Imperial College London

Sparse Sampling

Pier Luigi Dragotti[1]

December 17, 2012

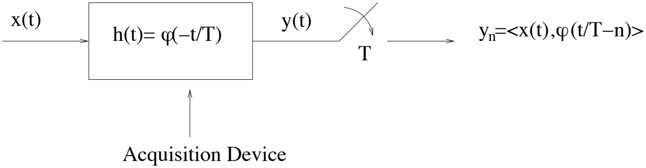
[1] Pier Luigi Dragotti is supported by the European Research Council (ERC) starting investigator award Nr. 277800 (RecoSamp).

Outline

- Problem Statement and Motivation
- Classical Sampling Formulation
- Sampling using expansion-based sparsity
  - Sparsity in Complete and Over-complete Dictionaries
  - Compressed Sensing
  - Applications
- Sampling using sparsity in parametric spaces
  - Signals with Finite Rate of Innovation (FRI)
  - Sampling Kernels: E-splines and B-splines
  - Sampling FRI Signals: the Basic Set-up and Extensions
  - Applications
- New Domains of Applications of the Sparsity and Sampling Paradigm
  - Diffusion Fields and Neuroscience
- Conclusions and Outlook

Problem Statement

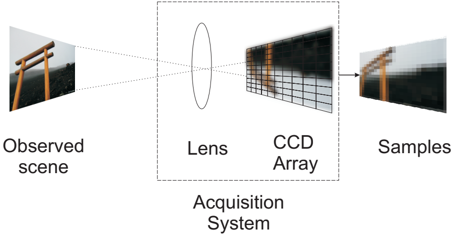
You are given a class of functions. You have a sampling device. Given the measurements $y_n = \langle x(t), \varphi(t/T - n)\rangle$, you want to reconstruct $x(t)$.



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?

Problem Statement



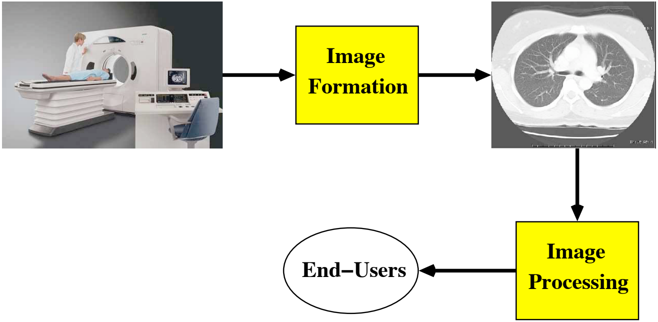
- The low-quality lens blurs the images.
- The images are sampled by the CCD array.

Outline

- Problem Statement and Motivation
- Classical Sampling Formulation
- Sampling using expansion-based sparsity
  - Sparsity in Complete and Over-complete Dictionaries
  - Compressed Sensing
  - Applications
- Sampling using sparsity in parametric spaces
  - Signals with Finite Rate of Innovation (FRI)
  - Sampling Kernels: E-splines and B-splines
  - Sampling FRI Signals: the Basic Set-up and Extensions
  - Applications
- New Domains of Applications of the Sparsity and Sampling Paradigm
  - Diffusion Fields and Neuroscience
- Conclusions and Outlook

Motivation: Sparsity and Sampling Everywhere

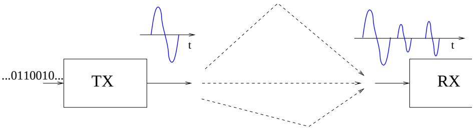
"In 2005, the U.S. spent 16% of its GDP on health care. It is projected that this will reach 20% by 2015." Goal: Individualized treatments based on low-cost and effective medical devices.



Imperial College London

Motivation: Sparsity and Sampling Everywhere

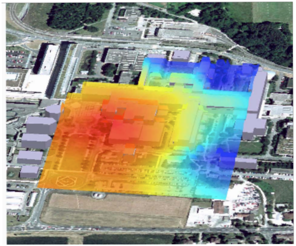
Wide-Band Communications:



- Current A-to-D converters in UWB communications operate at several gigaherz.
- This is a sparse parametric estimation problem, only the location and amplitude of the pulses need to be estimated.

Motivation: Sparsity and Sampling Everywhere

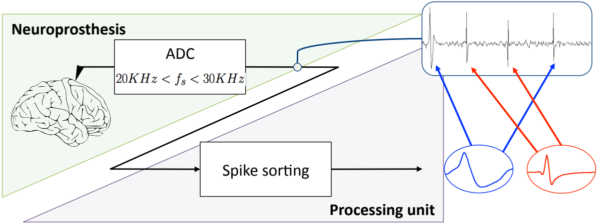
Sensor networks



- The source (phenomenon) is distributed in space and time.
- The phenomenon is sampled in space (finite number of sensors) and time.
- When the sources are localized the problem is sparse.

Motivation: Sparsity and Sampling Everywhere

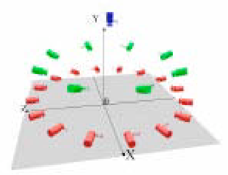
Applications in Neuroscience

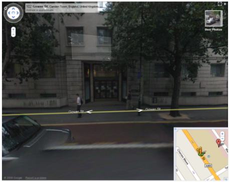


- Implanted neuronal prostheses require low-processing and low-sampling rate.
- Spike sorting is based on a sparse description of the action potentials.

Motivation: Free Viewpoint Video

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

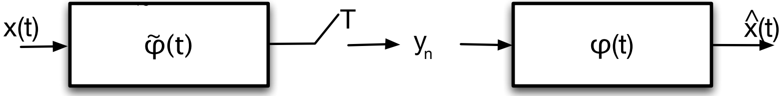




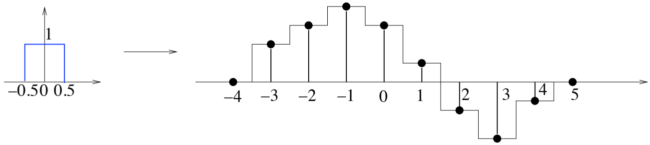
- This is a multi-dimensional sampling and interpolation problem.

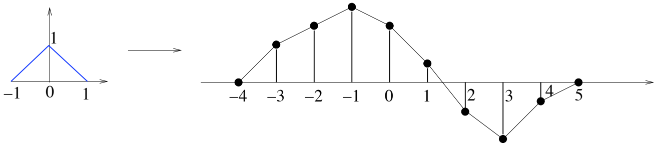
Classical Sampling Formulation

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



Sampling as Projecting into Shift-Invariant Sub-Spaces





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

Key elements in the novel sampling approaches

Classical Sampling Formulation:

- In classical sampling formulation, the reconstruction process is linear.
- Innovation is uniform.

New formulation:

- The reconstruction process can be non-linear.
- Innovation can be non-uniform.

Sparse Representations in a Basis

Wavelets provide sparse representations of images. In matrix/vector form

$$\alpha = W^{-1} Y$$

is sparse. Here the matrix W has size $N \times N$ and models the discrete-time wavelet transform of finite dimensional signals.

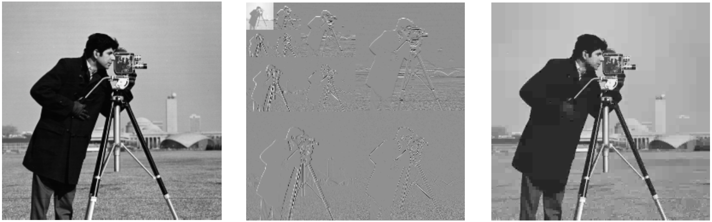


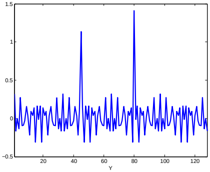
Figure: Cameraman is reconstructed using only 8% of the wavelet coefficients.

Notation

- The l_0 'norm' of a N-dimensional vector X is $\|X\|_0 =$ the number of i such that $x_i \neq 0$
- The $l_1$norm of a N-dimensional vector X is: $\|X\|_1 = \sum_{i=1}^{N} |x_i|$
- The *Mutual Coherence* of a given $N \times M$ matrix A is the largest absolute normalized inner product between different columns of A:

$$\mu(A) = \max_{1 \leq k,j \leq M; k \neq j} \frac{|\mathbf{a_k^T a_j}|}{\|\mathbf{a_k}\|_2 \cdot \|\mathbf{a_j}\|_2}$$

Sparsity in Redundant Dictionaries



The above signal, Y, is a combination of two spikes and two complex exponentials of different frequency (real part of Y plotted). In matrix vector form:

$$Y = \begin{pmatrix} I_N & F_N \end{pmatrix} \alpha = D\alpha,$$

where I_N is the $N \times N$ identity matrix and F_N is the $N \times N$ Fourier transform. The matrix D models an over-complete dictionary and has size $N \times M$ with $M > N$, α has only K non-zero coefficients (in the example $K = 4, N = 128, M = 2N$).

Sparsity in Redundant Dictionaries

- You are given Y and want to find its sparse representation.
- Ideally, you want to solve

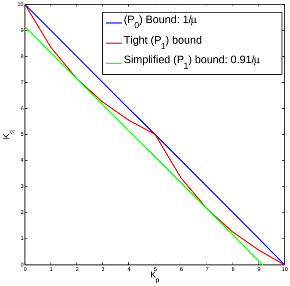
$$(P_0): \quad \min \|\alpha\|_0 \quad \text{s.t.} \quad Y = D\alpha.$$

- This is a combinatorial problem which requires N chooses K operations. You may instead solve the convex problem:

$$(P_1): \quad \min \|\alpha\|_1 \quad \text{s.t.} \quad Y = D\alpha.$$

- Key result due to Donoho et al.: (P_0) is unique when $K < 1/\mu(D) = \sqrt{N}$. (P_0) and (P_1) are equivalent when $K < (\sqrt{2} - 0.5)/\mu(D) \sim 0.9\sqrt{N}$.

Sparsity iBounds n Pairs of Bases



Uniqueness of (P_0) and the two l_1 bounds for the case of two orthogonal bases and $\mu(\mathbf{D}) = 0.1$. See [Elad 2010, page 59] for more details.

Sparsity in Redundant Dictionaries

Sketch on the proof of unicity of (P_0).

- (P_0) is unique when K is such that, given $Y_1 = D\alpha_1$ and $Y_2 = D\alpha_2$, then $Y_1 \neq Y_2$ for any possible K-sparse α_1, α_2.
- Consider $\alpha_n = \alpha_1 - \alpha_2$, this new vector has sparsity $2K$ and unicity is lost when $Y = D\alpha_n = 0$.
- α_n is in the null space of $D = \begin{pmatrix} I_N & F_N \end{pmatrix}$ when $\alpha_n = \begin{pmatrix} \hat{X} \\ -X \end{pmatrix}$, where $\hat{X} = F_N X$.
- In fact:
$$Y = D\alpha_n = \begin{pmatrix} I_N & F_N \end{pmatrix} \begin{pmatrix} \hat{X} \\ -X \end{pmatrix} = 0,$$
- X is an N dimensional vector and cannot be simultaneously sparse in both the time and the frequency domain. Donoho uncertainty principle says that the number of non-zero entries in α_n must be $2K \geq 2/\mu(D) = 2\sqrt{N}$. Thus, (P_0) can be solved when $K < \sqrt{N}$.

Sparsity in Redundant Dictionaries (cont'd)

Extensions [Tropp-04, GribonvalN:03, Elad-10]

- For a generic over-complete dictionary D, (P_1) is equivalent to (P_0) when[2]

$$K < \frac{1}{2}\left(1 + \frac{1}{\mu}\right).$$

- When D is a concatenation of J orthonormal dictionaries (P_1) is equivalent to (P_0) when

$$K < \left[\sqrt{2} - 1 + \frac{1}{2(J-1)}\right]\mu^{-1}$$

[2] Proof in Appendix.

Compressed Sensing

- The 'fat' matrix D now plays the role of the acquisition device and we denote it with Φ. The entries of $Y = \Phi\alpha$ are the samples.
- Based on the previous analysis, we want to reconstruct the signal α from the samples Y using l_1 minimization.
- We want maximum incoherence of the columns of Φ.
- We consider large M, N.

Key insight: Relax the condition of a 'deterministic' perfect reconstruction and accept that, with an extremely small probability, there might be an error in the reconstruction.

The power of randomness

- Key theorem due to Candès et al.[Candes:06-08]: if Φ is a proper random matrix (e.g., a matrix with normalized Gaussian entries), then with overwhelming probability the signal can be reconstructed from the samples Y when $N \geq C \cdot K \log(M/K)$ for some constant C.
- Assume that the measured signal X is not sparse but has a sparse representation: $X = D\alpha$. We have that $Y = \Phi X = \Phi D\alpha$. The new matrix ΦD is essentially as random as the original one. Therefore the theorem is still valid. Thus random matrices provides *universality*. However, very redundant dictionaries implies larger M and therefore larger N.

Restricted Isometry Property (RIP)

In order to have perfect reconstruction, Φ must satisfy the so called *Restricted Isometry Property*:

$$(1-\delta_S)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\delta_S)\|x\|_2^2$$

for some $0 < \delta_S < 1$ and for any S-sparse vector x.
Candes et al.:

- If x is K-sparse and $\delta_{2K} + \delta_{3K} < 1$ then the l_1 minimization finds x exactly.
- if Φ is a random Gaussian matrix, the above condition is satisfied with probability $1 - O(e^{-\gamma M})$ for some $\gamma > 0$, when $N \geq C \cdot K \log(M/K)$.
- if Φ is obtained by extracting at random N rows from the Fourier matrix, then perfect reconstruction is satisfied with high probability when:

$$N \geq C \cdot K(\log M)^4.$$

NB: When the signal x is not *exactly* sparse, solve:

$$\|y - \Phi\hat{x}\|_2 + \lambda\|\hat{x}\|_1$$

It is proved that linear programming achieve the best solution up to a constant factor.

Imperial College London

Compressed Sensing. Simulation Results

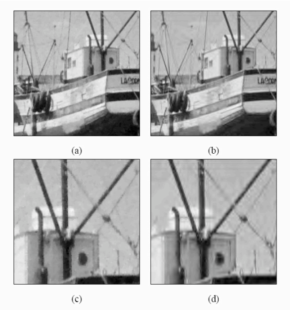


Image 'Boat'. (a) Recovered from 20000 random projections using Compressed Sensing. PSNR=31.8dB. (b) Optimal 7207-approximation using the wavelet transform with the same PSNR as (a). (c) Zoom of (a). (d) Zoom of (b). Images courtesy of Prof. J. Romberg.

Application in MRI

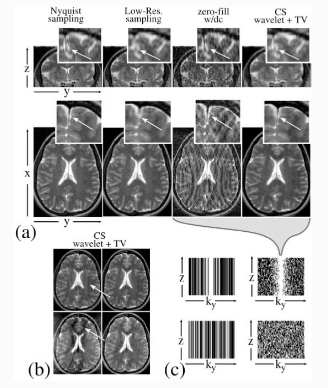
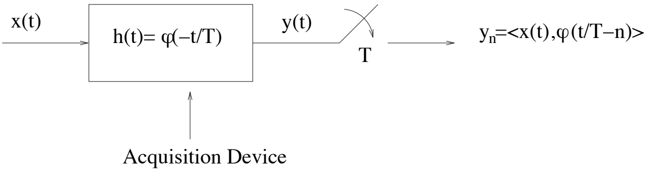


Image taken from Lustig, Donoho, Santos, Pauly-08.

Problem Statement

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 retrieve the degrees of freedom of $x(t)$.



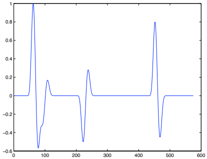
Natural questions:

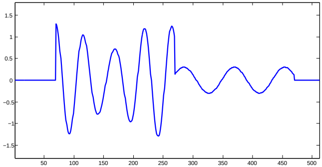
- When is there a one-to-one mapping between $x(t)$ and y_n?
- What kernels $\varphi(t)$ can be used?
- What reconstruction algorithm?

Imperial College London

Sparsity in Parametric Spaces

Consider a continuous-time stream of pulses or a piecewise sinusoidal 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.

Signals with Finite Rate of Innovation

Consider a signal of the form:

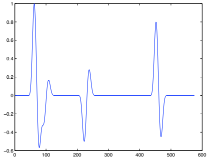
$$x(t) = \sum_{k \in \mathbb{Z}} \gamma_k g(t - t_k). \tag{1}$$

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

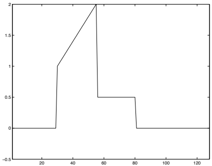
$$\rho = \lim_{\tau \to \infty} \frac{1}{\tau} C_x\left(-\frac{\tau}{2}, \frac{\tau}{2}\right), \tag{2}$$

where $C_x(-\tau/2, \tau/2)$ is a function counting the number of free parameters in the interval τ.
Definition 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).

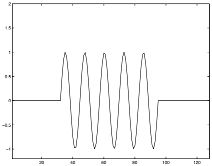
Examples of Signals with Finite Rate of Innovation



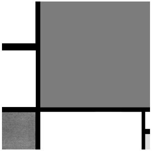
Filtered Streams of Diracs



Piecewise Polynomial Signals

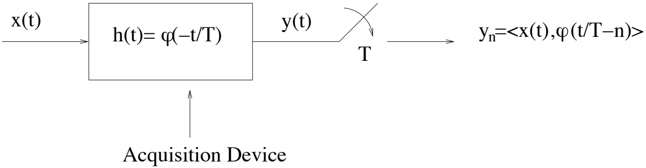


Piecewise Sinusoidal Signals



Mondrian paintings ;-)

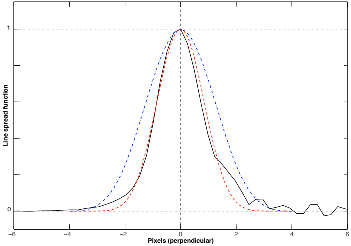
The Sampling Kernel



- 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.

The Sampling Kernel

(a)Original (2014 × 3039)



(b) Point Spread function

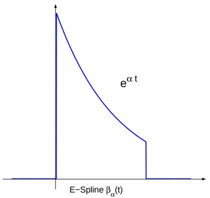
Imperial College London

Sampling Kernels

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

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

This includes any composite kernel of the form $\gamma(t) * \beta_{\vec{\alpha}}(t)$ where $\beta_{\vec{\alpha}}(t) = \beta_{\alpha_0}(t) * \beta_{\alpha_1}(t) * ... * \beta_{\alpha_L}(t)$ and $\beta_{\alpha_i}(t)$ is an Exponential Spline of first order [UnserB:05].

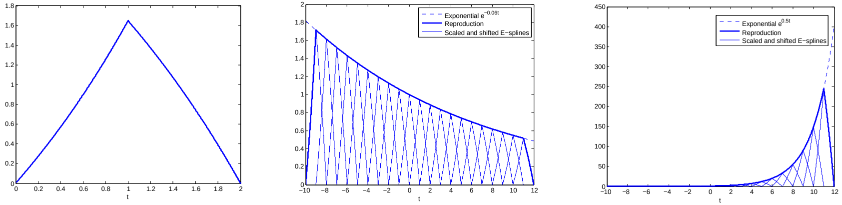


$$\beta_\alpha(t) \Leftrightarrow \hat{\beta}(\omega) = \frac{1 - e^{\alpha - j\omega}}{j\omega - \alpha}$$

Notice:

- α can be complex.
- E-Spline is of compact support.
- E-Spline reduces to the classical polynomial spline when $\alpha = 0$.

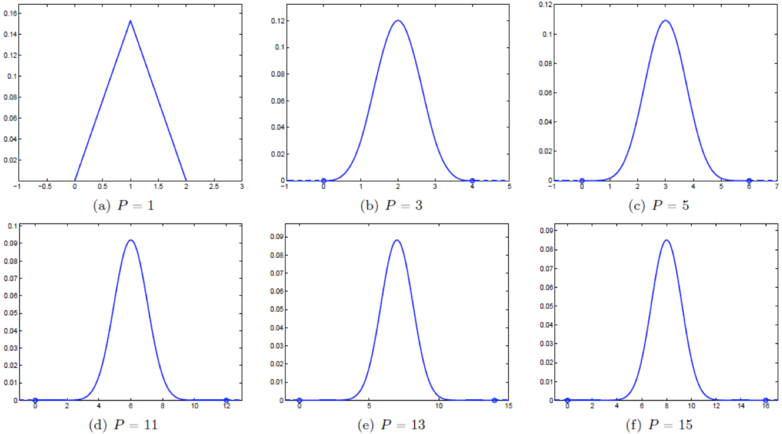
Kernels Reproducing Exponentials



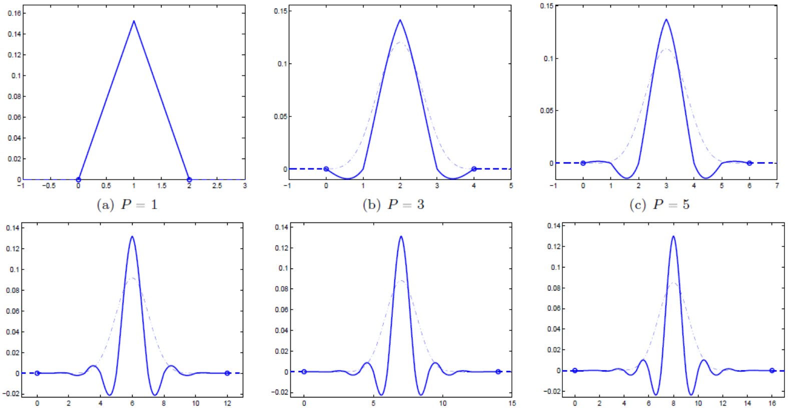
Here the E-spline is of second order and reproduces the exponential $e^{\alpha_0 t}$, $e^{\alpha_1 t}$: with $\alpha_0 = -0.06$ and $\alpha_1 = 0.5$.

Imperial College London

Examples of E-Splines Kernels



Examples of Best Kernels

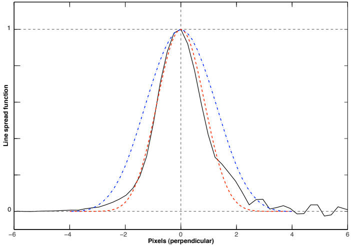


(a) $P = 1$ (b) $P = 3$ (c) $P = 5$

(d) $P = 11$ (e) $P = 13$ (f) $P = 15$

The Sampling Kernel

(a)Original (2014 × 3039)



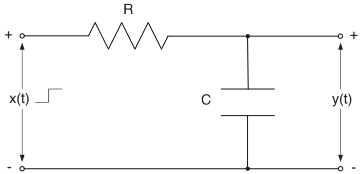
(b) Point Spread function

Kernel Reproducing Exponential

Any functions with rational Fourier transform:

$$\hat{\varphi}(\omega) = \frac{\prod_i (j\omega - b_i)}{\prod_m (j\omega - a_m)} \qquad m = 0, 1, ..., L.$$

is a *generalized* E-splines. This includes practical devices as common as an RC circuit:



Sparse Sampling: Basic Set-up

- Assume the sampling period $T = 1$.
- Consider any $x(t)$ with $t \in [0, N)$.
- Assume the sampling kernel $\varphi(t)$ is any function that can reproduce exponentials of the form

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

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

Sparse Sampling: Basic Set-up

We have that

$$
\begin{aligned}
s_m &= \sum_{n=0}^{N-1} c_{m,n} y_n \\
&= \langle x(t), \sum_{n=0}^{N-1} c_{m,n} \varphi(t-n) \rangle \\
&= \int_{-\infty}^{\infty} x(t) e^{\alpha_m t} dt, \qquad m = 0, 1, ..., L.
\end{aligned}
$$

- s_m is the bilateral Laplace transform of $x(t)$ evaluated at α_m.
- When $\alpha_m = j\omega_m$ then $s_m = \hat{x}(\omega_m)$ where $\hat{x}(\omega)$ is the Fourier transform of $x(t)$.
- When $\alpha_m = 0$, the s_m's are the polynomial moments of $x(t)$.

Sampling Streams of Diracs

- Assume $x(t)$ is a stream of K Diracs on the interval of size N: $x(t) = \sum_{k=0}^{K-1} x_k \delta(t - t_k)$, $t_k \in [0, N)$.
- We restrict $\alpha_m = \alpha_0 + m\lambda \quad m = 0, 1, ..., L$ and $L \geq 2K - 1$.
- We have N samples: $y_n = \langle x(t), \varphi(t-n) \rangle$, $n = 0, 1, ... N-1$:
- We obtain

$$
\begin{aligned}
s_m &= \sum_{n=0}^{N-1} c_{m,n} y_n \\
&= \int_{-\infty}^{\infty} x(t) e^{\alpha_m t} dt, \\
&= \sum_{k=0}^{K-1} x_k e^{\alpha_m t_k} \\
&= \sum_{k=0}^{K-1} \hat{x}_k e^{\lambda m t_k} = \sum_{k=0}^{K-1} \hat{x}_k u_k^m, \qquad m = 0, 1, ..., L.
\end{aligned}
$$

The Annihilating Filter Method

- The quantity

$$s_m = \sum_{k=0}^{K-1} \hat{x}_k u_k^m, \qquad m = 0, 1, ..., L$$

  is a sum of exponentials.
- We can retrieve the locations u_k and the amplitudes $\hat{x}_k$ with the annihilating filter method (also known as Prony's method since it was discovered by Gaspard de Prony in 1795).
- Given the pairs $\{u_k, \hat{x}_k\}$, then $t_k = (\ln u_k)/\lambda$ and $x_k = \hat{x}_k / e^{\alpha_0 t_k}$.

The Annihilating Filter Method

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

$$h_m * s_m = \sum_{i=0}^{K} h_i s_{m-i} = \sum_{i=0}^{K} \sum_{k=0}^{K-1} \hat{x}_k h_i u_k^{m-i} = \sum_{k=0}^{K-1} \hat{x}_k u_k^m \underbrace{\sum_{i=0}^{K} h_i u_k^{-i}}_{0} = 0.$$

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

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

Solve the above system to find the coefficients of the annihilating filter.

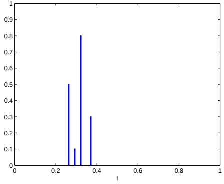
The Annihilating Filter Method

2. Given the coefficients $\{1, h_1, h_2, ..., h_k\}$, we get the locations u_k by finding the roots of $H(z)$.
3. Solve the first K equations in $s_m = \sum_{k=0}^{K-1} \hat{x}_k u_k^m$ to find the amplitudes $\hat{x}_k$.
In matrix/vector form

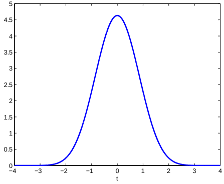
$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ u_0 & u_1 & \cdots & u_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_0^{K-1} & u_1^{K-1} & \cdots & u_{K-1}^{K-1} \end{bmatrix} \begin{pmatrix} \hat{x}_0 \\ \hat{x}_1 \\ \vdots \\ \hat{x}_{K-1} \end{pmatrix} = \begin{pmatrix} s_0 \\ s_1 \\ \vdots \\ s_{K-1} \end{pmatrix}. \qquad (3)$$

Classic Vandermonde system. Unique solution for distinct u_ks.

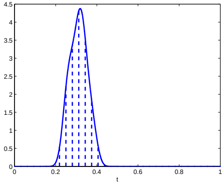
Sampling Streams of Diracs: Numerical Example



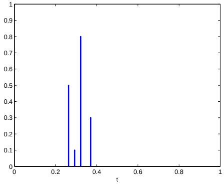
(a) Original Signal



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

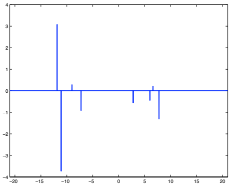


(c) Samples

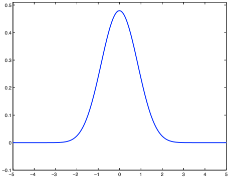


(d) Reconstructed Signal

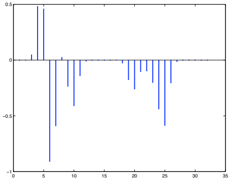
Sampling Streams of Diracs: Sequential Reconstruction



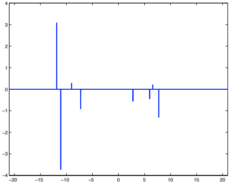
(a) Original Signal



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



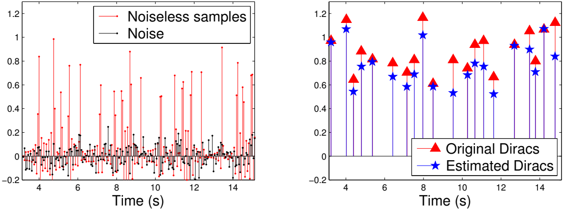
(c) Samples



(d) Reconstructed Signal

Imperial College London

Sampling Streams of Diracs: Sequential Reconstruction



(a) y_n samples **(b)** Reconstructed stream

In this example: $10K$ samples, 1000 Diracs, $SNR = 15$dB, Execution time: one minute, Success rate 100%, one false positive.

Note on the proof

Linear vs Non-linear

- Problem is **Non-linear** in t_k, but **linear** in x_k given t_k
- The key to the solution is the separability of the non-linear from the linear problem using the annihilating filter.

The proof is based on a constructive algorithm:

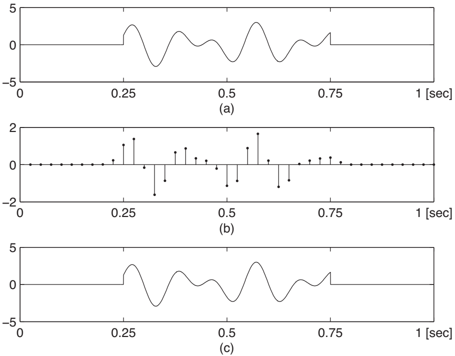
1. Given the N samples y_n, compute the moments s_m using the exponential reproduction formula. In matrix vector form $S = \mathbf{C}Y$.
2. Solve a $K \times K$ Toeplitz system to find $H(z)$
3. Find the roots of $H(z)$
4. Solve a $K \times K$ Vandermonde system to find the a_k

Complexity

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

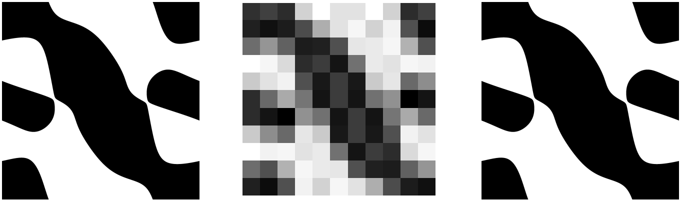
Thus, the algorithm complexity is polynomial with the signal innovation.

Sampling Piecewise Sinusoidal Signals: Numerical Example



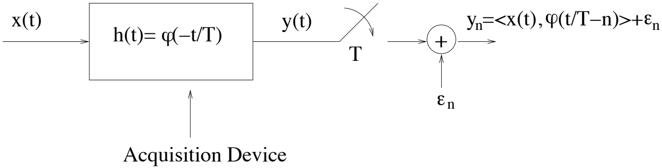
Imperial College London

Sampling 2-D domains



Imperial College London

Robust Sparse Sampling



- The measurements are noisy
- The noise is additive and i.i.d. Gaussian

Robust Sparse Sampling

In the presence of noise, 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 $\mathbf{S}$:

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

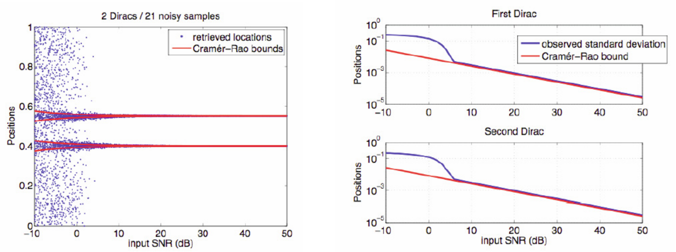
Then H is the last column of $\mathbf{V}$.
Notice: this is similar to Pisarenko's method in spectral estimation.

Robust Sparse Sampling: Cadzow's algorithm

For small SNR use Cadzow's method to denoise **S** before applying TLS.
The basic intuition behind this method is that, in the noiseless case, **S** is rank deficient (rank K) and Toeplitz, while in the noisy case **S** is full rank.
Algorithm:

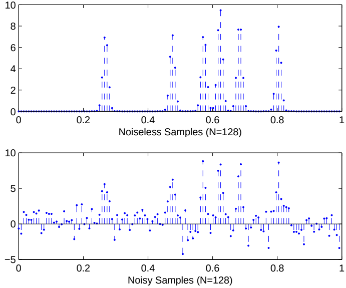
- SVD of $\mathbf{S} = \mathbf{U}\lambda\mathbf{V}^{\mathbf{T}}$.
- Keep the K largest diagonal coefficients of λ and set the others to zero.
- Reconstruct $\mathbf{S}' = \mathbf{U}\lambda'\mathbf{V}^{\mathbf{T}}$.
- This matrix is not Toeplitz, make it so by averaging along the diagonals.
- Iterate.

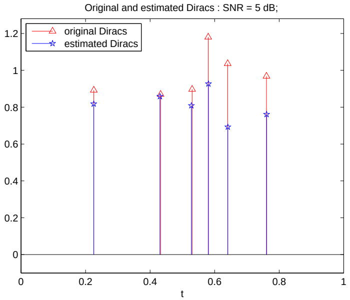
Robust Sparse Sampling



- 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.

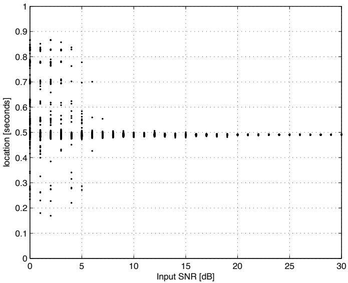
Robust Sparse Sampling

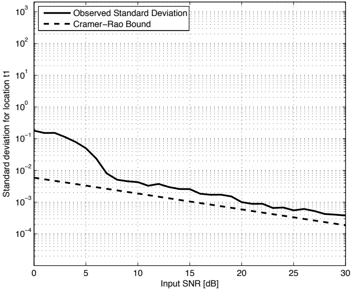




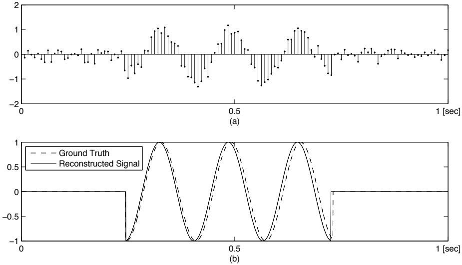
Robust Sparse Sampling

Piecewise sinusoidal signal





Robust Sparse Sampling



SNR= 8dB, N=128.

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)

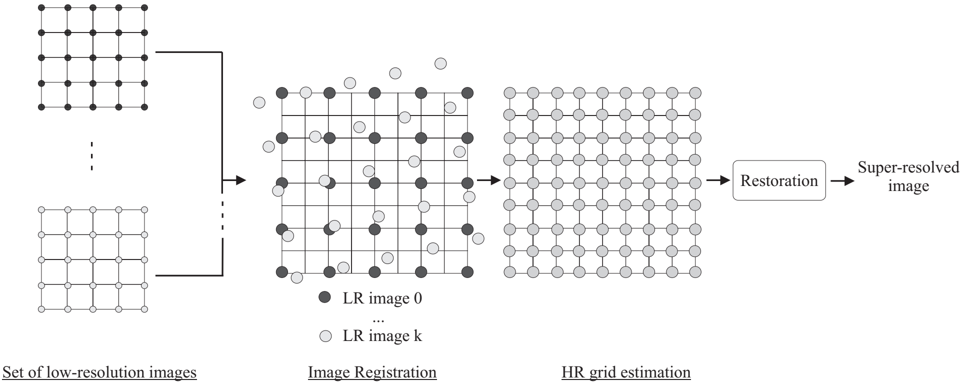
(c) Super-res (PSNR=24.2dB)

- Forty low-resolution and shifted versions of the original.
- The disparity between images has a finite rate of innovation and can be retrieved.
- Accurate registration is achieved by retrieving the continuous moments of the 'Tiger' from the samples.

Imperial College London

Application: Image Super-Resolution

Image super-resolution basic building blocks



Application: Image Super-Resolution

- For each blurred image $I(x, y)$:
  - A pixel $P_{m,n}$ 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:

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

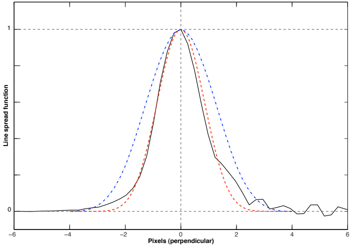
  - We retrieve the exact moments of $I(x, y)$ from $P_{m,n}$:

$$\tau_{l,j} = \sum_n \sum_m c_{m,n}^{(l,j)} P_{m,n} = \int \int I(x, y) x^l y^j dxdy.$$

- 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)

Imperial College London

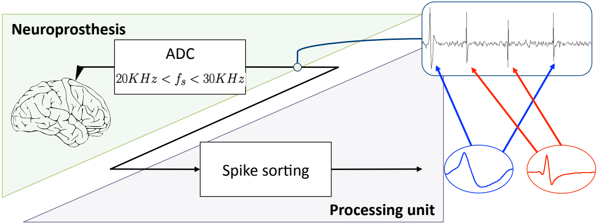
Application: Image Super-Resolution

(a)Original (48×48)

(b) Super-res (480×480)

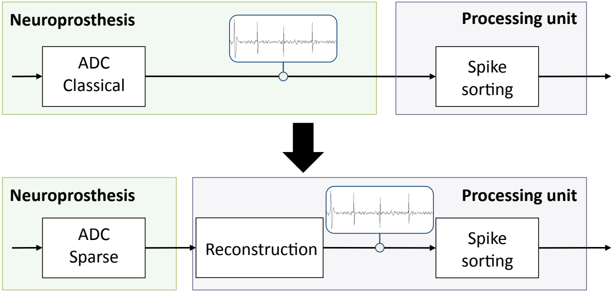
Application in Neuroscience

Applications in Neuroscience



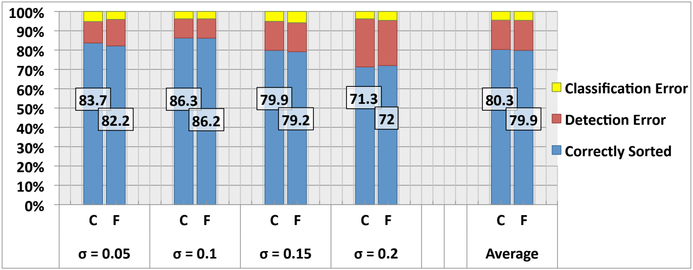
Application in Neuroscience

Insight: Sample at lower rate and reconstruct the signal outside the implant

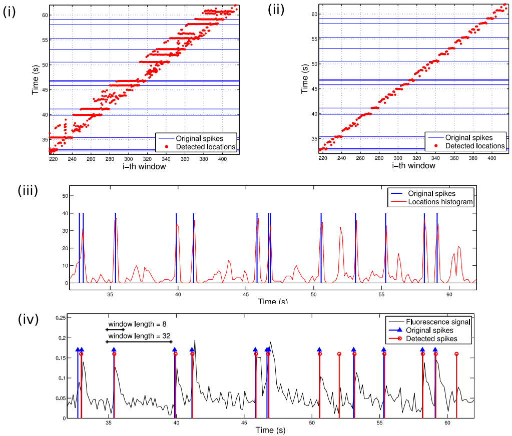


Application in Neuroscience

- Classical Sampling (C) $f_s = 24KHz$
- Sparse Sampling (F) $f_s = 5.8KHz$



Calcium Transient Detection



Imperial College London

Application in Sensor Networks

Localizing Point Sources in Diffusion Fields

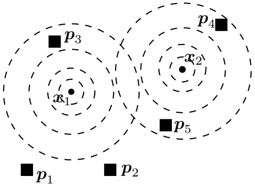
Localized and **instantaneous** sources:

$$f_t = \triangle f + \sum_{k=1}^{K} c_k \delta(\boldsymbol{x} - \boldsymbol{x}_k)\delta(t - t_k)$$

Diffusion field:

$$f(\boldsymbol{x}, t) = \sum_{k=1}^{K} \frac{c_k}{4\pi(t - t_k)} e^{-\frac{\|\boldsymbol{x}-\boldsymbol{x}_k\|^2}{4(t-t_k)}} U(t - t_k)$$

finite degrees of freedom



Goal:

Estimate the unknown parameters $\{c_k\}_k$, $\{t_k\}_k$, $\{\boldsymbol{x}_k\}_k$ from the spatiotemporal samples taken by distributed sensors.

Conclusions and Outlook

Sampling signals using sparsity models:

- New framework that allows the sampling and reconstruction of signals at a rate smaller than Nyquist rate.
- It is a non-linear problem
- Different possible algorithms with various degrees of efficiency and robustness

Applications:

- Many actual and potential applications:
- But you need to fit the right model!
- Carve the right algorithm for your problem: continuous/discrete, fast/ complex, redundant/ not-redundant

Still many open questions from theory to practice!

Imperial College London

References

On sparsity in over-complete dictionaries

- D. L. Donoho and P.B. Starck, ‘Uncertainty principles and signal recovery’, SIAM J. Appl. Math., 1989, pp. 906-931.
- D.L. Donoho and X. Huo, ‘Uncertainty principles and ideal atomic decomposition’, IEEE Trans. on Info.. Theory, vol.47(7, pp.2845-62, November 2001.
- M. Elad,‘Sparse and Redundant Representations’, Springer, 2010.

On Compressed Sensing and its applications

- E. J. Candès, J. Romberg, and T. Tao, ‘Robust Uncertainty Principle: Exact signal reconstruction from highly incomplete frequency information’, IEEE Trans. Info. Theory, vol. 52(2), pp. 489-509, February 2006.
- E.J. Candés and M.B. Wakin, ‘An introduction to compressive sampling’, IEEE Signal Processing Magazine, vol. 25(2), pp. 21-30, March 2008.
- M. Lustig, D.L. Donoho, J.M. Santos and J.M. Pauly,‘Compressed Sensing MRI’, IEEE Signal Processing Magazine, vol. 25(2), pp. 72-82, March 2008.

References

On sampling FRI Signals

- M. Vetterli, P. Marziliano and T.Blu, 'Sampling Signals with Finite Rate of Innovation', IEEE Trans. on Signal Processing, 50(6):14171428, June 2002.
- 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.
- J.Berent and P.L. Dragotti, and T. Blu,'Sampling Piecewise Sinusoidal Signals with Finite Rate of Innovation Methods,' IEEE Transactions on Signal Processing, Vol. 58(2),pp. 613-625, February 2010.
- J. Uriguen, P.L. Dragotti and T. Blu,'On the Exponential Reproducing Kernels for Sampling Signals with Finite Rate of Innovation' in Proc. of Sampling Theory and Application Conference, Singapore, May 2011.
- P.L. Dragotti, M. Vetterli and T. Blu, 'Exact Sampling Results for signals with finite rate of innovation using Strang-Fix conditions and local kernels' in Proc. of ICASSP, Philadelphia, March 2005.

References (cont'd)

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.

On Diffusion Fields

- Y. Lu, P.L. Dragotti and M. Vetterli, 'Localization of diffusive sources using spatio-temporal measurements', 49th Allerton Conference , Allerton 2011.
- D. Malioutov, M. Cetin and A. Willsky,'A sparse signal reconstruction perspective for source localization with sensor arrays', IEEE Trans, on Signal Processing, vol. 53(8) pp.3010-3022, August 2005.

On Neuroscience:

- J. Onativia, S. Schultz and P.L. Dragotti, 'A Finite Rate of Innovation algorithm for fast and accurate spike detection from two-photon calcium imaging', submitted to Journal of Neural Engineering, Nov. 2012.
- J. Caballero, J.A. Uriguen, S. Schultz and P.L. Dragotti, 'Spike Sorting at Sub-Nyquist Rates', in Proc. of IEEE (ICASSP), Kyoto, Japan, April 2012.

Appendix

Orthogonal matching pursuit (OMP) finds the correct sparse representation when

$$K < \frac{1}{2}\left(1 + \frac{1}{\mu}\right). \qquad (4)$$

Sketch of the Proof (Elad 2010, pages 65-67):

Assume the K non-zero entries are at the beginning of the vector in descending order with $y = Dx$. Thus

$$y = \sum_{l=1}^{K} x_l D_l \qquad (5)$$

First iteration of OMP work properly if $|D_1^T y| > |D_i^T y|$ for any $i > K$. Using (5)

$$|\sum_{l=1}^{K} x_l D_1^T D_l| > |\sum_{l=1}^{K} x_l D_i^T D_l|$$

Appendix (cont'd)

Sketch of the Proof (cont'd):

But

$$|\sum_{l=1}^{K} x_l D_1^T D_l| \geq |x_1| - \sum_{l=2}^{K} |x_l||D_1^T D_l| \geq |x_1| - \sum_{l=2}^{K} |x_l|\mu \geq |x_1|(1-\mu)(K-1).$$

Moreover,

$$|\sum_{l=1}^{K} x_l D_i^T D_l| \leq \sum_{l=1}^{K} |x_l||D_i^T D_l| \leq \sum_{l=1}^{K} |x_l|\mu \leq |x_1|\mu K$$

Using these two bounds, we conclude that $|D_1^T y| > |D_i^T y|$ is satisfied when condition (4) is met. □