Imperial College London

Sparse Signal Processing
Part 2: Sparse Sampling

Pier Luigi Dragotti[1]

September 8, 2015

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

Imperial College London

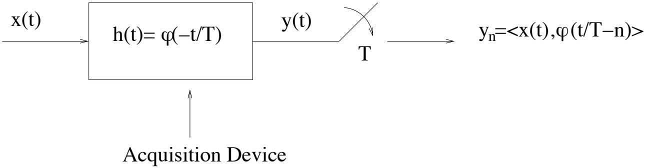
Outline

- Problem Statement and Motivation
- Classical Sampling Formulation
- Sampling using expansion-based sparsity
  - Compressed Sensing
  - Applications
- Sampling using parametric-based sparsity
  - Signals with Finite Rate of Innovation (FRI)
  - Sampling Kernels and Reproduction of Exponentials
  - Sampling Theorems for Continuous Sparse Signals
- Applications
- Conclusions and Outlook

Imperial College London

Problem Statement

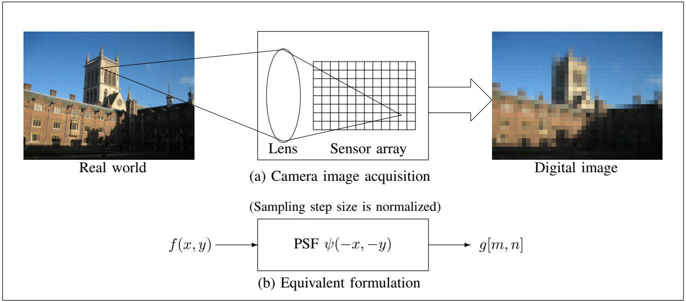
You are given a class of functions. You have a sampling device. Given the measurements $y_n = \langle x(t), \varphi(t/T - n) \rangle$, you want to reconstruct $x(t)$.



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 lens blurs the image.
- The image is sampled ('pixelized') by the sensor array.
- You want sharper and higher resolution images given the available pixels

Motivation: Image Resolution Enhancement

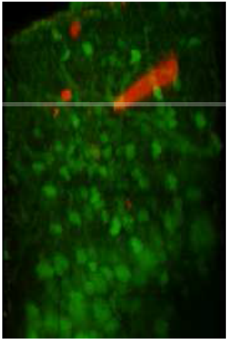
pixels

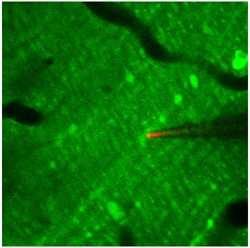
interpolation

enhancement with sparsity priors

Motivation: Application in Neuroscience

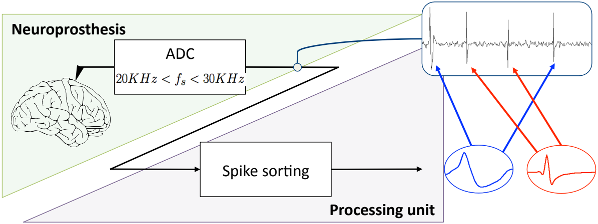
Time resolution enhancement and calcium transient detection in multi-photon calcium imaging.





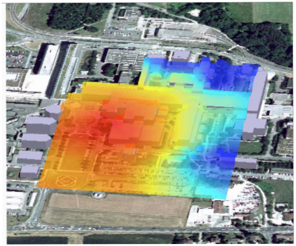
Motivation: Brain Machine Interface

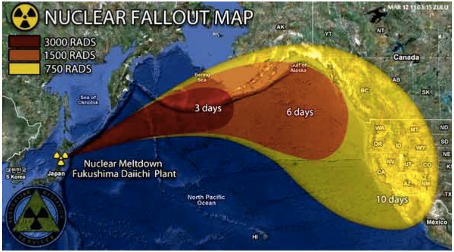
Applications in Neuroscience: Spike Sorting at sub-Nyquist rates



Wireless brain-machine interface place extreme limits on sampling.

Motivation: Sensor Networks

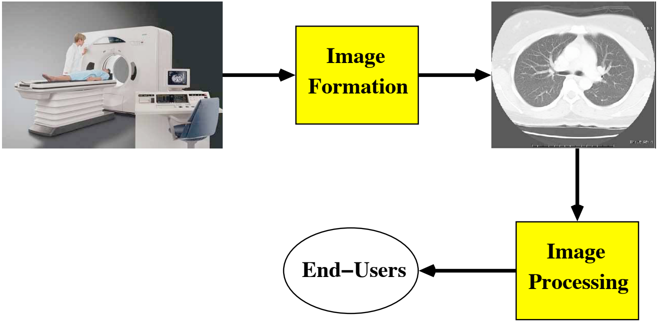




- Can we localise diffusion sources and estimate their activation time using sensor networks?
- Application:
  1. Check whether your government is lying ;-)
  2. Monitor dispersion in factories producing bio-chemicals

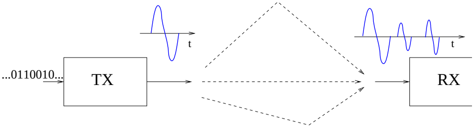
Motivation: MRI

"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.



Pulse Based Communication

Wide-Band Communications:

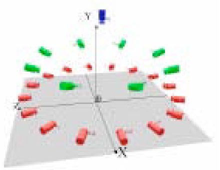


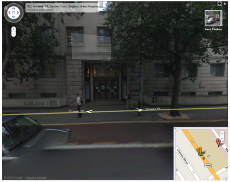
- Current A-to-D converters in UWB communications operate at several gigaherz.
- This is a sparse parametric estimation problem, only the location and amplitude of the pulses need to be estimated.

Imperial College London

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.

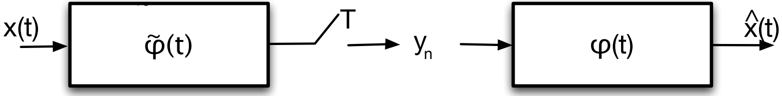


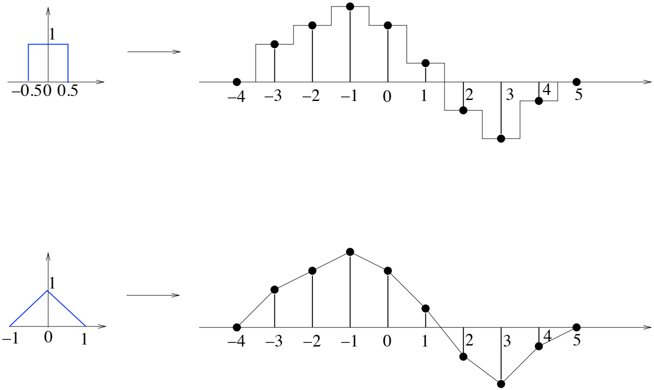
- This is a multi-dimensional sampling and interpolation problem.

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.

Compressed Sensing Case: Notation

Recall that:

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

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

- In the sparse representation case we were assuming that $\mathbf{y}$ was sparse in a redundant dictionary $\mathbf{D}$ and we were solving the following problem:

$$\min_{\alpha} \|\mathbf{y} - \mathbf{D}\alpha\|_2 + \lambda \|\alpha\|_1$$

Sparsity in Redundant Dictionaries

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

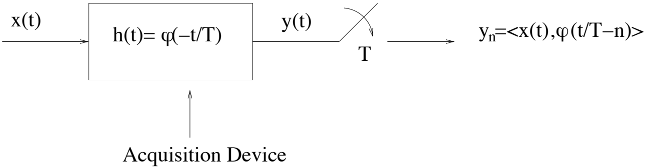
- For a generic over-complete dictionary D, (P_1) is equivalent to (P_0) when

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

So $K < \frac{1}{2}\sqrt{N}$. This is pretty bad...

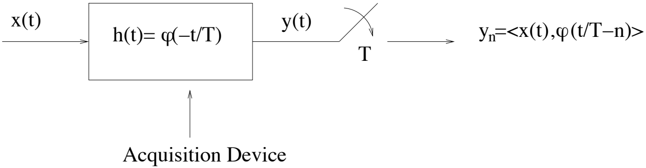
Imperial College London

Compressed Sensing Formulation



- In compressed sensing you discretize the sampling problem and assume $\mathbf{x}$ is a long vector of size M.
- For the time being call it α and assume it is K-sparse.
- The acquisition process stays linear and is modelled with a fat matrix leading to the samples $\mathbf{y}$. (short vector of size N)

Compressed Sensing Formulation



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

Compressed Sensing Formulation

Key Insights

- Since Φ is the ‘acquisition device’, you can choose the Φ you like
- Relax the condition of a ‘deterministic’ perfect reconstruction and accept that, with an extremely small probability, there might be an error in the reconstruction.
- From deterministic bounds to average case bounds

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 $\mathbf{y}$ when $N \geq C \cdot K \log(M/K)$ for some constant C.
- Assume that the measured signal $\mathbf{x}$ is not sparse but has a sparse representation: $\mathbf{x} = D\alpha$. We have that $\mathbf{y} = \Phi\mathbf{x} = \Phi D\alpha$. 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.

Compressed Sensing. Simulation Results

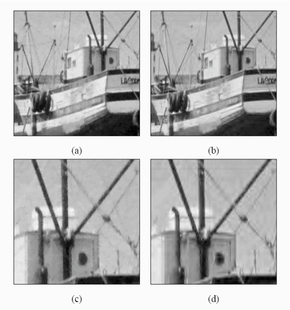


Image 'Boat'. (a) Recovered from 20000 random projections using Compressed Sensing. PSNR=31.8dB. (b) Optimal 7207-approximation using the wavelet transform with the same PSNR as (a). (c) Zoom of (a). (d) Zoom of (b). Images courtesy of Prof. J. Romberg.

Application in MRI

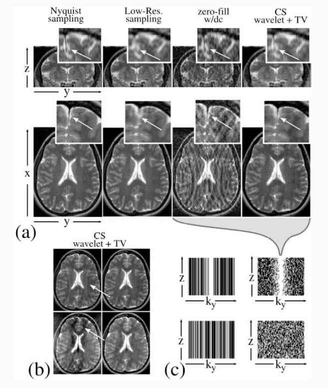


Image taken from Lustig, Donoho, Santos, Pauly-08.

Toward Sampling Continuous Sparse Signals

- In compressed sensing, we discretise a problem which is inherently '*analogue*'
- Once the size M of $\mathbf{x}$ is decided, this dictates resolution and complexity
- Complexity should be related to the sparsity of the problem (at least in the ideal case), not to M

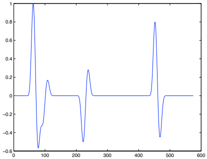
Key ingredients to overcome the above limitations

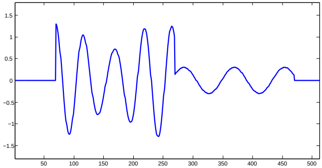
- Introduce '*analogue*' sparsity: sparsity for continuous-time signals
- Use wavelet theory and shift-invariant subspaces for hybrid analogue/digital processing
- Replace Basis Pursuit with Prony-like methods which can handle continuous-time problems

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

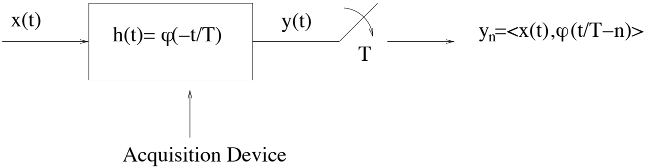
$$\rho = \lim_{\tau \to \infty} \frac{1}{\tau} C_x \left(-\frac{\tau}{2}, \frac{\tau}{2} \right), \tag{2}$$

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).

Imperial College London

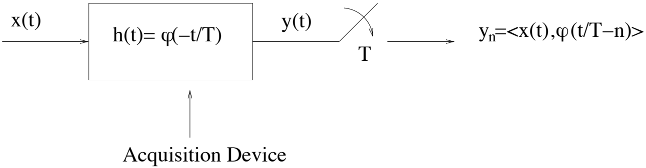
The Sampling Kernel



- We now have a good model for sparse continuous-time signals
- The samples however are discrete
- We need to map the discrete samples to some information of the continuous-time signal (e.g. Fourier transform)
- **Key Intuition:** Use the knowledge of the acquisition process to map the discrete samples to some information about $x(t)$

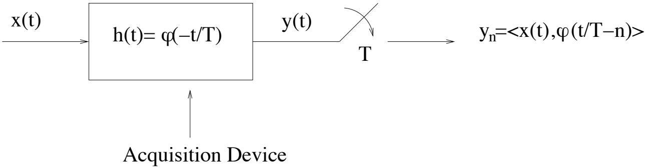
Imperial College London

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



It is reasonable to assume that the acquisition process is approximately linear and invariant. Therefore, the samples can be written as follows:

$$y_n = \langle x(t), \varphi(t/T - n) \rangle.$$

Compute a linear combination of the samples: $s_m = \sum_n c_{m,n} y_n$ for some choice of coefficients $c_{m,n}$

From Samples to Signals

Because of linearity of inner product, we have that

$$s_m = \sum_n c_{m,n} y_n$$
$$= \langle x(t), \sum_{n=0}^{N-1} c_{m,n}\varphi(t/T - n)\rangle \qquad m = 0, 1, ..., L.$$

Assume that $\sum_n c_{m,n}\varphi(t/T - n) \simeq e^{j\omega_m t/T}$ for some frequencies ω_m $m = 0, 1, ..., L$

From Samples to Signals

Then

$$
\begin{aligned}
s_m &= \sum_n c_{m,n} y_n \\
&= \langle x(t), \sum_n c_{m,n} \varphi(t/T - n) \rangle \\
&\simeq \int_{-\infty}^{\infty} x(t) e^{j\omega_m t} dt, \qquad m = 0, 1, ..., L.
\end{aligned}
$$

Note that s_m is the **Fourier transform** of $x(t)$ evaluated at $j\omega_m$.

Approximation of Exponentials

We want to find coefficients $c_{m,n}$ that give us a good approximation of the exponentials:

$$\sum_n c_{m,n}\varphi(t/T-n) \simeq e^{j\omega_m t/T}$$

- **Key Insight**: leverage from the theory of approximation in shift-invariant sub-spaces to find $c_{m,n}$ and to pick the best $\varphi(t)$.
- **Remark** we now use that theory for analysis and not for synthesis.

Approximation of Exponentials

For best approximation, we need to compute (orthogonal projection):

$$c_{m,n} = \langle e^{j\omega_m t/T}, \tilde{\varphi}(t/T - n)\rangle.$$

Since the kernel is shift-invariant, we have close-form expressions for the coefficients and the error.

- *Coefficients*

$$c_{m,n} = \frac{\hat{\varphi}(-j\omega_m)}{\hat{a}_\varphi(\mathrm{e}^{j\omega_m})}\mathrm{e}^{j\omega n},$$

  where $\hat{a}_\varphi(\mathrm{e}^{j\omega_m}) = \sum_{l\in\mathbb{Z}} a_\varphi[l]\mathrm{e}^{-j\omega_m l}$ with $a_\varphi[l] = \langle\varphi(t-l), \varphi(t)\rangle$.

- *Approximation error*

$$\varepsilon(t) = f(t) - e^{j\omega_m t} = \mathrm{e}^{j\omega_m t}\left[1 - c_0 \sum_{l\in\mathbb{Z}} \hat{\varphi}(j\omega_m + j2\pi l)\mathrm{e}^{j2\pi l t}\right].$$

Generalised Strang-Fix Conditions

A function $\varphi(t)$ can reproduce the exponential:

$$e^{j\omega_m t} = \sum_n c_{m,n}\varphi(t-n)$$

if and only if

$$\hat{\varphi}(j\omega_m) \neq 0 \text{ and } \hat{\varphi}(j\omega_m + j2\pi l) = 0 \qquad l \in \mathbb{Z} \setminus \{0\}$$

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

Also note that $c_{m,n} = c_{m,0}e^{j\omega_m n}$ with $c_{m,0} = \hat{\varphi}(j\omega_m)^{-1}$. (from now on we use this expression also for the approximate case).

Imperial College London

Approximate Strang-Fix

- Strang-Fix conditions are not restrictive
- Any low-pass or band-pass filter approximately satisfies them.

Approximate Strang-Fix

- Assume $\varphi(t)$ cannot reproduce exponentials, however, we still use the coefficients $c_n = \frac{1}{\hat{\varphi}(j\omega_m)} e^{j\omega_m n}$ such that:

$$\sum_{n\in\mathbb{Z}} c_n \varphi(t-n) \approxeq \mathrm{e}^{j\omega_m t}.$$

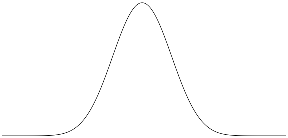
- Approximation error

$$\varepsilon(t) = f(t) - e^{j\omega_m t} = \mathrm{e}^{j\omega_m t}\left[1 - \frac{1}{\hat{\varphi}(j\omega_m)} \sum_{l\in\mathbb{Z}} \hat{\varphi}(j\omega_m + j2\pi l)\mathrm{e}^{j2\pi lt}\right].$$

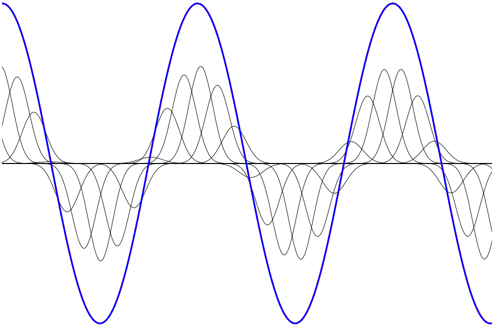
- We only need $\hat{\varphi}(j\omega_m + j2\pi l) \approxeq 0 \quad l \in \mathbb{Z} \setminus \{0\}$, which is satisfied when $\varphi(t)$ has an essential bandwidth of size 2π.

Reproduction of Exponentials (exact)

$$\sum_{n\in\mathbb{Z}} c_{m,n}\varphi(t-n) = e^{-j\omega_m t} \qquad \forall m \in \{1, 2, \ldots M\}$$

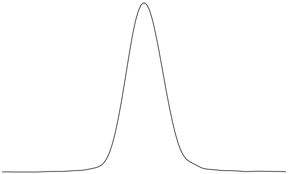


$\phi(t)$ is an E-spline

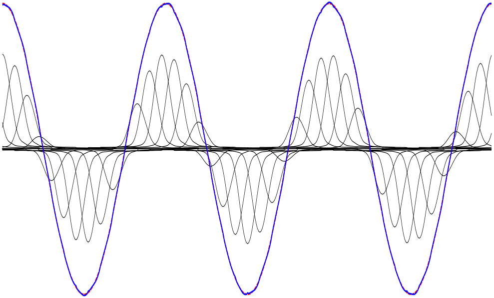


Approximate Strang-Fix

$$\sum_{n \in \mathbb{Z}} c_{m,n}\varphi(t-n) \simeq e^{-j\omega_m t} \qquad \forall m \in \{1,2,...M\}$$



$\phi(t)$ from a real camera



From Samples to Signals

$$
\begin{aligned}
s_m &= \sum_n c_{m,n} y_n \\
&= \langle x(t), \sum_n c_{m,n} \varphi(t/T - n) \rangle \\
&\simeq \int_{-\infty}^{\infty} x(t) e^{j\omega_m t} dt, \qquad m = 0, 1, ..., L.
\end{aligned}
$$

Note that s_m is the Fourier transform of $x(t)$ evaluated at $j\omega_m$.

From Samples to Signals

- We now have partial knowledge of $\hat{x}(j\omega)$:

$$y_n \Rightarrow \hat{x}(j\omega_m) \qquad m = 1, 2, ..., L$$

- Given $\hat{x}(j\omega_m)$, use your favourite sparsity model and reconstruction method to obtain a one-to-one mapping between the signal and its partial Fourier transform:

$$x(t) \Leftrightarrow \hat{x}(j\omega_m) \qquad m = 1, 2, ..., L$$

- For classes of parametrically sparse signals there is a one-to-one mapping between samples and signal:

$$x(t) \Leftrightarrow \hat{x}(j\omega_m) \qquad m = 1, 2, ..., L$$

- The number d of degrees of freedom of the signal must satisfy $d \leq L$

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 $j\omega_m = j\omega_0 + jm\lambda \quad m = 1, ..., L$ and $L \geq 2K$.
- We have N samples: $y_n = \langle x(t), \varphi(t-n) \rangle,\ n = 0, 1, ...N-1$:
- We obtain

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

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 $x_k = \hat{x}_k / e^{\alpha_0 t_k}$.

Imperial College London

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

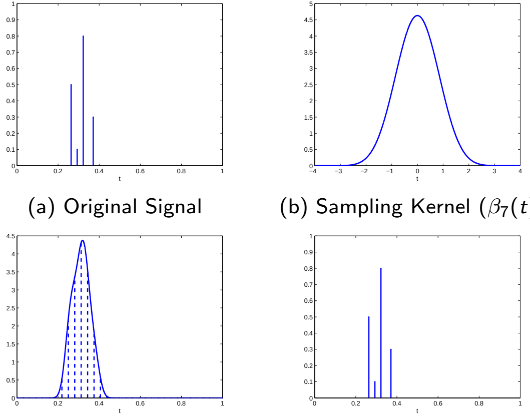
2. Given the coefficients $\{1, h_1, h_2, ..., h_k\}$, we get the locations u_k by finding the roots of $H(z)$.

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

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

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

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:

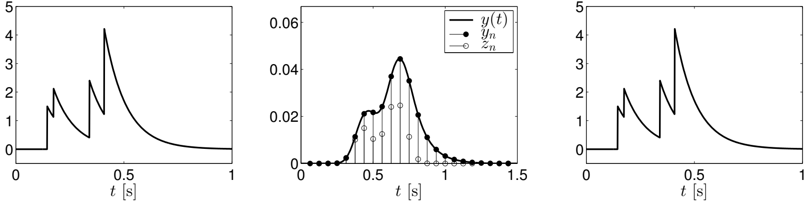
1. Given the N samples y_n, compute the new quantities s_m using the exponential reproduction formula. In matrix vector form $\mathbf{s} = \mathbf{C}\mathbf{y}$.
2. Solve a $K \times K$ Toeplitz system to find $H(z)$
3. Find the roots of $H(z)$
4. Solve a $K \times K$ Vandermonde system to find the a_k

Complexity

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

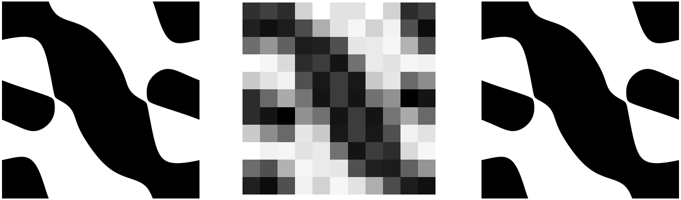
Thus, the algorithm complexity is polynomial with the signal innovation.

Stream of Decaying Exponentials



(a) Input signal, $x(t)$ **(b)** Filtered and sampled signal **(c)** Reconstructed signal

Sampling 2-D domains



The curve is implicitly defined through the equation [PanBluDragotti:11,14]:

$$f(x,y) = \sum_{k=1}^{K} \sum_{i=1}^{I} b_{k,i} e^{-j2\pi xk/M} e^{-j2\pi yi/N} = 0.$$

The coefficients $b_{k,i}$ are the only free parameters in the model.
This is a **non-separable** 2-D sparsity model.

Sampling 2-D domains

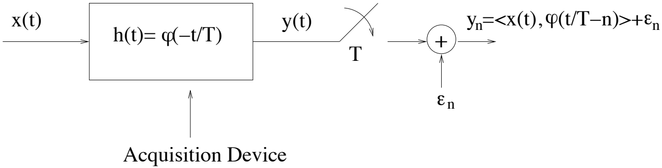
samples

interpolation

inter+ curve constraint

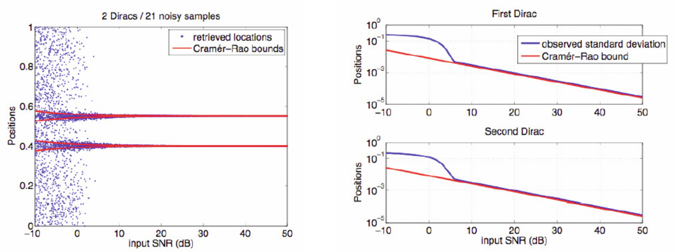
Imperial College London

Robust and Universal Sparse Sampling



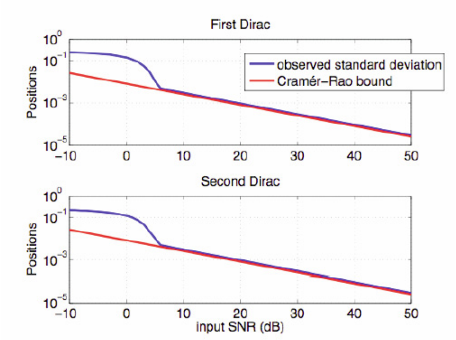
- The acquisition device is arbitrary
- The measurements are noisy
- The noise is additive and i.i.d. Gaussian
- Many robust versions of Prony's method exist (e.g., Cadzow, matrix pencil)

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

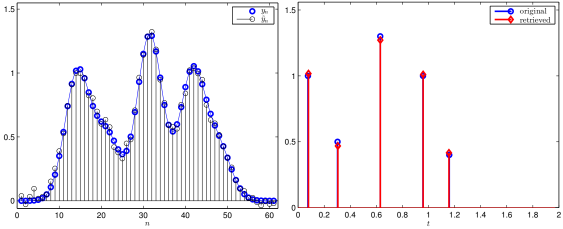


- Phase-transition
- The ‘cut-off’ SNR can be predicted precisely [Wei-Dragotti-15]

Imperial College London

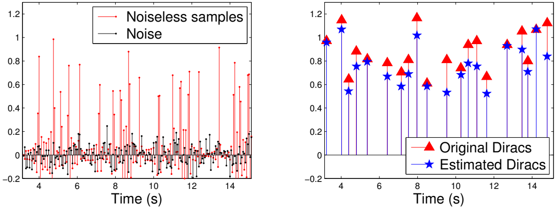
Approximate FRI recovery: Numerical Example

Gaussian Kernel



Approximate FRI with the Gaussian kernel. $K = 5$, $N = 61$, SNR=25dB. Recovery using the approximate method with $\alpha_m = j\frac{\pi}{3.5(P+1)}(2m - P)$, $m = 0, \ldots, P$ where $P + 1 = 21$.

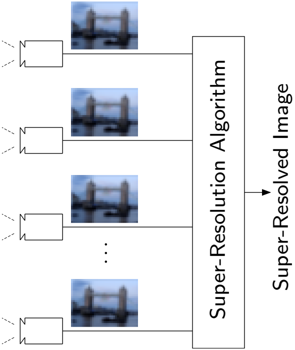
Retrieving 1000 Diracs with Strang-Fix Kernels

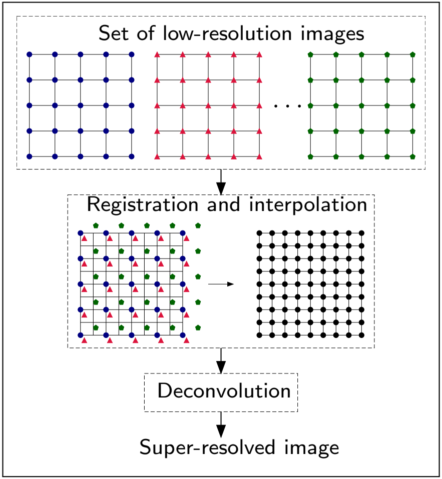


(a) y_n samples **(b)** Reconstructed stream

- $K = 1000$ Diracs in an interval of 630 seconds, $N = 10^5$ samples, $T = 0.06$ and $SNR = 10$dB
- 9997 Diracs retrieved with an error $\epsilon < T/2$
- Average accuracy $\Delta t = 0.005$, execution time 105 seconds.

Overview of Super-Resolution





Registration from Fourier information

Translation in space is a phase shift in frequency:

$$f_2(x,y) = f_1(x - s_x, y - s_y) \quad \Leftrightarrow \quad F_2(\omega_x,\omega_y) = e^{-j(\omega_x s_x + \omega_y s_y)} F_1(\omega_x,\omega_y).$$

Translation parameters can be found from the NCPS:

$$e^{j(\omega_x s_x + \omega_y s_y)} = \frac{F_1(\omega_x,\omega_y)F_2^*(\omega_x,\omega_y)}{|F_1(\omega_x,\omega_y)F_2^*(\omega_x,\omega_y)|}.$$

Construct an over-complete set of equations:

$$\omega_{m_x} s_x + \omega_{m_y} s_y = \arg\left(\frac{F_1(\omega_{m_x},\omega_{m_y})F_2^*(\omega_{m_x},\omega_{m_y})}{\left|F_1(\omega_{m_x},\omega_{m_y})F_2^*(\omega_{m_x},\omega_{m_y})\right|}\right),$$

$$\forall(\omega_{m_x},\omega_{m_y}) \text{ s.t. } \frac{1}{\left|\Phi(\omega_{m_x},\omega_{m_y})\right|} \sum_{l\in\mathbb{Z}\setminus\{0\}} \sum_{k\in\mathbb{Z}\setminus\{0\}} \left|\Phi(\omega_{m_x}+2\pi l, \omega_{m_y}+2\pi k)\right| \leq \gamma.$$

Results: Image registration

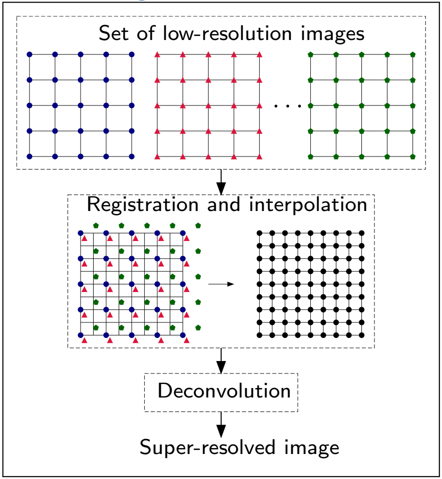
LR image from a particular viewpoint.

LR image from a different viewpoint.

100 shifts registered: RMSE is 0.012 pixels (DFT unable to distinguish the shift).

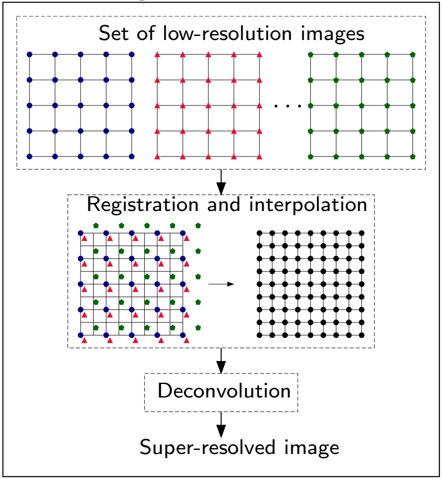
Sampling kernel - Canon EOS 40D.

Image super-resolution: Post registration



Set of LR images

Image super-resolution: Post registration



Interpolated HR image

Results: Image super-resolution

One of 100 LR images (40×40).

Interpolated image (400×400).

Deconvolution achieved using a sparse quad-tree based decomposition model [ScholefieldD:14]

Results: Image super-resolution

One of 100 LR images (40×40).

SR image (400×400).

Deconvolution achieved using a sparse quad-tree based decomposition model [ScholefieldD:14].

Application: Image Super-Resolution

Acquisition with Nikon D70

(a)Original (2014 × 3040)

(b) ROI (128 × 128)

(b) Super-res (1024 × 1024)

For more details [Baboulaz:D:09, ScholefieldD:14]

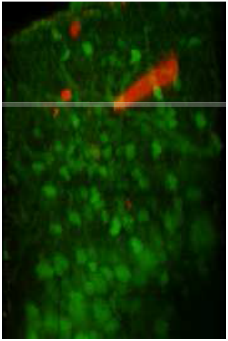
Application: Image Super-Resolution

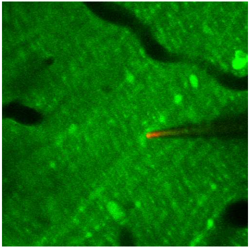
(a)Original (48 × 48)

(b) Super-res (480 × 480)

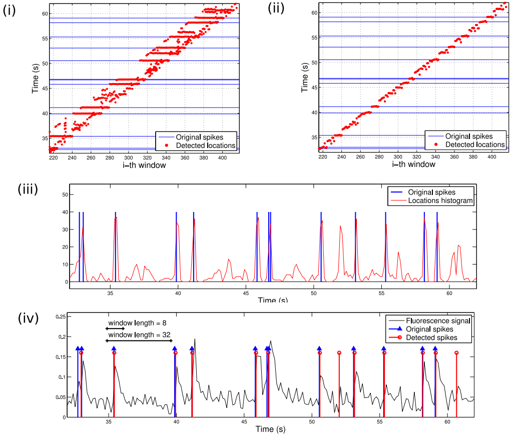
For more details [Baboulaz:D:09, ScholefieldD:14]

Neural Activity Detection [OnativiaSD:13]

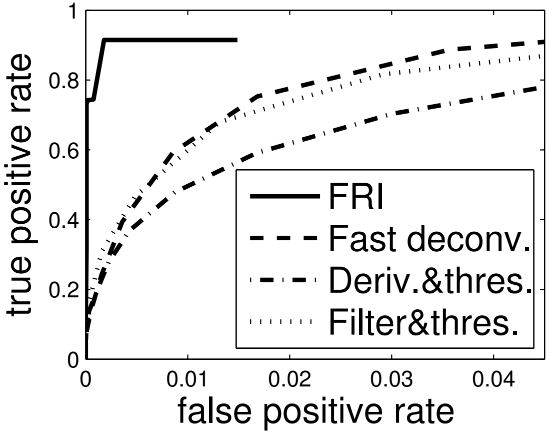




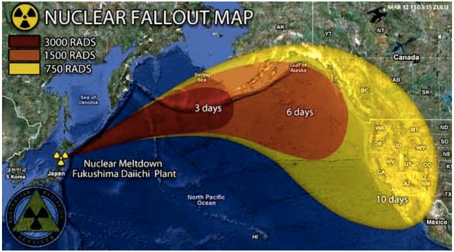
Calcium Transient Detection



Calcium Transient Detection



Localisation of Diffusion Sources using Sensor Networks [Murray-BruceD:14]



- The diffusion equation models the dispersion of chemical plumes, smoke from forest fires, radioactive materials
- The phenomenon is sampled in space and time using a sensor network.
- Sources often localised in space. Can we retrieve their location and the time of activation?

Localisation of Diffusion Sources using Sensor Networks

- The diffusion equation is

$$\frac{\partial}{\partial t}u(\mathbf{x},t) = \mu\nabla^2 u(\mathbf{x},t) + f(\mathbf{x},t),$$

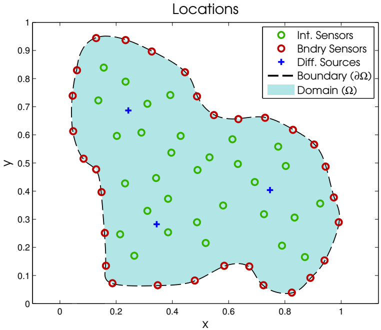
where $f(\mathbf{x},t)$ is the source.

- When sources are localised in space and time:

$$f(\mathbf{x},t) = \sum_{m=1}^{M} c_m \delta(\mathbf{x} - \xi_{\mathbf{m}}, \mathbf{t} - \tau_{\mathbf{m}}),$$

this field inversion problem is *sparse*.

- **Goal:** Estimate $\{c_m\}_m, \{\xi_m\}_m, \{\tau_m\}_m$ from the spatio-temporal sensor measurements.



Localisation of Diffusion Sources using Sensor Networks

Assume we have access to the following generalised measurements:

$$\mathcal{Q}(k,r)=\langle\Psi_k(\mathbf{x})\Gamma_r(t),f\rangle=\int_\Omega\int_t\Psi_k(\mathbf{x})\Gamma_r(t)f(\mathbf{x},t)\mathrm{d}t\mathrm{d}V,$$

with $\Psi_k=e^{-k(x+jy)}$, $k=0,1,,2M-1$ and $\Gamma_r(t)=e^{jrt/T}$, $r=0,1$. Since

$$f(\mathbf{x},t)=\sum_{m=1}^{M}c_m\delta(\mathbf{x}-\xi_{\mathbf{m}},\mathbf{t}-\tau_{\mathbf{m}}),$$

we obtain:

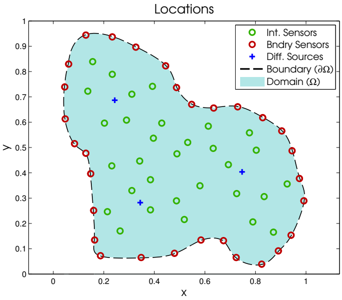
$$\mathcal{Q}(k,r)=\sum_{m=1}^{M}c_m e^{-k(\xi_{1,m}+j\xi_{2,m})}e^{-jrt_m}.$$

This quantity is a sum of exponentials and parameters $\{c_m\}_m,\{\xi_m\}_m,\{\tau_m\}_m$ can be recovered from it using Prony's method provided $k=0,1,2M-1$.

Localisation of Diffusion Sources using Sensor Networks

Assume $r = 0$, since Ψ_k is analytic, using Green's theorem, we obtain:

$$\int_t \left(\int_\Omega \frac{\partial}{\partial t}(u\Psi_k)\mathrm{d}V - \mu \oint_{\partial\Omega} (\Psi_k \nabla u - u \nabla \Psi_k) \cdot \hat{\mathbf{n}}_{\partial\Omega} \mathrm{d}S \right) dt = \int_t \int_\Omega \Psi_k f \mathrm{d}V \mathrm{d}t = \mathcal{Q}(k, 0).$$



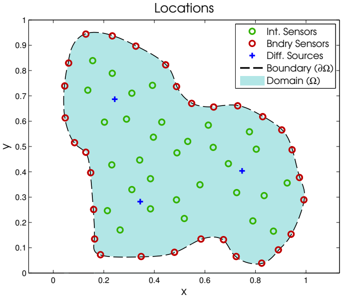
Imperial College London

Localisation of Diffusion Sources using Sensor Networks

Assume $r = 0$, since Ψ_k is analytic, using Green's theorem, we obtain:

$$\int_t \left(\int_\Omega \frac{\partial}{\partial t}(u\Psi_k)\mathrm{d}V - \mu \oint_{\partial\Omega} (\Psi_k \nabla u - u \nabla \Psi_k) \cdot \hat{\mathbf{n}}_{\partial\Omega} \mathrm{d}S \right) dt = \int_t \int_\Omega \Psi_k f \mathrm{d}V \mathrm{d}t = \mathcal{Q}(k, 0).$$

- The above equation provides a relationship between the generalised measurements and the induced field

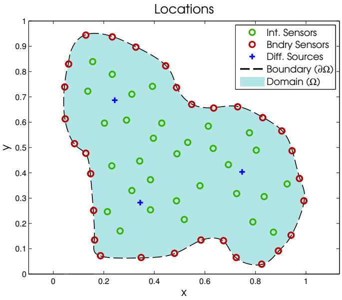


Localisation of Diffusion Sources using Sensor Networks

Assume $r = 0$, since Ψ_k is analytic, using Green's theorem, we obtain:

$$\int_t \left(\int_\Omega \frac{\partial}{\partial t}(u\Psi_k)\mathrm{d}V - \mu \oint_{\partial\Omega} (\Psi_k \nabla u - u \nabla \Psi_k) \cdot \hat{\mathbf{n}}_{\partial\Omega} \mathrm{d}S \right) dt = \int_t \int_\Omega \Psi_k f \mathrm{d}V \mathrm{d}t = \mathcal{Q}(k, 0).$$

- The above equation provides a relationship between the generalised measurements and the induced field
- We have only discrete spatio-temporal sensor measurements

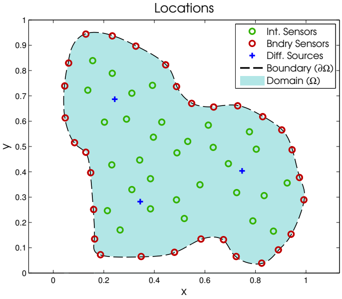


Localisation of Diffusion Sources using Sensor Networks

Assume $r = 0$, since Ψ_k is analytic, using Green's theorem, we obtain:

$$\int_t \left(\int_\Omega \frac{\partial}{\partial t}(u\Psi_k)\mathrm{d}V - \mu \oint_{\partial\Omega} (\Psi_k \nabla u - u \nabla \Psi_k) \cdot \hat{\mathbf{n}}_{\partial\Omega} \mathrm{d}S \right) dt = \int_t \int_\Omega \Psi_k f \mathrm{d}V \mathrm{d}t = \mathcal{Q}(k, 0).$$

- The above equation provides a relationship between the generalised measurements and the induced field
- We have only discrete spatio-temporal sensor measurements
- We build a mesh to approximate the full field integrals

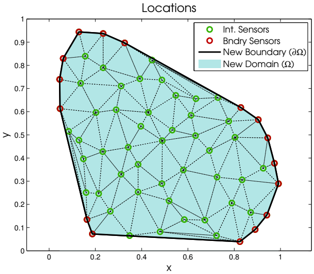


Localisation of Diffusion Sources using Sensor Networks

Assume $r = 0$, since Ψ_k is analytic, using Green's theorem, we obtain:

$$\int_t \left(\int_\Omega \frac{\partial}{\partial t}(u\Psi_k)\mathrm{d}V - \mu \oint_{\partial\Omega} (\Psi_k \nabla u - u \nabla \Psi_k) \cdot \hat{\mathbf{n}}_{\partial\Omega} \mathrm{d}S \right) dt = \int_t \int_\Omega \Psi_k f \mathrm{d}V \mathrm{d}t = \mathcal{Q}(k, 0).$$

- The above equation provides a relationship between the generalised measurements and the induced field
- We have only discrete spatio-temporal sensor measurements
- We build a mesh to approximate the full field integrals



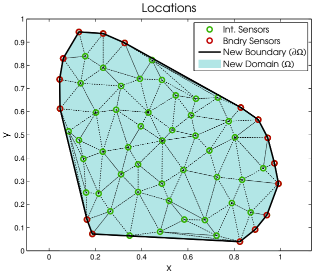
Imperial College London

Localisation of Diffusion Sources using Sensor Networks

Assume $r = 0$, since Ψ_k is analytic, using Green's theorem, we obtain:

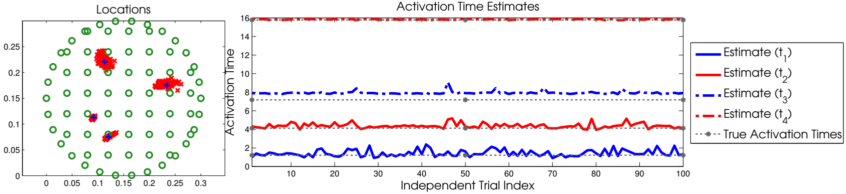
$$\int_t \left(\int_\Omega \frac{\partial}{\partial t}(u\Psi_k)\mathrm{d}V - \mu \oint_{\partial\Omega} (\Psi_k \nabla u - u \nabla \Psi_k) \cdot \hat{\mathbf{n}}_{\partial\Omega} \mathrm{d}S \right) dt = \int_t \int_\Omega \Psi_k f \mathrm{d}V \mathrm{d}t = \mathcal{Q}(k, 0).$$

- The above equation provides a relationship between the generalised measurements and the induced field
- We have only discrete spatio-temporal sensor measurements
- We build a mesh to approximate the full field integrals
- This is different from FEM because we use different priors



Imperial College London

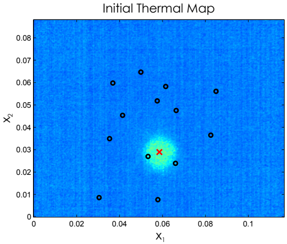
Localisation of Diffusion Sources: Numerical Results

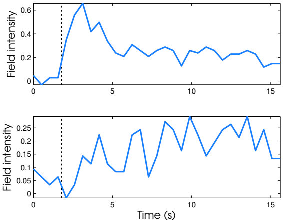


(b) 100 independent trials using noisy sensor measurement samples (SNR=15dB).

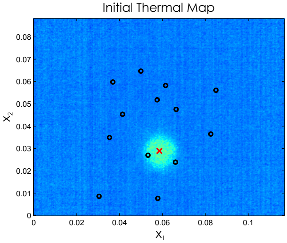
Imperial College London

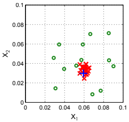
Localisation of Diffusion Sources: Real Data

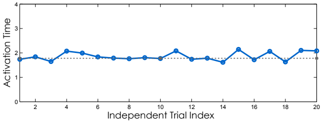




Localisation of Diffusion Sources: Real Data







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 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.A. Uriguen, T. Blu and P.L. Dragotti, 'FRI Sampling with Arbitrary Kernels', IEEE Trans. on Signal Processing, Vol 61(21), pp.5310-5323, November 2013.
- X. Wei and P.L. Dragotti, Guaranteed Performance in the FRI setting, IEEE Signal Processing Letters, 2015

Imperial College London

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

- J. Murray-Bruce and P.L. Dragotti, Estimating localized sources of diffusion fields using spatiotemporal sensor measurements, IEEE Transactions on Signal Processing, vol. 63(12), pp. 3018-3031, June 2015.

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.

Imperial College London

Appendix

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

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

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

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

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. □