EUSIPCO, 6.9.2007

Sampling Sparse Signals at Occams' Rate

Martin Vetterli,
EPFL & UC Berkeley

Joint work with Thierry Blu and Pier-Luigi Dragotti (ICL)

Audiovisual Communications
Laboratory

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Acknowledgements

- **Sponsors**
  - NSF Switzerland
- **Collaborations**
  - A.Hormati, EPFL
  - J. Kusuma, MIT/Schlumberger
  - I.Maravic, Deutsche Bank
  - P. Prandoni, Quividi
- **Discussions and Interactions**
  - V. Goyal, MIT
  - M. Unser, EPFL

Outline

1. Introduction

Shannon and beyond…

2. Signals with finite rate of innovation

Shift-invariant spaces

Poisson spikes

3. Sampling signals at their rate of innovation

The basic set up: periodic set of diracs, sampled at their rate of innovation

4. The noisy case

Total least squares and Cadzow

Lower bounds on SNR: Cramer-Rao and uncertainty principles

5. Extensions to non-periodic cases and other kernels

Generalized annihilation property

Sinc, Gauss, and Strang-Fix

6. Discussion and conclusions

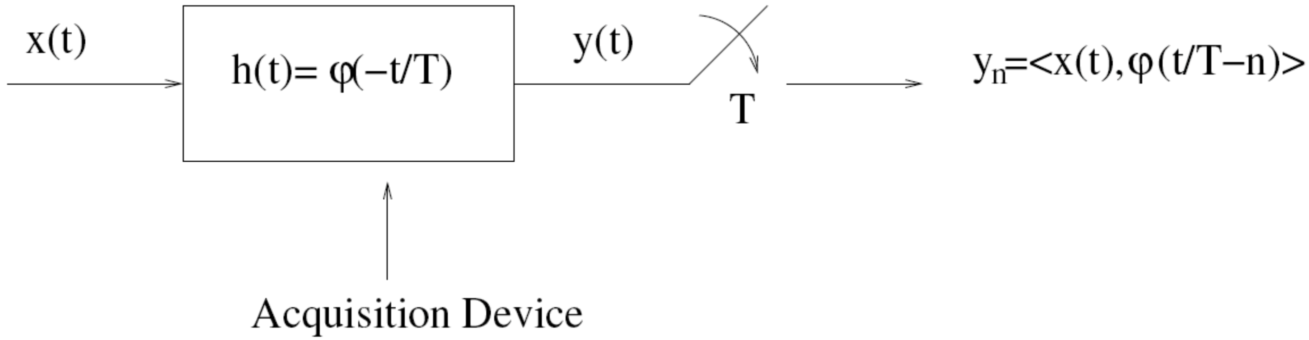
Applications, relation to compressed sensing, outlook

The Question!

You are given a class of objects, a class of functions (e.g. BL)

You have a sampling device, as usual to get acquire the real world

- Smoothing kernel or lowpass filter
- Regular, uniform sampling
- That is, the workhorse of sampling!



Obvious question:

When does a minimum number the samples uniquely specify the function?

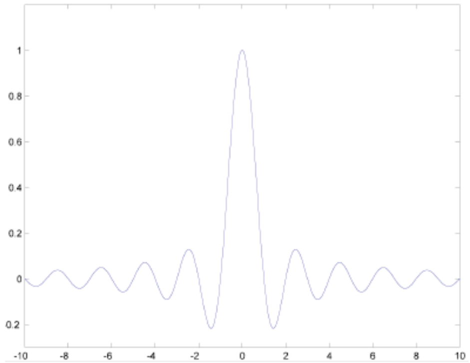
$$x(t) \quad \Longleftrightarrow \quad y_n$$

1. Shannon and Beyond… 1948 Foundation paper

If a function f(t) contains no frequencies higher than W cps, it is completely determined by giving its ordinates at a series of points spaced 1/(2W) seconds apart.
[if approx. T long, W wide, 2TW numbers specify the function]

It is a representation theorem:

- $\{sinc(t-n)\}_{n \in Z}$, is an orthogonal basis for BL[-π,π]
- f(t) ∈ BL[-π,π] can be written as $$f(t) = \sum_n f(n) \cdot sinc(t-n)$$



… slow convergence of the series!

Note:

- Shannon-BW, BL sufficient, not necessary.
- many variations, non-uniform etc
- Nyquist-28, Kotelnikov-33, Whittaker-35, Gabor-46, Shannon-48

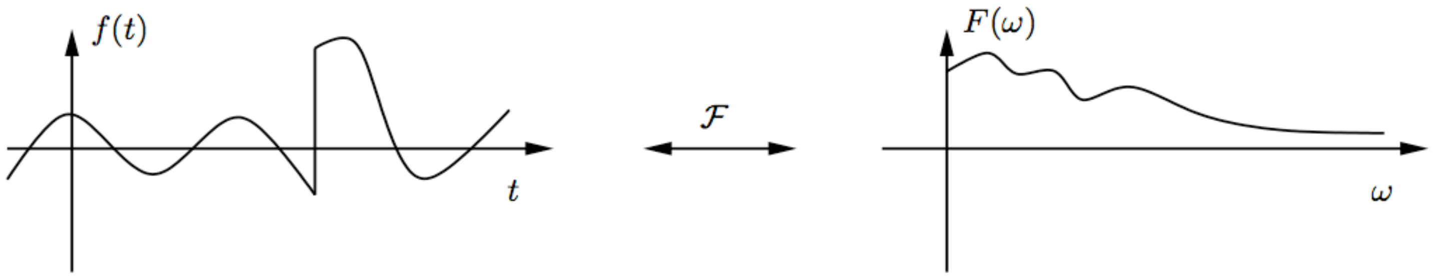
A Variation on a Theme by Shannon

Shannon BL case

$$x(t) = \sum_{n \in Z} x(nT) sinc(t/T - n)$$

or 1/T degrees of freedom per unit time

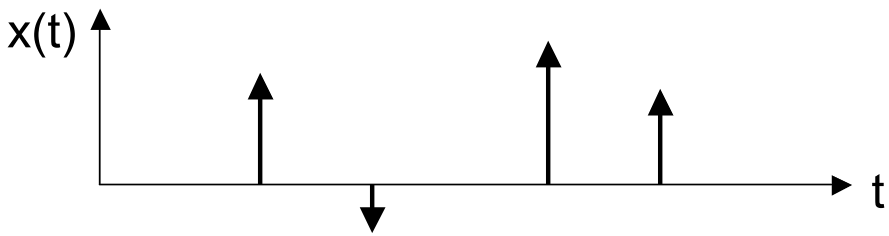
- **But: a single discontinuity, and no more sampling theorem…**



- Are there other signals with finite number or degrees of freedom per unit of time that allow exact sampling results?

1. Shannon and Beyond…

- **Is there a sampling theory beyond Shannon?**
  – Shannon: bandlimitedness is sufficient but not necessary
  – Shannon bandwidth: dimension of subspace
  – Shift-invariant spaces popular with wavelets etc
- **What about:**
  – Piecewise constant signals (CDMA)
  – Pulse position modulation (UWB)
  – Images with sharp edges
- **Thus, develop a sampling theory for classes of sparse non-bandlimited signals!**



1. Shannon and Beyond… Questions

Assume a sparse signal (CT or DT) observed through sampling:

- What is the minimum sampling rate?
- What classes of sparse signals are possible?
- What are good observation kernels, and what are efficient and stable recovery algorithms?
- How does observation noise influence recovery, and what algorithms will approach optimal performance?
- How will these new techniques impact practical applications?
- What is the relationship between these and classic methods
- How does this relate to recent methods in compressed sensing?

Apply Occam's razor to sampling sparse signals!

Occam's razor known as *Lex Parcimoniæ* or "Law of Parsimony": *Entia non svnt mvltiplicanda præter necessitatem*, or, "Entities should not be multiplied beyond necessity" (Wikipedia).

Outline

1. Introduction

2. Signals with finite rate of innovation

Band-limited spaces

Shift-invariant spaces

Poisson spikes

Rate of innovation ρ

The set up and the key questions

3. Sampling signals at their rate of innovation

4. The noisy case

5. Extensions to non-periodic cases and other kernels

6. Discussion and conclusions

2. Signals with Finite Rate of Innovation

1. Bandlimited spaces: Assume x(t) is bandlimited to [-B/2,B/2],

$$x(t) = \sum_{k \in Z} x_k sinc(Bt - k),$$

$$x_k = \langle Bsinc(Bt - k), x(t) \rangle = x(k/B)$$

Thus, x(t) is exactly defined by $\{x_k\}$, $k \in Z$, spaced T=1/B secs apart.
We call this the rate of innovation: $\rho = B$

2. Shift-invariant spaces: $\varphi(t-kT)$ is an ortho basis for S

$$x(t) = \sum_{k \in Z} x_k \varphi(t - kT), x_k = \langle \varphi(t - kT), x(t) \rangle$$

In both cases: $\rho = B = 1/T$

ρ : Rate of innovation or degrees of freedom per unit of time

Note: Shift-invariance and Multiresolution Analysis

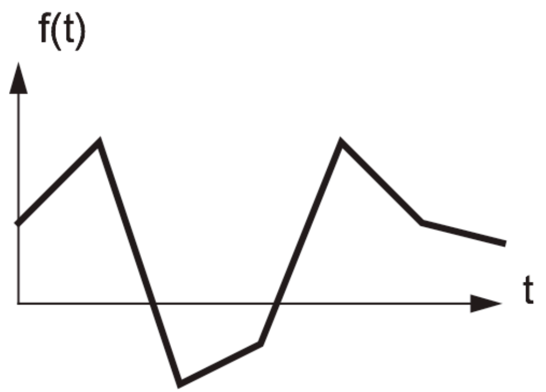
- Most sampling results require shift-invariance subspaces

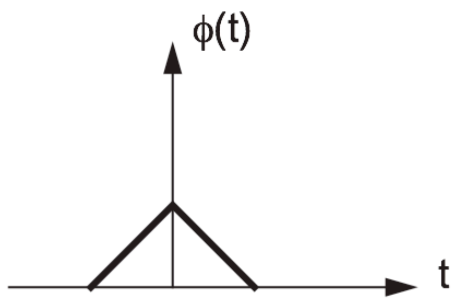
$$f(t) \in V \Leftrightarrow f(t - nT) \in V \quad n \in Z$$

- Wavelet constructions rely in addition on scale-invariance

$$f(t) \in V_0 \Leftrightarrow f(2^m t) \in V_{-m} \quad m \in Z$$

- Multiresolution analysis (Mallat, Meyer) gives a powerful framework. Yet it requires a subspace structure.
- Example: uniform or B-splines





- Question: Can sampling be generalized beyond subspaces?

2. Signals with Finite Rate of Innovation

3. Poisson Process: a set of Dirac pulses, $|t_k - t_{k+1}|$ exp. distrib

$$x(t) = \sum_{k \in Z} \delta(t - t_k), \quad t_k - t_{k+1} \quad p.d.f. \quad \lambda e^{-\lambda t}$$

Innovations: times $\{t_k\}$, rate of innovation average # of Diracs per sec.
Call C_T the number of Diracs in the interval [-T/2,T/2], then

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

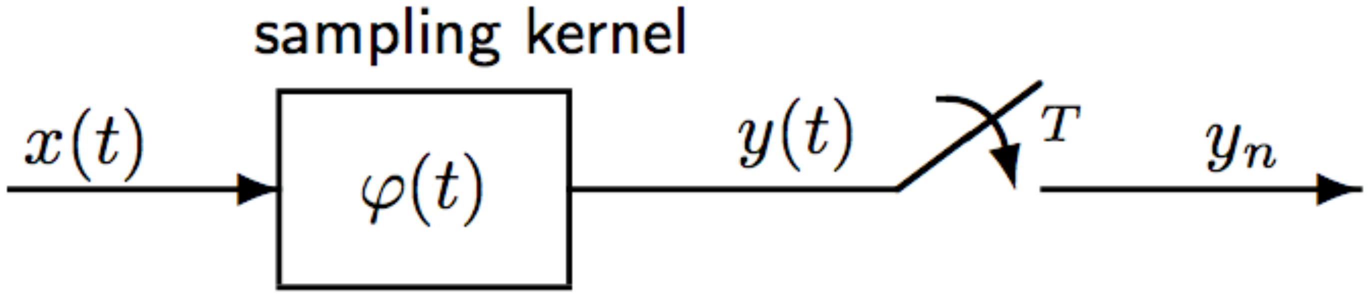
Note: Relation to information rate in Shannon-48
Poisson process: average time is $1/\lambda$, thus $\rho = \lambda$

Weighted Diracs: $$x(t) = \sum_{k \in Z} x_k \delta(t - t_k), \quad \Rightarrow \rho = 2\lambda$$

Obvious question: is there a sampling theorem at ρ samples per sec?
Necessity: Obvious! Sufficiency?

2. Signals with Finite Rate of Innovation

The set up:



Where x(t): signal, φ(t): sampling kernel,
y(t): filtering of x(t) and y_n: samples

For a sparse input, like a sum of Diracs (in DT and CT)

- When is there a one-to-one map $y_n \Leftrightarrow x(t)$?
- Efficient algorithm?
- Stable reconstruction?
- Robustness to noise?
- Optimality of recovery?

Outline

1. Introduction

2. Signals with finite rate of innovation

3. Sampling signals at their rate of innovation

- The basic set up: periodic set of diracs, sampled at their rate of innovation
- A representation theorem [VMB:02]
- Toeplitz system
- Annihilating filter and root finding
- Vandermonde system
- Total least squares and SVD

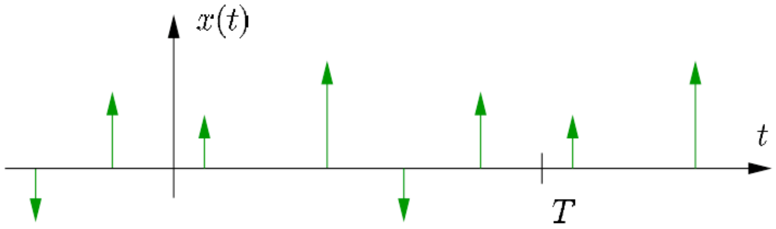
4. The noisy case

5. Extensions to non-periodic cases and other kernels

6. Discussion and conclusions

An Example and overview

K Dirацs on the interval: 2K degrees of freedom. Periodic case:



$$x(t) = \sum_{n \in Z} \sum_{k=0}^{K-1} x_k \delta(t - t_k - n\tau) = \sum_{k=0}^{K-1} x_k \frac{1}{\tau} \sum_{m \in Z} e^{\frac{j2\pi m(t-t_k)}{\tau}}$$

- **Key**: The Fourier series is a weighted sum of K exponentials

$$\hat{x}_m = \frac{1}{\tau} \sum_{k=0}^{K-1} x_k e^{\frac{-j2\pi m t_k}{\tau}}$$

- **Result**: taking 2K+1 samples from a lowpass version of bandwidth (2K+1) allows to perfectly recover x(t)
- **Method**: Yule-Walker, annihilating filter, Vandermonde system
- **Note**: Relation to spectral estimation and ECC (Berlekamp-Massey)

A Representation Theorem [VMB:02]

Theorem

Consider a periodic stream of K Diracs, of period τ, and weights $\{x_k\}$ and locations $\{t_k\}$.

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

Take a sampling kernel of bandwidth B, where $B\tau$ is an odd integer

$$\varphi(t) = \sum_{k' \in Z} sinc(B(t - k'\tau)) = \frac{\sin(\pi B t)}{B\tau \sin(\pi t/\tau)}$$

or a Dirichlet kernel. Then the samples

$$\begin{aligned} y_n &= \left\langle x(t), sinc\Big(B(nT - t)\Big) \right\rangle \\ &= \textstyle\sum_{k=1}^{K} x_k \varphi(nT - t_k), \end{aligned}$$

is a sufficient characterization of x(t).

Note:

Problem is non-linear in t_k, and linear in x_k given t_k

Consider two such streams of K Diracs, of period τ,
and weights and locations $\{x_k, t_k\}$ and $\{x'_k, t'_k\}$, respectively

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

The sum is in general a stream with 2K Diracs.

But, given a set of locations $\{t_k\}$ then the problem is linear in $\{x_k\}$.

The art of the solution:
Separability of non-linear from linear problem

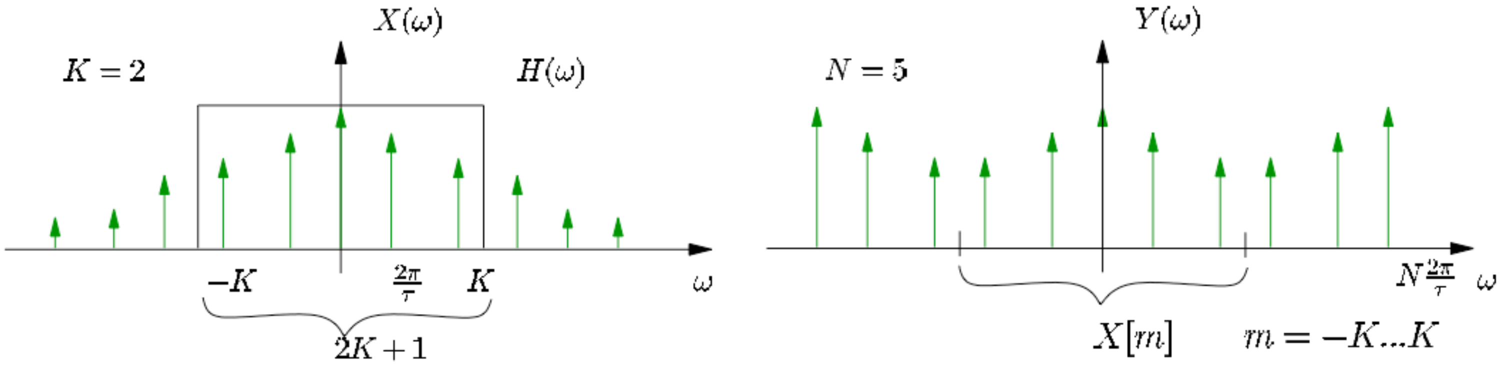
Method: Generalized Annihilation Property GAP

Proof

The signal is periodic, so consider its Fourier series

$$x(t) = \sum_{m \in Z} \hat{x}_m \, e^{j2\pi m t/\tau}, \quad \text{where} \quad \hat{x}_m = \frac{1}{\tau} \sum_{k=1}^{K} x_k \underbrace{e^{-j2\pi m t_k/\tau}}_{u_k^m}.$$

1. The samples y_n are a sufficient characterization of the central 2K+1 Fourier series coefficients. Use sampling theorem or graphically:



Proof

2. The Fourier series is a linear combination of K complex exponentials. These can be killed using a filter, which puts zeros at the location of the exponentials:

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

This filter satisfies:

$$h_m * \hat{x}_m = \sum_{k=0}^{K} h_k \hat{x}_{m-k} = 0$$

because $H(u_k) = 0$ and is thus called annihilating filter.
The filter has K+1 coefficients, but $h_0 = 1$.
Thus, the K degrees of freedom $\{u_k\}$, k = 1…K specify the locations.

Proof

3. To find the coefficients of the annihilating filter, we need to solve a convolution equation, which leads to a K by K Toeplitz system

$$\begin{bmatrix} \hat{x}_{-1} & \hat{x}_{-2} & \cdots & \hat{x}_{-K} \\ \hat{x}_{0} & \hat{x}_{-1} & \cdots & \hat{x}_{-K+1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}_{K-2} & \hat{x}_{K-3} & \cdots & \hat{x}_{-1} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_K \end{bmatrix} = - \begin{bmatrix} \hat{x}_0 \\ \hat{x}_1 \\ \vdots \\ \hat{x}_{K-1} \end{bmatrix}.$$

4. Given the coefficients $\{1, h_1, h_2, \ldots h_K\}$, we get the $\{t_k\}$'s by factorization of

$$H(z) = \prod_{k=1}^{K} (1 - u_k z^{-1}).$$

where

$$u_k = e^{-j2\pi t_k/\tau}$$

Proof

5. To find the coefficients $\{x_k\}$, we have a linear problem, since given the $\{t_k\}$'s or equivalently the $\{u_k\}$'s, the Fourier series is given by

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

This can be written as a Vandermonde system:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ u_1 & u_2 & \cdots & u_K \\ \vdots & \vdots & \ddots & \vdots \\ u_1^{K-1} & u_2^{K-1} & \cdots & u_K^{K-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{bmatrix} = \begin{bmatrix} \hat{x}_0 \\ \hat{x}_1 \\ \vdots \\ \hat{x}_{K-1} \end{bmatrix}.$$

which has always a solution for distinct $\{u_k\}$'s.
This completes the proof of unicity, using $B\tau \geq 2K$ samples.
Since $B\tau$ is odd, we need 2K+1 samples, which is as close as it gets to Occam's razor! Also, $B > 2K/\tau = \rho$.

Notes on Proof

The procedure is constructive, and leads to an algorithm:

1. Take 2K+1 samples y_n from the Dirichlet kernel output
2. Compute the DFT to obtain the Fourier series coefficients -K..K
3. Solve a Toeplitz system of equation of size K by K to get H(z)
4. Find the roots of H(z) by factorization, to get u_k and t_k
5. Solve a Vandermonde system of equation of size K by K to get x_k.

The complexity is:

1. Analog to digital converter
2. K Log K
3. K^2
4. K^3 (can be accelerated)
5. K^2

Or polynomial in K!

Note: For size N vector, with K Diracs, $O(K^3)$ complexity, noiseless

More on annihilation: Rectangular systems

What if more than 2K+1Fourier series? What if filter of length L $\geq$ K?

As long as H(z) has $\{u_k\}$ as roots, it will satisfy

$$\sum_{k=0}^{L} h_k \hat{y}_{m-k} = 0, \qquad \text{for all } |m| \leq \lfloor B\tau/2 \rfloor.$$

and conversely, if it satisfies the above, it has $\{u_k\}$ as roots.

This leads to a rectangular system of size 2M-L+1 by L+1:

$$A = \begin{bmatrix} \hat{y}_{-M+L} & \hat{y}_{-M+L-1} & \cdots & \hat{y}_{-M} \\ \hat{y}_{-M+L+1} & \hat{y}_{-M+L} & \cdots & \hat{y}_{-M+1} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \hat{y}_{M} & \hat{y}_{M-1} & \cdots & \hat{y}_{M-L} \end{bmatrix}$$

where M= $\lfloor B\tau/2 \rfloor$ and H=$[h_0,h_1,\ldots,h_L]^T$satisfy

$$\mathbf{A}\mathrm{H} = 0,$$

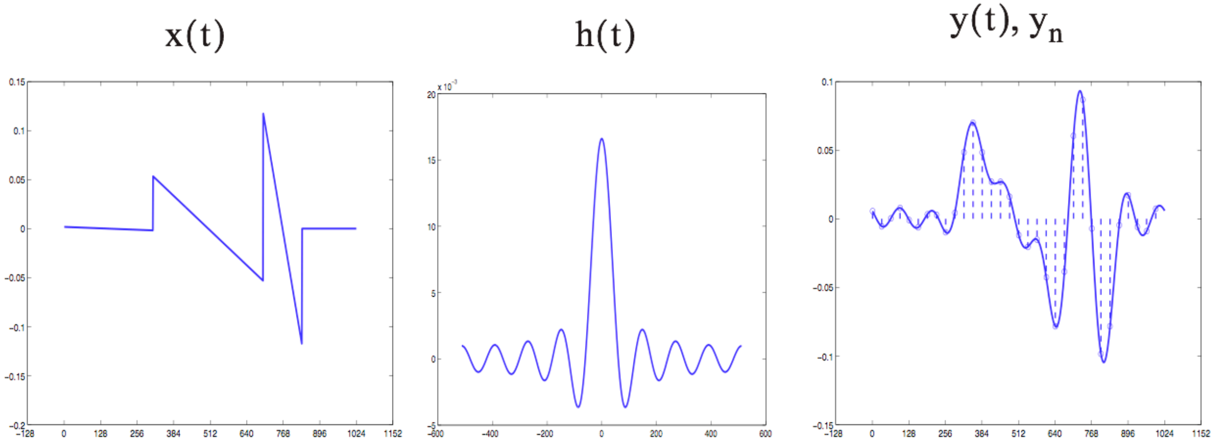
which can be solved using SVD methods (see noisy case….)

Generalizations [VMB:02]

For the class of periodic FRI signals which includes

- Sequences of Diracs
- Non-uniform or free knot splines
- Piecewise polynomials

There are sampling schemes with sampling at the rate of innovation with perfect recovery and polynomial complexity



- **Variations: finite length, 2D, local kernels etc**

Outline

1. Introduction

2. Signals with finite rate of innovation

3. Sampling signals at their rate of innovation

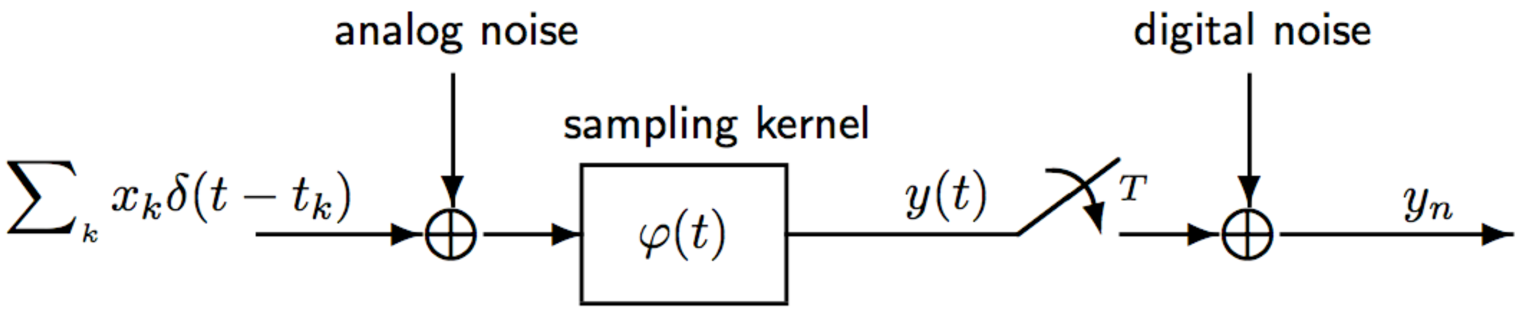
4. The noisy case

- Total least squares
- Cadzow or model based denoising
- Lower bounds on SNR: Cramer-Rao
- Uncertainty principles on location and amplitude of Diracs

5. Extensions to non-periodic cases and other kernels

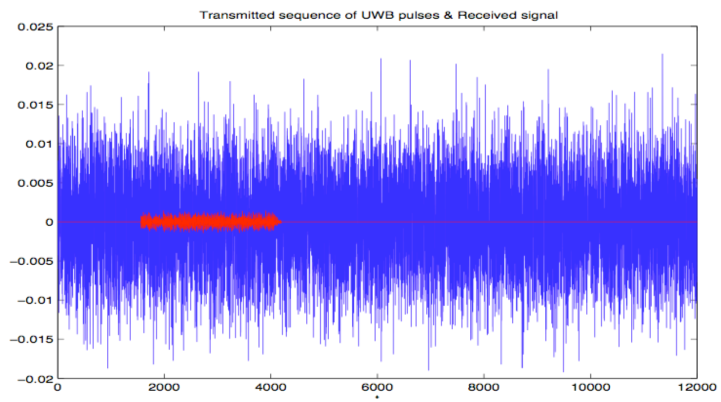
6. Discussion and conclusions

Acquisition in the noisy case:

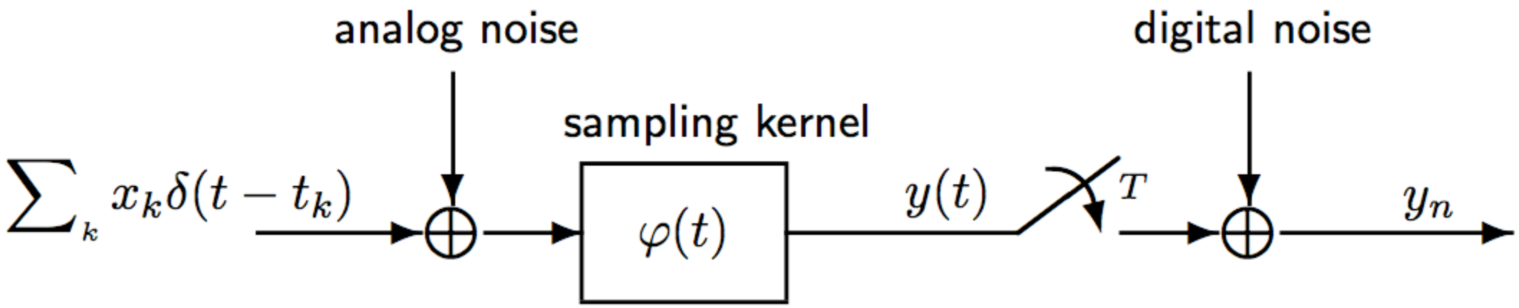


where ''analog'' noise is before acquisition (e.g. communication noise on a channel) and digital noise is due to acquisition (ADC, etc)

Example: Ultrawide band (UWB) communication….



The noisy case: The non-linear estimation problem



Data: $$y_n = \sum_{k=1}^{K} x_k \varphi(nT - t_k) + \varepsilon_n \quad \text{for } n = 1, 2, \ldots, N,$$

Where: T=τ/N, ϕ(t) Dirichlet kernel, N/(ρ τ) redundancy

Estimation: For $$\hat{y}_n = \sum_{k=1}^{K} \hat{x}_k \varphi(nT - \hat{t}_k) \quad n = 1, 2, \ldots, N,$$

Find the set $\{\hat{x}_k, \hat{t}_k\}_{k=1,2,\ldots,K}$ such that $\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2$ is minimized

This is: a non-linear estimation problem in $\{t_k\}$,
a linear estimation problem in $\{x_k\}$ given $\{t_k\}$.

Total least squares solution

Annihilation equation:

$$\mathbf{A}\mathrm{H} = 0,$$

can only be approximately satisfied.

Instead:

$$\text{Minimize}\|\mathbf{A}\mathrm{H}\|^2 \quad \text{under constraint}\|\mathrm{H}\|^2 = 1$$

Pick L=K for minimal size, and unique locations.

Method: SVD of **A**, or eigen-decomposition of $\mathbf{A}^T\mathbf{A}$.

That is:

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

$$\mathbf{U}: \ (B\tau - K) \times (K+1) \ \text{Unitary}$$

$$\mathbf{S}: \ (K+1) \times (K+1) \ \text{Diag. pos. decr.}$$

$$\mathbf{V}: \ (K+1) \times (K+1) \ \text{Unitary}$$

Then H is the last column of **V** .

Total least squares solution

From H, retrieve $\{t_k\}$ by factorization of H(z)

Then, solve for $\{x_k\}$ as the least squares solution of the model w.r.t the measurements.

$$\begin{bmatrix} \varphi(T-t_1) & \varphi(T-t_2) & \cdots & \varphi(T-t_K) \\ \varphi(2T-t_1) & \varphi(2T-t_2) & \cdots & \varphi(2T-t_K) \\ \vdots & \vdots & & \vdots \\ \varphi(NT-t_1) & \varphi(NT-t_2) & \cdots & \varphi(NT-t_K) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}.$$

Complexity: Linear in N, since

- Computation of (K+1) by (K+1) matrix $\mathbf{A}^T \mathbf{A}$: O(KN)
- Eigenvalue decomposition of $\mathbf{A}^T \mathbf{A}$: O(K^3)

Note:

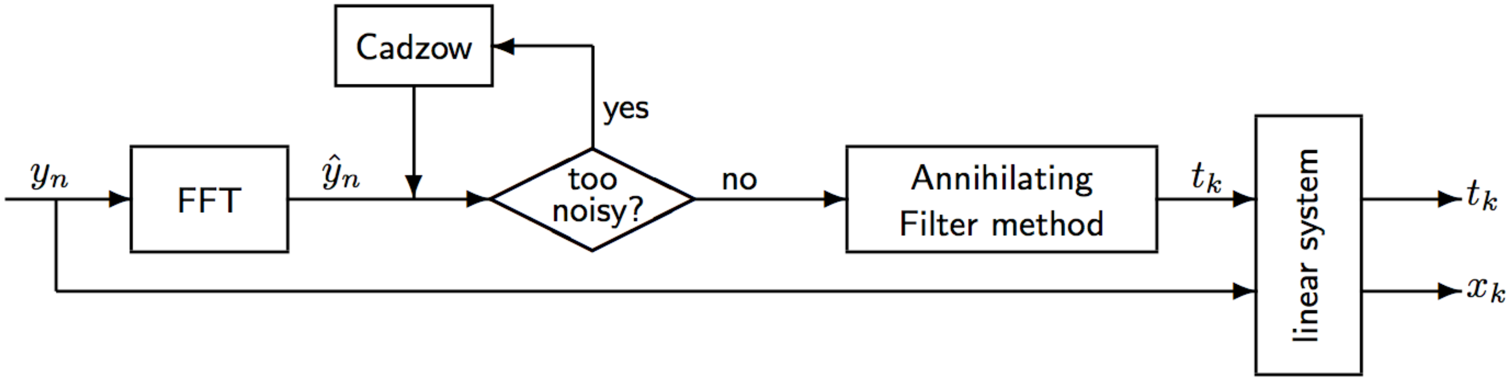
- This is similar to Pisarenko's method in spectral estimation
- If too much noise, preprocessing, or denoising, necessary

Denoising using Cadzow's iterated algorithm

For small SNR's, the TLS method can fail. Use a longer filter, of size L > K, but use knowledge that the matrix **A** is still only of rank K.

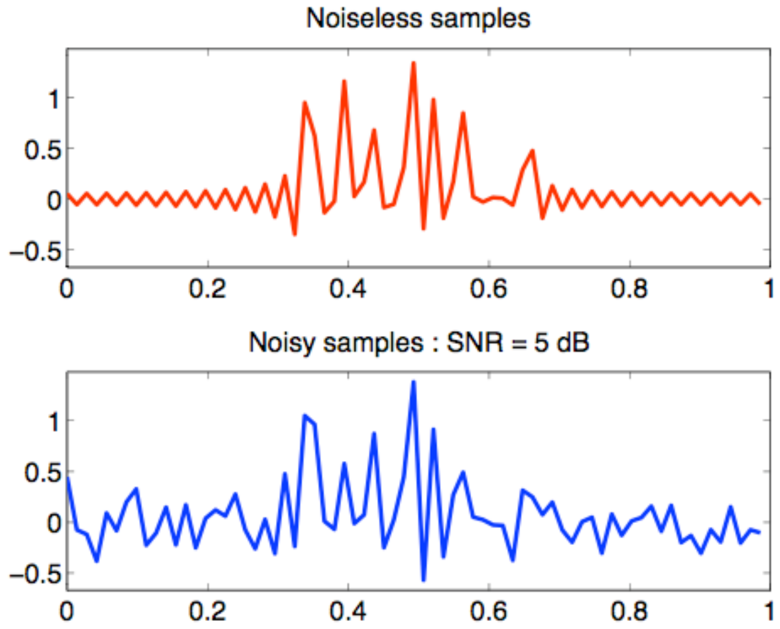
Typically, use L= $\lfloor B\tau/2 \rfloor$ or of order N/2. Then:

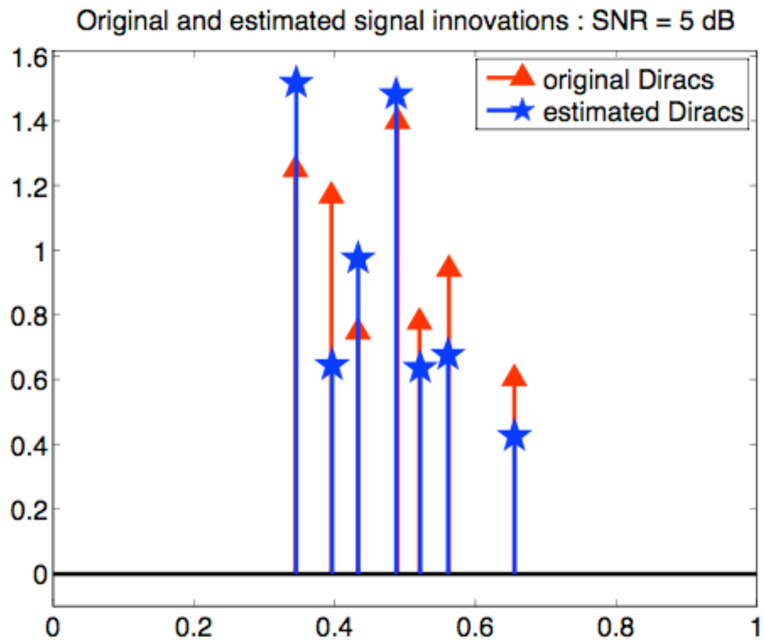
- SVD of **A** = **U S V**^{T}
- Put L+1-K smallest diagonal coefficients of **S** to zero, leading to **S**'
- New matrix **A**' = **U S**' **V**^{T}
- This matrix is not Toeplitz, make it so by averaging along diagonals
- This is a denoised version of the sequence, still fitting the model
- Iterate until convergence criterion



Example 1

7 Diracs in 5dB SNR, 71 samples



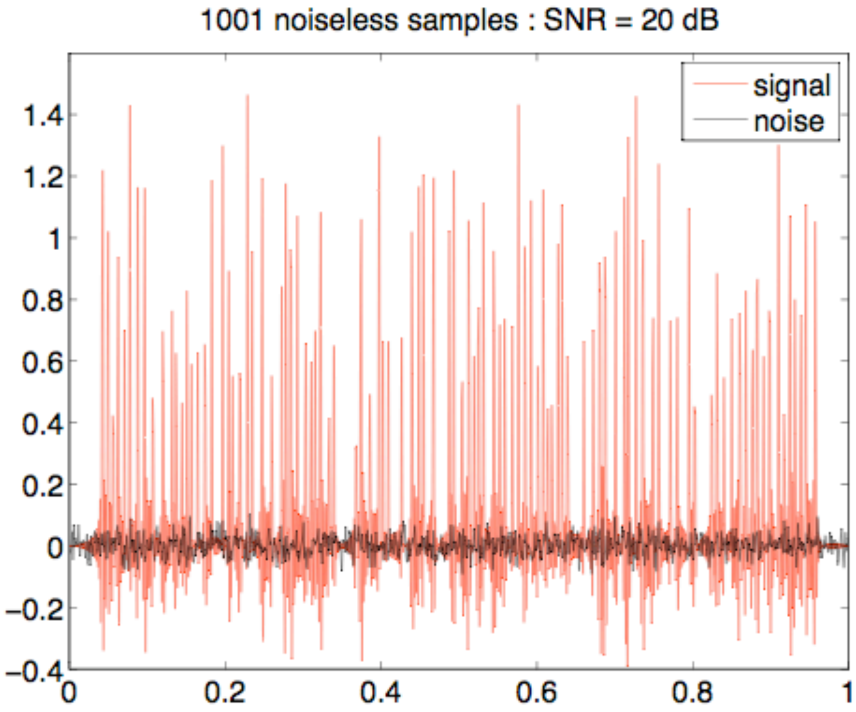


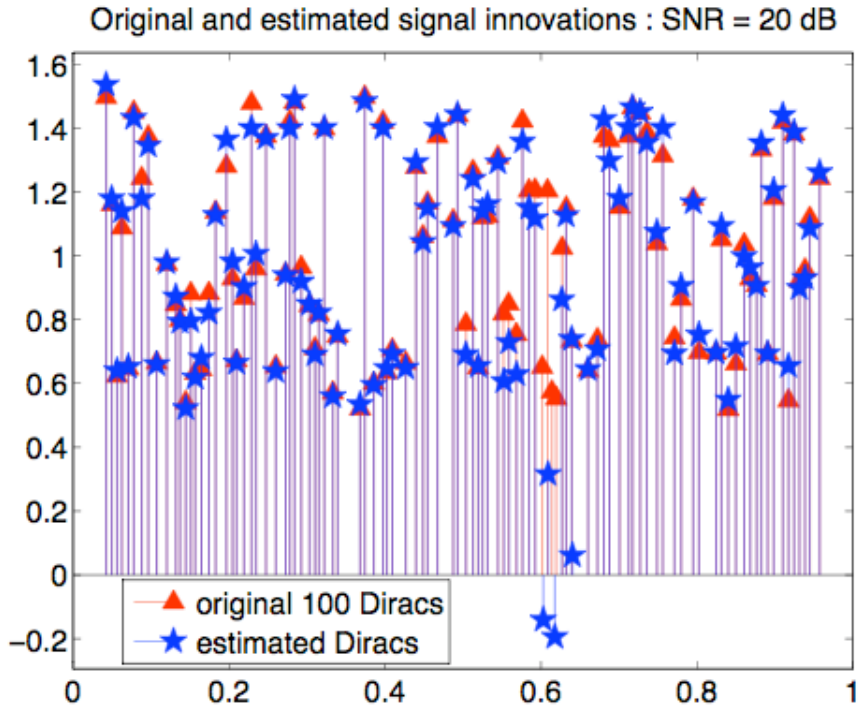
Original and noisy version

Original and retrieved Diracs

Example 2

100 Diracs in 20dB SNR, 1001 samples





Noisy version

Original and retrieved Diracs

Cramer-Rao bounds (CRB)

This is a parametric estimation problem with noise.

Unbiased algorithms have a covariance matrix lower bounded by CRB.

Consider the vector $Y=[y_1, y_2, \ldots y_N]$:

$$y_n = \sum_{k=1}^{K} x_k \varphi(nT - t_k) + \varepsilon_n$$

where ϵ_n is zero mean Gaussian stationary with cov. **R** .

Then any unbiased estimate $\Theta(Y)$ of $\{x_k, t_k\}$ has a covariance matrix lower bounded by the inverse of the Fisher information matrix:

$$\text{cov}\{\Theta\} \geq \left(\mathbf{\Phi}^T \mathbf{R}^{-1} \mathbf{\Phi}\right)^{-1},$$

where Φ is given by

$$\Phi = \left[\begin{array}{ccc|ccc} \varphi(T-t_1) & \cdots & \varphi(T-t_K) & -x_1\varphi'(T-t_1) & \cdots & -x_K\varphi'(T-t_K) \\ \varphi(2T-t_1) & \cdots & \varphi(2T-t_K) & -x_1\varphi'(2T-t_1) & \cdots & -x_K\varphi'(2T-t_K) \\ \vdots & & \vdots & \vdots & & \vdots \\ \varphi(NT-t_1) & \cdots & \varphi(NT-t_K) & -x_1\varphi'(NT-t_1) & \cdots & -x_K\varphi'(NT-t_K) \end{array}\right].$$

Cramer-Rao bound for 1 Dirac case

Assume N-periodic noise (circular **R**) and φ(t) periodic B-wide sinc:

$$\varphi(t) = \frac{\sin(\pi B t)}{B\tau \sin(\pi t/\tau)},$$

Because **R** is circular, it is diagonalized by DFT, leading to diagonal coefficients.

Then, the DFT of $\varphi(nT-t_k)$ and $\varphi'(nT-t_k)$ follows from the Dirichlet kernel

This leads to minimal uncertainties for location Δt_1 and Δx_1 :

$$\frac{\Delta t_1}{\tau} \geq \frac{B\tau}{2\pi |x_1| \sqrt{N}} \left(\sum_{|m| \leq \lfloor B\tau/2 \rfloor} \frac{m^2}{\widehat{r}_m} \right)^{-1/2} \quad \text{and} \quad \Delta x_1 \geq \frac{B\tau}{\sqrt{N}} \left(\sum_{|m| \leq \lfloor B\tau/2 \rfloor} \frac{1}{\widehat{r}_m} \right)^{-1/2}.$$

CRB for 1 Dirac case: Uncertainty relations

Find $[x_1,t_1]$ from N noisy measurements $[y_1,y_2, \ldots y_N]$

$$\begin{aligned} y_n &= \mu_n + \epsilon n \\ \text{with } \mu_n &= x_1 \varphi(n\tau/N - t_1) \end{aligned} \qquad \varphi(t) = \frac{\sin(\pi B t)}{B\tau \sin(\pi t/\tau)},$$

Use previous CRB formulae, consider 2 cases (location Δt_1 and Δx_1)

- Noise is white (''digital noise"), power σ^2 and PSNR $|x_1|^2 / \sigma^2$:
  Location ~ $(N B \tau)^{-1/2}$.
- Noise is white, filtered by $\varphi(t)$ (''analog noise")
  Location ~ $(B \tau)^{-1}$

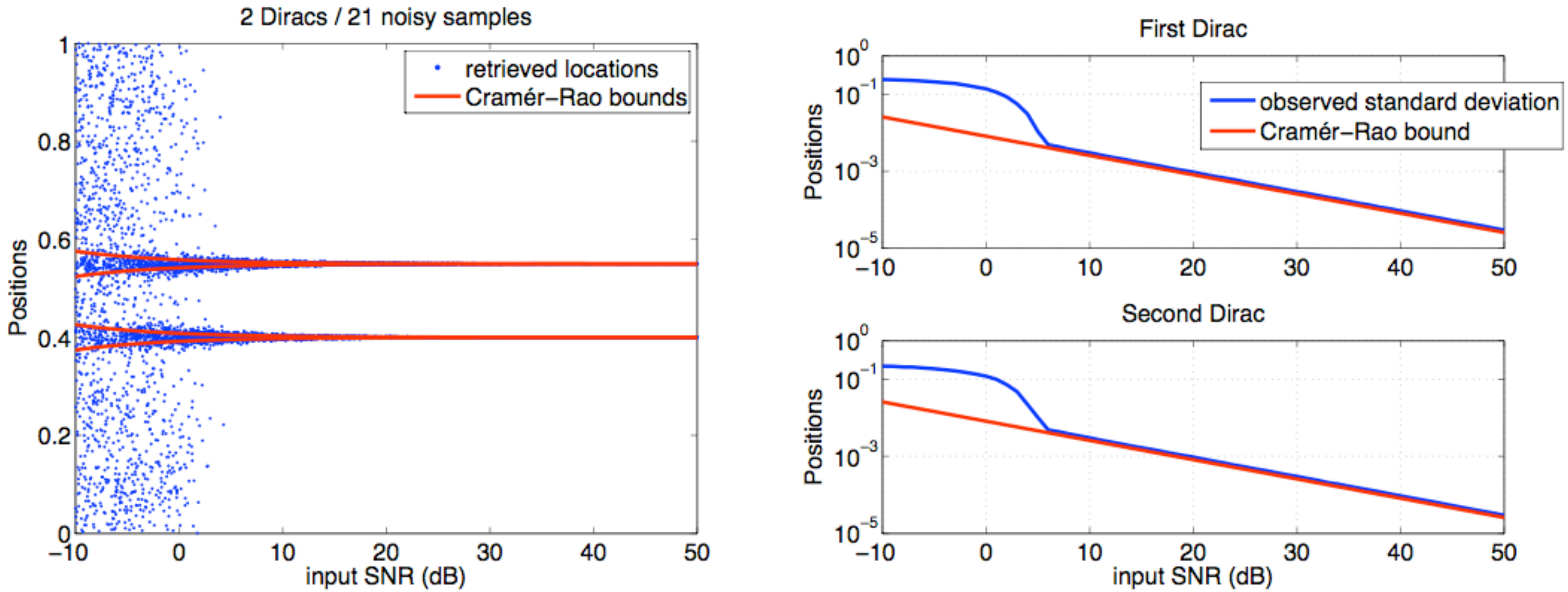
In both cases: to minimize uncertainty on Δt_1 , maximize bandwidth of Dirichlet kernel, $B\tau = N$ (odd) or $B\tau = N-1$ (even)

- Uncertainty relation: $$N \cdot \mathrm{PSNR}^{1/2} \cdot \frac{\Delta t_1}{\tau} \geq \frac{\sqrt{3}}{\pi},$$

Note: these formulae approx. hold for K > 1, for Diracs separated by at least 2/N (not adjacent in sampled domain)

CRB for 2 Dirac case: Experiment

Find $[x_1, t_1, x_2, t_2]$ from 21 noisy samples $[y_1, y_2, \ldots y_{21}]$



The CRB is reached down to SNR of 5dB, thus we have an optimal retrieval method with a polynomial algorithm

Note: SNR not a good measure, PSNR more appropriate

Outline

1. Introduction

2. Signals with finite rate of innovation

3. Sampling signals at their rate of innovation

4. The noisy case

5. Extensions to non-periodic cases and other kernels

Generalized annihilation property

Sinc case

Gaussian kernels

Strang-Fix and Spline kernels

6. Discussion and conclusions

Non-periodic case: Generalized Annihilation Property

Consider K Diracs with $\{x_k, t_k\}$, k=1..K, from N measurements $[y_1, y_2, \ldots y_N]$ at the output of a convolution with $\varphi(t)$:

$$y_n = \sum_{k=1}^{K} x_k \varphi(nT - t_k) \qquad \text{where } n = 1, 2, \ldots, N,$$

This leads to:

$$\underbrace{\begin{bmatrix} \varphi(T-t_1) & \varphi(T-t_2) & \cdots & \varphi(T-t_K) \\ \varphi(2T-t_1) & \varphi(2T-t_2) & \cdots & \varphi(2T-t_K) \\ \vdots & \vdots & & \vdots \\ \varphi(NT-t_1) & \varphi(NT-t_2) & \cdots & \varphi(NT-t_K) \end{bmatrix}}_{\mathbf{F}(\Omega)} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{bmatrix}}_{\mathsf{X}} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}}_{\mathsf{Y}}$$

where $\Omega = [\omega_1, \omega_2, \ldots \omega_K]$ are related 1-1 to the times t_k .
The general non-linear problem is thus:

$$\mathbf{F}(\Omega)\,\mathsf{X} = \mathsf{Y},$$

with unknowns $\Omega = [\omega_1, \omega_2, \ldots \omega_K]$ and $X = [x_1, x_2, \ldots x_N]$ and measurements $Y = [y_1, y_2, \ldots y_N]$.

Non-periodic case: Generalized Annihilation Property

This system satisfies a GAP if there exist K+1 constant matrices $\{\mathbf{A}_k\}$ and K+1 scalar functions of Ω, $h_k(\Omega)$ s.t.:

$$\sum_{k=0}^{K} h_k(\Omega)\,\mathbf{A}_k\,\mathbf{F}(\Omega) = \mathbf{0}, \quad \text{or} \quad \sum_{k=0}^{K} h_k(\Omega)\mathbf{A}_k \mathsf{Y} = 0$$

In matrix form $\quad \mathbf{A}\mathsf{H} = 0,$

With $\quad \mathsf{H} = [\,h_0(\Omega), h_1(\Omega), \ldots, h_K(\Omega)\,]^T \quad \mathbf{A} = \left[\,\mathbf{A}_0\mathsf{Y}, \mathbf{A}_1\mathsf{Y}, \ldots, \mathbf{A}_K\mathsf{Y}\,\right]$

Polynomial case: $\quad \sum_{k=0}^{K} h_k(\Omega) z^{-k} = \prod_{k=1}^{K} (1 - \omega_k z^{-1}),$

1. Compute solution $\mathsf{H} = [\,1, h_1, \ldots, h_{K-1}, h_K\,]^T$
$$\left[\,\mathbf{A}_0\mathsf{Y}, \mathbf{A}_1\mathsf{Y}, \ldots, \mathbf{A}_K\mathsf{Y}\,\right] \mathsf{H} = 0;$$
2. Compute the roots ω_k of the z-transform $H(z) = \sum_{k=0}^{K} h_k z^{-k}$;
3. Compute a solution X of $\mathbf{F}(\Omega)\,\mathsf{X} = \mathsf{Y}$.

Generalized Annihilation Property: Examples

1. Spectral estimation

$$\mathbf{F}(\Omega)\,\mathsf{X} = \mathsf{Y},$$

with

$$\mathbf{F}(\Omega) = \begin{bmatrix} \omega_1 & \omega_2 & \cdots & \omega_K \\ \omega_1^2 & \omega_2^2 & \cdots & \omega_K^2 \\ \vdots & \vdots & & \\ \omega_1^N & \omega_2^N & \cdots & \omega_K^N \end{bmatrix}$$

and where the frequencies to retrieve, f_k, are related to ω_k as

$$\omega_k = e^{j2\pi f_k}$$

The matrices $\mathbf{A}_k$ are band-diagonal, given by

$$\mathbf{A}_k = \left[\mathbf{0}_{\mathbf{N-K,k}} \;\; \mathbf{I}_{\mathbf{N-K}} \;\; \mathbf{0}_{\mathbf{N-K,K-k}} \right]$$

Note: the matrix $\mathbf{F}(\Omega)$ is Vandermonde of size N by K

Generalized Annihilation Property: Examples

2. Dirichlet kernel (seen above):

$$\mathbf{A}_k = \left[\mathbf{0}_{\mathrm{B}\tau-\mathrm{K},\mathrm{k}} \;\; \mathbf{I}_{\mathrm{B}\tau-\mathrm{K}} \;\; \mathbf{0}_{\mathrm{B}\tau-\mathrm{K},\mathrm{B}\tau-\mathrm{k}} \right] \mathbf{W}$$

with **W** the DFT Matrix of size Bτ by N

3. Infinite sinc filter (non-periodic) [VBM:02]

The matrices $\mathbf{A}_k$ are obtained through the product of a band-diagonal matrix and a diagonal one.

4. Gaussian filter [VMB:02]: $\varphi(t) = e^{-\frac{t^2}{\sigma^2}}$

Relation ω_k to t_k : $\omega_k = e^{\frac{2t_k T}{\sigma^2}}$

and the matrices $\mathbf{A}_k$ are given by

$$\mathbf{A}_k = \left[\mathbf{0}_{\mathrm{N}-\mathrm{K},\mathrm{k}} \;\; \mathbf{I}_{\mathrm{N}-\mathrm{K}} \;\; \mathbf{0}_{\mathrm{N}-\mathrm{K},\mathrm{K}-\mathrm{k}} \right] \mathrm{Diag}\left[\mathrm{e}^{\frac{(\mathrm{nT})^2}{\sigma^2}} \right]$$

Generalized Annihilation Property: Examples

5. Strang-Fix Filters [DBV:07]

The filter φ(t) and its shifts is able to reconstruct polynomials up to degree L-1 (or exponentials):

$$\sum_{n \in Z} c_{l,n} \varphi(nT - t) = t^l \quad \text{where } l = 0, 1, \ldots, L-1,$$

The matrices $\mathbf{A}_k$ are given by

$$\mathbf{A}_k = \begin{bmatrix} c_{k,1} & c_{k,2} & \cdots & c_{k,N} \\ c_{k-1,1} & c_{k-1,2} & \cdots & c_{k-1,N} \\ \vdots & \vdots & & \vdots \\ c_{k-L+1,1} & c_{k-L+1,2} & \cdots & c_{k-L+1,N} \end{bmatrix}.$$

Additionally, there is a density constraint (since φ(t) has compact support).

Generalized Annihilation Property: Measuring moments

Example of kernel satisfying Strang-Fix construction [DVB07]

The filter φ(t) and its shifts is able to reconstruct 1, t, t^2 … t^N :

$$\sum_{n \in Z} c_{m,n} \varphi(nT - t) = t^m \quad \text{where } m = 0, 1, \ldots, N,$$

1. Calculate weighted samples, which can be shown to lead to

$$\tau_m = \sum_n c_{m,n} y_n = \sum_{k=0}^{K-1} x_k t_k^m,$$

2. Solve for annihilating filter

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

where $$H(z) = \sum_{m=0}^{K} h_m z^{-m} = \prod_{k=0}^{K-1} (1 - t_k z^{-1}).$$

Generalized Annihilation Property: Measuring moments

2. (cont.) Solution for annihilating filter in matrix form:

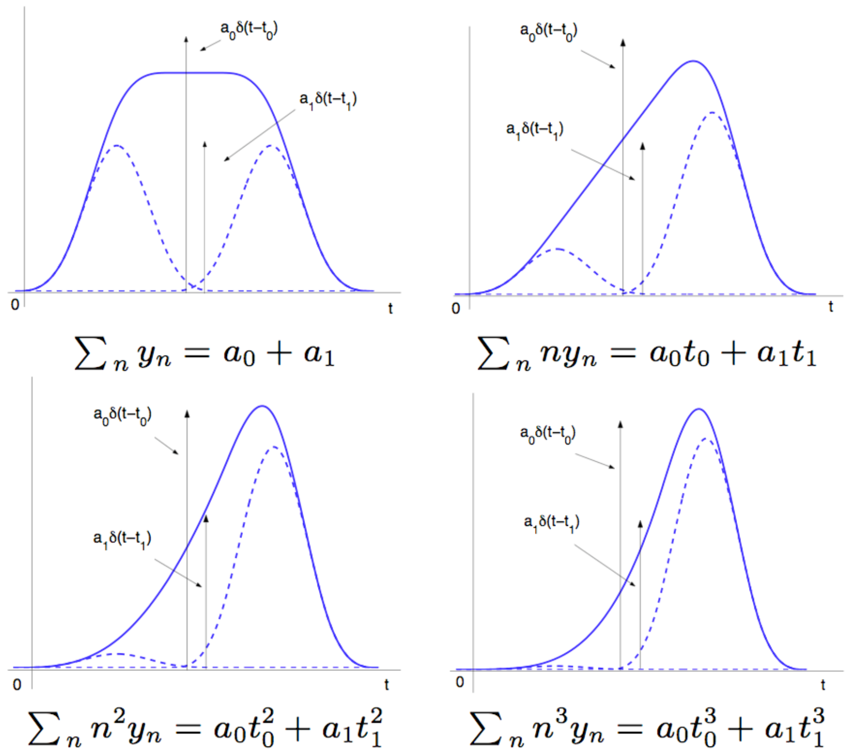
$$\begin{bmatrix} \tau_{K-1} & \tau_{K-2} & \cdots & \tau_0 \\ \tau_K & \tau_{K-1} & \cdots & \tau_1 \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{N-1} & \tau_{N-2} & \cdots & \tau_{N-K} \end{bmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_K \end{pmatrix} = - \begin{pmatrix} \tau_K \\ \tau_{K+1} \\ \vdots \\ \tau_N \end{pmatrix}.$$

3. Solve for weights x_k or Vandermonde system:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_0 & t_1 & \cdots & t_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ t_0^{K-1} & t_1^{K-1} & \cdots & t_{K-1}^{K-1} \end{bmatrix} \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{K-1} \end{pmatrix} = \begin{pmatrix} \tau_0 \\ \tau_1 \\ \vdots \\ \tau_{K-1} \end{pmatrix}.$$

A local algorithm for sparse sampling

The return of Strang-Fix!



Local, polynomial complexity reconstruction, for diracs and piecewise polynomials

Outline

1. **Introduction**
2. **Signals with finite rate of innovation**
3. **Sampling signals at their rate of innovation**
4. **The noisy case**
5. **Extensions to non-periodic cases and other kernels**
6. **Discussion and conclusions**
    - Applications
    - Relation to compressed sensing
    - Outlook
    - Publications

Applications: The proof of the pie is in …

1. Look for sparsity!
- In either domain, in any basis
- Different singularities
- Model is key
- Physics might be at play

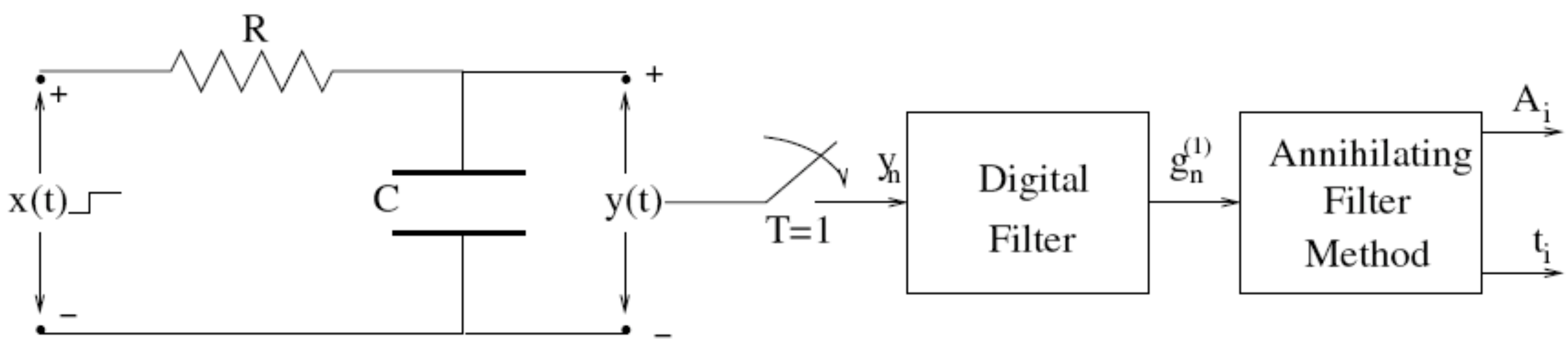
2. Choose operating point, algorithm,
- Low noise, high noise?
- Model precise of approximate?
- Choice of kernel?
- Algorithmic complexity? Iterative?

3. It is not a black box….
- It is a toolbox
- Handcraft the solution
- No free lunch

Applications: Sense real world, with acquisition circuit

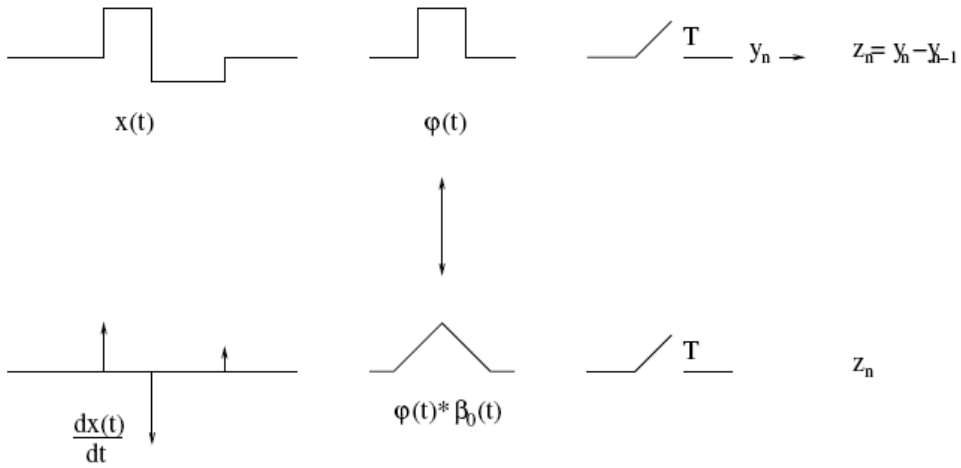
Example

– Circuits ore often ODE



A bit of spline algebra

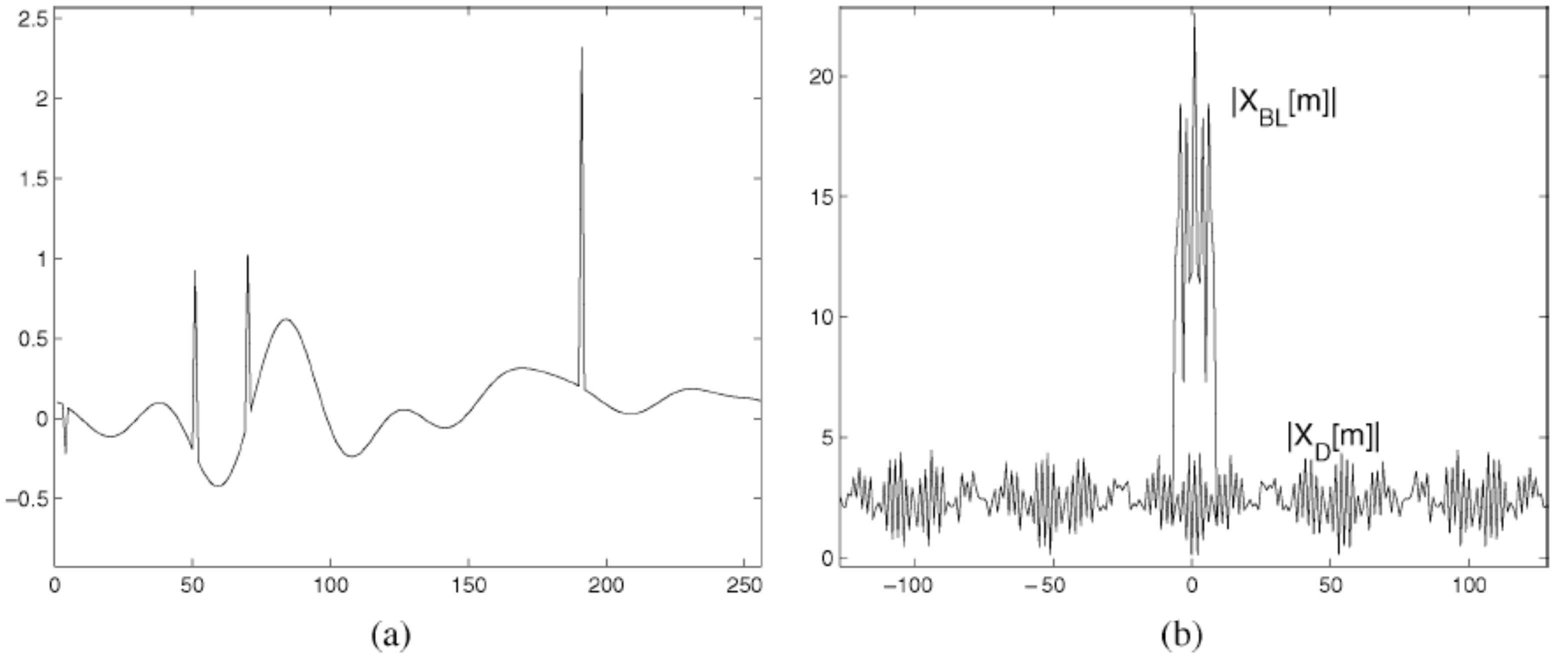
- back to Diracs
- perfect edge



Applications: Shot noise

FRI sampling results to remove shot noise from BL signal

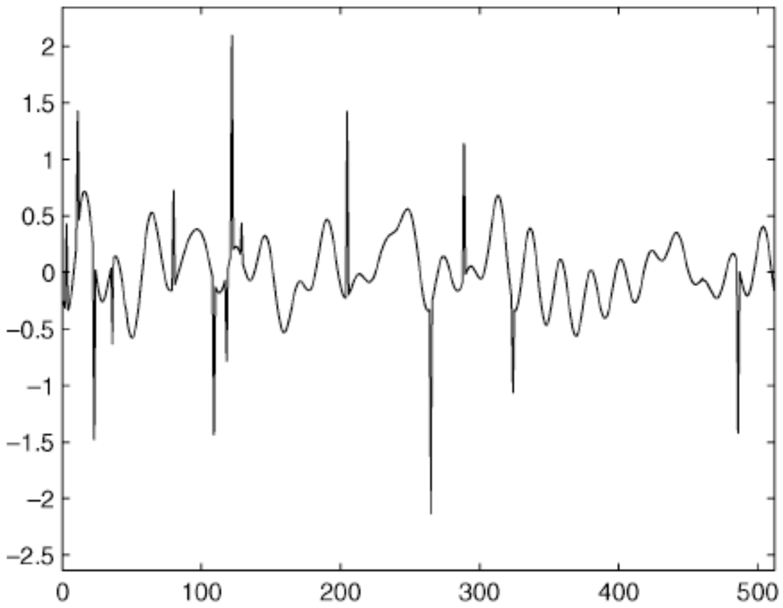
- $x(t) = x_{BL}(t) + x_D(t)$
- Estimate Diracs from outside of BL
- Remove Diracs influence of the BL spectrum
- Number of samples: order BL + number of Diracs



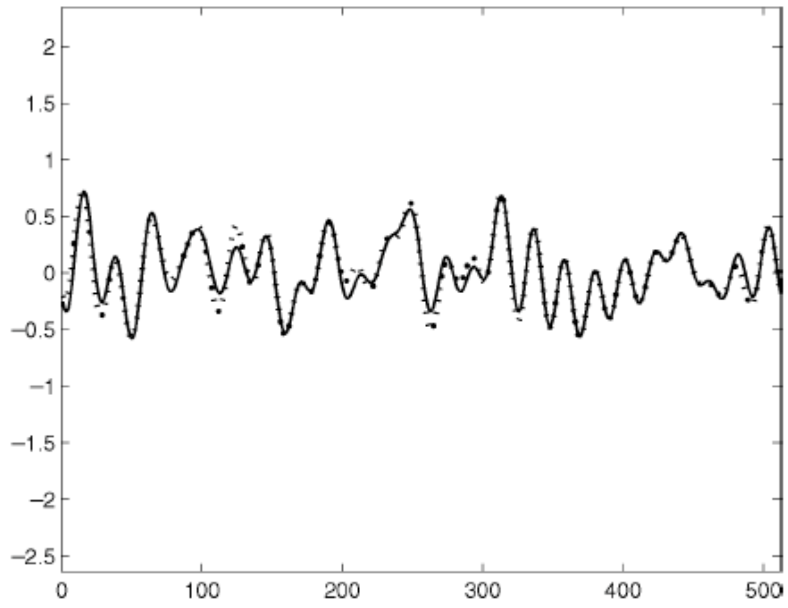
Applications: Shot noise

FRI sampling results to remove shot noise from BL signal

- $x(t) = x_{BL}(t) + x_D(t)$
- Example: 256 samples, 16 Diracs and 51 Fourier components
- Take 128 samples (> 50 + 4*16 + 1 = 115)
- Reconstruction 1: Lowpass
- Reconstruction 2 : Sparse sampling -> perfect



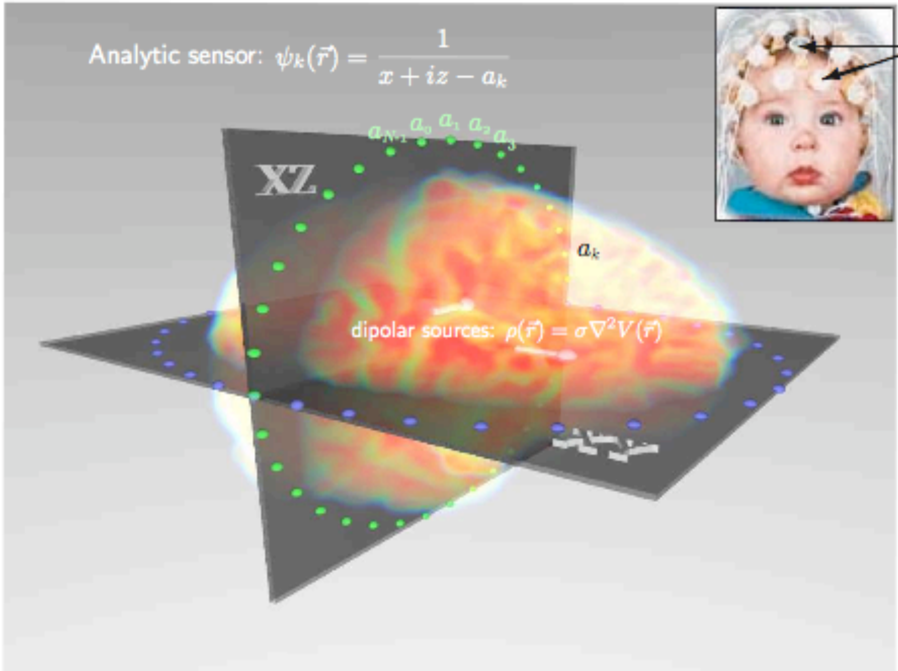
(a)



(b)

Applications: EEG and epylepsy [Blu et al]

Electroencephalographic source localization: an FRI problem



measured potential $V_{\text{skull}}(\vec{r})$

$$\underbrace{V_{\text{skull}}(\vec{r}) \xrightarrow{\text{Green's theorem}} \begin{pmatrix} \langle \psi_0, \rho \rangle \\ \langle \psi_1, \rho \rangle \\ \langle \psi_2, \rho \rangle \\ \vdots \\ \langle \psi_{N-1}, \rho \rangle \end{pmatrix}}_{\text{Analytic sensing}} \xrightarrow{\text{Generalized Annihilation}} \rho(\vec{r}) = \underbrace{\sum_k \vec{c}_k . \vec{\nabla} \delta(\vec{r} - \vec{r}_k)}_{\text{Finite Rate of Innovation}}$$

Applications: Superresolution [Dragotti et al]

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

- Large number of low-resolution and shifted versions of the original.
- Given the number of cameras and the low resolution, registration is critical.
- Accurate registration is achieved by modelling precisely the acquisition and compression process.
- The registered images are interpolated to achieve super-resolution.

Applications: Superresolution [Dragotti et al]

Problem: unknown shifts!
Method: retrieve moments to register images

- A pixel $P_{m,n}$ in the compressed image is given by

$$P_{m,n} = \langle I(x,y), \varphi(x/2^J - n, y/2^J - m) \rangle.$$

- The scaling function $\varphi(x,y)$ can reproduce polynomials:

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

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

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

- Register the compressed images using the retrieved moments.

Relation to compressed sensing

Cute work on sparsity and measurements [Candes, Tao, Donoho]

Model:

- Discrete, finite dimensional
- $x \in R^N$, but $|x|_0 = K << N$: K sparse

Method:

- Measure x with a fat matrix of size M x N, where M ~ K

$$y = Mx$$

Algorithm:

- From y, recover x using the prior that x is K-sparse
- Replace combinatorial problem (N choose K) by linear program

Strong results:

- l_1 minimization finds, with high probability, sparse solution, or l_0 and l_1 problems have same solution
- Redundancy small
- Universality of measurement matrices (e.g. random)

Relation to compressed sensing: Differences

Sparse Sampling of Signal Innovations

+ Continuous or discrete, infinite or finite dimensional
+ Lower bounds (CRB) provably optimal reconstruction
+ Close to “real” sampling
+ At Occam’s bound

± Deterministic

± High noise, SVD, complex

– Not universal, designer matrices

Compressed sensing

+ Universal and more general

± Probabilistic

± Can be complex

– Discrete

– Redundant

Conclusions and Outlook

Sampling at the rate of innovation

- Cool!
- Sharp theorems
- Robust algorithms
- Provable optimality over wide SNR ranges

Many actual and potential applications

- Fit the model (needs some work)
- Apply the ‘’right’’ algorithm
- Catch the essence, the wheat from the chaff

Still a number of good questions open, from the fundamental to the algorithmic and the applications

Sort the wheat from the chaff… at Occam’s rate!

Publications

Basic paper:

– M.Vetterli, P. Marziliano and T. Blu, “Sampling Signals with Finite Rate of Innovation,” IEEE Transactions on Signal Processing, June 2002.

Main paper, with comprehensive review:

– 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, Special issue on Compressed Sampling, to appear, 2008.

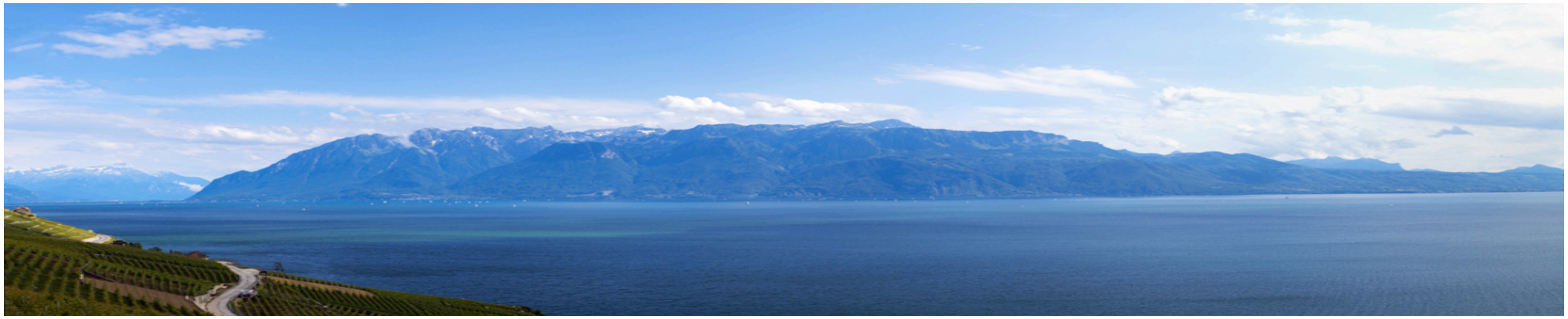
For more details, recent results:

– P.L. Dragotti, M. Vetterli and T. Blu, “Sampling Moments and Reconstructing Signals of Finite Rate of Innovation: Shannon Meets Strang-Fix,” IEEE Transactions on Signal Processing, May 2007.

– P. Marziliano, M. Vetterli and T. Blu, Sampling and exact reconstruction of bandlimited signals with shot noise, IEEE Transactions on Information Theory, Vol. 52, Nr. 5, pp. 2230-2233, 2006.

– I. Maravic and M. Vetterli, Sampling and Reconstruction of Signals with Finite Rate of Innovation in the Presence of Noise, IEEE Transactions on Signal Processing, Aug. 2005.

Thank you for your attention!



http://panorama.epfl.ch

Questions?