Linear measurements

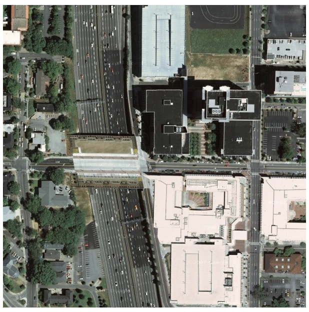
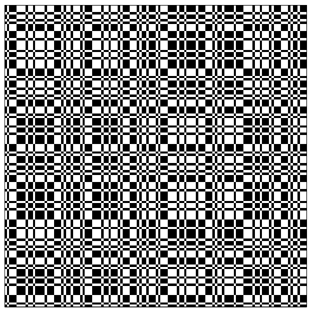
- Instead of samples, take *linear measurements* of signal/image x_0

$$y_1 = \langle x_0, \phi_1 \rangle, \quad y_2 = \langle x_0, \phi_2 \rangle, \quad \ldots, y_M = \langle x_0, \phi_K \rangle$$

$$y = \Phi x_0$$

- Equivalent to transform-domain sampling,
  $\{\phi_m\}$ = basis functions

- Example: coded imaging

$$y_m = \Big\langle$$  ,  $$\Big\rangle$$

Linear measurements

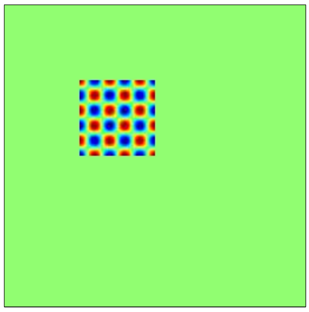
- Instead of samples, take *linear measurements* of signal/image x_0

$$y_1 = \langle x_0, \phi_1 \rangle, \quad y_2 = \langle x_0, \phi_2 \rangle, \quad \ldots, y_M = \langle x_0, \phi_K \rangle$$

$$y = \Phi x_0$$

- Equivalent to transform-domain sampling,
  $\{\phi_m\}$ = basis functions

- Example: DCT ?

$y_m = \Big\langle$  ,  $\Big\rangle$

Linear measurements

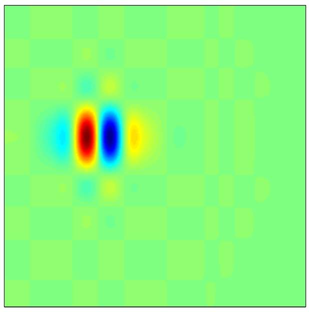
- Instead of samples, take *linear measurements* of signal/image x_0

$$y_1 = \langle x_0, \phi_1 \rangle, \quad y_2 = \langle x_0, \phi_2 \rangle, \quad \ldots, y_M = \langle x_0, \phi_K \rangle$$

$$y = \Phi x_0$$

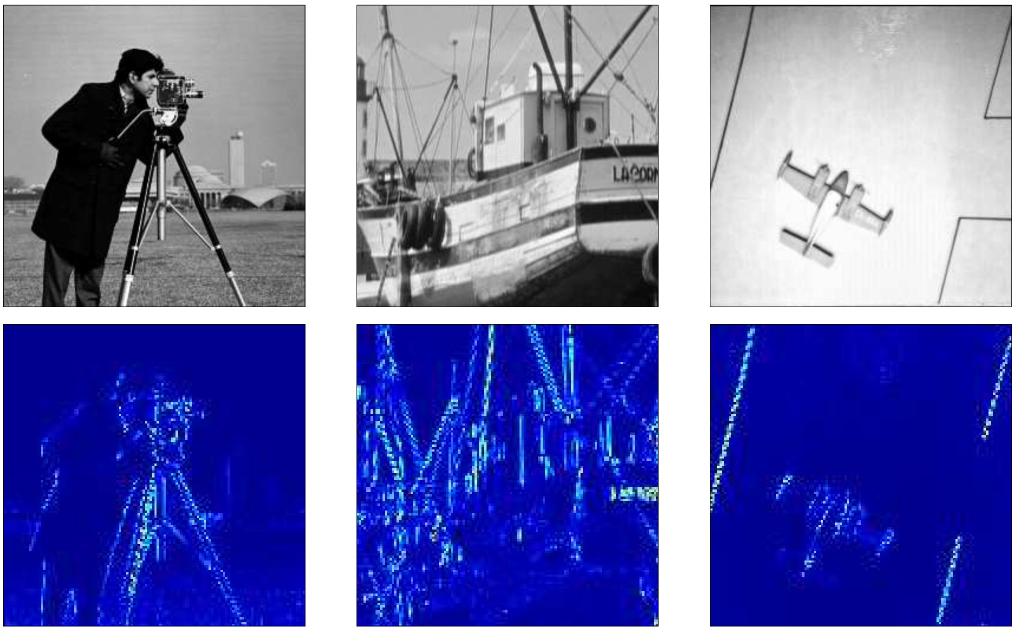
- Equivalent to transform-domain sampling,
  $\{\phi_m\}$ = basis functions

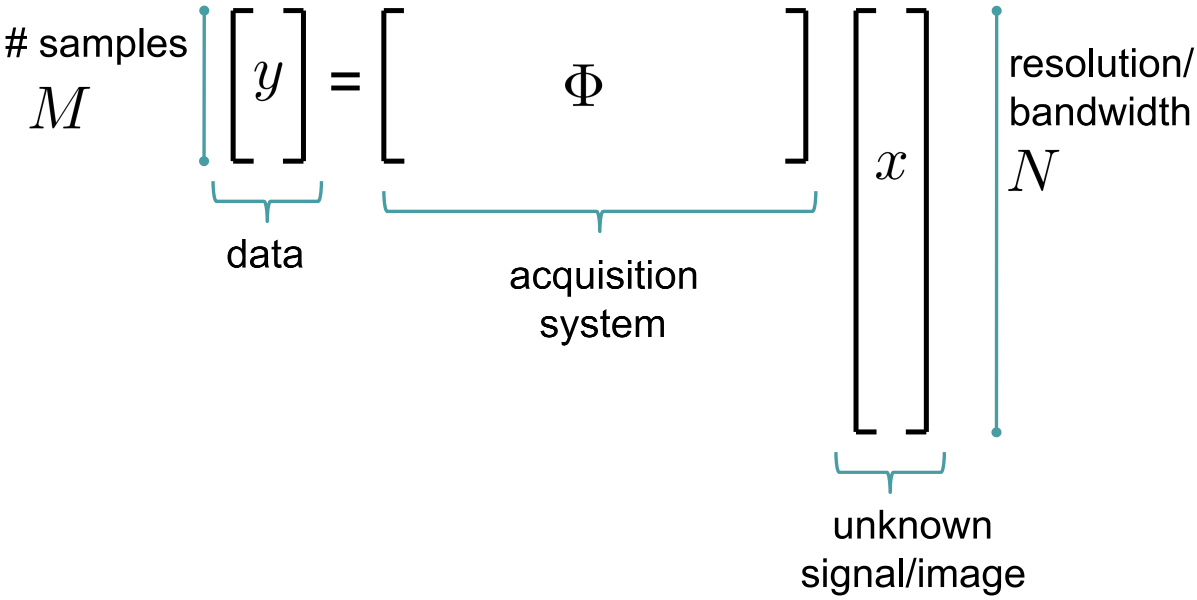
- Example: wavelets ?

$y_m = \Big\langle$  ,  $\Big\rangle$

Sparsity and Linear Measurements

- Since x_0 is sparse in Ψ, why don't we measure $\langle x_0, \psi_k \rangle$?
  Why not sample images in the wavelet domain?
- We'd love to sample wavelet coeffs, but which ones?





- If x is *sparse* and Φ is *diverse*, then these systems can be "inverted"

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_x \|y - Ax\|_2^2 \quad \Leftrightarrow \quad \hat{x} = (A^T A)^{-1} A^T y$$

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_x \|y - Ax\|_2^2 \quad \Leftrightarrow \quad \hat{x} = (A^T A)^{-1} A^T y$$

- Q: When is this recovery stable? That is, when is

$$\|\hat{x} - x_0\|_2^2 \sim \|\text{noise}\|_2^2 \quad ?$$

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_x \|y - Ax\|_2^2 \quad \Leftrightarrow \quad \hat{x} = (A^T A)^{-1} A^T y$$

- Q: When is this recovery stable? That is, when is

$$\|\hat{x} - x_0\|_2^2 \sim \|\text{noise}\|_2^2 \quad ?$$

- A: When the matrix A is an *approximate isometry*...

$$\|Ax\|_2^2 \approx \|x\|_2^2 \quad \text{for all } x \in \mathbb{R}^N$$

  i.e. A preserves *lengths*

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_x \|y - Ax\|_2^2 \quad \Leftrightarrow \quad \hat{x} = (A^T A)^{-1} A^T y$$

- Q: When is this recovery stable? That is, when is

$$\|\hat{x} - x_0\|_2^2 \sim \|\text{noise}\|_2^2 \quad ?$$

- A: When the matrix A is an *approximate isometry*...

$$\|A(x_1 - x_2)\|_2^2 \approx \|x_1 - x_2\|_2^2 \quad \text{for all } x_1, x_2 \in \mathbb{R}^N$$

  i.e. A preserves *distances*

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_x \|y - Ax\|_2^2 \quad \Leftrightarrow \quad \hat{x} = (A^T A)^{-1} A^T y$$

- Q: When is this recovery stable? That is, when is

$$\|\hat{x} - x_0\|_2^2 \sim \|\text{noise}\|_2^2 \quad ?$$

- A: When the matrix A is an *approximate isometry*...

$$(1 - \delta) \leq \sigma_{\min}^2(A) \;\leq\; \sigma_{\max}^2(A) \leq (1 + \delta)$$

i.e. A has *clustered singular values*

Classical: When can we stably "invert" a matrix?

- Suppose we have an $M \times N$ observation matrix A with $M \geq N$ (MORE observations than unknowns), through which we observe

$$y = Ax_0 + \text{noise}$$

- Standard way to recover x_0, use the *pseudo-inverse*

$$\text{solve } \min_x \|y - Ax\|_2^2 \quad \Leftrightarrow \quad \hat{x} = (A^T A)^{-1} A^T y$$

- Q: When is this recovery stable? That is, when is

$$\|\hat{x} - x_0\|_2^2 \sim \|\text{noise}\|_2^2 \quad ?$$

- A: When the matrix A is an *approximate isometry*...

$$(1-\delta)\|x\|_2^2 \;\leq\; \|Ax\|_2^2 \;\leq\; (1+\delta)\|x\|_2^2$$

  for some $0 < \delta < 1$

When can we stably recover an S-sparse vector?

- Now we have an underdetermined $M \times N$ system Φ (FEWER measurements than unknowns), and observe

$$y = \Phi x_0 + \text{noise}$$

When can we stably recover an S-sparse vector?

- Now we have an underdetermined $M \times N$ system Φ (FEWER measurements than unknowns), and observe

$$y = \Phi x_0 + \text{noise}$$

- We can recover x_0 when Φ is a *keeps sparse signals separated*

$$(1-\delta)\|x_1 - x_2\|_2^2 \;\leq\; \|\Phi(x_1 - x_2)\|_2^2 \;\leq\; (1+\delta)\|x_1 - x_2\|_2^2$$

  for all S-sparse x_1, x_2

When can we stably recover an S-sparse vector?

- Now we have an underdetermined $M \times N$ system Φ (FEWER measurements than unknowns), and observe

$$y = \Phi x_0 + \text{noise}$$

- We can recover x_0 when Φ is a *restricted isometry (RIP)*

$$(1-\delta)\|x\|_2^2 \;\leq\; \|\Phi x\|_2^2 \;\leq\; (1+\delta)\|x\|_2^2 \quad \text{for all } 2S\text{-sparse } x$$

When can we stably recover an S-sparse vector?

- Now we have an underdetermined $M \times N$ system Φ (FEWER measurements than unknowns), and observe

$$y = \Phi x_0 + \text{noise}$$

- We can recover x_0 when Φ is a *restricted isometry (RIP)*

$$(1-\delta)\|x\|_2^2 \;\leq\; \|\Phi x\|_2^2 \;\leq\; (1+\delta)\|x\|_2^2 \quad \text{for all } 2S\text{-sparse } x$$

- To recover x_0, we solve

$$\min_x \; \|x\|_0 \quad \text{subject to} \quad \Phi x \approx y$$

$\|x\|_0 =$ number of nonzero terms in x

- This program is intractable

When can we stably recover an S-sparse vector?

- Now we have an underdetermined $M \times N$ system Φ (FEWER measurements than unknowns), and observe

$$y = \Phi x_0 + \text{noise}$$

- We can recover x_0 when Φ is a *restricted isometry (RIP)*

$$(1-\delta)\|x\|_2^2 \;\leq\; \|\Phi x\|_2^2 \;\leq\; (1+\delta)\|x\|_2^2 \quad \text{for all } 2S\text{-sparse } x$$

- A relaxed (convex) program

$$\min_x \; \|x\|_1 \quad \text{subject to} \quad \Phi x \approx y$$

$\|x\|_1 = \sum_k |x_k|$

- This program is very tractable (linear program)

Sparse recovery algorithms

- Given y, look for a sparse signal which is consistent.
- One method: ℓ_1 minimization (or *Basis Pursuit*)

$$\min_x \ \|\Psi[x]\|_1 \quad \text{s.t.} \quad \Phi x = y$$

  Ψ = sparsifying transform, Φ = measurement system
  (need RIP for $\Phi\Psi^T$)

  Convex (linear) program, can relax for robustness to noise

  Performance has theoretical guarantees

- Other recovery methods include greedy algorithms and iterative thresholding schemes

Stable recovery

- Despite its nonlinearity, sparse recovery is stable in the presence of
  - *modeling mismatch* (approximate sparsity), and
  - *measurement error*
- **Theorem** (Candès, R, Tao '06)

  If we observe $y = \Phi x_0 + e$, with $\|e\|_2 \leq \epsilon$, the solution $\hat{x}$ to

  $$\min_x \|\Psi[x]\|_1 \quad \text{s.t.} \quad \|y - \Phi x\|_2 \leq \epsilon$$

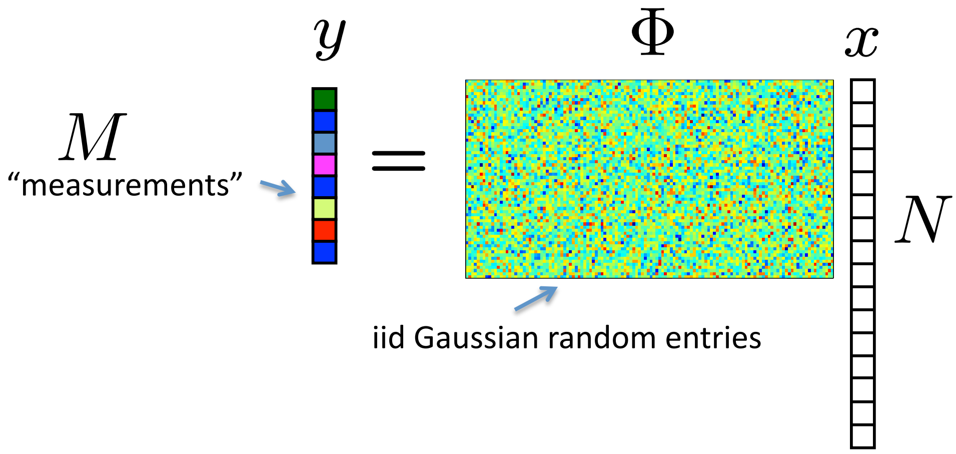
  will satisfy

  $$\|\hat{x} - x_0\|_2 \ \leq \ \mathrm{Const} \cdot \left(\epsilon + \frac{\|x_0 - x_{0,S}\|_1}{\sqrt{S}} \right)$$

  where
  - $x_{0,S}$ = S-term approximation of x_0
  - S is the largest value for which $\Phi\Psi^T$ satisfies the RIP
- Similar guarantees exist for other recovery algorithms
  - greedy (Needell and Tropp '08)
  - iterative thresholding (Blumensath and Davies '08)

What kind of matrices are restricted isometries?

- They are very hard to design, but they exist everywhere!



- For *any fixed* $x \in \mathbb{R}^N$, each measurement is

$$y_k \sim \mathrm{Normal}(0, \|x\|_2^2/M)$$