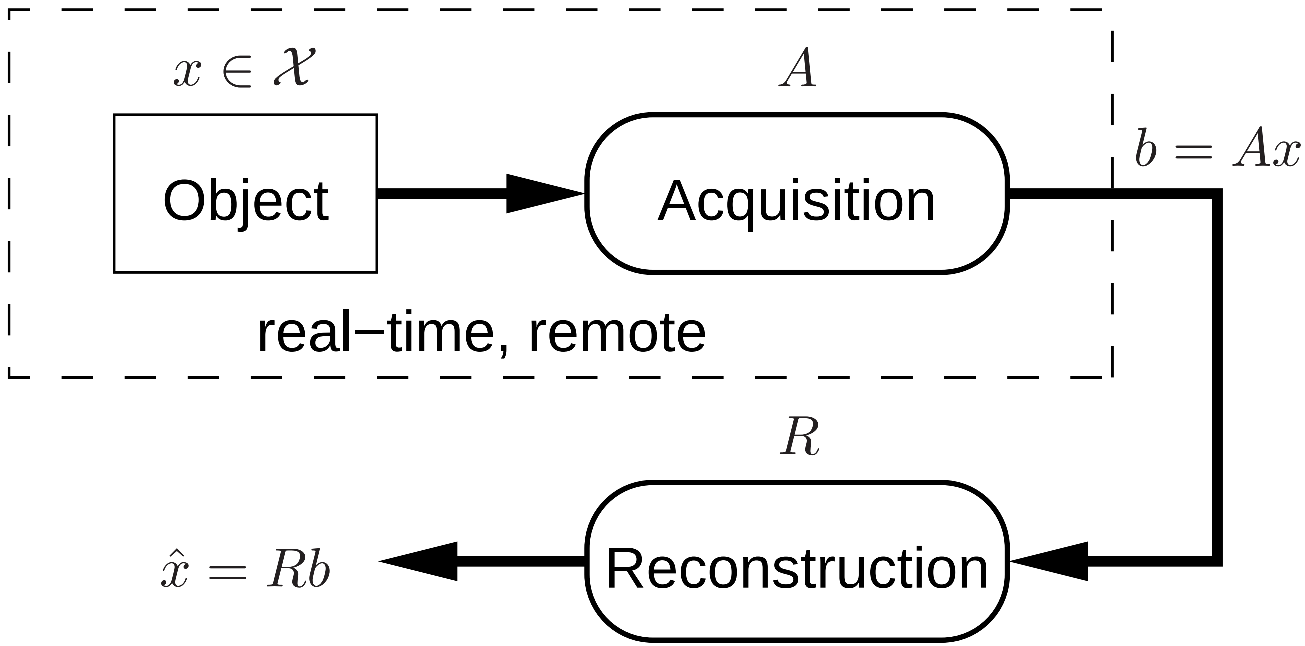
Practical Compressed Sensing

Minh N. Do

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

Collaborators: Yue Lu, Chinh La, and Yoram Bresler

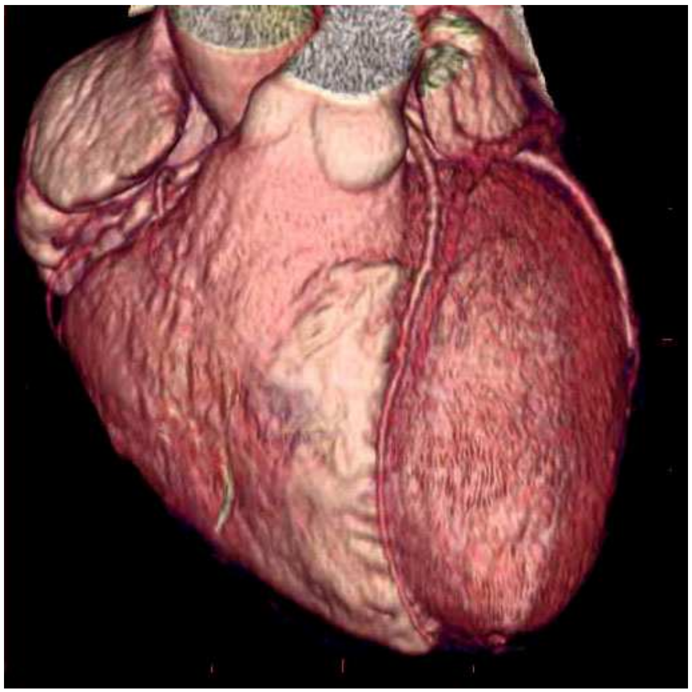
The Sensing/Sampling Problem



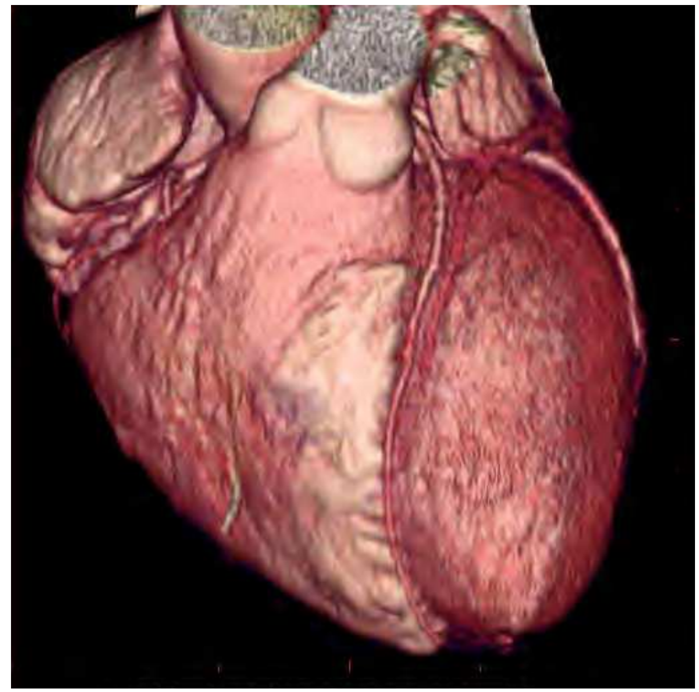
- Sensing $=$ Sampling $=$ Representing objects with a sequence of numbers.
- **Requirements:** length(b) is limited.
- **Goal:** use the prior information that $x \in \mathcal{X}$ to construct A and R.

A Driving Application

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



A heart image



Approximation using 3% of wavelet coefficients

Goal: Reconstruct a same quality image using a fraction of Fourier coefficients.

Classical Sampling

Whittaker, Shannon, Kotelnikov

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

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

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: 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,...

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

$$n = O(\log(m)k).$$

Proposed Sampling: Signals from a Union of Subspaces

Lu and Do, 2004

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

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

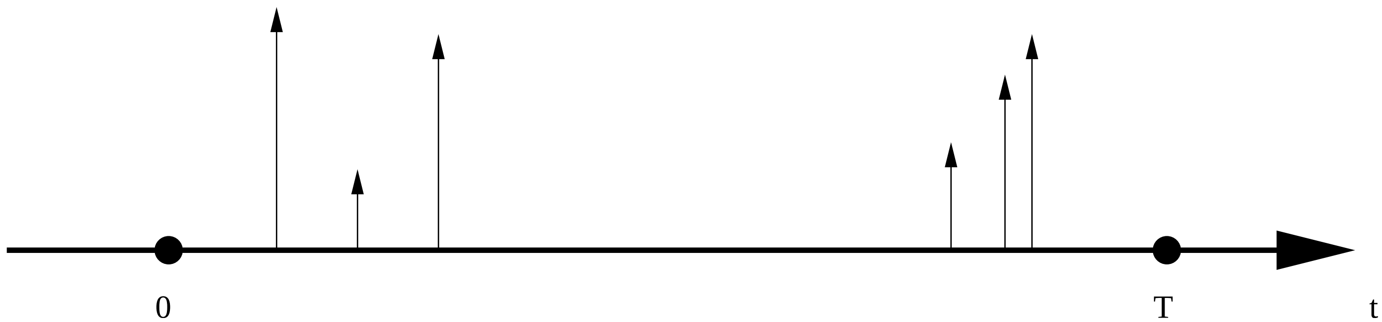
- A: sequence of linear functionals by $\{\psi_n\}_{n \in \Lambda}$ that return measurements

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

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

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

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

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

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: 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

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

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

- They lie exactly on a union of subspaces.
  - If we fix the set I_K of K selected basis vectors then

$$x \in \mathcal{S}_{I_K} \stackrel{\text{def}}{=} \text{span}\{\phi_n\}_{n \in I_K}, \quad \text{dim}(\mathcal{S}_{I_K}) = K.$$

  - With all possible sets of K selected basis vectors, the signal of interest leave on a union of subspaces

$$x \in \mathcal{X} \stackrel{\text{def}}{=} \bigcup_{I_K \text{ s.t. } |I_K| = K} \mathcal{S}_{I_K}.$$

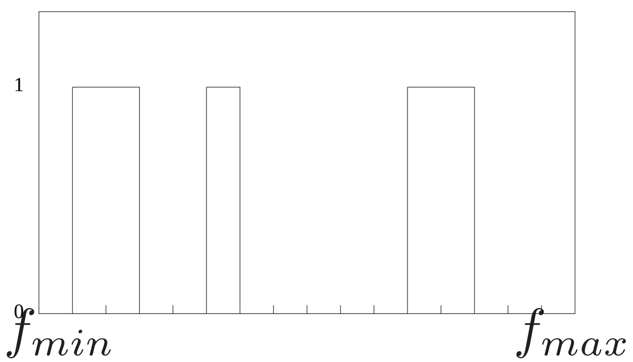
Example 4: Union of Shift-Invariant Spaces

$\mathcal{X} = \bigcup_{\Phi} \mathcal{S}_\Phi$ where $\mathcal{S}_\Phi$ is a (finitely generated) shift-invariant space

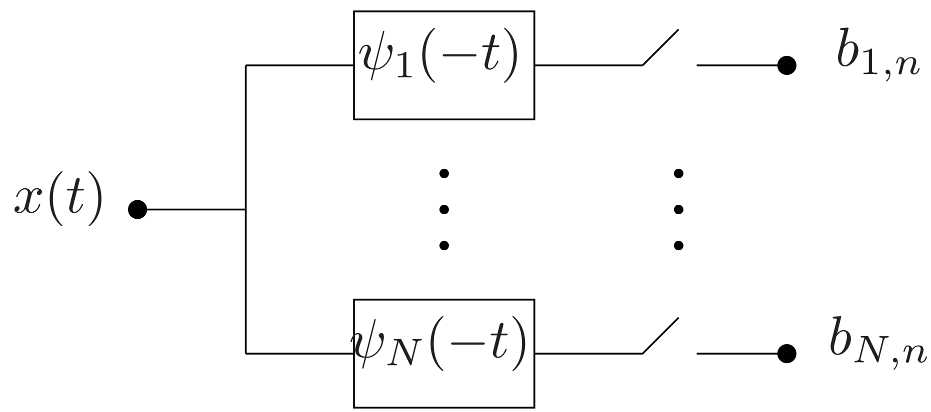
$$\mathcal{S}_\Phi = \left\{ \sum_{k=1}^{K} \sum_{n \in \mathbb{Z}} c_{k,n} \phi_k(t/T - n) \ : \ c_k \in l_2(\mathbb{Z}) \right\},$$

with $\Phi = \{\phi_k\}_{k=1}^{K}$ is a set of generating functions.

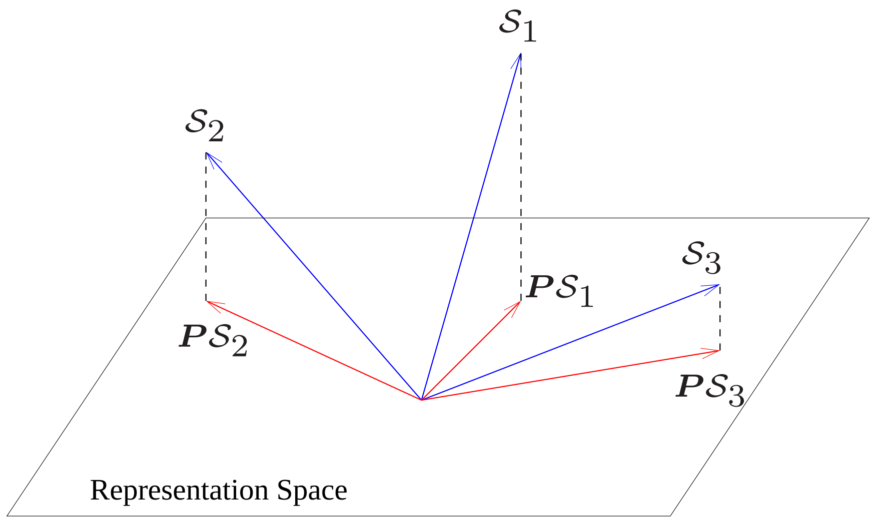
Example: signals with unknown spectral support



Multichannel sampling



A Geometrical Viewpoint



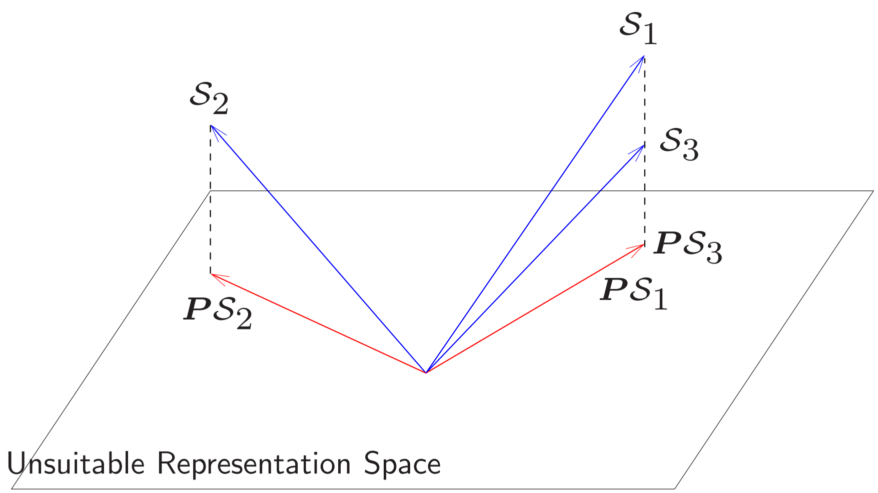
- $\mathcal{X} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$.

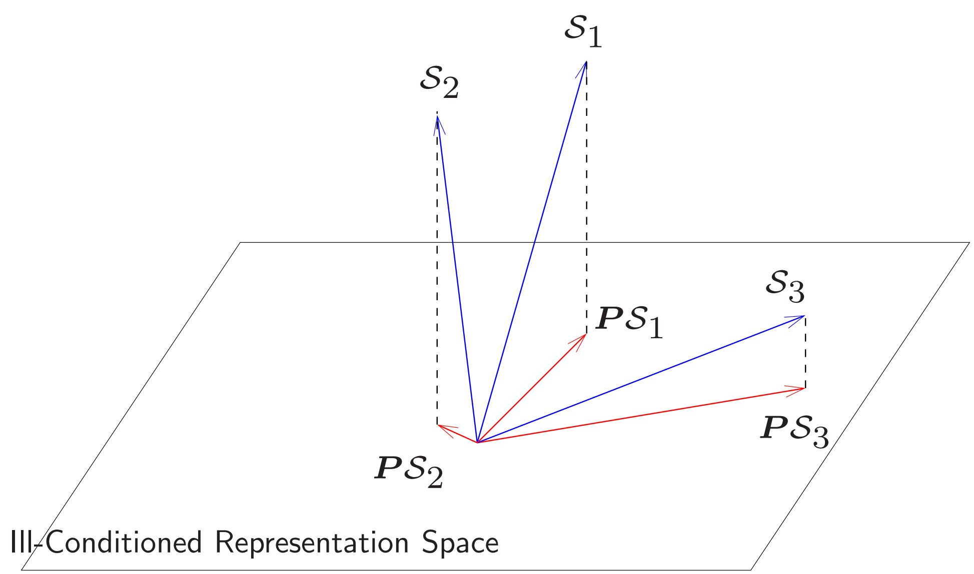
- Sampling by $\{\langle x, \psi_n \rangle\}_n$ is equivalent to projecting the signals to a lower dimensional representation space.

- 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





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

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

Definitions [LuD:06]

- 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 $x_1 \in \mathcal{X}, x_2 \in \mathcal{X}$

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

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:

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

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

Proposition [LuD:06] A linear sampling operator A is invertible for $\mathcal{X}$ *if and only if* A is invertible for each $\widetilde{\mathcal{S}}_{\gamma,\theta}$, $(\gamma, \theta) \in \Gamma \times \Gamma$.

Proposition [LuD:06] A linear sampling operator A is stably invertible for $\mathcal{X}$ with stability bounds α and β, *if and only if*

$$\alpha \|y\|^2_{\mathcal{H}} \leq \|Ay\|^2_{l_2(\Lambda)} \leq \beta \|y\|^2_{\mathcal{H}}, \quad \text{for all } y \in \widetilde{\mathcal{S}}_{\gamma,\theta},\ (\gamma, \theta) \in \Gamma \times \Gamma.$$

Minimum Sampling Requirement

Proposition [LuD:06] 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 [LuD:06] Let $\mathcal{X} = \bigcup_{\gamma \in \Gamma} \mathcal{S}_\gamma$ be a countable union of subspaces of $\mathcal{H}$. Suppose that

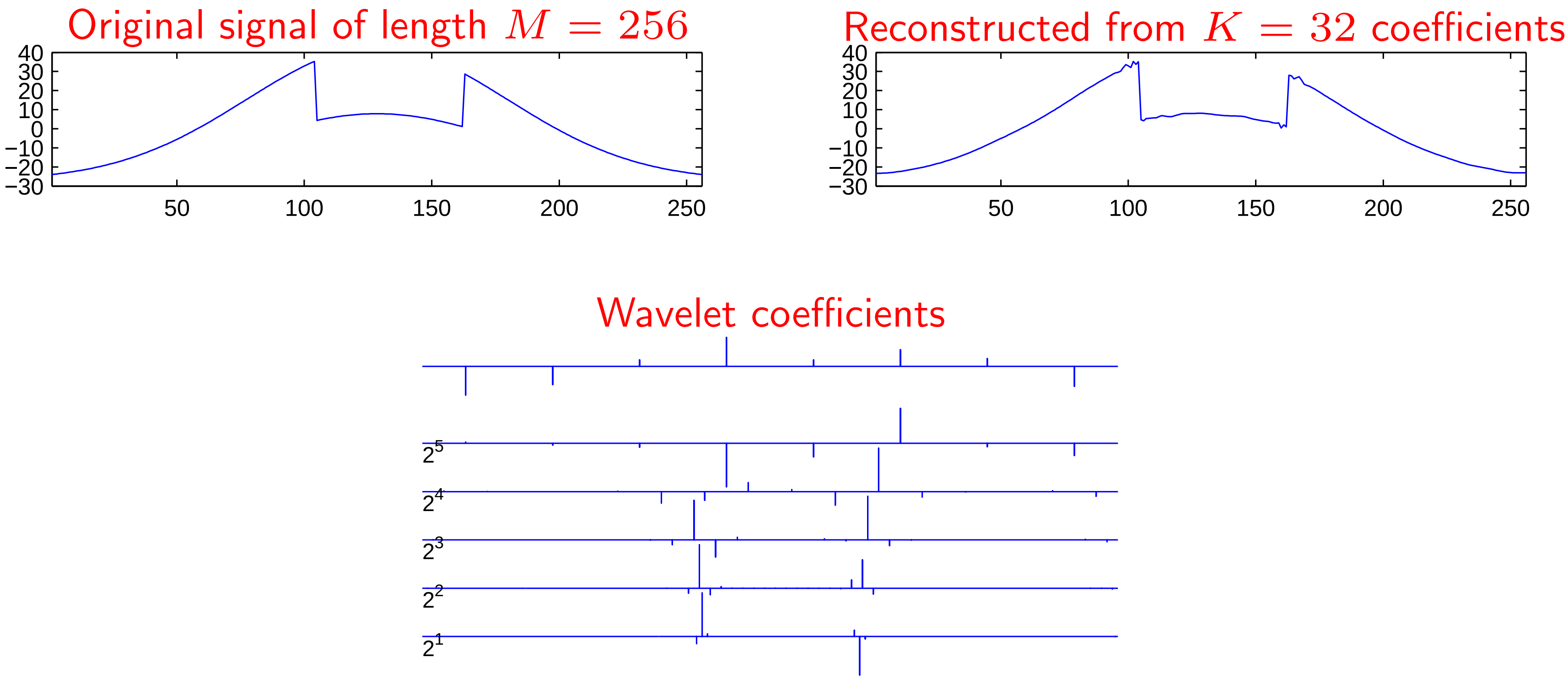
$$N_{\min} = \sup_{(\gamma,\theta) \in \Gamma \times \Gamma} \dim(\widetilde{\mathcal{S}}_{\gamma,\theta})$$

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 $\mathcal{H}^{N_{\min}}$.

Example: Consider $\mathcal{X}$ as the set of all sparse approximations using up to K vectors from a countable dictionary 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 provide to an invertible sampling operator.

Beyond Sparsity



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.

Existing Pursuit Algorithms for Signal Reconstruction

- **Basis Pursuit (BP):** Solve the l_1-norm minimization problem, using linear programming

$$\hat{\boldsymbol{x}}_{BP} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

- **Orthogonal Matching Pursuit (OMP):** Greedy algorithm, start with $\boldsymbol{r}_0 = \boldsymbol{b}$. At iteration k select out from $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m)$ the column that has largest inner product with the residual $\boldsymbol{r}_{k-1}$:

$$i_k = \arg\max_i |\langle \boldsymbol{r}_{k-1}, \boldsymbol{a}_i \rangle|$$

$$\boldsymbol{r}_k = \boldsymbol{r}_{k-1} - P_{\text{span}\left\{\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_{k-1}}\right\}} \boldsymbol{r}_{k-1}$$

Tree-based Orthogonal Matching Pursuit (TOMP)

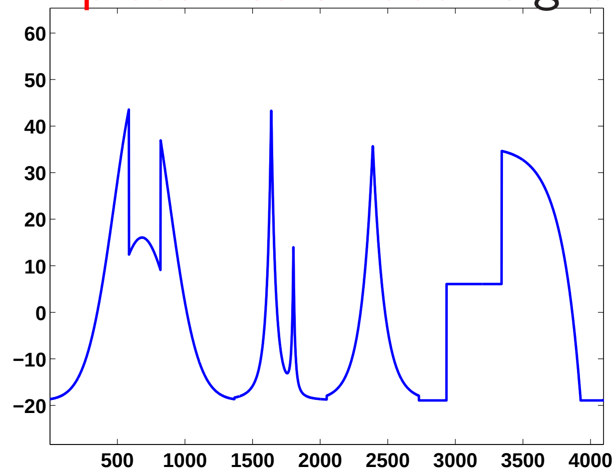
La and Do, 2005

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

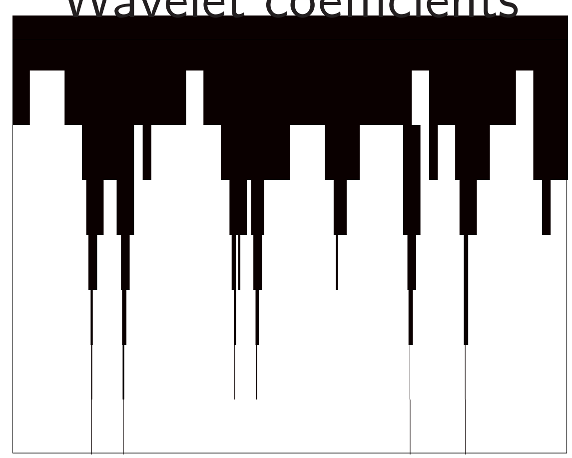
P1 *Vector x has sparse structure; i.e. only few entries in x are nonzero or significant.*

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

A piecewise smooth signal

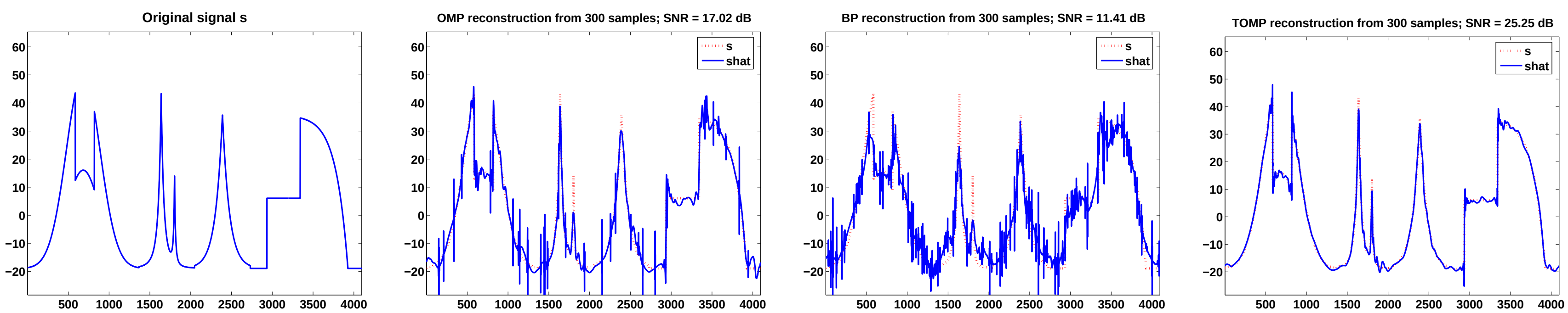


Wavelet coefficients

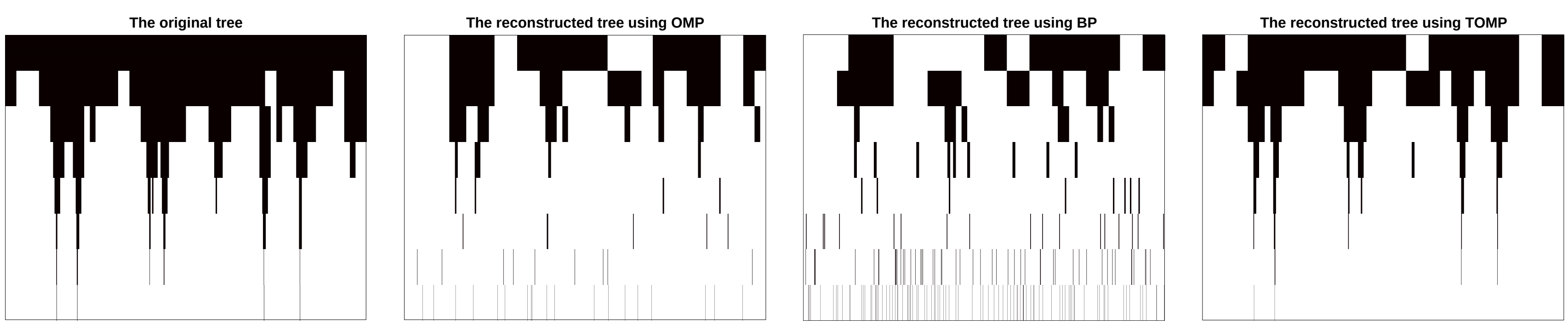


Experiment: A Single Case

Length-4096 piecewise smooth signal, 8-level Daubechies 4 wavelet, reconstruction from 300 linear measurements using OMP, BP and TOMP ($\alpha = 0.975$, $l = 2$).



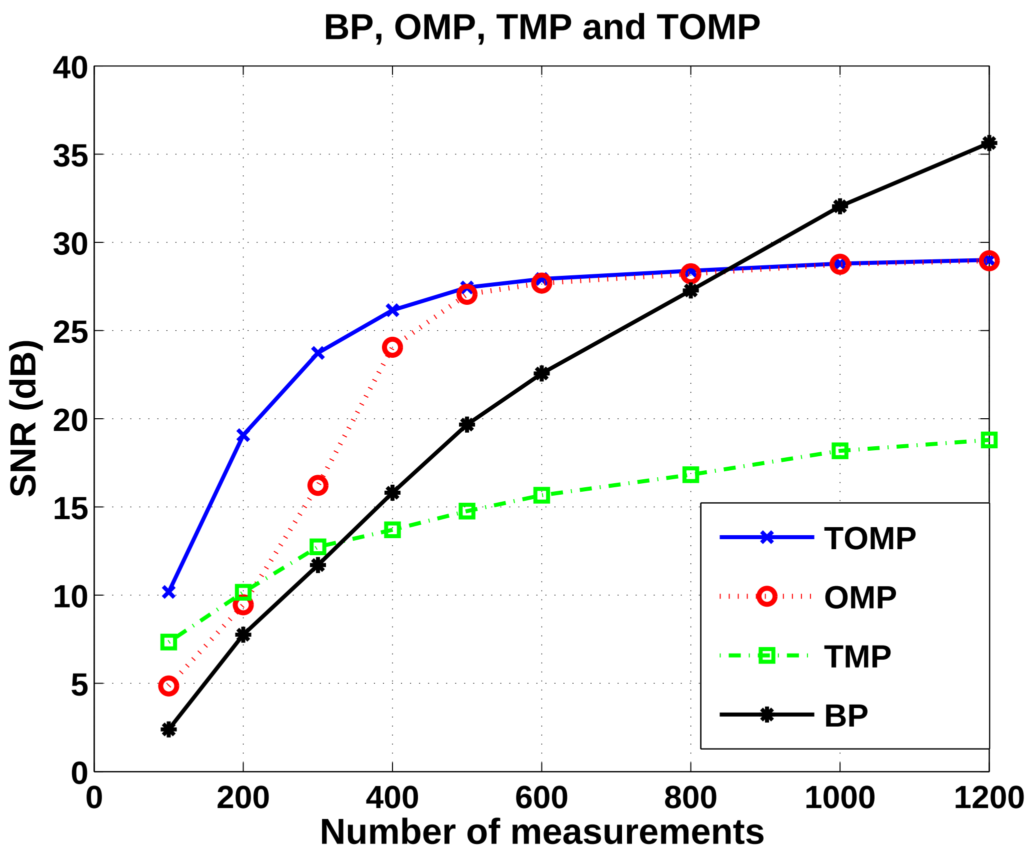
(a) Original (b) OMP (c) BP (d) TOMP



(e) Original (f) OMP (g) BP (h) TOMP

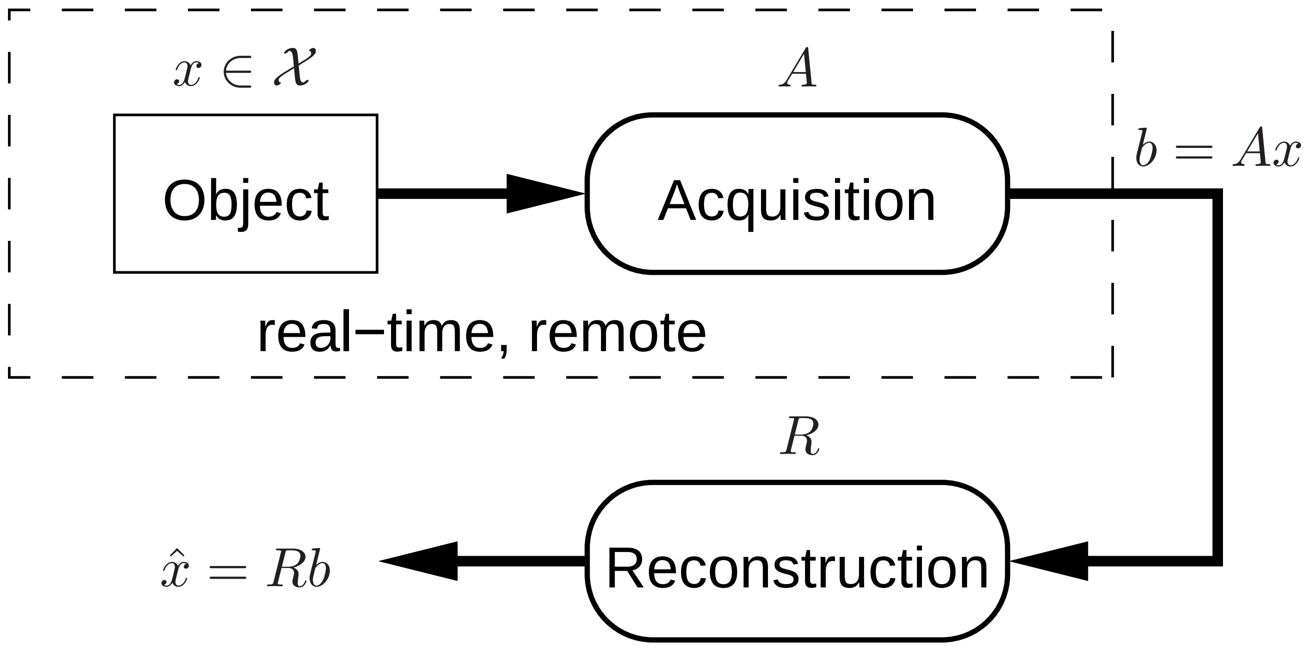
Experiment: Average Signal-to-Noise Ratios

Average reconstruction SNRs: length-4096 piecewise smooth signal, 8-level Db4 wavelet, random i.i.d. Gaussian measurement matrices.



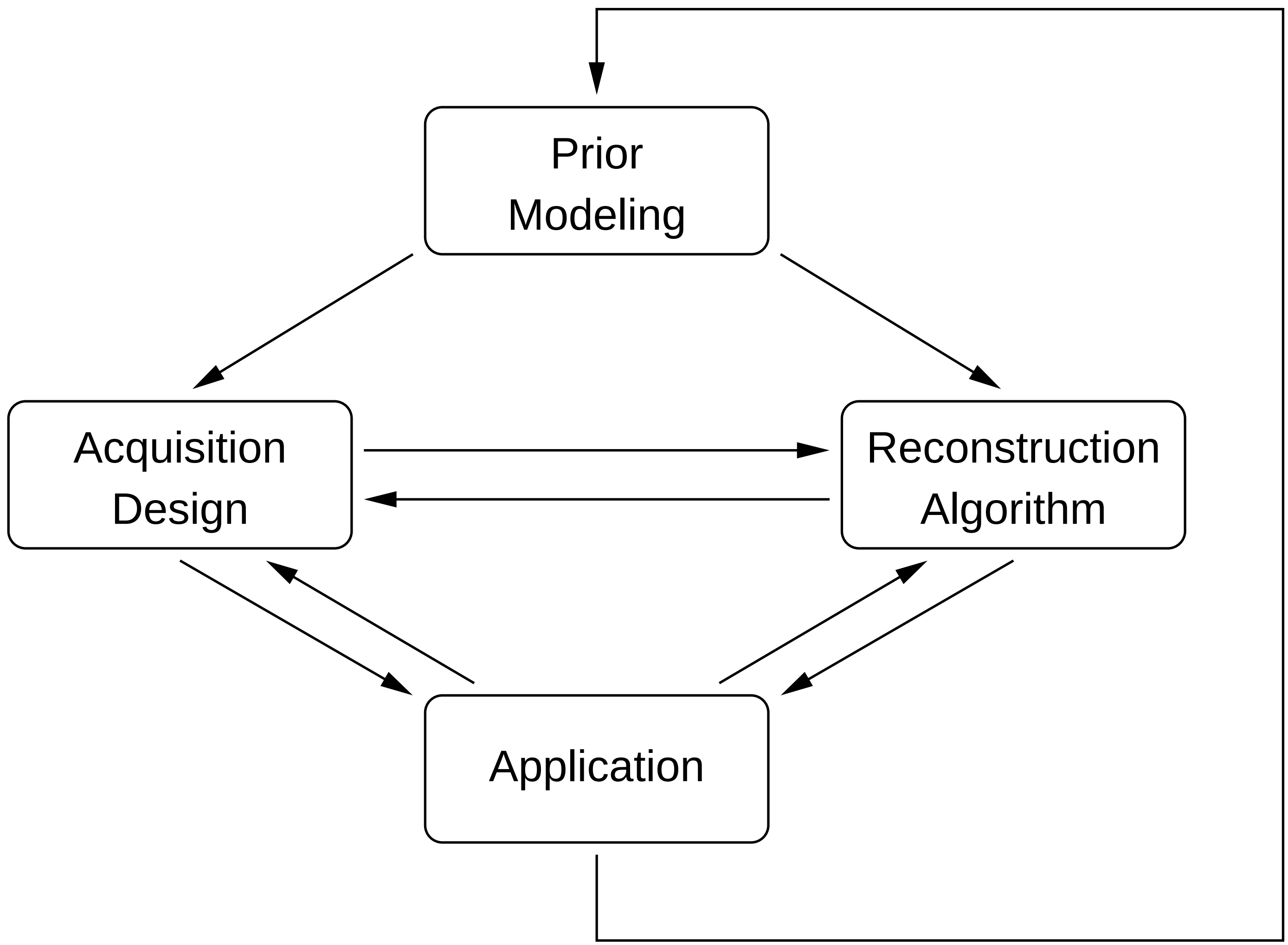
When number of measurements is limited, TOMP can gain around 7-10 dB over existing methods.

The Sensing/Sampling Problem

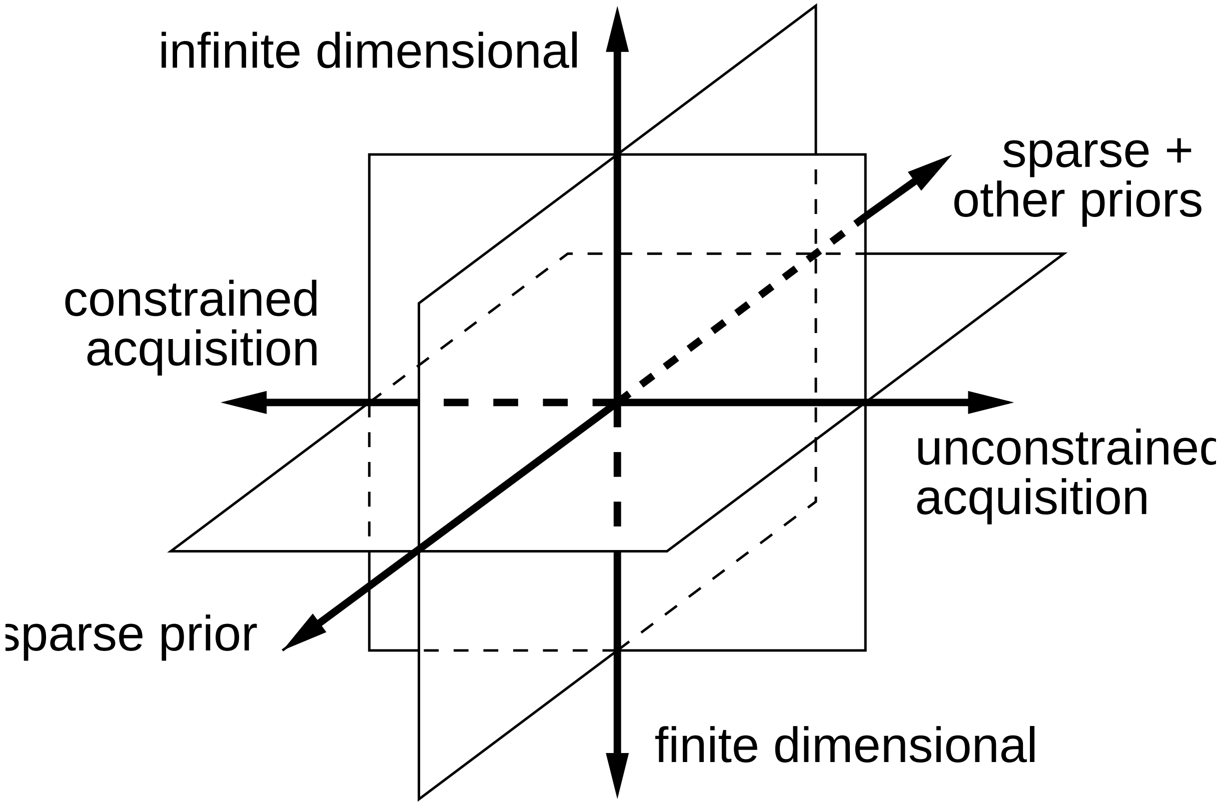


- Sensing = Sampling = Representing objects with a sequence of numbers.
- **Requirements:** length(b) is limited.
- **Goal:** use the prior information that $x \in \mathcal{X}$ to construct A and R.

Open Research Problems



Open Research Problems



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.