Sampling at unknown locations: Uniqueness 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. We show that, without further constraints, the problem is often hopeless. More precisely, we show that, for polynomial and bandlimited signals, it is possible to find two signals, arbitrarily far from each other, that fit the measurements.**

To overcome this, we propose to add constraints to the sample positions. As we show, this leads to a uniform sampling of a composite of functions. Our 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 prove 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, simulation results are provided to support our theoretical analysis.

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

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 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 doesn't 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

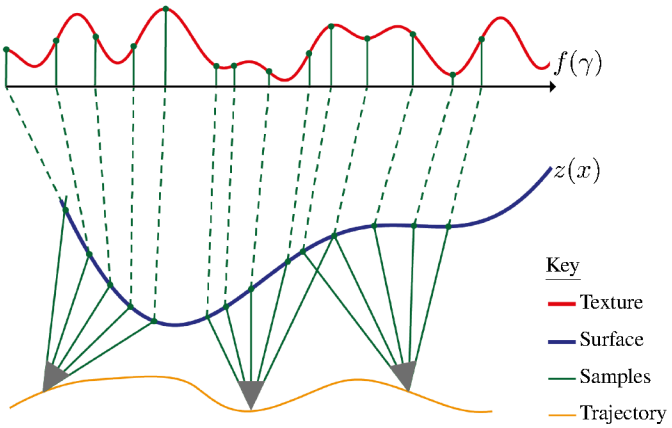


Fig. 1: The connection between sampling at unknown locations and SLAM/SfM. Here a camera moves along a trajectory and takes images of a surface. The surface is painted with a texture, which the camera measures. The locations of its measurements is dictated by the surface geometry, trajectory and camera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

†The 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. In 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 unicity 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 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].

In the discrete case, Marziliano et. al. investigated the recovery of bandlimited signals [19] and there is a connection to the recent work on unlabelled sensing [20]–[24].

Finally, since we consider a composite of functions, there is a connection to previous works on sampling time-warped signals [25]–[29]. In fact, in [30], we used a result from [29] 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 [31].

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.

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 regularized 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 $\boldsymbol{x} = [x_0, \ldots, x_{N-1}]$ with $x_n \in X$, $n = 0, \ldots, N-1$. Assume we observe $f \in \mathcal{F}$ at N unknown and distinct locations over the interval; that is, we measure $\boldsymbol{y} = \mathrm{T}_{\boldsymbol{x}}\, f = \{f(x_0), \ldots, f(x_{N-1})\}$, 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 $\mathcal{F}$.

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

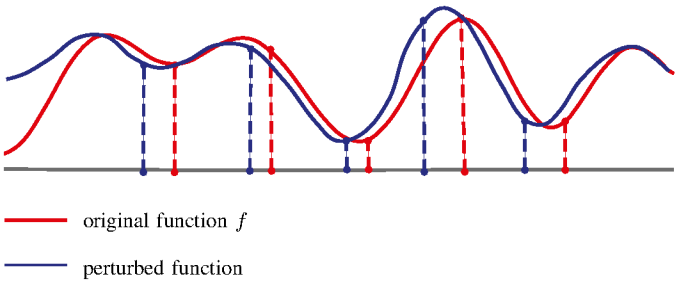


Fig. 2: If we sample function f from a linear space of continuous functions $\mathcal{F}$, we can add to it some small perturbation. If the perturbation is small enough the samples can be moved to match the perturbed function.

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 $\{\tilde{x}_1, \dots, \tilde{x}_m\}$ such that $\tilde{x}_n \in X$ and $\tilde{f}(\tilde{x}_n) = f(x_n)$ for all $i = 0, \dots, N-1$. 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 the L_2-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]. ■

Note that, in Lemma 1, the first and last samples are fixed ($\tilde{x}_0 = x_0$ and $\tilde{x}_{N-1} = x_{N-1}$). This is not necessary but, 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); however, 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]).

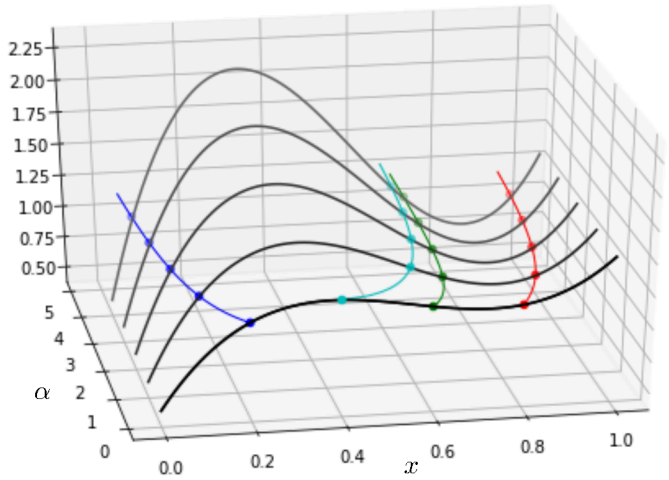


Fig. 3: Illustration of the path of solutions generated by moving along $g \in \mathcal{F}$ from the initial f. The path of solutions ($\tilde{f}_\alpha = f + \alpha g$) defines a trajectory along which samples can move.

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 allowing 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 the way 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 $\boldsymbol{h} = \mathrm{T}_{\{nT : n \in [0..N-1]\}}(f \circ \varphi)$. 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, lets 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

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

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

$$\frac{x_i \cos\theta}{x_i \sin\theta + d} = \frac{iT}{v} \quad \Rightarrow \quad x_i = \frac{iTd}{v\cos\theta - iT\sin\theta} = \varphi(iT);$$

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

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

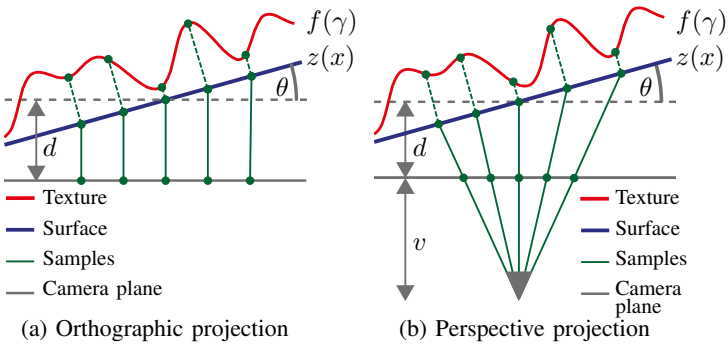
Let d and θ be the parameters of the true surface and $\tilde{d}$ and $\tilde{\theta}$ be the parameters of any other surface. Then, since

$$\frac{x_i \cos\theta}{x_i \sin\theta + d} = \frac{\tilde{x}_i \cos\tilde{\theta}}{\tilde{x}_i \sin\tilde{\theta} + \tilde{d}},$$

we can find the constraint function from

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

[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.

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. However, the framework we are proposing is very general and incorporates much of the existing works on sampling at unknown locations. For example, if Φ is 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 an affine 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

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 $f(x)$. Instead of $f(x)$ we observe samples of a warped

version of it. Let $h(x)$ denote the warped signal given by

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

where $\varphi(x) = bx + c$, with $b \neq 0, c \in \mathbb{R}$ is an affine transformation of the domain of $f(x)$. The signal $h(x)$ is sampled uniformly with a sampling rate of $T = \tau/N$ (e.g. N samples per period) producing the sequence $h_n = h(nT)$:

$$h_n = \sum_{k=-K}^{K} \underbrace{a_k \, e^{j2\pi bkc/\tau}}_{\widehat{a}_k} \, e^{j2\pi bkn/N}. \quad (4)$$

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, for reconstructing $f(x)$, and b, c to reconstruct $\varphi(x)$.

If we assume that $a_k \in \mathbb{R}$, it is easy to recover a_k and c from $\widehat{a}_k$: simple set $a_k = |\widehat{a}_k|$ and $c = \tau \angle \widehat{a}_k / 2\pi bk$. In the case of complex a_k's, there are methods such as Prony's method which enables us to retrieve the complex a_k and c from $\widehat{a}_k$.

We see immediately that we can only recover $\varphi(x)$ up to some trivial ambiguities. For instance, both $\varphi(x)$ and $\varphi(-x)$ would produce the same set of samples h_n.

Consider the case of $b = 1$ (the signal is just shifted by c). In that case, we can recover the FS coefficients of the signal $\widehat{a}_k$ provided $N \geq 2K + 1$ and we observe at least one period. In fact, we can recover $\widehat{a}_k$ from the DFT of h_n, $n = 0, \ldots, N-1$. From $\widehat{a}_k$, we can estimate the offset c by Prony's method [32], provided $c \leq \tau$.

Let us now look at the more interesting case of $b \neq 1$. Intuitively, one should be able to reconstruct the signal provided $N \geq b(2K+1)$ and we observe enough samples. We will show that it is indeed possible to recover the signal in most of the cases even when aliasing is present. The intuition behind the method is based on the observation that in the Fourier domain, the signal $f(x)$ corresponds to a set of Diracs uniformly spaced in frequency (see Fig. 5). By introducing a linear warping, we are effectively changing the spacing between those Diracs while still preserving the uniform (modulo 2π as a result of sampling) spacing structure. Depending on the warping we might have aliasing (e.g. Diracs warp around in the unit circle) but even in those cases we might be able to recover the signal. Of course, the first step for recovery is to identify the locations of the spikes in the unit circle. Note that (4) is a sum of complex exponentials and that line-spectral estimation methods [33] can be used to retrieve the angular frequencies (locations) of the spikes. From this discussion, it is easy to realize that a necessary condition for perfect recovery requires the set of spike locations to be uniquely specified. In other words, no two spikes can lie on top of each other as a result of the warping.

We are now ready to state the main result.

Theorem 2. *Let $f(x)$ be a periodic $(2K+1)/\tau$ bandlimited signal as in* (2)*. Let the affine warping function be $\varphi(x) = bx + c$, with $\{c, b \mid b/c \in [-1, 1], b \in \mathbb{R} - \{0\}\}$. Consider a finite sequence of $4K + 2$ samples of the following form:*

$$h_n = (f \circ \varphi)(nT), \qquad n = 0, \ldots, 4K + 1, \quad (5)$$

where $T = \tau/N$ is the sampling rate. Then, we can uniquely reconstruct $f(x)$ and find c and b, provided

$$N \geq \frac{2\pi b}{Y^{-1}\left(\sin\left(\frac{3\pi}{2(2K+1)}\right)^{-1}\right)}, \quad (6)$$

where $Y^{-1}(\cdot)$ is the inverse of the Dirichlet kernel as per (12)*. Furthermore, for smaller values of N if $b \in (0, N/2]$, both the set of non-recoverable cases and the set of non-unique cases have measure zero.*

In order to prove Theorem 2, we start by introducing an algorithm to unwarp $f(x)$ from samples h_n.

A. Unwarping bandlimited signals with DIRACHlet

As it has been already pointed out the Fourier transform of $f(x)$ corresponds to a set of uniformly spaced Diracs:

$$F(\omega) = \mathcal{F}\{f(x)\} = \sum_{k=-K}^{K} a_k \delta\left(\omega - \frac{2\pi k}{\tau}\right). \quad (7)$$

Because of (7), the the DTFT of h_n has the following form:

$$H(e^{j\omega}) = \frac{1}{Tb} \sum_{k=-K}^{K} \widehat{a}_k \delta\left(\omega - \frac{2\pi bk}{N}\right). \quad (8)$$

Figure 5 shows an example $F(\omega)$ and several examples of $H(e^{j\omega})$ for different values of b while keeping $c = 0$. As can be seen from the figure, $F(\omega)$ consists of $2K+1$ Diracs equally spaced on the interval $[-\pi, \pi]$. Similarly, $H(e^{j\omega})$ consists of $2K + 1$ Diracs in $[-\pi, \pi]$ but now at locations

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

In order to retrieve the warping parameters and the original signal we proceed as follows:

1) We use Prony's method [32,34] to find the locations of the delta Diracs θ_k from the observations h_n. In order to recover the $2K + 1$ locations, we need at least $2(2K + 1) = 4K + 2$ samples of h_n.
2) Then, to remove the effect of periodisation, we calculate the average phase:
$$s = \sum_{k=-K}^{K} e^{-j\theta_k}. \quad (9)$$
3) Next, find a solution for
$$s = \sum_{k=-K}^{K} e^{-jk\alpha} = Y(e^{j\alpha})\,, \quad (10)$$
where $\alpha = 2\pi b/N$. Note that $Y(e^{j\omega}) = \sum_{k=-K}^{K} e^{-jk\omega}$ can be seen as the DTFT of the following sequence:
$$y_n = \begin{cases} 1, & n = -K, \ldots, K \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

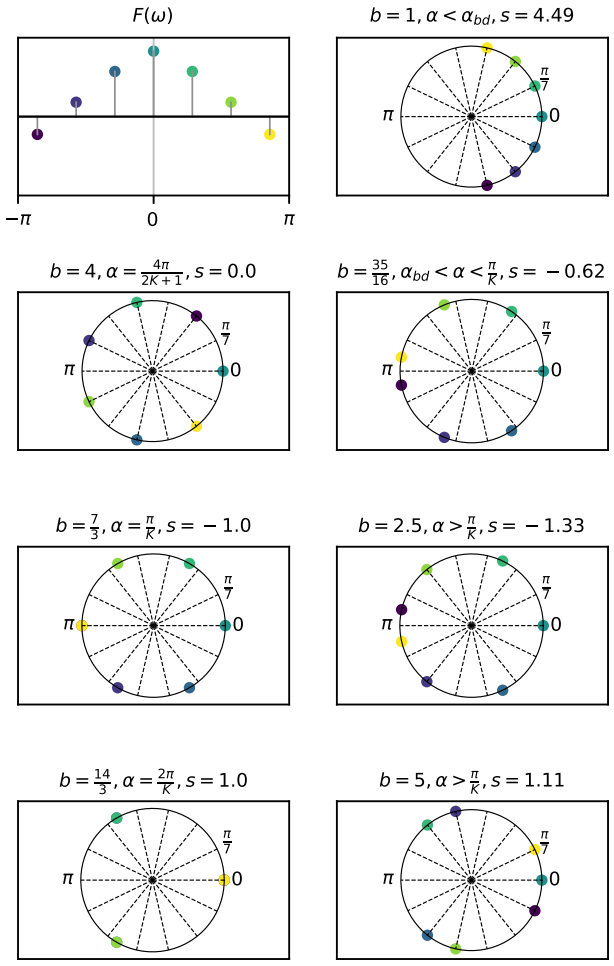


Fig. 5: An example of the function $F(\omega)$ and $H(e^{j\omega})$'s for different values of b. Here, we define $\alpha = \frac{2\pi b}{N}$. In all the examples, we set $K = 3$, $\tau = 2K+1$ and $N = 14$ ($T = 0.5$). In the titles we use $\alpha_{\text{bd}} = Y^{-1}(\sin(\frac{3\pi}{2(2K+1)})^{-1})$. 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.

The DTFT has a closed form and is referred to as the *Dirichlet kernel of order* K:

$$Y(e^{j\omega}) = \frac{\sin(\frac{\omega}{2}(2K+1))}{\sin(\frac{\omega}{2})}. \tag{12}$$

Therefore, we need to find the values of α such that

$$s = \frac{\sin(\frac{\alpha}{2}(2K+1))}{\sin(\frac{\alpha}{2})}. \tag{13}$$

Depending on s, (13) might have a different number of possible solutions, with the maximum being $2K$.

4) By just using the average phase to estimate α, we lose some information. The price is that we can recover some incorrect α's that do not warp the Diracs to the positions observed at the output of Prony's method. However, these are easy to detect and remove. The remaining α's are all valid. In the following section, we will explain uniqueness with respect to the number of possible valid α's.
5) Once we have estimated the valid α's, we can find the corresponding b's from $\hat{b} = N\hat{\alpha}/2\pi$.
We can use then use these estimates, to find the values of $\widehat{a}_k$, using a simple linear equation. Estimating c from $\widehat{a}_k$, is then straightforward, provided that $c \leq \tau$.

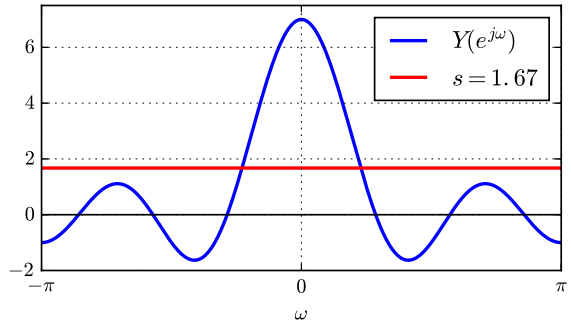


Fig. 6: Finding α using DIRAChlet.

Algorithm 1 DIRAChlet algorithm

Input: $2(2K+1)$ samples of sequence $h_n = h(nT)$.
Output: All possible values for b, c 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 values of α satisfying

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

4: Find the position of Diracs corresponding to the values of α from previous step.
5: Keep the valid α's that correctly warp the Diracs back to θ_k.
6: Find all the valid values of $b = \frac{N\alpha}{2\pi}$.
7: Solve a linear set of equations to find a_k and c.

The above procedure is summarized in Algorithm 1.

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. In our implementations, we find the intersection points using bisection methods.

Equivalence classes: Note that in (3), changing the value of b to $-b$ does not change $h(x)$, i.e., $f(c+bx) = f(c-bx)$. Therefore, both b and $-b$ are valid solutions. We observe this symmetry in Fig. 6, as the s-line intersects the Dirichlet kernel both at α and $-\alpha$. This shows that the unwarping problem with this formulation of $f(x)$ has symmetric solutions for b. We will

assume that either one of these solutions is valid.

Furthermore, if we replace b by $b+N\ell$ in (4), where $\ell \in \mathbb{Z}$, the expression for h_n does not change. This translates into changing α to $\alpha + 2\ell\pi$ in (10). Since the Dirichlet kernel is 2π-periodic, all the values $b + N\ell$ for $\ell \in \mathbb{Z}$ are also valid solutions. This results in another set of trivial solutions. Note that replacing b in (4) with $b+N\ell$ does not result in the same expression for $h(x)$ but we are given the sampled signal h_n and not $h(x)$. Thus, these values of b are not distinguishable given only the sampled sequence h_n.

In what follows, we will look for uniqueness beyond these trivial equivalences. To do this, we consider α in the interval $(0, \pi]$. Then, given a valid solution for α in this range, we can always find the trivial equivalent solutions corresponding to $\alpha + 2\pi\ell$ and $-\alpha + 2\pi\ell$ for $\ell \in \mathbb{Z}$.

In Theorem 2, we saw that if N is larger than a certain value, we can uniquely reconstruct the values of b up to the above equivalencies. Furthermore, we stated that for smaller values of N, we can uniquely reconstruct the functions almost surely. Now that we have introduced the required tools, we move to prove Theorem 2.

B. *Proof of Theorem 2*

First we will show that if (6) holds, then we always end up with the unique solution. The Dirichlet kernel lies between two envelopes $\pm 1/\sin(\omega/2)$. It is straightforward to see that the two curves and the kernel are tangent to each other when $\sin(\omega(2K+1)/2) = \pm 1$. Thus, the kernel is tangent to one or the other of these two curves at points $\omega_{t_n} = (2n+1)\pi/(2K+1)$. Particularly, the first and second tangent points are at $\omega_{t_1} = \pi/(2K+1)$ and $\omega_{t_2} = 3\pi/(2K+1)$. Let denote ω_{m_2} and ω_{m_3} as the second and third optima of the kernel, respectively (the first optimum happens at $\omega_{m_1} = 0$). In Figure 7, we can see the Dirichlet kernel of order 3, its absolute value, the envelope $\xi(\omega) = 1/\sin(\omega/2)$, as well as the tangent points and local optima. If the value of s was larger than $Y(e^{\omega_{m_3}})$, we could recover b uniquely. Additionally, as can be seen in Figure 7, ω_{m_3} is less than the first zero of the Dirichlet kernel; i.e., $\omega_{m_3} < 2\pi/(2K+1)$. Therefore, there is no overlap of Diracs and the value of s is known. Since there is no closed form solution for finding the value of ω_{m_3}, we will set the threshold to $|Y(e^{j\omega_{t_2}})|$, which is the value of the kernel at the second tangent point. Then, we only need to show that

$$|Y(e^{j\omega_{t_2}})| > Y(e^{j\omega_{m_3}})\,. \tag{14}$$

First, note that the envelope $\xi(\omega)$ is strictly decreasing in the interval $[0, \pi]$. Indeed,

$$\frac{\mathrm{d}\xi(\omega)}{\mathrm{d}\omega} = \frac{\frac{1}{2}\cos(\frac{\omega}{2})}{\sin^2(\frac{\omega}{2})} < 0, \quad \text{for } 0 < \omega \leq \pi\,.$$

We prove (14) by contradiction; Suppose that $|Y(e^{\omega_{t_2}})| \leq Y(e^{j\omega_{m_3}})$. Then,

$$\xi(\omega_{t_2}) \overset{(a)}{=} |Y(e^{j\omega_{t_2}})| \overset{(b)}{\leq} Y(e^{j\omega_{m_3}}) \overset{(c)}{\leq} \xi(\omega_{m_3}), \tag{15}$$

where (a) follows from the fact that, in the tangent point ω_{t_2}, the function and the envelope have the same value, (b) from the assumption and (c) from the fact that $\xi(\omega)$ is an envelope to the Dirichlet kernel. Finally, using (15) and the strict monotonicity of $\xi(\omega)$, we have $\omega_{t_2} \geq \omega_{m_3}$, which is a contradiction.

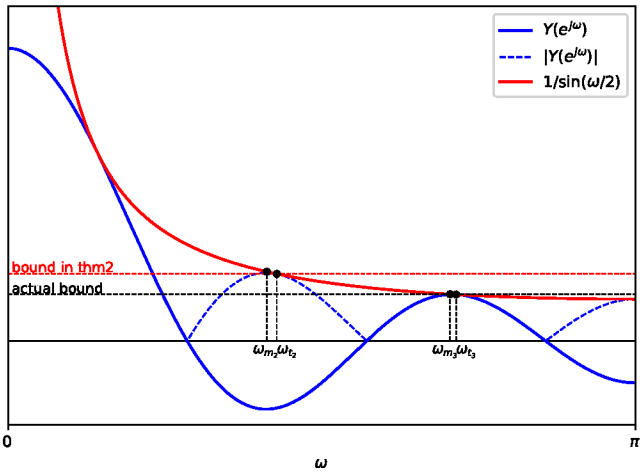


Fig. 7: The (absolute) Dirichlet kernel in (dashed-) blue and its envelope in red for $K = 3$. If the s-line is above the black dashed line, we are guaranteed to have a single solution for α. In Theorem 2 we loosen this by taking the red dashed line.

In the above, we showed that if the value of s is larger than $|Y(e^{j\omega_{t_2}})|$, it will also be larger than $Y(e^{j\omega_{m_3}})$ and thus we will have only one solution for α in $[0, \pi]$ (and hence for b). This condition translates into $s = Y(e^{j\alpha}) \geq \xi(\omega_{t_2}) = 1/\sin(3\pi/2(2K+1))$. Since $\alpha = 2\pi b/N$ then (6) follows from the previous inequality. This proves the first part of the theorem. Note that we could have stated instead a bound on the sampling rate N based on ω_{m_3}. That would provide a less restrictive constraint but for which there is no "closed-form" expression in terms of K.

Now, we want to show that the number of non-recoverable cases has measure zero. We consider jointly cases when the number of recovered Diracs is smaller than $2K+1$ and when the solution is not unique. Consider a set of points on a circle, $S_\alpha \subset S^1$, $S_\alpha = \{e^{j\alpha k} \,|\, k \in \{-K, -(K-1), \ldots, (K-1), K\}\}$, where $K \in \mathbb{N}$ is given and $\alpha \in \mathbb{R}$. We will say that $\widehat{\alpha} \neq \alpha$ are *equivalent* if $S_\alpha = S_{\widehat{\alpha}}$.

Note that in order to discard the trivial equivalent answers, we consider $\alpha \in (0, \pi]$ or equivalently $b \in (0, N/2]$. Assume that $\alpha, \widehat{\alpha}$ are equivalent. Since $S_\alpha = S_{\widehat{\alpha}}$, $e^{j\alpha} \in S_{\widehat{\alpha}}$, there exists $m \in \{-K, \ldots, K\}$, such that $\alpha = (m\widehat{\alpha})_{(0,\pi]}$. Similarly, we can write $\widehat{\alpha} = (\widehat{m}\alpha)_{(0,\pi]}$. This leads to $\alpha = (m\widehat{m}\alpha)_{(0,\pi]}$. Therefore, $\alpha(m\widehat{m} - 1) = \pi\ell$ for some $\ell \in \mathbb{Z}$. This can be true if:

1) $m\widehat{m} - 1 = 0$, or equivalently $m = \widehat{m} = 1$, since m and $\widehat{m}$ are integers. In this case $\widehat{\alpha} = \alpha$, meaning that the solution is unique on the interval $(0, \pi]$.
2) $\alpha = \ell\pi/(m\widehat{m} - 1)$, that is, α is a rational multiple of π. Then, $\widehat{\alpha}$ is also a rational multiple of π, which is clear because $\widehat{\alpha} = \pm\widehat{m}\alpha + 2\pi n$, $n \in \mathbb{Z}$.

Now, we only need to count how many pairs there are of the

form $(p\pi, q\pi)$, where $p, q \in \mathbb{Q}$. This is $|\mathbb{Q} \times \mathbb{Q}| = \aleph_0$, which is countable. Therefore, the set of non-recoverable cases has measure zero. ■

From the proof we can infer the following.

Corollary 1. There is only a finite number of equivalent pairs $\alpha, \widehat{\alpha}$ on the interval $(0, \pi]$.

Proof: Recall from the proof of Theorem 2 that, in order for α to have an equivalent, it has to have the form $\alpha = 2l/(m\widehat{m} - 1)$, for $m, \widehat{m} \in \{-K, \ldots, K\}$. The denominator can have values between $-K^2 - 1$ and $K^2 - 1$, possibly not all of them. The absolute value of the denominator $P = |m\widehat{m} - 1|$ defines a set of possible α's: $\alpha = 2n\pi/P$, for $p \in \{0, \lfloor P/2 \rfloor\}$. Additionally, $\widehat{\alpha}$ must also be in this set. This set has $\lfloor P/2 \rfloor + 1$ elements, and there can be at most $(\lfloor P/2 \rfloor + 1)\lfloor P/2 \rfloor$ equivalent pairs (for a given P). Thus, the total number of pairs is bounded by:

$$\sum_{P=2}^{K^2+1} (\lfloor P/2 \rfloor + 1)\lfloor P/2 \rfloor < (K^2 + 1)^3. \tag{16}$$

■

C. Behaviour of the bound in (6)

Theorem 2 provides a bound to guarantee that the DIRAChlet algorithm results in a unique solution for b. One may wonder about the tightness of this bound for different values of K. As K grows, the scale of the Dirichlet kernel also grows. In fact, the maximum value of the kernel (at $\omega = 0$) is equal to $2K+1$. In the following lemma, we show that the gap between the actual bound, shown in Figure 7 by a black dashed line, and the bound introduced in Theorem 2 converges to a small constant factor of the maximum of the kernel.

Lemma 3. *Define the gap between the bound introduced in Theorem 2 and the actual bound shown in Figure 7 as*

$$\gamma = |Y(e^{j\omega_{t_2}})| - Y(e^{j\omega_{m_3}})\,.$$

Then,

$$\lim_{K\to\infty} \frac{\gamma}{2K+1} \leq \frac{4}{15\pi} \approx 0.085\,.$$

Proof: We can use the strictly decreasing property of the envelope to show that the kernel satisfies $Y(e^{j\omega_{t_3}}) < Y(e^{j\omega_{m_3}})$. Thus,

$$\lim_{K\to\infty} \frac{\gamma}{2K+1} \leq \lim_{K\to\infty} \frac{\xi(\omega_{t_2}) - \xi(\omega_{t_3})}{2K+1} = \frac{4}{15\pi}.$$

■

This lemma is verified experimentally in Figure 8. As the figure shows, the bound converges quickly to $4/15\pi$ as K grows.

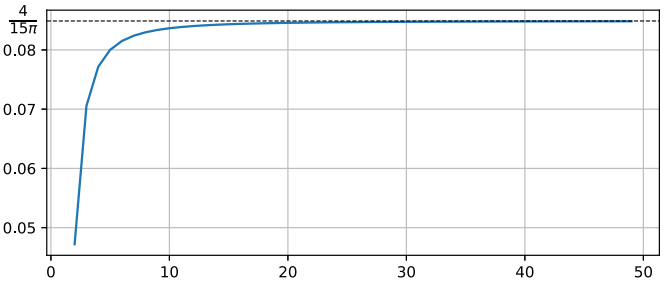


Fig. 8: The gap between the actual bound and the bound in Theorem 2. The x axis shows different values of K while the y axis shows the value of the gap, $\gamma/(2K+1)$.

D. Non-unique solutions

In Theorem 2, we showed that non-unique solutions have measure zero. We have also presented an algorithm that returns all valid solutions; recall that it actually first returns a set of candidate solutions and then retains only the valid ones in an additional step.

In the cases of multiple valid solutions, we have also considered their structure. Experimentally, we have observed that there are two cases. When $\alpha \leq \pi/K$, the valid solutions only occur when $s = -1, 0, 1$. It is easy to prove that these values of s lead to valid solutions, for $\alpha \leq \pi/K$; however, we have been unable to prove that they are the only cases. Finally, when $\alpha > \pi/K$, we have been unable to deduce anything about the structure of the valid solutions.

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 4. *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

$$\psi(x_n) = \frac{p(x_n)}{q(x_n)} \quad \textit{for all } n \in [0 \dots N-1], \tag{17}$$

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.

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:

$$\sum_{k=0}^{K} a_k x_n^k = \sum_{k=0}^{K} b_k \left(\frac{p(x_n)}{q(x_n)}\right)^k, \tag{18}$$

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

$$(q(x_n))^K \sum_{k=0}^{K} a_k x_n^k = \sum_{k=0}^{K} b_k (p(x_n))^k (q(x_n))^{K-k}. \tag{19}$$

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 (19) cannot be equal to the right hand side everywhere. Therefore, (18) 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 (19) 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:

$$C(\hat{\mathbf{x}}, \hat{\mathbf{a}}) = \|\mathbf{V}(\hat{\mathbf{x}})\hat{\mathbf{a}} - \mathbf{y}\|^2, \tag{20}$$

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

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

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

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

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

When the conditions of Lemma 4 are met, we have $\hat{\mathbf{a}} = \mathbf{a}$ and $\hat{x}_n = x_n$ for all $n \in [0 \dots N-1]$.

Unfortunately, (20) 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 $\mathbf{V}$ and solve for the coefficients $\hat{\mathbf{a}}$ using ordinary least squares (OLS).
2) Fix the vector $\hat{\mathbf{a}}$ and make one step of gradient descent with respect to $\hat{\mathbf{x}}$.

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

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

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

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

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

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:

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

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

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

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 an affine warping and a polynomial with sampling positions constrained by a rational function.

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

Algorithm 2 Alternating Least Squares Algorithm (ALS)

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

1: initialize sample transformation parameters α
2: **while** not converged **do**
3: for current matrix $\mathbf{V} := \mathbf{V}$ calculate:

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

4: calculate $\mathbf{\Phi'}$
5: update $\boldsymbol{\alpha}$:

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

6: calculate $\hat{\mathbf{x}} = \boldsymbol{\phi}(\boldsymbol{\alpha})$
7: **end while**

A. *Periodic bandlimited signals*

We start by evaluating the behavior of Algorithm 1 for unwarping periodic bandlimitted signals in the presence of noise. We ran the following simulations. We set $K = 5$ and $\tau = 2K + 1$ and fixed a_k to random arbitrary values (`a = [0.43, -0.15, -0.44, 0.67, -0.32, -0.76, -0.32, 0.67, -0.44, -0.15, 0.43]`). We also fixed the value of b to 4. Then, we took samples of the warped function $f(bx)$ with four different number of samples per period, N, each of which indicate one of the regimes we studied above:

1) $N = 3 \times 2\pi b / Y^{-1}(\sin(\frac{3\pi}{2(2K+1)})^{-1})$: According to Theorem 2, this value of N guarantees that in the noiseless case the intersection of $Y(\omega)$ and s has a unique solution in $[0, \pi]$.

2) $2Kb \leq N < 2\pi b / Y^{-1}(\sin(\frac{3\pi}{2(2K+1)})^{-1})$: We choose this value of T so that the line s has several intersections with the Dirichlet kernel $Y(\omega)$. The case $N \geq 2Kb$ corresponds to $\alpha \leq \frac{\pi}{K}$.

3) $N < 2Kb$: This value of N corresponds to $\alpha > \frac{\pi}{K}$.

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 of the above sampling periods results in a different set of samples $\boldsymbol{h_n}$. We contaminate the samples $\boldsymbol{h_n}$ by random Gaussian noise with zero mean and varying variance such that the value of SNR ranges from -10dB to 40dB, where we define the SNR as,

$$\text{SNR} = 10 \log_{10} \left(\frac{\sigma^2_{\boldsymbol{h_n}}}{\sigma^2_{\text{noise}}} \right). \tag{21}$$

This results in the noisy observations $\tilde{\boldsymbol{h}}_{\boldsymbol{n}}$. Then, we apply Algorithm 1 on these noisy samples to estimate the values of b and a_k (and thus $\boldsymbol{h_n}$); we call these estimated values $\hat{b}$, $\hat{a}_k$ and $\hat{\boldsymbol{h}}_{\boldsymbol{n}}$, respectively. For each value of SNR, we run 10000 simulations.

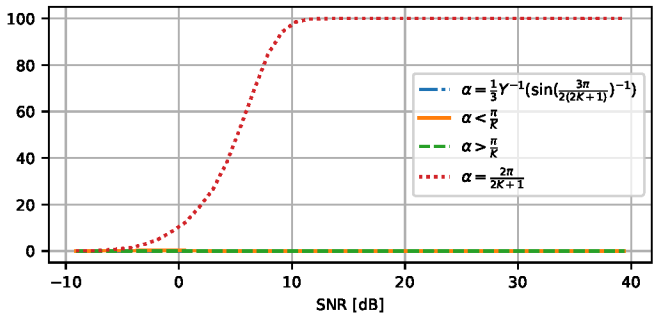


Fig. 9: Percentage of cases with more than one valid solutions for b.

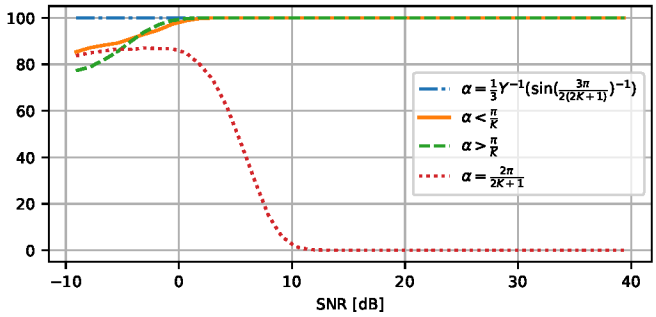


Fig. 10: Percentage of cases where Algorithm 1 results in a single solution.

Figure 9 depicts the percentage of cases that have multiple solutions. We see that, except for Case 4, there is always a unique solution, for the full SNR range considered. In Case 4 (s = 0), we expect to have multiple solutions when the noise level is low. This is due to the fact that each of the estimated values of α represents a multiple of b (in this case: $\widehat{b} \in \{4, 8, 12, 16, 20\}$), all of which produce identical $\hat{\boldsymbol{h}}_{\boldsymbol{n}}$.

Also note that it might happen that when the noise level is too large, the line s falls below the curve of $Y(\omega)$ and thus has no intersection with it. In the above study we have excluded these cases from the analysis and they do not contribute to the average error. Once we set aside the cases that Algorithm 1 generates no or multiple solutions, we can look at the percentage of cases where the algorithm produces a single solution. Figure 10 shows these cases.

and compute the average of the errors defined by

$$\text{error}(\widehat{b}) = \frac{|\widehat{b} - b|}{b} \tag{22}$$

$$\text{error}(\widehat{h}[n]) = \frac{\|\widehat{h}[n] - h[n]\|_2}{\|h[n]\|_2} \tag{23}$$

In the case that the algorithm generates multiple solutions, we choose the solution closest to the true value, when computing the above error. However, we can see in Fig. 11 represents the error in reconstructing b from noisy samples $\widehat{h}[n]$ for each of the four cases detailed above. As it is apparent in the simulation results, case 1 has lower reconstruction errors for large noise levels. This can be partly attributed to the fact that

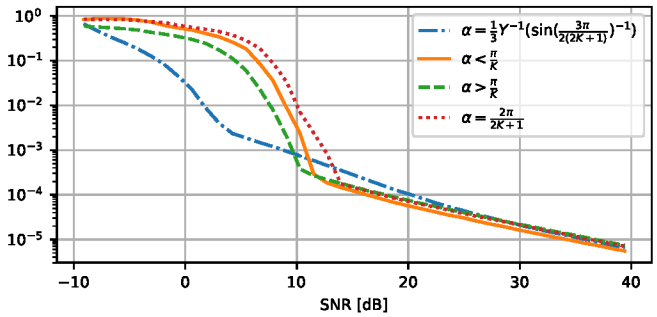


Fig. 11: Error in reconstruction of b from noisy observations $\widetilde{h}[n]$ versus SNR.

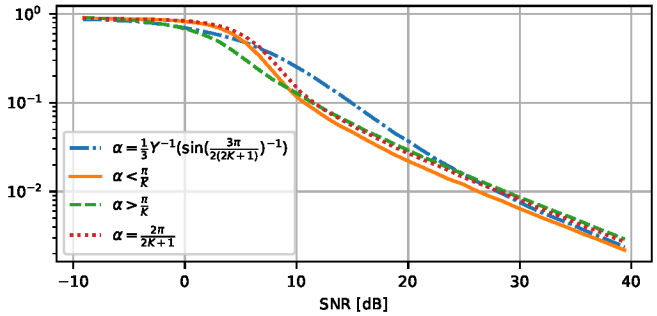


Fig. 12: Error in reconstruction of $h[n]$ from noisy observations $\widetilde{h}[n]$ versus SNR.

the Dirichlet function has a large slope in this part of its curve which results in small variation in $\widehat{\alpha}$ with changes in s. The rest of the cases have a very similar behavior. Also note the break point in the curve at around 10dB.

Figure 12 shows the average error for reconstructing $h[n]$. Interestingly, although the first case has a lower error for estimating b in high noise regimes, it has almost the same performance in estimating $h[n]$.

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

$$\varphi(iT) = \frac{iTd}{v\cos\theta - iT\sin\theta}.$$

Note that, from Lemma 4, we know that $2m$ 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 4 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 (21) – vary between -10dB and 200dB.

We report the error in position estimates defined as

$$\text{error}(\widehat{\theta}) = |\widehat{\theta} - \theta|. \tag{24}$$

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

$$C(\mathbf{x},\mathbf{a}) = \|\mathbf{V}(\mathbf{x})\mathbf{a} - \mathbf{y}\|^2. \tag{25}$$

This cost function is in general not convex, see Figure 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[3].

The problems with local minima can be seen for all kinds of polynomials. However, when the polynomial degree is small, those situations are rare. As the degree of polynomial increases they became more common and lead to increased error even in the noiseless case, see Figures 14 and 15.

The reconstruction is not very robust to noise, see Figure 17. This is not a problem with the alternating algorithm, but with the cost function itself. With noise, the minima of the cost function flatten out, because perfect fitting of the polynomial to the samples is no longer possible. This is also the reason why

[3]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.

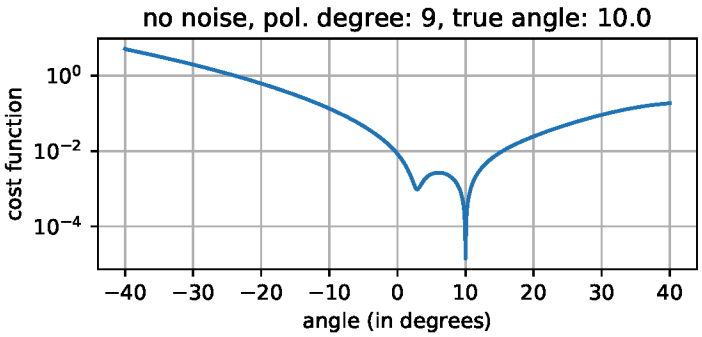


Fig. 13: The cost function is not convex. The plot shows the cost function for different estimated errors, for a fixed polynomial of degree 9, with no noise added. The true angle is 10 degrees. Since the algorithm is initialized at 0, it will stop at the local minimum.

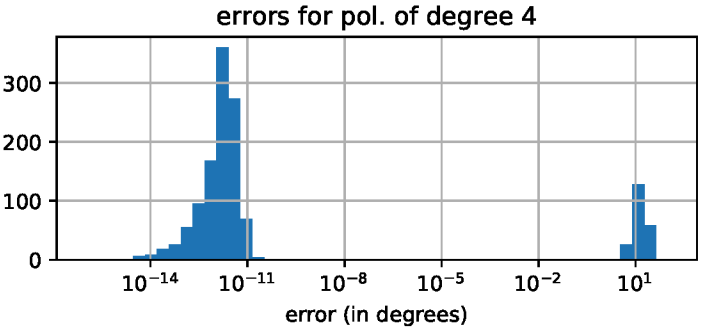


Fig. 14: Histogram of the errors for a polynomial of degree 4, with no additional noise added. The distribution of the error is clearly bi-modal. In more than 80% of cases the error is smaller than 10^{-10}, yet the mean is around 2.5.

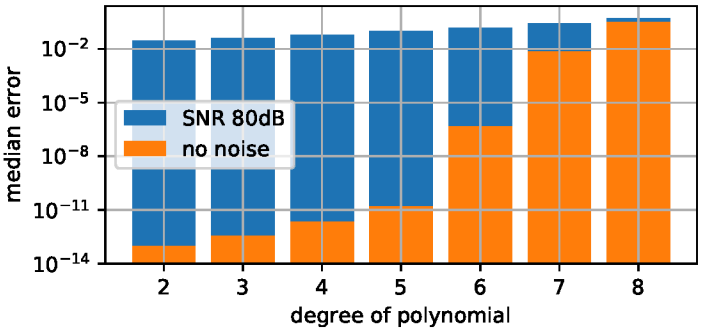


Fig. 15: Results of the ALS algorithm for different polynomial degrees. The median error with no noise is shown in orange and the median error with a small amount of additive noise (SNR 80dB) is shown in blue. As one can see, in the noiseless case, the algorithm breaks down at around degree 6. In the 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.

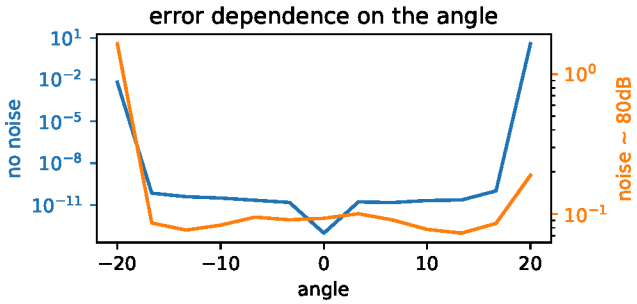


Fig. 16: Median error for different angles, aggregated for 7 different polynomial degrees (2,. . . ,8). The blue line (left scale) shows the noiseless case, and orange shows an SNR of around 80dB. The decrease of error around 0 for the noiseless case is due to algorithm initialization at 0.

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.

Inside the interval $[-20°, 20°]$, the error does not depend, see Figure 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 gradients that were too large. This is because 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 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 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. Furthermore, in one case we have provided an efficient algorithm that reaches this unique solution. We believe that there are many additional examples of sampling a composite of functions that can be solved.

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

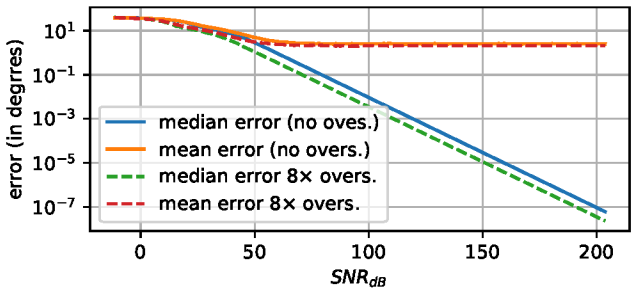


Fig. 17: Error for different signal to noise ratios, for polynomials of degree 4. Around 40dB SNR, the average error and mean error begin to differ. Unsurprisingly, the mean error flattens when the few large errors dominate the mean. Oversampling improves the results but not significantly. Oversampling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

to piecewise linear surfaces would already make a step in this direction. Furthermore, 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," *Proceedings 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," *Proceedings 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 Transactions 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] ——, "Bandlimited spatial field sampling with mobile sensors in the absence of location information," *CoRR*, vol. abs/1509.03966, 2015.

[19] 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.

[20] J. Unnikrishnan, S. Haghighatshoar, and M. Vetterli, "Unlabeled sensing with random linear measurements," *IEEE Trans. on Information Theory*, Feb 2018, to appear.

[21] 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.

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

[23] D. J. Hsu, K. Shi, and X. Sun, "Linear regression without correspondence," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 1531–1540.

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

[25] 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.

[26] 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.

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

[28] 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.

[29] 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.

[30] 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.

[31] 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.

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

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

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