7.22 and 7.23 but now applied to the homogenised system. Figures 7.26, 7.27, 7.28 and 7.29 show, using a continuous line, the results for the configurational first moment $\langle \boldsymbol{g}(0,s)\rangle$ of the sequence λ approximated numerically from the system in (7.3.8) evaluated at the position of base pairs, and, using the symbol $\circ$, the analogous results obtained from a Monte Carlo simulation. The relative error between the numerical solution obtained for the original and for the homogenised system, expressed at the base pair position, is presented in figure 7.30.

Since the coefficient obtained using the averaging principle are smooth and slowly varying, the number of numerical steps needed to solve the associated ODE are considerably less than for the original one for the same prescribed tolerance, specifically a decrease by a factor of 10^2 is obtained. Table 7.5 summarises the number of steps needed with and without homogenisation for the sequences λ , CF and T_L respectively. We observe, moreover, that the number of steps for the homogenised system is approximately a third of the actual number of base pairs in each case.

Table 7.5: Comparison between the number of reported steps to solve numerically the original and the homogenised system

	λ	CF	T_L
Without homogenisation	14832	14863	14984
With homogenisation	169	169	134
Ratio	1.1%	1.1%	0.8 %

[2] This method is *ode45* provided by the software *matlab*

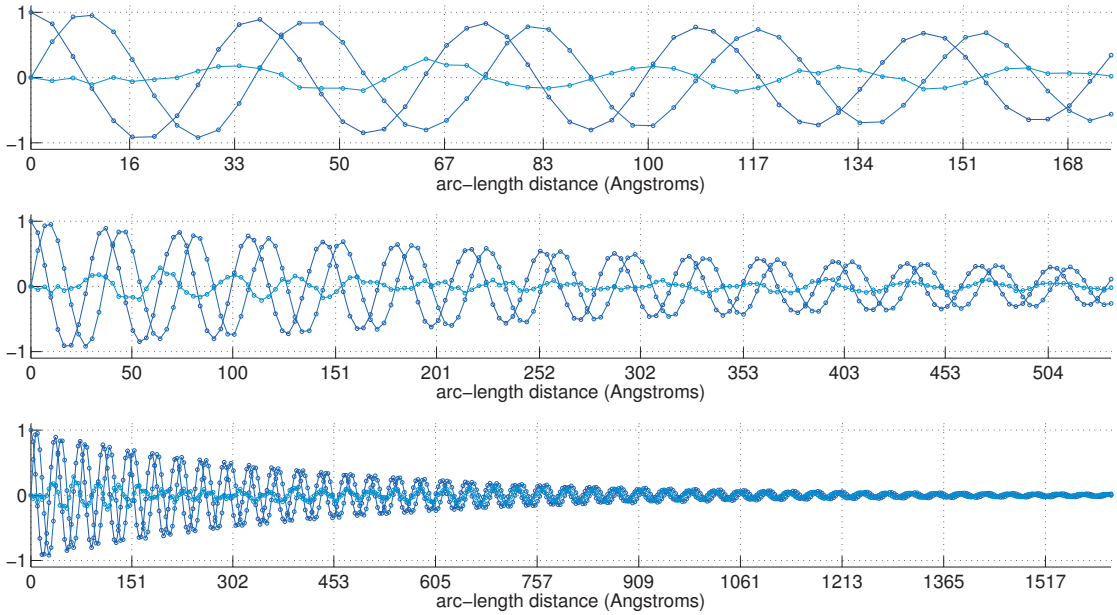


Figure 7.26: First column of the frame correlation matrix $\langle \boldsymbol{R}(0,s)\rangle$ for the sequence λ and for three different molecule lengths. Monte Carlo simulations ($\circ$) versus the analytical predictions using the averaging principle ($-$).

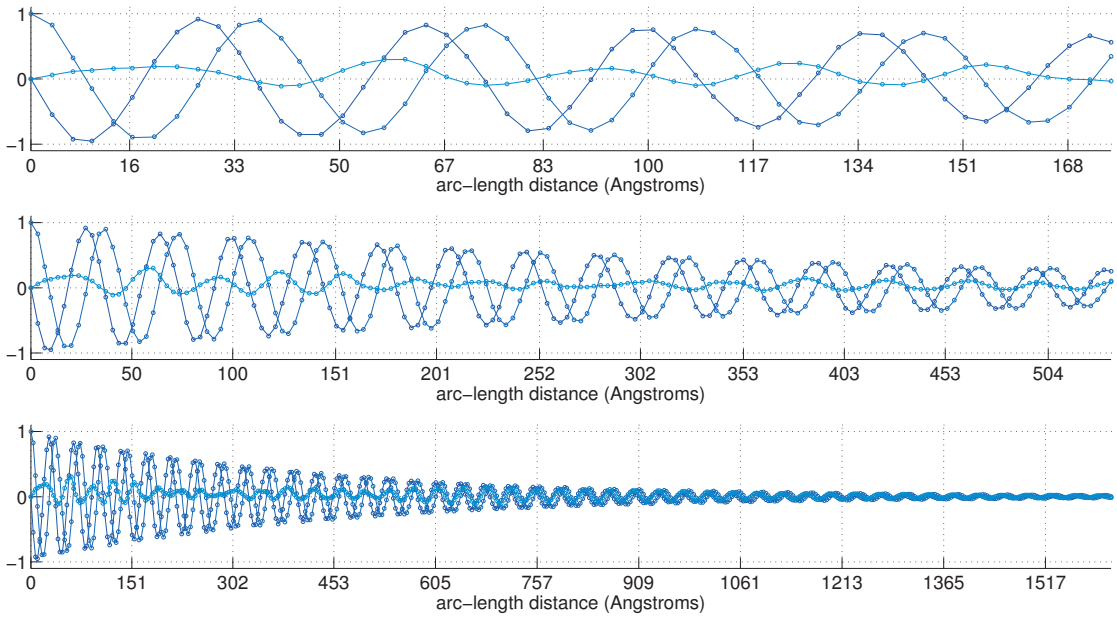


Figure 7.27: Second column of the frame correlation matrix $\langle \boldsymbol{R}(0,s)\rangle$ for the sequence λ and for three different molecule lengths. Monte Carlo simulations ($\circ$) versus the analytical predictions using the averaging principle ($-$).

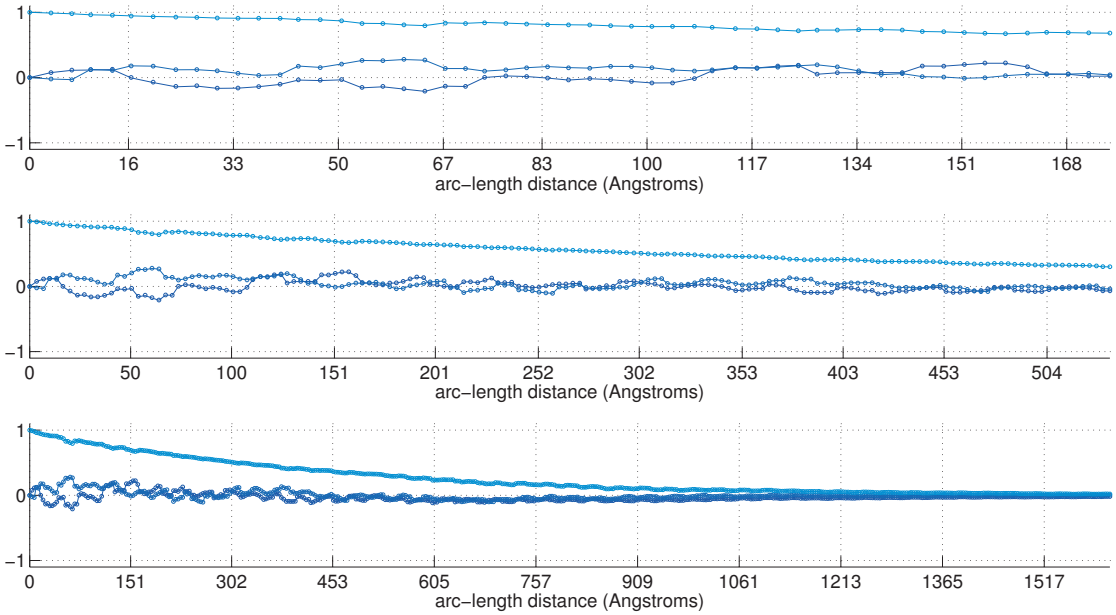


Figure 7.28: Third column of the frame correlation matrix $\langle \boldsymbol{R}(0,s)\rangle$ for the sequence λ and for three different molecule lengths. Monte Carlo simulations ($\circ$) versus the analytical predictions using the averaging principle ($-$).

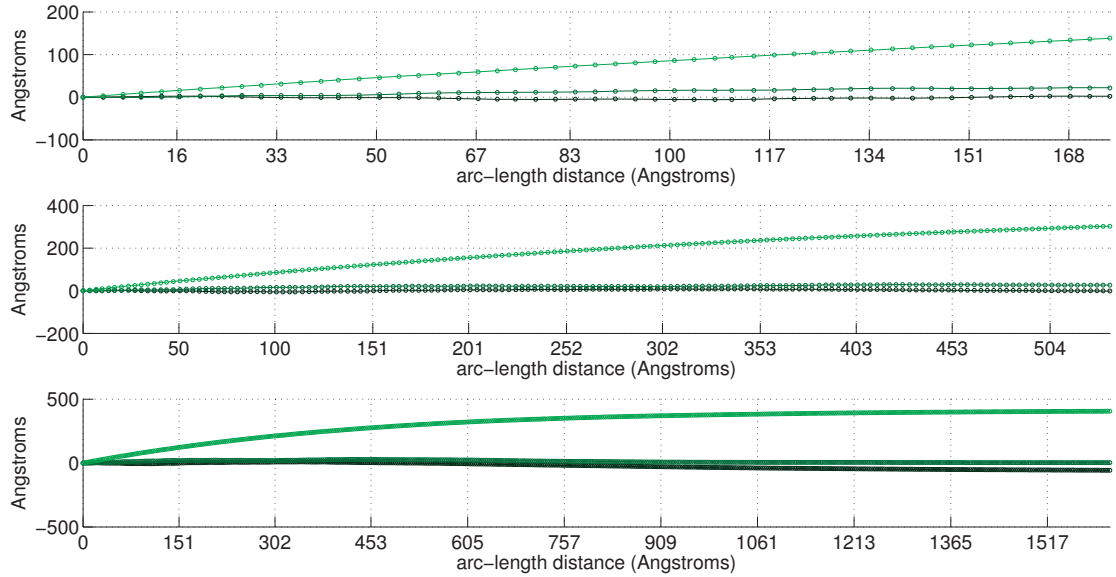


Figure 7.29: Component of the Flory persistence vector $\langle \boldsymbol{r}(0,s)\rangle$ for the sequence λ and for three different molecule lengths. Monte Carlo simulations ($\circ$) versus the analytical predictions using the averaging principle ($-$).

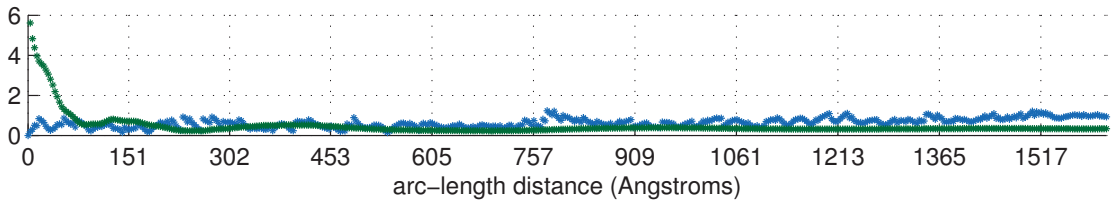


Figure 7.30: Relative errors between the frame correlation matrix $\langle \boldsymbol{R}(0,s)\rangle$ (blue) and for Flory persistence vector $\langle \boldsymbol{r}(0,s)\rangle$ (green) associated to the sequence λ computed using the original and the homogenised system, as a function of the molecule length.

It is remarkable that the use of the persistence matrix factorization together with the application of an averaging principle, as stated in (7.3.9), allows an accurate approximation of the sequence dependent configurational first moment $\langle \boldsymbol{g}_{1,n}\rangle$, obtained using a Monte Carlo simulation, by the solution $\langle \boldsymbol{g}(0,s)\rangle$ of a homogenised ODE involving a discretization strictly coarser than the original number of base pairs. The results, even if they require a more detailed analysis, are believed to be encouraging in the multi-scale modelling of DNA statistical physics and in the definition of sequence dependent large scale models.

Conclusions

In order to design a framework for the multi-scale modelling of the sequence dependent physical properties of DNA molecules, we have proposed, in part I, a study of the properties of chains and rods in both the context of classical statics and statistical mechanics, as well as a method to bridge these models. In part II, we have discussed the actual application to sequence dependent DNA models to compute most probable configurations satisfying prescribed end constraints and to approximate the value of configurational first moment.

More precisely, in part I we presented a general discussion on the equilibrium conditions and related variational principles, as well as on stationary configurational distributions and observables, for rigid body chains and bichains and continuum rods and birods. We also explained in which sense continuum models can provide approximations of discrete model properties. These results could be of potential interest in physical contexts other than DNA, since they comprise general statements on the multi-scale modelling of filamentary materials.

The second chapter has been concentrated on the rigid body chain and rigid body bichain models. We discussed precisely the construction of the bichain model from a double chain model and the statement of related equilibrium equations and variational principles, which in particular apply directly to the sequence dependent rigid base model in [Gonzalez et al., 2013]. For both chain and bichain equilibrium configurations the existence of six discrete first integrals is emphasised in (2.2.11) and in (2.2.43) respectively. Moreover, the Lagrange multiplier formulation of the equilibrium conditions is shown to precisely give multipliers corresponding to the external couple and force end loading needed to realise the desired Dirichlet conditions, according to (2.2.53). An interesting extension of these results would be to compute, in an analogous way, the second variation of the internal energy. The stationary configurational distributions for chains and bichains has been presented following arguments proposed in [Walter et al., 2010]. In the case of a single chain with a local deformation energy we have developed general analytical results on the configurational first moment which are of interest since it comprises the frame correlation matrix and the Flory persistence vector. The expression in (2.4.26) demonstrates the equal importance of the expected chain configuration $\bar{\boldsymbol{g}}_{m,n}$ and the local fluctuation matrices in the behaviour of the first moment $\langle \boldsymbol{g}_{m,n} \rangle$. It generalises the work in [Trifonov et al., 1988, Schellman and Harvey, 1995] on the notion of persistence length in DNA since it supports the idea of contributions of a deterministic part and of stochastic part to the local chain statistics, but it encodes it into two 4×4 matrices

instead of two scalar static and the dynamic persistence lengths. Moreover, this explicit formula motivates a factorization of the configurational first moment into the persistence matrix and the expected chain configuration, specifically $\langle \boldsymbol{g}_{m,n} \rangle = \langle \boldsymbol{\mathcal{D}}_{m,n} \rangle \, \bar{\boldsymbol{g}}_{m,n}$. The persistence matrix is shown to characterise the memory lost in the entries of the configurational first moment $\langle \boldsymbol{g}_{m,n} \rangle$, according to (2.4.44), and leads to a closed form expression in a short length expansion in (2.4.55). The extension of these discussions to the configurational first moment to the case of bichains would be a mathematically, and scientifically, important improvement in sequence dependent DNA modelling. However, further investigations are needed to handle the extra couplings, illustrated in figure 5.3, present in the bichain model.

The third chapter is dedicated to continuum rods and birods and shows that a completely analogous discussion to chains and bichains can be presented using continuum definitions. Following the original work in [Moakher and Maddocks, 2005], we defined birod configurations, which are the continuum analogue of bichains. The variational principle for rod equilibrium conditions, as originally discussed in [Cosserat and Cosserat, 1909], is extended to birods and, in addition, a coordinate independent Hamiltonian formulation is proposed in (3.2.71). The existence of six continuum first integrals is emphasised for both rod and birod equilibrium configurations in (3.2.12) and in (3.2.52). Similarly to chains, we present a Lagrange multiplier formulation of the equilibrium birod conditions with prescribed end configurations, and the set of multipliers are shown in (3.2.83) to correspond to external couple and force end loadings. It would be useful to study the second variation of the birod internal energy in a future work. Statistical physics properties of rods and birods was then studied. The configurational distribution for rods, and respectively birods, in the context of equilibrium statistical mechanics are introduced as a path integral density, and we discussed how these infinite dimensional distributions can be interpreted as the appropriate limit of a sequence of chain distributions for the discretised rod configurations in (3.3.16), and for birods in (3.3.32). Interestingly the effect of the Jacobian factor coming from the intrinsic geometry on the rotation group SO(3) is observed to vanish when the rod limit is reached, while it stays for the birod microstructure as discussed in (3.3.33). This development is then extended to observables and their expectations and, similarly to chains, we emphasize the special role of the configurational first moment which comprise, here in a continuum version, the frame correlation matrix and the Flory persistence vector. An explicit first order approximation is then obtained in (3.4.21) for the configurational first moment for a single rod which has a local internal energy and is assumed to be in a semi-flexible regime. A similar expression has been discussed in [Becker, 2007] where it was deduced by analogy to the Brownian dynamics of a rigid body. As in the case of chains, the analytical expressions show the equal importance of the expected rod configuration and of the fluctuation matrix. They also motivate the definition of the persistence matrix which is shown to characterise the memory lost in the entries of the configurational first moment, according to (3.4.39) and admit a closed form expression in a short length expansion (3.4.46). As in the case of chains, the extension of the results for the configurational first moment for birods is believed to be an important future objective.

In chapters 2 and 3, the strong analytic parallel between chain and rod formulations has been explicitly demonstrated. Many expressions, such as the equilibrium conditions for bichains (2.2.53) and birods (3.2.83) and their associated set of discrete and continuum first integrals (2.2.43) and (3.2.52), or the evolution equations for the configurational first moment chains (2.4.26) and rods (3.4.21), are literally the respective discrete and continuum analogues. The fundamental similarities between these two mechanical descriptions is not only appreciable, but is also encouraging us in the definition of multi-scale mechanical models of filamentary objects. One result which does not have its explicit discrete analogue is the Hamiltonian formulation of rods and birod equilibrium conditions. The missing step is the definition an appropriate of Legendre transform for the variational form of the chain constitutive relation in (2.2.12).

The fourth chapter first discussed the mathematical framework in which we believe that the mechanical results obtained for chains and rods in chapter 2 and 3 can be related. The underlying definitions to the idea of bridging the scales between chain and rod mechanics are presented, from the point of view of statics in (4.1.15) and of equilibrium statistics in (4.1.25). The entire discussion is based on the notion of consistency. We briefly discussed that in general deterministic exponential Cauchy-Born rules do not provide stochastic exponential Cauchy-Born rules. A more detailed analysis of questions concerning the consistency of stationary configurations for consistent discrete and continuum energies would be of great interest. The discussion of hypotheses leading to the equivalence of the deterministic and stochastic exponential Cauchy-Born rules also remains also to be considered. An explicit deterministic Cauchy-Born rule has been proposed for quadratic bichains and birods models in (4.2.11) and (4.2.14) in order to build a rigorous continuum analogue of the sequence dependent rigid base model in [Gonzalez et al., 2013]. As a corollary, the discussion made for bichains and birod provides an original deterministic exponential Cauchy-Born rule for single chains and rods. We then discussed the construction of an explicit stochastic Cauchy-Born rule (4.3.7) and in (4.3.10) for single chains and rods governed by normal configurational distributions and in a semi-flexible regime. The introduction of an explicit stochastic Cauchy-Born rule for bichains would be of interest, but also requires a better understanding of the normal distribution associated to the quadratic form (4.2.1) which has a banded, but non-local, structure. In addition to the work on carbon nanotubes already discussed in section 4.1, we note that continuum models for the macroscopic description of crystal lattices has been intensively studied in the literature as a limit of particle systems, see for instance [Askar, 1986, Triantafyllidis and Bardenhagen, 1993]. The appropriate notion of convergence, called $\Gamma-$*convergence*, has been discussed[3] in the limit when the lattice becomes dense in the continuum media, see for instance [Braides and Maria, 2006, E and Ming, 2006, Braides and Truskinovsky, 2008]. In the context of chains and rods, or bichains and birods, the macroscopic description requires an appropriate helical interpolation, as described in 4.2.4 for instance, because of the specific geometric structure of the special

[3]These approaches to derive continuum from discrete models should not be confused with other applications of the $\Gamma-$convergence in the derivation of rod models from three dimensional elasticity [Friesecke et al., 2006].

Euclidean group SE(3). The deterministic and stochastic exponential Cauchy-Born rules define respectively a sequence of models indexed by the number of base pairs N that are governed by an continuum internal energy of the form $\hat{E}^{(N)}$, in the deterministic case, or $\bar{E}^{(N)}$, in the stochastic case. The discussion of an appropriate $\Gamma-$limit $\hat{E}^{(\infty)}$ or $\bar{E}^{(\infty)}$ would then be an interesting extension of the analysis presented in this thesis. However, in the context of DNA modelling, it is not clear that these $\Gamma-$limits will provide accurate continuum models at the biologically important scale of a few hundreds of base pairs.

Even if the frontier between rigid body chain and continuum rod mechanical descriptions remains to be further analysed, we believe that the discussion made in part I provides original and interesting insights on the multi-scale modelling of filamentary materials, including DNA molecules. An interesting extension would be to discuss the construction of an effective rod model from a given birod model at long length scales in order to complete the multi-scale picture presented in figure 1.

Part II of the thesis is dedicated to the application of chain and rod theories to the multi-scale modelling of sequence dependent DNA mechanics. We presented a computation of the most probable bichain configuration using the sequence dependent birod model and the approximation of the chain configurational first moment using a sequence dependent rod model.

The fifth chapter discussed the sequence dependent rigid base pair and rigid base model and the introduction of their continuum analogues as rod and birod models. More precisely, the birod model has been parametrized using the deterministic exponential Cauchy-Born rule for bichains since we were targeting the approximation in chapter 6 of stationary configurations in the rigid base model, according to the argument given in section 4.1.2. Moreover, no definition of an explicit stochastic Cauchy-Born rule for bichains is available. Conversely, the continuum rod model has been defined using the stochastic Cauchy-Born rule for chains with the goal of approximating the configurational first moment in chapter 7 according to the construction in section 4.3. In both cases, the continuum constitutive coefficients are observed to be discontinuous at each base pair position and to have a strongly heterogeneous character, as illustrated in figures 5.11 and 5.15. However, concerning the sequence dependent birod model we are very satisfied with the results illustrated in figures 5.16 and 5.18 about consistency with the original rigid base energy definition.

The sixth chapter presented the application of the sequence dependent birod model to approximate the most probable configurations respecting end conditions in the rigid base model. We have, however, only discussed stationary configurations since the theory of the second variation is still under development. The Hamiltonian form of the equilibrium conditions is presented in (6.2.3) and the coefficient matrix obtained using the deterministic exponential Cauchy-Born rule corresponding to the sequence λ is illustrated in figure 6.1. The sequence dependence is reflected in the heterogeneous character of these coefficients. The numerical solutions of the continuum equilibrium conditions for the first 240 base pairs of the sequence λ , CF and T_L for under and over twisted end conditions, corresponding respectively to -300pNÅ and $+300$pNÅ, are then computed using the software bBDNA developed by *J. Glowacki*. A significantly different mechan-

ical response is observed both between different sequences and between under and over twisted end conditions. A more detailed analysis of the effect of sequence on the birod equilibrium configuration for prescribed end conditions remains to be carried out, but we believe that the ground state geometry has an important interplay with the applied end constraint. In order to assess the quality of these continuum approximations, we have then discussed the discrete equilibrium conditions obtained for the sequence dependent rigid base model in 6.3.1. Numerical solutions were computed by *A. Patelli* using the iterative root finding algorithm *fsolve* from the *matlab* software library. Even though comparison of the discrete and continuum solutions, as shown in figure 6.5 for instance, is satisfactory, a more detailed analysis of the ability of the continuum solutions to reproduce bichain stationary configurations remains to be investigated. The main negative conclusion of this approach is that the number of nodes used in the numerical discretization of the continuum model is considerably higher than the number of base pairs involved in the discrete model, which is because of the heterogenous character of the sequence dependent coefficients. Consequently, even if these continuum computations are very fast, and moreover provides an initial configuration for the discrete system, the interest of the continuum model is still questionable. We believe however that the definition of this sequence dependent continuum model is valuable, at least mathematically, since it provides the necessary first step in the definition of rigorous large scale DNA models, and, consequently, in the construction of the multi-scale modelling of DNA molecule statistical physics. An implementation of the local averaging principle, presented in the last chapter, has already been done and it allows the choice of different averaging window sizes along the sequence which is an encouraging method to describe some region in a very fine way and others more coarsely. Nevertheless, this is only a preliminary work. We would also like to mention that the software bBDNA provides the possibility of computing equilibrium birod configuration with periodic boundary conditions which are closed, and possibly knotted, loops. We believe this option to be interesting since circular DNA arise frequently in Nature, as in bacterial chromosomes for instance, or are used in vitro experiments. These special configurations are further discussed in the doctoral thesis[4] of *J. Glowacki*, an example of such a closed solution is shown in figure 7.31.

[4] in preparation

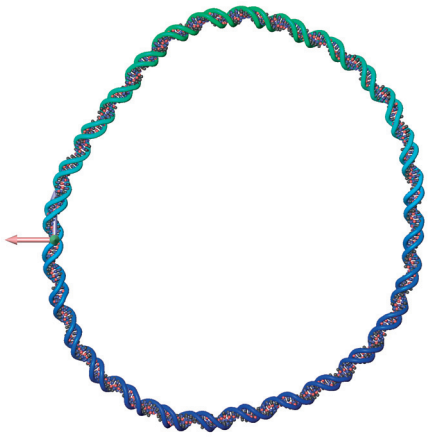


Figure 7.31: A closed loop sequence dependent birod equilibrium configuration.

The last chapter presented a detailed study of the configurational first moment for the sequence dependent rigid base pair model. We first discussed the application of the explicit recurrence relation obtained in (2.4.26) using an expansion justified for a rigid body chain in a semi-flexible regime and whose configurational distribution factorizes into local contributions. This last hypothesis is not verified for the rigid base model in [Gonzalez et al., 2013], but we proposed a modified version by an appropriate truncation of the covariance matrix, as described in (7.2.2), motivated by a maximum entropy principle. The resulting model is a sequence dependent rigid base pair model and, even if its constitutive parameters are believed to be less realistic than the original rigid base model, it constitutes a serious example of the theory for heterogeneous chains developed in section 2.4 since versions of this model are frequently used in the DNA literature. The non-dimensional parameter ϵ measuring the semi-flexibility defined in (2.4.28) is found to be of the order 10^{-2} in this model. The comparison between the numerical values obtained for the frame correlation matrix and for the Flory persistence using the deduced recurrence relation and a Monte Carlo simulation is illustrated in figures 7.2, 7.3, 7.4 and 7.5 for the sequence λ . The difference is hardly perceptible even after 500 base pairs and the relative errors are presented in figure 7.6 as a function of the base pair distance. While the actual behaviour of the entries of the configurational first moment $\boldsymbol{g}_{1,n}$ varies between the sequences λ , CF and T_L , the similar notably good agreement is provided by the explicit recurrence relation in each case. The proposed analytic expressions are thus believed to provide an accurate approximation of the sequence dependent configurational first moment $\boldsymbol{g}_{1,n}$. The factorization into the expected chain and the persistence matrix is then presented. Figure 7.10 illustrates this factorization for the frame correlation matrix associated to the sequence λ . Remarkably it shows that the fast variation in the configurational first moment reflects the expected chain configuration whereas the slow convergence of the entries is due to the persistence matrix. The analogous result for the Flory persistence vector is presented in figure 7.11. This factorization supports the idea, originally proposed in [Trifonov et al., 1988] and then discussed in a slightly

more general context in [Schellman and Harvey, 1995], that the statistics of semi-flexible chains have to be made in general of a deterministic part, reflecting the intrinsic geometry of the molecule, and of a stochastic part, reflecting thermal fluctuations. The proposed expression constitutes, in a mathematical sense, a generalisation of the studies [Trifonov et al., 1988, Schellman and Harvey, 1995] where these two contributions, also called static and dynamic parts, have been encoded through scalar values while here they are 4×4 matrices, namely the expected chain configuration $\bar{\boldsymbol{g}}_{1,n}$ and the persistence matrix $\langle \boldsymbol{\mathcal{D}}_{1,n} \rangle$. Moreover, this factorization allows, using properties of the persistence matrix, the derivation of a closed form expression for the configurational first moment in a short length expansion. It is observed to be valid until a distance of about 135 base pairs, as drawn in figures 7.14 and 7.15. We remark that all proposed analytic results are only rigorously valid for configurational distributions which factorizes into local contributions. In future work, it would be of interest, both mathematically and physically, to extend this approach to the sequence dependent rigid base model, but the objective requires a better understanding of the basic statistical properties of the normal distribution associated to the banded quadratic form in (4.2.1). The sequence dependent rod model, based on the application of the stochastic exponential Cauchy-Born rule, is used to provide continuum approximations of the chain configurational first moment. Based on the semi-flexible assumption for rods, we deduced in (3.4.21) a governing ODE system for the continuum configurational first moment. Using the standard *ode45* solver in *matlab*, we have computed numerically the solution of these ODE for the sequence λ . These results are compared with ones coming from a Monte Carlo simulation associated to the original rigid base pair description. Remarkably, the difference between the values is hardly perceptible, just as for the discrete approximation. The exponential stochastic Cauchy-Born rule for chains is thus believed to deliver an accurate continuum description of the sequence dependent rigid base pair model statistics and to comprise a significant step towards the multi-scale modelling of DNA physical properties. However, since these expressions are only valid for single chain models, the analogous results for bichains and birods configurational distributions would be of great interest. Similarly to the computation of continuum birod equilibrium configurations in chapter 6, the heterogeneous character of the coefficient matrix forced the use of a very dense numerical discretization to solve the configurational first moment ODE and consequently the continuum model appears to be more expensive than the original rigid base pairs model. To overcome this fundamental issue we proposed to use the homogenisation method in (7.3.6) based on a local averaging principle, with an explicit window size since none of the classical hypothesis, such as periodicity or ergodicity, are satisfied by the sequence dependent coefficients. Interestingly, a rigorous and direct application of the averaging principle to the ODE governing the configurational first moment is observed to not always be possible because of the presence in the coefficient matrix of the expected configuration generator vector, which is not always small as it is precisely the case for DNA continuum representations. However, the equivalent persistence matrix ODE for semi-flexible rods in (7.3.7) is shown to satisfy the appropriate hypotheses on scale separation. The definition of an optimal window size remains to be further investigated. We chose 30Å to illustrate our theory

which leads to the set of smooth and slowly varying homogenised coefficients presented in figures 7.24 and 7.25. The approximation of the configurational first moment $\langle \boldsymbol{g}(0,s)\rangle$ is then defined by an analytic change variables applied to the numerical solution of the homogenised system, as described in (7.3.8a). Figures 7.26, 7.27, 7.28 and 7.29 show the results for the configurational first moment $\langle \boldsymbol{g}(0,s)\rangle$ for the sequence λ and evaluated at the position of base pairs, and the analogous results obtained using the Monte Carlo simulation for the original sequence dependent rigid base pair model. The agreement is here again particularly good, but the significant additional characteristic in these examples is that the number of steps used in the numerical method is about a third of the original number of base pairs, as shown in table 7.5. Even if a more detailed analysis on the approximation of the original chain configurational first moment by the continuum solution of an homogenised ODE system remains to be done, these results, are believed to be not only encouraging in the multi-scale modelling of DNA statistical physics but also in the definition of sequence dependent large scale models.

The application of both the deterministic and the stochastic exponential Cauchy-Born rules to sequence dependent rigid base and rigid base pair models of DNA has allowed the definition of sequence dependent continuum birod and rod models. The use of these models to approximate discrete chains properties, such as equilibrium configurations or the configurational first moment, but in continuum settings has been shown to give results with a reasonable accuracy. In addition a promising route towards large scale descriptions has been established through a homogenisation method.

In conclusion, the study of the statistical physics properties of DNA molecules can only be done in an accurate and efficient way through sequence dependent and multi-scale mechanical models. The proper foundation of such models requires a fine understanding of both heterogeneous rigid body chains and continuum rod models, as well as practical ways to bridge, in a mathematical and in a physical sense, these two descriptions. According to the theoretical and concrete results presented in this thesis, we believe we have presented a framework which, even if it still requires improvement, is a significant contribution to the understanding and the modelling of sequence dependent DNA mechanics.

A. Appendices

We provide here a set of heterogenous sections each of which comprises technical developments related to the results of this thesis.

A.1 On the Expansion of the cgDNA Internal Coordinates

We present here a way to write down explicitly the expansion (1.3.4) for the specific choice of the variables $x_n = (u_n, v_n)$ used in (2.1.18), i.e. where u_n denotes the Cayley vector of the rotation matrix $Q_n = Q(u_n)$ and where v_n parametrized the translation vector $q_n = Q_n^{\frac{1}{2}} v_n$ expressed in the junction frame. Similarly to (2.1.5), we write the block structure of the terms in the expansion (1.3.4) for $k = 1, 2$ as

$$a_{\bar{x}_n}^{(k)} = \begin{pmatrix} Q_{\bar{x}_n}^{(k)} & q_{\bar{x}_n}^{(k)} \\ 0 & 0 \end{pmatrix} \tag{A.1.1}$$

where we have removed the explicit dependence on $(x_n - \bar{x}_n)$ for brevity. The expansion of the individual blocks of a_n reads

$$Q_n = \left(\mathrm{Id}_3 + Q_{\bar{x}_n}^{(1)} + \frac{1}{2} Q_{\bar{x}_n}^{(2)} \right) \bar{Q}_n \tag{A.1.2a}$$

$$q_n = \left(\mathrm{Id}_3 + Q_{\bar{x}_n}^{(1)} + \frac{1}{2} Q_{\bar{x}_n}^{(2)} \right) \bar{q}_n + q_{\bar{x}_n}^{(1)} + \frac{1}{2} q_{\bar{x}_n}^{(2)}. \tag{A.1.2b}$$

To compute the terms $Q_{\bar{x}_n}^{(1)}$ and $Q_{\bar{x}_n}^{(2)}$ we use an interesting feature of the Cayley parametrization of the rotation group in (2.1.16), which is that it allows a direct expansion around any Cayley vector $\bar{u}_n \in \mathbb{R}^3$ using the following *Neumann expansion*

$$(A - B)^{-1} = A^{-1} + A^{-1} B A^{-1} + A^{-1} B A^{-1} B A^{-1} + o\left(\left| A^{-1} B \right|^2 \right). \tag{A.1.3}$$

Precisely, applying the expansion (A.1.3) in the expression (2.1.16), with $A = \mathrm{Id}_3 - \frac{1}{2} [\bar{u}_n \times]$, gives

$$Q(u_n) = Q(\bar{u}_n) + \sum_{k=1}^{j} \frac{1}{2^{k-1}} E(\bar{u}_n) [\Delta u_n \times] \ldots E(\bar{u}_n) [\Delta u_n \times] E(\bar{u}_n) + \mathrm{o}\left(|\Delta u_n|^j \right) \tag{A.1.4}$$

where we have used the notation

$$E(\bar{u}_n) := \left(\mathrm{Id}_3 - \frac{1}{2}\left[\bar{u}_n\times\right]\right)^{-1} = \frac{1}{2}\left(Q(\bar{u}_n) + \mathrm{Id}_3\right) \tag{A.1.5}$$

and $\Delta u_n := u_n - \bar{u}_n$. Note that the term $[\Delta u_n\times]$ appears exactly k times in the k^{th} term of the sum in (A.1.4). If one uses the identity

$$E\left[y\times\right]E^T = \det E\left[(E^{-T}y)\times\right] \tag{A.1.6}$$

for all invertible $E \in \mathbb{R}^{3\times3}$ and $y \in \mathbb{R}^3$, one can rewrite (A.1.4), since $E(\bar{u}_n)^{-T}E(\bar{u}_n) = Q(\bar{u}_n)$, as

$$Q(u_n) = \left\{\mathrm{Id}_3 + \sum_{k=1}^{j}\frac{1}{2^{k-1}}Q^{(k)}_{\bar{u}_n}(\Delta u_n)\right\}Q(\bar{u}_n) + \mathrm{o}\left(|\Delta u|^j\right) \tag{A.1.7}$$

where $Q^{(k)}_{\bar{u}_n}(\Delta u_n)$ is k-multilinear in Δu_n and is given by

$$Q^{(k)}_{\bar{u}_n}(\Delta u_n) = \left[\mathbb{P}_1(\bar{u}_n)\Delta u_n\times\right]\left(\mathrm{Id}_3 + \frac{1}{2}\left[\bar{u}_n\times\right]\right)\left[\mathbb{P}_1(\bar{u}_n)\Delta u_n\times\right]\ldots\left(\mathrm{Id}_3 + \frac{1}{2}\left[\bar{u}_n\times\right]\right)\left[\mathbb{P}_1(\bar{u}_n)\Delta u_n\times\right] \tag{A.1.8}$$

with by definition

$$\mathbb{P}_1(\bar{u}_n) := \det E(\bar{u}_n)E^{-T}(\bar{u}_n) = \frac{1}{1+\left(\frac{|\bar{u}_n|}{2}\right)^2}\left(\mathrm{Id}_3 + \frac{1}{2}\left[\bar{u}_n\times\right]\right) \tag{A.1.9}$$

and where the term $[\mathbb{P}_1(\bar{u}_n)\Delta u_n\times]$ appears exactly k times. In particular, the two first matrices $Q^{(k)}_{\bar{u}_n}$ can be expressed as

$$Q^{(1)}_{\bar{u}_n}(\Delta u_n) = \left[\gamma^{(1)}_n\times\right] \tag{A.1.10a}$$

$$Q^{(2)}_{\bar{u}_n}(\Delta u_n) = \left[\gamma^{(1)}_n\times\right]^2 + \left[\gamma^{(2)}_n\times\right] \tag{A.1.10b}$$

with the triples $\gamma^{(1)}_n, \gamma^{(2)}_n \in \mathbb{R}^3$ defined as

$$\gamma^{(1)}_n = \mathbb{P}_1(\bar{u}_n)\Delta u_n \tag{A.1.11a}$$

$$\gamma^{(2)}_n = -\frac{1}{2}\mathbb{P}_1(\bar{u}_n)(\Delta u_n \otimes \Delta u_n)\bar{u}_n. \tag{A.1.11b}$$

Unfortunately, the computation of the terms $q^{(1)}_{\bar{x}_n}$ and $q^{(2)}_{\bar{x}_n}$ is less straight forward because of the presence of the matrix $Q_n^{\frac{1}{2}}$ in the definition of q_n. The difficulty is to compute a second order expansion of the matrix $Q_n^{\frac{1}{2}} = Q^{\frac{1}{2}}(u_n)$ with respect to the Cayley vector u_n. To do so, we will use the expansion (A.1.7) twice.

First, since $Q^{\frac{1}{2}}(u)$ is also a rotation matrix there exists, in analogy to (A.1.10), coefficient vectors $\zeta_n^{(1)}, \zeta_n^{(2)} \in \mathbb{R}^3$ such that

$$Q^{\frac{1}{2}}(u_n) = \left\{ \mathrm{Id}_3 + \left[\zeta_n^{(1)}\times\right] + \frac{1}{2}\left(\left[\zeta_n^{(1)}\times\right]^2 + \left[\zeta_n^{(2)}\times\right] \right) \right\} Q^{\frac{1}{2}}(\bar{u}_n) + \mathrm{o}\left(|\Delta u_n|^2\right). \quad \text{(A.1.12)}$$

Second, since the square of this expansion has to match, up to $\mathrm{o}\left(|\Delta u_n|^2\right)$ terms, the one in (A.1.7), one has to define

$$\zeta_n^{(1)} = \left(\mathrm{Id}_3 + Q^{\frac{1}{2}}(\bar{u}_n)\right)^{-1} \gamma_n^{(1)} \quad \text{(A.1.13a)}$$

$$\zeta_n^{(2)} = \left(\mathrm{Id}_3 + Q^{\frac{1}{2}}(\bar{u}_n)\right)^{-1} \left[\gamma_n^{(2)} - \zeta_n^{(1)} \times Q^{\frac{1}{2}}(\bar{u}_n)\zeta_n^{(1)}\right] \quad \text{(A.1.13b)}$$

where the triples $\gamma_n^{(1)}$ and $\gamma_n^{(2)}$ are defined in (A.1.11). Note that we have used the identity

$$[\gamma\times][\zeta\times] - [\zeta\times][\gamma\times] = [(\gamma\times\zeta)\times] \quad \text{(A.1.14)}$$

for any triple $\gamma, \zeta \in \mathbb{R}^3$. The relations (A.1.11) and (A.1.13) motivate the definition of the matrix

$$\mathbb{P}_2(\bar{u}_n) = \left(\mathrm{Id}_3 + Q^{\frac{1}{2}}(\bar{u}_n)\right)^{-1} \mathbb{P}_1(\bar{u}_n) \quad \text{(A.1.15)}$$

in order to write (A.1.13) as

$$\zeta_n^{(1)} = \mathbb{P}_2(\bar{u}_n)\Delta u_n \quad \text{(A.1.16a)}$$

$$\zeta_n^{(2)} = -\frac{1}{2}\mathbb{P}_2(\bar{u}_n)(\Delta u_n \otimes \Delta u_n)\bar{u}_n - \det\left(\mathrm{Id}_3 + Q^{\frac{1}{2}}(\bar{u}_n)\right)^{-1} \left[(Q^{\frac{T}{2}}(\bar{u}_n)\mathbb{P}_1(\bar{u}_n)\Delta u_n) \times (\mathbb{P}_1(\bar{u}_n)\Delta u_n)\right] \quad \text{(A.1.16b)}$$

similarly to (A.1.11). Note that we have used (A.1.6) to deduce (A.1.16b).

A second order expansion for the translation vector $q_n = Q^{\frac{1}{2}}(u_n)v_n$ is then

$$Q^{\frac{1}{2}}(u_n)v_n = \left\{ \mathrm{Id}_3 + \left[\zeta_n^{(1)}\times\right] + \frac{1}{2}\left(\left[\zeta_n^{(1)}\times\right]^2 + \left[\zeta_n^{(2)}\times\right] \right) \right\} Q^{\frac{1}{2}}(\bar{u}_n)\bar{v}_n + \left(\mathrm{Id}_3 + \left[\zeta_n^{(1)}\times\right]\right) Q^{\frac{1}{2}}(\bar{u}_n)\Delta v_n + \mathrm{o}\left(|\Delta u_n|^2|, \Delta v_n \cdot \Delta u_n\right) \quad \text{(A.1.17)}$$

where we have used $\Delta v_n = v_n - \bar{v}_n$. In order to match the terms in the expansion (A.1.2b), one has to define

$$q_{\bar{x}_n}^{(1)} = \left[\left(\zeta_n^{(1)} - \gamma_n^{(1)}\right)\times\right]\bar{q}_n + Q^{\frac{1}{2}}(\bar{u}_n)\Delta v_n \quad \text{(A.1.18a)}$$

$$q_{\bar{x}_n}^{(2)} = \left\{ \left[\zeta_n^{(1)}\times\right]^2 - \left[\gamma_n^{(1)}\times\right]^2 + \left[\left(\zeta_n^{(2)} - \gamma_n^{(2)}\right)\times\right] \right\} \bar{q}_n + 2\zeta_n^{(1)} \times \left(Q^{\frac{1}{2}}(\bar{u}_n)\Delta v_n\right) \quad \text{(A.1.18b)}$$

where $\bar{q}_n = Q^{\frac{1}{2}}(\bar{u}_n)\bar{v}_n$. The expressions (A.1.10) and (A.1.18) serve as explicit definitions of the blocks in the expansion (A.1.2), and then, equivalently, as definitions of the terms $a^{(1)}_{\bar{x}_n}(x_n - \bar{x}_n)$ and $a^{(2)}_{\bar{x}_n}(x_n - \bar{x}_n)$ in the expansion (1.3.4). The terms $\mathbb{L}^{(1)}_{\bar{x}_n}(x_n - \bar{x}_n)$ and $\mathbb{L}^{(2)}_{\bar{x}_n}(x_n - \bar{x}_n)$ are then obtained from the relations (1.3.6). Precisely, one finds

$$\mathbb{L}^{(1)}_{\bar{x}_n}(x_n - \bar{x}_n) = \begin{pmatrix} \gamma^{(1)}_n \\ q^{(1)}_{\bar{x}_n} \end{pmatrix} \tag{A.1.19a}$$

$$\mathbb{L}^{(2)}_{\bar{x}_n}(x_n - \bar{x}_n) = \begin{pmatrix} \gamma^{(2)}_n \\ q^{(2)}_{\bar{x}_n} - \gamma^{(1)}_n \times q^{(1)}_{\bar{x}_n} \end{pmatrix} \tag{A.1.19b}$$

From (A.1.19a) one defines the matrix operator $\mathbb{L}^{(1)}_{\hat{x}_n}$, which appears in (2.2.20), (A.5.3) and (A.5.4), by

$$\mathbb{L}^{(1)}_{\hat{x}_n} = \begin{pmatrix} \mathbb{P}_1(\hat{u}_n) & 0 \\ Q^{\frac{1}{2}}(\hat{u}_n)\,[\hat{v}_n\times]\,\mathbb{P}_2(\hat{u}_n) & Q^{\frac{1}{2}}(\hat{u}_n) \end{pmatrix} \tag{A.1.20}$$

according to the definitions (A.1.11a) and (A.1.18a). Its inverse is explicitly given by

$$\mathbb{L}^{(1)^{-1}}_{\hat{x}_n} = \begin{pmatrix} \mathbb{P}_1^{-1}(\hat{u}_n) & 0 \\ -\,[\hat{v}_n\times]\left(\mathrm{Id}_3 + Q^{\frac{1}{2}}(\hat{u}_n)\right)^{-1} & Q^{\frac{T}{2}}(\hat{u}_n) \end{pmatrix}. \tag{A.1.21}$$

A.2 On the Transformation Rule between Double Chain and Bichain Constitutive Relations

To translate the double chain constitutive relations (2.2.28), or equivalently (2.2.30), into the bichain ones (2.2.36) one uses the relations (2.1.29) together with the chain rule. One needs then to differentiate the matrix $P_n^{\pm\frac{1}{2}}$ which gives

$$\delta P_n^{\pm\frac{1}{2}} = \pm P_n^{\pm\frac{1}{2}}\left[M_n^{\pm}\phi_n\times\right] \tag{A.2.1}$$

for some triple $\phi_n \in \mathbb{R}^3$ and where $M_n^{\pm} = \left(\mathrm{Id}_3 + P_n^{\pm\frac{1}{2}}\right)^{-1}$ as in (2.2.39). This leads to the following identities

$$\mathsf{m}_n = P_n^{-}\mathsf{m}_n^{+} + P_n^{+}\mathsf{m}_n^{-} + \mathsf{w}_n \times \frac{1}{2}\left[P_n^{+}\mathsf{n}_n^{-} - P_n^{-}\mathsf{n}_n^{+}\right] \tag{A.2.2a}$$

$$\mathsf{n}_n = P_n^{-}\mathsf{n}_n^{+} + P_n^{+}\mathsf{n}_n^{-} \tag{A.2.2b}$$

and

$$\begin{aligned} \mathsf{C}_n^{\mathcal{P}} = {} & M_n^{-}\left[-\mathsf{m}_{n+1}^{-} - q_n^{-}\times\mathsf{n}_{n+1}^{-} + \mathsf{c}_n^{-}\right] - M_n^{+}(Q_{n-1}^{-})^T\mathsf{m}_n^{-} \\ & -M_n^{+}\left[-\mathsf{m}_{n+1}^{+} + \mathsf{m}_n^{+} - q_n^{+}\times\mathsf{n}_{n+1}^{+} + \mathsf{c}_n^{+}\right] + M_n^{-}(Q_{n-1}^{+})^T\mathsf{m}_n^{+} \end{aligned} \tag{A.2.2c}$$

$$\mathsf{F}_n^{\mathcal{P}} = \frac{1}{2}P_n^+\left(-\mathsf{n}_{n+1}^- + \mathsf{n}_n^- + \mathsf{f}_n^-\right) - \frac{1}{2}P_n^-\left(-\mathsf{n}_{n+1}^+ + \mathsf{n}_n^+ + \mathsf{f}_n^+\right) \tag{A.2.2d}$$

where we have used the short notation $P_n^+ = P_n^{\frac{1}{2}}$ and $P_n^- = P_n^{\frac{T}{2}}$ and the fact that $M_n^+ + M_n^- = \mathrm{Id}_3$.

A.3 On the Transformation Rule between Double Rod and Birod Constitutive Relations

To translate the double rod constitutive relations (3.2.32), or equivalently (3.2.33), into the birod ones (3.2.46) one uses the relations (3.1.22) together with the chain rule. One needs then to differentiate the matrix $P^{\pm\frac{1}{2}}(s)$ which gives

$$\delta P^{\pm\frac{1}{2}}(s) = \pm P^{\pm\frac{1}{2}}(s)\left[M^{\pm}(s)\phi(s)\times\right] \tag{A.3.1}$$

for some continuous function $\phi(s) \in \mathbb{R}^3$ and where $M^{\pm}(s) = \left(\mathrm{Id}_3 + P^{\pm\frac{1}{2}}(s)\right)^{-1}$ as in (3.1.25). The leads to the following identities

$$\begin{aligned}\mathsf{m}(s) &= P^-(s)\mathsf{m}^+(s) + P^+(s)\mathsf{m}^-(s)\\ &+\mathsf{w}(s)\times\frac{1}{2}\left[P^+(s)\mathsf{n}^-(s) - P^-(s)\mathsf{n}^+(s)\right]\end{aligned} \tag{A.3.2a}$$

$$\mathsf{n}(s) = P^-(s)\mathsf{n}^+(s) + P^+(s)\mathsf{n}^-(s) \tag{A.3.2b}$$

$$\mathsf{m}^{\mathcal{P}}(s) = M^-(s)\mathsf{m}^-(s) - M^+(s)\mathsf{m}^+(s) \tag{A.3.2c}$$

$$\mathsf{n}^{\mathcal{P}}(s) = \frac{1}{2}\left[P^+(s)\mathsf{n}^-(s) - P^-(s)\mathsf{n}^+(s)\right] \tag{A.3.2d}$$

and

$$\begin{aligned}\mathsf{c}^{\mathcal{P}} &= M^-(s)\left[\mathsf{m}^-(s)\times U^-(s) + \mathsf{n}^-(s)\times V^-(s) + \mathsf{c}^-(s)\right]\\ &-M^+(s)\left[\mathsf{m}^+(s)\times U^+(s) + \mathsf{n}^+(s)\times V^+(s) + \mathsf{c}^+(s)\right]\\ &+\frac{d}{ds}\left(M^+(s)\right)P^+(s)\mathsf{m}^-(s) - \frac{d}{ds}\left(M^-(s)\right)P^-(s)\mathsf{m}^+(s)\end{aligned} \tag{A.3.2e}$$

$$\mathsf{f}^{\mathcal{P}}(s) = \frac{1}{2}\left[P^+(s)\mathsf{f}^-(s) - P^-(s)\mathsf{f}^+(s)\right] - U(s)\times\frac{1}{2}\left[P^+(s)\mathsf{n}^-(s) - P^-(s)\mathsf{n}^+(s)\right] \tag{A.3.2f}$$

where we have used the short notation $P^+(s) = P^{\frac{1}{2}}(s)$ and $P^-(s) = P^{\frac{T}{2}}(s)$ and the fact that $M^+(s) + M^-(s) = \mathrm{Id}_3$.

A.4 Detail of the Formulation of the Quadratic Birod Energy

We discuss here in more details the quadratic birod energy presented in (5.3.3). We first discuss the property that any equivalent quadratic birod energy formulation can be written in terms of the absolute minimizer.

Suppose that the energy is given in the form

$$E[\boldsymbol{g},\mathcal{P}(y)] = \int_0^L \frac{1}{2}\begin{pmatrix} y(s)-a(s) \\ \frac{d}{ds}y(s)-b(s) \\ \xi(s)-c(s)\end{pmatrix}\cdot\mathcal{K}(s)\begin{pmatrix} y(s)-a(s) \\ \frac{d}{ds}y(s)-b(s) \\ \xi(s)-c(s)\end{pmatrix}ds \\ +\frac{1}{2}\left(y(0)-d_0)\right)\cdot\mathcal{K}_0\left(y(0)-d_0)\right)+\frac{1}{2}\left(y(L)-d_L\right)\cdot\mathcal{K}_L\left(y(L)-d_L\right). \tag{A.4.1}$$

For any birod variables $(\hat{\xi}(s),\hat{y}(s))$, the energy E in (A.4.1) is then identically equal to

$$E[\boldsymbol{g},\mathcal{P}(y)] = \int_0^L \frac{1}{2}\begin{pmatrix} y(s)-\hat{y}(s) \\ \frac{d}{ds}y(s)-\frac{d}{ds}\hat{y}(s) \\ \xi(s)-\hat{\xi}(s)\end{pmatrix}\cdot\mathcal{K}(s)\begin{pmatrix} y(s)-\hat{y}(s) \\ \frac{d}{ds}y(s)-\frac{d}{ds}\hat{y}(s) \\ \xi(s)-\hat{\xi}(s)\end{pmatrix}ds \\ +\frac{1}{2}\left(y(0)-\hat{y}(0)\right)\cdot\mathcal{K}_0\left(y(0)-\hat{y}(0)\right)+\frac{1}{2}\left(y(L)-\hat{y}(L)\right)\cdot\mathcal{K}_L\left(y(L)-\hat{y}(L)\right) \\ -\int_0^L\left(y(s)-\hat{y}(s)\right)\cdot\left[F_{\hat{y}}(s)-\frac{d}{ds}N_{\hat{y}}(s)\right]+\left(\xi(s)-\hat{\xi}(s)\right)\cdot M_{\hat{\xi}}(s)ds \\ -\frac{1}{2}\left(y(0)-\hat{y}(0)\right)\cdot\left[\mathcal{K}_0\left(d_0-\hat{y}(0)\right)-N_{\hat{y}}(0)\right]-\frac{1}{2}\left(y(L)-\hat{y}(L)\right)\cdot\left[\mathcal{K}_L\left(d_L-\hat{y}(L)\right)+N_{\hat{y}}(L)\right]+cste \tag{A.4.2}$$

after integration by parts and where we have defined

$$\begin{pmatrix} F_{\hat{y}}(s) \\ N_{\hat{y}}(s) \\ M_{\hat{\xi}}(s)\end{pmatrix} = \mathcal{K}(s)\begin{pmatrix} \hat{y}(s)-a(s) \\ \frac{d}{ds}\hat{y}(s)-b(s) \\ \hat{\xi}(s)-c(s)\end{pmatrix}. \tag{A.4.3}$$

Consequently, if the birod variables $(\hat{\xi}(s),\hat{y}(s))$ can be chosen such that

$$\begin{aligned} F_{\hat{y}}(s)-\frac{d}{ds}N_{\hat{y}}(s) &= 0 \\ M_{\hat{\xi}}(s) &= 0 \\ \mathcal{K}_0\left(d_0-\hat{y}(0)\right)-N_{\hat{y}}(0) &= 0 \\ \mathcal{K}_L\left(d_L-\hat{y}(L)\right)+N_{\hat{y}}(L) &= 0 \end{aligned} \tag{A.4.4}$$

according to the relations in (A.4.3) then the birod energy E becomes exactly of the wanted form. Such a birod configuration exists and is the global minimum of the energy in (A.4.1) since the equations in (A.4.3) are literally the birod stationary conditions in (3.2.59) for fully unstressed boundary conditions. Consequently the birod energy formulation in (5.3.3) and in (A.4.1) differ by a *null Lagrangian*.

A.5 Explicit Approximation of the Expected Junction and of the Local Fluctuation Matrix for Chains

The aim of this appendix is to present how an explicit approximation of the expected junction rigid body displacement $\bar{a}_n$ and to the local fluctuation matrix $\mathcal{C}_n$ can be deduced in the case of the normal configurational distribution in (5.2.1), or equivalently in (7.2.1), when it is concentrated enough.

Let $x = (x_1, ..., x_{N-1}) \in \mathbb{R}^{6(N-1)}$ be the internal coordinates of a rigid body chain as defined in (2.1.13) and denote by $a_n = a(x_n)$ the parametrization of a junction displacement in the chain, as defined in (2.1.11). We assume that there exists a unique rigid body motion $\bar{a}_n$ which satisfies the property of the expected junction (2.4.10) and we introduce $\bar{x}_n$ as its internal coordinates.

The relation between the variables Θ_n and x_n can be expanded around $\bar{x}_n$ as

$$\Theta_n(x_n) = \mathbb{L}^{(1)}_{\bar{x}_n}(x_n - \bar{x}_n) + \frac{1}{2}\mathbb{L}^{(2)}_{\bar{x}_n}(x_n - \bar{x}_n) + \mathrm{o}\left(|x_n - \bar{x}_n|^2\right) \tag{A.5.1}$$

where $\mathbb{L}^{(1)}_{\bar{x}_n}(x_n - \bar{x}_n)$ and $\mathbb{L}^{(2)}_{\bar{x}_n}(x_n - \bar{x}_n)$ stand respectively for the linear and quadratic contribution of $(x_n - \bar{x}_n)$ in this expansion. The fact that there is no constant term in this expansion comes from the definitions of $\bar{a}_n$ in (2.4.7) and of Θ_n in (2.4.8).

Taking both sides the expectation with respect to the normal Boltzmann distribution $d\rho_n(a(x_n))$ in (5.2.1), or in (7.2.1), leads to

$$\langle\Theta_n\rangle = \mathbb{L}^{(1)}_{\bar{x}_n}(\langle x_n\rangle - \bar{x}_n) + \frac{1}{2}\left\langle\mathbb{L}^{(2)}_{\bar{x}_n}(x_n - \langle x_n\rangle)\right\rangle + \frac{1}{2}\mathbb{L}^{(2)}_{\bar{x}_n}(\langle x_n\rangle - \bar{x}_n) + \left\langle\mathrm{o}\left(|x_n - \bar{x}_n|^2\right)\right\rangle. \tag{A.5.2}$$

Assuming that $\left|\mathbb{L}^{(1)}_{\hat{x}_n}\right|_{\mathrm{Sp}} = \mathrm{O}(1)$ and $\left|\mathbb{L}^{(2)}_{\hat{x}_n}\right|_{\mathrm{Sp}} = \mathrm{O}(1)$, and denoting the local covariance matrix $c^{xx}_{n,n}$, as defined in (2.3.19b), then the condition $\langle\Theta_n\rangle = 0$ in (2.4.10), which defines the expected junction $\bar{a}_n$, is then satisfied by

$$\bar{x}_n = \hat{x}_n + \frac{1}{2}\mathbb{L}^{(1)^{-1}}_{\hat{x}_n}\left\langle\mathbb{L}^{(2)}_{\hat{x}_n}(x_n - \hat{x}_n)\right\rangle \tag{A.5.3}$$

up to $\mathrm{o}\left(|c^{xx}_n|_{\mathrm{Sp}}\right)$ since $\left\langle\mathrm{o}\left(|x_n - \bar{x}_n|^2\right)\right\rangle = \mathrm{o}\left(|c^{xx}_n|_{\mathrm{Sp}}\right)$ and $\left\langle\mathbb{L}^{(2)}_{\hat{x}_n}(x_n - \hat{x}_n)\right\rangle = \mathrm{O}\left(|c^{xx}_n|_{\mathrm{Sp}}\right)$. Consequently, the covariance matrix $\mathcal{C}_n = \langle\Theta_n \otimes \Theta_n\rangle$ can also be approximated up to $\mathrm{o}\left(|c_n|_{\mathrm{Sp}}\right)$ as

$$\mathcal{C}_n = \mathbb{L}^{(1)}_{\hat{x}_n} c^{xx}_n \mathbb{L}^{(1)^T}_{\hat{x}_n} \tag{A.5.4}$$

which is symmetric and positive definite since the matrix c^{xx}_n is.

Moreover, for local internal chain coordinates x_n used in [Gonzalez et al., 2013, Petkeviciute et al., 2014], the matrix $\mathbb{L}^{(1)}_{\hat{x}_n}$ is explicitly given in (2.2.20), or in (A.1.19a), and the quadratic term $\mathbb{L}^{(2)}_{\hat{x}_n}(x_n - \hat{x}_n)$ in (A.1.19b). Observe that if the variable x_n is such that

its covariance c_n has the property that $|c_n|_{\mathrm{Sp}} << 1$ and the linear operator $\mathbb{L}^{(1)}_{\hat{x}_n}$ has the property that $\left|\mathbb{L}^{(1)}_{\hat{x}_n}\right|_{\mathrm{Sp}} = \mathrm{O}(1)$, then the hypothesis (2.4.14) is satisfied with $\epsilon_n = |c_n|_{\mathrm{Sp}}$ since

$$\mu_k(\Theta_n) = \mathrm{O}\left(|c_n|^{\frac{k}{2}}_{\mathrm{Sp}}\right) \text{ for } k > 2 \tag{A.5.5}$$

where $\mu_k(\Theta_n)$ denotes the k^{th} moment of the variable Θ_n.

Consequently, the approximation of the internal coordinates $\bar{x}_n = (\bar{u}_n, \bar{v}_n)$ for the expected junction in (A.5.3) can be shown to be

$$\bar{u}_n = \hat{u}_n - \frac{1}{4} c_n^{uu} \hat{u}_n \tag{A.5.6a}$$

$$\bar{v}_n = \left(\mathrm{Id}_3 - \mathrm{Tr}\left(Q^{\frac{T}{2}} \mathbb{P}_2 c_n^{uu} \mathbb{P}_2^T\right) \mathrm{Id}_3 - Q^{\frac{T}{2}} \mathbb{P}_2 c_n^{uu} \mathbb{P}_2^T\right) \hat{v}_n - \det\left(\mathrm{Id}_3 + Q^{\frac{1}{2}}\right)^{-1} \mathrm{Vect}\left(Q^{\frac{T}{2}} \mathbb{P}_1 c_n^{uu} \mathbb{P}_1^T Q\right) \times \hat{v}_n + \mathrm{Vect}\left(c_n^{uv} \mathbb{P}_1^T \left[\hat{u}_n^{\frac{1}{2}} \times\right] Q^{\frac{1}{2}}\right) \tag{A.5.6b}$$

where the matrices $Q, \mathbb{P}_1, \mathbb{P}_2$, defined in (2.1.16) and in (2.2.21), have to be evaluated at $\hat{u}_n$ and where $\hat{u}_n^{\frac{1}{2}}$ denotes the Cayley vector associated to the rotation matrix $Q^{\frac{1}{2}}(\hat{u}_n)$ as defined in (2.1.16).

A.6 Definition of the sequences λ and *CF*

In this appendix we give the precise composition of the sequences λ [Sanger et al., 1982] and *CF* [Kitchin et al., 1986] used in chapters 5, 6 and 7.

The sequence λ :

```
GGGCGGCGAC CTCGCGGGTT TTCGCTATTT ATGAAAATTT
TCCGGTTTAA GGCGTTTCCG TTCTTCTTCG TCATAACTTA
ATGTTTTTAT TTAAAATACC CTCTGAAAAG AAAGGAAACG
ACAGGTGCTG AAAGCGAGGC TTTTTGGCCT CTGTCGTTTC
CTTTCTCTGT TTTTGTCCGT GGAATGAACA ATGGAAGTCA
ACAAAAAGCA GCTGGCTGAC ATTTTCGGTG CGAGTATCCG
TACCATTCAG AACTGGCAGG AACAGGGAAT GCCCGTTCTG
CGAGGCGGTG GCAAGGGTAA TGAGGTGCTT TATGACTCTG
CCGCCGTCAT AAAATGGTAT GCCGAAAGGG ATGCTGAAAT
TGAGAACGAA AAGCTGCGCC GGGAGGTTGA AGAACTGCGG
CAGGCCAGCG AGGCAGATCT CCAGCCAGGA ACTATTGAGT
ACGAACGCCA TCGACTTACG CGTGCGCAGG CCGACGCACA
GGAACTGAAG AATGCCAGAG
```

The sequence *CF* :

```
 CAGGAAAGCG GTGAAAACAC CCCCACCAAA CCCAAGGCAG
GCCCAAAGTA CCAAACCAGC GCAAATCACC TCTGTCCAGC
ACAAACCCCG TCCAAACCAG CACTCAGACC CAGGAAAACC
CCTCCCGGAG GCCCCGAAAT CGGGCTAGAA CCCCGCCAAA
CCCCCTGCCA GGAGGCCTAA AATTCCAACC GAAAATCGCG
AGGTTACTTT TTTGGAGCCC GAAAACCACC CAAAATCAAG
GAAAAATGGC CAAAAAATGC CAAAAAATAG CGAAAATACC
CCGAAAATTG GCAAAAATTA ACAAAAAATA GCGAATTTCC
CTGAATTTTA GGCGAAAAAA CCCCCGAAAA TGGCCAAAAA
CGCACTGAAA ATCAAAATCT GAACGTCTAC GCCTCTGCTT
TAGAGTCTGT CTACCACCCG GGTCGGTTAA ATATGTCGGC
CGTATTAAAG CTCACATGAC AAAATCATGC GATGAGGTTA
GGAAAGGGAA AGACAATGTA C
```

A.7 Details on the averaging principle

We present here a brief summary, largely inspired from related sections in [Hartmann, 2007, Freidlin and Wentzell, 2012], of the averaging principle for problems with two distinct scales.

Let $0 < \epsilon << 1$ a scalar and consider the following ODE system for $s \in]0, L[$

$$\begin{aligned} \frac{d}{ds}x_\epsilon(s) &= \epsilon f(x_\epsilon(s), y_\epsilon(s)) \\ \frac{d}{ds}y_\epsilon(s) &= g(x_\epsilon(s), y_\epsilon)(s) \end{aligned} \tag{A.7.1}$$

where $f, g \sim \mathrm{O}(1)$. The scale separation is understood in the fact that the variable y_ϵ exhibits relatively strong variations with respect to the variable x_ϵ at the length scale $s \sim 1$. The variable x_ϵ is called the *slow variable* and y_ϵ is called the *fast variable*.

A *local averaging principle* can be stated as follows. Let $\Delta s > 0$ be a scalar value, which is not particularly small, and observe that for any fixed value of s_0 we have

$$\begin{aligned} x_\epsilon(s_0 + \Delta s) - x_\epsilon(s_0 - \Delta s) &= \epsilon \int_{s_0-\Delta s}^{s_0+\Delta s} f(x_\epsilon(s'), y_\epsilon(s'))ds' \\ &= \epsilon \int_{s_0-\Delta s}^{s_0+\Delta s} f(x_\epsilon(s_0), y_\epsilon(s))ds + \mathrm{O}\left(\epsilon^2 \Delta s^2\right) \\ &= \epsilon \int_{-\Delta s}^{\Delta s} f(x_\epsilon(s_0), y_\epsilon(s_0 + s'))ds' + \mathrm{O}\left(\epsilon^2 \Delta s^2\right) \end{aligned} \tag{A.7.2}$$

if f is Lipschitz in its first argument and where we have used (A.7.1). Introducing the scaled parameter $\sigma = \epsilon(s - s_0)$, and the rescaled variables

$$x_\epsilon^*(\sigma) = x_\epsilon\left(s_0 + \frac{\sigma}{\epsilon}\right) \text{ and } y_\epsilon^*(\sigma) = y_\epsilon\left(s_0 + \frac{\sigma}{\epsilon}\right) \tag{A.7.3}$$

leads then to the expression

$$\frac{d}{d\sigma}x_\epsilon^*(0) + \mathrm{O}\left(\Delta\sigma^2\right) = \frac{1}{2\Delta\sigma}\int_{-\Delta\sigma}^{\Delta\sigma} f(x_\epsilon(s_0), y_\epsilon^*(\sigma'))d\sigma' + \mathrm{O}\left(\Delta\sigma\right). \tag{A.7.4}$$

where $\mathrm{O}\left(\Delta\sigma\right) = \mathrm{O}\left(\epsilon\Delta s\right)$ by definition, and equivalently

$$\frac{d}{ds}x_\epsilon(s_0) + \mathrm{O}\left(\epsilon^3\Delta s^2\right) = \frac{\epsilon}{2\Delta s}\int_{s_0-\Delta s}^{s_0+\Delta s} f(x_\epsilon(s_0), y_\epsilon(s))ds + \mathrm{O}\left(\epsilon^2\Delta s\right). \tag{A.7.5a}$$

Similarly to (A.7.2) we also have

$$\frac{d}{ds}y_\epsilon(s_0) + \mathrm{O}\left(\Delta s^2\right) = \frac{1}{2\Delta s}\int_{s_0-\Delta s}^{s_0+\Delta s} g(x_\epsilon(s_0), y_\epsilon(s))ds + \mathrm{O}\left(\epsilon\Delta s\right) \tag{A.7.5b}$$

for any fixed value of s_0. These expressions motivate the definition, for a fixed value of $\Delta s = \mathrm{o}\left(\frac{1}{\epsilon}\right)$, of the *homogenised variable* $x^{(H)}$ as a solution of the *homogenised system*

$$\frac{d}{ds}x^{(H)}(s) = \epsilon f_{\Delta s}^{(H)}(x^{(H)}(s)) \tag{A.7.6a}$$

where

$$f_{\Delta s}^{(H)}(x^{(H)}(s_0)) = \frac{1}{2\Delta s}\int_{s_0-\Delta s}^{s_0+\Delta s} f(x^{(H)}(s_0), y_0(s; s_0))ds \tag{A.7.6b}$$

and where the variable $y_0(s; s_0)$ is defined as the solution of

$$\frac{d}{ds}y_0(s; s_0) = g(x^{(H)}(s_0), y(s; s_0)) \text{ with } y_0(s_0; s_0) = y(s_0) \tag{A.7.6c}$$

for any fixed value of s_0. The system (A.7.6) is said to constitute an averaging principle for the two scale ODE (A.7.1) and has the property that its solutions $x^{(H)}(s)$ and $y_0(s; s_0)$ provide, in some sense, approximations of the original solutions $x_\epsilon(s)$ and $y_\epsilon(s)$. The precise sense of the convergence, as well as necessary hypothesis on the function f and g, require a discussion which goes beyond this thesis, see [Freidlin and Wentzell, 2012, Abdulle et al., 2012] for instance. Intuitively, however, we expect the solution of the homogenised system be close to the original one if Δs is chosen such that $\epsilon\Delta s << 1$, since $s' = \mathrm{O}\left(1\right)$ one would have then for $|y_0(s_0 + s'; s_0) - y_\epsilon(s_0 + s')| = \mathrm{O}\left(\epsilon\Delta s\right)$, according to (A.7.5b), and $\left|x^{(H)}(s_0 + s') - x_\epsilon(s_0 + s')\right| = \mathrm{O}\left(\epsilon^2\Delta s\right)$ for $x^{(H)}(s_0) = x_\epsilon(s_0)$ and according to (A.7.5a). One of the crucial properties of the averaging principle is moreover that it delivers an approximation of the slow variable $x_\epsilon(s)$ at relatively large scale, as expressed in (A.7.4), or precisely

$$\left|x^{(H)}(s_0 + s') - x_\epsilon(s_0 + s')\right| = \mathrm{O}\left(\epsilon\Delta s\right) \text{ for } s' = \mathrm{O}\left(\frac{1}{\epsilon}\right) \tag{A.7.7}$$

and for $x^{(H)}(s_0) = x_\epsilon(s_0)$.

The averaging principle becomes particularly explicit when $g(x, y) = g(y)$ in (A.7.1) or if y is actually a known function in (A.7.5). In practice, it is difficult to guess a priori the optimal choice of the value of the window size Δs as a function of the value of ϵ. The smaller the value of Δs is, the closer are the solutions of (A.7.1) and (A.7.6a) respectively, but the less the original ODE system has been homogenised.

Index

Bibliography

[Abdulle et al., 2012] Abdulle, A., Weinan, E., Engquist, B., and Vanden-Eijnden, E. (2012). The heterogeneous multiscale method. *Acta Numerica*, 21:1–87.

[Antman, 1995] Antman, S. (1995). *Nonlinear Problems of Elasticity*. Applied Mathematical Sciences. Springer New York.

[Arroyo and Belytschko, 2004] Arroyo, M. and Belytschko, T. (2004). Finite crystal elasticity of carbon nanotubes based on the exponential Cauchy-Born rule. *Physical Review B*, 69(December 2003):1–11.

[Askar, 1986] Askar, A. (1986). *Lattice Dynamical Foundations of Continuum Theories: Elasticity, Piezoelectricity, Viscoelasticity, Plasticity*. Series in Theoretical and Applied Mechanics. World Scientific.

[Becker, 2007] Becker, N. (2007). *Sequence Dependent Elasticity of DNA*. PhD thesis, Technische Universität Dresden.

[Becker and Everaers, 2007] Becker, N. and Everaers, R. (2007). From rigid base pairs to semiflexible polymers: Coarse-graining DNA. *Physical Review E*, 76(2):021923.

[Becker and Everaers, 2009] Becker, N. B. and Everaers, R. (2009). DNA Nanomechanics in the Nucleosome. *Structure*, 17(4):579–589.

[Bednar et al., 1995] Bednar, J., Furrer, P., Katritch, V., Stasiak, A., Dubochet, J., and Stasiak, A. (1995). Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA. *Journal of molecular biology*, 254(4):579–94.

[Benham, 1977] Benham, C. J. (1977). Elastic model of supercoiling. *Proceedings of the National Academy of Sciences of the United States of America*, 74(6):2397–2401.

[Benham, 1979] Benham, C. J. (1979). An elastic model of the large-scale structure of duplex dna. *Biopolymers*, 18(3):609–623.

[Braides and Maria, 2006] Braides, A. and Maria, S. G. (2006). From discrete systems to continuous variational problems: an introduction. In *Topics on concentration phenomena and problems with multiple scales*, pages 3–77. Springer.

[Braides and Truskinovsky, 2008] Braides, A. and Truskinovsky, L. (2008). Asymptotic expansions by Γ-convergence. *Continuum Mechanics and Thermodynamics*, 20(1):21–62.

[Brenner, 1964] Brenner, H. (1964). Coupling between the Translational and Rotational Brownian Motions of Rigid Particles of Arbitrary Shape I: helicoidally isotropic particles. *Journal of Colloid Science*, 122(1955):104–122.

[Brenner, 1965] Brenner, H. (1965). Coupling between the Translational and Rotational Brownian Motions of Rigid Particles of Arbitrary Shape II. General Theory. *Journal of Colloid Science*, 20(2):104–122.

[Bresler and Frenkel, 1943] Bresler, S. E. and Frenkel, J. (1943). Character of the Thermal Motion of Long Organic Chains, with Reference to the elastic Properties of Rubber. *Rubber Chemistry and Technology*, 16(1):1–16.

[Bugl, 1969] Bugl, P. (1969). Dynamics of a Long Polymer Backbone. *The Journal of Chemical Physics*, 50(8):3137.

[Calladine et al., 2004] Calladine, C., Drew, H., Luisi, B., and Travers, A. (2004). *Understanding DNA: The Molecule and How it Works*. Elsevier Science.

[Chirikjian, 2011] Chirikjian, G. (2011). *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston.

[Chirikjian, 2010] Chirikjian, G. S. (2010). Group theory and biomolecular conformation: I. Mathematical and computational models. *Journal of physics. Condensed matter : an Institute of Physics journal*, 22(32):323103.

[Chouaïeb, 2003] Chouaïeb, N. (2003). *Kirchhoff's Problem of Helical Solutions of Uniform Rods and Their Stability Properties*. PhD thesis, EPFL.

[Coleman et al., 2003] Coleman, B. D., Olson, W. K., and Swigon, D. (2003). Theory of sequence-dependent DNA elasticity. *The Journal of Chemical Physics*, 118(15):7127.

[Cosserat and Cosserat, 1909] Cosserat, E. and Cosserat, F. (1909). *Thérie des Corps Défrormables*.

[Cotta-Ramusino, 2008] Cotta-Ramusino, L. (2008). *A path-integral formalism of DNA looping probability*. PhD thesis, EPFL.

[Cotta-Ramusino and Maddocks, 2010] Cotta-Ramusino, L. and Maddocks, J. H. (2010). Looping probabilities of elastic chains: A path integral approach. *Physical Review E*, 82(5):051924.

[Crothers et al., 1990] Crothers, D. M., Haran, T. E., and Nadeau, J. G. (1990). Intrinsically Bent DNA. *Journal of Biological Chemistry*, 265(13):7093–7096.

[Darboux, 1894] Darboux, G. (1894). *Lecons Sur La Theorie Generale Des Surfaces*.

[Dichmann et al., 1996] Dichmann, D. J., Li, Y., and Maddocks, J. (1996). Hamiltonian formulations and symmetries in rod mechanics. *Mathematical Approaches to Biomolecular Structure and Dynamics, IMA Volumes in Mathematics and its Applications*, 82:71 – 113.

[Doedel et al., 2009] Doedel, E. J., Champneys, A. R., Dercole, F., Fairgrieve, T. F., Kuznetsov, Y., Oldeman, B., Paffenroth, R. C., Sandstede, B., Wang, X., and Zhang, C. (2009). AUTO-07p: Continuation And Bifurcation Software For Ordinary Differential Equations.

[Doi and Edwards, 1986] Doi, M. and Edwards, S. F. (1986). *The Theory of Polymer Dynamics*. Oxford University Press.

[Dubochet et al., 1994] Dubochet, J., Bednar, J., Furrer, P., Stasiak, A. Z., Stasiak, A., and Bolshoy, A. A. (1994). Determination of the DNA helical repeat by cryo-electron microscopy. *Nature*.

[E and Ming, 2006] E, W. and Ming, P. (2006). Cauchy-Born Rule and the Stability of Crystalline Solids : Static Problems. *Archive for Rational Mechanics and Analysis*, 183:1–53.

[Ellis et al., 2010] Ellis, D. C., Gay-Balmaz, F., Holm, D. D., Putkaradze, V., and Ratiu, T. S. (2010). Symmetry reduced dynamics of charged molecular strands. *Archive for rational mechanics and analysis*, 197(3):811–902.

[Euler, 1765] Euler, L. (1765). *Theoria motus corporum solidorum seu rigidorum*.

[Eyring, 1932] Eyring, H. (1932). The resultant electric moment of complex molecules. *Physical Review*, 39(4):746–748.

[Fathizadeh et al., 2013] Fathizadeh, A., Berdy Besya, A., Reza Ejtehadi, M., and Schiessel, H. (2013). Rigid-body molecular dynamics of DNA inside a nucleosome. *The European physical journal. E, Soft matter*, 36(3):21.

[Flory, 1964] Flory, P. J. (1964). Mean-Square Moments of Chain Molecules. *Proceedings of the National Academy of Sciences of the United States of America*, 51(6):1060–1067.

[Flory, 1969] Flory, P. J. (1969). *Statistical Mechanics of Chain Molecules*. Hanser.

[Flory, 1973] Flory, P. J. (1973). Moments of the End-to-End Vector of a Chain Molecule, Its Persistence and Distribution. *Proceedings of the National Academy of Sciences of the United States of America*, 70(6):1819–1823.

[Flory and Jernigan, 1965] Flory, P. J. and Jernigan, R. L. (1965). Second and Fourth Moments of Chain Molecules. *The Journal of Chemical Physics*, 42(10):3509.

[Flory and Miller, 1966] Flory, P. J. and Miller, W. G. (1966). A General Treatment of Helix-Coil Equilibria in Macromolecular Systems. *Journal of molecular biology*.

[Freidlin and Wentzell, 2012] Freidlin, M. I. and Wentzell, A. D. (2012). *Random Perturbation of Dynamical Systems*, volume 338. Springer, springer edition.

[Friesecke et al., 2006] Friesecke, G., James, R. D., and Müller, S. (2006). A hierarchy of plate models derived from nonlinear elasticity by gamma-convergence. *Archive for rational mechanics and analysis*, 180(2):183–236.

[Friesecke and Theil, 2002] Friesecke, G. and Theil, F. (2002). Validity and failure of the Cauchy-Born hypothesis in a two-dimensional mass-spring lattice. *Journal of Nonlinear Science*, 12(5):445–478.

[Gonzalez and Maddocks, 2001] Gonzalez, O. and Maddocks, J. H. (2001). Extracting parameters for base-pair level models of DNA from molecular dynamics simulations. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 106(1-2):76–82.

[Gonzalez et al., 2013] Gonzalez, O., Petkeviciute, D., and Maddocks, J. H. (2013). A sequence-dependent rigid-base model of DNA. *The Journal of chemical physics*, 138(5):055102.

[Hagerman, 1990] Hagerman, P. J. (1990). Sequence-directed curvature of DNA. *Annual review of biochemistry*, 59(1):755–781.

[Hakobyan et al., 2012] Hakobyan, Y., Tadmor, E. B., and James, R. D. (2012). Objective quasicontinuum approach for rod problems. *Physical Review B*, 86(24):245435.

[Hartmann, 2007] Hartmann, C. (2007). *Model reduction in classical molecular dynamics*. PhD thesis, Freie Universität Berlin.

[Hubbard, 1972] Hubbard, P. S. (1972). Rotational Brownian Motion. *Physical Review A*, 6(6):2421–2433.

[Kabsch et al., 1982] Kabsch, W., Sander, C., and Trifonov, E. N. (1982). The ten helical twist angles of B-DNA. *Nucleic Acids Research*, 10(3):1097–1104.

[Kehrbaum, 1997] Kehrbaum, S. (1997). *Hamiltonian Formulations of the Equilibrium Conditions Governing Elastic Rods: Qualitative Analysis and Effective Properties*. PhD thesis, University of Maryland.

[Kendall, 1990] Kendall, W. S. (1990). Probability, Convexity, and Harmonic Maps With Small Image I: Uniqueness and Fine Existence. *Proceedings of the London Mathematical Society*, 61(August 1989):371–406.

[Kitchin et al., 1986] Kitchin, P. a., Klein, V. a., Ryan, K. a., Gann, K. L., Rauch, C. a., Kang, D. S., Wells, R. D., and Englund, P. T. (1986). A highly bent fragment of Crithidia fasciculata kinetoplast DNA. *Journal of Biological Chemistry*, 261(24):11302–11309.

[Koudelka et al., 2006] Koudelka, G. B., Mauro, S. A., and Ciubotaru, M. (2006). Indirect readout of dna sequence by proteins: The roles of dna sequence-dependent intrinsic and extrinsic forces. *Progress in nucleic acid research and molecular biology*, 81:143–177.

[Krakty and Porod, 1949] Krakty, O. and Porod, G. (1949). Röntgenuntersuchung gelöster fadenmoleküle. *Recueil des Travaux Chimiques des Pays-Bas*, 68(12):1106–1122.

[Kuhn, 1934] Kuhn, W. (1934). Über die gestalt fadenförmiger moleküle in lösungen. *Kolloid-Zeitschrift*, 68(1):2–15.

[Kumar et al., 2015] Kumar, A., Kumar, S., and Gupta, P. (2015). A Helical Cauchy-Born Rule for Special Cosserat Rod Modeling of Nano and Continuum Rods. *Journal of Elasticity*.

[Landau and Lifshitz, 1959] Landau, L. and Lifshitz, E. M. (1959). *Statistical Physics (Part. 1)*. Pergamon Press.

[Lankaš et al., 2009] Lankaš, F., Gonzalez, O., Heffler, L. M., Stoll, G., Moakher, M., and Maddocks, J. H. (2009). On the parameterization of rigid base and basepair models of DNA from molecular dynamics simulations. *Physical chemistry chemical physics : PCCP*, 11(45):10541–2.

[Lankaš et al., 2000] Lankaš, F., Šponer, J., Hobza, P., and Langowski, J. (2000). Sequence-dependent elastic properties of DNA. *Journal of Molecular Biology*, 299(3):695–709.

[Lankaš et al., 2003] Lankaš, F., Sponer, J., Langowski, J., and Cheatham, T. E. (2003). DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys J*, 85(5):2872–2883.

[Lavery and Hartmann, 1994] Lavery, R. and Hartmann, B. (1994). Modelling DNA conformational mechanics. *Biophys Chem*, 50(1-2):33–45.

[Lavery et al., 2014] Lavery, R., Maddocks, J. H., Pasi, M., and Zakrzewska, K. (2014). Analyzing ion distributions around DNA. *Nucleic acids research*, 42(12):8138–49.

[Lavery et al., 2009] Lavery, R., Zakrzewska, K., Beveridge, D., Bishop, T. C., Case, D. a., Cheatham, T., Dixit, S., Jayaram, B., Lankaš, F., Laughton, C., Maddocks, J. H., Michon, A., Osman, R., Orozco, M., Perez, A., Singh, T., Spackova, N., and Sponer, J. (2009). A systematic molecular dynamics study of nearest-neighbor effects

on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Research*, 38(1):299–313.

[Levene et al., 1986] Levene, S. D., Wu, H. M., and Crothers, D. M. (1986). Bending and flexibility of kinetoplast DNA. *Biochemistry*, 25(14):3988–3995.

[Liu et al., 2011] Liu, Y., Pérez, T., Li, W., Gunton, J. D., and Green, A. (2011). Statistical mechanics of helical wormlike chain model. *The Journal of chemical physics*, 134(6):065107.

[Long et al., 2012] Long, A., Wolfe, K., Mashner, M. J., and Chirikjian, G. S. (2012). The Banana Distribution is Gaussian: A Localization Study with Exponential Coordinates. In *Robotics: science and systems*.

[Love, 1927] Love, A. (1927). *A Treatise on the Mathematical Theory of Elasticity*. Cambridge University Press.

[Manning et al., 1996] Manning, R. S., Maddocks, J. H., and Kahn, J. D. (1996). A continuum rod model of sequence-dependent DNA structure. *The Journal of Chemical Physics*, 105(13):5626.

[Manning et al., 1998] Manning, R. S., Rogers, K. A., and Maddocks, J. H. (1998). Isoperimetric conjugate points with application to the stability of DNA minicircles. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 454(1980):3047–3074.

[Marini et al., 1982] Marini, J. C., Levene, S. D., Crothers, D. M., and Englund, P. T. (1982). Bent helical structure in kinetoplast DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 79(24):7664–7668.

[Marko, 1997] Marko, J. F. (1997). Stretching must twist DNA. *Europhysics Letters (EPL)*, 38(3):183–188.

[Marko and Siggia, 1994] Marko, J. F. and Siggia, E. D. (1994). Bending and twisting elasticity of DNA. *Macromolecules*, 27(4):981–988.

[Marky and Olson, 1994] Marky, N. L. and Olson, W. K. (1994). Configurational statistics of the DNA duplex: extended generator matrices to treat the rotations and translations of adjacent residues. *Biopolymers*, 34(1):109–120.

[Maroun and Olson, 1988] Maroun, R. C. and Olson, W. K. (1988). Base sequence effects in double-helical DNA. III. Average properties of curved DNA. *Biopolymers*, 27(4):585–603.

[Marsden and Ratiu, 2002] Marsden, J. and Ratiu, T. (2002). *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*. Texts in Applied Mathematics. Springer New York.

[Matthews, 1988] Matthews, B. W. (1988). Protein-DNA interaction. No code for recognition. *Nature*, 335:294–295.

[Meyer and Everaers, 2015] Meyer, S. and Everaers, R. (2015). Inferring coarse-grain histone-DNA interaction potentials from high-resolution structures of the nucleosome. *Journal of physics. Condensed matter : an Institute of Physics journal*, 27(6):064101.

[Moakher and Maddocks, 2005] Moakher, M. and Maddocks, J. H. (2005). A double-strand elastic rod theory. *Archive for Rational Mechanics and Analysis*, 177(1):53–91.

[Olson, 1996] Olson, W. K. (1996). Simulating DNA at low resolution. *Curr Opin Struct Biol*, 6(2):242–256.

[Olson et al., 2001] Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X. J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C. S., Westhof, E., Wolberger, C., and Berman, H. M. (2001). A standard reference frame for the description of nucleic acid base-pair geometry. *Journal of molecular biology*, 313(1):229–37.

[Olson et al., 1998] Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M., and Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 95(19):11163–11168.

[Olson et al., 1993] Olson, W. K., Marky, N. L., Jernigan, R. L., and Zhurkin, V. B. (1993). Influence of fluctuations on dna curvature: a comparison of flexible and static wedge models of intrinsically bent dna. *Journal of molecular biology*, 232(2):530–554.

[Öttinger, 1996] Öttinger, H. (1996). *Stochastic Processes in Polymeric Fluids: Tools and Examples for Developing Simulation Algorithms*. Springer.

[Padeken et al., 2015] Padeken, J., Zeller, P., and Gasser, S. M. (2015). Repeat DNA in genome organization and stability. *Current Opinion in Genetics & Development*, 31:12–19.

[Panyukov and Rabin, 2000] Panyukov, S. and Rabin, Y. (2000). Fluctuating filaments: statistical mechanics of helices. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, 62(5 Pt B):7135–46.

[Pasi et al., 2014] Pasi, M., Maddocks, J. H., Beveridge, D., Bishop, T. C., Case, D. A., Cheatham, T., Dans, P. D., Jayaram, B., Lankaš, F., Laughton, C., Mitchell, J., Osman, R., Orozco, M., Perez, A., Petkeviciute, D., Spackova, N., Sponer, J., Zakrzewska, K., and Lavery, R. (2014). μ−ABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Research*, 42(19):12272–12283.

[Petkeviciute, 2012] Petkeviciute, D. (2012). *A DNA Coarse-Grain Rigid Base Model and Parameter Estimation from Molecular Dynamics Simulations*. PhD thesis, EPFL.

[Petkeviciute et al., 2014] Petkeviciute, D., Pasi, M., Gonzalez, O., and Maddocks, J. H. (2014). cgDNA: a software package for the prediction of sequence-dependent coarse-grain free energies of b-form dna. *Nucleic Acids Research*, 42(20):e153.

[Phillips et al., 2009] Phillips, R., Kondev, J., and Theriot, J. (2009). *Physical Biology of the Cell*. Garland Science.

[Poincaré, 1901] Poincaré, H. (1901). Sur une forme nouvelle des équations de la mécanique . *Comptes rendus de l'Académie des Sciences*, 132:369–371.

[Porod, 1948] Porod, G. (1948). Zusammenhang zwischen mittlerem Endpunktsabstand und Kettenlänge bei Fadenmolekülen. *Acta Phys Austriaca*, 2:255–292.

[Rey, 2002] Rey, S. (2002). *Symmetry breaking, averaging and elastic rods with high intrinsic twist*. PhD thesis, EPFL.

[Rivetti et al., 1998] Rivetti, C., Walker, C., and Bustamante, C. (1998). Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *Journal of molecular biology*, 280(1):41–59.

[R.S. Manning, 1996] R.S. Manning, J.H. Maddocks, J. K. (1996). A continuum rod model of sequence-dependent dna structure. *J. Chem. Phys.*, 105:5626 – 5646.

[Sander, 2010] Sander, O. (2010). Geodesic finite elements for cosserat rods. *International Journal for Numerical Methods in Engineering*, 82(13):1645–1670.

[Sanger et al., 1982] Sanger, F., Coulson, A. R., Hong, G., Hill, D., and Petersen, G. d. (1982). Nucleotide sequence of bacteriophage λ DNA. *Journal of molecular biology*, 162(4):729–773.

[Schellman, 1974] Schellman, J. A. (1974). Flexibility of DNA. *Biopolymers*, 13(1):217–226.

[Schellman, 1980] Schellman, J. A. (1980). The flexibility of DNA, I. Thermal Fluctuation. *Biophysical Chemistry*, 11:321–328.

[Schellman and Harvey, 1995] Schellman, J. a. and Harvey, S. C. (1995). Static contributions to the persistence length of DNA and dynamic contributions to DNA curvature. *Biophysical chemistry*, 55(1-2):95–114.

[Schiehlen et al., 2006] Schiehlen, W., Guse, N., and Seifried, R. (2006). Multibody dynamics in computational mechanics and engineering applications. *Computer Methods in Applied Mechanics and Engineering*, 195(41–43):5509–5522.

[Segal and Widom, 2009] Segal, E. and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. *Current Opinion in Structural Biology*, 19(1):65–71.

[Shi et al., 1996] Shi, Y., He, S., and Hearst, J. E. (1996). Statistical mechanics of the extensible and shearable elastic rod and of DNA. *Journal of Chemical Physics*, 105(July):714–731.

[Steele, 1963a] Steele, W. a. (1963a). Molecular Reorientation in Liquids. I. Distribution Functions and Friction Constants. *The Journal of Chemical Physics*, 38(10):2404.

[Steele, 1963b] Steele, W. a. (1963b). Molecular Reorientation in Liquids. II. Angular Autocorrelation Functions. *The Journal of Chemical Physics*, 38(10):2411.

[Tadmor and Miller, 2011] Tadmor, E. and Miller, R. (2011). *Modeling Materials: Continuum, Atomistic and Multiscale Techniques*. Cambridge University Press.

[Thåström et al., 1999] Thåström, a., Lowary, P. T., Widlund, H. R., Cao, H., Kubista, M., and Widom, J. (1999). Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *Journal of molecular biology*, 288(2):213–229.

[Théveny et al., 1988] Théveny, B., Coulaud, D., Le Bret, M., and Révet, B. (1988). Local variations of curvature and flexibility along dna molecules analyzed from electron micrographs. In W. K. Olson, M.H. Sarma, R. H. S. and Sundaralingams, M., editors, *Structure and Expression (Vol. 3): DNA Bending & Curvature*, New York. Adenine Press.

[Triantafyllidis and Bardenhagen, 1993] Triantafyllidis, N. and Bardenhagen, S. (1993). On higher order gradient continuum theories in 1-d nonlinear elasticity. derivation from and comparison to the corresponding discrete models. *Journal of Elasticity*, 33(3):259–293.

[Trifonov and Sussman, 1980] Trifonov, E. N. and Sussman, J. L. (1980). The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proceedings of the National Academy of Sciences of the United States of America*, 77(7):3816–3820.

[Trifonov et al., 1988] Trifonov, E. N., Tan, R. K., and Harvey, S. C. (1988). Static persistence length of dna. In W. K. Olson, M.H. Sarma, R. H. S. and Sundaralingam, M., editors, *Structure and Expression (Vol. 3): DNA Bending & Curvature*, New York. Adenine Press.

[Vaillant, 2001] Vaillant, C. (2001). *Influence de la séquence sur les propriétés élastiques des chaînes ADN*. PhD thesis, Paris VI.

[Virstedt et al., 2004] Virstedt, J., Berge, T., Henderson, R. M., Waring, M. J., and Travers, A. A. (2004). The influence of DNA stiffness upon nucleosome formation. *Journal of Structural Biology*, 148(1):66–85.

[Vologodskaia and Vologodskii, 2002] Vologodskaia, M. and Vologodskii, A. (2002). Contribution of the intrinsic curvature to measured DNA persistence length. *Journal of molecular biology*, 317(2):205–13.

[Walter et al., 2010] Walter, J., Gonzalez, O., and Maddocks, J. H. (2010). On the stochastic modeling of rigid body systems with application to polymer dynamics. *Multiscale Modeling & Simulation*, 8(3):1018–1053.

[Yamakawa, 1976] Yamakawa, H. (1976). Statistical mechanics of helical wormlike chains. I. Differential equations and moments. *The Journal of Chemical Physics*, 64(12):5222.

[Yamakawa, 1997] Yamakawa, H. (1997). *Helical Wormlike Chains in Polymer Solutions*. Springer.

[Yang and E, 2006] Yang, J. and E, W. (2006). Generalized Cauchy-Born rules for elastic deformation of sheets, plates, and rods: Derivation of continuum models from atomistic models. *Physical Review B*, 74(18):184110.

[Zhang and Crothers, 2003] Zhang, Y. and Crothers, D. M. (2003). Statistical Mechanics of Sequence-Dependent Circular DNA and Its Application For DNA Cyclization. *Biophysical Journal*, 84(1):136–153.

[Zhou and Chirikjian, 2003] Zhou, Y. and Chirikjian, G. S. (2003). Conformational statistics of bent semiflexible polymers. *The Journal of Chemical Physics*, 119(9):4962.

Curriculum Vitae

Born in Lausanne on the 24^{th} of July 1987, I have grown up in the old town, around the cathedral. At the end of 2010, I have obtained my Master's degree in applied mathematics at the Ecole Polytechnique Fédérale de Lausanne (EPFL). My main interests were differential geometry and continuum mechanics. I have spent the following year working at *eauservice* for a civil service period where I studied the global tap water network metrology and especially the ageing of house flowmeters. I then started my Ph.D. on sequence dependent DNA modelling in the group of the Prof. J. H. Maddocks at the EPFL end of 2011. During these last four years I had the chance to go working for two months in the group of Dr. A. Stasiak at the Lausanne university on chromatin models, again for civil service period. Three months ago, I went back to eauservice to finish my civil service and to model the future tap water need in Lausanne and its close neighbourhood.