Eurographics Symposium on Geometry Processing 2013
Yaron Lipman and Richard Hao Zhang
(Guest Editors)

Volume 32 (2013), Number 5

Sparse Iterative Closest Point

Sofien Bouaziz Andrea Tagliasacchi Mark Pauly

École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

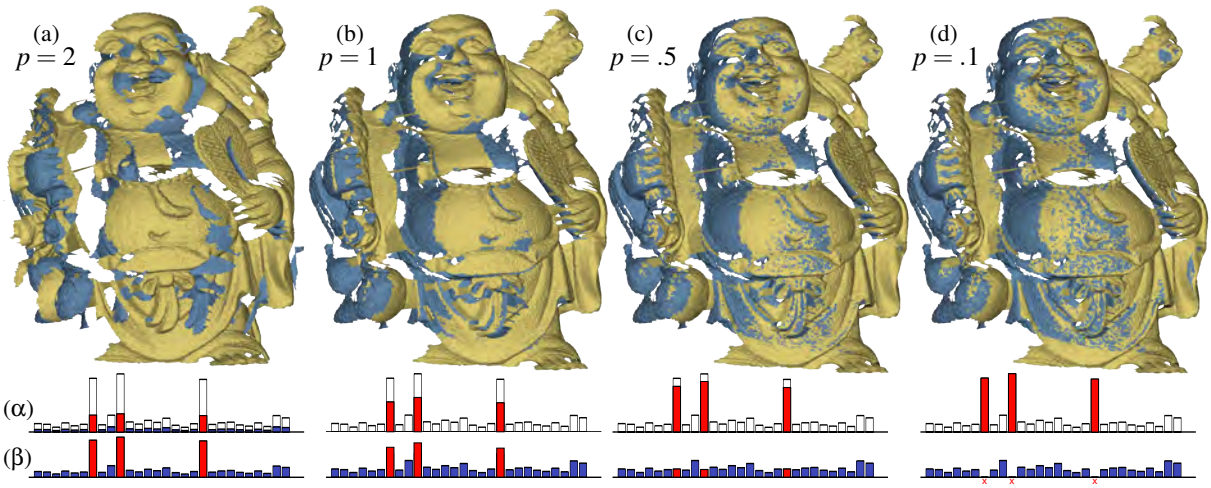


Figure 1: *Solving for optimal rigid alignment for incomplete geometry. (a) Traditional least-squares ICP does not distinguish between inliers and outliers, resulting in poor alignment. (b) The ℓ_1-ICP is more robust, but still cannot cope with the large amount of correspondence outliers. We show that ℓ_p-ICP, with $p \in [0,1]$ robustly handles large amounts of noise and outliers. Bottom rows: Illustration of sparsity-inducing norms. (α) Regularization of the input vector (black-framed) with an ℓ_p-norm leads to increased sparsity as we decrease the value of p. (β) The vector of complements provides an indication of how much its corresponding entry contributes in the optimization. Large outliers (red) contribute progressively less as we decrease p.*

Abstract

Rigid registration of two geometric data sets is essential in many applications, including robot navigation, surface reconstruction, and shape matching. Most commonly, variants of the Iterative Closest Point (ICP) algorithm are employed for this task. These methods alternate between closest point computations to establish correspondences between two data sets, and solving for the optimal transformation that brings these correspondences into alignment. A major difficulty for this approach is the sensitivity to outliers and missing data often observed in 3D scans. Most practical implementations of the ICP algorithm address this issue with a number of heuristics to prune or reweight correspondences. However, these heuristics can be unreliable and difficult to tune, which often requires substantial manual assistance. We propose a new formulation of the ICP algorithm that avoids these difficulties by formulating the registration optimization using sparsity inducing norms. Our new algorithm retains the simple structure of the ICP algorithm, while achieving superior registration results when dealing with outliers and incomplete data. The complete source code of our implementation is provided at `http://lgg.epfl.ch/sparseicp`.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Geometric algorithms, languages, and systems

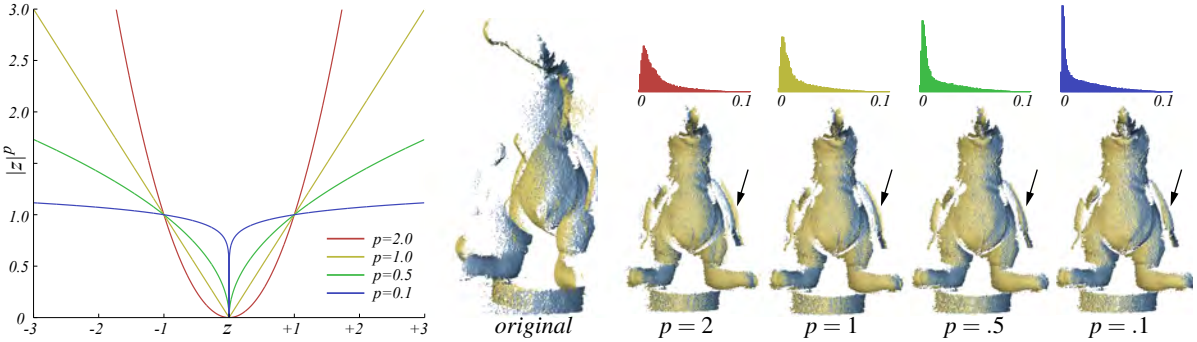


Figure 2: *(left) A plot of the penalty functions used to induce sparsity in our optimization; for small values of p, large outliers do not incur a large penalty in the optimization; this allows the optimization to effectively discard correspondence outliers when computing the optimal rigid transformation. (right) The alignment of the "coatie" dataset from [AMCO08] with several values of p matching the ones in the plot. By decreasing the value of p, the quality of the registration improves. The distribution of alignment residuals in the histograms highlights the sparse characteristics of our optimization.*

1. Introduction

The registration of digital geometry is a fundamental task in computer graphics and geometry processing. In this paper we focus on *pairwise registration*, where we aim to compute the optimal alignment of a *source* onto a *target* model. We additionally assume that the aligning transformation is *rigid*, that is, decomposable into a rotation and a translation.

Registration applications. Pairwise rigid alignment is widely employed as a sub-routine in a number of applications. In the acquisition of digital models, self-occlusion and limited sensor range create the need to scan the object from multiple directions; these scans have to be aligned into a common frame of reference to enable further processing [Pul99, GMGP05]. In digital quality inspection, a scan of a physical model needs to be registered with its ground truth CAD model, so to be able to measure and classify possible manufacturing errors [LG04]. In mobile robotics, matching a scan of range data to a digital representation of the scene is used to refine coarse satellite based localization [SHT09]. Even in non-rigid registration, a rigid alignment step is often used to bring a template of the deforming geometry into coarse alignment with the input data [PMG*05, LSP08, WBLP11]. Recently, consumer level 3D scanning devices (e.g. Microsoft Kinect) have lead to a growing interest in robust rigid alignment algorithms. The low cost of these acquisition devices comes at the expense of severely degraded data, which necessitates registration algorithms that can deal with large amounts of noise and outliers; see Figure 3.

Iterative closest point. Given two sets of points related by a correspondence relationship, there exist several ways of computing the optimal rigid transformation that aligns them with each other [ELF97]. The practical rigid registration problem consequently simplifies to finding a suitable set of corresponding points on source and target. The *Iterative Closest Point (ICP)* algorithm [BM92] addresses this problem by assuming the input data to be in *coarse alignment*. Under this assumption, a set of correspondences can be obtained by querying closest points on the target geometry. With a convergence guarantee, ICP computes a locally optimal registration by *alternately* solving for closest correspondences and optimal rigid alignment.

Correspondence outliers. As ICP is effectively performing local optimization, the quality of the solution is related to the quality of correspondences provided as input to the rigid transformation sub-routine. However, incorrect closest-point correspondences are particularly common in the registration of acquired geometry. Closest point queries are not only corrupted by measurement noise, but also by partial overlap of the source and target – many samples on the source simply do not have an ideal corresponding point on the target shape. To address this problem, various techniques rely on a set of heuristics to either *prune* or *downweigh* low quality correspondences. Typical criteria include discarding correspondences that are too far from each other, have dissimilar normals, or involve points on the boundary of the geometry; see [RL01] for details. These heuristics are often difficult to tune. For example, classifying boundary points in point cloud data is ill-posed and typically obtained by heuristic methods [BL04]. Furthermore, fixing a single parameter for distance thresholds is typically not sufficient, often leading to sub-optimal alignment; see Figure 4 for an example.

Sparsity inducing norms. In this paper, we propose a solution that implicitly models outliers using sparsity. Our work is based on recent advances in sparsity-inducing penalties [BJMO12, MS12] that have been successfully applied in compressive sensing [CW08]. We formulate the local alignment problem as recovering a rigid transformation that maximizes the number of zero distances between correspondences. This can be achieved by minimizing the ℓ_0 norm of

the vector of error residuals. In our work, we show how the ICP algorithm can be reformulated using ℓ_p norms, where $p \in [0,1]$, instead of the classical squared ℓ_2 norm (see Figure 2). Having the possibility to choose p between zero and one allows a tradeoff between efficiency and robustness.

Contributions. We propose a new technique for rigid alignment that robustly deals with significant amounts of noise and outliers, while not requiring any heuristic for correspondence pruning. We propose a sparse ℓ_p ($p \leq 1$) optimization problem that automatically learns the separation between data and outliers (see Figure 2 and Figure 6). Our optimization is the result of a careful design allowing to solve the alignment problem by iterating over a set of simple and tractable subproblems that can be solved efficiently. Moreover, our optimization builds on top of classical ICP components which permits us to reuse optimized closest point search [FBF77] and least-squares optimization for rigid transformation [ELF97]. We evaluate our approach by quantitative comparison to classical rigid registration methods [BM92] as well as recent robust variants. Our experiments demonstrate improvements in the quality of the registration on challenging datasets affected by noise, large number of outliers and incomplete data.

2. Background and related works

Over the last two decades, the registration of digital geometry has been an intensively studied problem. The recent survey by Tam *et al.* [TCL*12] provides a classification of several techniques for both rigid and non-rigid registration. In this section, we provide a simpler classification of rigid registration methods, where we distinguish solutions according to whether they seek a *global* optimum, or whether they attempt to perform a *local* refinement of an initial, possibly coarse, alignment. Roughly speaking, these optimizations attempt to minimize an alignment energy that measures the proximity of source and target models according to a given metric; see Section 3 for a detailed discussion.

Global optimization. Global approaches to registration exploit the small number of degrees of freedom of rigid transformations (6DoF). These transformations can be fully specified by two sets of matching points having cardinality of at least three. Global methods transform the continuous alignment problem into a discrete one, where the objective is to seek one of these pairs; see [vKZHCO11, Section 5.1]. A notable contribution in this domain is the paper by Aiger *et al.* [AMCO08] that exploits geometric invariants of rigid transformations. They proposed an $O(n^2)$ RANSAC algorithm capable of computing registration of geometry containing up to 40% outliers and having just 40% overlap. It is critical to note that global methods typically require *local refinement* in post processing in order to achieve high quality alignment. This creates a need for algorithms capable of performing robust local alignment – the problem we address in this paper.

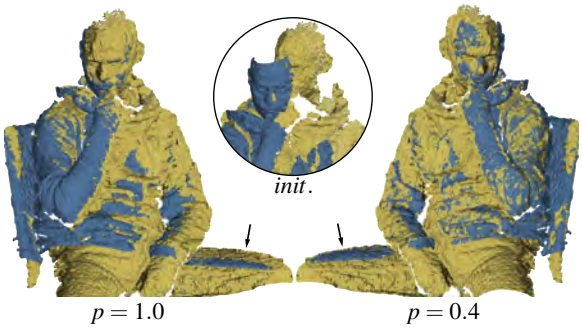


Figure 3: *"The Thinker": Alignment of two scans obtained by consumer-level depth cameras. Partial overlap and structured noise degrade the performance of ℓ_1-ICP. Robust ℓ_p-ICP achieves a more accurate registration, as can be observed in the leg and chair regions.*

Local optimization. Local ICP approaches refine the alignment assuming an initial coarse registration of source and target models is provided. Although commonly used in static registration, these algorithms are especially useful in *real-time* registration; see [RL01]. Indeed, assuming a sufficiently dense temporal sampling, the alignment achieved in the previous time-step can be used as initialization for the registration refinement. Registration based on local optimization effectively performs a descent optimization of the alignment energy [PHYH06]. Techniques in this class can be distinguished according to the way in which they approximate the alignment metric. When the distance to the closest point on the target is used, we obtain the *point-to-point* method of the original ICP algorithm by Besl and McKay [BM92]. Unfortunately, closest point distances only provide a good approximation of the distance function of the target geometry in *far-field* conditions [PH03]; a first-order Taylor expansion of the distance field can be used to improve this approximation in the *near-field*, resulting in the well known *point-to-plane* ICP variant [CM91]. Second-order approximations of the squared distance function [PH03] are also possible, resulting in schemes able to achieve quadratic convergence [PHYH06]. Note that although iteratively fetching closest points is the most common way of approaching local registration, there also exists the possibility of caching an approximation of the distance function at a desired precision [CLSB92, PLH04, MGPG04], or even directly aligning two distance functions [TK04, JV11].

Outliers and partial overlap. Various methods and heuristics have been proposed to improve the robustness of ICP at its different stages to cope with noisy and incomplete data [RL01]. Another prominent approach uses *robust functions* [Zha94, MY95, TFR99, Fit03] to reduce the importance of outliers, instead of explicitly pruning them. A notable side effect of this approach is an increase in the sparsity of the alignment residuals vector. This observation motivates the recent trend to model inliers/outliers by *explic-*

itly enforcing sparsity [FH10, HMS12] using ℓ_1 regularization [BJMO12]. Unlike the above-mentioned approaches, our regularizer is based on the ℓ_p norm with $p \in [0,1]$. It has been shown recently that ℓ_p norms with $p < 1$ outperform the ℓ_1 norm in inducing sparsity [Cha07], which makes our optimization more resilient to a large number of outliers. However, when $p < 1$, the resulting optimization problem is non-smooth and non-convex. It is therefore necessary to carefully design the optimization problem in order to obtain a robust and efficient solution [MS12, CW13]. We present a reformulation of ICP that takes advantage of p-norms, while still leading to a simple implementation that uses the core components of the original algorithm: optimized ℓ_2 closest point search [FBF77] and least-squares optimization for rigid transformation [ELF97].

3. Pairwise rigid registration

Given two surfaces $\mathcal{X}, \mathcal{Y}$ embedded in a k-dimensional space, we formulate the pairwise registration problem as

$$\underset{\mathbf{R},\mathbf{t}}{\arg\min} \quad \int_{\mathcal{X}} \varphi(\mathbf{R}\mathbf{x}+\mathbf{t},\mathcal{Y})d\mathbf{x} + I_{SO(k)}(\mathbf{R}), \tag{1}$$

where $\mathbf{R} \in \mathbb{R}^{k\times k}$ is a rotation matrix, $\mathbf{t} \in \mathbb{R}^k$ is a translation vector, and $\mathbf{x} \in \mathbb{R}^k$ is a point on the source geometry. The rigidity of the transformation is enforced by constraining $\mathbf{R}$ to the special orthogonal group $SO(k)$ using the indicator function $I_A(b)$ that evaluates to 0 if $b \in A$ and to $+\infty$ otherwise. The quality of a registration is evaluated by the metric φ that measures the distance to $\mathcal{Y}$ and is defined as

$$\varphi(\mathbf{x},\mathcal{Y}) = \min_{\mathbf{y}\in\mathcal{Y}} \varphi(\mathbf{x},\mathbf{y}) = \min_{\mathbf{y}\in\mathbb{R}^k} \varphi(\mathbf{x},\mathbf{y}) + I_{\mathcal{Y}}(\mathbf{y}). \tag{2}$$

Discrete pairwise registration. As we are solving the registration problem numerically, we sample the continuous surface $\mathcal{X}$ by a set of points $X = \{\mathbf{x}_i \in \mathcal{X}, i = 1\dots n\}$ to obtain

$$\underset{\mathbf{R},\mathbf{t}}{\arg\min} \quad \sum_{i=1}^{n} \varphi(\mathbf{R}\mathbf{x}_i+\mathbf{t},\mathcal{Y}) + I_{SO(k)}(\mathbf{R}). \tag{3}$$

Using Equation 2 and defining $Y = \{\mathbf{y}_i \in \mathbb{R}^k, i = 1\dots n\}$ we can then rewrite this energy as

$$\underset{\mathbf{R},\mathbf{t},Y}{\arg\min} \quad \sum_{i=1}^{n} \varphi(\mathbf{R}\mathbf{x}_i+\mathbf{t},\mathbf{y}_i) + I_{\mathcal{Y}}(\mathbf{y}_i) + I_{SO(k)}(\mathbf{R}). \tag{4}$$

Generalized ICP. In order to solve the non-linear problem of Equation 4, we can decouple the optimization by alternately solving two sub-problems as in the traditional *Iterative Closest Point* (ICP) algorithms:

$$\textbf{Step 1:} \quad \underset{Y}{\arg\min} \sum_{i=1}^{n} \varphi(\mathbf{R}\mathbf{x}_i+\mathbf{t},\mathbf{y}_i) + I_{\mathcal{Y}}(\mathbf{y}_i) \tag{5}$$

$$\textbf{Step 2:} \quad \underset{\mathbf{R},\mathbf{t}}{\arg\min} \sum_{i=1}^{n} \varphi(\mathbf{R}\mathbf{x}_i+\mathbf{t},\mathbf{y}_i) + I_{SO(k)}(\mathbf{R}) \tag{6}$$

In the first step, the alignment is fixed and a set of correspondences Y is computed; in the second step, the correspondences remain fixed and the *optimal* rigid transformation is solved. As in each iteration the total energy weakly decreases (and this energy is bounded from below) this algorithm converges to a local minima [BM92]. The correspondences in Equation 5 are computed by finding the closest points $\mathbf{y}_i \in \mathcal{Y}$ from $\mathbf{x}_i$ as defined by the selected metric φ. This is possible as the first step is separable with respect to Y

$$\min_{Y} \sum_{i=1}^{n} \varphi(\tilde{\mathbf{x}}_i,\mathbf{y}_i) + I_{\mathcal{Y}}(\mathbf{y}_i) = \sum_{i=1}^{n} \min_{\mathbf{y}_i} \varphi(\tilde{\mathbf{x}}_i,\mathbf{y}_i) + I_{\mathcal{Y}}(\mathbf{y}_i), \tag{7}$$

where $\tilde{\mathbf{x}}_i = \mathbf{R}\mathbf{x}_i + \mathbf{t}$. As a consequence, the individual elements of $\mathcal{Y}$ can be optimized for independently from each other.

Classical ICP. In the seminal paper of Besl and McKay [BM92] the authors chose φ to be the classical *squared Euclidean distance* $\varphi(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_2^2$. This particular choice of metric has two important consequences. Firstly, the optimization sub-problems can be solved in closed form. Equation 6 can be solved by following the ideas presented in [ELF97], while the closest point computation can be accelerated, for example, by a kd-tree data structure with an ℓ_2 distance metric [FBF77]. However, this choice of metric affects the capability of ICP to deal with noise and outliers. Indeed, employing ℓ_2 signifies that we are optimizing for Equation 6 in a *least-squares* sense, imposing a fundamental assumption that the error residuals assume a normal distribution – i.e. where outliers *rarely* happen. In our setting, measurement outliers and incomplete data substantially violate this assumption. As discussed in Section 1, this problem can be mitigated by filtering the set of input correspondences by a number of heuristics. Conversely, we approach the problem by identifying an appropriate *robust metric* φ, and present an algorithm to efficiently solve the associated optimization problem.

4. Robust alignment by sparsity enforcement

Given a generic optimization residual $\mathbf{z}_i \in \mathbb{R}^k$, an outlier-robust scheme attempts to automatically categorize a vector of residuals $\mathbf{z} = [\|\mathbf{z}_1\|_2, \dots, \|\mathbf{z}_n\|_2]^T$ into a large set of inliers having $\|\mathbf{z}_i\|_2 \approx 0$ and a small set of outliers having $\|\mathbf{z}_i\|_2 \gg 0$. This objective can be achieved by attempting to find a *sparse* vector $\mathbf{z}$. As the ℓ_0 norm counts the number of non-zero entries in a vector, optimizing for sparsity can be re-formulated as minimizing $\|\mathbf{z}\|_0$. However, due to the problems caused by the high non-convexity of the ℓ_0-norm, a popular choice is to optimize for sparsity by employing the ℓ_1 norm [CRT06]. The ℓ_1 norm *penalizes* the number of non-zero entries, thus inducing sparsity; it is the closest convex relaxation of the ℓ_0-norm.

Non-convex relaxation. In this paper, we employ non-convex $\ell_p, p < 1$ relaxations of the ℓ_0 norm and optimize it with the *Alternating Direction Method of Multipliers*

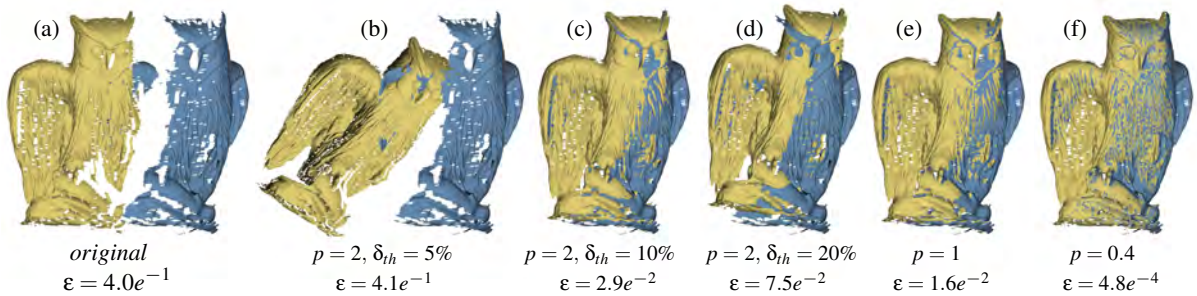


Figure 4: *(a) The alignment of the virtually scanned "owl" model is evaluated by the root mean square error (RMSE) ε w.r.t the ground truth alignment. (b,c,d) Traditional ICP registration is combined with correspondence pruning where correspondences with a distance above $\delta_{th}\%$ of the diagonal bounding box are rejected. As illustrated, it is difficult to find an appropriate threshold, which leads to sub-optimal alignment results. (e) ℓ_1-ICP without explicit outlier management converges to a better minimum, but the alignment is still poor. (f) Our ℓ_p-ICP outperforms all the previous methods.*

(ADMM). This approach has recently been shown to outperform its ℓ_1 counterpart in quality of results and performance [Cha07, MS12]. As we will demonstrate in Section 7, choosing $p < 1$ significantly improves the resilience of the method to large amounts of outliers.

Robust distance function. In our rigid alignment problem, given a pair of corresponding points $(\mathbf{x}_i, \mathbf{y}_i)$, the alignment residual is given by the vector $\mathbf{z}_i = \mathbf{R}\mathbf{x}_i + \mathbf{t} - \mathbf{y}_i$. Consequently, to optimize for alignment residual sparsity, we choose φ in Equation 4 to be $\varphi(\mathbf{x}, \mathbf{y}) = \phi(\|\mathbf{x} - \mathbf{y}\|_2)$, where $\phi(r) = |r|^p$ and $p \in [0, 1]$. Note that φ constructed in this fashion is still a metric [BK08, Chapter 1]. The curves of these functions, illustrated in Figure 2 for the simple scalar problem, can be interpreted as penalty curves; an outlier, having $\|\mathbf{z}\|_2 \gg 0$, will not be strongly penalized when a small value for p is chosen. This implies that the optimization will not skew the solution in order to reduce the large penalty associated with an outlier (see Figure 1). Please note that in this paper we are not solving for simple element-wise sparsity, but instead for *group sparsity* [EKB10], where each dimension of a residual vector $\mathbf{z}_i$ should vanish simultaneously (see Figure 5).

5. Numerical optimization

The introduction of sparse p-norms to increase robustness in the rigid alignment optimization results in a localized change to the classic ICP formulation. Hence, we still adopt the well-known two-step optimization:

$$\textbf{Step 1:} \quad \arg\min_{Y} \sum_{i=1}^{n} \|\mathbf{R}\mathbf{x}_i + \mathbf{t} - \mathbf{y}_i\|_2^p + I_{\mathcal{Y}}(\mathbf{y}_i) \tag{8}$$

$$\textbf{Step 2:} \quad \arg\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{n} \|\mathbf{R}\mathbf{x}_i + \mathbf{t} - \mathbf{y}_i\|_2^p + I_{SO(k)}(\mathbf{R}) \tag{9}$$

However, for $p \in [0, 1]$ these two problems are *non-convex* and *non-smooth*. We explain below how we can nevertheless obtain an efficient optimization algorithm.

5.1. Step 1 - Correspondences

$\phi(r) = |r|^p$ is a non-decreasing function on $\mathbb{R}^+$, $\phi(\|.\|_2)$ and thus achieves its minimum value at the same points as $\|.\|_2$. Consequently, the optimization in Equation 8 is equivalent

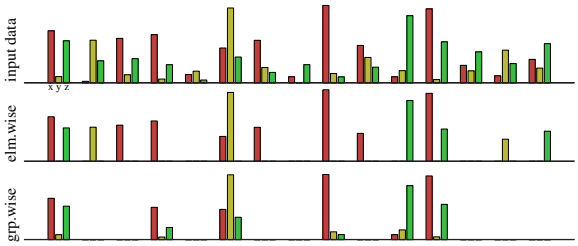


Figure 5: *Illustration of the difference between group sparsity and element-wise sparsity. For rigid alignment it is important to consider group sparsity, as a good corresponding pair will have small values in x, y, z simultaneously.*

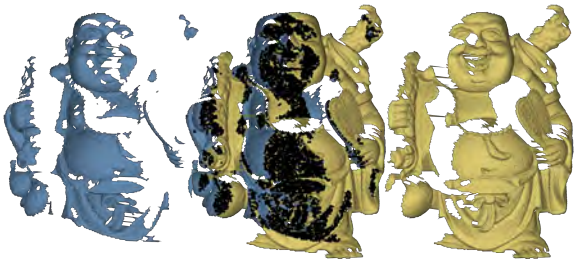
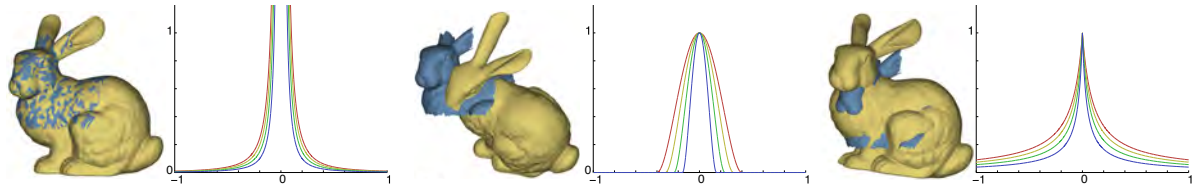


Figure 6: *Our shrinkage operator for $p \approx 0$ acts as a binary outlier classifier. Upon convergence, this classification simply identifies overlapping regions in the aligned geometry (marked in black).*



(a) p-norm weights (b) Tukey weights (c) fair weights

Figure 7: *A qualitative comparison of whole-in-part registration using point-to-point ICP. We compare two "robust weight functions" with the p-norm weight function presented in Equation 11. In (b) and (c) the weight functions fail to produce the correct registration because of the small amount of overlap, while our method succeeds in computing a perfect registration (a). It is interesting to note that the p-norm weight functions tend to infinity when approaching zero, strongly enforcing sparsity. We refer the reader to [MB93] for the definition of the robust weight functions.*

to

$$\arg\min_{Y} \sum_{i=1}^{n} \|\mathbf{R}\mathbf{x}_i + \mathbf{t} - \mathbf{y}_i\|_2 + I_{\mathcal{Y}}(\mathbf{y}_i), \tag{10}$$

allowing us to employ a kd-tree based on the l_2 metric for the first step of the optimization.

5.2. Step 2 - Alignment

Problems involving ℓ_p, similar to the one in Equation 9, can be approached by reweighting techniques [CY08]. For rigid registration this means iteratively solving the *weighted* least-squares problem

$$\arg\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{n} w_i^{p-2} \|\mathbf{R}\mathbf{x}_i + \mathbf{t} - \mathbf{y}_i\|_2^2 + I_{SO(k)}(\mathbf{R}), \tag{11}$$

where w_i is the ℓ_2 residual of the previous iteration. Given these weights, each iteration can be solved using classical methods for rigid transformation estimation. In practice, however, this approach suffers from instability when the residuals vanish, as $1/w_i^{2-p}$ goes to infinity; see Section 7. To optimize the problem in a robust manner, we introduce a new set of variables $Z = \{\mathbf{z}_i \in \mathbb{R}^k, i = 1 \dots n\}$ and then rewrite the problem as

$$\arg\min_{\mathbf{R},\mathbf{t},Z} \sum_{i=1}^{n} \|\mathbf{z}_i\|_2^p + I_{SO(k)}(\mathbf{R}) \quad \text{s.t} \quad \boldsymbol{\delta}_i = \mathbf{0}, \tag{12}$$

where $\boldsymbol{\delta}_i = \mathbf{R}\mathbf{x}_i + \mathbf{t} - \mathbf{y}_i - \mathbf{z}_i$ is introduced solely for compactness of notation. As detailed in Appendix A, *augmented Lagrangian* methods are an effective tool to approach the constrained optimization problem above. The augmented Lagrangian function for Equation 12 is defined as

$$\mathcal{L}_A(\mathbf{R},\mathbf{t},Z,\Lambda) = \sum_{i=1}^{n} \|\mathbf{z}_i\|_2^p + \boldsymbol{\lambda}_i^T \boldsymbol{\delta}_i + \frac{\mu}{2}\|\boldsymbol{\delta}_i\|_2^2 + I_{SO(k)}(\mathbf{R}),$$

where $\Lambda = \{\boldsymbol{\lambda}_i \in \mathbb{R}^k, i = 1 \dots n\}$ is a set of Lagrange multipliers and $\mu > 0$ is a penalty weight. We optimize this function by employing the *Alternating Direction Method of Multipliers* (ADMM); see Appendix B. ADMM effectively decomposes our problem into three simple steps:

Step 2.1: $$\arg\min_{Z} \sum_{i} \|\mathbf{z}_i\|_2^p + \frac{\mu}{2}\|\mathbf{z}_i - \mathbf{h}_i\|_2^2 \tag{13}$$

Step 2.2: $$\arg\min_{\mathbf{R},\mathbf{t}} \sum_{i} \|\mathbf{R}\mathbf{x}_i + \mathbf{t} - \mathbf{c}_i\|_2^2 + I_{SO(k)}(\mathbf{R}) \tag{14}$$

Step 2.3: $$\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_i + \mu\boldsymbol{\delta}_i \tag{15}$$

where $\mathbf{c}_i = \mathbf{y}_i + \mathbf{z}_i - \boldsymbol{\lambda}_i/\mu$ and $\mathbf{h}_i = \mathbf{R}\mathbf{x}_i + \mathbf{t} - \mathbf{y}_i + \boldsymbol{\lambda}_i/\mu$. In *Step 2.1*, like for Equation 7, the problem is separable and each $\mathbf{z}_i$ can be optimized independently. Each sub-problem can then be solved efficiently by applying the following *shrinkage operator* [PB13] to each vector $\mathbf{h}_i$:

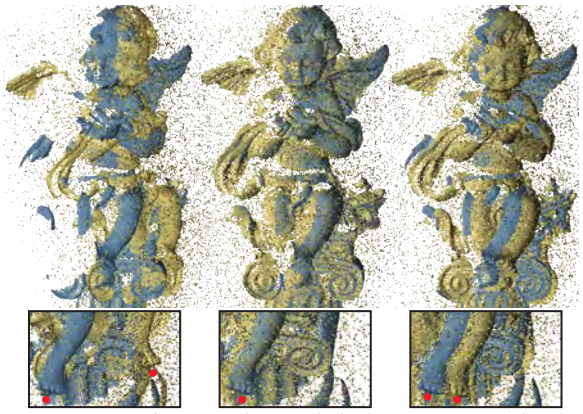
$$\mathbf{z}_i^* = \begin{cases} \mathbf{0} & \text{if} \quad \|\mathbf{h}_i\|_2 \leq \tilde{h}_i \\ \beta\mathbf{h}_i & \text{if} \quad \|\mathbf{h}_i\|_2 > \tilde{h}_i \end{cases} \tag{16}$$

The values of β and $\tilde{h}_i$ are detailed in Appendix C. The shrink operator can be interpreted as a *classifier* acting on residual vectors. For example, when $p = 0$, β from Equation 16 will always evaluate to one; this results in a *binary* classification: the operator either rejects the value h_i or accepts it fully. In Figure 6, we highlight correspondences classified as inliers which are naturally located in regions of overlap between source and target. Since the only free variables in *Step 2.2* are $\mathbf{R},\mathbf{t}$, this least-squares problem can approached by classical rigid transformation estimation techniques.

6. Higher order metric approximants

Rusinkiewicz and Levoy [RL01] experimentally observed how the point-to-plane variant of ICP has better convergence speed than its point-to-point variant. Subsequently, Pottmann *et al.* [PHYH06] formally demonstrated how under appropriate conditions, these classical algorithms possess linear resp. quadratic convergence properties. These observations were made possible by expressing rigid alignment as an optimization problem involving the Taylor expansion of the ℓ_2 metric $\varphi(\mathbf{x},\bar{\mathbf{y}})$ for $\bar{\mathbf{y}} \in \mathcal{Y}$, the "foot point" of $\mathbf{x}$, i.e. the closest point of $\mathbf{x}$ onto $\mathcal{Y}$ [FH10, pg.63], as

$$\tilde{\varphi}_{\bar{\mathbf{y}}}(\mathbf{x})|_{\mathbf{x}_0} \approx \|\mathbf{x}_0 - \bar{\mathbf{y}}\|_2 + \left(\frac{\mathbf{x}_0 - \bar{\mathbf{y}}}{\|\mathbf{x}_0 - \bar{\mathbf{y}}\|_2}\right)^T (\mathbf{x} - \bar{\mathbf{y}}). \tag{17}$$



(a) *original* (b) $p = 0.5$ (c) $p = 2.0$

Figure 8: *Partial registration of two synthetically corrupted laser scans (a) can be achieved even in the presence of a large amount of outliers in the source geometry by using a robust metric (b). The classical least-squares ICP ($p = 2$) fails to align the scans, as in this case the outliers heavily bias the estimation of the rigid alignment (c).*

As $\mathbf{x}_0$ approaches $\mathcal{Y}$ at the foot point $\bar{\mathbf{y}}$, the term $\|\mathbf{x}_0 - \bar{\mathbf{y}}\|_2$ vanishes and $\tilde{\varphi}_{\bar{\mathbf{y}}}(\mathbf{x})$ can be rewritten as

$$\tilde{\varphi}_{\bar{\mathbf{y}}}(\mathbf{x}) \approx \mathbf{n}^T(\mathbf{x} - \bar{\mathbf{y}}), \tag{18}$$

where $\mathbf{n}$ is the normal of the surface $\mathcal{Y}$ at $\bar{\mathbf{y}}$. Consequently, we can linearize the metric in Equation 8 as $\varphi(\mathbf{x}, \bar{\mathbf{y}}) = \phi(\|\mathbf{x} - \bar{\mathbf{y}}\|_2) \approx \phi(\tilde{\varphi}_{\bar{\mathbf{y}}}(\mathbf{x}))$ and rewrite the optimization as

$$\arg\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{n} |\tilde{\varphi}_{\mathbf{y}_i}(\mathbf{R}\mathbf{x}_i + \mathbf{t})|^p + I_{SO(k)}(\mathbf{R}). \tag{19}$$

This problem can be optimized using the same techniques introduced in Section 5, resulting in the ADMM steps:

$$\textbf{Step 2.1:} \quad \arg\min_{Z} \sum_i |z_i|^p + \frac{\mu}{2}(z_i - h_i)^2 \tag{20}$$

$$\textbf{Step 2.2:} \quad \arg\min_{\mathbf{R},\mathbf{t}} \sum_i (\delta_i - c_i)^2 + I_{SO(k)}(\mathbf{R}) \tag{21}$$

$$\textbf{Step 2.3:} \quad \lambda_i = \lambda_i + \mu\delta_i \tag{22}$$

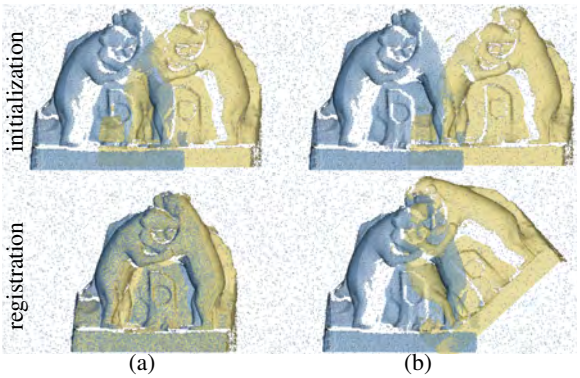
where $\delta_i = \mathbf{n}_i^T(\mathbf{R}\mathbf{x}_i + \mathbf{t} - \mathbf{y}_i)$, $h_i = \delta_i + \lambda_i/\mu$ and $c_i = z_i - \lambda_i/\mu$.

By comparing the optimization step in Equation 20 with the one of Equation 13, it becomes clear that the sparsity formulation of point-to-plane ICP involves a simple scalar problem. Furthermore, Equation 21 is nothing but a traditional least-squares point-to-plane optimization that we can solve by an elementary linearization of rotation matrices [Rus13].

7. Evaluation

In this section we empirically evaluate the characteristics of our method in comparison to some variations of the traditional ICP approach. In Figures 1, 2, and 4, we compare against the recent ℓ_1-ICP technique from [FH10]. Throughout the paper, unless otherwise specified, we employ the linearized point-to-plane distance metric; for Figure 8 and Figure 9, where surface normals are not well defined, we employ the point-to-point metric. We will focus our comparisons on *partial overlap* and *outliers* in the data, which are the most challenging problems for the ICP algorithms. Note that it is difficult to perform an exhaustive comparison of ICP methods considering the large amount of heuristics and variations available, and the size of the parameter space: 6 dimensional rigid transformation space for the initialization, p-parameter, noise parameters, etc. We therefore limit our analysis to a set of experiments that we believe best illustrate the unique properties of our approach.

Rejecting v.s. penalizing outliers. Figure 1 shows how least-squares ICP registration ($p = 2$) performs poorly unless appropriate heuristics for outlier management are adopted. Typical heuristics include rejecting pairs on boundaries, having distance above a given threshold or with dissimilar normals; see [RL01]. Outlier rejection approaches can be difficult to tune and even be harmful as they drastically increase the number of local minima. This is especially visible when the source is far from the target at the initialization. In this case, selecting a distance threshold too small will lead to the rejection of most of the correspondences, which in turn will increase the probability of converging to a bad local minimum. Selecting a distance threshold too big will not lead to a good final alignment as not enough outliers will be rejected. Conversely, our approach *weakly penalizes* outliers leading to a more stable approach. In Figure 4, we compare the alignment performance of our approach with outlier rejection techniques using a distance threshold. To evaluate the quality, we simulate the scan of a digital model, regis-



(a) (b)

Figure 9: *We align two models with partial overlap and a large number of outliers in the target geometry. As our ℓ_p algorithm still employs closest point correspondences, we converge to the correct solution only when source and target are in relatively close proximity (a). Placing the models too far from each other will drive the optimization towards a bad local minimum (b).*

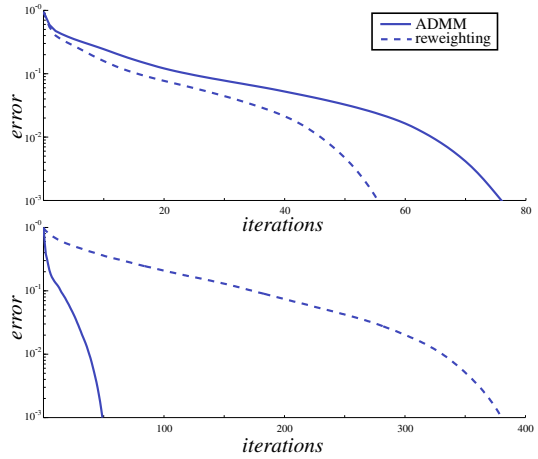


Figure 10: *Convergence of the reweighting approach compared to the ADMM method. When using point-to-point distances (top), iterative reweighting converges faster that ADMM. However, when using point-to-plane distances ADMM outperforms the iterative reweighting approach. The slow convergence of iterative reweighting in this case is due to the ill-conditioning of the associated linear system, as the weights vary in the range* $[0, \infty]$.

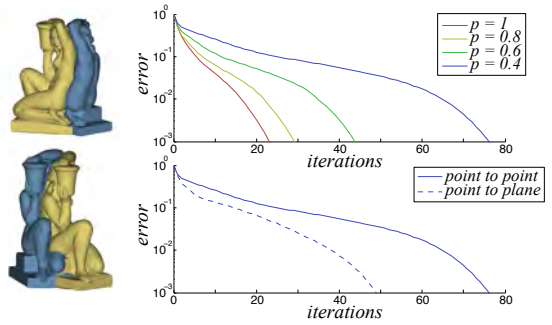


Figure 11: *An analysis of the convergence rate of our algorithm for different values of p (top) and first order approximations (bottom). When p decreases, the number of iterations of the ICP algorithm increases. Higher order approximations drastically speed-up the convergence of the ICP algorithm. The initial alignment is shown on the left.*

ter it, and evaluate the RMSE of the registered point locations w.r.t. the ground truth. As demonstrated by the values of RMSE in Figure 4 and in the supplemental video, our technique outperforms classical ICP with outliers rejection. Another approach for robust registration selects the best k-percent of the set of correspondences at each step of the registration [CSK05]. As illustrated in Figure 12, contrary to this approach, where the percentage of inliers k needs to be known beforehand, our method automatically selects the set of inliers for the registration.

Iterative reweighting. Discarding unreliable correspondences is undoubtedly the simplest and most common way of dealing with outliers. Another common approach is to reweight correspondences using "robust functions" [MB93] resembling the weight function presented in Equation 11; see Figure 7. An important observation is that only the weight function of p-norms, amongst the ones shown, tend to infinity as we approach zero. Clearly this has immediate implications in terms of sparsity, as only p-norms will greatly reward values approaching zero. Unfortunately, as p-norm weights are defined in the $[0, \infty]$ range, both stability and convergence speed of ICP algorithms employing reweighting are affected due to the ill-conditioning of the system; see [PBB11] and Figure 10. Fortunately, the ADMM optimization described in Section 5 does not suffer from this drawback as it does not employ a reweighting scheme. This makes our optimization both stable and efficient as demonstrated in Figure 10. It is interesting to note the similarity of the Tukey reweighting function in Figure 7-b and a weight function for outlier pruning that is a step function.

Outliers source v.s. target. In Figure 8, we show the effects of outliers in the *source* model. These outliers can be seen as samples lacking a proper *ground truth* match on the target shape. Outliers in the *target* geometry have a completely different effect. Consider the example in Figure 9, where the source geometry is outlier-free while the target has been corrupted with uniform environment noise. In this scenario, when a coarse initialization is given (small overlap between source and target), most closest-correspondence queries are incorrectly matched to outliers, thus severely affecting the effectiveness of the closest-point step of Equation 5.

Convergence speed. Figure 11-a illustrates the a trade-off between reducing p to achieve robust registration and convergence speed. As p is reduced, our registration becomes more resilient to outliers; however, the convergence speed decreases. This is mainly due to the fact that when p decreases, the contribution of pairs of correspondences with large distances also decreases forcing the optimization to take smaller steps. Figure 11-b reports a comparison between our sparse implementation of point-to-point and point-to-plane for $p = 0.4$. Similarly to what was previously discovered for standard ICP [RL01,PHYH06], our optimization also greatly benefits from the extra degree of freedom (i.e. tangential motion) of the point-to-plane metric approximation.

Selecting "p". Our experiments show how the performance of the algorithm varies with p, progressively trading-off between performance and robustness. Throughout our experiments, unless otherwise stated, we selected $p = 0.4$ as it seemed to offer a good trade-off.

Limitations. In our approach we essentially focused our efforts to robustify the second of the two ICP steps, i.e. Equa-

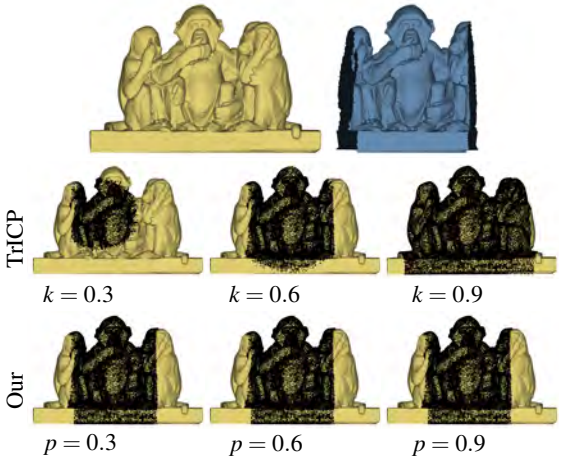


Figure 12: *Contrary to TrICP [CSK05], where the percentage of inliers k is a parameter of the algorithm, our approach automatically selects the set of inliers. As shown in this example, the set of inliers used at the last iteration of the registration (shown in black) is stable and accurate in our approach.*

tion 6. However, as we demonstrated in Section 7, the optimization step seeking closest correspondents (Equation 5) can still be affected by outliers. The performance of our algorithm is consequently degraded when there are too many outliers in the target geometry; see Figure 9. To address this problem, we would like to extend our formulation in order to relax the one-to-one correspondence assumption to one-to-many allowing *fuzzy* correspondences [GP02, CR03].

Source code and integration. An important advantage of our new ICP formulation is that already deployed implementations can easily be adapted. Code implementing heuristics for outlier selection can be improved using our ℓ_p optimization. We provide a C++ implementation at the URL `http://lgg.epfl.ch/sparseicp`.

8. Conclusions

We presented an extension of the classical ICP algorithm that systematically addresses the problem of outliers commonly observed in acquired 3D data. We express the ICP registration problem as a sparse ℓ_p optimization, obtaining an *heuristic-free*, *robust* rigid registration algorithm having only *one* free parameter. We believe that our method can replace or extend most existing ICP implementations that often are a crucial component in numerous geometry processing applications.

Acknowledgements

The authors thank Bailin Deng, Rick Chartrand, Dror Aiger and the reviewers for their valuable comments. We would also like to thank Niloy J. Mitra, Mario Deuss, Yuliy Schwartzburg, Boris Neubert, Cheryl Lau for proof-reading the paper and providing valuable comments. The model in (Figure 2) was scanned by Thibaut Weise. The models of Figures 4, 9, 11, 12 can be downloaded at `http://lgg.epfl.ch/statues`. This work has been supported by Swiss National Science Foundation (SNSF) grant #20PA21L_129607.

References

[AMCO08] AIGER D., MITRA N. J., COHEN-OR D.: 4-points congruent sets for robust pairwise surface registration. *ACM Trans. Graph. 27*, 3 (2008), 85:1–85:10. 2, 3

[BJMO12] BACH F. R., JENATTON R., MAIRAL J., OBOZINSKI G.: Optimization with sparsity-inducing penalties. *Foundations and Trends in Machine Learning 4*, 1 (2012), 1–106. 2, 4

[BK08] BANERJEE A., K A.: *Metric Space & Complex Analysis*. New Age International (P) Limited, 2008. 5

[BL04] BAE K.-H., LICHTI D. D.: Automated registration of unorganised point clouds from terrestrial laser scanners. In *In: International Archives of Photogrammetry and Remote Sensing, Vol. XXXV, Part B5, Proceedings of the ISPRS working group V/2* (2004), pp. 222–227. 2

[BM92] BESL P., MCKAY H.: A method for registration of 3D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence 14*, 2 (1992), 239–256. 2, 3, 4

[BPC*11] BOYD S., PARIKH N., CHU E., PELEATO B., ECKSTEIN J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning 3*, 1 (2011), 1–122. 11

[BV04] BOYD S., VANDENBERGHE L.: *Convex optimization*. Cambridge University Press, 2004. 11

[Cha07] CHARTRAND R.: Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Processing Letters 14*, 10 (2007), 707–710. 4, 5

[CLSB92] CHAMPLEBOUX G., LAVALLEE S., SZELISKI R., BRUNIE L.: From accurate range imaging sensor calibration to accurate model-based 3d object localization. In *Computer Vision and Pattern Recognition* (1992), pp. 83–89. 3

[CM91] CHEN Y., MEDIONI G.: Object modeling by registration of multiple range images. In *Proc. of the IEEE Intern. conf. on Robotics and Automation* (1991), pp. 2724–2729. 3

[CR03] CHUI H., RANGARAJAN A.: A new point matching algorithm for non-rigid registration. *Comput. Vis. Image Understanding 89*, 2-3 (2003), 114–141. 9

[CRT06] CANDÈS E. J., ROMBERG J., TAO T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory 52*, 2 (2006), 489–509. 4

[CSK05] CHETVERIKOV D., STEPANOV D., KRSEK P.: Robust euclidean alignment of 3d point sets: the trimmed iterative closest point algorithm. *Image and Vision Computing 23* (2005), 299–309. 8, 9

[CW08] CANDÈS E. J., WAKIN M. B.: An introduction to compressive sampling. *IEEE Signal Process. Mag. 25*, 2 (2008). 2

[CW13] CHARTRAND R., WOHLBERG B.: A Nonconvex Admm Algorithm for Group Sparsity With Sparse Groups. *Submitted to IEEE ICASSP* (2013). 4

[CY08] CHARTRAND R., YIN W.: Iteratively reweighted algorithms for compressive sensing. In *ICASSP* (2008), pp. 3869–3872. 6

[EKB10] ELDAR Y. C., KUPPINGER P., BOLCSKEI H.: Block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Trans. Signal Process. 58*, 6 (2010), 3042–3054. 5

[ELF97] EGGERT D. W., LORUSSO A., , FISHER R. B.: Estimating 3-d rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications*, 9 (1997), 272–290. 2, 3, 4

[FBF77] FRIEDMAN J. H., BENTLEY J. L., FINKEL R. A.: An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software (TOMS) 3*, 3 (1977), 209–226. 3, 4

[FH10] FLÖRY S., HOFER M.: Surface fitting and registration of point clouds using approximations of the unsigned distance function. *Computer Aided Geometric Design 27*, 1 (2010), 60–77. 4, 6, 7

[Fit03] FITZGIBBON A. W.: Robust registration of 2d and 3d point sets. *Image Vision Comput. 21* (2003), 1145–1153. 3

[GMGP05] GELFAND N., MITRA N. J., GUIBAS L., POTTMANN H.: Robust global registration. *Computer Graphics Forum (Proc. of the EG/SIGGRAPH Symposium on Geometry processing)* (2005), 197–206. 2

[GP02] GRANGER S., PENNEC X.: Multi-scale em-icp: A fast and robust approach for surface registration. In *Proc. of the European Conf. on Computer Vision* (2002), ECCV, pp. 418–432. 9

[HMS12] HONTANI H., MATSUNO T., SAWADA Y.: Robust non-rigid ICP using outlier-sparsity regularization. In *Computer Vision and Pattern Recognition* (2012). 4

[JV11] JIAN B., VEMURI B.: Robust point set registration using gaussian mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence 33*, 8 (2011), 1633–1645. 3

[LG04] LI Y., GU P.: Free-form surface inspection techniques state of the art review. *Computer-Aided Design 36*, 13 (2004), 1395–1417. 2

[LSP08] LI H., SUMNER R. W., PAULY M.: Global correspondence optimization for non-rigid registration of depth scans. In *Proc. of the EG/SIGGRAPH Symposium on Geometry processing* (2008). 2

[MB93] MIRZA M., BOYER K.: Performance evaluation of a class of m-estimators for surface parameter estimation in noisy range data. *IEEE Transactions on Robotics and Automation 9* (1993), 75–85. 6, 8

[MGPG04] MITRA N. J., GELFAND N., POTTMANN H., GUIBAS L.: Registration of Point Cloud Data from a Geometric Optimization Perspective. *Computer Graphics Forum (Proc. of the EG/SIGGRAPH Symposium on Geometry processing)* (2004), 22–31. 3

[MS12] MARJANOVIC G., SOLO V.: On ℓ_q optimization and matrix completion. *IEEE Trans. Signal Process. 60*, 11 (2012), 5714–5724. 2, 4, 5, 11

[MY95] MASUDA T., YOKOYA N.: A robust method for registration and segmentation of multiple range images. *Computer Vision and Image Understanding 61*, 3 (1995), 295 – 307. 3

[NW06] NOCEDAL J., WRIGHT S. J.: *Numerical optimization*. Springer Verlag, 2006. 11

[PB13] PARIKH N., BOYD S.: Proximal Algorithms. *Found. and Trends in Optimization* (2013), to appear. 6, 11

[PBB11] PYZARA A., BYLINA B., BYLINA J.: The influence of a matrix condition number on iterative methods convergence. In *Conf. on Comp. Science and Information Systems* (2011), pp. 459–464. 8

[PH03] POTTMANN H., HOFER M.: Geometry of the squared distance function to curves and surfaces. *Visualization and Mathematics III, Springer* (2003), 221–242. 3

[PHYH06] POTTMANN H., HUANG Q.-X., YANG Y.-L., HU S.-M.: Geometry and convergence analysis of algorithms for registration of 3D shapes. *International Journal of Computer Vision 67*, 3 (2006), 277–296. 3, 6, 8

[PLH04] POTTMANN H., LEOPOLDSEDER S., HOFER M.: Registration without ICP. *Computer Vision and Image Understanding 95*, 1 (2004), 54–71. 3

[PMG*05] PAULY M., MITRA N. J., GIESEN J., GROSS M., GUIBAS L. J.: Example-based 3d scan completion. In *Proc. of the EG/SIGGRAPH Symposium on Geometry processing* (2005). 2

[Pul99] PULLI K.: Multiview registration for large data sets. In *Proc. of 3DIM* (1999), pp. 160–168. 2

[RL01] RUSINKIEWICZ S., LEVOY M.: Efficient variants of the ICP algorithm. *3-D Digital Imaging and Modeling* (2001), 145–152. 2, 3, 6, 7, 8

[Rus13] RUSINKIEWICZ S.: Derivation of point to plane minimization, 2013. http://www.cs.princeton.edu/~smr/papers/icpstability.pdf. 7

[SHT09] SEGAL A., HAEHNEL D., THRUN S.: Generalized-icp. In *Proc. of Robotics: Science and Systems (RSS)* (2009), pp. 26–27. 2

[TCL*12] TAM G., CHENG Z.-Q., LAI Y.-K., LANGBEIN F., LIU Y., MARSHALL A., MARTIN R., SUN X.-F., ROSIN P.: Registration of 3d point clouds and meshes: A survey from rigid to non-rigid. *Trans. on Visualization and Computer Graphics 99*, PrePrints (2012), 1–20. 3

[TFR99] TRUCCO E., FUSIELLO A., ROBERTO V.: Robust motion and correspondence of noisy 3-D point sets with missing data. *Pattern Recognition Letters 20*, 9 (1999), 889 – 898. 3

[TK04] TSIN Y., KANADE T.: A correlation-based approach to robust point set registration. In *In ECCV* (2004), pp. 558–569. 3

[vKZHCO11] VAN KAICK O., ZHANG H., HAMARNEH G., COHEN-OR D.: A survey on shape correspondence. *Computer Graphics Forum 30*, 6 (2011), 1681–1707. 3

[WBLP11] WEISE T., BOUAZIZ S., LI H., PAULY M.: Realtime performance-based facial animation. *ACM Trans. Graph. (Proc. SIGGRAPH) 30*, 4 (2011), 77. 2

[Zha94] ZHANG Z.: Iterative point matching for registration of free-form curves and surfaces. *Int. J. Comput. Vision 13*, 2 (1994), 119–152. 3

Appendix A: Augmented Lagrangian Method (ALM)

We briefly discuss constrained optimization methods to provide a suitable background for the optimization approach taken in this paper. Consider the *equality-constrained* optimization problem having $x \in \mathbb{R}^k$:

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to } c_i(\mathbf{x}) = 0 \quad i = 1 \dots n. \tag{23}$$

Given a vector of Lagrange multipliers $\boldsymbol{\lambda} \in \mathbb{R}^n$ and $\mu \in \mathbb{R}^+$, we can transform the problem into an *unconstrained* optimization expressed by the *augmented Lagrangian* function

$$\mathcal{L}_A(\mathbf{x},\boldsymbol{\lambda},\mu) = f(\mathbf{x}) + \sum_{i=1}^{n} \lambda_i c_i(\mathbf{x}) + \frac{\mu}{2}\sum_{i=1}^{n} c_i^2(\mathbf{x}). \quad (24)$$

A close inspection of this expression reveals that it is a combination of a *linear* penalty typical of *dual ascent methods* [PB13, Ch. 2] and a *quadratic* one from *penalty methods* [NW06, Ch. 17]

$$\mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) = f(x) + \sum_{i=1}^{n} \lambda_i c_i(\mathbf{x}) \quad Q(\mathbf{x},\mu) = f(\mathbf{x}) + \frac{\mu}{2}\sum_{i=1}^{n} c_i^2(\mathbf{x}).$$

Intuitively, dual ascent methods exploit the fact that the Lagrange dual function $g(\boldsymbol{\lambda}) = \min_{\mathbf{x}} \mathcal{L}(\mathbf{x},\boldsymbol{\lambda})$ is a *lower bound* on the primal problem [BV04, pp. 216]. A gradient ascent w.r.t. $\boldsymbol{\lambda}$ of appropriate size gradually closes the *duality gap*, effectively optimizing Equation 23 as shown in Equation 26. Conversely, penalty methods gradually increase the penalty parameter μ, resulting in progressively increased constrain satisfaction. Unfortunately, both approaches suffer severe limitations; dual ascent methods place strong assumptions on the problem, like strict convexity and boundedness of $g(.)$; penalty methods offer limited control on constraint satisfaction, running into numerical issues as μ is increased to very large values. These issues are addressed by the augmented Lagrangian approach; the optimization involves an algorithm strongly resembling dual ascent:

Step 1: $$\mathbf{x}^{t+1} := \arg\min_{x} \mathcal{L}(\mathbf{x},\boldsymbol{\lambda}^t,\mu) \quad (25)$$

Step 2: $$\lambda_i^{t+1} := \lambda_i + \mu\, c_i(\mathbf{x}^{t+1}) \quad i = 1\ldots n \quad (26)$$

providing an asymptotic approximate constraint satisfaction in the form [NW06, Thm 17.2]

$$c_i(x) \approx (\lambda_i^* - \lambda_i)/\mu. \quad (27)$$

Consequently, constraints can be satisfied by either increasing μ, or alternatively by providing multipliers λ having values close to the optimal λ^*; note that this is achieved by the Lagrange multiplier ascent step inherited from dual ascent; see Equation 26.

Appendix B: Alternating Direction Method of Multipliers

The ADMM method is an extension of the augmented Lagrangian method to optimize a compound problem $f(\mathbf{x}) = \tilde{f}(\tilde{\mathbf{x}}) + \hat{f}(\hat{\mathbf{x}})$ by decoupling it into simpler subproblems [BPC*11]. Under appropriate conditions [PB13, Ch. 3], as the optimizer is separated in $\mathbf{x} = [\tilde{\mathbf{x}}, \hat{\mathbf{x}}]$, the problem can be approached by iteratively solving the following three steps:

Step 1: $$\tilde{\mathbf{x}}^{t+1} := \arg\min_{\tilde{\mathbf{x}}} \mathcal{L}([\tilde{\mathbf{x}}, \hat{\mathbf{x}}^t], \boldsymbol{\lambda}^t, \mu) \quad (28)$$

Step 2: $$\hat{\mathbf{x}}^{t+1} := \arg\min_{\hat{\mathbf{x}}} \mathcal{L}([\tilde{\mathbf{x}}^{t+1}, \hat{\mathbf{x}}], \boldsymbol{\lambda}^t, \mu) \quad (29)$$

Step 3: $$\lambda_i^{t+1} := \lambda_i^t + \mu\, c_i(\mathbf{x}^{t+1}) \quad i = 1\ldots n \quad (30)$$

Note that alternately looping over *Steps 1/2* until convergence would simply correspond to a *block coordinate descent* decomposition of the joint minimization in Equation 25. Consequently, ADMM can be interpreted as applying a single step of block coordinate descent optimization applied to the augmented Lagrangian problem.

Appendix C: Shrink operator for $f(z) = \|\mathbf{z}\|_2^p + \frac{\mu}{2}\|\mathbf{z}-\mathbf{h}\|_2^2$

The minimization of $f(z)$ can be simplified into a scalar problem by noticing that $\mathbf{z}$ can be expressed as $\mathbf{z} = \alpha\mathbf{h}$ (see Appendix D) leading to

$$\min_{\alpha} \|\alpha\mathbf{h}\|_2^p + \tfrac{\mu}{2}\|\alpha\mathbf{h}-\mathbf{h}\|_2^2 = \min_{\alpha} \|\mathbf{h}\|_2^{q-2}|\alpha|^p + \tfrac{\mu}{2}|\alpha-1|^2. \quad (31)$$

As proven in [MS12], the optimal α^* is given by:

$$\alpha^* = \begin{cases} 0 & \text{if } \|\mathbf{h}\|_2 \le \tilde{h} \\ \beta & \text{if } \|\mathbf{h}\|_2 > \tilde{h} \end{cases} \quad (32)$$

where the threshold $\tilde{h}$ is computed as

$$\tilde{h} = \alpha_a + \tfrac{p}{\mu}\alpha_a^{p-1}, \quad \alpha_a = \left(\tfrac{2}{\mu}(1-p)\right)^{\frac{1}{2-p}}. \quad (33)$$

β is found by using the following update scheme

$$\beta_{t+1} = 1 - \tfrac{p}{\mu}\|\mathbf{h}\|_2^{p-2}\beta_t^{p-1} \quad (34)$$

by initializing $\beta_0 \in [\alpha_a\|\mathbf{h}\|_2^{-1}, 1]$. Similarly to [MS12], we noticed that this scheme converges in two or three iterations. The optimal $\mathbf{z}$ is then given by $\mathbf{z}^* = \alpha^*\mathbf{h}$.

Appendix D: Scalar version of $f(z) = \|\mathbf{z}\|_2^p + \frac{\mu}{2}\|\mathbf{z}-\mathbf{h}\|_2^2$

The minimization of $f(z)$ can be re-interpreted into a simpler scalar problem where the optimal solution can be expressed as $\mathbf{z}^* = \alpha^*\mathbf{h}$, $\alpha \in \mathbb{R}$. We verify this by expressing $\mathbf{z}$ as a linear combination of a component $\alpha\mathbf{h}$ lying in the space defined by $\mathbf{h}$, and κ orthogonal to it, that is, $\kappa^T\mathbf{h} = 0$. All that is necessary is to prove that $\forall\kappa \in \mathbb{R}^3, f(\alpha\mathbf{h}+\kappa) \ge f(\alpha\mathbf{h})$. This can be verified by checking that both of the two following inequalities $\|\alpha\mathbf{h}+\kappa\|_2^p \ge \|\alpha\mathbf{h}\|_2^p$ and $\|\alpha\mathbf{h}+\kappa-\mathbf{h}\|_2^2 \ge \|\alpha\mathbf{h}-\mathbf{h}\|_2^2$ hold. We can easily convert the first of these into one involving quadratic exponents only. We first raise both sides to the power $2/p$ and then expand the norm as an inner product obtaining $(\alpha\mathbf{h}+\kappa)^t(\alpha\mathbf{h}+\kappa) \ge \|\alpha\mathbf{h}\|^2$. As $\mathbf{h}$ and κ are orthogonal we get $\|\alpha\mathbf{h}\|_2^2 + \|\kappa\|_2^2 \ge \|\alpha\mathbf{h}\|_2^2$; this is always verified as $\|\kappa\|_2^2 \ge 0$. To verify the second expression, we again expand the norm as an inner product, then, after removing $\|\mathbf{h}\|_2^2$ from both sides, we obtain $\|\alpha\mathbf{h}+\kappa\|_2^2 - (\alpha\mathbf{h}+\kappa)^T\mathbf{h} \ge \|\alpha\mathbf{h}\|_2^2 - (\alpha\mathbf{h})^T\mathbf{h}$. Similarly to what we did before, we expand the norm and exploit orthogonality, simplifying the expression to $\|\kappa\|^2 \ge 0$, that, again, is always verified.