Sparsity Through Annihilation

Algorithms and Applications

Thierry Blu

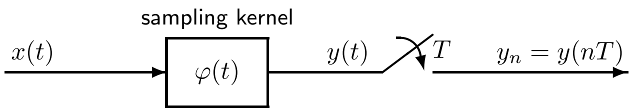
Department of Electronic Engineering
The Chinese University of Hong Kong

ICSP, 25 October 2010

Outline

A very old problem

Given a sampling device that provides smooth, *uniform* samples y_n of a "real-world" function $x(t)$



How to reconstruct $x(t)$ *exactly*, and under which conditions?

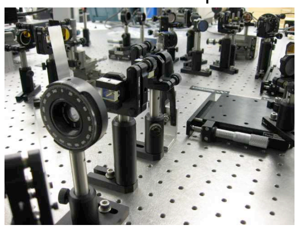
NOTE: Implicitely, there is the assumption that if the samples are *shifted*, then the reconstruction should also be shifted by the *same amount*.

Observation kernel $\varphi(t)$: often given partly by nature, partly by design.

Hubble telescope

Electro-EncephaloGraphy

OCT Set-Up



MRI scanner

Standard solution (from Shannon, Whittaker, Kotel'nikov, Nyquist,. . .)

If $x(t)$ is *band-limited* in $]-\pi/T, \pi/T[$ and $\hat{\varphi}(\omega) \neq 0$ in that band, then the knowledge of its samples y_n at the frequency $1/T$ allows to reconstruct $x(t)$ *uniquely* by

$$x(t) = \sum_{n \in \mathbb{Z}} y(nT)\, \psi(t - nT)$$

where $(\varphi * \psi)(t) = \mathrm{sinc}(t/T)$.

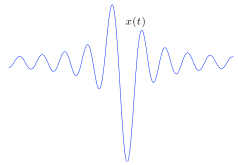
Problems $\rightsquigarrow$ need for a better adapted signal model

- the samples are almost always in *finite* number
- a natural signal is *never* band-limited
- noise sensitivity of Shannon's formula

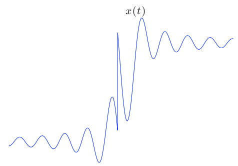
NOTE: Replacing sinc by other "basis" functions (e.g., splines) addresses these issues, but fails to produce *shift-invariant* solutions.

Shannon's nightmare

An *ideal band-limited* signal $x(t)$ can be represented exactly by its samples $x(nT)$



But a *single discontinuity* and no more sampling theorem.



NOTE: Bandlimited signals are represented using $1/T$ degrees of freedom per unit of time.

Are there other shift invariant signal families with finite numbers of degrees of freedom per unit of time, and allowing perfect reconstruction?

Signals with Finite Rate of Innovation

A novel signal model, that emphasizes the *duality* of the "information" —the innovation— conveyed by a signal

- A *linear* aspect : e.g., the amplitude of a sample
- A *nonlinear* aspect: e.g., a time of change of the signal

The FRI hypothesis[1] 

A Finite Rate of Innovation signal can be expressed as the convolution of an acquisition window with a stream of Diracs

$$y(t) = \left(\sum_{k=-\infty}^{+\infty} x_k\, \delta(t - t_k) \right) * \varphi(t) = \sum_{k=-\infty}^{+\infty} x_k\, \varphi(t - t_k)$$

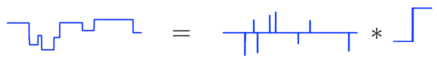
x_k and t_k are called the *innovations* of the signal.

Rate of innovation: the average number of innovations per unit of time.

[1] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. on Signal Processing*, vol. 50, pp. 1417–1428, June 2002.

Examples

- Piecewise-constant signals



- OCT signals: convolution with a Gabor window

- . . . and many more "sparse" signals

Are there interpolation formulas for such signals?

Annihilation of periodic signals

Consider the case

- τ-*periodic* signal $x(t) = x(t+\tau)$, where $\tau = NT$, N integer
- $\varphi(t) = \mathrm{sinc}(Bt)$ with $BT = \frac{2M+1}{N} \leq 1$, M integer
- rate of innovation, $2K/\tau \leq B$ (K = number of Diracs in $[0, \tau]$)

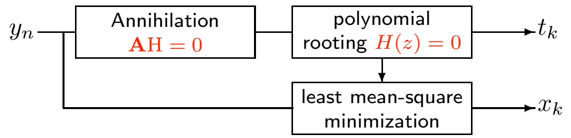
Then the filter of transfer function $H(z) = \prod_{k=1}^{K}(1 - \mathrm{e}^{-j2\pi\frac{t_k}{\tau}} z^{-1})$ *annihilates* the N-DFT coefficients of y_n

$$\sum_{k=0}^{K} h_k \hat{y}_{m-k} = 0, \quad m = -M+K, \dots M$$

Under algebraic form, the annihilation equation becomes $\mathbf{A}\mathrm{H} = 0$, where $\mathbf{A}$ is a Tœplitz matrix

$$\mathbf{A} = \begin{bmatrix} \hat{y}_{-M+K} & \hat{y}_{-M+K-1} & \cdots & \hat{y}_{-M+1} & \hat{y}_{-M} \\ \hat{y}_{-M+K+1} & \hat{y}_{-M+K} & \cdots & \hat{y}_{-M+2} & \hat{y}_{-M+1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \hat{y}_{M} & \hat{y}_{M-1} & \cdots & \hat{y}_{M-K+1} & \hat{y}_{M-K} \end{bmatrix}$$

Hence, an exact reconstruction algorithm looks like



A non-iterative solution to a non-linear problem:
two linear systems to solve + polynomial root extraction

Other annihilation examples

Consider the Gaussian case

- $\varphi(t) = \mathrm{e}^{-t^2/(2\sigma^2)}$
- K Diracs to retrieve from N samples $n \in [-N/2, N/2]$

Then the filter of transfer function $H(z) = \prod_{k=1}^{K}(1 - \mathrm{e}^{\frac{t_k T}{\sigma^2}} z^{-1})$ *annihilates* the samples $\tilde{y}_n = \mathrm{e}^{(nT/\sigma)^2/2} y_n$

$$\sum_{k=0}^{K} h_k \tilde{y}_{n-k} = 0, \quad m = -N/2+K, \dots N/2$$

Consider the non-periodic sinc case

- $\varphi(t) = \mathrm{sinc}(t/T)$
- K Diracs to retrieve from N samples $n \in [-N/2, N/2]$

Then the filter of transfer function $H(z) = (1 - z^{-1})^K$ *annihilates* the samples $\tilde{y}_n = (-1^n)P(n)y_n$ where $P(n) = \prod_{k=1}^{K}(n - t_k/T)$

$$\sum_{k=0}^{K} h_k \tilde{y}_{n-k} = 0, \quad m = -N/2+K, \dots N/2$$

Consider kernels that satisfy Strang-Fix conditions of order $L \geq 2K$

- either $\{1, t, t^2, \ldots t^{L-1}\} \in \operatorname{span}_n\{\varphi(nT - t)\}$
- or $\{e^{at}, e^{(a+b)t}, e^{(a+2b)t}, \ldots e^{(a+(L-1)b)t}\} \in \operatorname{span}_n\{\varphi(nT - t)\}$

Then the filter of transfer function $H(z) = \prod_{k=1}^{K}(1 - e^{bt_k} z^{-1})$ *annihilates* modified samples $\tilde{y}_n$

$$\sum_{k=0}^{K} h_k \tilde{y}_{n-k} = 0, \quad m = K, K+1, \ldots L$$

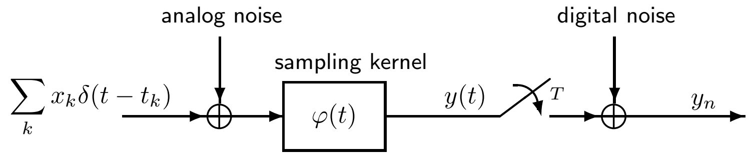
The $\tilde{y}_n$ are obtained by an adequate linear transformation of the y_n.

A very large range of of observation/analysis kernels (wavelets, etc.)

[1]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, pp. 1741–1757, May 2007.

FRI with noise

Schematical acquisition of a τ-periodic FRI signal with noise



Modelization

$$y_n = \sum_k x_k \varphi(nT - t_k) + \varepsilon_n$$

The noisy periodic case

- *τ-periodic* signal $x(t) = x(t + \tau)$, where $\tau = NT$, N integer
- $\varphi(t) = \operatorname{sinc}(Bt)$ with $BT = \frac{2M+1}{N} \leq 1$, M integer
- rate of innovation, $2K/\tau \leq B$ (K = number of Diracs in $[0, \tau]$)

Estimation problem

Find estimates $\bar{y}_n$, $\bar{x}_k$ and $\bar{t}_k$ of y_n, x_k and t_k such that

- $\bar{y}_n = \sum_k \bar{x}_k \varphi(nT - \bar{t}_k)$
- $\|\bar{\mathbf{y}} - \mathbf{y}\|_{\ell^2}$ is as small as possible

Total least-squares

Replace the annihilation equation $\mathbf{A}\mathrm{H} = 0$ by

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

Solution

Perform a *Singular Value Decomposition*

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

and choose the last column of $\mathbf{V}$ for H.

- $\mathbf{U}$ is *unitary* of same size as $\mathbf{A}$
- $\mathbf{S}$ is *diagonal* (with decreasing coefficients) and of size $(K+1) \times (K+1)$
- $\mathbf{V}$ is *unitary* and of size $(K+1) \times (K+1)$

Total least-squares

The estimation of the innovations are then obtained as follows

- t_k: by finding the roots of the polynomial $H(z)$
- x_k: by least-square minimization of

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

NOTE:

- Related to Pisarenko method
- Not robust with respect to noise $\rightsquigarrow$ need for extra denoising

Cadzow iterated denoising

Without noise, the annihilation property $\mathbf{A}\mathrm{H} = 0$ still holds if length(H) $= L+1$ is *larger* than $K+1$. We have the properties

- $\mathbf{A}$ is still of rank K
- $\mathbf{A}$ is a Tœplitz matrix
- conversely, if $\mathbf{A}$ is Tœplitz and has rank K, then y_n are the samples of an FRI signal

Rank K "projection" algorithm

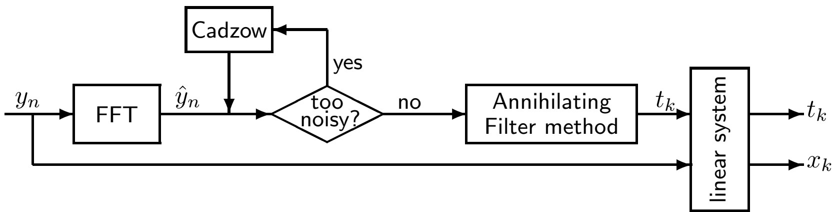
1. Perform the SVD of $\mathbf{A}$: $\mathbf{A} = \mathbf{USV}^{\mathrm{T}}$
2. Set to zero the $L-K+1$ smallest diagonal elements of $\mathbf{S} \rightsquigarrow \mathbf{S}'$
3. build $\mathbf{A}' = \mathbf{US}'\mathbf{V}^{\mathrm{T}}$
4. find the Tœplitz matrix that is closest to $\mathbf{A}'$ and goto step 1

Cadzow iterated denoising

Essential details

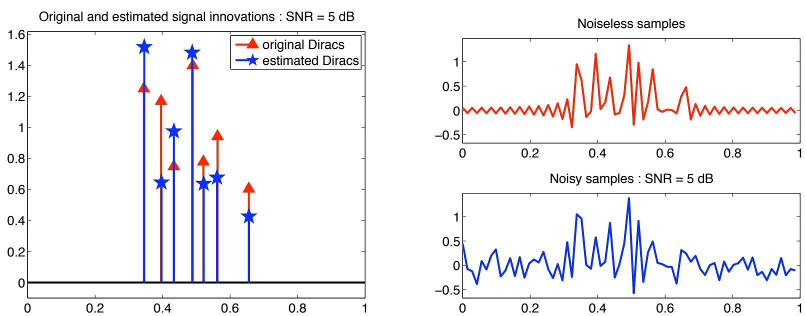
- Iterations of the projection algorithm are performed *until the matrix $\mathbf{A}$ is of "effective" rank K*
- L is chosen *maximal*, i.e., $L = M$

Schematical view of the whole retrieval algorithm



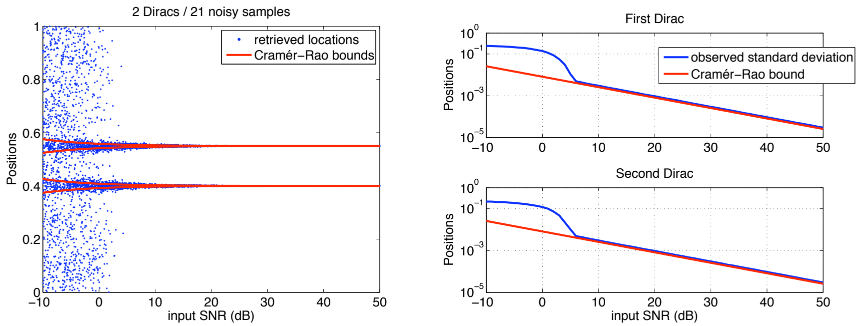
[1]T. Blu et al., "Sparse Sampling of Signal Innovations," *IEEE Signal Processing Magazine*, vol. 25, pp. 31–40, March 2008.

Examples



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

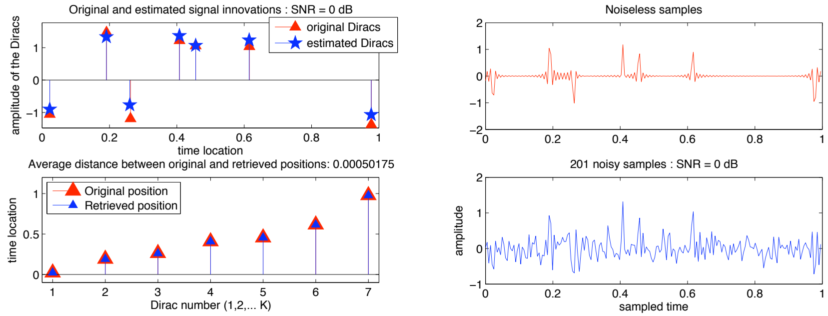
Simulations: Quasi-optimality



Retrieval of the locations of a FRI signal. Left: scatterplot of the locations; right: standard deviation (averages over 10000 realizations) compared to Cramér-Rao lower bounds.

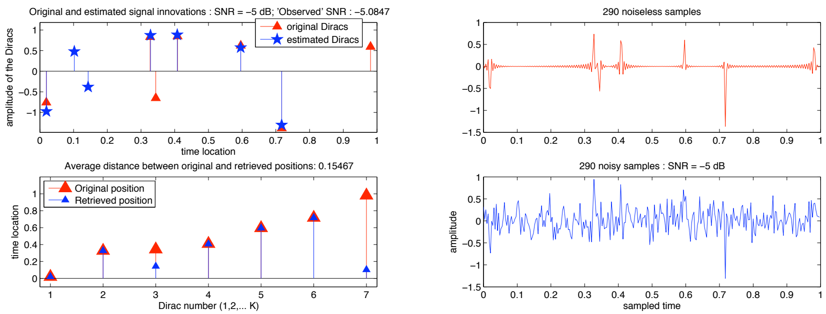
Quasi-optimality of the algorithm.

Simulations: Robustness



201 samples of an FRI signal in 0 dB noise. Right: noiseless and noisy signal. Left: retrieved locations and amplitudes.

Simulations: Robustness



290 samples of an FRI signal in -5 dB noise. Right: noiseless and noisy signal. Left: retrieved locations and amplitudes.

In high noise levels, the algorithm is still able to find *accurately* a substantial proportion of Diracs

Optical Coherence Tomography: Principle

Detection of coherent backscattered waves from an object by making *interferences* with a *low-coherence* reference wave. Measurement performed with a standard Michelson interferometer.

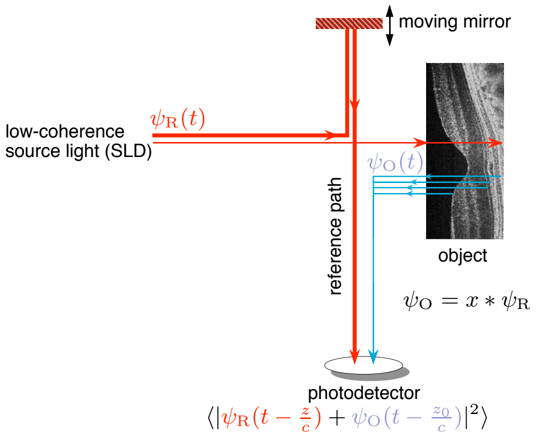
Axial (depth) resolution: $\propto$ *coherence length* of the reference wave;
Transversal resolution: width of the optical beam.

NOTE: *Very high sensitivity* (low SNR), noninvasive, low-depth penetration
$\rightsquigarrow$ biomedical applications (ophthalmology, dentistry, skin).
Axial resolution[2]: 10→20μm. Better resolution → better diagnoses.

OCT is a ranging application!

[2]with SuperLuminescent Diods: low cost, compact, easy to use.

OCT: Experimental Setup



NOTE: z, z_0 are the optical path lengths of the reference and the object wave.

OCT: Mathematical setting

Measured intensity:

$$I_{\text{photo}}(z_0 - z) = \left\langle\left|\psi_R(t-\tfrac{z}{c}) + \psi_O(t-\tfrac{z_0}{c})\right|^2\right\rangle$$
$$= \text{const} + \underbrace{2\Re\left\{(x*\varphi)\left(\tfrac{z-z_0}{c}\right)\right\}}_{\text{OCT signal}}.$$

variable (moving mirror)

$\varphi(t)$ is the temporal *coherence function* of the reference wave:

$$\varphi(t'-t) = \langle\psi_R(t'), \psi_R(t)\rangle$$

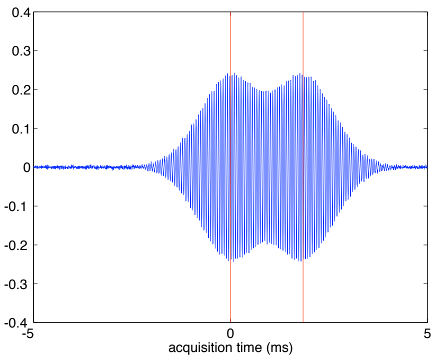
Typically, $\varphi(t) \propto e^{-t^2/(2\sigma^2)+2i\pi\nu_0 t}$ and $x(t)$ is a stream of Diracs characterizing the depth of the interfaces, and the refractive index jumps.

An FRI interpolation problem

Retrieve $x(t)$ from the uniform samples of the OCT signal.

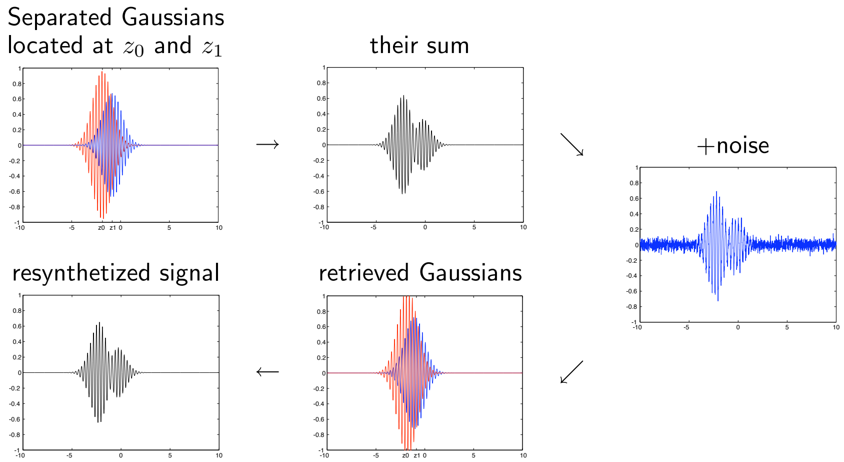
OCT: Resolution

Resolution limit of OCT (two interfaces): $L_c = c \times \text{FWHM of } A(t)$



NOTE: Because the light travels twice (forward then backward) in the object, the actual *physical resolution* is $L_c/2$. Moreover, a larger value of refraction index inside the object further divides the resolution limit.

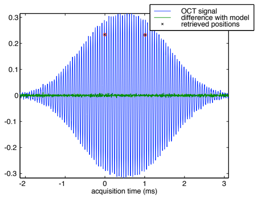
OCT: Example of Processing



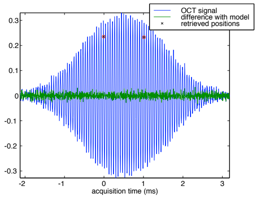
OCT: Simulation example

Simulation examples: two interfaces distant by 7μm (1ms below)

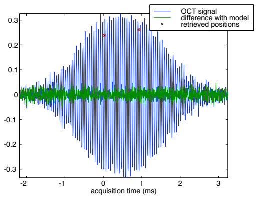
PSNR 40dB



PSNR 30dB



PSNR 25dB



OCT: Real Data Processing

SLD source of central wavelength 0.814μm and coherence length 25μm.
$\rightsquigarrow$ OCT resolution of 12.5μm.

Depth scan of a 4μm thick pellicle beamsplitter of an optical[3] depth of 6.6μm $\rightsquigarrow$ approximately *half* the OCT resolution.

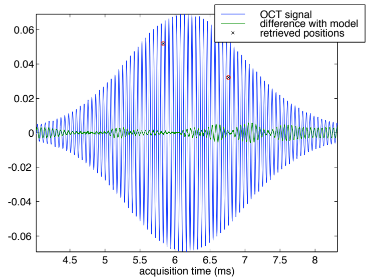
Calibration part:

- Depth scan of 1 interface $\rightsquigarrow$ effective coherence function;
- High-coherence interferometer $\rightsquigarrow$ accurate position of the moving mirror.

[3] refractive index 1.65.

OCT: Real Data Processing

Example of superresolution: data-model PSNR=31dB



The two retrieved interfaces are distant by 17 interference fringes
$\rightsquigarrow 17 \times 0.814/2 = 6.9\mu\text{m}$.

Presentation of a generic framework for interpolating samples under sparsity assumptions

- Super-resolution applications with noise-robust behaviour
- Unique solution as soon as $2K$ measurements for $2K$ unknowns
- Patents on the Dirichlet kernel transferred to Qualcomm

Papers available at `http://www.ee.cuhk.edu.hk/~tblu/`