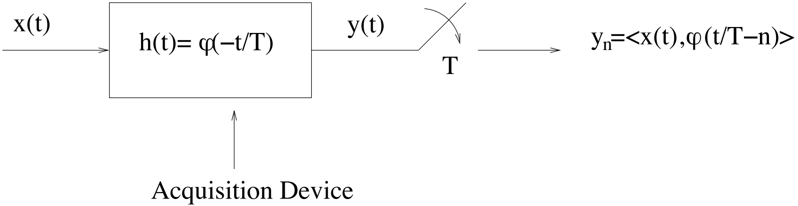
Imperial College London

Sparse Sampling: Theory and Applications

Pier Luigi Dragotti
Imperial College London

The Sampling Problem

You are given a class of functions. You have a sampling device. Given the measurements $y_n = \langle x(t), \varphi(t/T - n) \rangle$, you want to reconstruct $x(t)$.



Natural questions:

- When is there a one-to-one mapping between $x(t)$ and y_n?
- What signals can be sampled and what kernels $\varphi(t)$ can be used?
- What reconstruction algorithm?

A few Interesting Applications: Digital Imaging

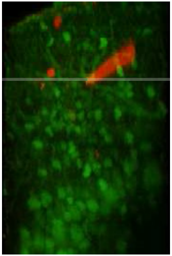
Real Scene

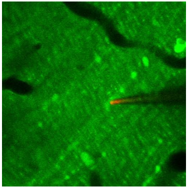
Digital Image

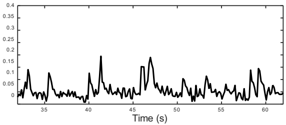
- The real scene is **analogue** and is turned into a **digital** image by the camera
- Can we overcome the limitation of the camera and, given the pixels, obtain a sharper image with arbitrary resolution?
- Sparsity: Edges and contours can be modelled using a small number of parameters

Application: Neuroscience

Goal: Understanding how information processing occurs in neural circuits.

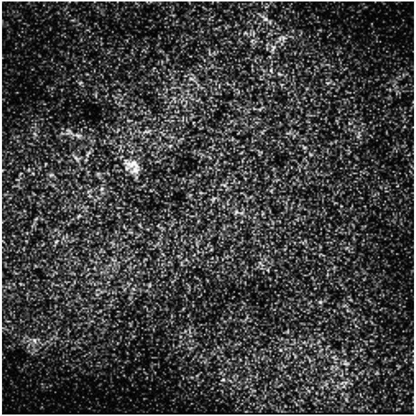






- Brain activity is **analogue** and it is observed with **digital** two-photon microscopes
- Neurones communicate through pulses called Action Potentials (AP)
- The problem of reconstructing the activity of a neurone is sparse when the shape of the AP is approximately known.

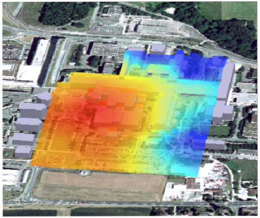
Application: Neuroscience



This is a multidimensional problem that requires simultaneously segmentation, denoising and time-resolution enhancement.

Note: Video from Simon Peron, Jeremy Freeman, Vijay Iyer, Karel Svoboda (2014); Calcium imaging data from vibrissal S1 neurons in adult mice performing a pole localization task.

Application: Estimation of Diffusion Fields



- Can we localise diffusion sources or acoustic sources using sensor networks?
- Application:
  1. Monitor dispersion in factories producing bio-chemicals
  2. Check whether your government is lying ;-)
- The diffusion field is **analogue**, spatio-temporal sensor measurements are **digital**

Problem Statement

What do all these problems have in common?

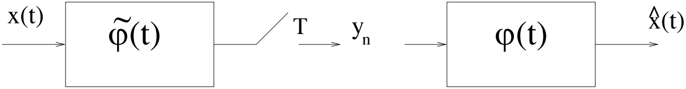
- The world is **analogue**, computation is **digital**
- The source is continuous in time and/or space, measurements are discrete (e.g., pixels in a camera, sensors measurements)
- The **data formation process** is well understood and involves **deterministic** smoothing functions normally known a priori (e.g., point spread function in a camera, the diffusion kernel for diffusion fields)
- **Strong priors** on the source to be estimated

Our Approach

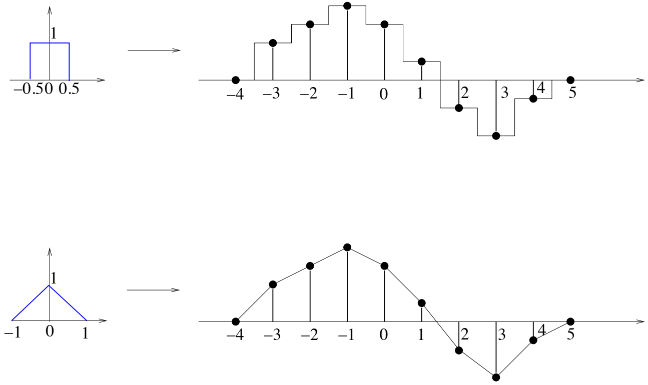
- From the samples, using the knowledge of the observation process, estimate proper integral measurements of the source (e.g., estimate the Fourier transform at specific frequencies, or estimate a few moments of the function)
- Given the integral measurements, solve the inverse/reconstruction problem using proper sparsity priors

Classical Sampling Formulation

- Sampling of $x(t)$ is equivalent to projecting $x(t)$ into the shift-invariant subspace $V = \text{span}\{\varphi(t/T-n)\}_{n\in\mathbb{Z}}$.
- If $x(t) \in V$, perfect reconstruction is possible.
- Reconstruction process is linear: $\hat{x}(t) = \sum_n y_n \varphi(t/T - n)$.
- For bandlimited signals $\varphi(t) = \text{sinc}(t)$.



Classical Sampling Formulation



Signals with Finite Rate of Innovation

Consider a signal of the form:

$$x(t) = \sum_{k \in \mathbb{Z}} \sum_{r=0}^{R-1} \gamma_{k,r} g_r(t - t_k). \tag{1}$$

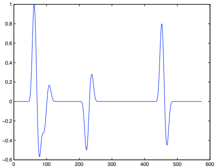
The rate of innovation of $x(t)$ is then defined as

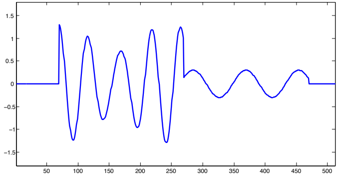
$$\rho = \lim_{\tau \to \infty} \frac{1}{\tau} C_x \left(-\frac{\tau}{2}, \frac{\tau}{2} \right). \tag{2}$$

Definition [VetterliMB:02] A signal with a finite rate of innovation is a signal whose parametric representation is given in (1) and with a finite ρ as defined in (2).

Signals with Finite Rate of Innovation

Consider a continuous-time stream of pulses or a piecewise regular signal.





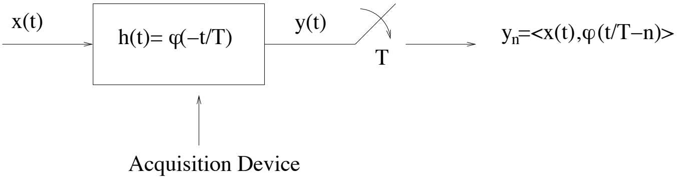
These signals

- are not bandlimited.
- are not sparse in a basis or a frame.

However:

- they are completely determine by a finite number of free parameters.

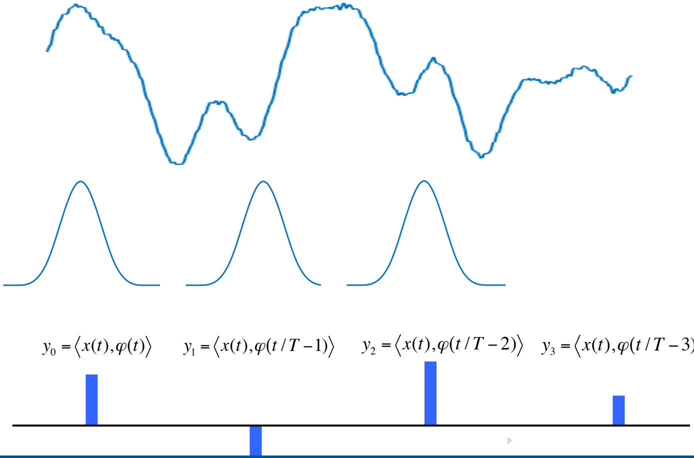
The Sampling Kernel



Note that

- $y(t) = h(t) * x(t) = \int_{-\infty}^{\infty} x(\tau) h(t-\tau) d\tau$
- $$\begin{aligned} y_n &= y(nT) = \int_{-\infty}^{\infty} x(\tau) h(nT-\tau) d\tau = \int_{-\infty}^{\infty} x(\tau) \varphi(\tau/T - n) d\tau \\ &= \langle x(t), \varphi(t/T - n) \rangle \end{aligned}$$
- $\varphi(t)$ is the time reversed version of the acquisition device and is called **sampling kernel**.

Sampling Process



$$y_0 = \langle x(t), \varphi(t) \rangle \quad y_1 = \langle x(t), \varphi(t/T-1) \rangle \quad y_2 = \langle x(t), \varphi(t/T-2) \rangle \quad y_3 = \langle x(t), \varphi(t/T-3) \rangle$$

Sampling Kernels

Possible classes of kernels (we want to be as general as possible):

1. Any kernel $\varphi(t)$ that can reproduce polynomials:

$$\sum_n c_{m,n}\varphi(t-n) = t^m \quad m = 0, 1, \ldots, L,$$

   for a proper choice of coefficients $c_{m,n}$.

2. Any kernel $\varphi(t)$ that can reproduce exponentials:

$$\sum_n c_{m,n}\varphi(t-n) = e^{a_m t} \quad a_m = \alpha_0 + m\lambda \text{ and } m = 0, 1, \ldots, L,$$

   for a proper choice of coefficients $c_{m,n}$.

3. Any kernel with rational Fourier transform:

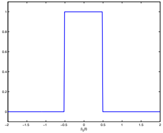
$$\hat{\varphi}(\omega) = \frac{\prod_i (j\omega - b_i)}{\prod_m (j\omega - a_m)} \quad a_m = \alpha_0 + m\lambda \text{ and } m = 0, 1, \ldots, L.$$

Sampling Kernels

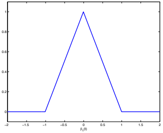
Class 1 is made of all functions $\varphi(t)$ satisfying Strang-Fix conditions. This includes any scaling function generating a wavelet with $L+1$ vanishing moments (e.g., Splines and Daubechies scaling functions).
Strang-Fix Conditions:

$$\hat{\varphi}(0) \neq 0 \text{ and } \hat{\varphi}^{(m)}(2n\pi) = 0 \text{ for } n \neq 0 \text{ and } m = 0, 1, .., L,$$

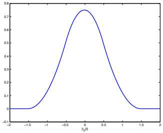
where $\hat{\varphi}(\omega)$ is the Fourier transform of $\varphi(t)$.



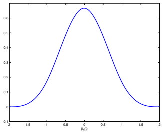
$\beta_0(t)$



$\beta_1(t)$



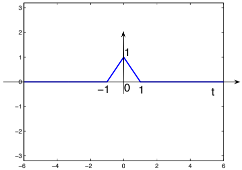
$\beta_2(t)$



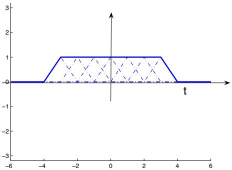
$\beta_3(t)$

B-splines satisfy Strang-Fix conditions since $\hat{\beta}_n(\omega) = \text{sinc}(\omega/2)^{n+1}$.

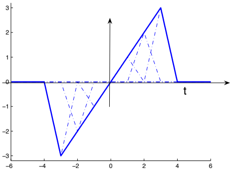
Sampling Kernels



$\beta_1(t)$



$c_{0,n} = (1, 1, 1, 1, 1, 1, 1)$

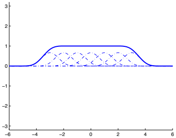


$c_{1,n} = (-3, -2, -1, 0, 1, 2, 3)$

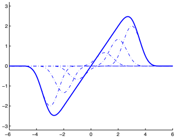
The linear spline reproduces polynomials up to degree L=1: $\sum_n c_{m,n}\beta_1(t-n) = t^m \quad m = 0, 1,$ for a proper choice of coefficients $c_{m,n}$ (in this example $n = -3, -2, ..., 1, 2, 3$).

Notice: $c_{m,n} = \langle \tilde{\varphi}(t-n), t^m \rangle$ where $\tilde{\varphi}(t)$ is biorthogonal to $\varphi(t)$: $\langle \tilde{\varphi}(t), \varphi(t-n) \rangle = \delta_n$.

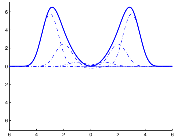
Sampling Kernels



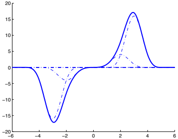
$c_{0,n} = (1,1,1,1,1,1,1)$



$c_{1,n} = (-3,-2,-1,0,1,2,3)$



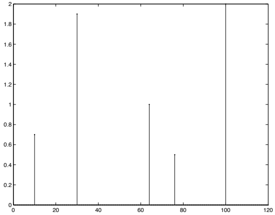
$c_{2,n} \sim (8.7, 3.7, 0.7, -0.333, 0.7, 3.7, 8.7)$



$c_{3,n} \sim (-24, -6, -0.001, 0, 0.001, 6, 24)$

The cubic spline reproduces polynomials up to degree L=3: $\sum_n c_{m,n}\beta_3(t-n) = t^m \quad m = 0,1,2,3.$

Sampling Streams of Diracs



- Assume that $x(t)$ is a stream of K Diracs on the interval $[0, \tau[$:

$$x(t) = \sum_{k=0}^{K-1} x_k \delta(t - t_k).$$

  The signal has $2K$ degrees of freedom.

- Assume that $\varphi(t)$ is any function that can reproduce polynomials of maximum degree $L \geq 2K - 1$:

$$\sum_n c_{m,n} \varphi(t - n) = t^m, \quad m = 0, 1, ..., L,$$

  where $c_{m,n} = \langle \tilde{\varphi}(t - n), t^m \rangle$ and $\tilde{\varphi}(t)$ is biorthogonal to $\varphi(t)$.

- We want to retrieve $x(t)$, from the samples $y_n = \langle x(t), \varphi(t/T - n) \rangle$ with $n = 0, 1, ..., N - 1$ and $TN = \tau$.

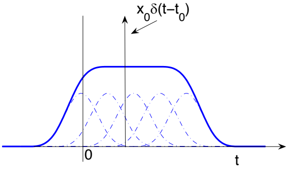
Sampling Streams of Diracs

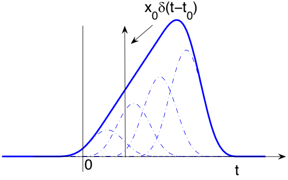
Assume for simplicity that $T = 1$ and define $s_m = \sum_n c_{m,n} y_n$, $m = 0, 1, \ldots, L$, we have that

$$\begin{aligned} s_m &= \sum_{n=0}^{N-1} c_{m,n} y_n \\ &= \langle x(t), \sum_{n=0}^{N-1} c_{m,n} \varphi(t-n) \rangle \\ &= \int_{-\infty}^{\infty} x(t) t^m dt \\ &= \sum_{k=0}^{K-1} x_k t_k^m \quad m = 0, 1, \ldots, L \end{aligned} \tag{3}$$

We thus observe $s_m = \sum_{k=0}^{K-1} x_k t_k^m$, $m = 0, 1, \ldots, L$.

Sampling Streams of Diracs

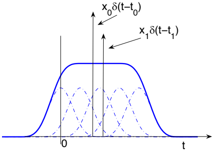




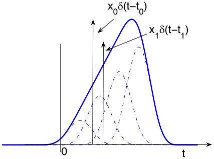
$$\sum_n y_n = \langle x_0\delta(t-t_0), \sum_n \varphi(t-n)\rangle = \int_{-\infty}^{\infty} x_0\delta(t-t_0)\sum_n \varphi(t-n)dt = x_0\sum_n \varphi(t_0-n) = x_0$$

$$\sum_n c_{1,n}y_n = \langle x_0\delta(t-t_0), \sum_n c_{1,n}\varphi(t-n)\rangle = x_0\sum_n c_{1,n}\varphi(t_0-n) = x_0 t_0$$

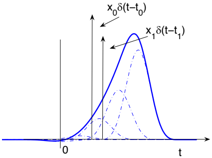
Sampling Streams of Diracs



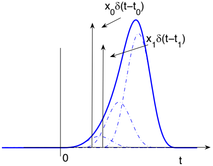
$$s_0 = \sum_n y_n = x_0 + x_1$$



$$s_1 = \sum_n c_{1,n} y_n = x_0 t_0 + x_1 t_1$$



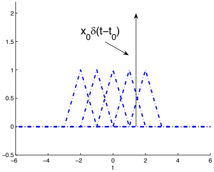
$$s_2 = \sum_n c_{2,n} y_n = x_0 t_0^2 + x_1 t_1^2$$

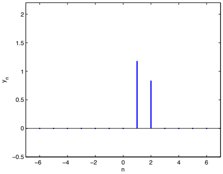


$$s_3 = \sum_n c_{3,n} y_n = x_0 t_0^3 + x_1 t_1^3$$

Sampling Streams of Diracs

Example: $T = 1$, $x(t) = x_0\delta(t - t_0)$ with $x_0 = 2$ and $t_0 = \sqrt{2}$.
The sampling kernel is a linear spline.





In this case

$$y_n = \langle x(t), \varphi(t - n)\rangle = x_0\varphi(t_0 - n) = \begin{cases} 0 & n \neq 1, 2 \\ 2(2 - \sqrt{2}), & n = 1, \\ 2(\sqrt{2} - 1) & n = 2. \end{cases}$$

and

$$\begin{aligned} s_0 &= \textstyle\sum_n c_{0,n} y_n = \sum_n y_n = y_1 + y_2 = x_0 = 2, \\ s_1 &= \textstyle\sum_n c_{1,n} y_n = \sum_n n y_n = y_1 + 2y_2 = x_0 t_0 = 2\sqrt{2}. \end{aligned}$$

Prony's Method

The quantity

$$s_m = \sum_{k=0}^{K-1} x_k t_k^m, \quad m = 0, 1, ..., L$$

is a power-sum series.

We can retrieve the locations t_k and the amplitudes x_k with the annihilating filter method (also known as Prony's method since it was discovered by Gaspard de Prony in 1795).

Overview of Prony's Method

1. Call h_m the filter with z-transform $H(z) = \sum_{i=0}^{K} h_i z^{-i} = \prod_{k=0}^{K-1}(1 - t_k z^{-1})$.

We have that

$$h_m * s_m = \sum_{i=0}^{K} h_i s_{m-i} = \sum_{i=0}^{K} \sum_{k=0}^{K-1} x_k h_i t_k^{m-i} = \sum_{k=0}^{K-1} x_k t_k^m \underbrace{\sum_{i=0}^{K} h_i t_k^{-i}}_{0} = 0.$$

This filter is thus called the annihilating filter.
In matrix/vector form we have that $\mathbf{S}H = \mathbf{H}S = 0$ or alternatively using the fact that $h_0 = 1$

$$\begin{bmatrix} s_{K-1} & s_{K-2} & \cdots & s_0 \\ s_K & s_{K-1} & \cdots & s_1 \\ \vdots & \vdots & \ddots & \vdots \\ s_{L-1} & s_{L-2} & \cdots & s_{L-K} \end{bmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_K \end{pmatrix} = - \begin{pmatrix} s_K \\ s_{K+1} \\ \vdots \\ s_L \end{pmatrix}.$$

Solve the above Toeplitz system to find the coefficient of the annihilating filter.

Overview of Prony's Method

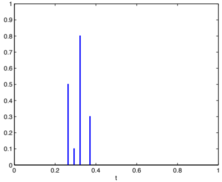
2. Given the coefficients $\{1, h_1, h_2, ..., h_k\}$, we get the locations t_k by finding the roots of $H(z)$.

3. Solve the first K equations in $s_m = \sum_{k=0}^{K-1} x_k t_k^m$ to find the amplitudes x_k. In matrix/vector form

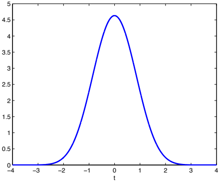
$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_0 & t_1 & \cdots & t_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ t_0^{K-1} & t_1^{K-1} & \cdots & t_{K-1}^{K-1} \end{bmatrix} \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{K-1} \end{pmatrix} = \begin{pmatrix} s_0 \\ s_1 \\ \vdots \\ s_{K-1} \end{pmatrix}. \tag{4}$$

Classic Vandermonde system. Unique solution for distinct t_ks.

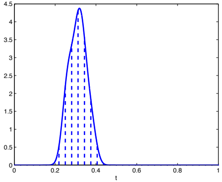
Sampling Streams of Diracs: Numerical Example



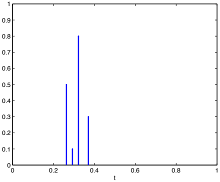
(a) Original Signal



(b) Sampling Kernel ($\beta_7(t)$)

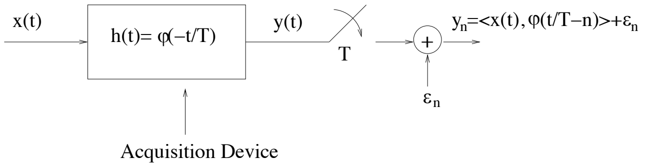


(c) Samples



(d) Reconstructed Signal

The Noisy Scenario



- The measurements are noisy
- The noise is additive and i.i.d. Gaussian
- The sampling kernel is the sinc function or B-splines or E-splines.

Total Least Squares Algorithm

The annihilation equation

$$\mathbf{S}H = 0$$

is only approximately satisfied.

Minimize: $\|\mathbf{S}H\|_2$ under the constraint $\|H\|_2 = 1$.

This is achieved by performing an SVD of $\mathbf{S}$:

$$\mathbf{S} = \mathbf{U}\lambda\mathbf{V}^{\mathbf{T}}.$$

Then H is the last column of $\mathbf{V}$.

Notice: this is similar to Pisarenko's method in spectral estimation.

Cadzow's Algorithm

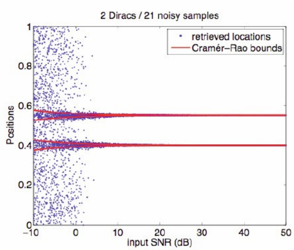
For small SNR use Cadzow's method to denoise $\mathbf{S}$ before applying TLS.

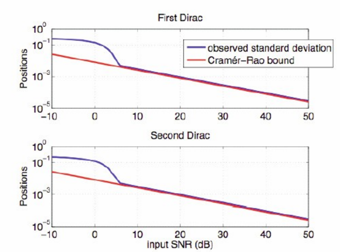
The basic intuition behind this method is that, in the noiseless case, $\mathbf{S}$ is rank deficient (rank K) and Toeplitz, while in the noisy case $\mathbf{S}$ is full rank.

Algorithm:

- SVD of $\mathbf{S} = \mathbf{U}\lambda\mathbf{V}^{\mathbf{T}}$.
- Keep the K largest diagonal coefficients of λ and set the others to zero.
- Reconstruct $\mathbf{S}' = \mathbf{U}\lambda'\mathbf{V}^{\mathbf{T}}$.
- This matrix is not Toeplitz, make it so by averaging along the diagonals.
- Iterate.

Simulation Results





- Samples are corrupted by additive noise.
- This is a parametric estimation problem.
- Unbiased algorithms have a covariance matrix lower bounded by CRB.
- The proposed algorithm reaches CRB down to SNR of 5dB.

Application: Image Super-Resolution

Super-Resolution is a multichannel sampling problem with unknown shifts. Use moments to retrieve the shifts or the geometric transformation between images.

(a)Original (512×512)

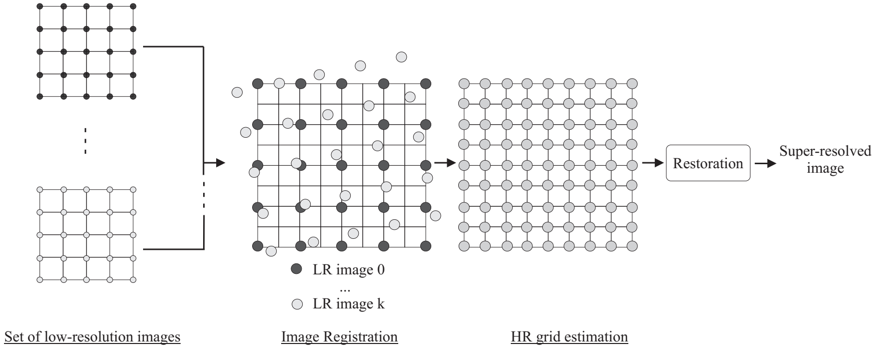
(b) Low-res. (64×64)

(c) Super-res (PSNR=24.2dB)

- Forty low-resolution and shifted versions of the original.
- Intuition: The disparity between images has a finite rate of innovation and can be retrieved exactly.
- Accurate registration is achieved by retrieving the continuous moments of the 'Tiger' from the samples.
- The registered images are interpolated to achieve super-resolution.

Application: Image Super-Resolution

Image super-resolution basic building blocks



Application: Image Super-Resolution

- For each blurred image $I(x, y)$:
  - A sample $P_{m,n}$ in the blurred image is given by

$$P_{m,n} = \langle I(x, y), \varphi(x/T - n, y/T - m) \rangle,$$

    where $\varphi(t)$ represents the point spread function of the lens.
  - We assume $\varphi(t)$ is a spline that can reproduce polynomials:

$$\sum_n \sum_m c_{m,n}^{(l,j)} \varphi(x - n, y - m) = x^l y^j \quad l = 0, 1, ..., N; j = 0, 1, ..., N.$$

  - We retrieve the exact moments of $I(x, y)$ from $P_{m,n}$:

$$\tau_{l,j} = \sum_n \sum_m c_{m,n}^{(l,j)} P_{m,n} = \langle I(x, y), \sum_n \sum_m c_{m,n}^{(l,j)} \varphi(x/2^J - n, y/2^J - m) \rangle = \int \int I(x, y) x^l y^j dx dy.$$

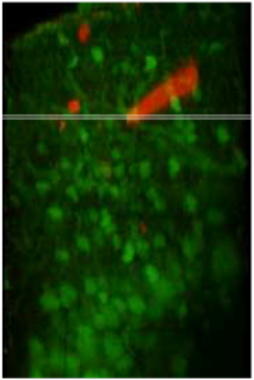
- Given the moments from two or more images, we estimate the geometrical transformation and register them. Notice that moments of up to order three along the x and y coordinates allows the estimation of an affine transformation.

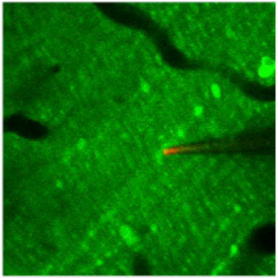
Application: Image Super-Resolution

(a)Original (128×128)

(b) Super-res (1024×1024)

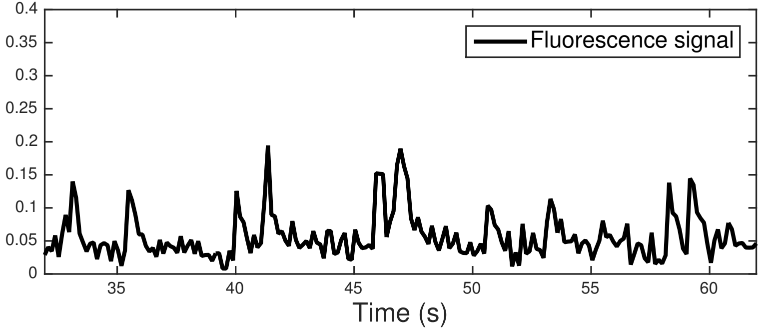
Sparse Sampling and Neuroscience





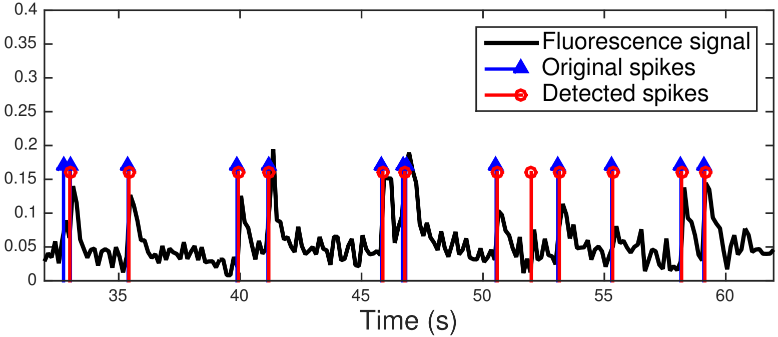
[OnativiaSD:13, ReynoldsSD:17]

Sparse Sampling and Neuroscience

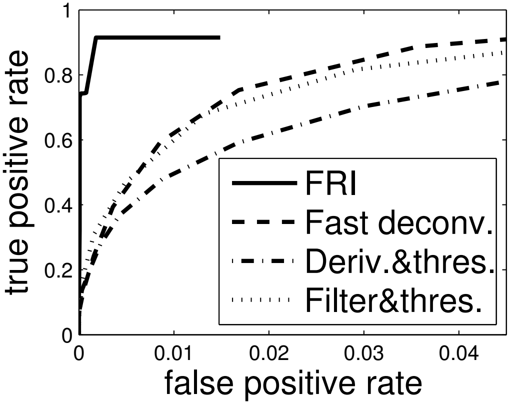


Signal Model: Stream of decaying exponentials

Sparse Sampling and Neuroscience

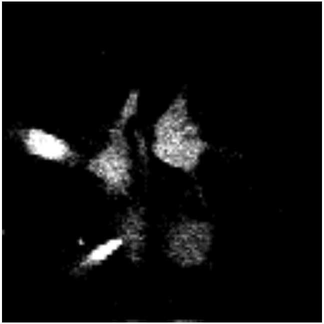


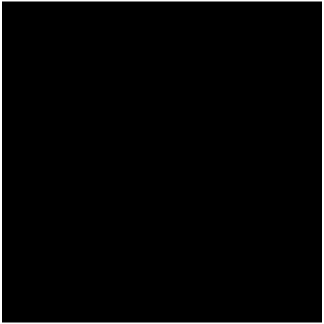
Sparse Sampling and Neuroscience



- The algorithm **outperforms** state-of-the art methods
- Can operate in **real-time** simultaneously on 80 streams
- Increase in **resolution** by factor 3

Segmentation and Calcium Transient Detection





[ReynoldsSD:17]

Nobel Prizes in Signal Processing ;-)

- 1908 (phy), Gabriel Lippmann
- 1914 (phy), Max von Laue (X-ray diffraction crystallography)
- 1953 (phy), Frits Zernike (phase-contrast microscopy)
- 1962 (med), Watson, Crick and Wilkins (DNA)
- 1971 (phy), Dennis Gabor (holographic method)
- 1979 (med), Allan M. Cormack, Godfrey N. Hounsfield (CAT)
- 1991 (chem), Richard R. Ernst (high-resolution NMR spectroscopy)
- 2002 (chem), Kurt Wüthrich (3D NMR spectroscopy)
- 2003 (med), Paul C. Lauterbur and Sir Peter Mansfield (MRI)
- 2014 (chem), Eric Betzig, Stefan W. Hell, William E. Moerner (super-resolved fluorescence microscopy)
- 2017 (chem), Jacques Dubochet (cryo-electron microscopy)

- One of the legacy of wavelet and sampling theory is the emergence of **computational imaging** which refers to techniques that combine the design of the hardware layer with signal processing techniques in order to go beyond physical limitations of traditional image formation systems.
- In computational imaging, computation plays an integral role in the image formation process, for this reason, this research area is intimately related to sampling theory and aspects of sparse representation.

Single Image Super-Resolution: Numerical Comparisons

(a) Original | (b) Linear (25.9dB) | (c) A+ (27.3dB) | (d) FRESH (27.7dB)

FRESH Results: Real Data

Low-res input
64 x64 pixels

Final result
256x256 pixels

Conclusions

Sparse Sampling:

- New framework that allows the sampling and reconstruction of continuous-time non-bandlimited signals.
- Use the knowledge of the acquisition process to map discrete measurements to specific integral measurements

Key Insights:

- No need to discretize the input signals (use the mathematics of Hilbert and shift-invariant spaces)
- Non-convex reconstruction algorithms (i.e., structured least-squares methods) often superior to other methods
- Be creative with the notion of sparsity!
- Carve the right solution for the problem at hand: discrete/continuous, convex/non-convex, noise/no-noise

Outlook: This is the era of inverse problem! Be prepared to witness and contribute to surprising new results and applications!

References

On sampling

- J. Uriguen, T. Blu, and P.L. Dragotti ‘FRI Sampling with Arbitrary Kernels’, IEEE Trans. on Signal Processing, November 2013

- T. Blu, P.L. Dragotti, M. Vetterli, P. Marziliano and L. Coulot ‘Sparse Sampling of Signal Innovations: Theory, Algorithms and Performance Bounds,’ IEEE Signal Processing Magazine, vol. 25(2), pp. 31-40, March 2008

- P.L. Dragotti, M. Vetterli and T. Blu, ‘Sampling Moments and Reconstructing Signals of Finite Rate of Innovation: Shannon meets Strang-Fix’, IEEE Trans. on Signal Processing, vol.55 (5), pp.1741-1757, May 2007.

- H. Pan, T. Blu, and P.L. Dragotti, ‘Sampling Curves with Finite Rate of Innovation’ IEEE Trans. on Signal Processing, January 2014.

- X. Wei and P.L. Dragotti, Guaranteed Performance in the FRI setting, IEEE Signal Processing Letters, 2015

References (cont'd)

On Single-Image Image Super-Resolution

- X. Wei and P.L. Dragotti, FRESH -FRI-based single image super-resolution algorithm, IEEE Trans on Image Processing, Vol.25(8), pp. 3723-3735, August 2016.

On Calcium Transient Detection

- Jon Onativia, Simon R. Schultz, and Pier Luigi Dragotti, A Finite Rate of Innovation algorithm for fast and accurate spike detection from two-photon calcium imaging, Journal of Neural Engineering, August 2013.
- S. Reynolds, C.S. Copeland, S. Schultz and P.L. Dragotti, An Extension of the FRI framework for Calcium Transient Detection, International Symposium on BioMedical Imaging, (ISBI), Prague, Czech Republic, April 2016.

On Diffusion Fields and Sensor Networks

- John Murray-Bruce and Pier Luigi Dragotti, Estimating localized sources of diffusion fields using spatiotemporal sensor measurements, IEEE Trans. on Signal Processing, June 2015.