Finite Rate of Innovation Sampling

Mathematical Foundations of Signal Processing

Dr. Matthieu Simeoni

November 30, 2020

EPFL

Table of contents

Introduction

Beyond Shannon Sampling Theorem

Sampling theorems lie at the foundation of modern digital signal processing: they permit the navigation between the analogue and digital worlds.

The most famous is undoubtedly the Shannon sampling theorem [1]. The latter states that bandlimited signals can be exactly recovered after uniform sampling with rate at least twice their maximum frequency and interpolation with a sinus cardinal kernel with same bandwidth.

This major result has had tremendous impact on the field of signal processing and by extension on many fields of natural sciences. But this unanimous celebration lead many scientists to start thinking about sampling theory exclusively in terms of bandlimitedness, which is only a sufficient condition for a signal to admit a discrete representation.

In fact, sampling theorems can also be devised for non-bandlimited signals as long as they possess finitely many degrees of freedom.

Signals with Finite Rates of Innovation (FRI)

Definition (Rate of Innovation of a Signal)

The rate of innovation of a signal is defined as its number of degrees of freedom, or innovations, per unit of time.

In this lesson, we will investigate sampling algorithms for signals with finite rates of innovation –i.e. finitely-many degrees of freedom per unit of time. Intuitively, if the signal is sampled below this critical rate not all the degrees of freedom will be fixed, leading to an underdetermined system and hence unidentifiability.

Example: (Rate of Innovation of Bandlimited Signals)

For a bandlimited signal with bandwidth B, Shannon theorem tells us than the signal can be identified with a sequence of samples spaced $T_s = 2\pi/B = T_{max}/2$ seconds apart.There is hence one degree of freedom every T_s seconds or a rate of innovation of $\rho = 1/T_s = 2/T_{max} = 2f_{max} = f_c$. Note that this rate of innovation coincides with the critical sampling rate announced by the theorem.

Stream of Diracs and the FRI Framework

Similarly, consider the prototypical sparse signal, a T-periodic stream of Diracs:

$$x(t) = \sum_{k' \in \mathbb{Z}} \sum_{k=1}^{K} x_k \delta(t - t_k - Tk'), \quad \forall t \in \mathbb{R}, \tag{1}$$

with $x_k \in \mathbb{C}$ and $t_k \in [0, T)$, $k = 1, \ldots, K$. It is easy to see that this signal has $2K$ innovations $\{(x_k, t_k)\}_{k=1,\ldots,K}$ over a period of time T, leading to a finite rate of innovation of $\rho = 2K/T$. Hence, despite being non-bandlimited, it seems that sampling some transformation of $x(t)$ at a finite rate ρ could, at least theoretically, yield a complete characterisation of the signal in terms of discrete measurements.

The classical FRI framework, introduced in [2], aims at estimating the innovations $\{(x_k, t_k), k = 1, \ldots, K\} \subset \mathbb{C} \times [0, T[$, of the T-periodic stream of Diracs (1).

The estimation procedure is divided into two stages. The locations t_k are first estimated by a nonlinear method, and then used to form a Vandermonde system whose solution yields the Dirac amplitudes x_k.

Stream of Diracs: Example

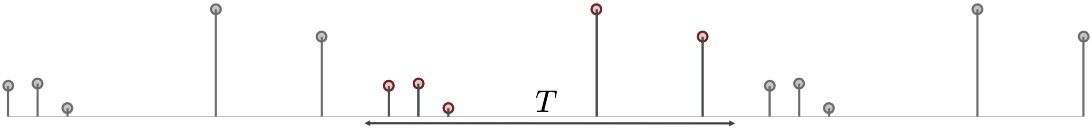


Figure: Example of Dirac stream with rate of innovation $\rho = 10/T$.

The Finite Rate of Innovation Framework

The Annihilating Filter and Annihilating Equation

The recovery of the locations t_k relies on the so-called annihilating equation which cancels out the Fourier series coefficients of x by convolving them with a particular filter, called the annihilating filter.

Definition (Annihilating Filter)

Let x be a Dirac stream with innovations $\{(x_k, t_k),\, k = 1, \ldots, K\} \subset \mathbb{C} \times [0, T[$. Then the annihilating filter of x is defined as the finite-tap sequence $h = [\cdots, 0, \boxed{h_0}, h_1, \ldots, h_K, 0, \cdots] \in \mathbb{C}^{\mathbb{Z}}$, with z-transform vanishing at roots $\left\{u_k := e^{-j2\pi t_k/T},\, k = 1, \ldots, K\right\}$:

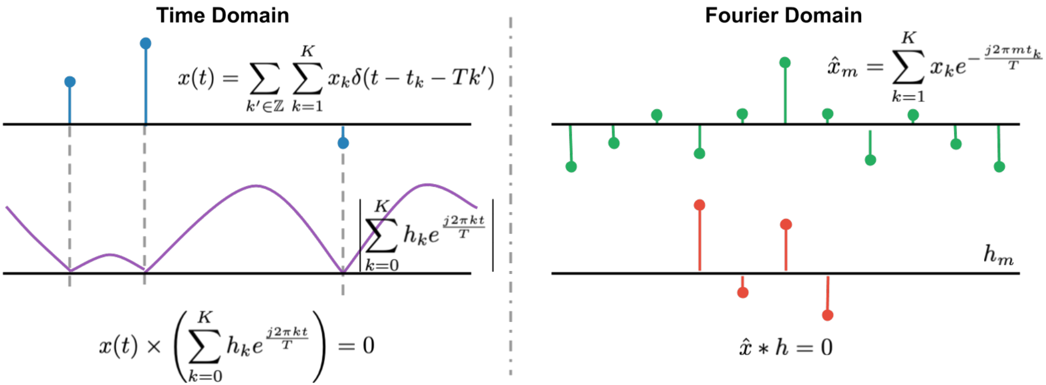
$$H(z) := \sum_{k=0}^{K} h_k z^{-k} = h_0 \prod_{k=1}^{K} (1 - u_k z^{-1}). \tag{2}$$

The annihilating filter satisfies the annihilating equation:

$$(\hat{x} * h)_m = \sum_{k=0}^{K} h_k \hat{x}_{m-k} = \sum_{k'=1}^{K} x_{k'} \left(\sum_{k=0}^{K} h_k u_{k'}^{-k} \right) u_{k'}^{m} = 0, \qquad \forall m \in \mathbb{Z}, \tag{3}$$

where $\hat{x}_m = \sum_{k=1}^{K} x_k u_k^m,\, m \in \mathbb{Z},$ are the Fourier coefficients of x in (1).

The Annihilating Filter and Annihilating Equation



Leveraging the Annihilating Equation

We can recover the Dirac locations t_k as follows:

1. Estimate the coefficients $\boldsymbol{h} = [h_0, \dots, h_K] \in \mathbb{C}^{K+1}$ of h by extracting $K+1$ independent equations from the annihilating equation (3).
2. Find the K roots u_k of the polynomial $\sum_{k=0}^{K} h_k z^k$ (using for example `numpy.roots()`).
3. Recover the Dirac locations t_k with:

$$t_k = \frac{T\theta(u_k)}{2\pi}, \qquad k = 1, \dots, K, \tag{4}$$

where $\theta : \mathbb{C} \to [0, 2\pi)$ maps a complex number to its phase modulo 2π.[1] From the definition of the annihilating filter we have indeed that $u_k = e^{-j2\pi t_k/T}$.

[1]See `numpy.angle()`.

Solving the Annihilating Equation

Assume that we have access to $N = 2M + 1$ consecutive Fourier coefficients of x, $\boldsymbol{x} = [\hat{x}_{-M}, \ldots, \hat{x}_M] \in \mathbb{C}^{2M+1}$, with $M \geq K$. Then, we can extract the $N - K$ equations from (3) corresponding to the convolution indices $m = -M + K, \ldots, M$, and form the following matrix equation:

$$\begin{bmatrix} \hat{x}_{-M+K} & \hat{x}_{-M+K-1} & \cdots & \hat{x}_{-M} \\ \hat{x}_{-M+K+1} & \hat{x}_{-M+K} & \cdots & \hat{x}_{-M+1} \\ \vdots & \ddots & \ddots & \vdots \\ \hat{x}_{M-1} & \hat{x}_{M-2} & \cdots & \hat{x}_{M-K-1} \\ \hat{x}_M & \hat{x}_{M-1} & \cdots & \hat{x}_{M-K} \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_K \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad \Longleftrightarrow \quad T_K(\boldsymbol{x})\boldsymbol{h} = \boldsymbol{0}_{N-K}. \tag{5}$$

The operator T_K, called the Toeplitzification operator, embeds $\boldsymbol{x}$ into the space $\mathbb{T}_K$ of Toeplitz matrices of $\mathbb{C}^{(N-K)\times(K+1)}$. It is defined as:

$$T_K : \begin{cases} \mathbb{C}^N & \to \mathbb{T}_K \subset \mathbb{C}^{(N-K)\times(K+1)} \\ \boldsymbol{x} & \mapsto [T_K(\boldsymbol{x})]_{i,j} := x_{-M+K+i-j}, \quad i = 1, \ldots, N-K,\ j = 1, \ldots, K+1. \end{cases} \tag{6}$$

Observe that the value of an entry of $T_K(\boldsymbol{x})$ depends only on the distance $i - j$ between the row and column indices: $T_K(\boldsymbol{x})$ is indeed a Toeplitz matrix and the vector $\boldsymbol{x}$ is called its generator.

Total Least-Squares

Note that $\boldsymbol{h} = \mathbf{0}_{K+1}$ is trivially a solution to (5). To avoid this degenerate case, we constrain $\|\boldsymbol{h}\| \neq 0$. Estimating the coefficients of the annihilating filter amounts then to solving:

$$T_K(\boldsymbol{x})\boldsymbol{h} = \mathbf{0}_{N-K}, \qquad \text{such that} \qquad \|\boldsymbol{h}\| \neq 0. \tag{7}$$

Observe that any nontrivial element of the nullspace of $T_K(\boldsymbol{x})$ is a solution to (7).

For $M \geq K$, it can be shown [3] that $T_K(\boldsymbol{x}) \in \mathbb{C}^{(N-K)\times(K+1)}$ has rank K and therefore has a nontrivial nullspace with dimension 1. Up to a multiplicative constant, the annihilating equation (7) admits hence a unique solution. The latter can be obtained numerically by means of total least-squares [3], which computes the eigenvector associated to the smallest[2] eigenvalue of $T_K(\boldsymbol{x})$.

In the critical case $M = K$, the matrix $T_K(\boldsymbol{x})$ is square, while in the oversampling case $M > K$ it is rectangular and tall. As explained in [3], oversampling makes the estimation procedure more resilient to potential noise perturbations in the Fourier coefficients.

[2] An eigenvalue exactly equal to zero may in practice be impossible to obtain due to numerical inaccuracies.

Estimating the Amplitudes

Once the coefficients $\boldsymbol{h} = [h_0, \dots, h_K]$ of the annihilating filter recovered by total least-squares, we can compute its roots $u_k = e^{-j2\pi t_k/T}$ and recover the Dirac locations t_k as explained on Slide 11.

To recover the Dirac amplitudes, we leverage the formula $\hat{x}_m = \sum_{k=1}^K x_k u_k^m$ (see Slide 9) which yields the following matrix equation:

$$\begin{bmatrix} u_1^{-M} & u_2^{-M} & \cdots & u_K^{-M} \\ u_1^{-M+1} & u_2^{-M+1} & \cdots & u_K^{-M+1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ u_1^{M-1} & u_2^{M-1} & \cdots & u_K^{M-1} \\ u_1^{M} & u_2^{M} & \cdots & u_K^{M} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{bmatrix} = \begin{bmatrix} \hat{x}_{-M} \\ \hat{x}_{-M+2} \\ \vdots \\ \hat{x}_0 \\ \vdots \\ \hat{x}_{M-1} \\ \hat{x}_M \end{bmatrix}. \tag{8}$$

Note that the entries of the right-hand side vector in (8) corresponds to the same $N = 2M + 1$ consecutive Fourier coefficients of x that were used to estimate the annihilating filter.

Estimating the Amplitudes (continued)

Notice that:

$$\begin{bmatrix} u_1^{-M} & u_2^{-M} & \cdots & u_K^{-M} \\ u_1^{-M+1} & u_2^{-M+1} & \cdots & u_K^{-M+1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ u_1^{M-1} & u_2^{M-1} & \cdots & u_K^{M-1} \\ u_1^{M} & u_2^{M} & \cdots & u_K^{M} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ u_1^{1} & u_2^{1} & \cdots & u_K^{1} \\ \vdots & \vdots & \cdots & \vdots \\ u_1^{M+1} & u_2^{M+1} & \cdots & u_K^{M+1} \\ \vdots & \vdots & \cdots & \vdots \\ u_1^{2M} & u_2^{2M} & \cdots & u_K^{2M} \\ u_1^{2M+1} & u_2^{2M+1} & \cdots & u_K^{2M+1} \end{bmatrix}}_{:=\boldsymbol{V}} \underbrace{\begin{bmatrix} u_1^{-M} & 0 & \cdots & 0 \\ 0 & u_2^{-M} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & u_K^{-M} \end{bmatrix}}_{:=\boldsymbol{D}},$$

which is the product of a transposed rectangular Vandermonde matrix[3] $\boldsymbol{V} \in \mathbb{R}^{(2M+1)\times K}$ and a diagonal matrix $\boldsymbol{D} \in \mathbb{R}^{K\times K}$ both full column ranks. Equation (8) hence admits indeed a solution.

[3] We have indeed $V_{ik} := u_k^{i-1}$, $i = 1,\ldots,2M+1$, $k = 1,\ldots,K$.

Finite Rate of Innovation Sampling

We have hence described a procedure for recovering the innovations $\{(x_k, t_k)\}$ of a Dirac stream x from $2M+1$ consecutive Fourier coefficients of x, $M \geq K$. In practice, these Fourier coefficients can be obtained via the following sampling scheme:

1. Pre-filter the signal $x(t)$ by an ideal low-pass filter with bandwidth $N = 2M+1$, and collect N samples $y_n \in \mathbb{R}$ from the resulting signal, taken uniformly at a sampling period $T_s = T/N$:

$$y_n = \sum_{k=1}^{K} x_k \varphi_N(nT_s - t_k) = \sum_{m=-M}^{M} T\hat{x}_m e^{j2\pi mn/N}, \quad n = 1, \ldots, N, \tag{9}$$

   where $\varphi_N(t) := \frac{\sin(N\pi t)}{NT\sin(\pi t/T)}$, $t \in \mathbb{R}$ is the T-periodic sinc function, or Dirichlet kernel.

2. Take the discrete Fourier transform (DFT) of the time samples (9). This yields, up to a multiplicative constant, the N consecutive Fourier coefficients $[\hat{x}_{-M}, \cdots, \hat{x}_M]$ of x:

$$\hat{y}_m = \sum_{n=1}^{N} y_n e^{-j2\pi nm/N} = \begin{cases} T\hat{x}_m, & \text{if } |m| \leq M, \\ 0, & \text{otherwise.} \end{cases}$$

This sampling scheme has sampling rate $F_s = 1/T_s = (2M+1)/T$. Note that when $M = K$, this is almost the rate of innovation $\rho = 2K/T$ of the Dirac stream.

Finite Rate of Innovation Sampling: Example

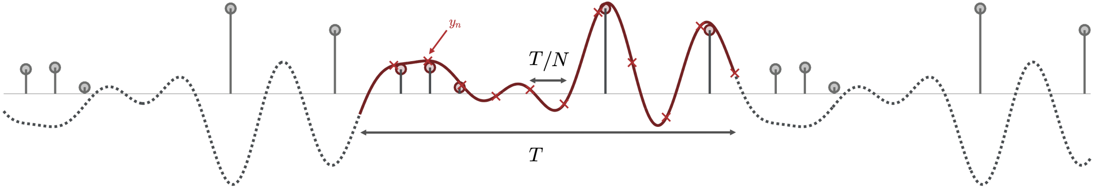


Figure: Sampling of a Dirac stream with rate of innovation $\rho = 10/T$. The sampling rate is $F_s = 11/T$.

Finite Rate of Innovation Sampling: Summary

Let $x(t) = \sum_{k' \in \mathbb{Z}} \sum_{k=1}^{K} x_k \delta(t - t_k - Tk')$. The latter can be sampled/reconstructed according to the following schemes:

- **Sampling:**
  1. Pre-filter the signal $x(t)$ by an ideal low-pass filter with bandwidth $N = 2M+1,\ M \geq K$;
  2. Sample uniformly the low-pass filtered signal with sampling rate $F_s = N/T$;
  3. Take the DFT of the time samples to get $2M+1$ consecutive Fourier coefficients $[\hat{x}_{-M}, \cdots, \hat{x}_M]$ of x.
- **Reconstruction:**
  1. Solve the annihilating equation (7) via total least-squares to get the annihilating filter $\boldsymbol{h} = [h_0, \ldots, h_K] \in \mathbb{C}^{K+1}$;
  2. Compute the K roots u_k of the polynomial $\sum_{k=0}^{K} h_k z^k$;
  3. Recover the Dirac stream innovations (x_k, t_k) from (4) and (8).

18

Cadzow Denoising

The Need for Denoising

In practical setups, the low-pass filtered time samples (9) are often polluted by noise. Consequently, the Fourier coefficients $\boldsymbol{x} = [\hat{x}_{-M}, \cdots, \hat{x}_M]$ obtained by taking the DFT of the time samples are noisy too. For strong noise perturbations, the annihilating equation

$$T_K(\boldsymbol{x})\boldsymbol{h} = \boldsymbol{0}_{N-K}, \qquad \text{such that} \qquad \|\boldsymbol{h}\| \neq 0,$$

may fail to admit a nontrivial solution. Indeed, noisy generators $\boldsymbol{x}$ can yield full column rank matrices $T_K(\boldsymbol{x})$ with trivial nullspace.

A potential cure consists in denoising the Fourier coefficients $\boldsymbol{x}$ prior to solving the annihilating equation. This denoising step attempts to transform $T_K(\boldsymbol{x})$ into a Toeplitz matrix with rank at most K, thus guaranteeing the existence of nontrivial solutions to the annihilating equation.

The Method of Alternating Projections (MAP)

The method of alternating projections (MAP) is used in computational mathematics to approximate projections onto intersecting sets. In its simplest form proposed by von Neumann in 1933 [4], the MAP performs a cascade of n projection steps onto subsets $\{\mathscr{M}_1, \ldots, \mathscr{M}_L\}$ of some Hilbert space $\mathscr{H}$, starting from a point $z \in \mathscr{H}$:

$$\check{z} = \left[\Pi_{\mathscr{M}_L} \cdots \Pi_{\mathscr{M}_1}\right]^n (z). \tag{10}$$

In (10), $\Pi_{\mathscr{M}_k}$ denotes the projection map onto $\mathscr{M}_k$, defined for $k = 1, \ldots, L$ as

$$\Pi_{\mathscr{M}_k} : \begin{cases} \mathscr{H} \to \mathscr{M}_k, \\ z \mapsto \arg\min\limits_{x \in \mathscr{M}_k} \|z - x\|, \end{cases}$$

for some norm $\|\cdot\|$ on $\mathscr{H}$.

Convergence of the MAP

In the case of closed linear subspaces $\{\mathscr{M}_1, \ldots, \mathscr{M}_L\}$, von Neumann and Halperin showed that [4, 5, 6]

$$\lim_{n\to\infty} \left\| \left[\Pi_{\mathscr{M}_L} \cdots \Pi_{\mathscr{M}_1}\right]^n (z) - \Pi_{\bigcap_{k=1}^L \mathscr{M}_k}(z) \right\| = 0, \quad \forall z \in \mathscr{H}. \tag{11}$$

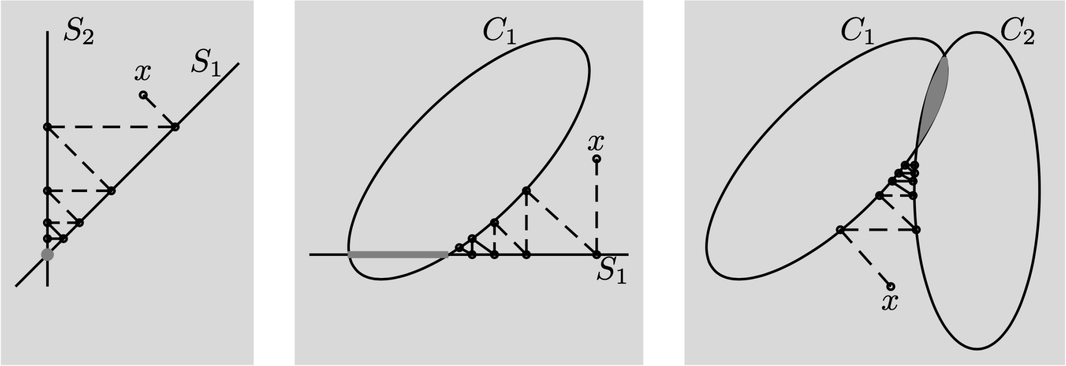
The MAP equation (10) can therefore be used to approximate the complex projection map $\Pi_{\bigcap_{k=1}^L \mathscr{M}_k}$.

For closed convex sets $\{\mathscr{M}_1, \ldots, \mathscr{M}_L\}$, Bregman [7] showed moreover the convergence of the MAP $\left[\Pi_{\mathscr{M}_L} \cdots \Pi_{\mathscr{M}_1}\right]^n (z)$ towards a point[4] in the intersection $\check{z} \in \bigcap_{k=1}^L \mathscr{M}_k$.

In the case of non convex intersecting sets, the MAP is often used as a heuristic with no convergence guarantees.

[4]This point may not necessarily be the projection $\Pi_{\bigcap_{k=1}^L \mathscr{M}_k}(z)$ however. Convergence towards the actual projection is achieved by Dysktra's MAP [8], one of the most popular variant to von Neumann's original algorithm.

Illustration of MAP



(a) $C = S_1 \cap S_2$. (b) $C = C_1 \cap S_1$. (c) $C = C_1 \cap C_2$.

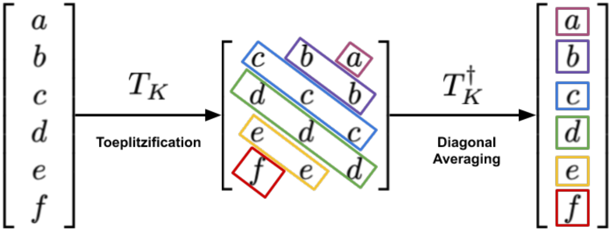
Figure: Iterative solution of finding the nearest vector in a convex set using MAP. (a) The convex sets are affine subspaces. (b) Intersection of a general convex set and an affine subspace. (c) Intersection of two general convex sets.

Cadzow denoising

The standard denoiser in FRI is called Cadzow denoising [9]. The latter processes the noisy coefficients $\boldsymbol{x}$ as follows:

$$\check{\boldsymbol{x}} = T_K^{\dagger}\left[\Pi_{\mathbb{T}_K}\Pi_{\mathscr{H}_K}\right]^n T_K(\boldsymbol{x}). \tag{12}$$

The operator $T_K^{\dagger} : \mathbb{C}^{(N-K)\times(K+1)} \to \mathbb{C}^N$ in (12) is the Moore-Penrose pseudoinverse of the Toeplitzification operator T_K. The latter maps a Toeplitz matrix on its generator by averaging across each diagonal of the matrix.



$$T_K^{\dagger} T_K(\mathbf{x}) = \mathbf{x}$$

Cadzow denoising (continued)

The operators $\Pi_{\mathbb{T}_K} = T_K T_K^\dagger$ and $\Pi_{\mathscr{H}_K}$ in (12) are the projections onto respectively the subspace $\mathbb{T}_K$ of Toeplitz matrices and the subset $\mathscr{H}_K$ of matrices with rank at most K:

$$\mathscr{H}_K := \left\{ \boldsymbol{M} \in \mathbb{C}^{(N-K)\times(K+1)} \mid \operatorname{rank} \boldsymbol{M} \leq K \right\}. \tag{13}$$

The projection operator onto the space $\mathscr{H}_K$ of matrices with rank at most K is given by the Eckart-Young-Minsky theorem [10]. The latter states that the projection map

$$\Pi_{\mathscr{H}_K}(\boldsymbol{X}) = \arg\min_{\boldsymbol{H}\in\mathscr{H}_K} \|\boldsymbol{X} - \boldsymbol{H}\|_F, \quad \boldsymbol{X} \in \mathbb{C}^{(N-K)\times(K+1)}, \tag{14}$$

can be computed in closed-form as:

$$\Pi_{\mathscr{H}_K}(\boldsymbol{X}) = \boldsymbol{U}\boldsymbol{\Lambda}_K\boldsymbol{V}^H, \quad \boldsymbol{X} \in \mathbb{C}^{(N-K)\times(K+1)}, \tag{15}$$

where $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^H$ is the singular value decomposition of $\boldsymbol{X}$, and $\boldsymbol{\Lambda}_K$ is the diagonal matrix of sorted singular values truncated to the K strongest ones.

Cadzow Denoising as a MAP

Note that (12) leverages a MAP to approximate the complex projection $\Pi_{\mathscr{H}_K \cap \mathbb{T}_K}$:

$$\Pi_{\mathscr{H}_K \cap \mathbb{T}_K} \simeq \left[\Pi_{\mathbb{T}_K} \Pi_{\mathscr{H}_K}\right]^n, \quad n \in \mathbb{N}.$$

If this approximation is sufficiently good (n large enough) then the denoised Fourier coefficients $\check{x}$ should generate a Toeplitz matrix with rank at most K, hence guaranteeing the feasibility of the annihilating equation.

Since $\mathscr{H}_K$ is a non convex set the convergence of the MAP above is however not guaranteed. Nevertheless, experimental results [3, 9] suggest that Cadzow denoising almost always converges after a few iterations (typically $n \leq 20$).

Cadzow Denoising: Example I [3]

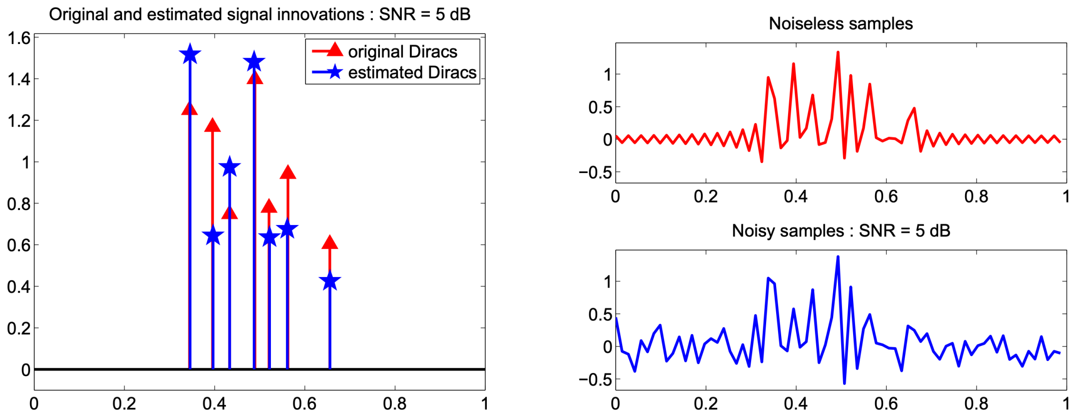
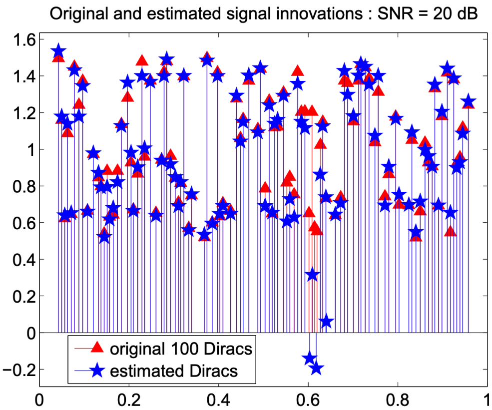


Figure: Retrieval of an FRI signal with 7 Diracs (left) from 71 noisy (SNR = 5 dB) samples (right).

Cadzow Denoising: Example II [3]



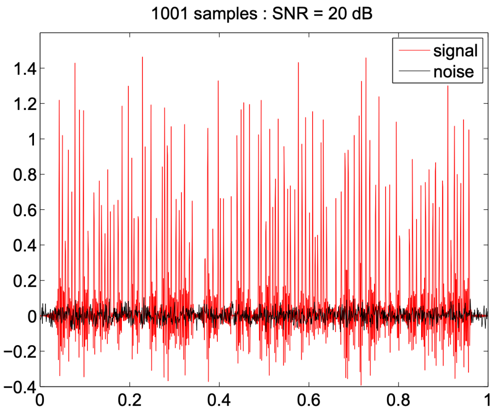


Figure: Retrieval of an FRI signal with 100 Diracs (left) from 1001 noisy (SNR = 20 dB) samples (right).

Generalised FRI Sampling

Generalised FRI Sampling

We have seen that the innovations of a Dirac stream could be recovered from consecutive Fourier coefficients, obtained by applying a discrete Fourier transform to the low-pass filtered uniform time samples.

In some applications however, the measurements available to us may not consist in low-pass filtered time samples. In which case, the situation is more complex, and the Fourier coefficients $\boldsymbol{x} \in \mathbb{C}^N$ must in general be estimated from the measurements $\boldsymbol{y} \in \mathbb{C}^L$ by solving a linear inverse problem:

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{x} + \boldsymbol{n}, \tag{16}$$

where the application-dependent forward matrix $\boldsymbol{G} \in \mathbb{C}^{L \times N}, L \geq N$, is assumed injective, and $\boldsymbol{n}$ is additive noise, usually assumed to be a white Gaussian random vector.

The Explicit Generalised FRI Problem

In [11], Pan *et al.*proposed the *explicit* generalised FRI (genFRI) optimisation problem to deal with (16). The latter is a non convex constrained optimisation problem whose objective is to jointly recover the Fourier coefficients $\boldsymbol{x} \in \mathbb{C}^N$ –required to minimise a quadratic data-fidelity term– and their corresponding annihilating filter coefficients $\boldsymbol{h} \in \mathbb{C}^{K+1}$. The annihilating equation linking the two unknowns is explicitly enforced as a constraint, yielding an optimisation problem of the form:

$$\min_{\substack{\boldsymbol{x}\in\mathbb{C}^N \\ \boldsymbol{h}\in\mathbb{C}^{K+1}}} \|\boldsymbol{G}\boldsymbol{x}-\boldsymbol{y}\|_2^2 \quad \text{subject to} \quad \begin{cases} T_K(\boldsymbol{x})\,\boldsymbol{h} = \boldsymbol{0}_{N-K}, \\ \langle \boldsymbol{h}, \boldsymbol{h}_0 \rangle = 1, \end{cases} \tag{17}$$

where $\boldsymbol{h}_0 \in \mathbb{C}^{K+1}$ is an arbitrary vector generated randomly.

The annihilating constraint regularises the genFRI problem by making sure that the recovered Fourier coefficients $\boldsymbol{x}$ can indeed be annihilated. The normalisation constraint[5] $\langle \boldsymbol{h}, \boldsymbol{h}_0 \rangle = 1$ is used to exclude trivial solutions to the annihilating equation in (17) [11, 12].

[5] In [11], the authors have also considered the more natural normalisation constraint $\|\boldsymbol{h}\| = 1$. They claim however that this normalisation strategy is less successful experimentally.

The Implicit Generalised FRI Problem

The annihilating equation constraint in (17) complicates significantly the optimisation procedure. Indeed, it requires the introduction of an extra unknown variable with non linear dependency on the data, namely the annihilating filter $\boldsymbol{h}$.

To circumvent this issue, Simeoni *et al.* proposed in [13] an implicit formulation of the genFRI problem, in which only the Fourier coefficients are recovered:

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} \|\boldsymbol{G}\boldsymbol{x}-\boldsymbol{y}\|_2^2 \quad \text{subject to} \quad \operatorname{rank} T_K(\boldsymbol{x}) \leq K. \tag{18}$$

The regularising rank constraint on $T_K(\boldsymbol{x})$ guarantees that solutions to (18) can be annihilated. Unlike (17), this constraint does not explicitly involve the unknown annihilating filter.

This implicit regularisation hence greatly simplifies the genFRI problem: it decouples the problem of estimating the Fourier coefficients from the problem of estimating the annihilating filter.

Optimisation Algorithm

The optimisation problem (18) can be rewritten in unconstrained form as:

$$\min_{\boldsymbol{x}\in\mathbb{C}^N} \underbrace{\|\boldsymbol{G}\boldsymbol{x}-\boldsymbol{y}\|_2^2}_{:=F(\boldsymbol{x})} + \underbrace{\iota_{\mathcal{H}_K}(T_K(\boldsymbol{x}))}_{:=H(\boldsymbol{x})}, \tag{19}$$

where $\mathcal{H}_K$ is the non convex set of matrices with rank lower than or equal to K defined in (13), and $\iota_{\mathcal{H}_K} : \mathbb{C}^{(N-K)\times(K+1)} \to \{0,+\infty\}$, is the indicator function of $\mathcal{H}_K$. It is possible to optimise (19) by means of proximal gradient descent (PGD), which proceeds iteratively as follows:

$$\boldsymbol{x}_{k+1} = \text{prox}_{\tau H}\left(\boldsymbol{x}_k - \tau\nabla F(\boldsymbol{x}_k)\right) = \text{prox}_{\tau H}\left(\boldsymbol{x}_k - \tau \boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{x}_k - \boldsymbol{y})\right), \tag{20}$$

for $k \geq 0$, $\boldsymbol{x}_0 \in \mathbb{C}^N$, and $\tau > 0$. We have the following convergence result [13, Theorem 2]:

Theorem: (Convergence of PGD for Injective G)

Assume that $\tau < 1/\beta$ with $\beta = 2\|\boldsymbol{G}\|_2^2$ and $\boldsymbol{G} \in \mathbb{C}^{L\times N}$ in (19) is injective, i.e. $\mathcal{N}(\boldsymbol{G}) = \{\boldsymbol{0}_N\}$. Then, any *limit point* $\boldsymbol{x}_\star$ of the sequence $\{\boldsymbol{x}_k\}_{k\in\mathbb{N}}$ generated by (20) is a *local minimum* of (19).

Cadzow Plug-and-Play Gradient Descent (CPGD)

The proximal step is hard to compute exactly but can be approximated via Cadzow denoising [13, Section IV.B]:

$$\mathrm{prox}_{\tau H}(\boldsymbol{x}) = \underset{\boldsymbol{z}\in\mathbb{C}^N}{\arg\min} \frac{1}{2\tau}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 + \iota_{\mathscr{H}_K}(T_K(\boldsymbol{z})) \simeq T_K^\dagger \left[\Pi_{\mathbb{T}_K}\Pi_{\mathscr{H}_K}\right]^n T_K(\boldsymbol{x}), \quad \forall \boldsymbol{x}\in\mathbb{C}^N, \tag{21}$$

for some $n \geq 0$. This yields an inexact PGD method, called Cadzow Plug-and-Play Gradient Descent:

Algorithm 1 Cadzow PnP Gradient Descent (CPGD)

1: **procedure** CPGD($\tau > 0, \boldsymbol{x}_0 \in \mathbb{C}^N, K \in \mathbb{N}, n \in \mathbb{N}$)
2:     **for all** $n \geq 1$ **do**
3:         $\boldsymbol{z}_{k+1} := \boldsymbol{x}_k - 2\tau \boldsymbol{G}^H(\boldsymbol{G}\boldsymbol{x}_k - \boldsymbol{y})$
4:         $\boldsymbol{x}_{k+1} := T_K^\dagger \left[\Pi_{\mathbb{T}_K}\Pi_{\mathscr{H}_K}\right]^n T_K(\boldsymbol{z}_{k+1})$
5:     **return** $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$

Such an approach is reminiscent of the plug-and-play (PnP) framework [14, 15] from image processing, which leverages generic denoisers to approximate complex proximal operators [16].

Generalised FRI: Irregular Time Samples [11]

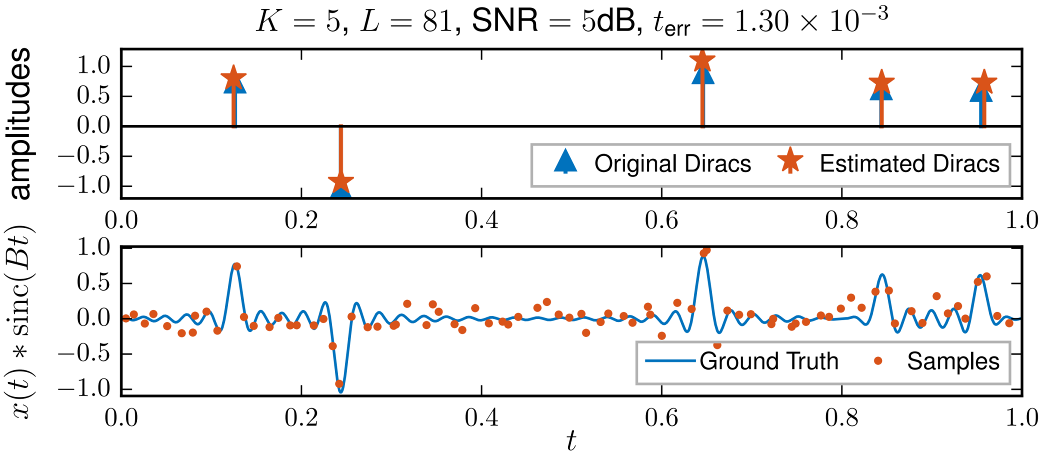


Figure: Reconstruction of a stream of Diracs from ideally low-pass filtered samples taken at irregular time instances.

Generalised FRI: Irregular Fourier Samples [11]

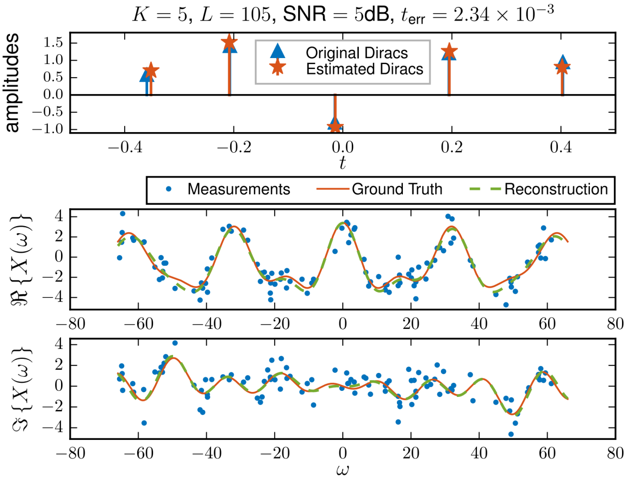


Figure: Reconstruction of weighted Diracs from non-uniform Fourier samples.

Generalised FRI: Direction of Arrival (DOA) Estimation [17]

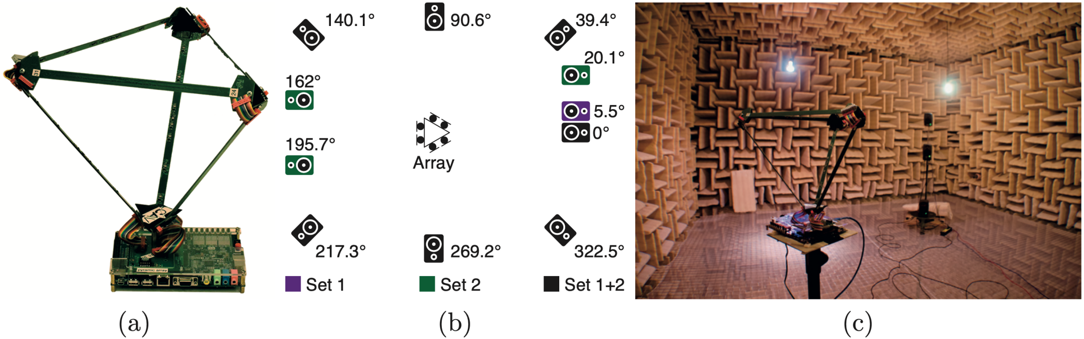
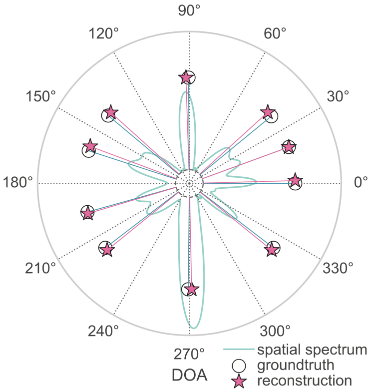
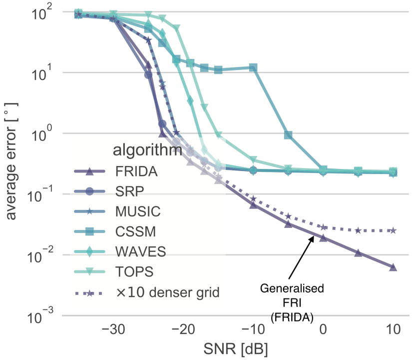


Figure: (a) Microphone array with 48 microphones. (b) Locations of the loudspeakers and microphone array in experiments. (c) Anechoic chamber used for the experiments.

Generalised FRI: Direction of Arrival (DOA) Estimation [17]



(a) Reconstruction of 10 acoustic sources.



(b) Average DOA reconstruction error.

References I

[1] Claude Elwood Shannon.
A mathematical theory of communication.
ACM SIGMOBILE Mobile Computing and Communications Review, 5(1):3–55, 2001.

[2] M. Vetterli, P. Marziliano, and T. Blu.
Sampling signals with finite rate of innovation.
IEEE Trans. Signal Process., 50(6):1417–1428, 2002.

[3] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot.
Sparse sampling of signal innovations.
IEEE Signal Process. Mag., 25(2):31–40, 2008.

[4] John von Neumann.
The geometry of orthogonal spaces, functional operators-vol. ii.
Annals of Math. Studies, 22, 1950.

[5] Heinz Bauschke, Frank Deutsch, Hein Hundal, and Sung-Ho Park.
Accelerating the convergence of the method of alternating projections.
Transactions of the American Mathematical Society, 355(9):3433–3461, 2003.

References II

[6] Israel Halperin.
The product of projection operators.
Acta Sci. Math.(Szeged), 23(1):96–99, 1962.

[7] René Escalante and Marcos Raydan.
***Alternating projection methods*.**
SIAM, 2011.

[8] Heinz H Bauschke and Jonathan M Borwein.
Dykstra's alternating projection algorithm for two sets.
Journal of Approximation Theory, 79(3):418–443, 1994.

[9] Laurent Condat and Akira Hirabayashi.
Cadzow denoising upgraded : A new projection method for the recovery of dirac pulses from noisy linear measurements, 2015.

[10] Carl Eckart and Gale Young.
The approximation of one matrix by another of lower rank.
Psychometrika, 1(3):211–218, 1936.

References III

[11] Hanjie Pan, Thierry Blu, and Martin Vetterli.
Towards generalized FRI sampling with an application to source resolution in radioastronomy.
IEEE Trans. Signal Process., 65(4):821–835, 2017.

[12] Hanjie Pan, Matthieu Simeoni, Paul Hurley, Thierry Blu, and Martin Vetterli.
Leap: Looking beyond pixels with continuous-space estimation of point sources.
Astronomy & Astrophysics, 608:A136, 2017.

[13] Matthieu Simeoni, Adrien Besson, Paul Hurley, and Martin Vetterli.
Cpgd: Cadzow plug-and-play gradient descent for generalised fri.
arXiv preprint arXiv:2006.06374, 2020.

[14] Ernest K Ryu, Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin.
Plug-and-play methods provably converge with properly trained denoisers.
In *International Conference on Machine Learning*, pages 5546–5557, 2019.

[15] Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg.
Plug-and-play priors for model based reconstruction.
In *2013 IEEE Global Conference on Signal and Information Processing*, pages 945–948. IEEE, 2013.

References IV

[16] Harshit Gupta, Kyong Hwan Jin, Ha Q Nguyen, Michael T McCann, and Michael Unser.
Cnn-based projected gradient descent for consistent ct image reconstruction.
IEEE transactions on medical imaging, 37(6):1440–1453, 2018.

[17] Hanjie Pan, Robin Scheibler, Eric Bezzam, Ivan Dokmanić, and Martin Vetterli.
FRIDA: FRI-based DOA estimation for arbitrary array layouts.
In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3186–3190. IEEE, 2017.