Sampling at unknown locations: Uniqueness and reconstruction under constraints

Golnoosh Elhami†, *Student Member, IEEE,* Michalina Pacholska†, *Student Member, IEEE,*
Benjamín Béjar Haro, *Member, IEEE,* Martin Vetterli, *Fellow, IEEE,* Adam Scholefield, *Member, IEEE*

***Abstract*—Traditional sampling results assume that the sample locations are known. Motivated by simultaneous localization and mapping (SLAM) and structure from motion (SfM), we investigate sampling at unknown locations. Without further constraints, the problem is often hopeless. For example, we recently showed that, for polynomial and bandlimited signals, it is possible to find two signals, arbitrarily far from each other, that fit the measurements. However, we also showed that this can be overcome by adding constraints to the sample positions.**

In this paper, we show that these constraints lead to a uniform sampling of a composite of functions. Furthermore, the formulation retains the key aspects of the SLAM and SfM problems, whilst providing uniqueness, in many cases.

We demonstrate this by studying two simple examples of constrained sampling at unknown locations. In the first, we consider sampling a periodic bandlimited signal composite with an unknown linear function. We derive the sampling requirements for uniqueness and present an algorithm that recovers both the bandlimited signal and the linear warping. Furthermore, we prove that, when the requirements for uniqueness are not met, the cases of multiple solutions have measure zero.

For our second example, we consider polynomials sampled such that the sampling positions are constrained by a rational function. We previously proved that, if a specific sampling requirement is met, uniqueness is achieved. In addition, we present an alternate minimization scheme for solving the resulting non-convex optimization problem.

Finally, fully reproducible simulation results are provided to support our theoretical analysis.

***Keywords*—Sampling, sampling at unknown locations, SLAMping.**

I. Introduction

As we navigate through our surroundings, we are able to visually map the 3-D structure of the environment and at the same time localize ourselves within it. As humans, we do this so naturally that it is tempting to assume that the problem is trivial. However, theoretically understanding this process is far from easy.

The most obvious existing work in this direction comes from the robotics and computer vision communities in the form of simultaneous localization and mapping (SLAM) [1,2] and structure from motion (SfM) [3,4]. In the traditional SLAM problem, one considers a robot measuring distances and/or directions between itself and a set of landmarks. Each time the robot moves, it obtains an estimate of this movement (from odometry sensors) and takes new measurements to the landmarks. The aim is to use this data to estimate both the location of the landmarks and the robot's trajectory.

SfM is very similar. In this case, one typically has a set of images of the same scene taken from different viewpoints. The aim is to build a 3-D model of the scene and estimate the pose of the camera for each of the input images. This is traditionally done by extracting key feature points from the scene that can be matched between views. Reconstructing the scene geometry and camera poses is then a problem in multi-view geometry.

We see that in both these cases, the features/landmarks that are considered are discrete. While this simplification leads to practical algorithms, it does not fully model the underlying continuous world. To do this, we argue that one needs to consider the problem from a sampling perspective.

Sampling results have two main components: a signal model and a sampling scheme. For example, in Nyquist-Shannon sampling [5]–[7], one assume that the signal belongs to the shift-invariant space of band-limited functions and the signal

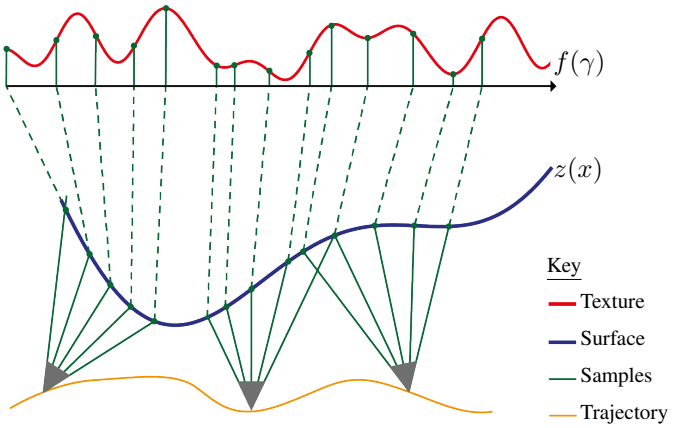


Fig. 1: The connection between sampling at unknown locations and SLAM/SfM. Here a camera moves along a trajectory and takes images of a surface. The surface is painted with a texture, which the camera measures. The locations of its measurements is dictated by the surface geometry, trajectory and camera orientations. Since none of these are known the sample locations are unknown.

†The first two authors have equal contribution to this work and the order is alphabetical.

This work was supported by the Swiss National Science Foundation grant number 20FP-1 151073, "Inverse Problems regularized by Sparsity".

Golnoosh Elhami, Michalina Pacholska, Martin Vetterli and Adam Scholefield are with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland, email: firstname.lastname@epfl.ch. Benjamín Béjar Haro is with the Vision Lab, Johns Hopkins University, Baltimore, MD, e-mail: bbejar@jhu.edu.

is sampled at uniform known locations. Extensions have been made in both of these directions, leading to additional sampling results [8]–[10]. For example, on the signal model side, sampling results have been developed for general-shift invariant spaces and other more complex spaces [11,12]. On the sampling scheme side, the known uniform sampling positions have been generalized to the case of a small unknown additive perturbation as well as non-uniform known locations [13,14].

In SLAM and SfM, we take measurements to landmarks/features at unknown locations. Therefore, to develop a sampling theory for these problems, we need to consider the problem of sampling at unknown locations.

In this case, unsurprisingly, uniqueness is not guaranteed in general. In fact, in [15], we show that, for polynomial and bandlimited signals, it is possible to find a valid solution arbitrarily far from the original signal (we review this result in Lemma 1 of this paper).

However, despite this result, we know that algorithms exist that can solve SLAM and SfM; therefore, given the correct constraints, it is possible to recover the measurement positions and underlying function from samples at unknown locations. In this paper, we formulate a set of constraints on the sampling positions, which both retain applicability to SLAM and SfM and, at least in some cases, lead to uniqueness.

To see this, consider the toy problem depicted in Fig. 1. Here we show a surface, which we assume is painted with an unknown texture, being sampled by a camera at three positions along an unknown trajectory. We could also remove the trajectory and view this as three cameras viewing the same surface. Note that here we assume that we are in flatland but the general idea extends to higher dimensions. As the figure shows, the cameras take samples of the texture at non-uniform locations. Furthermore, these locations are unknown, since they are governed by the unknown surface and unknown camera poses. However, if we assume that the surface and trajectory belong to some known function space, the sample positions are no longer arbitrary[1]. In this paper, we consider problems of this form; that is, functions sampled at unknown locations but where the locations of the samples are constrained by another function. As we show in the next section, this can also be interpreted as a uniform sampling of a composite of functions.

To emphasize, in this paper, we are proposing sampling of a composite of functions as a problem with previously unseen practical relevance. As a first analysis in this direction, we do not analyze the full SLAM and SfM setups and the algorithms we propose are not in anyway intended to be practical algorithms that compete with the state of the art in these fields. Instead, we study two simple incarnations of constrained sampling at unknown locations:

1) We show that periodic bandlimited signals can be efficiently recovered from an unknown linear warping.
2) We show uniqueness for polynomial signals constrained by a rational function. This result originally appeared in [15] but we present it here under the more general framework we are proposing.

We believe that these two incarnations provide a first step towards a deeper theoretical understanding of the more complex SLAM and SfM problems.

In relation to prior work, sampling at unknown locations is a relatively unexplored topic. For the continuous problem that we consider in this paper, Browning proposed an alternating least squares algorithm that converges to a local minimum [16] and Kumar considered the case where the unknown sample positions are governed by a stochastic model. He was able to show that the reconstruction error is asymptotically inversely proportional to the number of samples [17] [18].

Following the statistical theme, there is an interesting connection between sampling composites of functions and single index models [19]. These models, which are commonly used in econometrics, add more flexibility to parametric models by composing a linear function with a non-parametric function. Typically, it is assumed that the underlying signal follows a Gaussian distribution [20,21], although this condition has been recently relaxed [22]. However, only the linear part of the composite is recovered and only up to a scaling.

In the discrete case, Marziliano et. al. investigated the recovery of bandlimited signals [23] and there is a connection to the recent work on unlabelled sensing [24]–[28].

Additionally, since we consider a composite of functions, there is a connection to previous works on sampling time-warped signals [29]–[33]. In fact, in [34], we used a result from [33] to show that, for particular warpings of bandlimited signals, uniqueness can be guaranteed. We also proposed an algorithm based on local bandwidth to recover the shape of a surface from an image. Clerc et. al. introduced 'warplets' to perform surface retrieval in a similar spirit [35].

In this paper, we also consider toy examples of surface retrieval but using the two sampling results we develop for composites of functions.

The rest of this manuscript is organized as follows. We first define the problem of sampling at unknown locations and show that in many cases it is ill-posed. We use this as motivation to introduce the constraints that lead to a sampling of a composite of functions. Then, we consider the two incarnations previously mentioned. Finally, we present simulations results supporting our theoretical findings and conclude. The code used for simulations is available at github.com/LCAV/surface-reconstruction.

II. Problem formulation

In this section, we formalize the problem of recovering a function from a finite set of samples at unknown locations, show that the problem is in general ill-posed and show how additional constraints on the sample positions can be used to regularize it. In doing so we are effectively transforming the problem of signal recovery from unknown irregular sample positions to that of regular sampling of a composite of functions.

Consider the following setup: let $\mathcal{F}$ be a linear space of functions defined over some interval $X \subseteq \mathbb{R}$ and let $\mathrm{T}_{\boldsymbol{x}}$ be a sampling or acquisition device that records the value of a function, $f \in \mathcal{F}$, at the set of locations $\boldsymbol{x} = [x_0, \ldots, x_{N-1}]$

[1]In the general case depicted in Fig. 1, we need an additional function enforcing a 'trajectory' for the camera's orientation.

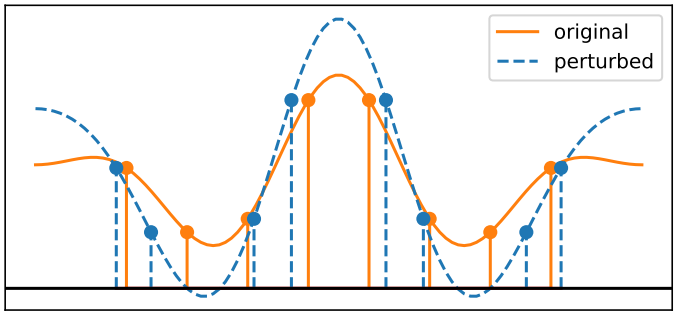


Fig. 2: If we sample function f from some linear space (here bandlimited) of continuous functions $\mathcal{F}$, we can add to it some perturbation. If the perturbation is small enough the samples can be moved to match the perturbed function.

with $x_n \in X$, $n = 0, \dots, N-1$. Assume we observe $f \in \mathcal{F}$ at N unknown and distinct locations over the interval; that is, we measure $\boldsymbol{y} = \mathrm{T}_{\boldsymbol{x}} f = \{f(x_0), \dots, f(x_{N-1})\}$, where $x_n \neq x_m$ for $i \neq j$. The knowledge on the sampling device is limited. In the most general case we consider, the only knowledge about the sampling instants is their linear order, that is $x_0 < x_1 < \cdots < x_{N-1}$.

The question is whether we can recover the original f from the set of observations. Since $\mathcal{F}$ is a linear space, recovering functions is understood as finding the expansion coefficients of the function f in the space $\mathcal{F}$.

We call *a solution* any function $\tilde{f} \in \mathcal{F}$ that could have been a source for the observed samples; that is, a function for which there exists an ordered sequence $\{\tilde{x}_0, \dots, \tilde{x}_{N-1}\}$ such that $\tilde{x}_n \in X$ and $\tilde{f}(\tilde{x}_n) = f(x_n)$ for all $n = 0, \dots, N-1$. Of course, f itself is a solution.

A. Non-uniqueness

It is clear that, without any further constraints on the sample positions, many solutions may exist. Except for trivial cases, the problem is ill-posed, since every measurement introduces a new unknown—its location. For instance, in the case of sampling bandlimited signals at unknown locations one can find many valid solutions by just adding a small perturbation to the original samples [16] (see Fig. 2). However, in many cases, the situation is even worse. As proved in [15, Lemma 1], for polynomial and bandlimited functions, one can find a solution arbitrarily far from the original (in any norm). We restate this result here using the notation adopted in this paper.

Lemma 1 (Pacholska *et al*, 2017)**.** *Let $f \in \mathcal{F}$ and let $\boldsymbol{y} = \mathrm{T}_{\boldsymbol{x}} f = \{f(x_0), \dots, f(x_{N-1})\}$ be the samples of f. Furthermore, suppose $x_0 < x_1 < \cdots < x_{N-1}$. If*

1) $\mathcal{F}$ is the class of polynomials of degree at most m, or
2) $\mathcal{F}$ is the class of real-valued, m-bandlimited functions,

then for any $C > 0$ there exists a function $\tilde{f} \in \mathcal{F}$ such that $\|f - \tilde{f}\| \geq C$ and points $\tilde{\boldsymbol{x}} = [\tilde{x}_0, \dots, \tilde{x}_{N-1}]$, with $\tilde{x}_0 = x_0$ and $\tilde{x}_{N-1} = x_{N-1}$, such that $f(x_n) = \tilde{f}(\tilde{x}_n)$.

Proof: See [15]. ■

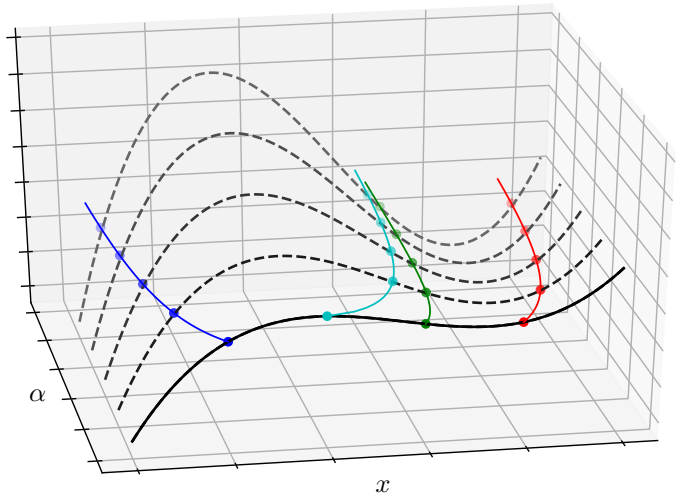


Fig. 3: Illustration of the path of solutions generated by moving along $g \in \mathcal{F}$ from the initial f. The path of solutions ($\tilde{f}_\alpha = f + \alpha g$, dashed lines) defines a trajectories along which samples can move (solid lines).

In the statement of Lemma 1, the bracket $\|\cdot\|$ can denote any norm on the linear space $\mathcal{F}$. Additionally, note that the fixing of the first and last samples ($\tilde{x}_0 = x_0$ and $\tilde{x}_{N-1} = x_{N-1}$) is not necessary. However, without this restriction, it is very easy to find another function that could have produced the samples (e.g. shift the domain of the original function). The lemma shows that, even with this additional restriction, it is still possible to find a function, arbitrarily far from the original, that could have produced the samples.

The proof of the lemma is based on the construction of a function $\tilde{f}$, or equivalently on finding a direction g in the linear space $\mathcal{F}$, such that if we move in that direction (e.g. take $\tilde{f} = g + f$) the values of the maxima do not decrease and the values of the minima do not increase. Such a function g then defines a *path of solutions*, on which we can find a solution arbitrary far from the original (see [15]). An illustration of this path of solutions is shown in Fig. 3.

B. Constraining the sample positions

Lemma 1 gives one way in which sampling at unknown locations can break. Whilst this is a negative result, it also gives us some intuition of how we can fix the problem. More precisely, the lemma gives us a path of solutions, which defines a trajectory for the sample positions, see Fig. 3. If the samples can move freely, they will adjust to any function on the path. However, if we restrict the way samples can move, there is a high chance that at least one of the samples will not lie on the trajectory defined by the path. Therefore, adding any constraint will remove at least some of the large scale ambiguity. This observation motivates us to regularize the inverse problem by adding a constraint on the allowed sample trajectories.

To do this, let's constrain the sample positions. Let $\boldsymbol{x} = [x_0, \dots, x_{N-1}]$ be the true sample positions. Instead of allow-

ing the sample positions to move arbitrarily, let's only allow sample positions $\tilde{\boldsymbol{x}} = [\tilde{x}_0, \dots, \tilde{x}_{N-1}]$ satisfying the constraint $\tilde{x}_n = \psi(x_n)$, for all n; that is, we only allow sample positions that are a function of the true sample positions. If ψ is unknown but comes from some known family of functions, maybe we can recover the original function f.

An alternative way to constrain how samples can move is to consider a uniform sampling of a composite of functions. Let $\varphi \in \Phi$ be a function from a known space of functions, Φ, and assume that the sample positions are $x_n = \varphi(nT)$ where $n \in [0, \dots, N-1]$. That is, we uniformly sample the composite $f \circ \varphi$, obtaining $\boldsymbol{y} = \mathrm{T}_{\{nT : n \in [0..N-1]\}}(f \circ \varphi)$. Now, although the true φ is unknown, we still know that any valid set of sample positions must satisfy $\tilde{x}_n = \tilde{\varphi}(nT)$, for some $\tilde{\varphi} \in \Phi$. Furthermore, to maintain the order of samples, let us restrict φ to be a monotonically increasing function. It follows that φ is invertible and we can define $\psi := \tilde{\varphi} \circ \varphi^{-1}$. With this definition, we have $\tilde{x}_n = \tilde{\varphi}(\varphi^{-1}(x_n)) = \psi(x_n)$, showing the equivalence to the previous formulation.

Therefore, by constraining the sampling positions, we have changed the problem from sampling at arbitrary unknown locations to sampling a composite of functions $(f \circ \varphi)$ at known uniform locations. We often think about φ as a warping of f. We thus sample a warped version of f and wish to recover both f and the warping φ.

To summarize, let $f \in \mathcal{F}$ be the signal of interest, and let $\varphi \in \Phi$ be a warping function. The problem to solve is

$$\begin{aligned} &\text{find} && \{f \in \mathcal{F}, \varphi \in \Phi\} \\ &\text{s.t.} && \boldsymbol{y} = \mathrm{T}_{\{nT : n \in [0,\dots,N-1]\}}(f \circ \varphi). \end{aligned} \tag{1}$$

As a motivating example of the proposed framework, consider a camera in flatland, i.e. a 2-D world, viewing a linear surface $z(x)$ painted with an unknown texture f as illustrated in Fig. 4. We would like to recover both the texture and the surface from a set of observations. Under this setup we can distinguish between the following two scenarios:

1) Orthographic projection: In the orthographic projection case, depicted in Fig. 4a, the sample positions are simply $x_n = nT\cos\theta = \varphi(nT)$; i.e., the warping function is a scaling: $\varphi(x) \in \Phi = \{x \mapsto bx \text{ for } b \in \mathbb{R}\}$, where in our example b is the cosine of the unknown surface orientation θ[2]. Note that, in this example, the distance d of the surface from the camera does not affect the measurements and is thus unrecoverable.

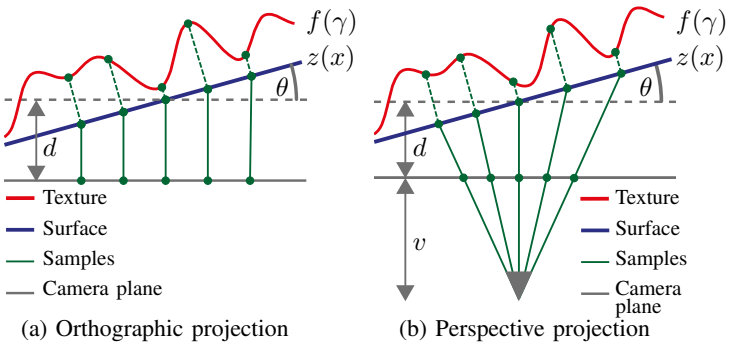
To find the corresponding constraint function $\psi \in \Psi$, let θ be the true surface orientation. The true sample positions, x_n, are related to the sample positions, $\tilde{x}_n$, for a surface with angle $\tilde{\theta}$, by $x_n \cos\theta = \tilde{x}_n \cos\tilde{\theta}$. Therefore,

$$\tilde{x}_n = \frac{x_n \cos\theta}{\cos\tilde{\theta}} = \psi(x_n);$$

i.e., the samples are constrained to move according to

$$\psi(x) \in \Psi = \left\{ x \mapsto \frac{x\cos\theta}{\cos\tilde{\theta}} \text{ for } \theta \in (-\pi/2, \pi/2) \right\}.$$

[2]Since $b = \cos\theta$, $b \in [-1, 1]$. However, for generality, we consider the case $b \in \mathbb{R}$.



(a) Orthographic projection (b) Perspective projection

Fig. 4: Orthographic and perspective projections. Examples of sampling a warped signal, where the warping is define by the camera. Note that in the orthographic projection the warped samples are equally spaced, what is not the case for perspecitve projection.

2) Perspective projection: Similarly, in the perspective projection case, depicted in Fig. 4b, for odd N, we have

$$\frac{x_n\cos\theta}{x_n\sin\theta + d} = \frac{nT}{v} \quad \Rightarrow \quad x_n = \frac{nTd}{v\cos\theta - nT\sin\theta} = \varphi(nT);$$

i.e., the warping function is $\varphi(x) \in \Phi$, where

$$\Phi = \left\{ x \mapsto \frac{dx}{v\cos\theta - x\sin\theta} \text{ for } d, v \in \mathbb{R}^+ \text{ and } \theta \in (-\pi/2, \pi/2) \right\}.$$

Let d and θ be the parameters of the true surface and $\tilde{d}$ and $\tilde{\theta}$ be the parameters of any other surface. Then, since

$$\frac{x_n\cos\theta}{x_n\sin\theta + d} = \frac{\tilde{x}_n\cos\tilde{\theta}}{\tilde{x}_n\sin\tilde{\theta} + \tilde{d}},$$

we can find the constraint function from

$$\tilde{x}_n = \frac{\tilde{d}x_n\cos\theta}{d\cos\tilde{\theta} + x_n\sin(\theta - \tilde{\theta})} = \psi(x_n);$$

i.e., the constraint function satisfies $\psi \in \Psi$, where

$$\Psi = \left\{ x \mapsto \frac{\tilde{d}x\cos\theta}{d\cos\tilde{\theta} + x\sin(\theta - \tilde{\theta})} \text{ for } d, \tilde{d} \in \mathbb{R}^+ \text{ and } \theta, \tilde{\theta} \in (-\pi/2, \pi/2) \right\}.$$

For the majority of this paper, we investigate setups related to these two examples, and we focus on deterministic families Ψ and Φ. However, the framework we are proposing can be naturally extended to describe much of the existing works on sampling at unknown locations. For example, replacing Φ with a space of random functions, we align with the probabilistic framework of Kumar. When Φ is the space of i.i.d. random variables independent of the input, we have the model analyzed in [17] and when it is a random process we have the model analyzed in [18]. The framework can also be used to describe measurements taken approximately at known positions.

In the following two sections, we consider two simple cases intimately connected to the previous two examples: first, we consider periodic bandlimited signals warped by a linear function and, second, we consider polynomials with sample locations constrained to be a rational function of the true sample positions. In the first case, we show when the function and sample locations can be retrieved and present an algorithm that performs this recovery. In the second case, we present a uniqueness result and an iterative algorithm that attempts to find this unique solution.

III. Periodic bandlimited signals warped by a linear function

In this section, we consider periodic bandlimited signal with a linear warping. We describe the problem in terms of its Fourier representation and show when the warping parameter is unique. Finally, we propose an algorithm to recover the linear warping and the signal with provable guarantees.

Let $f(x)$ be a τ-periodic and bandlimited signal given by

$$f(x) = \sum_{k=-K}^{K} a_k \, e^{j2\pi kx/\tau}. \tag{2}$$

Note that a_k corresponds to the *Fourier Series* (FS) coefficients of $f(x)$. The warped signal is

$$(f \circ \varphi)(x) = f(bx) = \sum_{k=-K}^{K} a_k \, e^{j2\pi kbx/\tau}, \tag{3}$$

where $\varphi(x) = bx$, with $b \neq 0$, is a linear transformation of the domain of $f(x)$. This composite is sampled uniformly with a sampling rate of $T = \tau/N$ (e.g. N samples per period) producing the sequence $y_n = (f \circ \varphi)(nT)$:

$$y_n = \sum_{k=-K}^{K} a_k e^{j2\pi kbn/N}. \tag{4}$$

The goal is then to recover both $f(x)$ and $\varphi(x)$ from the set of observations. In other words, we would like to find a_k, to reconstruct $f(x)$, and b to reconstruct $\varphi(x)$.

The main insight of our approach is that, in the Fourier domain, the signal $f(x)$ corresponds to a set of Diracs uniformly spaced in frequency:

$$F(\omega) = \mathcal{F}\{f(x)\} = \sum_{k=-K}^{K} a_k \delta\left(\omega - \frac{2\pi k}{\tau}\right). \tag{5}$$

By introducing a linear warping, we are effectively expanding or contracting the spacing between those Diracs (modulo 2π as a result of sampling). It follows that the DTFT of y_n has the following form:

$$Y(e^{j\omega}) = \frac{1}{Tb} \sum_{k=-K}^{K} a_k \delta\left(\omega - \theta_k\right), \ \omega \in [-\pi, \pi], \tag{6}$$

where the locations of the Diracs are given by

$$\theta_k = \frac{2\pi bk}{N} \mod 2\pi, \quad k = -K, \ldots, K\,. \tag{7}$$

For simplicity, we will often use $\alpha = 2\pi b/N$, as a basic angular distance between Diracs.

To further help gain intuition, Fig. 5 visually shows an example $F(\omega)$ from (5) and the corresponding $Y(e^{j\omega})$ from (6) for different values of b. To recover b from y_n, we need to calculate the Dirac positions observed in $Y(e^{j\omega})$ and then find the angular spacing α such that $k\alpha \mod 2\pi, \quad k = -K, \ldots, K$, matches these positions.

Therefore, we propose to solve the problem with the following three steps:

1) Find the positions of the Diracs in $Y(e^{j\omega})$;
2) Recover the warping parameter b;
3) Reconstruct the coefficients of f.

The next subsection focuses solely on the second step and the uniqueness of b. The last subsection details all the steps.

Intuitively, one should be able to reconstruct the signal provided the Diracs cannot overlap – that is when $N \geq 2Kb$ – and we observe enough samples. We will show that it is indeed possible to recover the signal under this assumption, as well in other less restrictive cases as long as the set of Dirac locations is uniquely identifiable.

Before we move to the details, let us introduce some useful notation and observations. Throughout the section, we will be using the average phase of the Diracs:

$$s = \sum_{k=-K}^{K} e^{-j\theta_k} = \sum_{k=-K}^{K} e^{-jk\alpha} = D(e^{j\alpha}). \tag{8}$$

Average phase is useful, because it does not depend on the periodicity, and thus does not depend on the order of the Diracs on the circle; however, it still takes into account the mutliplicity of Dirac locations. It also has a closed form and is referred to as the *Dirichlet kernel of order* K:

$$D(e^{j\omega}) = \sum_{k=-K}^{K} e^{-jk\omega} = \frac{\sin(\frac{\omega}{2}(2K+1))}{\sin(\frac{\omega}{2})}\,. \tag{9}$$

We will also be using S_α to denote the set of Dirac positions: $S_\alpha = \{e^{jk\alpha} \,|\, k \in \{-K, \ldots, K\}\}$. Note that S_α can have less than $2K+1$ elements, if some of the Diracs overlap. If the Diracs do not overlap, the center of mass of S_α is exactly $D(e^{j\alpha})$.

A. Uniqueness of the warping

In this subsection, we formalize the notion of a unique α, and therefore a unique b. In addition, we show that, with reasonable assumptions, the number of non-unique cases is finite.

Throughout the subsection, we assume that we know the positions of all the elements of S_α. We will see when it is sufficient to reconstruct the signal.

Definition 1 (Equivalence)**.** We say that α and $\widehat{\alpha}$ are *equivalent*, if $\alpha \neq \widehat{\alpha}$ but $S_\alpha = S_{\widehat{\alpha}}$.

If α and $\widehat{\alpha}$ are equivalent, it is impossible to distinguish the parameter of the warping, and therefore (in general) impossible to recover the signal.

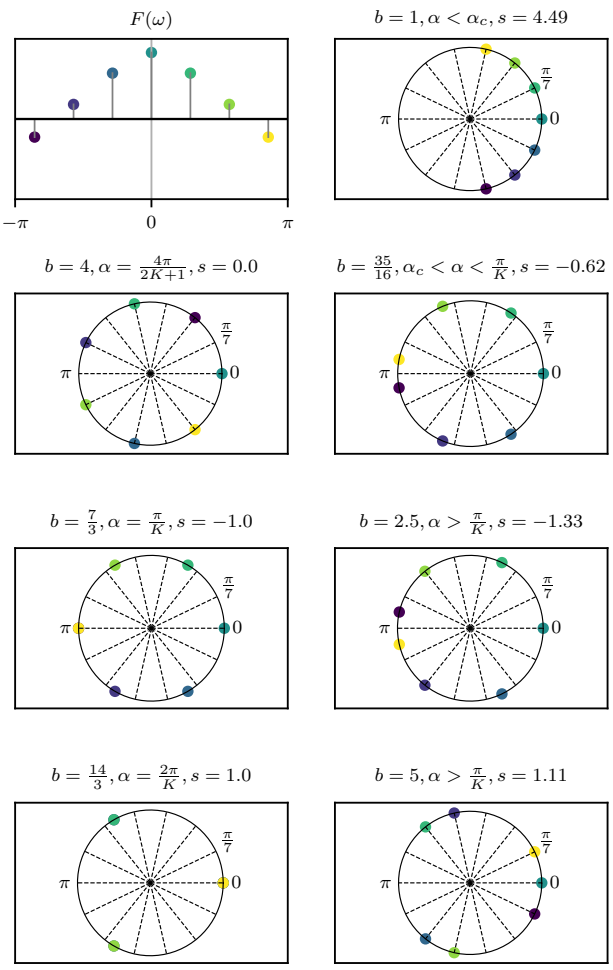


Fig. 5: An example of the function $F(\omega)$ and $Y(e^{j\omega})$'s for different values of b. Here, we define $\alpha = \frac{2\pi b}{N}$. In all the examples, we set $K = 3$, $\tau = 2K + 1$ and $N = 14$ ($T = 0.5$). α_c is defined in Theorem 5. For small enough values of b, there is no aliasing in the Fourier domain. However, for large values of b (for example $b = 5$), we have aliasing and thus retrieving the value of b is not trivial.

Note, that for any $l \in \mathbb{N}$, α and $\widehat{\alpha} = 2\pi l \pm \alpha$ are equivalent. In order to avoid this trivial equivalence, we will now assume that $\alpha \in (0, \pi)$. One could relax this assumption, but then α would only be recovered modulo π.

Lemma 2 (Characterization of non-unique warping)**.** *For fixed K, there is only a finite number of equivalent pairs $(\alpha, \widehat{\alpha})$, such that $\alpha, \widehat{\alpha} \in (0, \pi)$. Additionally, there are no equivalent pairs such that $\alpha, \widehat{\alpha} \in (0, \pi/K)$.*

Proof: We begin by proving the second statement of the lemma. Let us assume that $\alpha, \widehat{\alpha}$ are equivalent. Note that if $\alpha \in (0, \pi/K)$, then there are no overlaps between Diracs, which is equivalent to the condition $N > 2Kb$, stated earlier.

Therefore, we know that the Dirichlet kernel $D(e^{j\alpha})$ is exactly the center of mass of S_α. It follows that, since α and $\widehat{\alpha}$ are equivalent, the sets they generate are equal and their centers of mass are equal; i.e., $D(e^{j\alpha}) = D(e^{j\widehat{\alpha}})$. But on the interval $(0, \pi/K)$, the Dirichlet kernel is monotonic[3], so it can be true only when $\widehat{\alpha} = \alpha$.

For the first part of the lemma, again assume that $\alpha, \widehat{\alpha}$ are equivalent and thus $S_\alpha = S_{\widehat{\alpha}}$. It follows that $e^{j\alpha} \in S_{\widehat{\alpha}}$ and there exists $m \in \{-K, \dots, K\}$ such that $\alpha = m\widehat{\alpha} \mod \pi$. Similarly, we can write $\widehat{\alpha} = \widehat{m}\alpha \mod \pi$. This leads to $\alpha = m\widehat{m}\alpha \mod \pi$. Therefore,

$$\alpha(m\widehat{m} - 1) = \pi\ell, \tag{10}$$

for some $\ell \in \mathbb{Z}$. This can only be true in the following two cases:

1) $m\widehat{m} - 1 = 0$, or equivalently $m = \widehat{m} = 1$, since m and $\widehat{m}$ are integers. In this case, $\widehat{\alpha} = \alpha$, which is a contradiction.
2) $\alpha = \ell\pi/(m\widehat{m} - 1) = p\pi/P$, $p, P \in \mathbb{Z}$, that is, α is a rational multiple of π. Then, $\widehat{\alpha}$ is also a rational multiple of π, which is clear because $\widehat{\alpha} = \pm\widehat{m}\alpha + \pi n$, $n \in \mathbb{Z}$. What is more, $|m\widehat{m} - 1| \leq K^2 + 1$. Then, $P = |m\widehat{m} - 1|$ defines a set of possible α's: $\alpha = \pi p/P$, for $p \in \{1, P - 1\}$. Additionally, $\widehat{\alpha}$ must also be in this set. This set has $P - 1$ elements, and there can be at most $(P-1)P/2$ equivalent pairs (for a given P). Thus, the total number of pairs is bounded by

$$\sum_{P=2}^{K^2+1} \frac{(P-1)P}{2} < (K^2 + 1)^3. \tag{11}$$

This proves the first part of the lemma. ∎

From the proof of Lemma 2, we can infer the following observation.

Corollary 3. *The number of $\alpha \in (0, \pi)$ such that some of the Diracs overlap is finite.*

Indeed, if two Diracs overlap, their positions have be equal modulo π, that is $k\alpha = \widehat{k}\alpha \mod \pi$, or equivalently

$$\alpha(k - \widehat{k}) = \ell\pi,$$

for some $\ell \in \mathbb{Z}$. However, this is the same form as (10) and in the proof we have counted all α's of this form.

Finally, by substituting $\alpha = 2\pi b/N$ in Lemma 2, we observe the following.

Corollary 4. *Fix the sampling rate $T = \tau/N$. For a given (feasible) set of Diracs, there is unique $b \in (0, N/2K)$, that explains the Diracs.*

Equivalently, assuming that $b \in (0, B]$ for some $B \in \mathbb{R}_+$. If $N > 2KB$, then b can be reconstructed uniquely from the Dirac positions.

B. Unwarping bandlimited signals with DIRAChlet

In this section, we present an algorithm to find the warping parameters and the original signal from the set of samples.

[3]To see that Dirichlet kernel is monotonic, it is sufficient to check that the derivative is not zero on this interval.

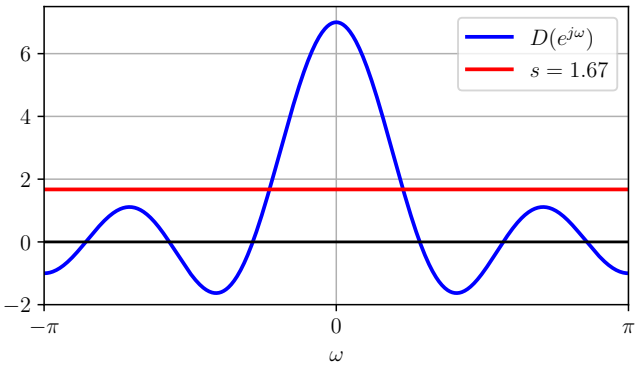


Fig. 6: Finding α using DIRAChlet.

Since the method is based on the properties of the Dirichlet kernel, we name our method DIRAChlet (see Algorithm 1).

As mentioned earlier, we propose recovery via a three step approach. The first step is to identify the locations of the Diracs on the unit circle. Since (4) is a sum of complex exponentials, line-spectral estimation methods [36] can be used to retrieve the angular frequencies (locations) of the Diracs. In particular, we propose to use Prony's method [37,38] to find the θ_k locations of the Diracs from the observations y_n. In order to recover $2K+1$ distinct locations, we need at least $2(2K+1) = 4K+2$ samples of y_n.

The second step is to calculate b from the Dirac positions. If Prony's method returns $2K+1$ locations, we know there is no overlap between Diracs and we can easily calculate the average phase s, as in (8).

From s, we can recover α (and thus b) by solving

$$D(e^{j\alpha}) = s. \tag{12}$$

As an example, Fig. 6 shows the Dirichlet kernel and the horizontal line $s \simeq 1.67$, which is the value of s resulting from $K = 3$, $b = 1.6$ and $N = 14$. To calculate α, we need to intersect the Dirichlet function with this horizontal line for which root search methods can be used (e.g. Newton's method). In Fig. 6, we can immediately see that, depending on the value of the average phase s, a different number of solutions can exists.

In the case when there are multiple solutions to (12), on the interval $(0, \pi)$, we find them all. Then, for each candidate solution $\widehat{\alpha}$, we recalculate the corresponding positions of the Diracs. If they match the positions recovered by Prony's method, we accept $\widehat{\alpha}$ as a solution and find the corresponding b.

Finally, in the third step, we use the calculated b (or b's) to solve for the amplitudes a_k by solving a linear system of equations. The complete estimation procedure is summarized in Algorithm 1.

We will see in the following theorem that the DIRAChlet algorithm exactly recovers unique solutions and, in the non-unique case, it returns all valid solutions as long as $Y(e^{j\omega})$ has

Algorithm 1 DIRAChlet algorithm

Input: $4K+2$ samples of the sequence $y_n = (f \circ \varphi)(nT)$.
Output: All possible values for b and a_k.

1: Find the position of warped Diracs, θ_k using Prony's method.
2: Calculate $s = \sum_{k=-K}^{K} e^{-j\theta_k}$.
3: Find all $\alpha \in (0, \pi)$ satisfying

$$\frac{\sin(\frac{\alpha}{2}(2K+1))}{\sin(\frac{\alpha}{2})} = s\,.$$

These are the candidate α's.
4: Prune the candidate α's by keeping only those satisfying

$$\theta_k = k\alpha \mod 2\pi, \quad k = -K, \ldots, K.$$

5: Find all the valid values of $b = N\alpha/2\pi$.
6: Solve a linear set of equations to find a_k.

$2K+1$ distinct Diracs.[4] Note that $Y(e^{j\omega})$ fails to have $2K+1$ distinct Diracs in two cases: either two or more Diracs overlap or there are some missing frequencies ($a_k = 0$ for some k).

With the described procedure in mind, we are now ready to state the main result of this section.

Theorem 5 (Recovery of the signal)**.** *Let $f(x)$ be a τ-periodic $2K+1$ bandlimited signal as defined in* (2)*. Let $T = \tau/N$ be a fixed sampling rate and let the warping function be $\varphi(x) = bx, b \in (0, N/2)$. Consider a finite sequence of $4K+2$ samples of the following form:*

$$y_n = (f \circ \varphi)(nT), \qquad n = 0, \ldots, 4K+1. \tag{13}$$

From this sequence, if $Y(e^{j\omega})$ has $2K+1$ distinct Diracs, the DIRAChlet algorithm will recover all possible values of b. Furthermore, if b is unique, the algorithm will recover the function f exactly.

Moreover, let α_c be the smallest α such that $D(e^{j\alpha})$ attains the value of its second maximum (see Fig. 7). Then, if

$$N \geq \frac{2\pi b}{\alpha_c}, \tag{14}$$

the DIRAChlet algorithm finds only one, correct, candidate solution. For smaller N, the algorithm finds multiple candidate solutions, which are later pruned.

Proof: We have outlined a method for retrieving the signal and linear warping from the set of observations. From the assumptions of the theorem, we know that we have enough samples to recover all the Diracs and that indeed $2K+1$ of them are recovered. Therefore, we can calculate s and find its intersection with the Dirichlet kernel which returns all candidate solutions of $\alpha \in (0, \pi)$. To complete the proof,

[4] We believe that the DIRACHlet algorithm can be extended to always recover all possible solutions. However, we believe that handling these corner cases would raise a large number of additional complications, which would increase complexity and greatly reduce clarity.

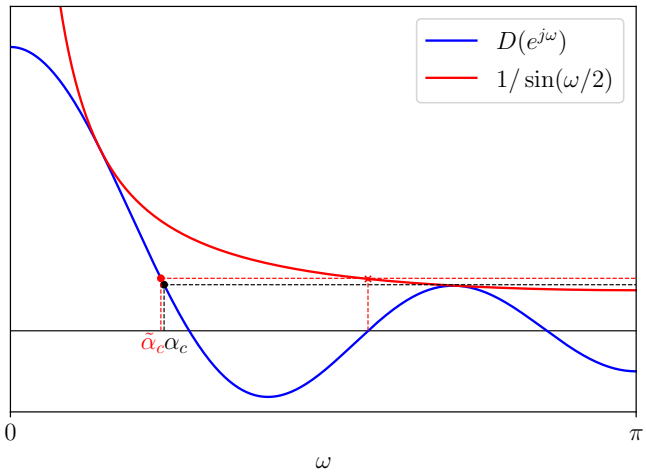


Fig. 7: The (absolute) Dirichlet kernel in blue and its envelope in red for $K = 3$. As Theorem 5 states, if the s-line is above the black dashed line, we are guaranteed to have a single candidate solution for α. In (15), we provide a closed form but looser bound shown in red.

we just need to show that below the bound $N \geq 2\pi b/\alpha_c$, there is only one candidate α.

Note that, for $\alpha \in (0, \alpha_c)$, the Dirichlet kernel D is monotonically decreasing, $D(e^{j\alpha}) > D(e^{j\alpha_c})$. In contrast, since $D(e^{j\alpha}) \leq |1/\sin(\alpha/2)|$, $D(e^{j\alpha_c}) = 1/\sin(\alpha_c/2)$ and $1/\sin(\alpha/2)$ is strictly decreasing for $\alpha \in (0, \pi]$, $D(e^{j\alpha}) \leq D(e^{j\alpha_c})$, for $\alpha \in [\alpha_c, \pi)$. Therefore, there can only be one intersection for $\alpha \in (0, \alpha_c)$ and thus the solution is unique. ■

Since there is no closed form for the locations of the extrema of the Dirichlet kernel (except at $\alpha = 0$), we can modify (14) to have a closed form bound in the expense of slightly loosening it. We can use the value of the Dirichlet kernel's envelope at the kernel's second root to derive

$$\tilde{\alpha}_c = D^{-1}\left(1/\sin\left(\frac{2\pi}{2K+1}\right)\right) < \alpha_c. \tag{15}$$

Here, $D^{-1}(\cdot)$ is the inverse of Dirichlet kernel defined just over the region $[0, 2\pi/(2K+1)]$, where it is monotonic. The value of $\tilde{\alpha}_c$ substituted in (14), provides a closed form, but slightly looser bound. The value of $\tilde{\alpha}_c$ is also shown in Fig. 7.

IV. Polynomials and Rational Functions

In this section, we consider polynomials warped such that the sampling positions are constrained by a rational function. We give a uniqueness result and propose an iterative algorithm that aims to find the unique solution.

Let f be a polynomial of degree K:

$$f(x) = \sum_{k=0}^{K} a_k x^k,$$

where the coefficients $a_k \in \mathbb{R}$ are unknown. Next, assume that we uniformly sample the composite function $h = f \circ \varphi$; i.e., our measurements are $y_n = h(nT) = f(\varphi(nT))$. As explained earlier, as well as thinking of $\varphi(x)$ warping f, we can also consider $\psi = \hat{\varphi} \circ \varphi^{-1}$ constraining the possible sampling positions.

We now prove that the polynomial exactly fitting the samples is unique, if ψ is a rational function with the degree of its denominator not smaller than the degree of its numerator. In addition, we will propose a simple iterative algorithm, which attempts to find this unique solution. The algorithm employs a simple Alternating Least Squares strategy similar to Browning [16].

The uniqueness result is the following lemma.

Lemma 6 (Pacholska *et al*, 2017)**.** *Let $\mathcal{F}$ be the space of polynomials of degree at most K. Let f be the sampled polynomial and let $0 \leq x_0 < x_1 < \cdots < x_{N-1} \leq T$ be the original sample positions. Let $\hat{x}_n$ be any other sample positions satisfying the constraint ψ; i.e. $\hat{x}_n = \psi(x_n)$. Let*

$$\psi(x_n) = \frac{p(x_n)}{q(x_n)} \quad \textit{for all } n \in [0 \dots N-1], \tag{16}$$

where p and q are irreducible polynomials with degrees satisfying: $\deg(p) \leq \deg(q)$. If the number of samples $N > K(deg(q)+1)$, then there is no polynomial $g \in \mathcal{F}$, $f \not\equiv g$ such that $f(x_n) = g(\hat{x}_n)$ for all n.

Note that this lemma does not use the ordering of the samples, and thus does not require monotonicity of the constrain function ψ. However, one could restrict the parameters of the rational function describing ψ, so that it is monotone on some interval containing $[x_0, x_{N-1}]$.

To prove the lemma, we use the fact that the polynomial g would have to have a higher degree than f in order to match $N > K(deg(q)+1)$ samples.

Proof: Let $g \in \mathcal{F}$ be a polynomial such that

$$g(\hat{x}_n) = g\left(\frac{p(x_n)}{q(x_n)}\right) = f(x_n) \quad \text{for all } n \in [0 \dots N-1],$$

and let $K_p = \deg(p)$ and $K_q = \deg(q)$. For every x_n, the following equation is satisfied:

$$\sum_{k=0}^{K} a_k x_n^k = \sum_{k=0}^{K} b_k \left(\frac{p(x_n)}{q(x_n)}\right)^k, \tag{17}$$

where a_k and b_k, $k = 1, \dots, K$ are the coefficients of the polynomials f and g, respectively. We can rewrite this as

$$(q(x_n))^K \sum_{k=0}^{K} a_k x_n^k = \sum_{k=0}^{K} b_k (p(x_n))^k (q(x_n))^{K-k}. \tag{18}$$

This equation defines a polynomial with degree at most $\kappa = \max(K_q K + K, K_p K)$. But, since $K_q \geq K_p$, $\kappa = (K_q+1)K$.

If the degree of f is not zero, the left hand side of (18) cannot be equal to the right hand side everywhere. Therefore, (17) has at most κ solutions and hence the polynomial f is unique, provided that $n > (K_q+1)K$.

If f is a constant (degree 0), it is possible that both sides of (18) are equal everywhere but this can only occur if $f \equiv g$. ■

Once the solution is unique for a certain constraint ψ, it is also unique for the corresponding warping function φ. Therefore, theoretically, a non-convex optimization method can be used to recover the sample positions and warping parameters.

To begin with, the error we can optimize is the difference between the true sample values and the re-estimated sample values. We choose the standard Mean Squared Error (MSE). In the constrained case, it has the following form:

$$C(\hat{\boldsymbol{x}}, \hat{\boldsymbol{a}}) = \|\boldsymbol{V}(\hat{\boldsymbol{x}})\hat{\boldsymbol{a}} - \boldsymbol{y}\|^2, \tag{19}$$

where $\hat{\boldsymbol{a}}$ is an estimated vector of coefficients of $\boldsymbol{f}$ and $\boldsymbol{V}$ is an interpolation matrix at points $\hat{\boldsymbol{x}} = [\hat{x}_0, \ldots, \hat{x}_{N-1}]$. In the polynomial case, $\boldsymbol{V}(\hat{\boldsymbol{x}})$ is the Vandermonde matrix consisting of the powers of $\hat{\boldsymbol{x}} = [\hat{x}_0, \ldots, \hat{x}_{N-1}]$:

$$\boldsymbol{V}(\hat{\boldsymbol{x}}) = \begin{bmatrix} | & | & & | \\ \hat{\boldsymbol{x}}^0 & \hat{\boldsymbol{x}}^1 & \ldots & \hat{\boldsymbol{x}}^K \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & \hat{x}_1 & \ldots & \hat{x}_1^K \\ 1 & \hat{x}_2 & \ldots & \hat{x}_2^K \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \hat{x}_{N-1} & \ldots & \hat{x}_{N-1}^K \end{bmatrix}.$$

For simplification from now on we shall use $\boldsymbol{V}$ for $\boldsymbol{V}(\boldsymbol{x})$.

We wish to find the sample positions $\hat{\boldsymbol{x}}$ and polynomial coefficients $\hat{\boldsymbol{a}}$ that solve the following optimization problem:

$$\hat{\boldsymbol{x}}, \hat{\boldsymbol{a}} = \arg\min_{\hat{\boldsymbol{x}}, \hat{\boldsymbol{a}}} C(\hat{\boldsymbol{x}}, \hat{\boldsymbol{a}}).$$

When the conditions of Lemma 6 are met, we have $\hat{\boldsymbol{a}} = \boldsymbol{a}$ and $\hat{x}_n = x_n$ for all $n \in [0 \ldots N-1]$.

Unfortunately, (19) is non-convex and thus the problem is difficult to solve in practice. We utilize an alternating least squares (ALS) algorithm with the following two steps:

1) Fix the matrix $\boldsymbol{V}$ and solve for the coefficients $\hat{\boldsymbol{a}}$ using ordinary least squares (OLS).
2) Fix the vector $\hat{\boldsymbol{a}}$ and make one step of gradient descent with respect to $\hat{\boldsymbol{x}}$.

The gradient step is the part of the algorithm that depends on the warping. In the general case, with no constraints but fixed $\hat{\boldsymbol{a}}$, the derivative of C in the direction x_n is

$$\frac{\partial C}{\partial x_n} = 2\Big(\big(\boldsymbol{V}\hat{\boldsymbol{a}}\big)_n - f_n\Big)\Big(\frac{\partial \boldsymbol{V}}{\partial x_n}\hat{\boldsymbol{a}}\Big)_n,$$

where $(\cdot)_n$ denotes the n-th element in the vector. Therefore, the gradient can be written as a column vector:

$$\nabla_{\hat{\boldsymbol{x}}} C = 2(\boldsymbol{V}\hat{\boldsymbol{a}} - \boldsymbol{f}) \circ (\boldsymbol{V}'\hat{\boldsymbol{a}}),$$

where $\circ$ is the entrywise (Hadamard) product, and the entries of $\boldsymbol{V}'$ are $(\boldsymbol{V}')_{n,k} = k\hat{x}_n^{k-1}$, counting from 0, so $\boldsymbol{V}'\hat{\boldsymbol{a}}$ is the derivative of the polynomial $\hat{f}$ evaluated at the points $\hat{x}_0, \ldots, \hat{x}_{N-1}$.

In order to include the warping function $\hat{x}_n = \varphi(nT, \boldsymbol{\alpha})$, we use the chain rule to replace the derivative over x_n with the derivative over $\boldsymbol{\alpha}$–the parameters of the transformation:

$$\nabla_{\boldsymbol{\alpha}} C = \boldsymbol{\Phi}'\nabla_{\boldsymbol{x}} C,$$

Algorithm 2 Alternating Least Squares Algorithm (ALS)

Input: Sampled vector $\boldsymbol{f}$, initial sample positions $\hat{\boldsymbol{x}}$
Output: Sample positions $\hat{\boldsymbol{x}}$ and polynomial coefficients $\hat{\boldsymbol{p}}$.

1: Initialize sample transformation parameters α
2: **while** not converged **do**
3: For current matrix $\boldsymbol{V} := \boldsymbol{V}$ calculate:

$$\hat{\boldsymbol{p}} := (\boldsymbol{V}^\top\boldsymbol{V})^{-1}\boldsymbol{V}^\top\boldsymbol{f}$$

4: Calculate $\boldsymbol{\Phi}'$
5: Update $\boldsymbol{\alpha}$:

$$\boldsymbol{\alpha} := \boldsymbol{\alpha} - \beta\boldsymbol{\Phi}'\hat{\boldsymbol{x}}\big((\boldsymbol{V}\boldsymbol{p} - \boldsymbol{f}) \circ (\boldsymbol{V}'\boldsymbol{p})\big)$$

6: Calculate $\hat{\boldsymbol{x}} = \boldsymbol{\varphi}(\boldsymbol{\alpha})$
7: **end while**

where $\boldsymbol{\Phi}'$ is a matrix of partial derivatives of $\varphi(Tn, \boldsymbol{\alpha})$ with respect to the parameters:

$$\big(\boldsymbol{\Phi}'\big)_{i,n} = \frac{\partial \varphi}{\partial \alpha_i}(Tn, \boldsymbol{\alpha}).$$

The matrix form of the gradient allows fast calculations. The derivative matrix has to be recomputed every time, but one expects the number of parameters to be small compared to the number of samples. The OLS part is generally the most expensive computationally and most sensitive to numerical errors.

Naturally, full specification of the (ALS) algorithm requires details of the step size and stopping criteria. This is described in Section V with reference to the specific application. A summary of the final algorithm is given in Algorithm 2.

V. Simulation results

We have presented two main scenarios for sampling at unknown locations with constrained sampling positions: a periodic bandlimited signal with a linear warping and a polynomial with sampling positions constrained by a rational function.

We now present simulation results for these two problems separately. The simulation code will be available online.

A. Periodic bandlimited signals

We start by evaluating the behavior of Algorithm 1 for unwarping periodic bandlimitted signals in the presence of noise. We set $\tau = 2K + 1$ and fixed $\boldsymbol{a} = [a_{-K}, \ldots, a_K]^T$ and b to the following values:

$$\begin{aligned} \boldsymbol{a} =& [0.43, -0.15, -0.44, 0.67, -0.32, -0.76, -0.32, \\ & 0.67, -0.44, -0.15, 0.43]^T, \\ b =& 4. \end{aligned}$$

Here, the values of $\boldsymbol{a}$ are just arbitrary random values and $K = 5$, since there are $2K + 1$ values in $\boldsymbol{a}$. As set out above, we sample $f(bx)$ with N samples per period. We study four different values of N, each corresponding to one of the regimes studied above:

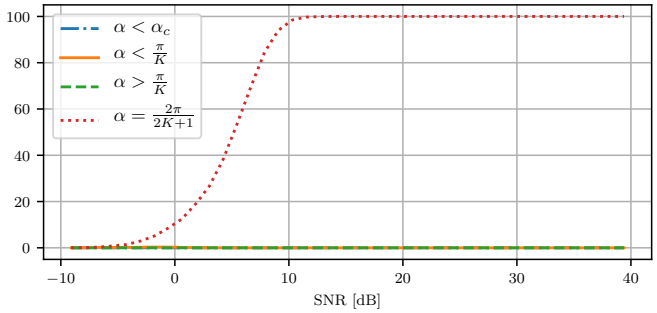


Fig. 8: Percentage of cases with more than one valid solution for b.

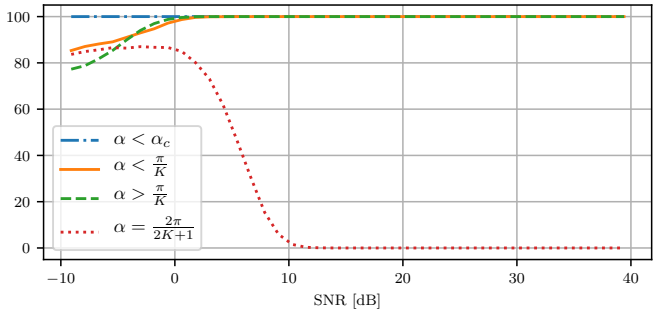


Fig. 9: Percentage of cases where Algorithm 1 results in a single solution.

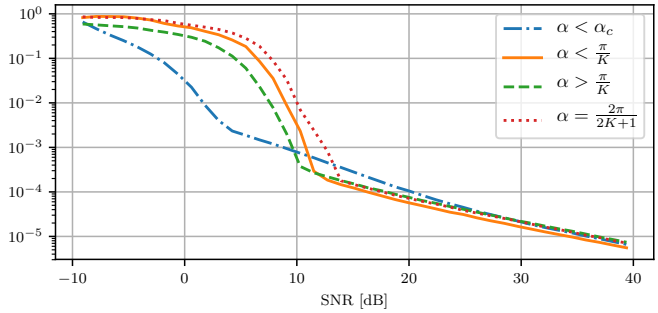


Fig. 10: Error in reconstructing b from noisy observations $\tilde{y}_n$ versus SNR. Here, the original b was set to 4.

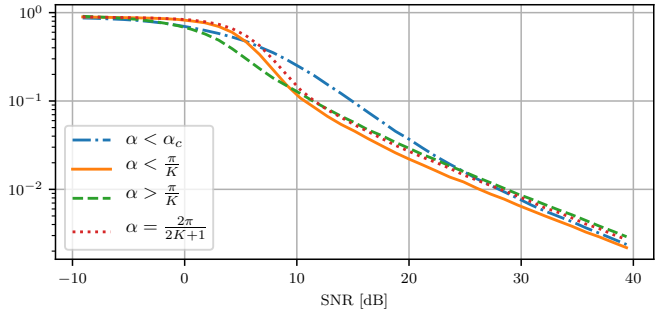


Fig. 11: Error in reconstruction of y_n from noisy observations $\hat{y}_n$ versus SNR.

Case 1: $N > 2\pi b/\alpha_c$: According to Theorem 5, this value of N guarantees that in the noiseless case the DIRACh-let algorithm will return one, correct, candidate $\alpha \in (0, \pi)$. Note that this value of N corresponds to $\alpha < \alpha_c$.

Case 2: $2Kb \leq N < 2\pi b/\alpha_c$: We choose this value of N so that the line s has several intersections with the Dirichlet kernel $D(e^{j\omega})$. In terms of α , this corresponds to $\alpha_c < \alpha \leq \pi/K$.

Case 3: $N < 2Kb$: This value of N corresponds to $\alpha > \pi/K$.

Case 4: $N = b\tau$: This case results in $s = 0$ in the noiseless case when $\tau = 2K + 1$ (our assumption in the simulations).

In each case, we fix the value of N in the interval of interest and run the simulation. Each value of N results in different samples y_n. We contaminate these samples by random Gaussian noise with zero mean and varying variance: $\tilde{y}_n = y_n + \epsilon_n$. We choose the noise such that the value of SNR ranges from -10dB to 40dB, where we define the SNR as

$$\text{SNR} = 10 \log_{10} \left(\sigma^2_{y_n} / \sigma^2_{\epsilon_n} \right) . \tag{20}$$

Here, σ_{y_n} and σ_{ϵ_n} are the empirical standard deviation of the signal y_n and noise ϵ_n, respectively. This results in the noisy observations $\tilde{y}_n$. Then, we apply Algorithm 1 on these noisy samples to estimate the values of b and a_k (and thus y_n); we call these estimated values $\hat{b}$, $\hat{a}_k$ and $\hat{y}_n$, respectively. For each value of SNR, we run $10,000$ simulations.

When running the algorithm, one of three things can happen:

1) The algorithm can return no solution, for example when the s-line does not intersect the kernel.
2) The algorithm can return multiple valid solutions (even after pruning), for example when $s = 0$.
3) The algorithm can return a single solution (after pruning).

Figure 8 depicts the percentage of cases that have multiple solutions and Fig. 9 shows the percentage of cases that have only a single solution.

In Fig. 8, we see that, except for Case 4, there is always a unique solution, for the full SNR range considered. This is expected, since the non-unique cases form a finite set for $\alpha \in (0, \pi)$. In Case 4 (s = 0), we expect to have multiple solutions for high SNR, since we have fixed α to one of the cases from this finite set. In Fig. 9, we see a similar behaviour except that, at low SNR, we fail to always return a single solution. From Fig. 8, we know that this occurs when the algorithm returns no solution.

Next, in Figs. 10 and 11, we show the error in reconstructing b and y_n, respectively. Again, we use the four cases detailed above, with the original value of b set to 4. We compute the average of the errors defined by

$$\text{error}(\hat{b}) = \frac{|\hat{b} - b|}{b} ,$$

$$\text{error}(\hat{y}_n) = \frac{\|\hat{y}_n - y_n\|_2}{\|y_n\|_2} .$$

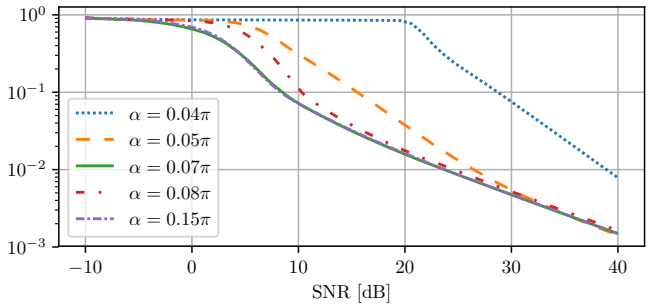


Fig. 12: Error in reconstruction of y_n from noisy observations $\hat{y}_n$ versus SNR for different values of α.

In the case that the algorithm generates multiple solutions, we choose the solution closest to the true value, when computing the above error. In Fig. 10, we observe that all cases, except Case 1, have a similar behaviour, with a break point in the curve at around 10dB. Case 1 has lower reconstruction errors for low SNR values, which is not surprising since this case has more oversampling. Also, the Dirichlet function has a small slope in this part of its curve which results in smaller variation in $\widehat{\alpha}$ with changes in s. Additionally, we observe that Case 1 does slightly worse for mid-SNR's. This can be explained by ill-conditioning, which occurs as $\alpha \to 0$. As shown in Fig. 11, although Case 1 has a lower error for estimating b in low SNR regimes, it has almost the same performance in estimating y_n. However, the ill-conditioning still makes the error slightly worse at mid-SNR's.

To further investigate this ill-conditioning as Diracs become close ($\alpha \to 0$), Fig. 12 shows a further simulation with five different values of α, all less than (or equal to) $\alpha_c = 0.15\pi$. In this simulation, $K = 5$, $\tau = 2K + 1$, $N = 30$ and $\alpha = 0.07\pi$ corresponds to the case of $b = 1$. We see that, as α decreases, the conditioning becomes worse and the performance degrades at mid to high SNR's.

B. Polynomials

To evaluate the performance of the ALS algorithm in the polynomial case (see Algorithm 2), we simulate the surface retrieval problem, introduced in Subsection II-B. We describe how to alter Algorithm 2 to retrieve the angle.

We assume a polynomial texture and linear surface, with unknown angle and offset. Recall that, assuming the pinhole camera model, the sample positions are defined by

$$\varphi(nT) = \frac{nTd}{v\cos\theta - nT\sin\theta}.$$

Note that, from Lemma 6, we know that $2K$ samples are sufficient to distinguish between different angles of the surface. On the other hand, if the angle is found, the constraint ψ becomes a linear function, and Lemma 6 does not tell us anything about the recovery of the offset d. Note that changing the distance of the surface from the camera is equivalent to scaling the polynomial. But a scaled polynomial is also a polynomial and therefore it is impossible to recover this offset. This also suggests that it is difficult to relax the assumptions of the lemma.

From the above reasoning, we know that we will be unable to recover the offset of the surface. We thus ran a number simulations with different polynomial degrees, surface orientations and noise levels. We set the irretrievable distance d and the focal length v to 1.

For each polynomial degree, surface orientation and noise level, we ran 100 experiments with arbitrary random polynomials. The polynomials were generated in the standard polynomial basis. The coefficient of the highest power was fixed to 0.5 and the remaining coefficients generated randomly from a standard normal distribution $\mathcal{N}(0, 1)$. We needed to fix the first coefficient to ensure that it is not zero. This is because, if the polynomial is similar to a polynomial of smaller degree, the model becomes too powerful with respect to the data.

If not stated otherwise, each of the 100 tests was done for 13 different angles uniformly spaced between $-20°$ and $20°$. The alternating algorithm is always initialized with $b = 0$. The initialized sample values range from -1 to 1, and therefore angles close to $45°$ cannot be recovered, because the line from the origin to the last sample would also be $45°$ and would thus never cross the surface. We are restricting the angles even further because of stability problems, see Figure 16.

We add to the sample values noise generated from the normal distribution $\mathcal{N}(0, \sigma)$, for different values of σ. The signal to noise ratio (SNR) – defined in (20) – varies between -10dB and 200dB.

We report the error in position estimates defined as

$$\text{error}(\widehat{\theta}) = |\widehat{\theta} - \theta|. \tag{21}$$

Recall also that, since the algorithm knows only the sample values, the cost function it minimizes is

$$C(\boldsymbol{x}, \boldsymbol{a}) = \|\boldsymbol{V}(\boldsymbol{x})\boldsymbol{a} - \boldsymbol{y}\|^2. \tag{22}$$

This cost function is in general not convex, see Fig. 13. This means that, without any additional modifications, the algorithm will sometimes miss the global minimum. This can be fixed by choosing a number of different starting positions, or other standard methods. In the noiseless case we know that the cost function is equal to 0 if and only if we found the global minimum. However, in the noisy case, distinguishing between local minima might be difficult[5].

In the simulations, we have seen the problems with local minima for all degrees of polynomials. In our experience, when the polynomial degree is small, local minima do not appear often. As the degree of the polynomial increases they became more common and lead to increased error even in the noiseless case, see Figs. 14 and 15.

The reconstruction is not robust to noise, see Figs 15 and 17. This is not a problem with the alternating algorithm, but with the chosen cost function. With noise, the minima of the cost function flatten out, because perfect fitting of the polynomial to

[5]Note that, for this 1D problem, it is easy to come up with a more robust scheme, such as a simple grid-search. However, we have chosen the ALS algorithm, because it generalizes easily to higher dimensions, and is therefore more illustrative.

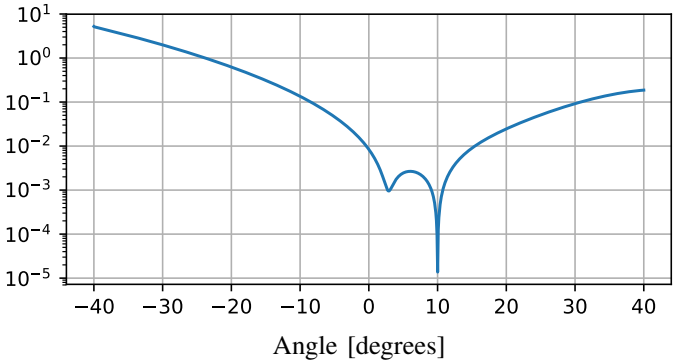


Fig. 13: The cost function is not convex. The plot shows the cost function for different estimated errors, for a fixed polynomial of degree 9, with no noise added. The true angle is 10 degrees. Since the algorithm is initialized at 0, it will stop at the local minimum.

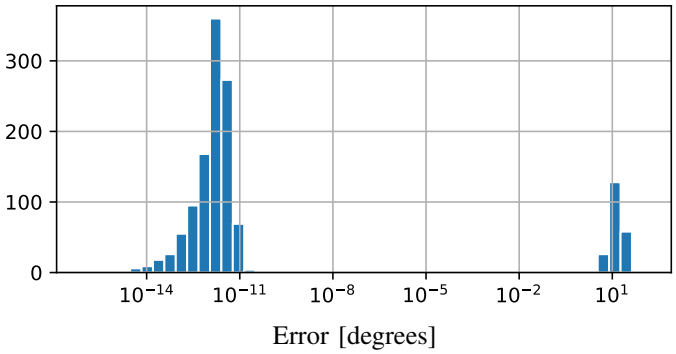


Fig. 14: Histogram of the errors for a polynomial of degree 4, with no additional noise added. The distribution of the error is clearly bi-modal. In more than 80% of cases the error is smaller than 10^{-10}, yet the mean is around 2.5.

the samples is no longer possible. This is also the reason why oversampling does not give a big improvement — although the oversampling reduces the relative power of the noise, it does not prevent minima from flattening out.

For angles inside the interval $[-20°, 20°]$, the error does not vary significantly, see Fig. 16. Outside this interval the algorithm becomes unstable. This is due to the geometry of the problem and the fact that a small change in θ leads to a big change of the estimated sample positions and therefore a big change in the estimated coefficients of the polynomial. One can imagine that a change of variables or introducing a varying step size depending on the current angle could widen the stable region.

Finally, we needed to adjust the step size and the stopping criteria of Algorithm 2. We chose step size β to be inversely proportional to the oversampling factor, in order to prevent too large gradients. Large gradients can cause the algorithm to move to angles θ outside the allowed $(-45°, 45°)$ interval. Therefore, we multiplicatively decrease β every time θ would become too extreme. We use different stopping criteria: when

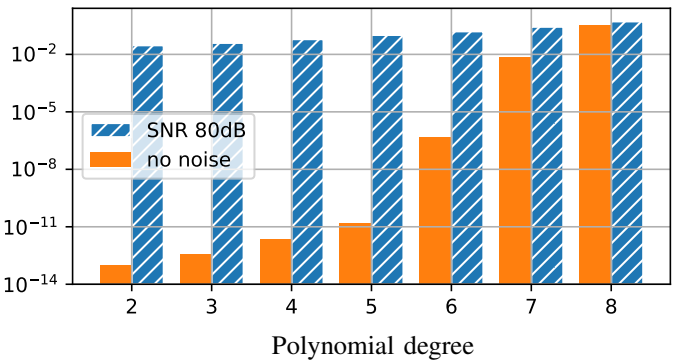


Fig. 15: Results of the ALS algorithm for different polynomial degrees. The median error with no noise is shown in orange (solid) and the median error with a small amount of additive noise (SNR 80dB) is shown in blue (hatched). As one can see, in the noiseless case, the algorithm breaks down at around degree 6. In the noisy case, the errors are much bigger even when the algorithm finds the global minimum. This makes the error less dependent on the number of local minima, and thus less dependent on the degree of the polynomial.

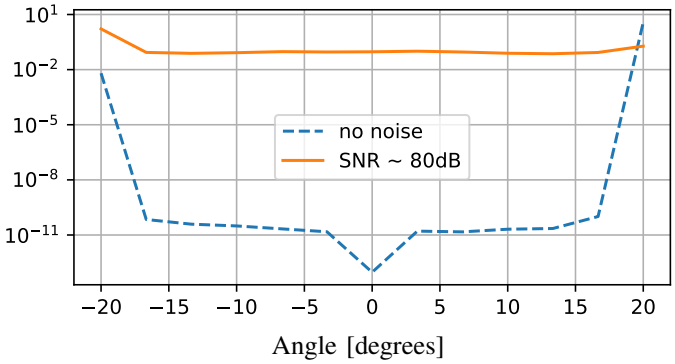


Fig. 16: Median error for different angles, aggregated for 7 different polynomial degrees (2,...,8). The blue dashed line (left scale) shows the noiseless case, and orange (solid) line shows an SNR of around 80dB. The decrease of error around 0 for the noiseless case is due to algorithm initialization at 0.

the cost function is small enough, when the cost function stops changing and after a certain number of iterations. We noticed that increasing the number of iterations does not improve the results, and limiting the number of iterations might be seen as a version of early stopping.

VI. Conclusion

We have proposed the problem of uniformly sampling a composite of functions as a regularizer for sampling at unknown locations. As we have shown, this formulation maintains many of the key aspects of practical problems such as simultaneous localization and mapping (SLAM) and structure from motion (SfM).

In addition, we have studied two simple examples and demonstrated uniqueness in both cases. In one case, we have

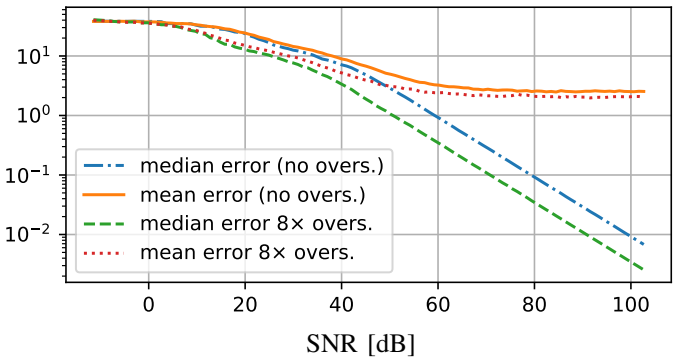


Fig. 17: Error for different signal to noise ratios, for polynomials of degree 4. Around 40dB SNR, the mean error and median error begin to differ. Unsurprisingly, the mean error flattens when the few large errors dominate the mean. Oversampling improves the results but not significantly. Oversampling 8 times gives error equivalent to no oversampling with SNR 10dB bigger, but this results in the small difference in error.

also provided an efficient algorithm that reaches this unique solution and, in the other, we saw that the problem was not stable in the presence of noise. An interesting topic for future work is to attempt to improve stability by regularizing the parameter values.

Moreover, we believe that there are many additional examples of sampling a composite of functions that can be solved. For example, a promising research direction is probabilistic warping functions, which would further connect the excellent work of Kumar [17,18].

In terms of the connection to SLAM and SfM, much work needs to be done to create practical algorithms from this type of approach. However, simple extensions such as moving to piecewise linear surfaces would already make a step in this direction. In addition, ideas from single index models may bridge the gap to higher dimensions.

Finally, we believe that it is important to understand the fundamental limit of such problems and our analysis contributes to this understanding.

References

[1] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. Cambridge, Mass.: MIT Press, 2005.

[2] M. Montemerlo and S. Thrun, *FastSLAM: A Scalable Method for the Simultaneous Localization and Mapping Problem in Robotics*, 1st ed. Springer Publishing Company, Incorporated, 2010.

[3] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[4] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*. SpringerVerlag, 2003.

[5] H. Nyquist, "Certain topics in telegraph transmission theory," *American Institute of Electrical Engineers, Trans. of the*, vol. 47, no. 2, pp. 617–644, Apr. 1928.

[6] C. E. Shannon, "Communication in the presence of noise," *Proc. Institute of Radio Engineers*, vol. 37, no. 1, pp. 10–21, 1949.

[7] E. T. Whittaker, "On the functions which are represented by the expansions of the interpolation-theory," *Proc. of the Royal Society of Edinburgh*, vol. 35, p. 181194, 1915.

[8] A. J. Jerri, "The Shannon sampling theorem - its various extensions and applications: A tutorial review," *Proc. of the IEEE*, vol. 65, no. 11, pp. 1565–1596, Jun. 1977.

[9] M. Unser, "Sampling-50 years after Shannon," *Proc. of the IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

[10] M. Vetterli, J. Kovacevic, and V. K. Goyal, *Foundations of Signal Processing*. Cambridge: Cambridge Univ. Press, 2014.

[11] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. on Signal Processing*, vol. 50, pp. 1417–1428, Aug. 2002.

[12] Y. Lu and M. N. Do, "A theory for sampling signals from a union of subspaces," *IEEE Trans. on Signal Processing*, vol. 56, no. 6, p. 2334–2345, Jun. 2008.

[13] H. Feichtinger and K. Gröchenig, "Theory and practice of irregular sampling," in *Wavelets: Mathematics and Applications*, 01 1994.

[14] A. Aldroubi and K. Gröchenig, "Nonuniform sampling and reconstruction in shift-invariant spaces," *SIAM Rev.*, vol. 43, no. 4, pp. 585–620, Apr. 2001.

[15] M. Pacholska, B. Béjar Haro, A. Scholefield, and M. Vetterli, "Sampling at unknown locations, with an application in surface retrieval," in *Proc. of the 12th Int. Conf. on Sampling Theory and its Applications (SampTA)*, jul 2017, pp. 364–368.

[16] J. Browning, "Approximating signals from nonuniform continuous time samples at unknown locations," *IEEE Trans. on Signal Processing*, vol. 55, no. 4, pp. 1549–1554, April 2007.

[17] A. Kumar, "On bandlimited signal reconstruction from the distribution of unknown sampling locations," *IEEE Trans. on Signal Processing*, vol. 63, no. 5, pp. 1259–1267, March 2015.

[18] ——, "On bandlimited field estimation from samples recorded by a location-unaware mobile sensor," *IEEE Trans. on Information Theory*, vol. 63, no. 4, pp. 2188–2200, April 2017.

[19] J. L. Horowitz, *Semiparametric and Nonparametric Methods in Econometrics*. Springer US, 2009.

[20] A. K. Han, "Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator," *Journal of Econometrics*, vol. 35, no. 2, pp. 303 – 316, 1987.

[21] Y. Plan and R. Vershynin, "The generalized lasso with non-linear observations," *IEEE Trans. Information Theory*, vol. 62, no. 3, pp. 1528–1537, 2016.

[22] Z. Yang, K. Balasubramanian, and H. Liu, "High-dimensional non-Gaussian single index models via thresholded score function estimation," in *Proc. of Machine Learning Research (PMLR)*, vol. 70, Aug 2017, pp. 3851–3860.

[23] P. Marziliano and M. Vetterli, "Reconstruction of irregularly sampled discrete-time bandlimited signals with unknown sampling locations," *IEEE Trans. on Signal Processing*, vol. 48, no. 12, pp. 3462–3471, Dec 2000.

[24] J. Unnikrishnan, S. Haghighatshoar, and M. Vetterli, "Unlabeled sensing with random linear measurements," *IEEE Trans. on Information Theory*, vol. 64, no. 5, pp. 3237–3253, May 2018.

[25] G. Elhami, A. J. Scholefield, B. Bejar Haro, and M. Vetterli, "Unlabeled sensing: Reconstruction algorithm and theoretical guarantees," *Proc. of the Int. Conf. on Acoustics Speech and Signal Processing (ICASSP)*, p. 5, 2017.

[26] A. Pananjady, M. J. Wainwright, and T. A. Courtade, "Linear regression with shuffled data: Statistical and computational limits of permutation recovery," *IEEE Trans. on Information Theory*, pp. 1–15, 2017.

[27] D. J. Hsu, K. Shi, and X. Sun, "Linear regression without correspondence," in *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., 2017, pp. 1531–1540.

[28] S. Haghighatshoar and G. Caire, "Signal recovery from unlabeled samples," *IEEE Trans. on Signal Processing*, pp. 1242 – 1257, 03 2018.

[29] S. Azizi, D. Cochran, and J. McDonald, "On the preservation of bandlimitedness under non-affine time warping," in *Proc. Int. Workshop on Sampling Theory and Applications*, 1999.

[30] J. Clark, M. Palmer, and P. Lawrence, "A transformation method for the reconstruction of functions from nonuniformly spaced samples," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 33, no. 5, pp. 1151–1165, 1985.

[31] K. Horiuchi, "Sampling principle for continuous signals with time-varying bands," *Information and Control*, vol. 13, no. 1, pp. 53–61, 1968.

[32] D. Cochran and J. J. Clark, "On the sampling and reconstruction of time warped band-limited signals," in *Proc. of the Int. Conf. on Acoustics Speech and Signal Processing (ICASSP)*, April 1990.

[33] X.-G. Xia and Z. Zhang, "On a conjecture on time-warped band-limited signals," *IEEE Trans. on Signal Processing*, vol. 40, no. 1, pp. 252–254, 1992.

[34] A. Scholefield, B. Béjar Haro, and M. Vetterli, "Shape from bandwidth: the 2-D orthogonal projection case," in *Proc. of the Int. Conf. on Acoustics Speech and Signal Processing (ICASSP)*, 2017.

[35] M. Clerc and S. Mallat, "The texture gradient equation for recovering shape from texture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 536–549, 2002.

[36] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice Hall, 1997.

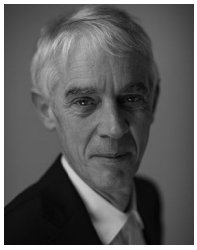
[37] G. Prony, "Essai experimental et analytic," *J. d'Ecole Polytech. (Paris)*, vol. 1, pp. 24–76, 1795.

[38] J. F. Hauer, C. J. Demeure, and L. L. Scharf, "Initial results in Prony analysis of power system response signals," *IEEE Trans. on Power Systems*, vol. 5, no. 1, pp. 80–89, 1990.

Golnoosh Elhami started her Ph.D. studies in the Audiovisual Communications Laboratory at the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland in 2015. She received BS in Electrical and Telecommunications Engineering from University of Tehran, in 2013, and MS in communication systems, from Sharif University of Technology, in 2015. Her research interests include sampling theory, reconstruction algorithms, computer vision, machine learning and localization.

Michalina Pacholska started her Ph.D. studies in the Audiovisual Communications Laboratory at the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland in 2016. She finished Inter-faculty Studies in Mathematics and Natural Sciences obtaining BS in Mathematics in 2014, and received MS in Mathematics in 2016, both from the University of Warsaw. Her research interests include sampling theory, computational imaging, Fourier optics and localization.

Benjamín Béjar received the Electrical Engineering degree from both the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, and the Technische Universität Darmstadt (TUD), Darmstadt, Germany, in 2006 under the framework of the Double Degree Exchange program. He received the Ph.D. degree in Electrical Engineering from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2012. From Sept. 2011 to Sept. 2013 he was a member of the Vision, Dynamics and Learning lab at the Johns Hopkins University (JHU), Baltimore, MD, USA, as part of the MSE program in Biomedical Engineering. He has held research appointments as a visiting Ph.D. student at the Universit degli Studi di Udine, Udine, Italy, in 2009 and, at the Hong Kong University of Science and Technology (HKUST), Hong Kong, China, in 2011. In 2012, he was awarded with the 2012 Best Paper Award in Medical Robotics and Computer Assisted Intervention Systems from the Medical Image Computing and Computer Assisted Intervention Society (MICCAI). He served as a postdoctoral researcher and lecturer in the Audiovisual Communications Laboratory (LCAV) at EPFL from Oct. 2013 to Oct. 2017. Since Oct. 2017 he is with The Department of Biomedical Engineering at The Johns Hopkins University where he currently holds an appointment as an Assistant Research Professor. His research interest include digital signal processing, sampling and reconstruction of sparse signals, image processing, convex optimization, machine learning, and inverse problems in biomedical imaging.



Martin Vetterli received the EE degree from ETHZ in 1981, the MS from Stanford University in 1982, and the Doctorate from EPFL in 1986. He was an Assistant/Associate Professor in EE at Columbia University, and then an Associate/Full Professor in EECS at the UC Berkeley. In 1995, he joined the EPFL as a Full Professor. From 2004 to 2011 he was VP of EPFL for international affairs, and from 2011 to 2012, he was the Dean of IC. From 2013 to 2016 he was President of the Swiss NSF and since January 2017, he is President of EPFL. He works in the areas of electrical engineering, computer sciences and applied mathematics and this led to about 170 journals papers, as well as about 30 patents that led to technology transfer to high-tech companies and the creation of several start-ups. He is also the co-author of three textbooks. His prizes include best paper awards from EURASIP in 1984 and of the IEEE Signal Processing Society in 1991, 1996 and 2006, the Swiss National Latsis Prize in 1996, the SPIE Presidential award in 1999, the IEEE SP Technical Achievement Award in 2001 and the IEEE SP Society Award in 2010 and the IEEE Jack S. Kilby Signal Processing Medal in 2017. He is a Fellow of IEEE, of ACM, ISI highly cited researcher and a foreign member of the NAE.

Adam Scholefield is currently a post-doctoral researcher in the Audiovisual Communications Laboratory at the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. He received the MEng. degree in Electrical and Electronic Engineering from Imperial College London, in 2007 and, after a brief interruption of studies to represent Team GB in the 2012 Olympics, the Ph.D. degree, from the same university, in 2013. His research interests include computational imaging, computer vision, localization, optics and sampling theory.