Robust Parametric Signal Estimations

Hanjie Pan
LCAV, I&C, EPFL

***Abstract*—In this research proposal, we generalise the theory of sampling signals with *finite rate of innovation (FRI)* [1] to a specific class of two-dimensional curves, which are implicitly defined as the roots of a mask function. Here the mask function is parametrised as a weighted summation of complex exponentials and hence has finite rate of innovation. It is possible to reconstruct the curve parameters from a set of linear annihilation equations. Reconstruction algorithms in our current work [2] give reliable reconstructions for SNR as low as 20dB. However, in order to apply the curve sampling framework for real applications, e.g., with natural images, it is essential that we have a more robust reconstruction algorithm. We start with the review of the general FRI sampling framework and then focus on a robust reconstruction algorithm. Next we study a 2D shape reconstruction problem, which is shown to be an FRI problem as well. We conclude the proposal by summarising the current work on sampling FRI curves and its possible extensions.**

***Index Terms*—Sampling, finite rate of innovation (FRI), annihilating filter, parametric curve model**

I. Introduction

SAMPLING plays an essential role in signal processing and communications, which aims at representing a continuous domain signal with a few discrete samples. A major concern is whether this set of samples is a faithful representation of the original signal. In the case of bandlimited signals, the answer is given by the fundamental Shannon sampling theorem [3]: if the samples are taken at a rate that is at least twice the signal bandwidth, then we can perfectly reconstruct the continuous domain signal from these samples.

Proposal submitted to committee: June 25th, 2014; Candidacy exam date: July 3rd, 2014; Candidacy exam committee: Prof. Sabine Süsstrunk, Prof. Martin Vetterli, Prof. Thierry Blu, Prof. Dimitri Van De Ville.

This research plan has been approved:

Date: July 3rd, 2014

Doctoral candidate: Hanjie Pan
(name and signature)

Thesis director: Martin Vetterli
(name and signature)

Thesis co-director: Thierry Blu
(name and signature)

Doct. prog. director: ______________
(B. Falsafi) (signature)

EDIC-ru/05.05.2009

The classic sampling theorem, however, cannot deal with cases where the signal has infinite bandwidth. But not everything is lost as bandlimitedness is only a sufficient condition for perfect reconstructions. Under certain assumptions on the underlying signal, e.g., satisfying a specific parametric form, we may still be able to reconstruct the non-bandlimited signal from a finite number of samples. One such framework is sampling and reconstruction for signals with *finite rate of innovation (FRI)* [1], [4], where the signal can be represented with a few parameters (Section II-A). A typical example in 1D is a periodic stream of Diracs. It has been shown that by applying the annihilating filter method, we can perfectly reconstruct signal parameters (and hence the signal itself).

After reviewing the general FRI framework in Section II-A and a robust reconstruction algorithm in Section II-B, we turn to a polygonal shape estimation problem and show that it also falls into the FRI framework in Section II-C, which motivates us to consider a different 2D shape parametrisation. We propose a generalisation of the FRI framework to a class of two-dimensional curves, which are defined implicitly as the roots of a certain mask function (Section III). We show that by sampling an associated continuous domain edge image, we can recover the curves. Possible extensions of the current work is also briefly discussed towards the end of the research proposal.

II. Survey of the Selected Papers

In this section, we review three papers that lay the foundations for the proposed PhD project. We first present the work by Blu *et al.* [4] on the sampling and reconstruction framework for signals with *finite rate of innovation*. Both the noiseless and noisy scenarios have been considered. Then, a reconstruction algorithm [6], which achieves improved reconstruction robustness, is studied. Finally, we introduce a parametric shape (a polygon) estimation problem by Milanfar *et al.* [8].

A. Sparse Sampling of Signal Innovations [4]

Consider a classic problem in signal processing: suppose a sparse signal (e.g., in time or another transformation domain) is measured through a sampling device, which consists of a smoothing filter and uniform sampling. What is the minimum number of samples required in order to have a perfect reconstruction? The sparse signals are usually of infinite bandwidth. Hence, based on Shannon sampling theorem, it is impossible to recover the signal from countable number of samples. However, we also have a strong prior knowledge that the signal is sparse, based on which we may still hope to reconstruct the signal from a few samples.

In this section, we review the sampling and reconstruction framework that addresses a specific class of sparse signals, which can be parametrised with a few parameters (i.e., a sparse representation). It is proved that it is still possible to reconstruct the signal from a few measurements.

1) Signals with Finite Rate of Innovation: Typical examples of signals with finite rate of innovation (FRI), include periodic streams of Diracs [1], [4], piecewise polynomials [1], and piecewise sinusoids [5]. A common feature of such signals is that they all can be represented in parametric forms that have finite degrees of freedom per unit time (or space in the 2D cases).

Definition 1 (Signal Rate of Innovation)**.** *Let C_T denotes a counting function, which counts the number of degrees of freedom of a signal $x(t)$ over a time interval of length T. Then, the rate of innovation is*

$$\rho = \lim_{T\to\infty} \frac{C_T}{T}. \tag{1}$$

A signal with finite rate of innovation is the one with finite ρ as in (1)*.*

For bandlimited signal, the signal rate of innovation is simply the signal bandwidth B — any signal bandlimited to $[-B/2, B/2]$ can be expressed as a weighted summation of the shifted sinc basis [3]:

$$x(t) = \sum_{n\in\mathbb{Z}} x_n \operatorname{sinc}\left(Bt - n\right),$$

where $x_n = x(n/B)$ and $\operatorname{sinc}(t) \overset{\text{def}}{=} \frac{\sin \pi t}{\pi t}$. Hence, $x(t)$ has 1 degree of freedom every $1/B$ second, i.e., $\rho = B$.

Compared with signal bandwidth, the notion of signal rate of innovation is a more general concept that is also applicable to cases other than bandlimited signals, e.g., the τ-periodic streams of Diracs

$$x(t) = \sum_{k'\in\mathbb{Z}} \sum_{k=1}^{K} x_k \delta(t - t_k - k'\tau) \tag{2}$$

has a rate of innovation $\rho = 2K/\tau$ (K for amplitudes x_k and K of positions t_k of the Diracs within each period τ).

2) Sampling Signals at the Rate of Innovation: It turns out that we can devise a sampling setup that reconstructs FRI-type signals, like (2), at the rate of signal innovation.

For the sake of simplicity, assume the τ-periodic signal $x(t)$ is convolved with a sinc window of bandwidth B (ideal low-pass filtering) and is uniformly sampled with sampling step $T = \tau/N$:

$$y_n = \langle x(t), \operatorname{sinc}(B(nT - t))\rangle = \sum_{k=1}^{K} x_k \varphi(nT - t_k), \tag{3}$$

where $\varphi(t) = \sum_{k'\in\mathbb{Z}} \operatorname{sinc}\left(B\left(t - k'\tau\right)\right) = \frac{\sin(\pi Bt)}{B\tau \sin(\pi t/\tau)}$ is the Dirichlet kernel. Note that we can represent the periodic signal $x(t)$ with its Fourier series as:

$$x(t) = \sum_{m\in\mathbb{Z}} \hat{x}_m \mathrm{e}^{j2\pi mt/\tau},$$

where

$$\hat{x}_m = \frac{1}{\tau} \sum_{k=1}^{K} x_k \underbrace{\mathrm{e}^{-j2\pi m t_k/\tau}}_{u_k^m}. \tag{4}$$

If a sufficient number of samples is available (i.e., $N \geq B\tau$), then we can relate the samples $\{y_n\}_{n=1}^{N}$ with the Fourier series coefficients $\hat{x}_m$ by:

$$y_n = \frac{1}{N} \sum_{|m|\leq\lfloor B\tau/2\rfloor} \frac{N}{B} \hat{x}_m \mathrm{e}^{j2\pi mn/N}. \tag{5}$$

Observe that we can treat (5) as a *bandlimited* (to low-pass frequencies) inverse discrete Fourier transform (IDFT) of $\frac{N}{B}\hat{x}_m$, i.e., the DFT coefficients of y_n are

$$\hat{y}_m = \begin{cases} \frac{N}{B}\hat{x}_m & \text{if } |m| \leq \lfloor B\tau/2 \rfloor \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$

Theorem 1 (Annihilating Filter)**.** *There exists a filter*[1] *$H(z)$, whose roots correspond to the locations of the Diracs $u_k = \mathrm{e}^{-j2\pi t_k/\tau}$:*

$$H(z) = \sum_{k=0}^{K} h_k z^{-k} = \prod_{k=1}^{K} (1 - u_k z^{-1}),$$

such that its convolution with the DFT coefficients of the τ-periodic streams of Diracs (2) *are zero: $h_m * \hat{x}_m = 0$.*

Put it differently: the filter h_m annihilates the DFT coefficients $\hat{x}_m$. That is why the filter is also known as *annihilating filter*.

On the other hand, we have shown in (6) that $\hat{x}_m$, has one to one correspondence with the DFT coefficients of the samples. Based on Theorem 1, we can build K equations from $2K$ consecutive DFT coefficients, e.g., $\{\hat{x}_m\}_{m=-K}^{K-1}$ (or equivalently $\{\hat{y}_m\}_{m=-K}^{K-1}$ because of (6)):

$$\mathbf{A}\mathbf{h} = \mathbf{0}, \tag{7}$$

where

$$\mathbf{A} = \begin{bmatrix} \hat{y}_0 & \hat{y}_{-1} & \cdots & \hat{y}_{-K} \\ \hat{y}_1 & \hat{y}_0 & \cdots & \hat{y}_{-K+1} \\ \vdots & \ddots & \ddots & \vdots \\ \hat{y}_{K-1} & \hat{y}_{K-2} & \cdots & \hat{y}_{-1} \end{bmatrix} \quad \text{and} \quad \mathbf{h} = \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_K \end{bmatrix}.$$

Since $h_0 = 1$, h_m is found by solving a $K \times K$ linear system of equations. The Diracs' positions $\{t_k\}_{k=1}^{K}$ are obtained by taking the roots of the filter $H(z)$.

The amplitudes of the Diracs x_k are given by solving a Vandermonde system that involves K consecutive DFT coefficients, e.g., $\hat{x}_m = \frac{N}{B}\hat{y}_m$ for $m = 0, 1, \cdots, K-1$:

$$\begin{bmatrix} u_1^0 & u_2^0 & \cdots & u_K^0 \\ u_1^1 & u_2^1 & \cdots & u_K^1 \\ \vdots & \vdots & \vdots & \vdots \\ u_1^{K-1} & u_2^{K-1} & \cdots & u_K^{K-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{bmatrix} = \frac{N}{B} \begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \vdots \\ \hat{y}_{K-1} \end{bmatrix}.$$

To summarize, we have shown an algorithm that manages to recover both the Diracs' positions and the amplitudes with

[1]Here we have assumed without loss of generality that $h_0 = 1$ as any scaling to h_0 does not change the annihilation results in the Theorem.

no more than $2K$ samples[2], i.e., the signal is reconstructed from samples obtained at the signal's rate of innovation.

3) Robust Reconstructions from Noisy Samples: In the presence of noise (or more generally model mismatch), the annihilation equation (7) is only satisfied approximately. In order to have a reliable reconstruction, we need to have access to more samples than the critically sampled cases considered in the previous section.

Two reconstruction algorithms were proposed in [4] to cope with cases with different levels of noise:

- Total least-square approach — The annihilating filter is the solution that minimize the annihilation error in the least-square sense:
$$\min_{\mathbf{h}} \quad \|\mathbf{A}\mathbf{h}\|_2^2$$
$$\text{subject to} \quad \|\mathbf{h}\|^2 = 1,$$
where $\mathbf{A}$ is the Toeplitz matrix built from the noisy samples. The constraint is to avoid trivial solutions, i.e., $\mathbf{h} = \mathbf{0}$. The total least-square solution is given by the singular value decomposition of $\mathbf{A}$ and $\mathbf{h}$ is the singular vector that corresponds to the smallest singular value. Experimentally, the least-square approach achieves relatively robust reconstructions in the presence of mild-levels of noise.
- Cadzow's iterative denoising — Observe that the annihilation equation (7) is also satisfied for *any* filter $\{h_k\}_{k=0}^{L}$, where $L \geq K$ as long as the Diracs' positions $u_k = \mathrm{e}^{-j2\pi t_k/\tau}$ for $k = 1, 2, \cdots, K$, are its roots:
$$H(z) = \sum_{k=0}^{L} h_k z^{-k} = P_{L-K}(z^{-1}) \prod_{k=1}^{K} (1 - u_k z^{-1}),$$
where $P_{L-K}(z^{-1})$ is a polynomial of order $L - K$.
Corollary 1 (Rank of Annihilation Matrix)**.** *The Toeplitz-structured matrix* $\mathbf{A}$*, which is built from the DFT coefficients of the noiseless samples as if the annihilating filter were of length L,* $L \geq K$*:*
$$\mathbf{A} = \begin{bmatrix} \hat{y}_{-M+L} & \hat{y}_{-M+L-1} & \cdots & \hat{y}_{-M} \\ \hat{y}_{-M+L+1} & \hat{y}_{-M+L} & \cdots & \hat{y}_{-M+1} \\ \vdots & \cdots & \cdots & \vdots \\ \hat{y}_{M} & \hat{y}_{M-1} & \cdots & \hat{y}_{M-L} \end{bmatrix},$$
has rank K *for* $M = \lfloor B\tau/2 \rfloor$ *and* $M \geq L$*.*
The reason is that we have $L - K + 1$ different choices for the L-length annihilating filter $\mathbf{h}$ such that the annihilation equations $\mathbf{A}\mathbf{h} = \mathbf{0}$ are satisfied. Hence, the dimension of the null space of $\mathbf{A}$ is $L - K + 1$, i.e., $\mathrm{rank}(\mathbf{A}) = (L + 1) - (L - K + 1) = K$.
The Cadzow's iterative denoising algorithm exploits the result stated in Corollary 1 and iterates back and forth between
  (i) Thresholding step — Set the smallest singular values of $\mathbf{A}$ to zero based on Corollary 1;
  (ii) Projection step — Find the closest Toeplitz-structured matrix to the thresholded matrix obtained in step (i).
We can then apply the usual total least-square approach, where $\mathbf{A}$ is built based on the *denoised* DFT coefficients from the Cadzow's iteration, to reconstruct the signal innovations.

[2]To be precise, the minimum number of samples required is $2K + 1$ because $B\tau$ is assumed to be an *odd* integer in (3) for the consideration of the convergence of the τ-periodized sum of sinc kernels.

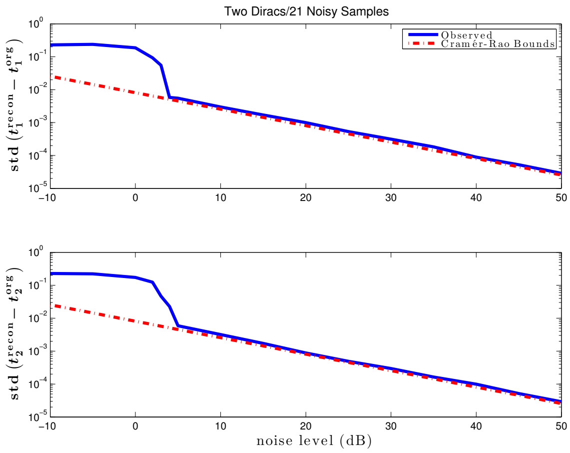


Fig. 1. Comparisons of Cramér-Rao bounds with the standard deviations of the estimated Diracs' positions.

4) Estimation Error Bounds: The performance of the reconstruction algorithms in the previous section can be evaluated objectively by comparing with the theoretical estimation error bounds for any un-biased risk estimator (see Theorem 2).

Theorem 2 (Cramér-Rao Bounds)**.** *Let the noisy samples of the periodic streams of Diracs* (2) *be*
$$y_n = \sum_{k=1}^{K} x_k \varphi(nT - t_k) + \varepsilon_n,$$
where ε_n *is a zero-mean stationary Gaussian noise. Then the covariance matrix of any unbiased estimate based on the measurements* $\mathbf{\Theta}$ *for the unknown parameters* $\{x_k\}_{k=1}^{K}$ *and* $\{t_k\}_{k=1}^{K}$ *are lower bounded by the inverse of the Fisher information matrix:*
$$\mathrm{cov}(\mathbf{\Theta}) \succeq \left(\mathbf{\Phi}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{\Phi}\right)^{-1},$$
where
$$\mathbf{\Phi} = \left[\begin{array}{ccc|ccc} \varphi(T - t_1) & \cdots & \varphi(T - t_K) & -x_1\varphi'(T - t_1) & \cdots & -x_K\varphi'(T - t_K) \\ \varphi(2T - t_1) & \cdots & \varphi(2T - t_K) & -x_1\varphi'(2T - t_1) & \cdots & -x_K\varphi'(2T - t_K) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \varphi(NT - t_1) & \cdots & \varphi(NT - t_K) & -x_1\varphi'(NT - t_1) & \cdots & -x_K\varphi'(NT - t_K) \end{array}\right]$$
and $\mathbf{R}$ *is the covariance matrix of the stationary Gassian noise:* $\mathbf{R}_{n,m} = r_{n-m} = \mathbb{E}[\varepsilon_n \varepsilon_m]$.

In the specific case where $x(t)$ contains one Dirac in each period (i.e., $K = 1$), we can evaluate explicitly the estimation uncertainties in the Dirac's position Δt and amplitude Δx.

Corollary 2 (One-Dirac Periodic Sinc). *Assume the stationary Gaussian noise $\{\varepsilon_n\}_{n=1}^{N}$ is N-periodic and the sampling kernel is the Dirichlet kernel $\varphi(t) = \frac{\sin(\pi B t)}{B\tau \sin(\pi t/\tau)}$, then the uncertainties in the estimated Dirac's position and amplitude are lower bounded as:*

$$\frac{\Delta t}{\tau} \geq \frac{B\tau}{2\pi |x_1| \sqrt{N}} \left(\sum_{|m| \leq \lfloor B\tau/2 \rfloor} \frac{1}{\hat{r}_m} \right)^{-1/2}$$

and

$$\Delta x \geq \frac{B\tau}{\sqrt{N}} \left(\sum_{|m| \leq \lfloor B\tau/2 \rfloor} \frac{1}{\hat{r}_m} \right)^{-1/2},$$

where $\hat{r}_m$ are the DFT coefficients of r_n.

For cases with more than one Diracs, the results in Corollary 2 still hold approximately provided that the Diracs are located sufficiently apart. Experimentally, the reconstruction algorithms introduced in Section II-A3 reaches the Cramér-Rao bounds in the high SNR regions (as low as SNR $= 5$dB, see Fig. 1).

5) Summary: This paper addresses sampling and reconstruction problems of a class of sparse signals that have finite rate of innovation. Perfect reconstructions are possible by sampling the signals at the rate of innovation despite of the non-bandlimitedness. The performance of the reconstruction algorithms in the presence of model mismatch are also analyzed with the optimal performances given by the Cramér-Rao bounds.

B. Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise [6]

In the previous section we have reviewed the sampling and reconstruction framework for signals with finite rate of innovation, where we want to reconstruct the Diracs' amplitudes and positions from the Fourier series (4). The reconstruction problem bears remarkable similarities to array processing [7], which has wide applications in sonar, radar, seismology and audio processing.

Formally, the array processing problem aims at estimating the unknown parameters x_k and u_k from the measurements a_n:

$$a_n = \sum_{k=1}^{K} x_k u_k^n + \varepsilon_n \qquad \text{for } n = 1, 2, \cdots, N \tag{8}$$

where u_k's represent the unknown sources and are in general complex exponentials (possibly damped): $u_k = \mathrm{e}^{-j\phi_k + \alpha_k}$; x_k's are the amplitudes of the sources; and ε_n is additive noise. We can treat the FRI reconstruction problem in Section II-A as a specific type of array processing problem, where the measurements $\{a_n\}_{n=0}^{N}$ correspond to the Fourier series of the streams of Diracs $\hat{x}_m$ as in (4) (or equivalently the DFT coefficients of the samples $\hat{y}_m$); and $\{u_k\}_{k=1}^{K}$ are pure complex exponentials, i.e., $\alpha_k \equiv 0$.

Because of such direct connections with array processing, we can apply various mature algorithms developed in array processing to FRI reconstructions. In this section, we review one efficient array processing algorithm, namely *Iterative Quadratic Maximum Likelihood (IQML)* algorithm [6].

1) Problem Formulation: Consider a "multi-experiment" setting of the typical array processing problem (8), where M snapshots are taken, say, at different times:

$$\boldsymbol{a}^{(i)} = \boldsymbol{\Psi}_{\boldsymbol{u}} \boldsymbol{x}^{(i)} + \boldsymbol{\varepsilon}^{(i)} \qquad \text{for } i = 1, 2, \cdots, M.$$

Here the superscript denotes the measurements/parameters at the i-th snapshot; $\boldsymbol{a} = [a_1, a_2, \cdots, a_N]^{\mathrm{T}}$ are the measurements, $\boldsymbol{u} = [u_1, u_2, \cdots, u_K]^{\mathrm{T}}$ the signal parameters, $\boldsymbol{x} = [x_1, x_2, \cdots, x_K]^{\mathrm{T}}$ the signal amplitudes, $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N]^{\mathrm{T}}$ the additive noise respectively; and $\boldsymbol{\Psi}_{\boldsymbol{u}}$ is a Vandermonde matrix defined as:

$$\boldsymbol{\Psi}_{\boldsymbol{u}} = \begin{bmatrix} u_1 & u_2 & \cdots & u_K \\ u_1^2 & u_2^2 & \cdots & u_K^2 \\ \vdots & \vdots & \vdots & \vdots \\ u_1^N & u_2^N & \cdots & u_K^N \end{bmatrix}.$$

The goal is to estimate the signal parameters $\{u_k\}_{k=1}^{K}$ (i.e., $\boldsymbol{\Psi}_{\boldsymbol{u}}$) and amplitudes $\boldsymbol{x}^{(i)}$ from the given measurements $\boldsymbol{a}^{(i)}$. Note that since $\boldsymbol{\Psi}_{\boldsymbol{u}}$ also depends on the unknown parameters u_k's that we want to reconstruct, the reconstruction problem is less straight-forward than it may appear.

2) The Maximum Likelihood Criterion: If we assume that the additive noise is (complex) Gaussian white noise with zero mean and a covariance matrix $\sigma^2\mathbf{I}$: $\varepsilon_n \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$, then the maximum likelihood estimate (MLE) of the parameters $\left(\boldsymbol{x}^{(i)}, \boldsymbol{u}\right)$ are given by solving a *non-linear* optimization problem:

$$\min_{\boldsymbol{x}^{(i)}, \boldsymbol{u}} \quad \sum_{i=1}^{M} \left\| \boldsymbol{a}^{(i)} - \boldsymbol{\Psi}_{\boldsymbol{u}} \boldsymbol{x}^{(i)} \right\|^2. \tag{9}$$

For a given set of complex exponentials $\boldsymbol{u}$, (9) reduces to a simple least-square minimization and the optimal $\boldsymbol{x}^{(i)}$ is

$$\boldsymbol{x}^{(i)} = \left(\boldsymbol{\Psi}_{\boldsymbol{u}}^{\mathrm{H}} \boldsymbol{\Psi}_{\boldsymbol{u}}\right)^{-1} \boldsymbol{\Psi}_{\boldsymbol{u}}^{\mathrm{H}} \boldsymbol{a}^{(i)}. \tag{10}$$

Hence, (9) reduces to

$$\min_{\boldsymbol{u}} \quad \mathrm{tr}\left(\mathbf{P}_{\boldsymbol{u}}^{\perp} \mathbf{R}_{\boldsymbol{a}}\right),$$

where $\mathrm{tr}(\cdot)$ is the trace; $\mathbf{P}_{\boldsymbol{u}}^{\perp} = \mathbf{I} - \boldsymbol{\Psi}_{\boldsymbol{u}} \left(\boldsymbol{\Psi}_{\boldsymbol{u}}^{\mathrm{H}} \boldsymbol{\Psi}_{\boldsymbol{u}}\right)^{-1} \boldsymbol{\Psi}_{\boldsymbol{u}}^{\mathrm{H}}$ is the orthogonal complement of the projection onto the column space of $\boldsymbol{\Psi}_{\boldsymbol{u}}$; and $\mathbf{R}_{\boldsymbol{a}} = \sum_{i=1}^{M} \boldsymbol{a}^{(i)} \left[\boldsymbol{a}^{(i)}\right]^{\mathrm{H}}$ is the sample autocorrelation matrix of the measurements $\boldsymbol{a}^{(i)}$ for $i = 1, 2, \cdots, M$ (up to a constant scaling factor).

However, the implicit dependencies of $\mathbf{P}_{\boldsymbol{u}}^{\perp}$ on $\boldsymbol{u}$ make the direct minimization difficult. We can rewrite the objective function in terms of another variable based on the observations in the following propositions.

Proposition 1. *Let the polynomial $H(z) = \sum_{k=0}^{K} h_k z^{-k}$ have roots at $z = u_k$ for $k = 1, 2, \cdots, K$, then $\mathbf{H}\boldsymbol{\psi}_k = \mathbf{0}$ for all $k = 1, 2, \cdots, K$, where the Toeplitz matrix $\mathbf{H}$ is*

$$\mathbf{H} = \begin{bmatrix} h_K & \cdots & h_0 & 0 & \cdots & 0 \\ 0 & h_K & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & h_K & \cdots & h_0 \end{bmatrix}, \tag{11}$$

and $\boldsymbol{\psi}_k$ is the k-th column of $\boldsymbol{\Psi}_{\boldsymbol{u}}$: $\boldsymbol{\psi}_k = \left[u_k, u_k^2, \cdots, u_k^N\right]^{\mathrm{T}}$.

That is to say the rows of $\mathbf{H}$ are orthogonal to the columns of $\mathbf{\Psi}_{\boldsymbol{u}}$. Because $\mathbf{H}$ is a Toeplitz matrix, we can treat $\mathbf{H}\boldsymbol{\psi}_k$ as the convolution between $[h_0, h_1, \cdots, h_K]$ and $\boldsymbol{\psi}_k$, i.e., $H(z)$ is the *annihilating filter* (see Theorem 1) of $\boldsymbol{\psi}_k$ for all $k = 1, 2, \cdots, K$.

Following the statement in Proposition 1 we can express $\mathbf{P}_{\boldsymbol{u}}^{\perp}$ with the annihilating filter coefficients h_k.

Proposition 2. *The orthogonal complement $\mathbf{P}_{\boldsymbol{u}}^{\perp}$ of the projection onto the column space of $\mathbf{\Psi}_{\boldsymbol{u}}$ is $\mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\mathbf{H}$.*

The objective function in the maximum likelihood estimate reduces to

$$\operatorname{tr}(\mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\mathbf{H}\mathbf{R}_{\boldsymbol{a}}) = \sum_{i=1}^{M} \left(\mathbf{H}\boldsymbol{a}^{(i)}\right)^{\mathrm{H}} (\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\mathbf{H}\boldsymbol{a}^{(i)}.$$

On the other hand, because of the commutativity of convolutions $\boldsymbol{a}^{(i)} * \mathbf{h} = \mathbf{h} * \boldsymbol{a}^{(i)}$, we have $\mathbf{A}^{(i)}\mathbf{h} = \mathbf{H}\boldsymbol{a}^{(i)}$, where $\mathbf{A}^{(i)}$ is the convolution matrix associated with $\boldsymbol{a}^{(i)}$ that follows the same Toeplitz structure for $\mathbf{H}$ as in (11), i.e., the MLE is

$$\min_{\mathbf{h}\in\mathcal{H}} \quad \mathbf{h}^{\mathrm{H}} \left(\sum_{i=1}^{M} \left[\mathbf{A}^{(i)}\right]^{\mathrm{H}} \left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{-1} \mathbf{A}^{(i)} \right) \mathbf{h}. \tag{12}$$

Here $\mathcal{H}$ is the feasible set for the annihilating filter $\mathbf{h}$, which is application-dependent.

3) Reconstruction Algorithm: We have reformulated the maximum likelihood estimate into (12), whose solution can be obtained iteratively: at each iteration, we need to solve a quadratic minimization (subject to a proper constraint). The signal parameters $\{u_k\}_{k=1}^{K}$ are reconstructed at the end of iterations by taking the roots of the polynomial $H(z)$ and the amplitudes $\boldsymbol{x}^{(i)}$ for $i = 1, 2, \cdots, M$ are given by (10). The iterative algorithm is summarized in Algorithm 1.

Algorithm 1: Iterative Quadratic Maximum Likelihood

Input: $\boldsymbol{a}^{(i)}$
Output: $\boldsymbol{x}^{(i)}$, $\mathbf{h}$

Initialize $\mathbf{h}$ with a $(K+1)\times 1$ vector;
for `loop` $= 1$ **to** `max_iter` **do**
  Build the convolution matrix $\mathbf{H}$ from $\mathbf{h}$;
  Solve the quadratic optimization (12);
  Exit the loop if the convergence criterion is met: $\|\mathbf{h} - \mathbf{h}_{\mathrm{prev}}\| < \epsilon$;
end
Reconstruct $\{u_k\}_{k=1}^{K}$ by taking the roots of the polynomial $H(z)$ and the amplitudes $\boldsymbol{x}^{(i)}$ from (10).

4) Summary: This paper proposes a reconstruction algorithm by deriving the exact maximum likelihood criterion for an array processing problem. It has been shown to have a close relation with the FRI signal reconstructions introduced in Section II-A. An IQML-based algorithm will be developed and has been observed to give more robust reconstructions in the proposed thesis research project (see details in Section III-B1).

C. Reconstructing Polygons from Moments with Connections to Array Processing [8]

In this section, we review the work by Milanfar *et al.*, which reconstructs the shape of a polygon from a few complex moments (Definition 4) of the associated binary polygonal areas. We show that this problem is actually also an FRI reconstruction problem. The paper considers the polygonal shape estimation problem in the complex plane, where the horizontal and vertical coordinates correspond to the real and imaginary part of a complex variable $z = x + jy$ respectively. We will first review some basic results of complex plane geometry before we formulate exactly what is the shape reconstruction problem to tackle.

1) Background on Complex Plane Geometries: Green's Theorem provides a powerful tool to associate the surface integration with a line integration along the contour that encompasses the surface area. Combined with the Cauchy-Riemann equation for an analytic function f, we can prove that

$$\iint_S f'(z)\mathrm{d}x\mathrm{d}y = \frac{j}{2}\int_{\partial S} f\mathrm{d}\overline{z}, \tag{13}$$

where j is the imaginary unit: $\sqrt{-1}$ and $\mathrm{d}\overline{z} = \mathrm{d}x - j\mathrm{d}y$. Here S is a simply connected area in the complex plane and ∂S denotes the boundary of S. This formulation provides a direct link between a function and its derivatives.

Lemma 1 (Motzkin-Schoenberg Formula [9])**.** *Given a function $f(z)$, which is analytic inside a triangle T, we have*

$$\frac{1}{2A}\iint_T f''(z)\mathrm{d}x\mathrm{d}y = \frac{f(z_1)}{(z_1 - z_2)(z_1 - z_3)} + \frac{f(z_2)}{(z_2 - z_1)(z_2 - z_3)} + \frac{f(z_3)}{(z_3 - z_1)(z_3 - z_2)},$$

where A is the area of the triangle T.

The lemma is proved by applying (13) to $f'(z)$ (i.e., $f(z) \to f'(z)$ and $f'(z) \to f''(z)$) over a triangular area T. The contour integration on the right hand side reduces to three integrations along the sides of the triangle and can be evaluated analytically.

Lemma 1 shows that the integration of the second derivative of any analytic function over a (fixed) triangular area depends solely on the analytic function value on the vertices. This result can be generalised to a polygonal surface.

Theorem 3. *Let $z_1, z_2, \cdots, z_K$ designate the vertices of a polygon P. Then, we can find constants $a_1, a_2, \cdots, a_K$ depending on the vertices z_k for $k = 1, 2, \cdots, K$ but independent of $f(z)$ such that for all $f(z)$ analytic in the closure of P:*

$$\iint_P f''(z)\mathrm{d}x\mathrm{d}y = \sum_{k=1}^{K} a_k f(z_k).$$

This is because a polygon can be partitioned into several non-overlapping triangles (Fig. 2), on which Lemma 1 is applied. We can tell exactly what the coefficients a_n's are as shown in Lemma 2.

Lemma 2. *Let $z_1, z_2, \cdots, z_K$ designate vertices in counter-clockwise order of a convex polygon P. We extend the index*

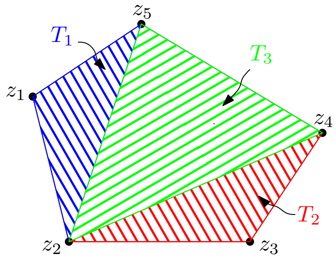


Fig. 2. Partition of a polygon into non-overlapping triangles.

z_k cyclically, i.e., $z_0 = z_K$ and $z_{K+1} = z_1$, etc. Then for all $f(z)$ that is analytic in the closure of P,

$$\iint_P f''(z)\mathrm{d}x\mathrm{d}y = \sum_{k=1}^{K} \frac{2A_k}{(z_k - z_{k-1})(z_k - z_{k+1})} f(z_k),$$

where A_k is the area formed by vertices z_{k-1}, z_k and z_{k+1}.

If we expand the expression for the coefficients a_k, we have

$$a_k = \frac{2}{(z_k - z_{k-1})(z_k - z_{k+1})} \frac{j}{4} \begin{vmatrix} z_{k-1} & \overline{z}_{k-1} & 1 \\ z_k & \overline{z}_k & 1 \\ z_{k+1} & \overline{z}_{k+1} & 1 \end{vmatrix}$$
$$= \frac{j}{2}\left(\frac{\overline{z}_{k-1} - \overline{z}_k}{z_{k-1} - z_k} - \frac{\overline{z}_k - \overline{z}_{k+1}}{z_k - z_{k+1}} \right).$$

Denote the angle between the side $\langle z_k z_{k+1} \rangle$ and the positive real-axis as ϕ_k, then

$$\alpha_k = \frac{\overline{z}_k - \overline{z}_{k+1}}{z_k - z_{k+1}} = \mathrm{e}^{-2j\phi_k},$$

which can be interpreted as the *slope* in the complex plane. Hence, the coefficients $a_k = \frac{j}{2}\left(\mathrm{e}^{-2j\phi_{k-1}} - \mathrm{e}^{-2j\phi_k}\right)$ are the differences in slopes of the two sides $\langle z_{k-1} z_k \rangle$ and $\langle z_k z_{k+1} \rangle$ meeting at the vertex z_k.

2) Problem Formulation: We can relate the vertices of a polygon with moments of the associated binary indicator image, which has value 1 for the interior of the polygon and 0 otherwise:

$$f(x, y) = \begin{cases} 1 & \text{if } (x, y) \in \mathring{P} \\ 0 & \text{otherwise} \end{cases}, \qquad (14)$$

where $\mathring{P}$ denotes the closure of the interior of the polygonal area P. We need to use different types of moments in the problem formulation as well as in the application of the shape estimation algorithm (Section II-C4).

Definition 2. *The geometric moments of a K-sided polygonal region P are $\rho_{m,n} \stackrel{\text{def}}{=} \iint_P x^m y^n \mathrm{d}x\mathrm{d}y$.*

Definition 3. *The harmonic moments of a K-sided polygonal region P are $c_m \stackrel{\text{def}}{=} \iint_P z^m \mathrm{d}x\mathrm{d}y$.*

The harmonic moments can be expressed as linear combinations of several geometric moments by observing that the binomial expansion of z^m is

$$z^m = \sum_{l=0}^{m} \binom{m}{l} x^{m-l} (jy)^l, \qquad (15)$$

Therefore

$$c_m = \left[\boldsymbol{w}^{(m)}\right]^{\mathrm{T}} \boldsymbol{\rho}^{(m)},$$

where $\boldsymbol{w}^{(m)} = \left[\binom{m}{0} j^0 \ \binom{m}{1} j^1 \ \cdots \ \binom{m}{m} j^m\right]^{\mathrm{T}}$ and $\boldsymbol{\rho}^{(m)} = \left[\rho_{m,0} \ \rho_{m-1,1} \ \cdots \ \rho_{0,m}\right]^{\mathrm{T}}$.

Note that if we take $f(z) = z^m$ in Theorem 3, then

$$m(m-1) \iint_P z^{m-2} \mathrm{d}x\mathrm{d}y = \sum_{k=1}^{K} a_k z_k^m.$$

For the ease of notations, we may define the left hand side, which is a re-weighted harmonic moment of order $m-2$, as another type of moment.

Definition 4. *The complex moments τ_m are defined as weighted harmonic moments: $\tau_m \stackrel{\text{def}}{=} m(m-1)c_{m-2}$.*

Finally, we are ready to present the polygonal shape estimation problem:

Suppose that M complex moments $\{\tau_m\}_{m=0}^{M-1}$ are measured. How can we recover the polygon vertices $z_1, z_2, \cdots, z_K$ from the relations $\tau_m = \sum_{k=1}^{K} a_k z_k^m$?

This paper focuses on the reconstruction of vertices alone. A more general problem, where we also need to "connect the dots" once the vertices have been reconstructed, is more involved and is beyond the scope of the discussions in the paper.

3) Reconstruction Algorithm: Observe that the complex moment τ_m are weighted summations of the vertices' coordinates z_k (with power m). This is exactly the same reconstruction problem that we have encountered in Section II-A and Section II-B. For instance, in the shape estimation problem a_k corresponds to the Diracs' amplitudes x_k; z_k corresponds to the complex exponentials u_k; and the measured complex moments τ_m correspond to the Fourier series of the Dirac stream $\hat{x}_m$ in (4).

We can directly apply the same annihilating filter method (as well as the robust algorithms aforementioned) to recover the positions of the vertices. Specifically, the annihilating filter in this case is

$$H(z) = \sum_{k=0}^{K} h_k z^{-k} = \prod_{k=1}^{K} (1 - z_k z^{-1}),$$

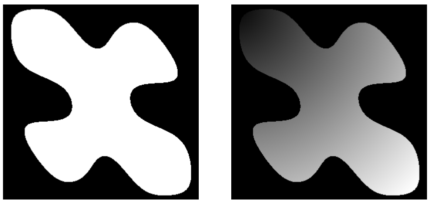
and h_m annihilates the complex moments: $h_m * \tau_m = 0$.

4) Application on Tomographic Reconstructions: One application of the polygonal shape estimation is in tomographic reconstructions, where moments of the underlying polygonal region are related with the projections in a *linear* manner. Specifically the tomography projection is modelled with the Radon transform.

Definition 5. *The Radon transform $Rf(t, \theta)$ of a compactly supported continuous function $f(x, y)$ is*

$$Rf(t, \theta) \stackrel{\text{def}}{=} \iint f(x, y)\delta(t - x\cos\theta - y\sin\theta)\mathrm{d}x\mathrm{d}y,$$

where $\delta(\cdot)$ is the Dirac delta function.



(a) (b)

Fig. 3. A continuous domain "edge image", which is discontinuous on some predefined curve. The discontinuity across the curve can be (a) uniform or (b) varying in general (see texts before Theorem 4 for details).

It can be proved that for any $F(t)$ that is square-summable, we have

$$\int Rf(t,\theta)F(t)\mathrm{d}t = \iint f(x,y)F(x\cos\theta + y\sin\theta)\mathrm{d}x\mathrm{d}y$$

If we take $F(t) = t^m$ and assume $f(x,y)$ is the binary indicator image associated with certain polygon P as in (14), then

$$\int Rf(t,\theta)t^m \mathrm{d}t = \sum_{l=0}^{m} \binom{m}{l} \rho_{m-l,l} \cos^{m-l}\theta \sin^l \theta,$$

where ρ is the *geometric moment* as in Definition 2. That means the m-th order geometric moment of the Radon transform at angle θ (i.e., the left hand side) is a linear combination of the m-th order geometric moments of the indicator image $\rho_{m-l,l}$ for $l = 0, 1, \cdots, m$ (and hence the harmonic moments c_m from (15)). Therefore, we can obtain the complex moments from the projections and then apply the shape reconstruction algorithm to recover the underlying polygon shape.

5) Summary: The paper addresses an interesting problem where vertices of a polygon can be recovered from several complex moments of the associated indicator image. We have shown in the review that the polygonal shape reconstruction problem is also an FRI problem. The work has motivated us to consider a shape recovery problem with a different boundary parametrisation than a polygon (see details in the next section).

III. Research Proposal

Previous efforts to generalise the sampling and reconstruction frameworks for multi-dimensional FRI signals have led to the extensions to 2-D Diracs [10], lines of finite length [10], and polygons [10]–[13]. As we have discussed in Section II-C that the polygonal shape reconstruction from the complex moments of the associated indicator image also falls into the FRI framework. These multi-dimensional signals are usually of very simple geometry and may not have enough descriptive power to cope with various shapes encountered in higher dimensions. Can we have a more general parametrisation for two-dimensional curves (as opposed to line segments in a polygon)?

A. Summary of Current Work

We have considered a specific class of curves C, that are implicitly defined as the roots of a mask function $\mu(x,y)$ in [2]:

$$C: \overbrace{\sum_{k=-K_0}^{K_0} \sum_{l=-L_0}^{L_0} c_{k,l} \mathrm{e}^{j\frac{2\pi k}{\tau_x}x + j\frac{2\pi l}{\tau_y}y}}^{\mu(x,y)} = 0, \text{ where } \begin{cases} 0 \le x < \tau_x, \\ 0 \le y < \tau_y. \end{cases} \tag{16}$$

Here τ_x and τ_y are some positive real numbers, which specify the periods along x-y directions respectively. The curve is uniquely specified by the set of coefficients $c_{k,l}$, which is of finite dimensions. Hence, the family of curves also have finite rate of innovation.

This formulation leads to curves with very diverse topologies: we may have multiply connected curves, open curves, crossings, and non-smooth connections, etc. Moreover, this representation is potentially very rich (because the mask function is a Fourier expansion of finite length): if we increase the number of degrees of freedom, we can approximate arbitrary mask functions as accurately as we want.

For each curve defined implicitly this way, we have an associated *edge image*[3], which is analytic almost everywhere except on the predefined curve C where it is discontinuous with certain jump values (see examples in Fig. 3).

We have showed that the Fourier transform of the edge images, whose positions of discontinuities are defined by curves (16) can be annihilated by the curve coefficients $c_{k,l}$.

Theorem 4 (Annihilation of the Edge Image)**.** *Consider an annihilable curve C defined in* (16) *and its associated edge image $I_C(x,y)$, then for any frequencies ω_x and ω_y we have:*

$$\sum_{k=-K_0}^{K_0} \sum_{l=-L_0}^{L_0} c_{k,l} \widehat{I}'_C \left(\omega_x - \frac{2\pi k}{\tau_x}, \omega_y - \frac{2\pi l}{\tau_y}\right) = 0, \tag{17}$$

where $\widehat{I}'_C(\omega_x, \omega_y) = j(\omega_x + j\omega_y)\hat{I}_C(\omega_x, \omega_y)$.

If we discretize (17) with $\omega_x = \frac{2\pi k}{\tau_x}$ and $\omega_y = \frac{2\pi l}{\tau_y}$ for $k, l \in \mathbb{Z}$, then we have a set of *linear* equations about the (unknown) curve coefficients: $c_{k,l} * h_{k,l} = 0$ and the (discrete) annihilating filter is

$$h_{k,l} = \left(\frac{2\pi k}{\tau_x} + j\frac{2\pi l}{\tau_y}\right) \hat{I}_C \left(\frac{2\pi k}{\tau_x}, \frac{2\pi l}{\tau_y}\right). \tag{18}$$

Based on the annihilation equations we can devise a similar sampling and reconstruction framework for FRI curves (16). Please refer to [2] for details.

Another important contribution is that we have cast a new spatial domain interpretation of the annihilation equations. Specifically, we have shown that the annihilation equation, which is a Fourier domain convolution, corresponds to a simple (continuous) spatial domain multiplication:

$$\sum_{k,l} c_{k,l} \widehat{I}'_C \left(\omega_x - \frac{2\pi k}{\tau_x}, \omega_y - \frac{2\pi l}{\tau_y}\right) \xleftrightarrow{\mathcal{F}^{-1}} \mu(x,y) \times I'_C(x,y).$$

From this perspective, the function $\mu(x,y)$ serves as a "mask" that automatically annihilates whatever is different from zero in the derivative image $I'_C \overset{\text{def}}{=} \left(\frac{\partial}{\partial x} + j\frac{\partial}{\partial y}\right) I_C$ (Fig. 4).

[3]In the specific case where the image has constant amplitude 1 for the interior areas of the curve and 0 otherwise, the binary edge image follows the same definition of the indicator image (14) in Section II-C.

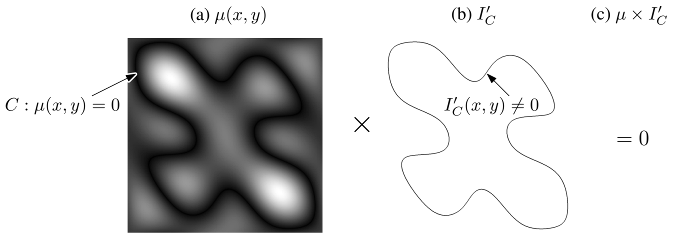


Fig. 4. Spatial domain interpretation of the annihilation equation (17). The function $\mu(x, y)$, whose roots defines the curve, serves as a mask to annihilate whatever is different from zero in the derivative image I'_C.

Thanks to this new spatial domain interpretation, we can apply the annihilation idea to cases where the cure model (16) is not satisfied exactly, e.g., with natural images. Preliminary tests have shown the effectiveness by enforcing the FRI curve model as a linear constraint in image up-sampling [2].

B. Future Directions

We may extend the current work on FRI curves in the following directions.

1) Robust Reconstructions of Curve Model: In [2], we have applied both total least-square approach and the Cadzow's iterative denoising algorithm for the reconstruction of curve coefficients. The algorithms work reasonably well and give reliable reconstructions for noise levels as low as 20dB. One reason that these algorithms become less robust compared with the 1D Diracs cases is that the annihilating filter in (18) is built from the Fourier transform of the *derivative* image: $\widehat{I'_C} = j(\omega_x + j\omega_y)\hat{I}_C(\omega_x, \omega_y)$ — errors in high-frequency Fourier data are amplified leaving the algorithm biased and inefficient for the FRI curve reconstructions. Instead, a more appropriate goal should be to find the denoised Fourier data such that the ℓ_2 distance with the given noisy Fourier data $\hat{I}_C$ is minimized:

$$\begin{aligned} \min_{\mathbf{b},\mathbf{c}} \quad & \|\mathbf{a} - \mathbf{b}\|_2^2 \\ \text{subject to} \quad & \mathbf{T}(\mathbf{b})\mathbf{c} = \mathbf{0} \\ & \|\mathbf{c}\|_2^2 = 1, \end{aligned}$$

where $\mathbf{a}$ and $\mathbf{b}$ are the noisy and denoised Fourier transforms of the edge image $\hat{I}_C$ respectively; $\mathbf{T}(\mathbf{b})$ constructs the convolution matrix associated with (18) from a given set of Fourier data $\mathbf{b}$; and we have rearranged the curve coefficients $c_{k,l}$ in a vector form $\mathbf{c}$. It turns out that we can solve this optimisation problem with an IQML-like algorithm and achieve more robust reconstructions (Fig. 5).

2) Unified View of the Local and Global Curve Model: Note that the curve model (16) is a global one — it may not be the most suitable choice to characterize edges of a natural image, which are primarily a local feature. On the other hand, based on the spatial domain interpretation of annihilation equations, we may obtain an edge model directly by minimizing the product between the derivative image the mask function in e.g., the least-square sense:

$$\min_{\mathbf{c}} \iint \|\mu_{\mathbf{c}}(x - x_0, y - y_0) \cdot I'(x, y)\|_2^2 w(x - x_0, y - y_0)\mathrm{d}x\mathrm{d}y. \tag{19}$$

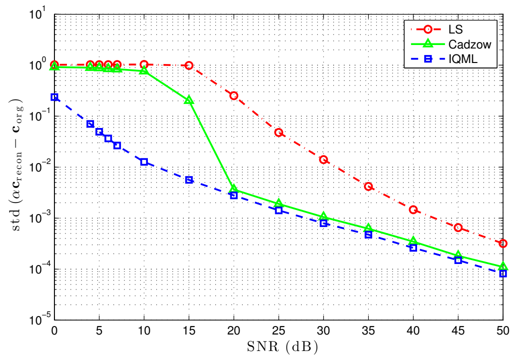


Fig. 5. Reconstruction errors for an FRI curve generated from 3×3 coefficients. The reconstruction errors under different levels of noise are averaged over 100 runs of different noise realizations.

Here $\mu_{\mathbf{c}}$ is a certain curve model parametrised by coefficients $\mathbf{c}$; I' is the derivative image; and $w(\cdot, \cdot)$ is a window to restrict the minimization to a localized block centred at (x_0, y_0). In the specific case, where we choose the mask function as a simple linear function: $\mu_{\mathbf{c}} = c_1 x + c_2 y + c_3$, then the solution of (19) can be implemented efficiently with a few linear filterings. We would like to explore whether we can bridge the global edge model (16) with the local one (19), which serves as further "refinements", in e.g., a multi-resolution setting.

References

[1] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.
[2] H. Pan, T. Blu, and P. L. Dragotti, "Sampling curves with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 458–471, 2014.
[3] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.
[4] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31–40, 2008.
[5] J. Berent, P. L. Dragotti, and T. Blu, "Sampling piecewise sinusoidal signals with finite rate of innovation methods," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 613–625, 2010.
[6] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 5, pp. 1081–1089, 1986.
[7] L. L. Scharf, *Statistical signal processing*. Addison-Wesley Reading, MA, 1991, vol. 98.
[8] P. Milanfar, G. C. Verghese, W. C. Karl, and A. S. Willsky, "Reconstructing polygons from moments with connections to array processing," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 432–443, 1995.
[9] P. J. Davis, "Triangle formulas in the complex plane," *Mathematics of Computation*, vol. 18, no. 88, pp. 569–577, 1964.
[10] I. Maravić and M. Vetterli, "Exact sampling results for some classes of parametric nonbandlimited 2-D signals," *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 175–189, 2004.
[11] P. Shukla and P. L. Dragotti, "Sampling schemes for multidimensional signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3670–3686, 2007.
[12] I. Maravić and M. Vetterli, "A sampling theorem for the Radon transform of finite complexity objects," in *2002 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, vol. 2. IEEE, 2002, pp. II–1197.
[13] C. Chen, P. Marziliano, and A. C. Kot, "2D finite rate of innovation reconstruction method for step edge and polygon signals in the presence of noise," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2851–2859, 2012.