Signal Reconstruction from Limited Number of Measurements: Theory and Algorithms

Minh N. Do

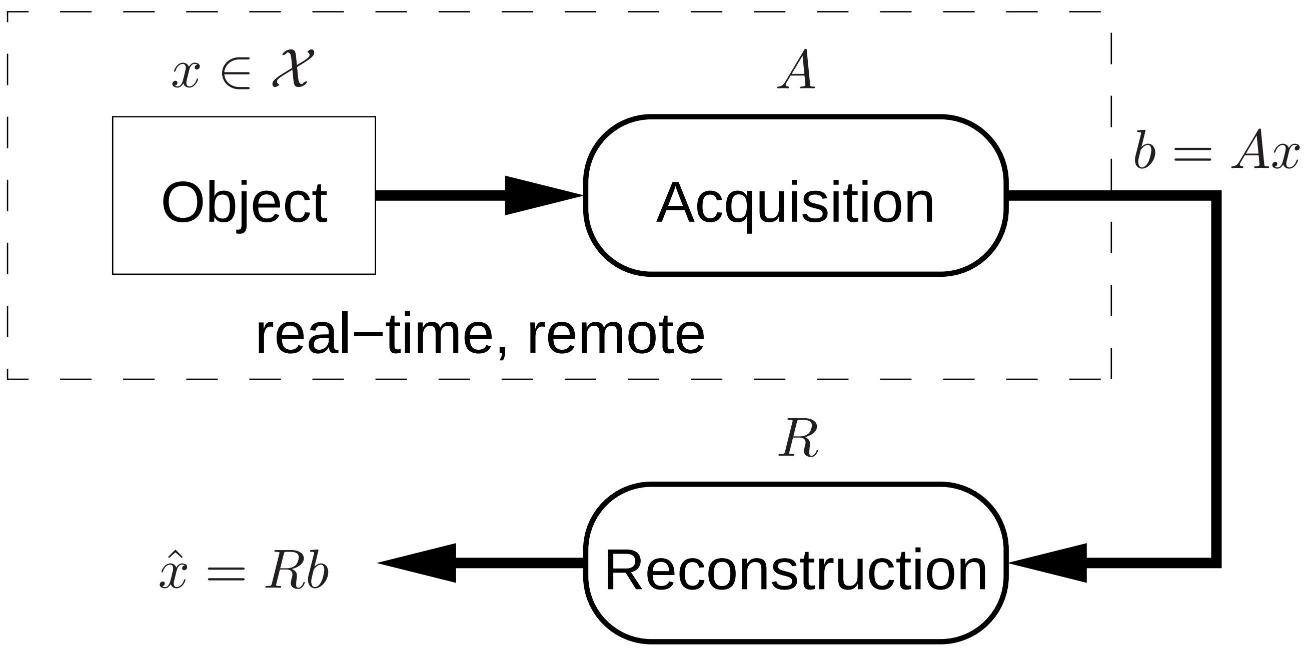
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

Joint work with **Yue Lu** and **Chinh La** (UIUC)

Outline

- Introduction
- Sampling signals from a union of subspaces (with Lu)
- Signal reconstruction using sparse tree representations (with La)
- Conclusion and outlook

The Sensing Problem



- Sensing = Sampling = Representing objects with sequence of real numbers.

- **Requirements:** length(b) is small; A is fast and simple.

- **Goal:** use the prior information that $x \in \mathcal{X}$ to construct A and R.

Classical Sampling

Shannon, 1948

- $\mathcal{X} = \mathsf{BL}([-\frac{\pi}{T}, \frac{\pi}{T}])$.

- A: uniform sampling

$$\begin{aligned} x(t) \mapsto b_n &= x(nT) \\ &= (x * \frac{1}{T}\mathsf{sinc}_T)(nT) \quad \text{if } x \in \mathcal{X} \\ &= \langle x, \frac{1}{T}\mathsf{sinc}_T(\cdot - nT)\rangle_{L_2(\mathbb{R})}, \end{aligned}$$

where $\mathsf{sinc}_T(t) = \frac{\sin(\pi t/T)}{\pi t/T}$.

- R: sinc-interpolation

$$x(t) = \sum_{n\in\mathbb{Z}} x(nT)\mathsf{sinc}_T(t - nT).$$

General Sampling

Unser and Aldroubi, 1994; Unser, 2000

- $\mathcal{X} = \overline{\text{span}}\,\{\phi(t - nT),\ n \in \mathbb{Z}\}$

- A: filtering and sampling

$$x(t) \mapsto b_n = (x * \tilde{\psi})(nT) = \langle x, \psi(\cdot - nT)\rangle_{L_2(\mathbb{R})}$$

- R:

$$\{b_n\}_{n\in\mathbb{Z}} \mapsto \{c_n\}_{n\in\mathbb{Z}}$$
$$x(t) = \sum_{n\in\mathbb{Z}} c_n \phi(t - nT)$$

- **Key:** Sampling signals from a shift-invariant or spline-like space.

More General Sampling: Frames

- $\mathcal{X}$: a Hilbert space.
- A: sequence of linear functionals (including Fourier imaging, tomography,...)

$$x \mapsto b_n = \langle x, \psi_n \rangle_{\mathcal{X}}, \quad n \in \Lambda.$$

- If $\{\psi_n\}_{n\in\Lambda}$ in a frame of $\mathcal{X}$; i.e. there exist two constants (frame bounds) $\alpha > 0$ and $\beta < \infty$ such that for all $x \in \mathcal{X}$

$$\alpha \|x\|_{\mathcal{X}}^2 \leq \sum_{n\in\Lambda} |\langle x, \psi_n \rangle|^2 \leq \beta \|x\|_{\mathcal{X}}^2,$$

  then we can reconstruct x in a numerically stable way from $\{\langle x, \psi_n \rangle\}_{n\in\Lambda}$.

  The tightest frame ratio β/α provides a metric for this stability.

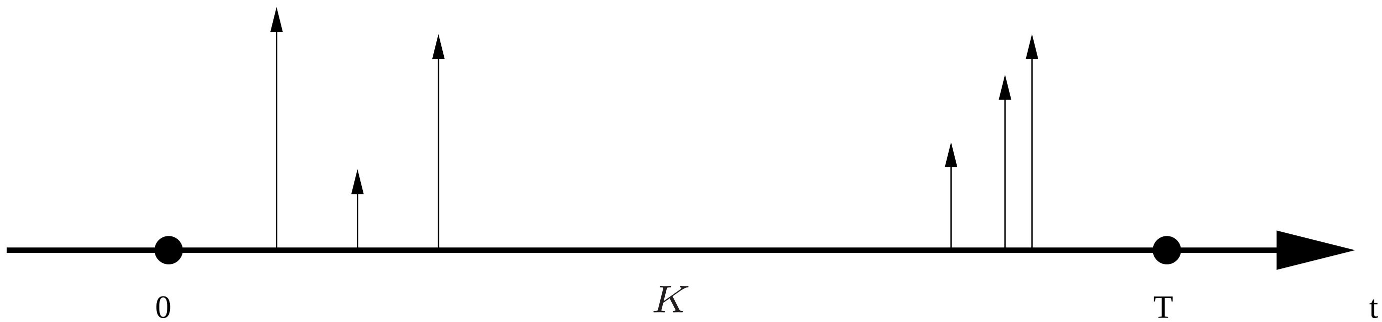
  Example: for matrix multiplication $x \mapsto Ax$, frame ratio $\beta/\alpha = (\kappa(A))^2$.

- R: using dual frame, frame algorithm, conjugate gradient, consistency,...

New Sampling: Signals with Finite Rate of Innovations

Vetterli, Marziliano, and Blu, 2001; also with Maravic and Dragotti

- **Basic model:** stream of Diracs



$$x(t) = \sum_{k=1}^{K} c_k \delta(t - t_k),$$

where the weights $\{c_k\}$ and the locations $\{t_k\}$ are unknown.

- **Key feature:** There is a known finite rate of innovations, but we have find out where are these innovations (e.g. locations of the Diracs).

  $\Rightarrow$ Signals of interest do not fill a vector space.

- **Key result:** Exact reconstruction algorithms for certain signal models and sampling kernels.

Compressed Sensing

Bresler et al., 1999; Donoho, 2004; Candès, Romberg, Tao, 2004; Tropp, 2004; and many others

- $\mathcal{X}$: objects x in $\mathbb{R}^m$ that are compressible by a fixed basis

$$x \approx \mathbf{\Phi} c, \quad \text{where } c \text{ is sparse (i.e. few non-zero entries).}$$

- A: take n $(n \ll m)$ linear non-adaptive measurements; i.e. $A \in \mathbb{R}^{n \times m}$

$$b = Ax \approx \underbrace{A\mathbf{\Phi}}_{M} c$$

- R: solve c from $b = Mc$ with known M and knowing that c is sparse.

- **Key result:** All k-sparse c is recoverable from $b = Mc$ for 'most' random $M \in \mathbb{R}^{n \times m}$, where $k \log(m/k) \ll n \ll m$.

- **Provably good reconstruction algorithms:** Basic Pursuit and Orthogonal Matching Pursuit.

Outline

- Introduction
- Sampling signals from a union of subspaces (with Lu)
- Signal reconstruction using sparse tree representations (with La)
- Conclusion and outlook

Proposed Sampling: Signals from a Union of Subspaces

Lu and Do, 2004

- $\mathcal{X}$: a union of subspaces

$$\mathcal{X} = \bigcup_{\gamma \in \Gamma} \mathcal{S}_\gamma, \quad \text{where } \mathcal{S}_\gamma \text{ are subspaces of a Hilbert space } \mathcal{H}.$$

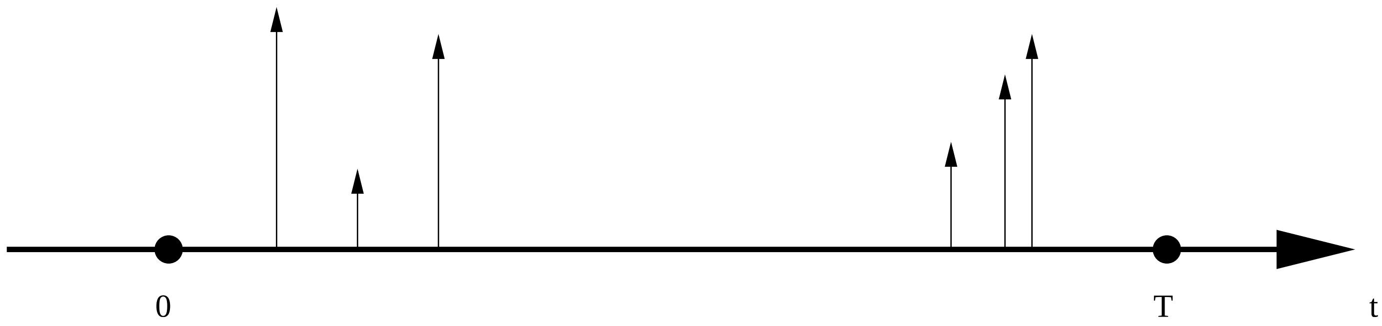
- A: sequence of linear functionals by $\{\psi_n\}_{n \in \Lambda}$ that return measurements

$$b_n = \langle x, \psi_n \rangle_{\mathcal{H}}, \quad n \in \Lambda.$$

  E.g. ψ_n is the point spread function of the n-sensing device.

- **Goals:**
  - Fundamentally extend traditional sampling theorems which are based on the single vector space model.
  - More efficient sampling/sensing schemes for an unknown object x by exploring the prior information that $x \in \mathcal{X}$, instead of just $x \in \mathcal{H}$.

Example 1: Stream of Diracs



$$x(t) = \sum_{k=1}^{K} c_k \delta(t - t_k),$$

- If we fix the locations of Diracs $\gamma \stackrel{\text{def}}{=} (t_1, t_2, \ldots, t_K)$ then

$$x \in \mathcal{S}_\gamma \stackrel{\text{def}}{=} \text{span}\{\delta(t - t_1), \ldots, \delta(t - t_K)\}, \quad \text{dim}(S_\gamma) = K.$$

- With all possible unknown locations, the unknown signal exactly lies on a union of subspaces

$$x \in \mathcal{X} \stackrel{\text{def}}{=} \bigcup_{\gamma \in \mathbb{R}^K} \mathcal{S}_\gamma, \quad \text{dim}(S_\gamma) = K.$$

Example 2: Overlapping Echoes

- Return signal contains up to K overlapping echoes:

$$x(t) = \sum_{k=1}^{K} c_k \phi(t - t_k),$$

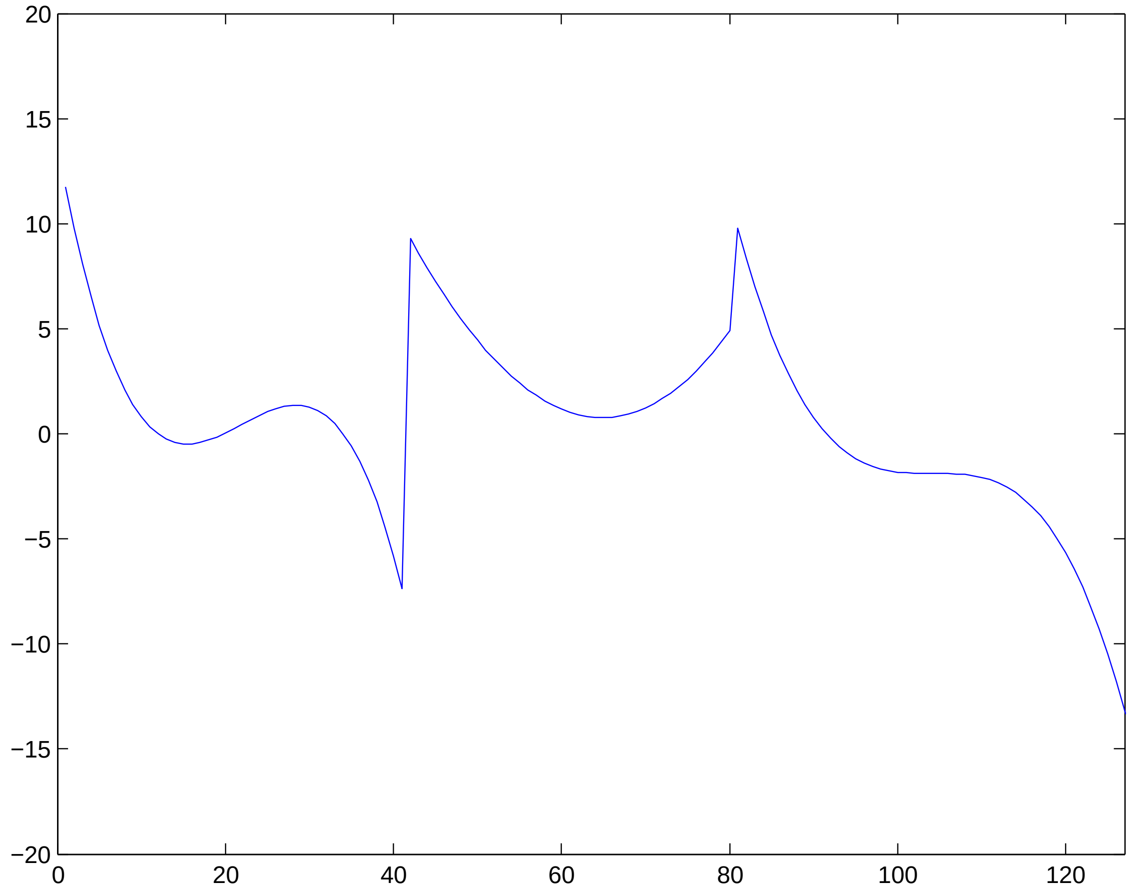
where $\phi(t)$ is a known pulse shape, but delays $\{t_k\}_{k=1}^{K}$ and amplitudes $\{c_k\}_{k=1}^{K}$ are unknown.

- **Applications:** geophysics, radar, sonar, communications,...

- The inverse problem: find out the delays and amplitudes from a limited number of samples of the return signal, have been extensively studied.

- **Note:** the sampling problem for overlapping echos using $\{\psi_n(t)\}_{n=1}^{N}$ is equivalent for stream of Diracs using $\left\{\mathring{\psi}_n(t)\right\}_{n=1}^{N}$, where $\mathring{\psi}_n(\tau) = \int \psi_n(t)\phi(t-\tau)dt$.

Example 3: Piecewise Polynomials or Non-uniform Splines



Similarly: Fix break-points / knots $\Rightarrow$ one subspace
With unknown break-points / knots $\Rightarrow$ union of subspaces.

Example 4: Sparse Approximations

- Consider all K-term approximations using a fixed basis or dictionary $\{\phi_n\}_{n=1}^{\infty}$ (e.g. a Fourier or wavelets basis) as

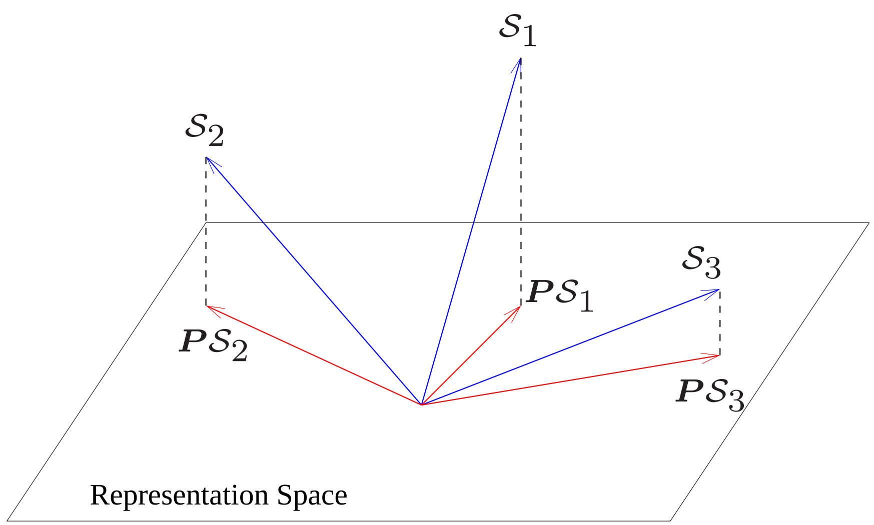
$$\hat{x}_K = \sum_{n \in I_K} c_n \phi_n,$$

where I_K is a set of K selected basis functions or atoms.

- These represent all compressed signals by transform coding or denoised signals by thresholding.

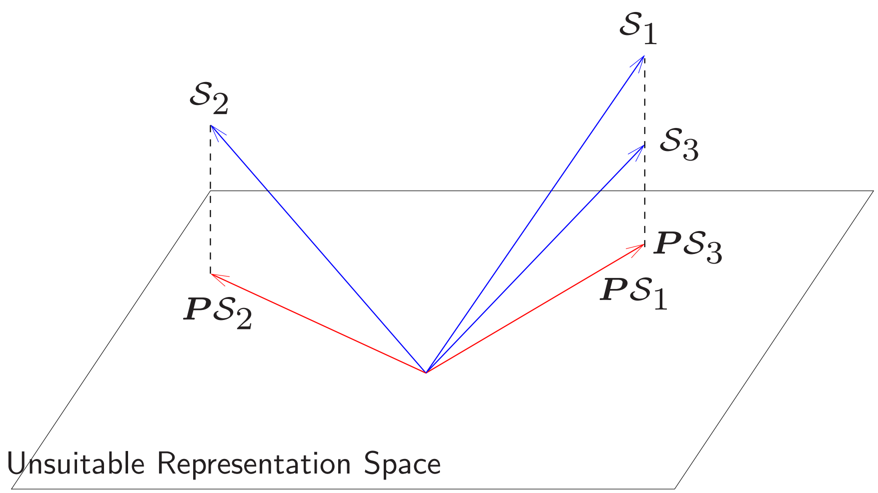
- They lie exactly on a union of subspaces.

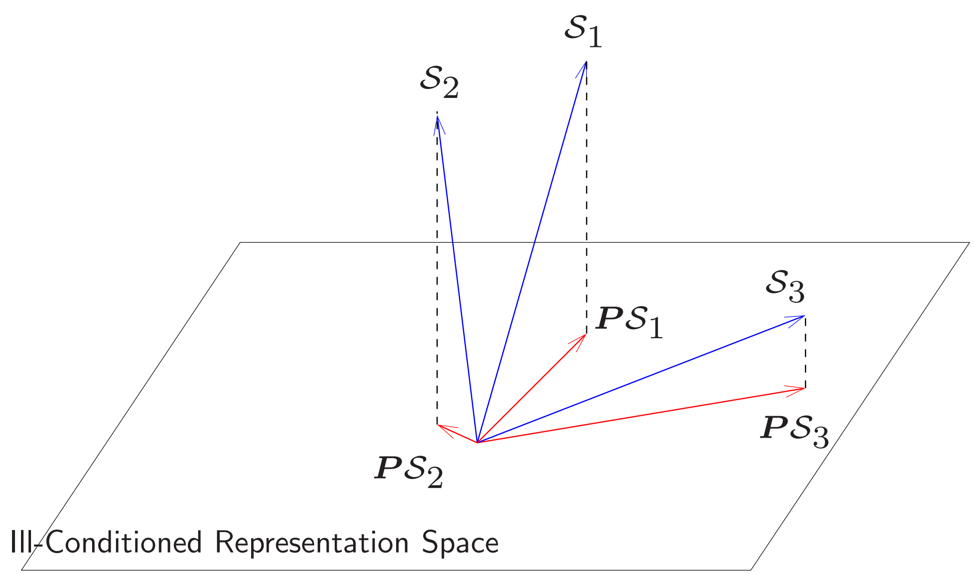
A Geometrical Viewpoint



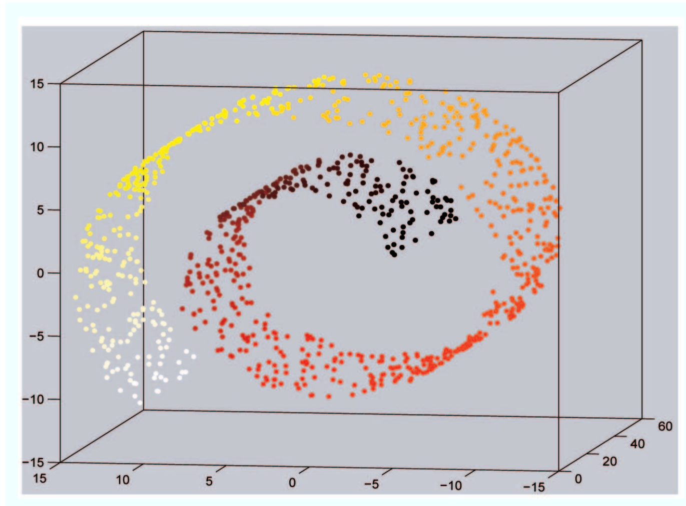
- $\mathcal{X} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$.
- Sampling by $\{\langle x, \psi_n \rangle\}_n$ is equivalent to projecting the signals to a lower dimensional representation space.
- The union of subspace "structure" is preserved $\Longleftrightarrow$ Signals are uniquely determined by their projections.
- Dimension is reduced, without loss of information.

Not All Samplings (Representation Spaces) Are The Same

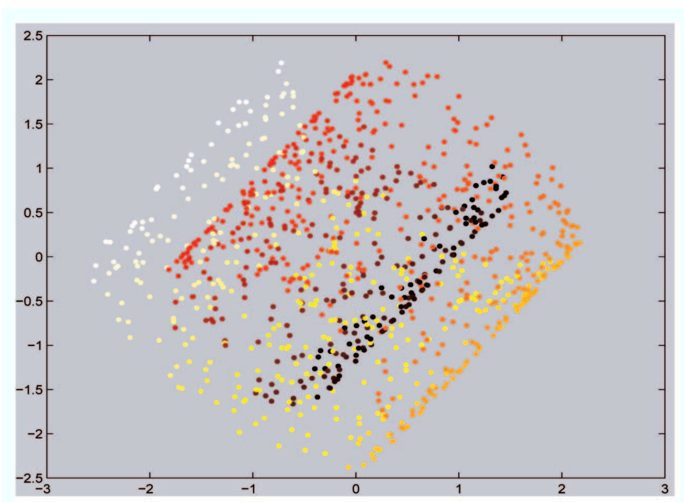




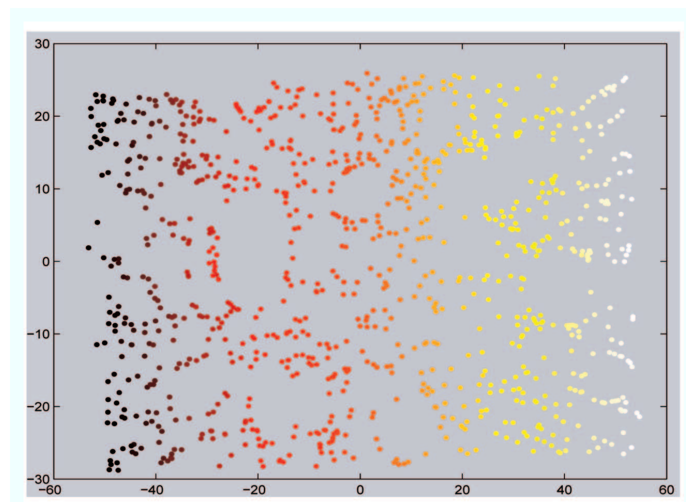
Connection to Dimensionality Reduction



“Swiss roll” data set



After PCA



After ISOMAP

Key Questions

$$\mathcal{X} = \bigcup_{\gamma \in \Gamma} \mathcal{S}_\gamma, \quad \text{where } \mathcal{S}_\gamma \text{ are subspaces of a Hilbert space } \mathcal{H},$$

$$A : x \mapsto b_n = \langle x, \psi_n \rangle_{\mathcal{H}}, \quad n \in \Lambda.$$

- When each object $x \in \mathcal{X}$ is uniquely represented by its sampling data $\{\langle x, \psi_n \rangle\}_{n \in \Lambda}$?
- What is the minimum sampling requirement for a signal class $\mathcal{X}$?
- What are the optimal sampling functions $\{\psi_n\}_{n \in \Lambda}$?
- What are algorithms to reconstruct a signal $x \in \mathcal{X}$ from its sampling data $\{\langle x, \psi_n \rangle\}_{n \in \Lambda}$?
- How stable is the reconstruction in the presence of noise and model mismatch?

Conditions on the Sampling Operator

$$\mathcal{X} = \bigcup_{\gamma \in \Gamma} \mathcal{S}_\gamma, \qquad\qquad A \;:\; x \mapsto b_n = \langle x, \psi_n \rangle_{\mathcal{H}}, \quad n \in \Lambda,$$

Definitions:

- We call A an invertible sampling operator for $\mathcal{X}$ if each $x \in \mathcal{X}$ is uniquely determined by its sampling data Ax; i.e.

$$Ax_1 \neq Ax_2, \quad \text{whenever } x_1 \neq x_2,\ x_1 \in \mathcal{X},\ x_2 \in \mathcal{X}.$$

- We call A a stably invertible sampling operator for $\mathcal{X}$ if there exist two constants $\alpha > 0$ and $\beta < \infty$ such that for all $x_1 \in \mathcal{X}, x_2 \in \mathcal{X}$

$$\alpha \|x_1 - x_2\|_{\mathcal{H}}^2 \leq \|Ax_1 - Ax_2\|_{l_2(\Lambda)}^2 \leq \beta \|x_1 - x_2\|_{\mathcal{H}}^2.$$

We call α and β stability bounds and the tightest ratio β/α provides a metric for the stability of the sampling operator.

Key Observation

- The difficulty in dealing with union of subspaces is that in the previous definitions, x_1 and x_2 can be in two different subspaces.

- We introduce the following subspaces:

$$\widetilde{\mathcal{S}}_{\gamma,\theta} \stackrel{\text{def}}{=} \mathcal{S}_\gamma + \mathcal{S}_\theta = \{y : y = x_1 + x_2, \text{where } x_1 \in \mathcal{S}_\gamma,\ x_2 \in \mathcal{S}_\theta\},$$

and

$$\widetilde{\mathcal{X}} = \bigcup_{(\gamma,\theta)\in\Gamma\times\Gamma} \widetilde{\mathcal{S}}_{\gamma,\theta}.$$

- For example, in the case with streams of K Diracs, $\widetilde{\mathcal{S}}_{\gamma,\theta}$ is a subspace of up to $2K$ Diracs.

Proposition: A linear sampling operator A is stably invertible for $\mathcal{X}$ with stability bounds α and β, *if and only if* for all $y \in \widetilde{\mathcal{X}}$

$$\alpha\|y\|^2_{\mathcal{H}} \leq \|Ay\|^2_{l_2(\Lambda)} \leq \beta\|y\|^2_{\mathcal{H}},$$

Minimum Sampling Requirement

Proposition: A linear sampling operator A is invertible for $\mathcal{X}$ *if and only if* A is invertible for every $\widetilde{\mathcal{S}}_{\gamma,\theta}$, $(\gamma, \theta) \in \Gamma \times \Gamma$.

Proposition: Suppose that $A : x \mapsto \{\langle x, \psi_n \rangle\}_{n=1}^{N}$ is an invertible sampling operator for $\mathcal{X}$. Then

$$N \geq N_{\min} \stackrel{\text{def}}{=} \sup_{(\gamma,\theta)\in\Gamma\times\Gamma} \dim(\widetilde{\mathcal{S}}_{\gamma,\theta}).$$

- **Example:** Streams of K Diracs

  $N_{\min} = 2K$ compare to # of free parameters $= 2K$

- **Example:** Piecewise polynomials on an interval with K pieces, each of degree less than d

  $N_{\min} = (2K - 1)d$ compare to # of free parameters $= Kd + K - 1.$

- **Note:** The reconstruction algorithm in Vetterli et al., 2004 achieves the minimum sampling in both cases.

Existence of Minimal Sampling Operators

Proposition Let $\mathcal{X} = \bigcup_{\gamma \in \Gamma} \mathcal{S}_\gamma$ be a countable union of subspaces of $\mathcal{H}$. Suppose that

$$N_{\min} = \sup_{(\gamma,\theta) \in \Gamma \times \Gamma} \dim(\widetilde{\mathcal{S}}_{\gamma,\theta})$$

is finite. Then the set of sampling vectors $\{\psi_n\}_{n=1}^{N_{\min}}$ such that the associated sampling operator A is invertible for $\mathcal{X}$ is dense in $\mathcal{H}^{N_{\min}}$.

As a result, consider $\mathcal{X}$ as the set of sparse approximations using up to K basis vectors from a countable basis of $\mathcal{H}$.

- An invertible linear sampling operator requires at least $2K$ sampling vectors $\{\psi_n\}_n$.
- An arbitrary set of $2K$ vectors $\{\psi_n\}_n$ will almost surely leads to an invertible sampling operator.

Case Study 1: Streams of Diracs

Consider $\mathcal{X} = \{\text{streams of } K \text{ Diracs}\}$. If $y \in \widetilde{\mathcal{S}}_{\gamma,\theta}$ then

$$y(t) = \sum_{k=1}^{M} c_k \delta(t - t_k), \quad \text{where } \ldots < t_k < t_{k+1} < \ldots, \; M \leq 2K.$$

Let $\{\psi_n\}_{n=1}^{N}$ be the set of continuous sampling functions for A, then

$$(Ay)_n = \langle y, \psi_n \rangle = \sum_{k=1}^{M} c_k \psi(t_k)$$

$$\Rightarrow Ay = Gc, \quad \text{where } G \in \mathbb{R}^{N \times M}, \; G_{n,k} = \psi_n(t_k).$$

Thus, A is invertible for $\mathcal{X}$ *if and only if* G is invertible, or

$$\det([\psi_n(t_k)]_{n,k=1}^{M}) \neq 0, \quad \text{for all } 1 \leq M \leq N, \; t_1 < t_2 < \ldots < t_M.$$

For $N = N_{\min} = 2K$, the set $\{\psi_n\}_{n=1}^{N}$ that satisfies the last condition is called a complete Tchebycheff system.

Tchebycheff systems play an important role in several areas; notably, theory of approximation, methods of interpolation, numerical analysis.

Numerous examples of Tchebycheff systems, including: power functions, Gauss kernel, spline polynomials, $\sin$ and $\cos$ functions.

Case Study 2: Sparse Approximations

$$\mathcal{X} = \left\{ x \;:\; x = \sum_{k \in I,\ |I| = K} c_k \phi_k \right\}, \quad \text{where } \{\phi_k\}_{k=1}^{\infty} \text{ is an orthonormal basis.}$$

$$A \;: \qquad x \mapsto b = Ax, \quad \text{where } b_n = \langle x, \psi_n \rangle$$

$$x = \sum_k c_k \phi_k \mapsto b = Gc, \quad \text{with } G_{n,k} = \langle \phi_k, \psi_n \rangle.$$

Reconstruction problem: solve c from $b = Gc$ subject to $\|c\|_0 \leq K$.

Denote $g_k = [\langle \phi_k, \psi_1 \rangle, \ldots, \langle \phi_k, \psi_N \rangle]^T$ the k-th column of G, and $G_I = [g_m]_{m \in I}$. Then from our propositions, the stability bounds are

$$\alpha = \inf_{|I| = 2K} \lambda_{\min}(G_I^T G_I)$$

$$\beta = \sup_{|I| = 2K} \lambda_{\max}(G_I^T G_I)$$

Lemma:

$$\alpha \geq \inf_{|I|=2K-1, k \notin I} \left(\langle g_k, g_k \rangle - \sum_{l \in I} |\langle g_k, g_l \rangle| \right)$$

$$\beta \geq \inf_{|I|=2K-1, k \notin I} \left(\langle g_k, g_k \rangle + \sum_{l \in I} |\langle g_k, g_l \rangle| \right)$$

Proposition: Suppose $\|g_k\| = 1$ and denote $\mu_1(m) = \sup_{|I|=m, k \notin I} \sum_{l \in I} |\langle g_k, g_l \rangle|$. Then A is a stably invertible sampling operator if

$$\mu_1(2K-1) < 1.$$

Note: Tropp (2004) shows if

$$\mu_1(K-1) + \mu_1(K) < 1.$$

the OMP and BP exactly reconstruct the signal.

Case Study 3: Union of Shift-Invariant Spaces (Infinite-Dimensional Case)

Definition: $\mathcal{S}$ is called a (finitely generated) shift-invariant space, if

$$\mathcal{S}_\Phi = \left\{ \sum_{n\in\mathbb{Z}} \sum_{k=1}^{D} c_{kn}\phi_k(t/T-n) \right\},$$

where $\Phi = \{\phi_k\}_{k=1}^{K}$ are the generating functions.

Signals of interest:

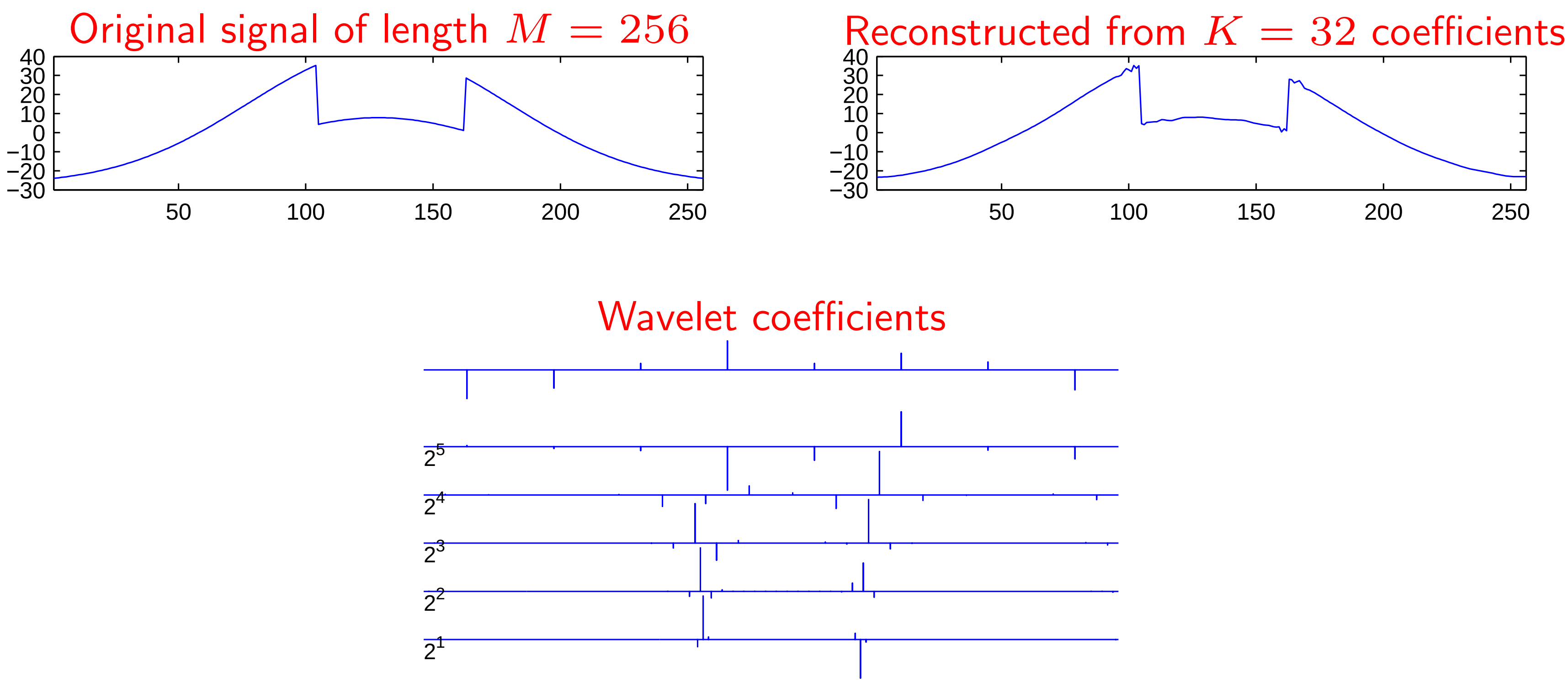
$$\mathcal{X} = \bigcup_{\Phi} \mathcal{S}_\Phi,$$

Example: signals with unknown spectral support.

Outline

- Introduction
- Sampling signals from a union of subspaces (with Lu)
- Signal reconstruction using sparse tree representations (with La)
- Conclusion and outlook

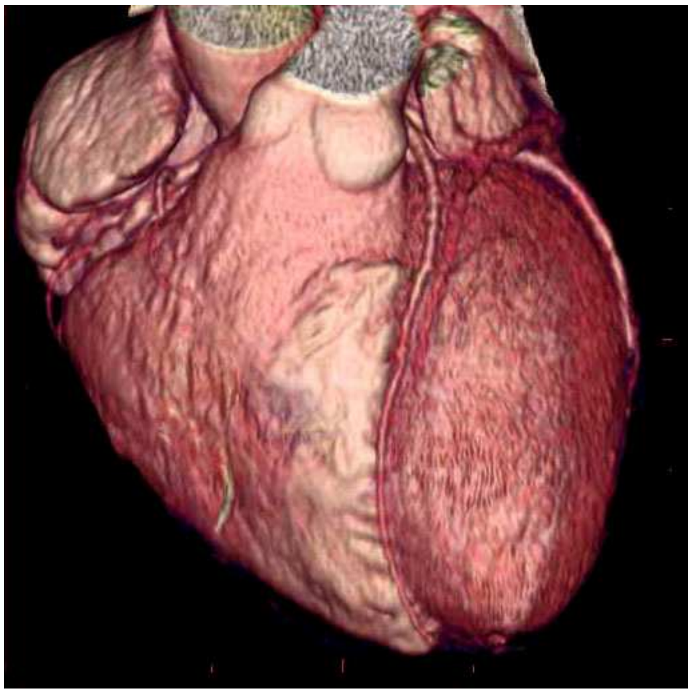
Sparse Tree Representations



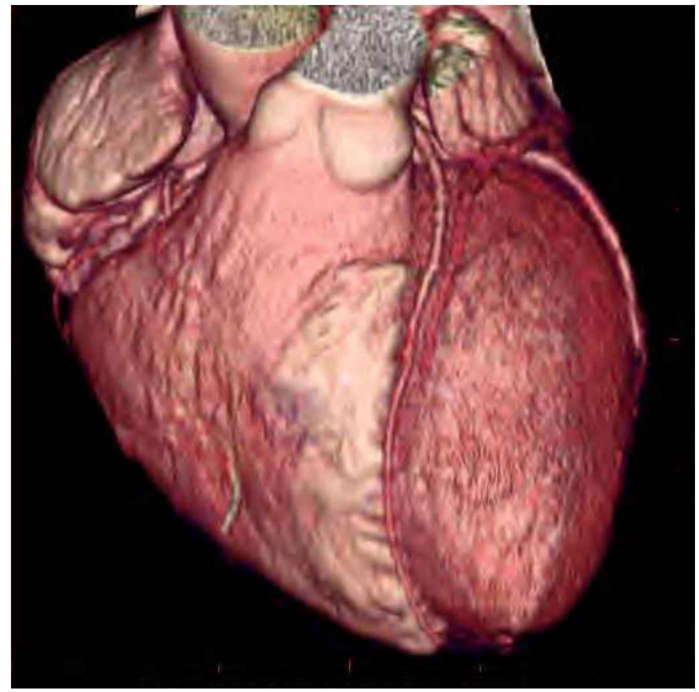
In many multiscale bases (e.g. wavelets), signals of interest (e.g. piecewise-smooth) not only have few significant coefficients, but also those significant coefficients are well-organized in trees.

A Driving Application

MRI with limited number of measurements: MRI measures Fourier coefficients of the unknown image sequentially.



A heart image



Nonlinear approximation using 3% of wavelet coefficients

Goal: Reconstruct a same quality image using about 10% of Fourier coefficients.

Signal Reconstruction using Sparse Tree Representations

- We propose to exploit the sparse tree representation as additional prior information for signal reconstruction with limited numbers of measurements.

- Intuitively, a general sparse representation with K coefficients can be described with $2K$ numbers: K for the values and another K for the locations.

- If these K significant coefficients are known to be organized in trees then the indexing cost is significantly reduced and hence the total description of the unknown signal.

- Exploiting this embedded tree structure in addition to the sparse representation prior in inverse problems would potentially lead to:

  1. better reconstructed signals;
  2. reconstruction using fewer measurements; and
  3. faster reconstruction algorithms.

Proposed: TOMP – Tree-based Orthogonal Matching Pursuit

La and Do, 2005

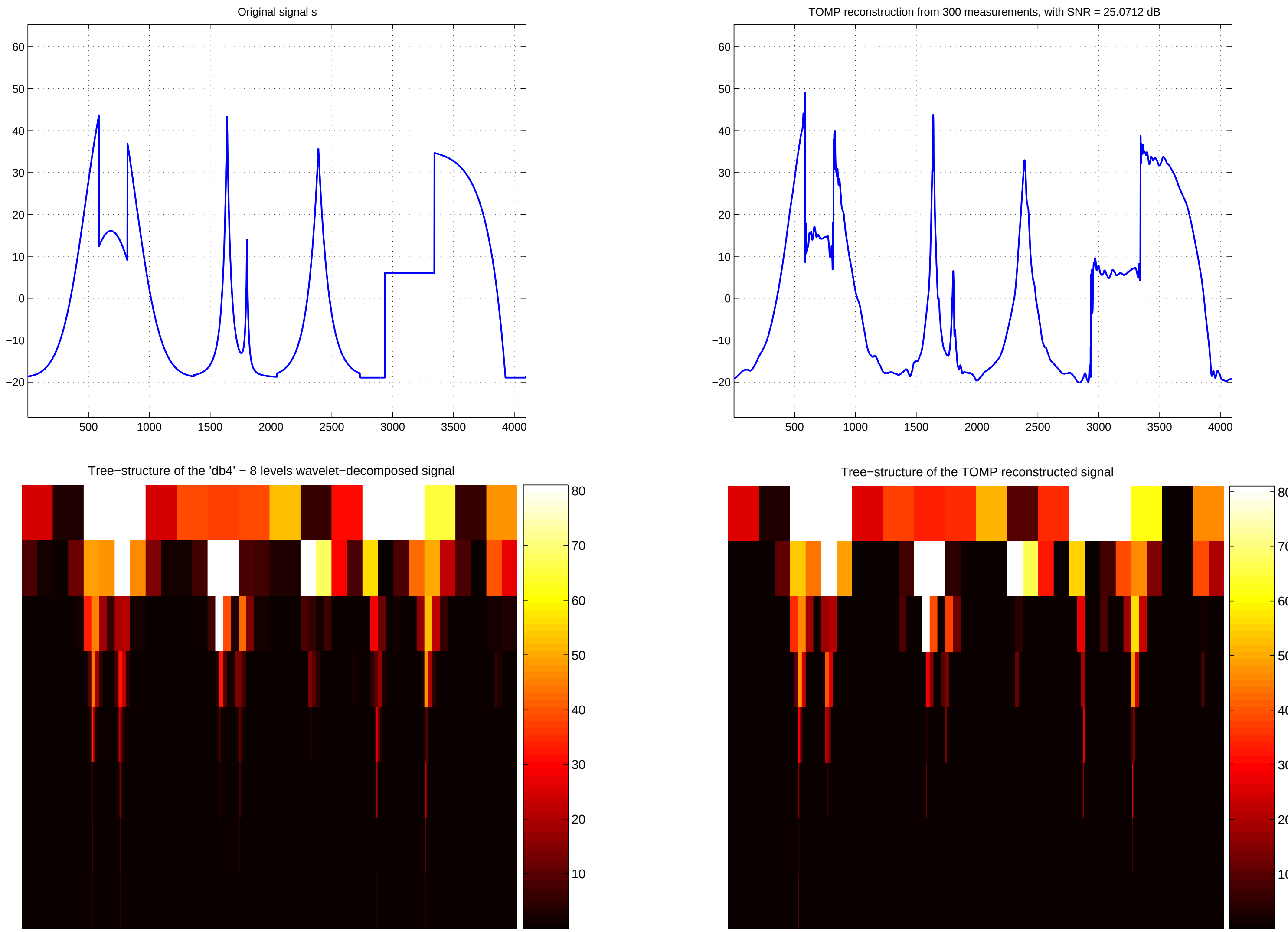
Extension of OMP for solving $Ax = b$ that exploits:

P1 *Vector* x *has sparse structure; i.e. only few entries in* x *are nonzero or significant.*

P2 *Those significant entries of* x *are well organized in a tree structure.*

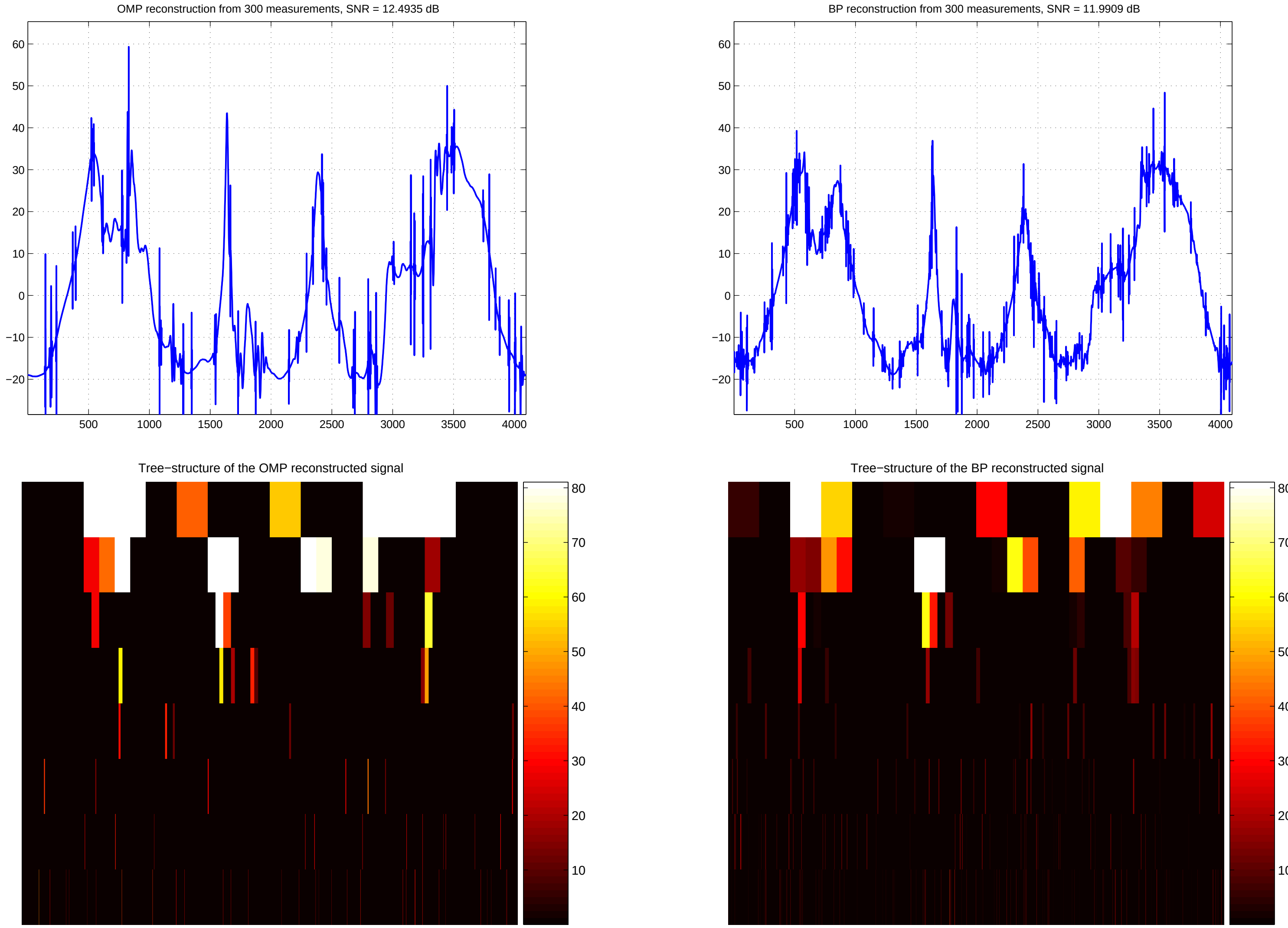
Example

Piecewise-smooth signal of length 4096 and take 300 random measurements.
Reconstruction using TOMP (Tree-based Orthogonal Matching Pursuit).

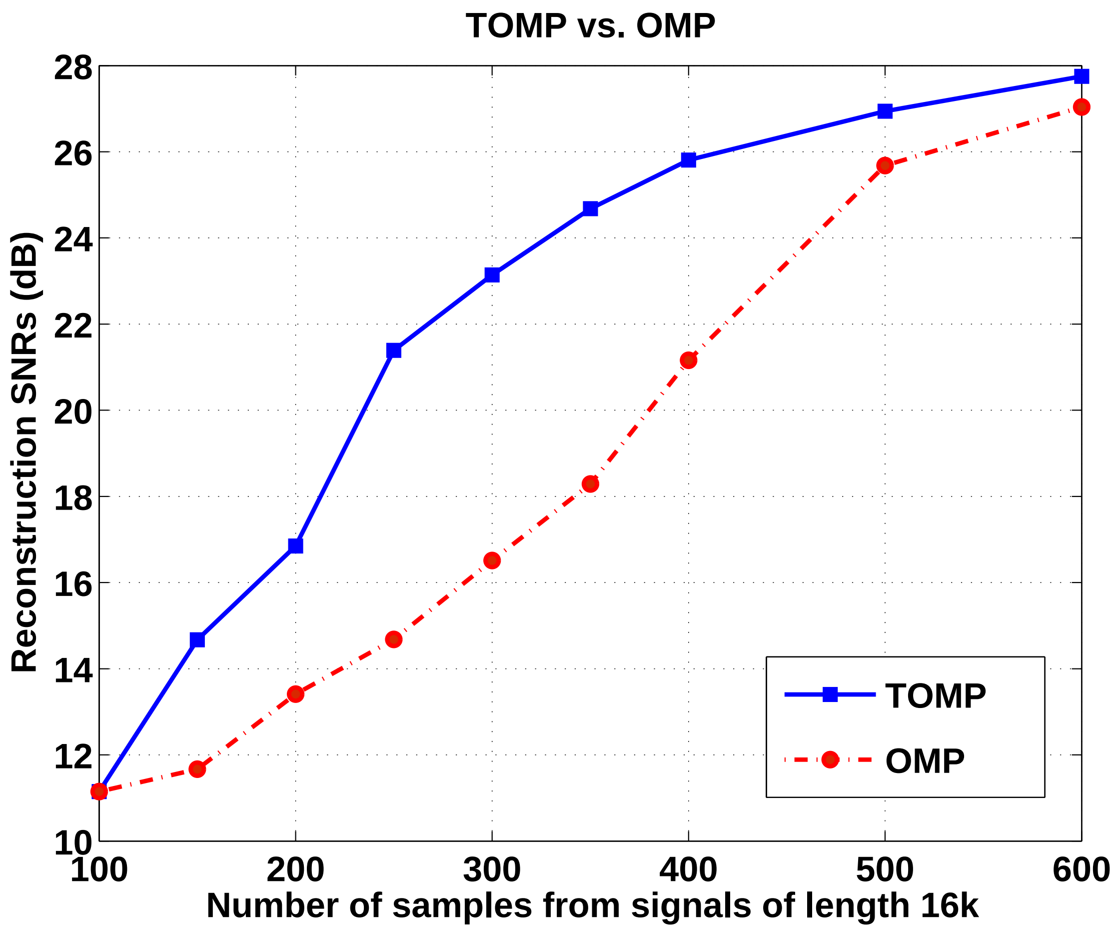


Using Other Methods

Reconstructions from the same set of measurements using OMP (Orthogonal Matching Pursuit) and BP (Basis Pursuit).



Comparison: OMP vs. TOMP



There is a critical sampling region where TOMP improves reconstruction by more than 7 dB, or achieves the same reconstruction quality but using nearly half of number of measurements.

Conclusion

- Sampling signals from a union of subspaces
  - Fundamentally extend traditional sampling theorems which are based on the single vector space model.
  - Sharp results on sampling requirements.
- Signal reconstruction using sparse tree representations
  - Significant gains by exploit the additional sparse tree prior.
- Great opportunity for developing new theory and algorithms that could have impact on applications.

References

- Y. Lu and M. N. Do, A geometrical approach to sampling signals with finite rate of innovation, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Canada, May 2004.

- Y. Lu and M. N. Do, Sampling signals from a union of subspaces, *Manuscript*, 2005.

- C. La and M. N. Do, Signal reconstruction using sparse tree representations, *SPIE Conference on Wavelet Applications in Signal and Image Processing*, San Diego, Aug. 2005.

- C. La and M. N. Do, Tree-based orthogonal matching pursuit algorithm for signal reconstruction, *IEEE International Conference on Image Processing*, Atlanta, Oct. 2006.

www.ifp.uiuc.edu/~minhdo/publications