Sparse Sampling: Theory, Algorithms and Applications

Pier Luigi Dragotti

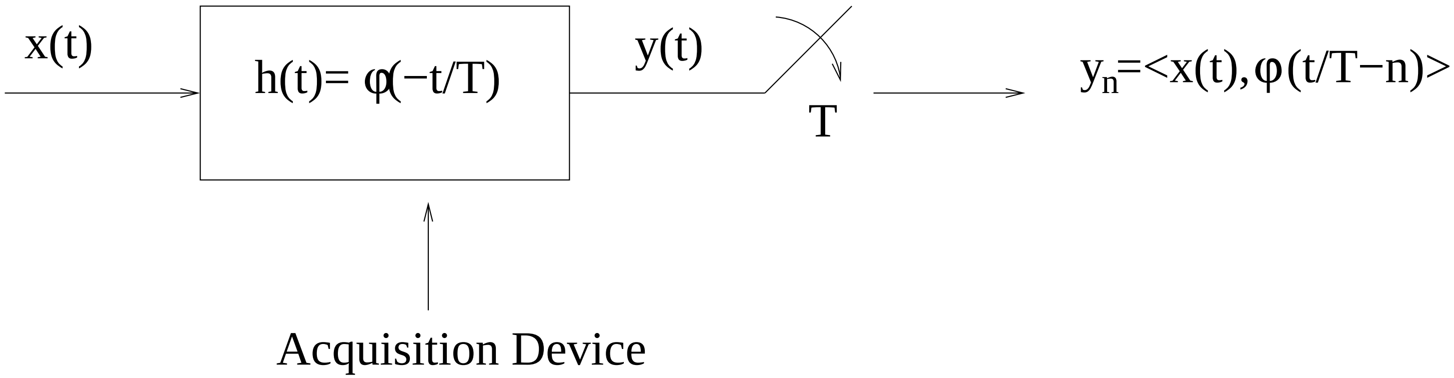
Communications and Signal Processing Group
Imperial College London

Outline

1. Problem Statement and Motivation
2. Signals with Finite Rate of Innovation (FRI)
3. Kernels Reproducing Polynomials
4. Sampling FRI Signals with Polynomial Reproducing Kernels
5. Noisy Scenario
6. Application: Image Super-Resolution and Neuroscience
7. Conclusions and Outlook

Problem Statement and Motivation

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

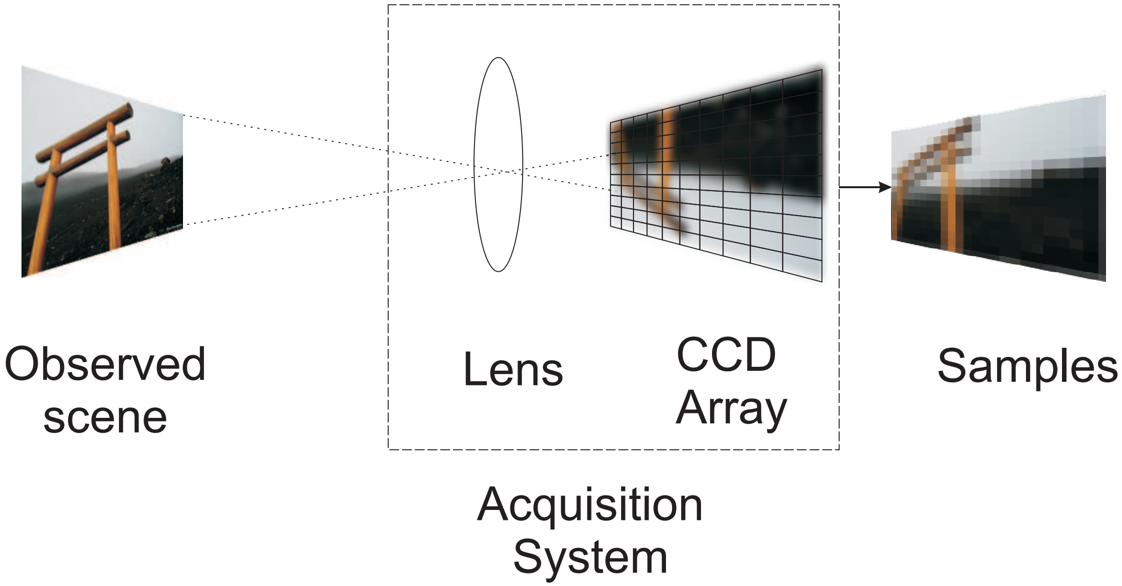


Natural questions:

- 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 algorithms?

Problem Statement and Motivation

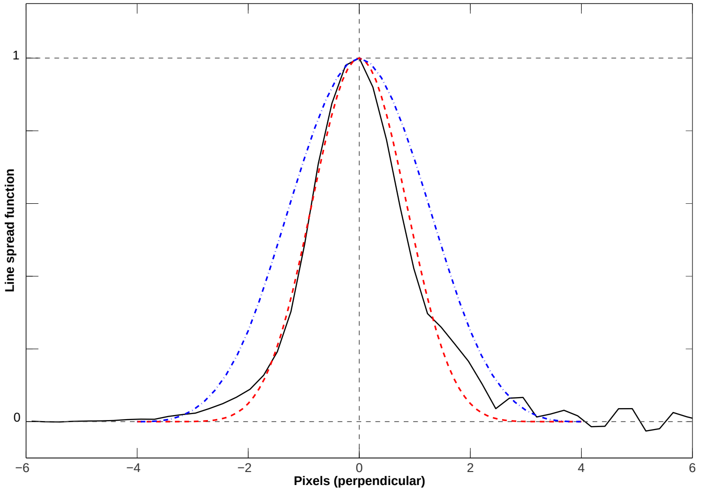
Digital Cameras:



- The low-quality lens blurs the images. Lenses are, to some extent, equivalent to the sampling kernel. Lenses do not necessarily behave like the sinc function.

Problem Statement and Motivation

(a)Original (2014×3039)

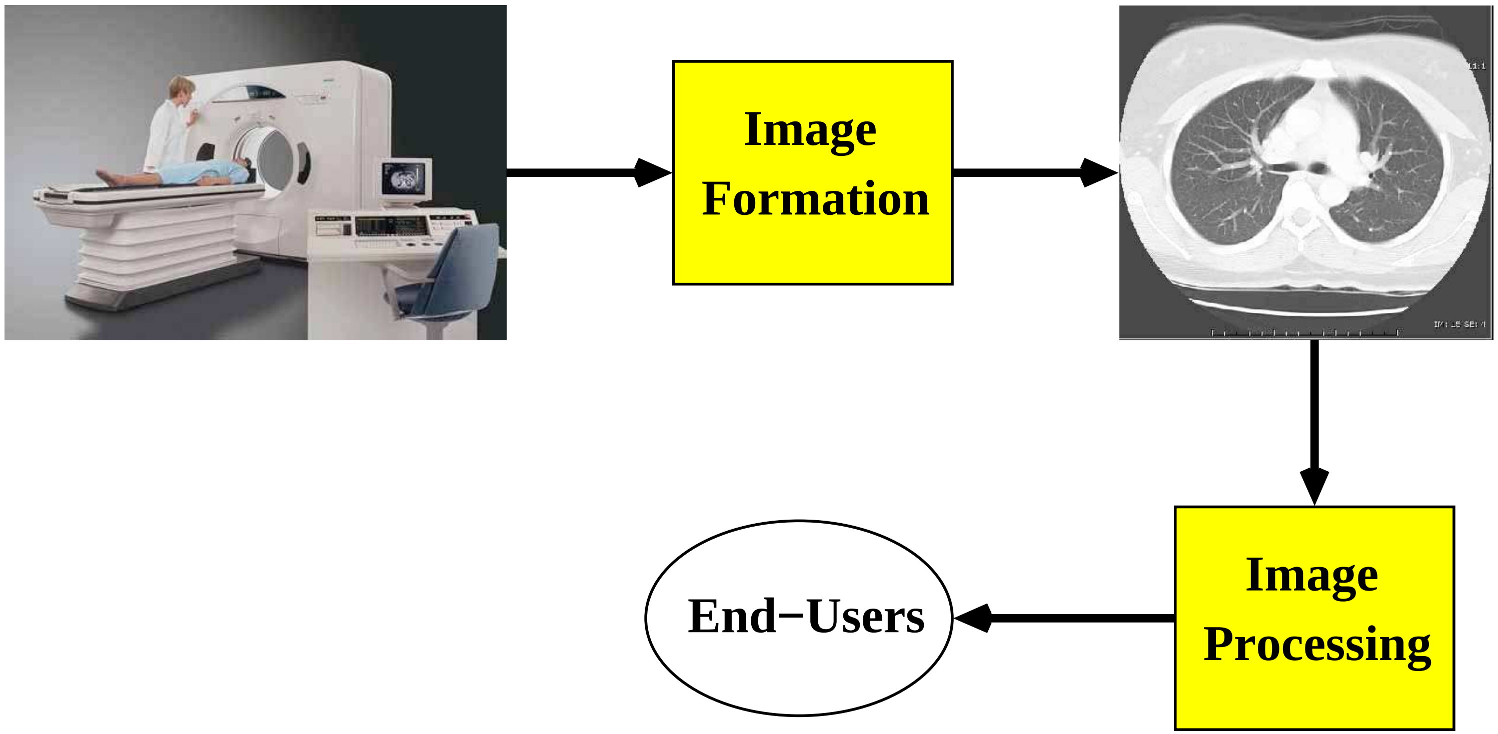


(b) Point Spread function in black

The distortion introduced by a lens can be well approximated with splines.

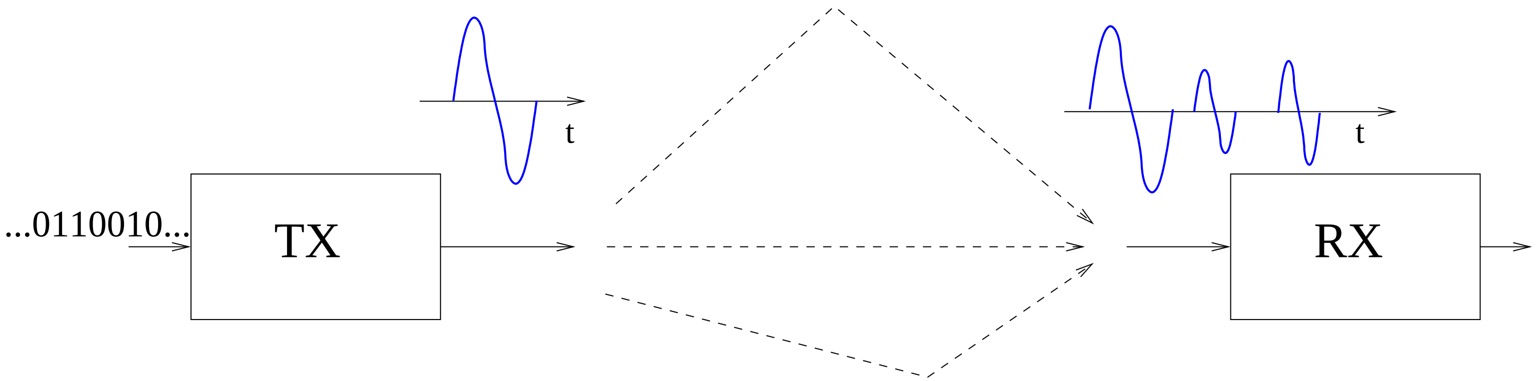
Problem Statement and Motivation

"In 2005, the U.S. spent 16% of its GDP on health care. It is projected that this will reach 20% by 2015." Goal: Individualized treatments based on low-cost and effective medical devices.



Problem Statement and Motivation

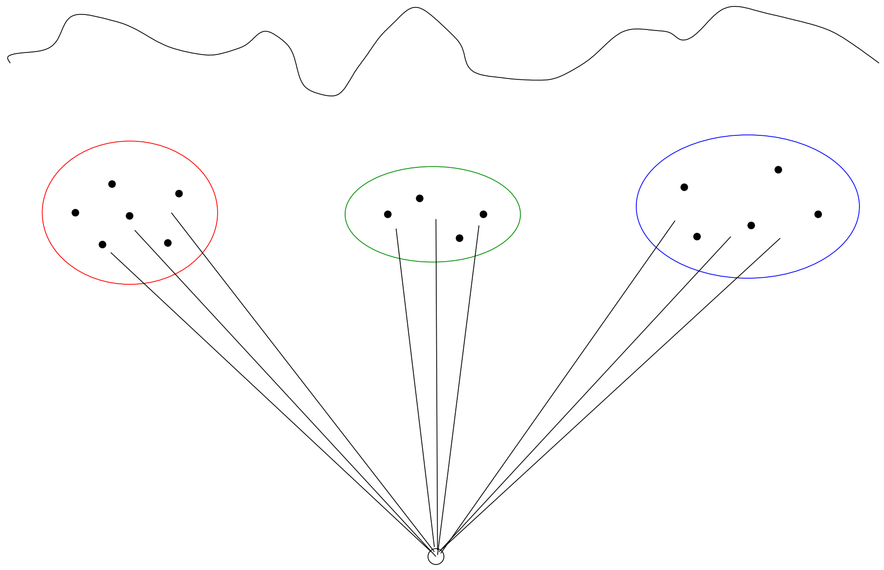
Wide-Band Communications:



Current A-to-D converters in UWB communications operate at several gigaherz.

This is a parametric estimation problem, only the location and amplitude of the pulses need to be estimated.

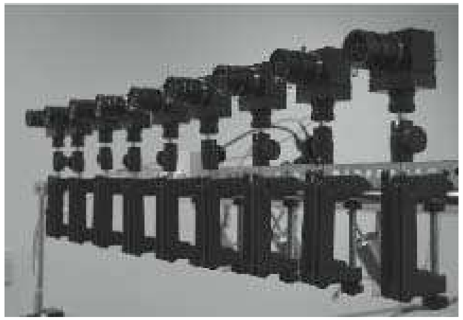
Problem Statement: Distributed Sensing of Diffusive Sources

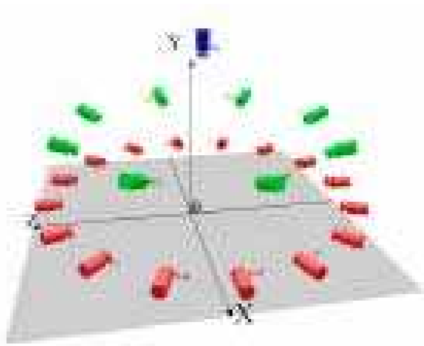


- The source (phenomenon) is distributed in space and time.
- The phenomenon is sampled in space (finite number of sensors) and time.
- Sensors need to be low-power.

Emerging Application: Free Viewpoint Video

Multiple cameras are used to record a dynamic scene. Users can freely choose an arbitrary viewpoint for 3D viewing.





- Standard approaches are supervised, model-based and use a relatively small number of high-resolution cameras (6-8 cameras). They do not achieve photorealistic rendering.
- This is a multi-dimensional sampling problem.

Classical Sampling Formulation

The Shannon sampling theorem provides sufficient but not necessary conditions for perfect reconstruction.

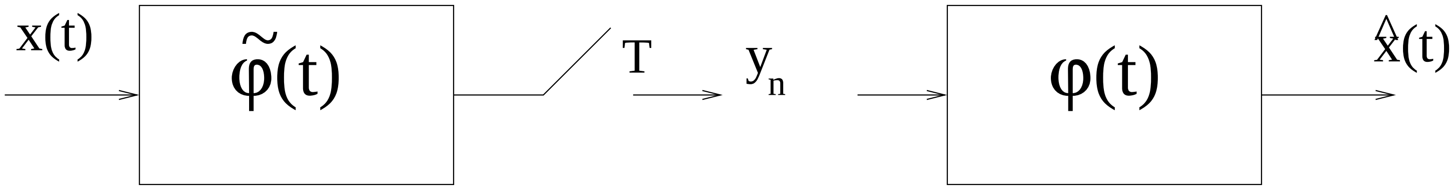
Moreover: How many real signals are bandlimited? How many realizable filters are ideal low-pass filters?

By the way, who discovered the sampling theorem? The list is long ;-)

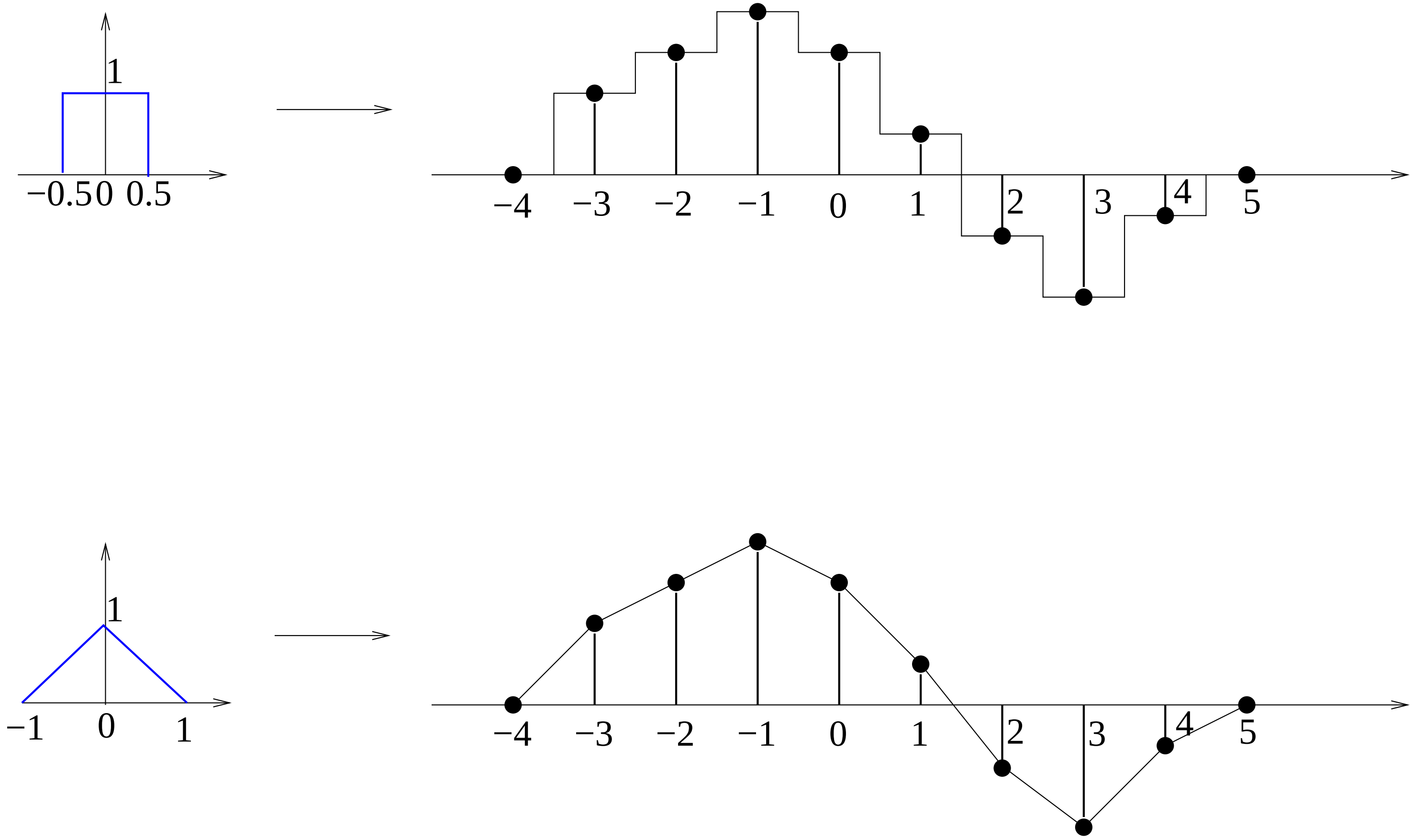
- Whittaker 1915, 1935
- Kotelnikov 1933
- Nyquist 1928
- Raabe 1938
- Gabor 1946
- Shannon 1948
- Someya 1948

Sub-Space Interpretation of the Sampling Theorem

- 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 $\varphi(t) = \text{sinc}(t)$.



Sampling as Projecting into Shift-Invariant Sub-Spaces



Signals with Finite Rate of Innovation

Can we go beyond this subspace interpretation and extend the family of signals we can sample?
The Shannon-Whittaker-Kotelnikov theorem is essentially a signal representation theorem.
What is so special about those signals?
The signal $x(t) = \sum_n y_n \varphi(t/T - n)$ is exactly specified by one parameter y_n every T seconds, $x(t)$ has a finite number $\rho = 1/T$ of degrees of freedom per unit of time.

In the classical formulation, innovation is uniform. How about signals where the rate of innovation is finite but non-uniform? E.g.

- Piecewise sinusoidal signals (Frequency Hopping modulation)
- Piecewise Constant Signals (CDMA)
- Pulse position modulation (UWB)
- Edges in images

Signals with Finite Rate of Innovation

Consider a signal of the form:

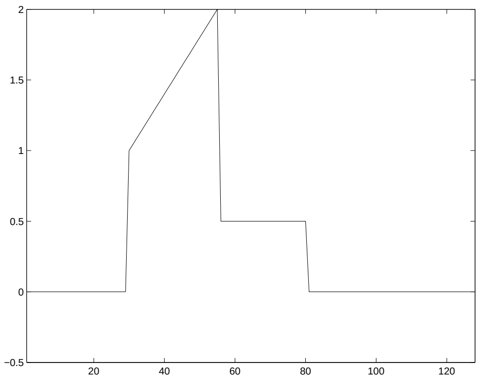
$$x(t) = \sum_{k\in\mathbb{Z}} \sum_{r=0}^{R-1} \gamma_{k,r} g_r(t - t_k). \tag{1}$$

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

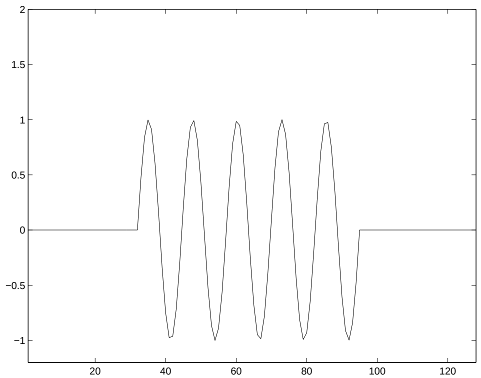
$$\rho = \lim_{\tau\to\infty} \frac{1}{\tau} C_x\left(-\frac{\tau}{2}, \frac{\tau}{2}\right). \tag{2}$$

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).

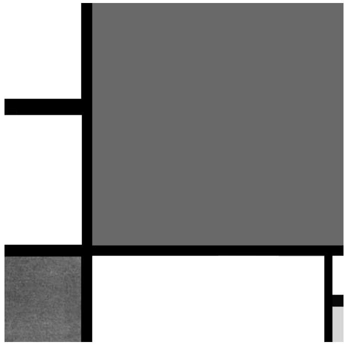
Examples of Signals with Finite Rate of Innovation



Piecewise Polynomial Signals

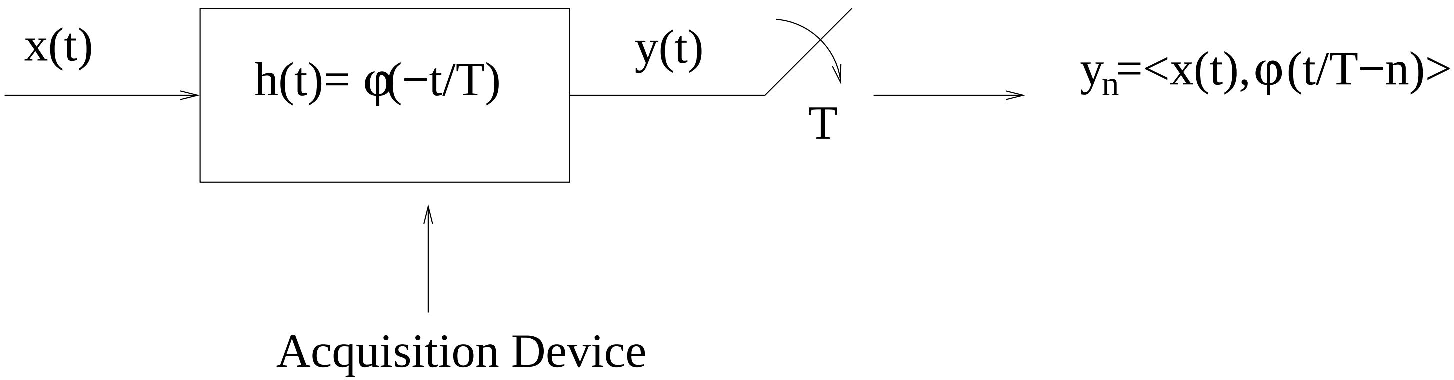


Piecewise Sinusoidal Signals



Mondrian paintings ;-)

The Observation Kernel



- Given by nature
  - Diffusion equation, Green function. Ex: sensor networks.
- Given by the set-up
  - Designed by somebody else. Ex: Hubble telescope, digital cameras.
- Given by design
  - Pick the best kernel. Ex: engineered systems.

About the sampling rate: Given by problem. Ex: sensor networks. Given by design (usually, as low as possible). Ex: UWB

Sampling Kernels

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

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

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

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

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

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

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

3. Any kernel with rational Fourier transform:

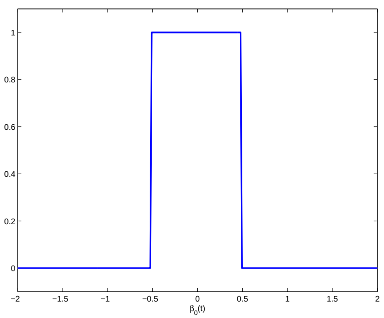
$$\hat{\varphi}(\omega) = \frac{\prod_i (j\omega - b_i)}{\prod_m (j\omega - a_m)} \quad a_m = \alpha_0 + m\lambda \text{ and } m = 0, 1, \dots, L.$$

Sampling Kernels

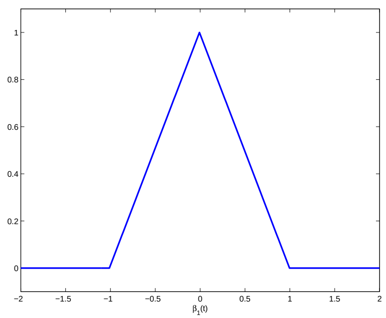
Class 1 is made of all functions $\varphi(t)$ satisfying Strang-Fix conditions. This includes any scaling function generating a wavelet with $L+1$ vanishing moments (e.g., Splines and Daubechies scaling functions).
Strang-Fix Conditions:

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

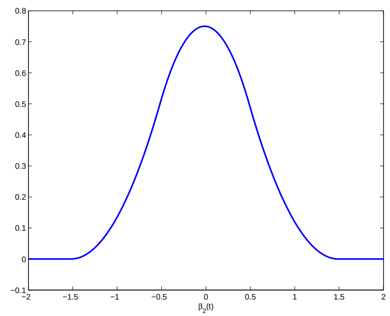
where $\hat{\varphi}(\omega)$ is the Fourier transform of $\varphi(t)$.



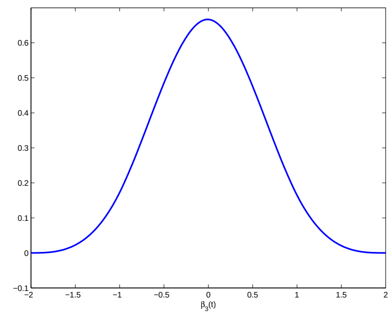
$\beta_0(t)$



$\beta_1(t)$



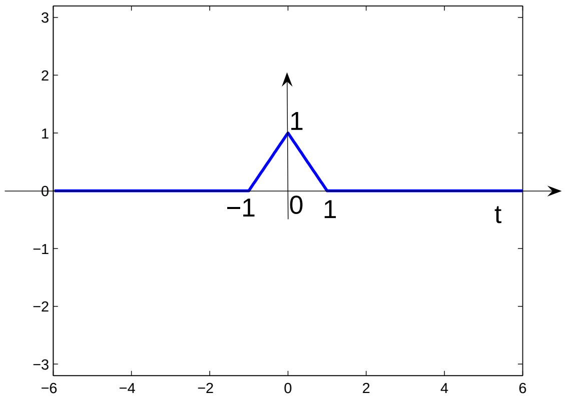
$\beta_2(t)$



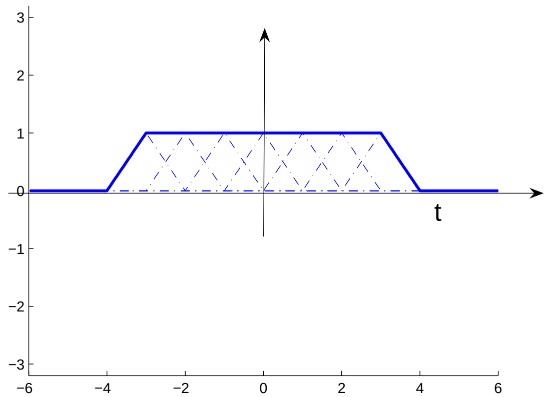
$\beta_3(t)$

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

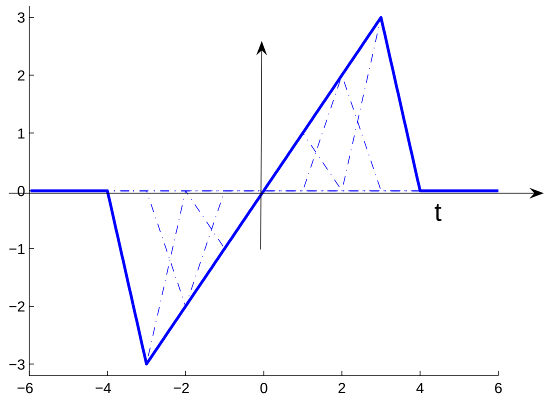
B-splines Reproduce Polynomials



$\beta_1(t)$



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

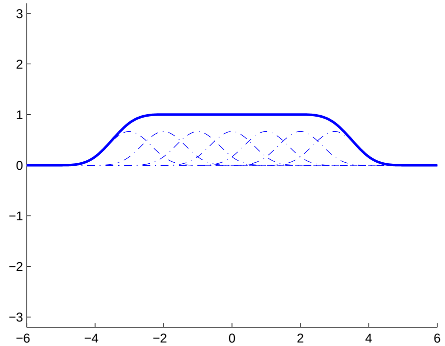


$c_{1,n} = (-3, -2, -1, 0, 1, 2, 3)$

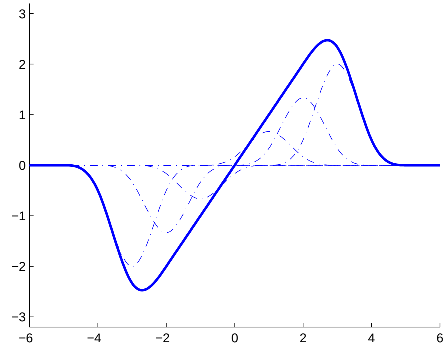
The linear spline reproduces polynomials up to degree L=1: $\sum_n c_{m,n}\beta_1(t-n) = t^m \quad m = 0, 1$, for a proper choice of coefficients $c_{m,n}$ (in this example $n = -3, -2, \ldots, 1, 2, 3$).

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

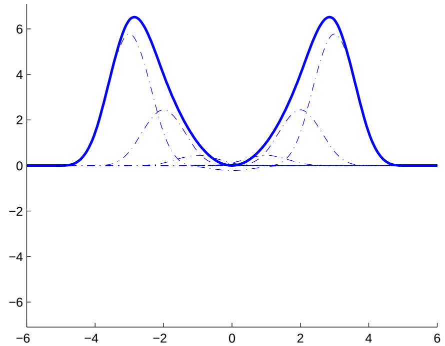
B-splines Reproduce Polynomials



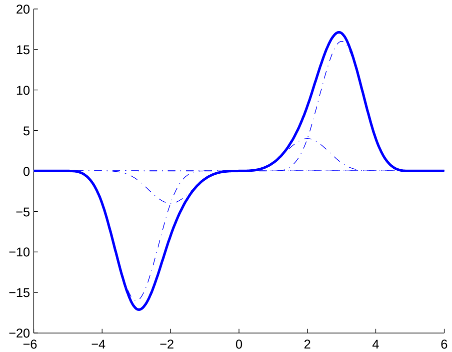
$$c_{0,n} = (1, 1, 1, 1, 1, 1, 1)$$



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



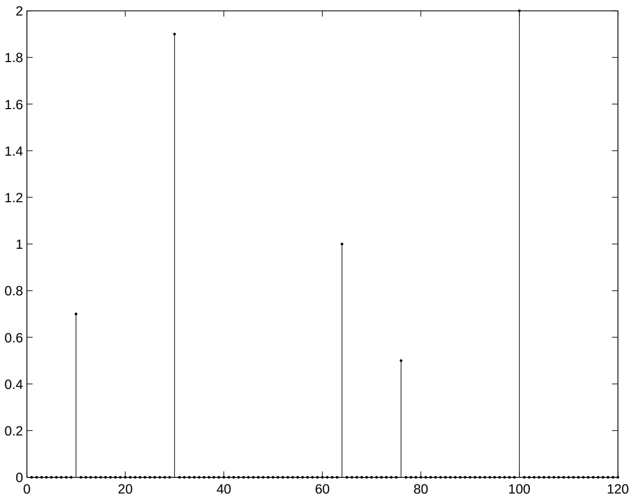
$$c_{2,n} \sim (8.7, 3.7, 0.7, -0.333, 0.7, 3.7, 8.7)$$



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

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

Basic Set-up: 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 $TN = \tau$.

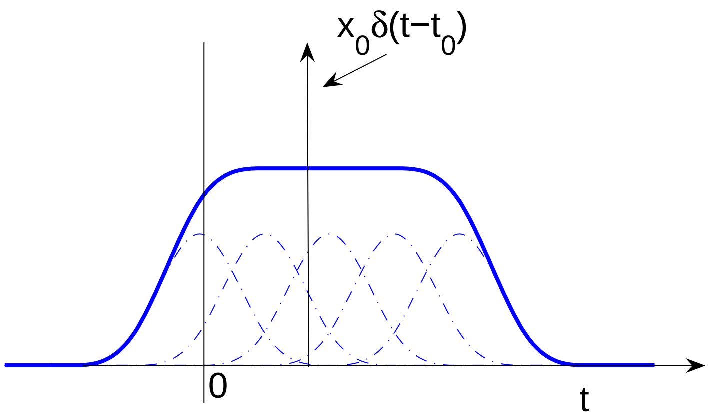
Sampling Streams of Diracs

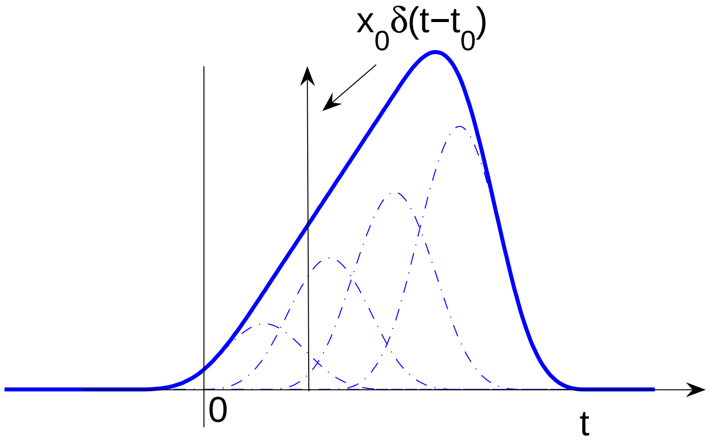
Assume for simplicity that $T = 1$ and define $s_m = \sum_n c_{m,n} y_n$, $m = 0, 1, ..., L$, we have that

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

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

Sampling Streams of Diracs

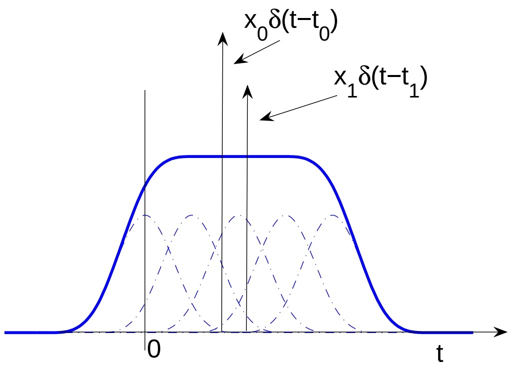


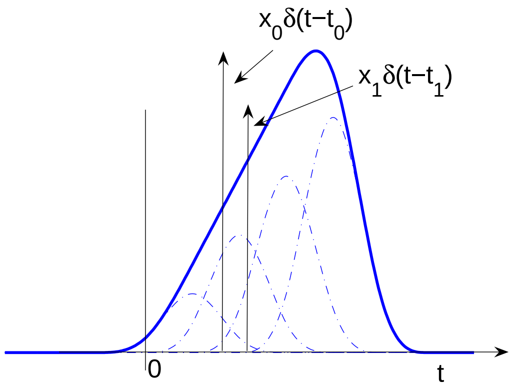


$$\sum_n y_n = \langle x_0\delta(t-t_0), \sum_n \varphi(t-n)\rangle = \int_{-\infty}^{\infty} x_0\delta(t-t_0)\sum_n \varphi(t-n)dt = x_0\sum_n \varphi(t_0-n) = x_0$$

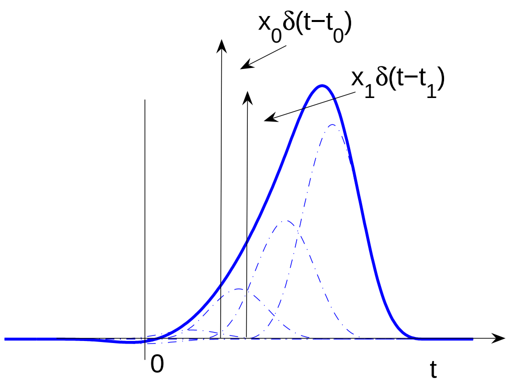
$$\sum_n c_{1,n}y_n = \langle x_0\delta(t-t_0), \sum_n c_{1,n}\varphi(t-n)\rangle = x_0\sum_n c_{1,n}\varphi(t_0-n) = x_0t_0$$

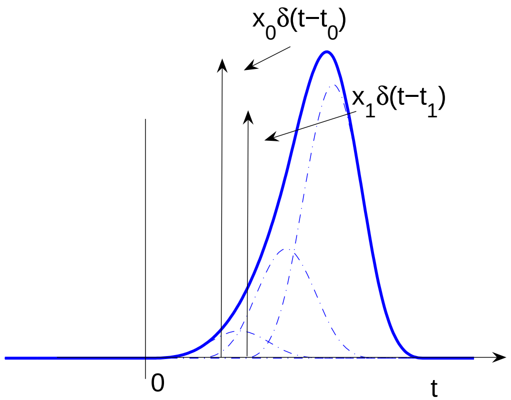
Sampling Streams of Diracs





$$s_0 = \sum_n y_n = x_0 + x_1 \qquad s_1 = \sum_n c_{1,n} y_n = x_0 t_0 + x_1 t_1$$

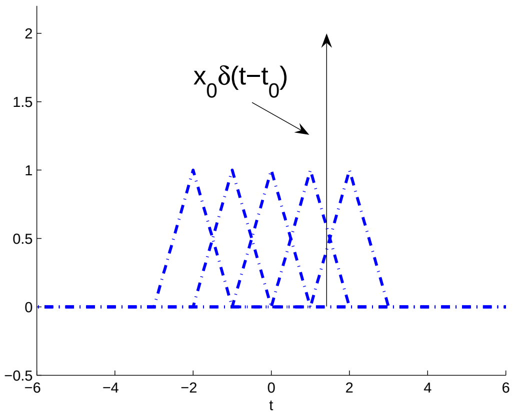


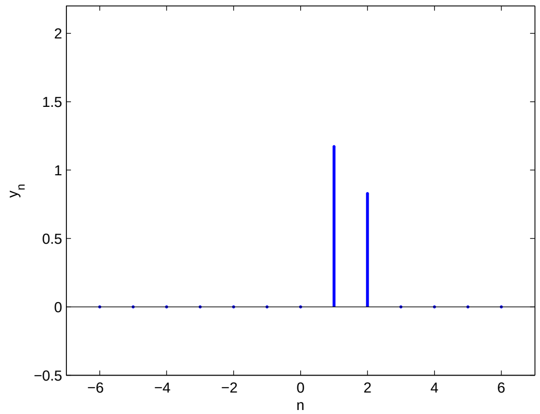


$$s_2 = \sum_n c_{2,n} y_n = x_0 t_0^2 + x_1 t_1^2 \qquad s_3 = \sum_n c_{3,n} y_n = x_0 t_0^3 + x_1 t_1^3$$

Sampling Streams of Diracs - Toy Examples

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

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

The Annihilating Filter 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).

The Annihilating Filter 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.

The Annihilating Filter Method

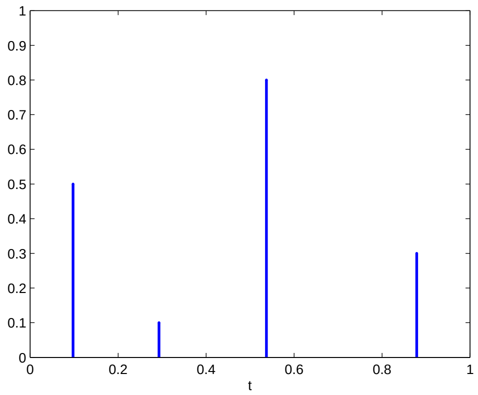
2. Given the coefficients $\{1, h_1, h_2, \ldots, h_k\}$, we get the locations t_k by finding the roots of $H(z)$.

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

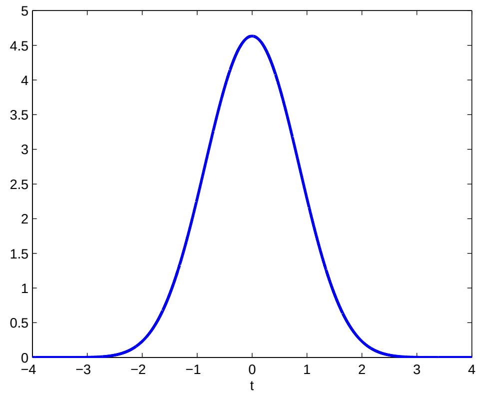
$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_0 & t_1 & \cdots & t_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ t_0^{K-1} & t_1^{K-1} & \cdots & t_{K-1}^{K-1} \end{bmatrix} \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{K-1} \end{pmatrix} = \begin{pmatrix} s_0 \\ s_1 \\ \vdots \\ s_{K-1} \end{pmatrix}. \tag{4}$$

Classic Vandermonde system. Unique solution for distinct t_ks.

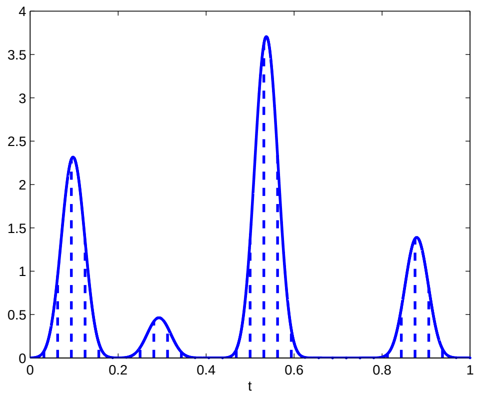
Sampling Streams of Diracs: Numerical Example



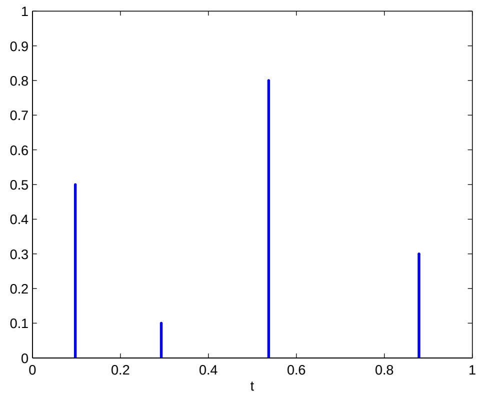
(a) Original Signal



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

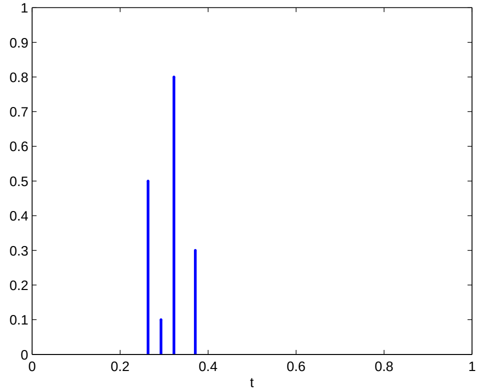


(c) Samples

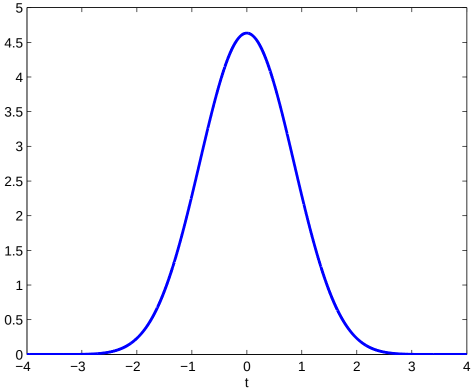


(d) Reconstructed Signal

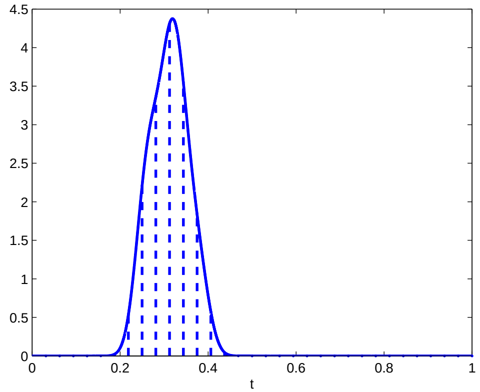
Sampling Streams of Diracs: Closely Spaced Diracs



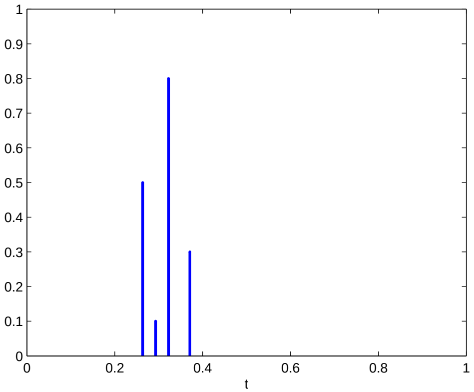
(a) Original Signal



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



(c) Samples



(d) Reconstructed Signal

Complexity

Notice the proof is based on a constructive algorithm:

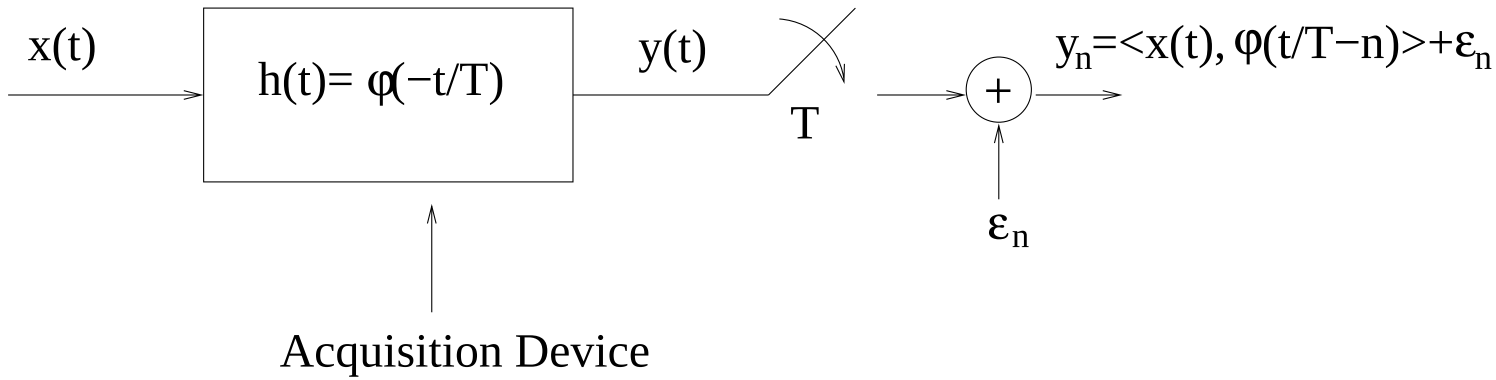
1. Given the N samples y_n, compute the moments s_m using the polynomial reproduction formula
2. Solve a $K \times K$ Toeplitz system to find $H(z)$
3. Find the roots of $H(z)$
4. Solve a $K \times K$ Vandermonde system to find the a_k

Complexity

1. $O(KN)$
2. $O(K^2)$
3. $O(K^3)$
4. $O(K^2)$

Thus, the algorithm complexity is polynomial with the signal innovation.

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 Square Solution

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

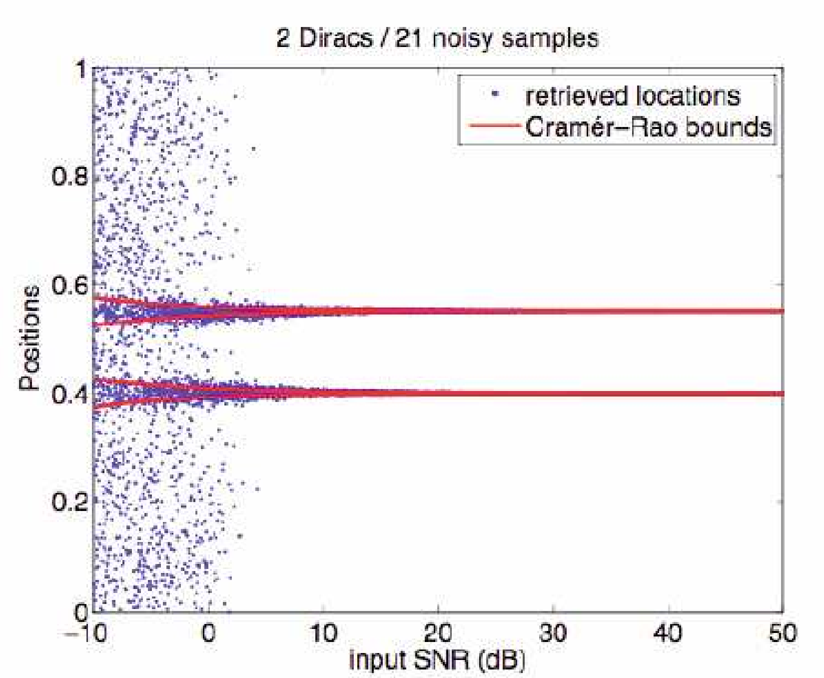
For small SNR use Cadzow's method to denoise $\mathbf{S}$ before applying TLS.

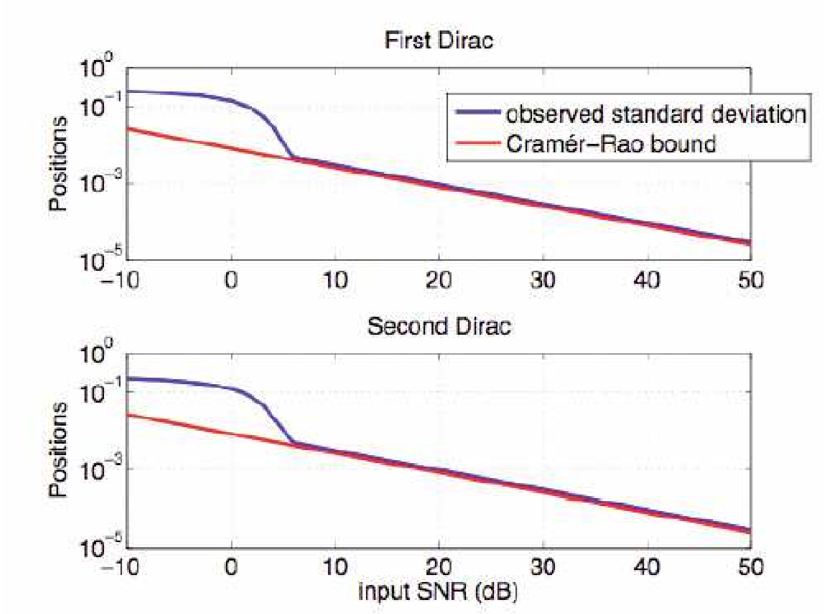
The basic intuition behind this method is that, in the noiseless case, $\mathbf{S}$ is rank deficient (rank K) and Toeplitz, while in the noisy case $\mathbf{S}$ is full rank.

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.

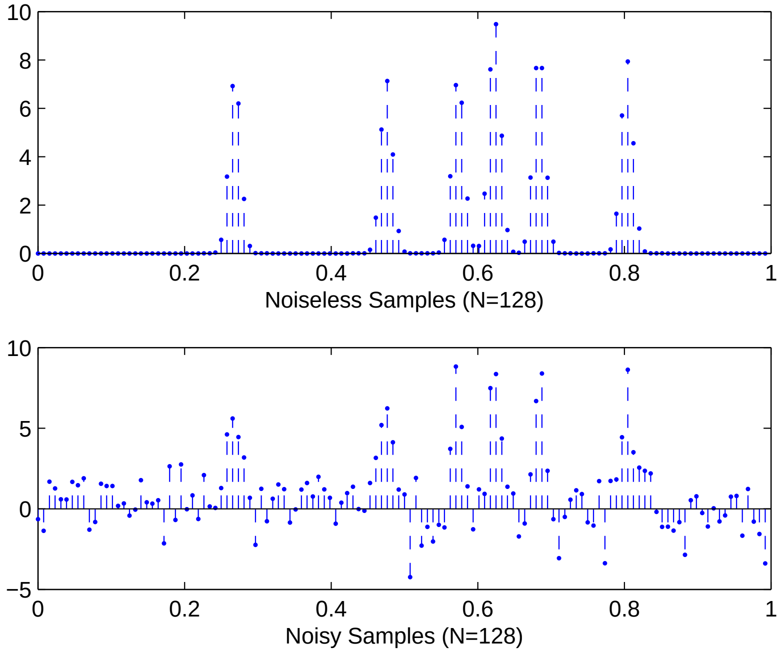
Simulation Results

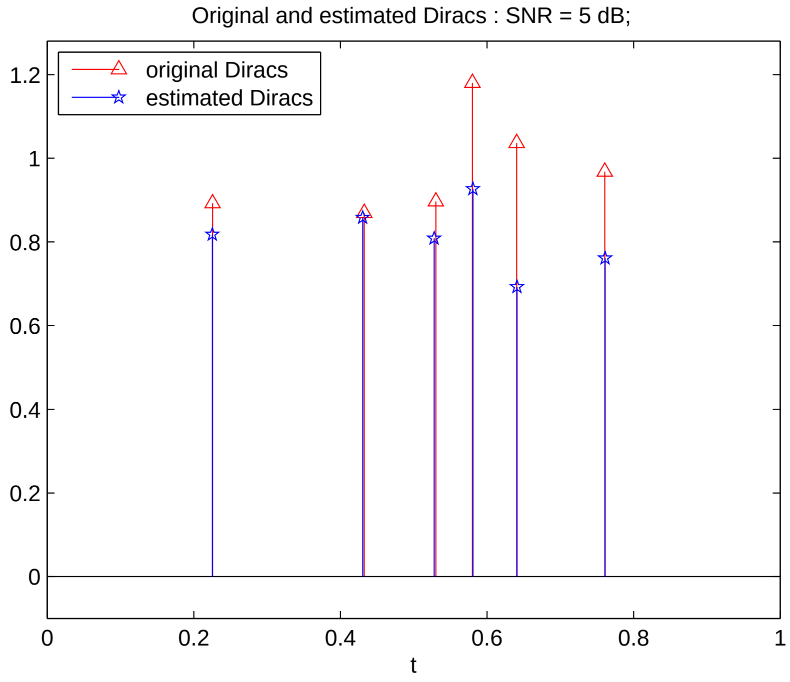




- This is a parametric estimation problem.
- Unbiased algorithms have a covariance matrix lower bounded by CRB.
- The proposed algorithm reaches CRB down to SNR of 5dB.

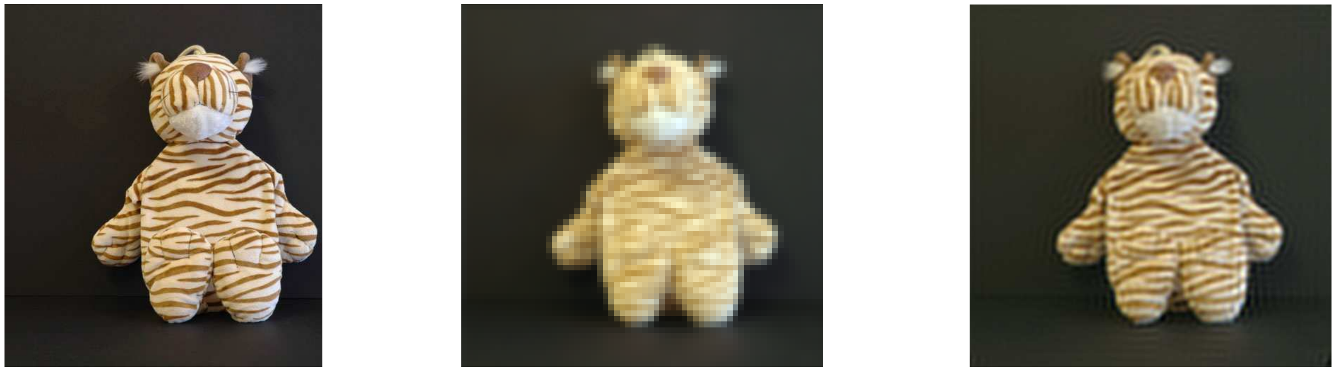
Simulation Results (cont'd)





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.

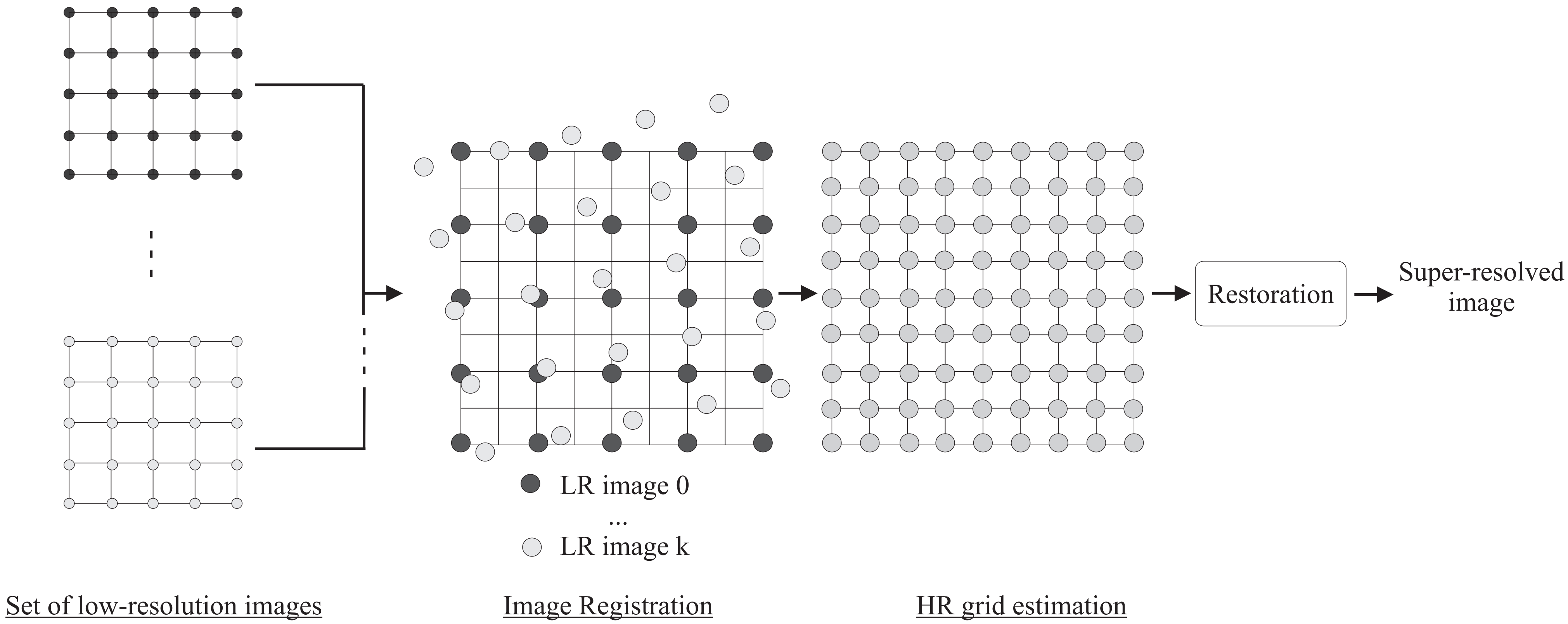


(a)Original (512×512) (b) Low-res. (64×64) (c) Super-res (PSNR=24.2dB)

- 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:

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

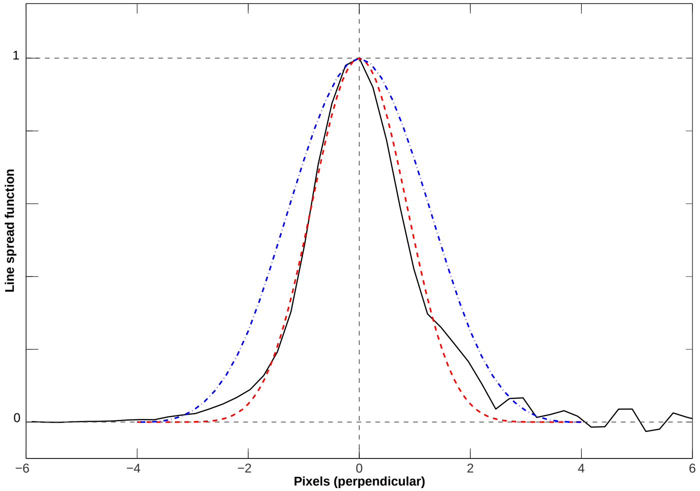
  - We retrieve the exact moments of $I(x, y)$ from $P_{m,n}$:

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

- 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 (2014×3039)



(b) Point Spread function

Application: Image Super-Resolution

(a)Original (128×128)

(b) Super-res (1024×1024)

Application: Image Super-Resolution

(a)Original (48×48)

(b) Super-res (480×480)

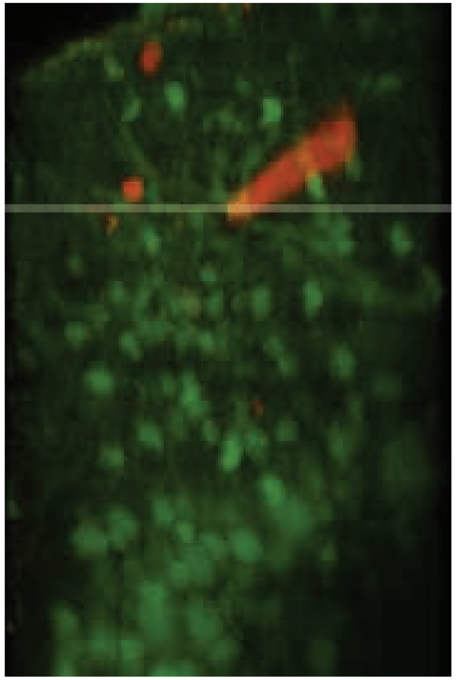
Application: Image Super-Resolution

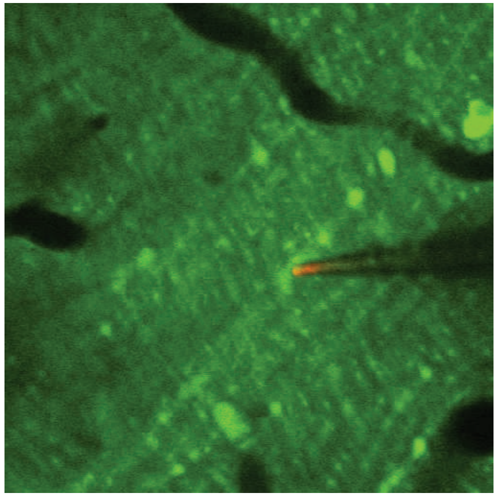
(c) Super-res using Harris (PSNR=14.9dB)

(d) Super-res using FRI (PSNR=16dB)

Local features (e.g., edges and corners) are estimated using FRI algorithms.

Application: Two-photon Microscopy for Neuroscience

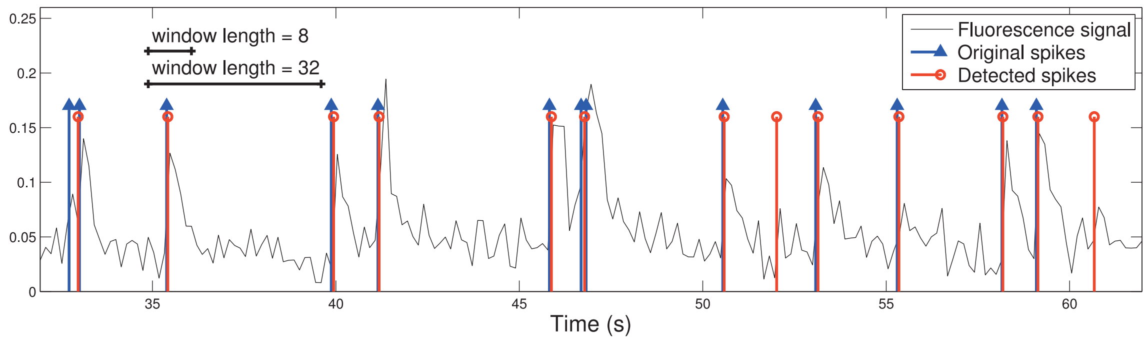




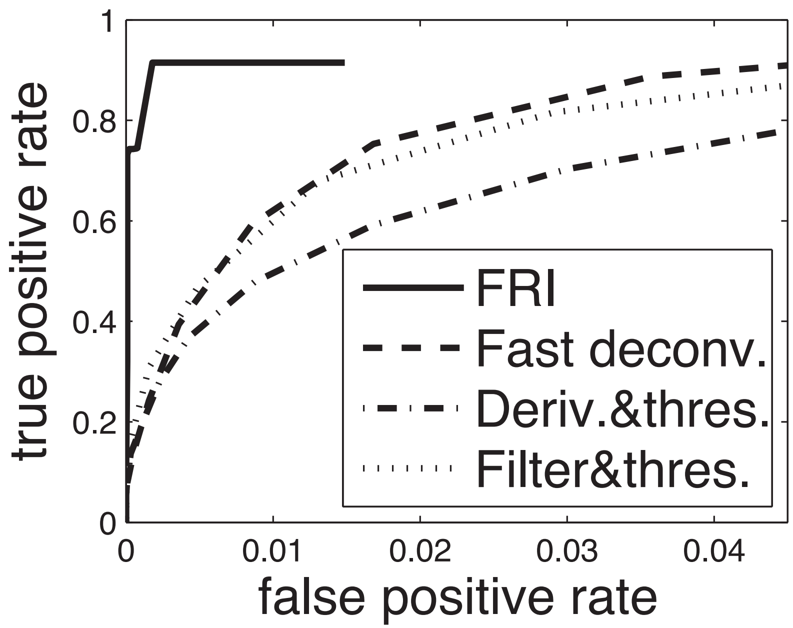
Application: Two-photon Microscopy for Neuroscience

Challenges:

- Increase temporal resolution
- Detect calcium transients



Application: Calcium Transient Detection



Conclusions

Sampling signals at their rate of innovation:

- New framework that allows the sampling and reconstruction of signals at a rate smaller than Nyquist rate.
- Robust and fast algorithms with complexity proportional to the number of degrees of freedom, i.e., $O(K^3)$.
- Provable optimality (i.e. CRB achieved over wide SNR ranges)
- Intriguing connections with multi-resolution analysis, Fourier theory and analogue circuit theory.

But also

- There is no such thing as a free lunch ;-)
- Core application is actually difficult.
- More work need to be done to fit the model and devise the right algorithm for the given application.
- Still many open questions from theory to practice.

References

- On sampling
  - 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.
  - P.Shukla and P.L. Dragotti, ‘Sampling Schemes for Multidimensional Signals with Finite Rate of Innovation’, IEEE Trans. on Signal Processing, vol. 55 (7), pp. 3670-3686, July 2007.
  - J.Berent and P.L. Dragotti, ‘Sampling Piecewise Sinusoidal Signals with Finite Rate of Innovation Methods,’ IEEE Transactions on Signal Processing, February 2010
- On Image Super-Resolution
  - L. Baboulaz and P.L. Dragotti, ‘Exact Feature Extraction using Finite Rate of Innovation Principles with an Application to Image Super-Resolution’, IEEE Trans. on Image Processing, vol.18(2), pp. 281-298, February 2009.