Sparse Signal Processing
Part 2: Sparse Sampling

Pier Luigi Dragotti

September 8, 2015

1Pier 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
  ▶ 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
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 gigahertz.
- This is a \textit{sparse} parametric estimation problem, only the location and amplitude of the pulses need to be estimated.
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 $\varphi(t) = \text{sinc}(t)$.

![Diagram of Classical Sampling Formulation](image-url)
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 $x$ is $\|x\|_0 = \text{the number of } i \text{ such that } x_i \neq 0$
- 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 $A$:

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

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

$$
\min_{\alpha} \|y - D\alpha\|_2 + \lambda \|\alpha\|_1
$$
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...
In compressed sensing you discretize the sampling problem and assume \( x \) is a long vector of size \( M \).

- For the time being call it \( \alpha \) and assume it is \( K \)-sparse.
- The acquisition process stays linear and is modelled with a fat matrix leading to the samples \( y \). (short vector of size \( N \))
The ‘fat’ matrix \( D \) now plays the role of the acquisition device and we denote it with \( \Phi \). The entries of \( y = \Phi \alpha \) are the samples.

Based on the previous analysis, we want to reconstruct the signal \( \alpha \) from the samples \( y \) using \( l_1 \) minimization.

We want maximum incoherence of the columns of \( \Phi \).

We consider large \( M, N \).
Compressed Sensing Formulation

Key Insights

▶ Since $\Phi$ is the ‘acquisition device’, you can choose the $\Phi$ 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 $\Phi$ 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 $\Phi D$ is essentially as random as the original one. Therefore the theorem is still valid. Thus random matrices provide 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 \(\Phi\) 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 \(\Phi\) 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.
In compressed sensing, we discretise a problem which is inherently ‘analogue’ once the size $M$ of $x$ 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.
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 determined by a finite number of free parameters.
Signals with Finite Rate of Innovation

Consider a signal of the form:

\[ x(t) = \sum_{k \in \mathbb{Z}} \gamma_k g(t - t_k). \]  

(1)

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

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

(2)

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

**Definition** A signal with a **finite rate of innovation** is a signal whose parametric representation is given in (1) and with a finite \( \rho \) as defined in (2).
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 $x(t)$.
The Sampling Kernel

Given by nature

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

\[ x(t) \xrightarrow{\text{h}(t)=\varphi(-t/T)} y(t) \xrightarrow{T} y_n=\langle x(t), \varphi(t/T-n) \rangle \]

Acquisition Device

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 \( c_{m,n} \). 

Pier Luigi Dragotti
Sparse Signal Processing Part 2: Sparse Sampling
From Samples to Signals

Because of linearity of inner product, we have that

\[ s_m = \sum_n c_{m,n} y_n \]

\[ = \langle x(t), \sum_{n=0}^{N-1} c_{m,n} \varphi(t/T - n) \rangle \quad m = 0, 1, \ldots, L. \]

Assume that \( \sum_n c_{m,n} \varphi(t/T - n) \simeq e^{j\omega_m t/T} \) for some frequencies \( \omega_m \)
\( m = 0, 1, \ldots, L \)
From Samples to Signals

Then

\[ s_m = \sum_n c_{m,n} y_n \]

\[ = \langle x(t), \sum_n c_{m,n} \varphi(t/T - n) \rangle \]

\[ \simeq \int_{-\infty}^{\infty} x(t) e^{j\omega_m t} dt, \quad m = 0, 1, \ldots, L. \]

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**

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

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

- **Approximation error**

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

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

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).
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:

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

- Approximation error

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

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

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

\( \varphi(t) \) is an E-spline
Approximate Strang-Fix

\[
\sum_{n \in \mathbb{Z}} c_{m,n} \varphi(t - n) \simeq e^{-j\omega_m t} \quad \forall m \in \{1, 2, \ldots, M\}
\]

\(\varphi(t)\) from a real camera
From Samples to Signals

\[ s_m = \sum_n c_{m,n} y_n \]

\[ = \langle x(t), \sum_n c_{m,n} \varphi(t/T - n) \rangle \]

\[ \simeq \int_{-\infty}^{\infty} x(t) e^{j\omega_m t} dt, \quad m = 0, 1, \ldots, L. \]

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 $\hat{x}(j\omega)$:
  
  $$y_n \Rightarrow \hat{x}(j\omega_m) \quad m = 1, 2, \ldots, L$$

- 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:
  
  $$x(t) \Leftrightarrow \hat{x}(j\omega_m) \quad m = 1, 2, \ldots, L$$

- For classes of parametrically sparse signals there is a one-to-one mapping between samples and signal:
  
  $$x(t) \Leftrightarrow \hat{x}(j\omega_m) \quad m = 1, 2, \ldots, L$$

- 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 $N$:
  \[ x(t) = \sum_{k=0}^{K-1} x_k \delta(t - t_k), \quad t_k \in [0, N). \]
- We restrict $j\omega_m = j\omega_0 + jm\lambda \quad m = 1, ..., L \quad \text{and} \quad L \geq 2K$.
- We have $N$ samples: $y_n = \langle x(t), \varphi(t - n) \rangle$, $n = 0, 1, ... N - 1$:
- We obtain

\[
\begin{align*}
    s_m &= \sum_{n=0}^{N-1} c_{m,n} y_n \\
    &= \int_{-\infty}^{\infty} x(t) e^{i\omega_m t} dt, \\
    &= \sum_{k=0}^{K-1} x_k e^{i\omega_m t_k} \\
    &= \sum_{k=0}^{K-1} \hat{x}_k e^{i\lambda m t_k} = \sum_{k=0}^{K-1} \hat{x}_k u_k^m, \quad m = 1, ..., L.
\end{align*}
\]
The Annihilating Filter Method

▶ The quantity

\[ s_m = \sum_{k=0}^{K-1} \widehat{x}_k u_k^m, \quad m = 0, 1, \ldots, L \]

is a sum of exponentials.

▶ We can retrieve the locations \( u_k \) and the amplitudes \( \widehat{x}_k \) with the annihilating filter method (also known as Prony’s method since it was discovered by Gaspard de Prony in 1795).

▶ Given the pairs \( \{u_k, \widehat{x}_k\} \), then \( t_k = (\ln u_k)/\lambda \) and \( x_k = \widehat{x}_k/e^{\alpha_0 t_k} \).
The Annihilating Filter Method

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

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

This filter is thus called the annihilating filter. In matrix/vector form, we have that $SH = 0$ and using the fact that $h_0 = 1$, we obtain

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

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

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

Classic Vandermonde system. Unique solution for distinct \( u_k \)s.
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 $s = Cy$.
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
Algorithm 1: Annihilation of indicator function

1. Compute Fourier transform of the indicator plane $\chi$ evaluated at uniformly sampled frequency grids

2. Create annihilation filter according to (6) where the elements are rearranged lexicographically;

3. Solve the annihilation coefficients $c$.

4. Retrieve the coefficients in (2).

In order to apply the annihilation Algorithm 1, Fourier transform has to be evaluated exactly. FRI is directly either. Because the indicator plane $\chi$ has infinities, no explicit expression of the Fourier transform can be computed exactly. However, in the case of signals investigated previously in 1-D cases, are either streams of annihilable curves here, no explicit expression of the Fourier transform can be computed exactly. However, in the case of simulations, the number of noiseless annihilations, it is better to denoise the samples prior to applying the least square approach with e.g. Cadzow's method [2] for the same reasoning as is in 1-D.

5.1. Exact annihilation for noiseless case

For noise-free cases, the algorithm should exactly annihilate the interior indicator function. In simulations, the number of noise deviations is ubiquitous in signal processing, which may arise from linear system of equations (8) for noise-free cases. However, the annihilating coefficients for noise deviations are not satisfied exactly, yet we can still obtain the solution with the annihilable curve. Thus the algorithm can easily cope with large scale problems.

To evaluate the relative accuracy, we define the percentile error of annihilation algorithm is also investigated in our experiments.

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

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

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  interpol+ curve constraint
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]
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

- $K = 1000$ 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

Super-Resolution Algorithm

Set of low-resolution images

Registration and interpolation

Deconvolution

Super-resolved image
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:

\[ \omega_{mx} s_x + \omega_{my} s_y = \arg \left( \frac{F_1(\omega_{mx}, \omega_{my}) F_2^*(\omega_{mx}, \omega_{my})}{|F_1(\omega_{mx}, \omega_{my}) F_2^*(\omega_{mx}, \omega_{my})|} \right), \]

\[ \forall (\omega_{mx}, \omega_{my}) \text{ s.t.} \quad \frac{1}{|\Phi(\omega_{mx}, \omega_{my})|} \sum_{l \in \mathbb{Z}\setminus\{0\}} \sum_{k \in \mathbb{Z}\setminus\{0\}} |\Phi(\omega_{mx} + 2\pi l, \omega_{my} + 2\pi k)| \leq \gamma. \]
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 low-resolution images
- Registration and interpolation
- Deconvolution
- Super-resolved image

Set of LR images
Image super-resolution: Post registration

Set of low-resolution images

Registration and interpolation

Deconvolution

Super-resolved image

Interpolated HR image
Results: Image super-resolution

One of 100 LR images \((40 \times 40)\).  
Interpolated image \((400 \times 400)\).

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

Figure 6: Double consistency spike search. (i) and (ii) show the detected locations in red and the locations of the original spikes in green for two different window sizes. In (i) the algorithm runs estimating the number of spikes within the sliding window. In (ii) the algorithm runs assuming a fixed number of spikes equal to one for each position of the sliding window. (iii) shows the joint histogram of the detected locations. (iv) shows the fluorescence signal in blue with the original spikes in green and the detected spikes in red.

2.4 Generating surrogate data

We generated surrogate data with similar properties to the experimental data, in order to investigate the changes in performance of the spike detection algorithm in terms of parameters such as data signal to noise ratio and the sampling frequency. We assume that the spike occurrence follows a Poisson distribution with parameter \( \lambda \) spikes/s. Experimental data presents occurrences between 0.45 and 0.5 spikes per second. The probability of having \( k \) spikes in the interval considered in parameter \( \lambda \) (one second) is given by the probability mass function of the Poisson distribution:

\[
 f_{\lambda}(k) = \frac{\lambda^k e^{-\lambda}}{k!}
\]

(17)

To generate a spike train for a time interval \( L \) we divide this interval in \( N \) slots. Each slot corresponds to a time interval of \( \Delta t = \frac{L}{N} \) seconds. The \( \lambda_1 \) parameter that corresponds to this new time interval is \( \lambda_1 = \lambda \frac{1}{\Delta t} \).

We then generate a vector \( k = (k_1, \ldots, k_N) \) of size \( N \) where each \( k_i \) are independent random variables. The \( i \)-th element of this vector, \( k_i \), gives the number of spikes that occurred during the \( i \)-th time slot. We then have to generate the precise instant of time when the spike occurred. For a given time slot, we generate the \( k_i \) spike instants according to a uniform distribution. The total number of spikes in the time interval \( L \) is \( K = \sum_{i=1}^{N} k_i \). Once we have generated the locations of the \( K \) spikes \( p_t \), the waveform given...
Calcium Transient Detection

false positive rate
true positive rate

- FRI
- Fast deconv.
- Deriv.&thres.
- Filter&thres.

Pier Luigi Dragotti
Sparse Signal Processing Part 2: Sparse Sampling
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?
The diffusion equation is

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

where $f(x, t)$ is the source.

When sources are localised in space and time:

$$f(x, t) = \sum_{m=1}^{M} c_m \delta(x - \xi_m, t - \tau_m),$$

denoting the left-hand side of (9) by $\mathcal{L}$ and given the instantaneous source parameterization of $M$ instantaneous sources, providing 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:

\[ Q(k, r) = \langle \Psi_k(x) \Gamma_r(t), f \rangle = \int_{\Omega} \int_t \Psi_k(x) \Gamma_r(t) f(x, t) dt dV, \]

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

\[ f(x, t) = \sum_{m=1}^{M} c_m \delta(x - \xi_m, t - \tau_m), \]

we obtain:

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

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 \( \Psi_k \) is analytic, using Green’s theorem, we obtain:

\[
\int_t \left( \int_\Omega \frac{\partial}{\partial t} (u \psi_k) dV - \mu \oint_{\partial \Omega} (\psi_k \nabla u - u \nabla \psi_k) \cdot \hat{n}_{\partial \Omega} dS \right) dt = \int_t \int_\Omega \psi_k f dV dt = Q(k, 0).
\]
Localisation of Diffusion Sources using Sensor Networks

Assume $r = 0$, since $\Psi_k$ is analytic, using Green’s theorem, we obtain:

$$\int_t \left( \int_\Omega \frac{\partial}{\partial t} (u\psi_k) \, dV - \mu \int_{\partial \Omega} (\psi_k \nabla u - u \nabla \psi_k) \cdot \hat{n}_{\partial \Omega} \, ds \right) \, dt = \int_t \int_\Omega \psi_k f \, dV \, dt = Q(k, 0).$$

▶ 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 \( \Psi_k \) is analytic, using Green’s theorem, we obtain:

\[
\int_t \left( \int_{\Omega} \frac{\partial}{\partial t} (u \psi_k) \, dV - \mu \int_{\partial\Omega} (\psi_k \nabla u - u \nabla \psi_k) \cdot \hat{\mathbf{n}}_{\partial\Omega} \, dS \right) \, dt = \int_t \int_{\Omega} \psi_k f \, dV \, dt = Q(k, 0).
\]

- 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 \( \Psi_k \) is analytic, using Green's theorem, we obtain:
\[
\int_{t_0}^{t_f} \left( \int_{\Omega} \frac{\partial}{\partial t} (u\psi_k) \, dV - \mu \int_{\partial\Omega} (\psi_k \nabla u - u \nabla \psi_k) \cdot \hat{n} \, dS \right) \, dt = \int_{t_0}^{t_f} \int_{\Omega} \psi_k f \, dV \, dt = \mathcal{Q}(k, 0).
\]

- 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

\[
\int_t \left( \int_{\Omega} \frac{\partial}{\partial t} (u\psi_k) \, dV - \mu \int_{\partial\Omega} (\psi_k \nabla u - u \nabla \psi_k) \cdot \hat{n} \, dS \right) \, dt = \int_{t_0}^{t_f} \int_{\Omega} \psi_k f \, dV \, dt = \mathcal{Q}(k, 0).
\]
Localisation of Diffusion Sources using Sensor Networks

Assume \( r = 0 \), since \( \Psi_k \) is analytic, using Green’s theorem, we obtain:

\[
\int_t \left( \int_\Omega \frac{\partial}{\partial t} (u \Psi_k) \, dV - \mu \int_{\partial\Omega} (\Psi_k \nabla u - u \nabla \Psi_k) \cdot \hat{n}_{\partial\Omega} \, dS \right) \, dt = \int_t \int_\Omega \psi_k \, f \, dV \, dt = Q(k, 0).
\]

- 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 $\Psi_k$ is analytic, using Green’s theorem, we obtain:

$$
\int_t \left( \int_\Omega \frac{\partial}{\partial t} (u \Psi_k) \, dV - \mu \int_{\partial \Omega} (\Psi_k \nabla u - u \nabla \Psi_k) \cdot \hat{n} \, dS \right) \, dt = \int_t \int_\Omega \Psi_k f \, dV \, dt = Q(k,0).
$$

- 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
Localisation of Diffusion Sources: Numerical Results

- Figure 2a: Ideal (noiseless) measurement samples.
- Figure 2b: 100 independent trials using noisy sensor measurement samples (SNR=15dB).

<table>
<thead>
<tr>
<th>Source Index</th>
<th>Estimate ($t_1$)</th>
<th>Estimate ($t_2$)</th>
<th>Estimate ($t_3$)</th>
<th>Estimate ($t_4$)</th>
<th>True Activation Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.21</td>
<td>4.126</td>
<td>7.178</td>
<td>15.781</td>
<td>1.218</td>
</tr>
<tr>
<td>2</td>
<td>1.21</td>
<td>4.2775</td>
<td>7.87</td>
<td>15.8475</td>
<td></td>
</tr>
</tbody>
</table>

In this section, we provide simulation results showing the performance of our multi-source estimation algorithm. We simulate the 2-D field governed by the diffusion equation, in particular we consider the setting where the field is induced by four sources activated at different times. Samples of the integrated field, as well as the averaging of the multiple activation time estimates from the nearest sensors to the source.

This algorithm is evaluated in Section 5 using synthetic field measurements. For activation times see legend in Figure 2a. Source Estimation from circular (radius = 4 m) and uniformly inside (Ω) the bounded region ($\Omega$)

$\frac{\partial}{\partial t}c(x, t) = \Delta c(x, t) + k \delta(x - x_0(t))$ for the test function family $\phi(x, t) = \phi(x, t_0)$ (0 to $10^6$ Hz).

The simulation parameters are summarized below:

- $\Omega$: Field sampled over $[0, 100] \times [0, 100]$.
- $M$: Field are then collected, at different times.
- $K$: Samples of the concentrations estimates vary marginally around these estimates. For the noisy measurements (SNR=15dB), the accuracy and also reduces the spread of the estimates in the higher degree.

Localization of Diffusion Sources: Numerical Results

<table>
<thead>
<tr>
<th>Source Index</th>
<th>Estimate ($t_1$)</th>
<th>Estimate ($t_2$)</th>
<th>Estimate ($t_3$)</th>
<th>Estimate ($t_4$)</th>
<th>True Activation Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.21</td>
<td>4.126</td>
<td>7.178</td>
<td>15.781</td>
<td>1.218</td>
</tr>
<tr>
<td>2</td>
<td>1.21</td>
<td>4.2775</td>
<td>7.87</td>
<td>15.8475</td>
<td></td>
</tr>
</tbody>
</table>

For the test function family $\phi(x, t) = \phi(x, t_0)$, for the source intensities given ideal measurements, $c_\hat{}(m) = 1$ for $m = 1, 2, \ldots, 3$.

Integrated field, as well as the averaging of the multiple activation time estimates from the nearest sensors to the source.

Source Estimation from circular (radius = 4 m) and uniformly inside (Ω) the bounded region ($\Omega$). Locating and instantaneous. Simulations demonstrate that the estimation problem when the sources are spatially-localized and instantaneous. The boundary sensors however have little effect on the estimation accuracy. In addition, we retrieve the following measurements, for the source intensities given ideal measurements, $c_\hat{}(m) = 1$ for $m = 1, 2, \ldots, 3$.

The simulation parameters are summarized below:

- $\Omega$: Field sampled over $[0, 100] \times [0, 100]$.
- $M$: Field are then collected, at different times.
- $K$: Samples of the concentrations estimates vary marginally around these estimates. For the noisy measurements (SNR=15dB), the accuracy and also reduces the spread of the estimates in the higher degree.
Localisation of Diffusion Sources: Real Data

- Initial Thermal Map
- 2D spatial distribution with a source indicated.

- Field intensity over time
- Plots showing the intensity of the field over time for different activation times.

- Activation Time
- Graphs showing the activation time with error.

- Sampling and Reconstruction of Diffusion Fields
  - Consideration of two types of spatially localized sources.
  - Temporally instantaneous sources.

- Thermal Camera
  - Used to capture thermal images.
  - Frame rate and sampling frequency considerations.

- Heat Source
  - Using heat gun and silicon wafer (1mm thick).

- Independent Trial Index
- Data summarization and error analysis.

- Algorithm 2
  - Processing of spatiotemporal measurements.

- Validation
  - Robustness to the choice of parameters.
  - Different source setups considered.

- Expression for the test function family
  - Novel expressions presented.

- Table III
  - Summary of location and activation time estimates.
Localisation of Diffusion Sources: Real Data

Initial Thermal Map

Activation

Activation Time

MAE

f

s = 0.22m
d = 0.1m
Thermal Camera
Heat source using a heat gun
Silicon wafer (1mm thick)

Sparse Signal Processing Part 2: Sparse Sampling
Pier Luigi Dragotti

Field intensity

Time (s)

MAE of the activation time is around 0.0036.

Ω    s

f sampling frequency

MAE of the location and activation time estimates in Table III. For this estimate has been obtained by applying our proposed algorithm on spatiotemporal measurements obtained at the 13 ◦ locations marked by black circles.

Moreover, the estimated source location is shown as the (light) region of the map indicating the true source location.

Fig. 9. Measurements of two monitoring sensors obtained by two different activations.

Fig. 8. Experimental Setup.

In this paper we have presented novel expressions for the field induced by the single instantaneous source with spatial dimensions of the monitored region, and also smaller than an order of magnitude smaller than the temporal sampling interval for the test function family. (a): Shows the complete temperature distribution of the monitored region, and also smaller than an order of magnitude smaller than the temporal sampling interval for the test function family. (b): Summarizes the results of 20 repetitions of the spatiotemporal measurements. The thermal camera is used to capture a thermal image immediately after source activation, the locations of the spatial locations (the circles) are indicated by the black circles.

TABLE III

Source Parameter

Independent Trial Index

MAE 0.0036 0.0050 0.1544

(s) The Initial Thermal Map.

(b) 20 Independent trials.

(a) and (b) at random, and then downsampling in time by a factor of 13. The true source location is s = 11 7800 m≈.

In addition, the parameter estimates remain close to the true values.

Spatiotemporal measurements are taken for different source setups. We now consider recordings for different source setups.
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!
On Compressed Sensing and its applications


References

On sampling FRI Signals


References (cont’d)

On Image Super-Resolution

On Diffusion Fields

On Neuroscience:
Appendix

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

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

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

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

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