Primal Dual Pursuit
A Homotopy based Algorithm for the Dantzig Selector

Muhammad Salman Asif

Thesis Committee: *Justin Romberg (Advisor),*
James McClellan,
Russell Mersereau

School of Electrical and Computer Engineering,
Georgia Institute of Technology.

Problem statement

Measurement model

$$y = Ax + e$$

- x : n dimensional unknown signal.
- y : m dimensional measurement vector.
- A : $m \times n$ measurement matrix.
- e : error vector.

The Dantzig selector: A robust estimator for recovery of sparse signals from linear measurements.

$$\text{DS :} \quad \text{minimize } \|\tilde{x}\|_1 \quad \text{subject to} \quad \|A^T(A\tilde{x} - y)\|_\infty \leq \epsilon,$$

for some $\epsilon > 0$. This is convex and can be recast into an LP.

Primal Dual pursuit: a homotopy approach to solve the Dantzig selector.

Outline

- Background
  - Sparse representation
  - Compressed sensing
    - Introduction
    - ℓ_1 minimization
    - Random sensing
    - ℓ_1 estimators
- Primal dual pursuit
  - Dantzig selector primal dual formulation
  - Primal-dual homotopy
  - Main algorithm
    - Primal and dual update
    - Update directions
  - Numerical implementation
- S-step solution property
  - Optimality conditions
  - S-step solution property analysis
    - Random matrices
    - Incoherent ensemble
- Simulation results
- Conclusion & future work

Sparse Representation

- Signal/image $x(t)$ in time/spatial domain can be represented in some basis ψ as

$$x(t) = \sum_i \alpha_i \psi_i \quad \text{or} \quad x = \Psi\alpha$$

$$\psi_i = \text{basis function}$$

$$\alpha_i = \text{expansion coefficients in } \psi \text{ domain}$$

  e.g., sinusoids, wavelets, curvelets,. . .

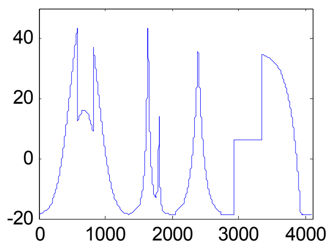
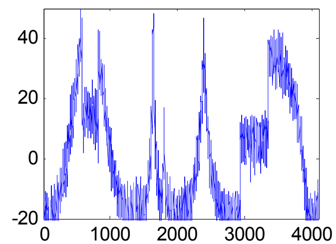
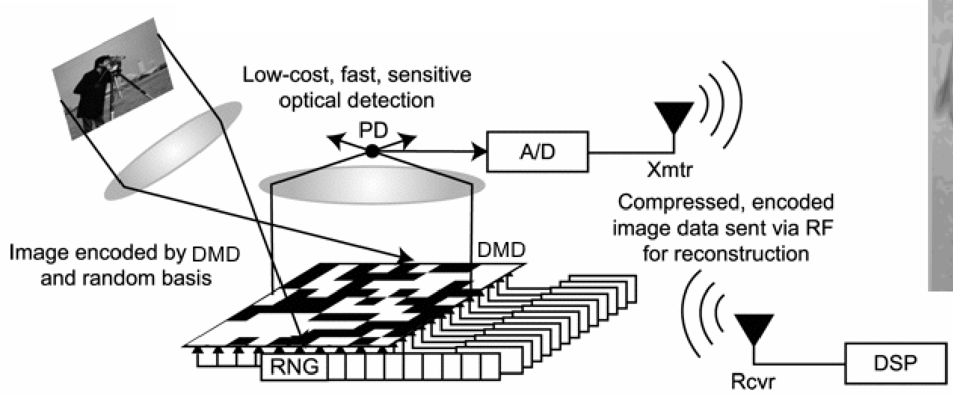
- Most signals of interest can be well-represented by a small number of transform coefficients in some appropriate basis.
- Magnitude of the transform coefficients decay rapidly, usually following some power law

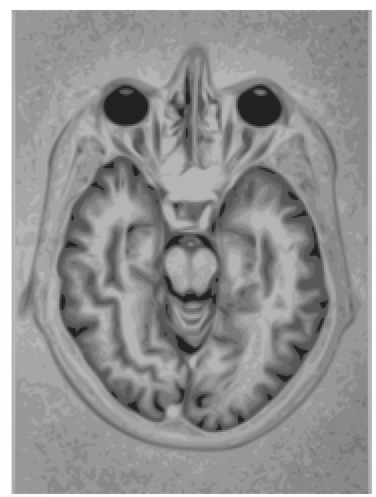
$$|\alpha|_{(k)} \sim k^{-r}, \quad \text{for some} \quad r > 0$$

Benefits of sparsity

Sparsity plays an important role in many signal processing applications such as

- Signal estimating in the presence of noise (thresholding)
- Inverse problems for signal reconstruction (tomography)
- Compression (transform coding)
- Signal acquisition (compressed sensing)



Best K-term approximation with 25k wavelet coeffs.

Compressed Sensing

- Data acquisition (Sampling) Shannon-Nyquist sampling theorem
  - Exact reconstruction of a band-limited signal possible if sampling rate is atleast twice the maximum frequency.
  - Reconstruction phenomenon is linear.
- Compression (Transform coding)
  - Transform signal/image in some suitable basis e.g., wavelets or discrete cosine (DCT)
  - Select few best coefficients without causing much perceptual loss.
  - Transfrom back to the canonical basis.

Sample a signal at or above Nyquist rate, transform into some sparsifying basis, adaptively encode a small portion out of it and throw away the rest. Seems very wasteful!

Is it possible to recover a signal by acquiring only as many nonadaptive samples as we will be keeping eventually? Answer is YES and lies in compressed sensing!

Compressed Sensing

"`entia non sunt multiplicanda praeter necessitatem` -- *entities must not be multiplied beyond necessity".*

-Occam's Razor (wiki)

"Consider projecting the points of your favorite sculpture first onto a plane and then onto a single line". This is the power of dimensionality reduction!

– [Achlioptas 2001]

Compressed Sensing Model

- Take m linear measurements of n dimensional signal x

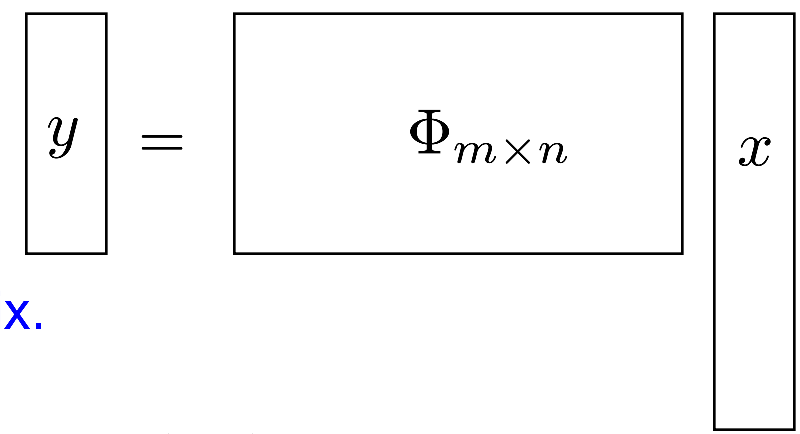
$$y_1 = \langle x, \varphi_1 \rangle, \; y_2 = \langle x, \varphi_2 \rangle, \; \ldots \; y_m = \langle x, \varphi_m \rangle \quad \text{or} \quad y = \Phi x$$

- Generalized sampling/sensing, equivalent to sampling in transform domain φ. Call φ_k - sensing functions.
- Choice of φ_k gives flexibility in acquisition
  - Dirac delta functions : conventional sampling.
  - Block indicator functions : pixels values collected from CCD arrays.
  - Sinusoids : Fourier measurements in MRI.

Compressed sensing model

$y = \Phi x$ where $m \ll n$.

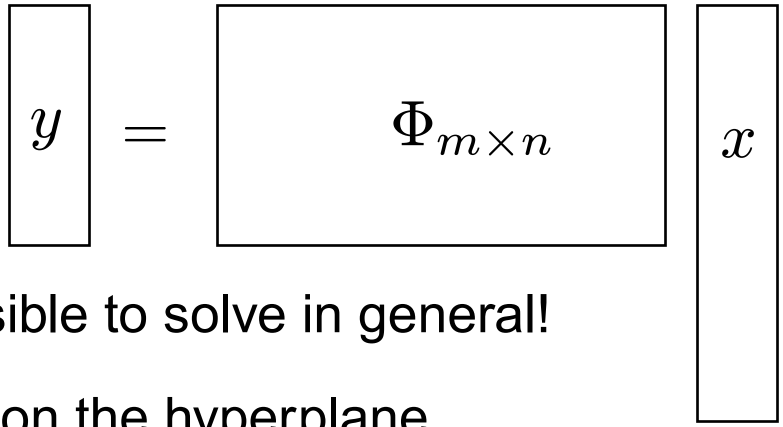
Φ is the measurement/sensing matrix.



Signal Reconstruction

Compressed sensing model

$y = \Phi x$ where $m \ll n$.



- Underdetermined system : Impossible to solve in general!
- Infinetely many possible solutions on the hyperplane

$$\mathcal{H} := \{\hat{x} : \Phi\hat{x} = y\} = \mathcal{N}(\Phi) + x$$

- However situation is different if x is sufficiently sparse and Φ obeys some incoherence properties.
- Combinatorial search : Find sparsest vector in hyperplane $\mathcal{H}$.

$$\text{P}_0 : \quad \text{minimize } \|\tilde{x}\|_0 \quad \text{subject to} \quad \Phi\tilde{x} = y$$

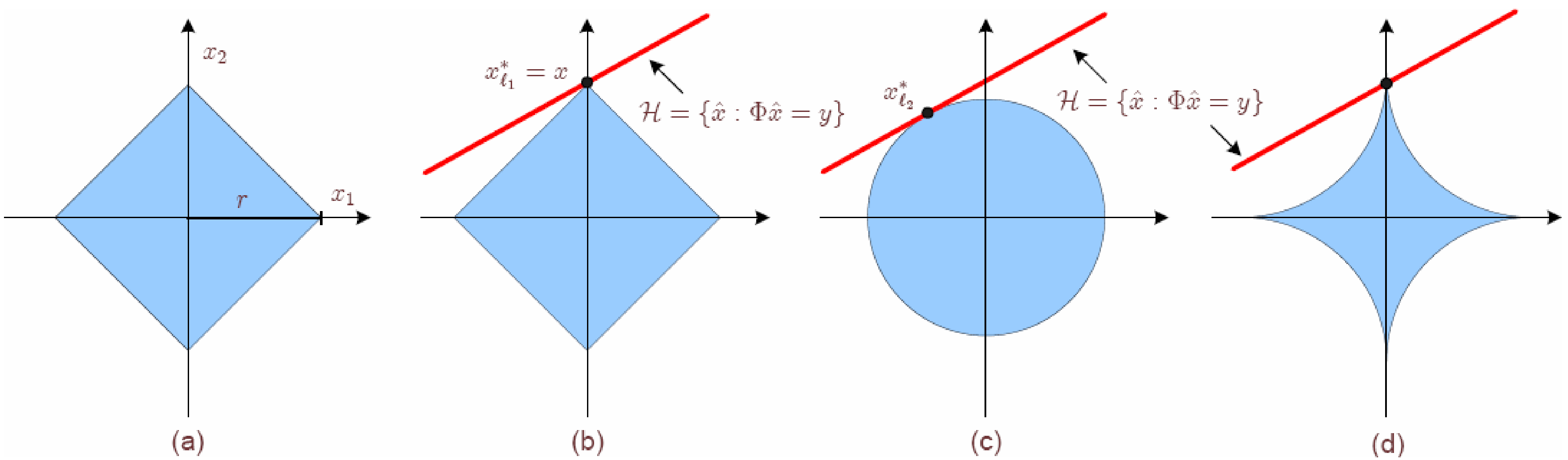
Combinatorial optimization problem, known to be NP-hard.

Convex Relaxation

- minimum ℓ_1 norm

$$\text{P}_1: \quad \text{minimize } \|\tilde{x}\|_1 \quad \text{subject to} \quad \Phi\tilde{x} = y$$

Can be recast into an LP for real data and SOCP for complex data.



Geometry of ℓ_1 minimization

Uniform Uncertainty Principle

- Uniform uncertainty principle (UUP) or Restricted Isometry Property

$$(1-\delta_S)\|c\|_2^2 \leq \|\Phi_T c\|_2^2 \leq (1+\delta_S)\|c\|_2^2 \qquad \text{(RIP)}$$

  - δ_S : S-restricted isometry constant.
  - Φ_T : columns of Φ indexed by set T.

- UUP requires Φ to obey (RIP) for every subset of columns T and coefficient sequence $\{c_j\}_{j\in T}$ such that $|T| \leq S$.

- This essentially tells that every subset of columns with cardinality less than S behaves almost like an orthonormal system.

- An equivalent form of Uniform uncertainty principle

$$\frac{1}{2}\frac{m}{n}\|c\|_2^2 \leq \|\Phi_T c\|_2^2 \leq \frac{3}{2}\frac{m}{n}\|c\|_2^2$$

where columns of Φ are not normalized, $\delta_S = \frac{1}{2}$.

Uniform Uncertainty Principle

- Uniform uncertainty principle (UUP) or Restricted Isometry Property

$$(1-\delta_S)\|c\|_2^2 \leq \|\Phi_T c\|_2^2 \leq (1+\delta_S)\|c\|_2^2 \qquad \text{(RIP)}$$

  - δ_S : S-restricted isometry constant.
  - Φ_T : columns of Φ indexed by set T.

- UUP requires Φ to obey (RIP) for every subset of columns T and coefficient sequence $\{c_j\}_{j\in T}$ such that $|T| \leq S$.
- This essentially tells that every subset of columns with cardinality less than S behaves almost like an orthonormal system.

Matrices which obey UUP

- Gaussian matrix $m \gtrsim S \cdot \log n$
- Bernoulli matrix $m \gtrsim S \cdot \log n$
- Partial Fourier matrix $m \gtrsim S \cdot \log^5 n$
- Incoherent ensemble $m \gtrsim \mu^2 S \cdot \log^5 n$

Stable Recovery

Compressed sensing model in the presence of noise

$$y = \Phi x + e,$$

ℓ_1 minimization with data fidelity constraints

- Quadratic constraints

$$\text{minimize } \|\tilde{x}\|_1 \quad \text{subject to} \quad \|\Phi\tilde{x} - y\|_2 \leq \epsilon$$

  Also known as Lasso in statistics community.

- Bounded residual correlation : The Dantzig selector

$$\text{minimize } \|\tilde{x}\|_1 \quad \text{subject to} \quad \|\Phi^*(\Phi\tilde{x} - y)\|_\infty \leq \epsilon,$$

  where ϵ is usually chosen close to $\sqrt{2\log n} \cdot \sigma$, assuming that entries in e are i.i.d. $N(0, \sigma^2)$.

Dantzig selector

- The Dantzig selector

$$\text{minimize } \|\tilde{x}\|_1 \quad \text{subject to} \quad \|\Phi^*(\Phi\tilde{x} - y)\|_\infty \le \epsilon,$$

where ϵ is usually chosen close to $\sqrt{2\log n} \cdot \sigma$, assuming that entries in e are i.i.d. $N(0, \sigma^2)$.

- If columns of Φ are unit normed then solution x^* to DS obeys

$$\|x - x^*\|_2^2 \le C \cdot 2\log n \cdot \left(\sigma^2 + \sum_{i=1}^{n} \min(x_i^2, \sigma^2)\right),$$

- Soft thresholding (ideal denoising) : estimate x from noisy samples $y = x + e$

$$\text{minimize } \|\tilde{x}\|_1 \quad \text{subject to} \quad \|y - \tilde{x}\|_\infty \le \lambda \cdot \sigma.$$

Primal Dual Pursuit

Dantzig Selector Primal and Dual

- System model

$$y = Ax + e$$

  - $x \in \mathbb{R}^n$: unknown signal
  - $y \in \mathbb{R}^m$: measurement vector
  - $A \in \mathbb{R}^{m \times n}$: measurement matrix
  - $e \in \mathbb{R}^m$: noise vector.

- Dantzig selector (Primal)

$$\text{minimize } \|\tilde{x}\|_1 \quad \text{subject to} \quad \|A^T(A\tilde{x} - y)\|_\infty \leq \epsilon,$$

- Dantzig selector (Dual)

$$\text{maximize } -(\epsilon\|\lambda\|_1 + \langle \lambda, A^T y \rangle) \quad \text{subject to} \quad \|A^T A\lambda\|_\infty \leq 1$$

where $\lambda \in \mathbb{R}^n$ is the dual vector.

Dantzig Selector Primal and Dual

- Strong duality tells us that at any primal-dual solution pair (x^*, λ^*) corresponding to certain ϵ

$$\|x^*\|_1 = -(\epsilon\|\lambda^*\|_1 + \langle\lambda^*, A^T y\rangle)$$

or equivalently

$$\|x^*\|_1 + \epsilon\|\lambda^*\|_1 = -\langle x^*, A^T A\lambda^*\rangle + \langle\lambda^*, A^T(Ax^* - y)\rangle. \qquad (1)$$

- Using (1), the complementary slackness and feasibility conditions for primal and dual problems we get the following four optimality conditions for the solution pair (x^*, λ^*)

K1. $A^T_{\Gamma_\lambda}(Ax^* - y) = \epsilon z_\lambda$

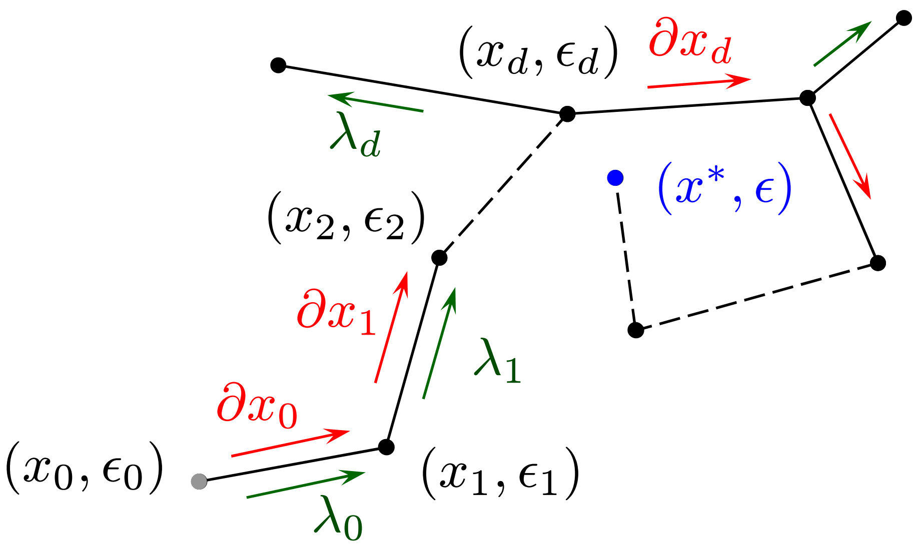
K2. $A^T_{\Gamma_x} A\lambda^* = -z_x$

K3. $|a^T_\gamma(Ax^* - y)| < \epsilon$ for all $\gamma \in \Gamma^c_\lambda$

K4. $|a^T_\gamma A\lambda^*| < 1$ for all $\gamma \in \Gamma^c_x$

Primal Dual Pursuit

- Homotopy Principle: Start from an artificial initial value and iteratively move towards the desired solution by gradually adjusting the homotopy parameter(s).
- In our formulation homotopy parameter is ϵ.
- Follow the path traced by sequence of primal-dual solution pairs (x_k, λ_k) for a range of ϵ_k as $\epsilon_k \downarrow \epsilon$, and consequently $x_k \to x^*$.



Primal Dual Pursuit

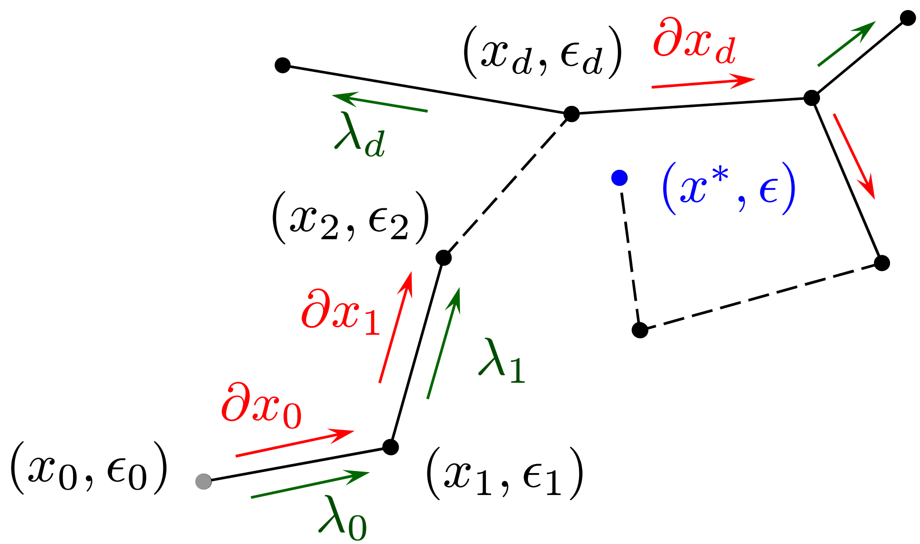
- Active primal constraints give us sign and support of dual vector λ
- Active dual constraints give us sign and support of primal vector x

$$A_{\Gamma_\lambda}^T (Ax_k - y) = \epsilon_k z_\lambda$$
$$A_{\Gamma_x}^T A\lambda_k = -z_x$$

And we keep track of supports Γ_x, Γ_λ and sign sequences z_x, z_λ all the time.

- Start at sufficiently large ϵ_k such that $x_k = 0$ and only one dual constraint is active (use $\epsilon_k = \|A^T y\|_\infty$).
- Move in the direction which reduces ϵ by most, until there is some change in the supports.
- Update the supports and sign sequences for primal and dual vectors and find new direction to move in.

Primal Dual Pursuit



- The homotopy path for x_k is piecewise linear and the kinks in this path represent some critical values of ϵ_k where primal and/or dual supports change.
- Either a new element enters the support or an element from within the support shrinks to zero.
- At any instant, the optimality conditions (K1-K4) must be obeyed by the primal-dual solution pair (x_k, λ_k).

Primal Dual Pursuit

At every kth step we have primal-dual vectors (x_k, λ_k), respective support $(\Gamma_x, \Gamma_\lambda)$ and sign sequence (z_x, z_λ).

We can divide each step into two main parts

- Primal update: Compute update direction ∂x and smallest step size δ such that either a new element enters Γ_λ or an existing element leaves Γ_x.
- Dual update: Compute update direction $\partial \lambda$ and smallest step size θ such that either a new element enters Γ_x or an existing element leaves Γ_λ.

Set

- $x_{k+1} = x_k + \delta \partial x$
- $\lambda_{k+1} = \lambda_k + \theta \partial \lambda$

and update primal-dual supports and sign sequences.

Primal Update

$$|a_\gamma^T(Ax_{k+1}-y)| = \epsilon_{k+1} \qquad \text{for all } \gamma \in \Gamma_\lambda$$

$$|a_\gamma^T(Ax_{k+1}-y)| \le \epsilon_{k+1} \qquad \text{for all } \gamma \in \Gamma_\lambda^c$$

$$|\underbrace{a_\gamma^T(Ax-y)}_{p_k(\gamma)} + \delta \underbrace{a_\gamma^T A\partial x}_{d_k(\gamma)}| \le \epsilon_k - \delta \qquad \text{for all } \gamma \in \Gamma_\lambda^c$$

$$|p_k(\gamma) + \delta d_k(\gamma)| \le \epsilon_k - \delta \qquad \text{for all } \gamma \in \Gamma_\lambda^c$$

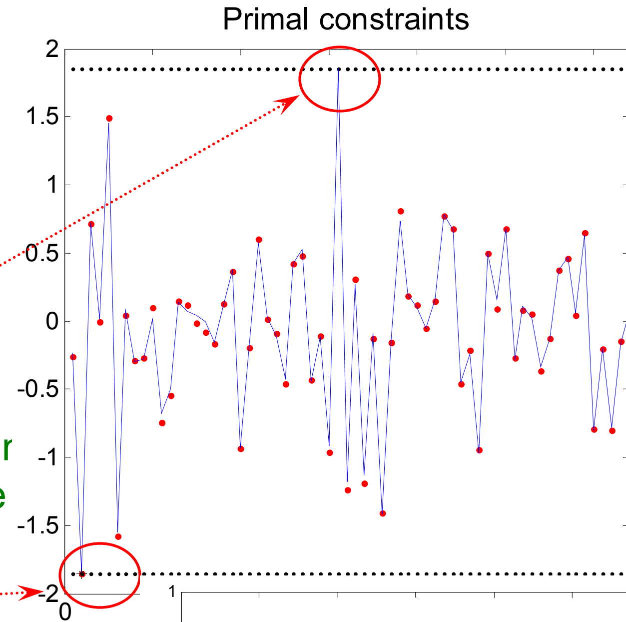
$$\delta^+ = \min_{i \in \Gamma_\lambda^c} \left(\frac{\epsilon_k - p_k(i)}{1 + d_k(i)}, \frac{\epsilon_k + p_k(i)}{1 - d_k(i)} \right)$$

$$i^+ = \arg\min_{i \in \Gamma_\lambda^c} \left(\frac{\epsilon_k - p_k(i)}{1 + d_k(i)}, \frac{\epsilon_k + p_k(i)}{1 - d_k(i)} \right)$$

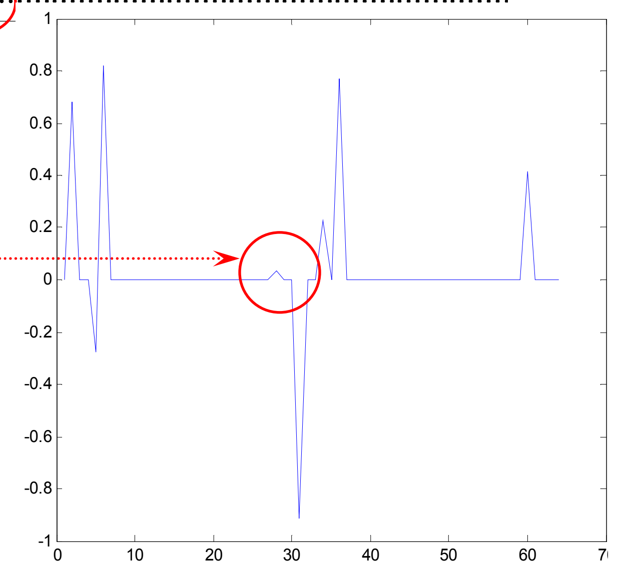
$$\delta^- = \min_{i \in \Gamma_x} \left(-\frac{x_k(i)}{\partial x(i)} \right)$$

$$i^- = \arg\min_{i \in \Gamma_x} \left(-\frac{x_k(i)}{\partial x(i)} \right)$$

$$\delta = \min(\delta^+, \delta^-)$$



New constrain
getting active
ϵ_k reduces



an element
from within Γ_x
shrinks to zero.

Dual Update

$$|a_\nu^T A\lambda_{k+1}| = 1 \qquad \text{for all } \nu \in \Gamma_x$$

$$|a_\nu^T A\lambda_{k+1}| \le 1 \qquad \text{for all } \nu \in \Gamma_x^c$$

$$|\underbrace{a_\nu^T A\lambda_k}_{a_k(\nu)} + \theta \underbrace{a_\nu^T A\partial\lambda}_{b_k(\nu)}| \le 1 \qquad \text{for all } \nu \in \Gamma_x^c$$

$$|a_k(\nu) + \theta b_k(\nu))| \le 1 \qquad \text{for all } \nu \in \Gamma_x^c$$

$$\theta^+ = \min_{j \in \Gamma_x^c} \left(\frac{1 - a_k(j)}{b_k(j)}, \frac{1 + a_k(j)}{-b_k(j)} \right)$$

$$j^+ = \arg\min_{j \in \Gamma_x^c} \left(\frac{1 - a_k(j)}{b_k(j)}, \frac{1 + a_k(j)}{-b_k(j)} \right)$$

$$\theta^- = \min_{j \in \Gamma_\lambda} \left(\frac{-\lambda(j)}{\partial\lambda(j)} \right)$$

$$j^- = \arg\min_{j \in \Gamma_\lambda} \left(\frac{-\lambda(j)}{\partial\lambda(j)} \right)$$

$$\theta = \min(\theta^+, \theta^-)$$

Update Directions

- Primal update direction

$$\partial x = \begin{cases} -(A_{\Gamma_\lambda}^T A_{\Gamma_x})^{-1} z_\lambda & \text{on } \Gamma_x \\ 0 & \text{elsewhere} \end{cases}$$

- Dual update direction

$$\partial \lambda = \begin{cases} -z_\gamma (A_{\Gamma_x}^T A_{\Gamma_\lambda})^{-1} A_{\Gamma_x}^T a_\gamma & \text{on } \Gamma_\lambda \\ z_\gamma & \text{on } \gamma \\ 0 & \text{elsewhere} \end{cases}$$

Why? $$A_{\Gamma_x}^T A(\lambda + \theta \partial \lambda) = -z_x$$
$$A_{\Gamma_x}^T A_{\Gamma_\lambda} \tilde{\partial \lambda} + A_{\Gamma_x}^T a_\gamma z_\gamma = 0.$$

Primal Dual Pursuit Algorithm

Primal update:

compute the primal update direction ∂x

compute p_k, d_k and δ

$x_{k+1} = x_k + \delta\partial x$

$\epsilon_{k+1} = \epsilon_k - \delta$

if $\delta = \delta^-$ **then**

$\Gamma_x \leftarrow \Gamma_x \setminus i^-$ {remove i^- from supp(x) and update Γ_x}

$\tilde{\Gamma}_\lambda = \Gamma_\lambda$ {store the current Γ_λ in a dummy variable}

$\Gamma_\lambda \leftarrow \Gamma_\lambda \setminus \gamma$ {select an index γ from supp(λ) and remove it from Γ_λ}

$z_\gamma = z_\lambda(\gamma)$ {treat γ as the new element to supp(λ)}

update z_x, z_λ {update sign sequences on updated supports}

else

$\tilde{\Gamma}_\lambda = \Gamma_\lambda \cup \{i^+\}$ {store i^+ but do not update Γ_λ}

$z_\lambda = \text{sign}[A^T_{\tilde{\Gamma}_\lambda}(Ax_{k+1} - y)]$ {update z_λ}

$\gamma = i^+$

$z_\gamma = z_\lambda(\gamma)$

end if

Primal Dual Pursuit Algorithm

Dual update:

compute the dual update direction $\partial\lambda$

compute a_k and b_k

if $\delta = \delta^-$ && $\text{sign}[a_k(i^-)] = \text{sign}[b_k(i^-)]$ **then**

  $\partial\lambda \leftarrow -\partial\lambda$ {a check needed due to uncertainty in sign}

  $b_k \leftarrow -b_k$ {flip the sign of $\partial\lambda$ and in turn b_k}

end if

compute θ

$\lambda_{k+1} = \lambda_k + \theta\partial\lambda$

if $\theta = \theta^-$ **then**

  $\Gamma_\lambda \leftarrow \tilde{\Gamma}_\lambda \setminus j^-$ {remove j^- from supp(λ) and update Γ_λ}

  update z_λ {update sign sequence on updated support}

else

  $\Gamma_x \leftarrow \Gamma_x \cup \{j^+\}$ {add j^+ to supp(x) and update Γ_x}

  $\Gamma_\lambda \leftarrow \tilde{\Gamma}_\lambda$ {set Γ_λ to supp(λ) determined in Primal update}

  $z_x = \text{sign}[A^T_{\Gamma_x} A\lambda_{k+1}]$ {update z_x}

end if

Numerical Implementation

- Main computational cost
  - Update direction $(\partial x, \partial \lambda)$.
  - Step size (δ, θ).

- No need to solve a new system at every step.

- Just update the most recent inverse matrix whenever supports change.
  - Matrix inversion lemma.
  - Rank one update.

$$\tilde{A}^T \tilde{B} = \begin{bmatrix} A^T & a^T \end{bmatrix} \begin{bmatrix} B \\ b \end{bmatrix} = \begin{bmatrix} A^T B & A^T b \\ a^T B & a^T b \end{bmatrix} =: \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Numerical Implementation

- Just update the most recent inverse matrix whenever supports change.
  - Matrix inversion lemma.
  - Rank one update.

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1} A_{12} S^{-1} A_{21} A_{11}^{-1} & -A_{11}^{-1} A_{12} S^{-1} \\ -S^{-1} A_{21} A_{11}^{-1} & S^{-1} \end{bmatrix},$$

where $S = A_{22} - A_{21} A_{11}^{-1} A_{12}$ is the Schur complement of A_{11}.

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix},$$

$$A_{11}^{-1} = Q_{11} - Q_{12} Q_{22}^{-1} Q_{21}.$$

Computational cost for one step is just few matrix vector multiplications.

S-step Solution

- S-sparse signal can be recovered in S primal-dual step !

$$y = Ax_0$$

- Random measurements

$$m \gtrsim S^2 \cdot \log n$$

Gaussian entries of A independently selected to be i.i.d. Gaussian $N(0, 1/m)$.
Bernoulli entries of A independently selected to be $\pm 1/\sqrt{m}$ with equal probability

- Incoherent measurements

$$S \leq \frac{1}{2}\left(1 + \frac{1}{M}\right)$$

where $M = \max_{i \neq j} |\langle a_i, a_j \rangle|$

Optimality Condition

K1. $A^T_{\Gamma_\lambda}(Ax^* - y) = \epsilon z_\lambda$

K2. $A^T_{\Gamma_x} A\lambda^* = -z_x$

K3. $\|A^T_{\Gamma^c_\lambda}(Ax^* - y)\|_\infty < \epsilon$

K4. $\|A^T_{\Gamma^c_x} A\lambda^*\|_\infty < 1$

$(x^*_\epsilon, \lambda^*)$ is a solution pair

for all $0 \leq \epsilon \leq \epsilon_{crit} := \min_{\gamma \in \Gamma} \left(-\frac{x_0(\gamma)}{\lambda(\gamma)} \right)$

$\Gamma := \Gamma_x$
$z := z_x$

$$\lambda^* = \begin{cases} -(A^T_\Gamma A_\Gamma)^{-1} z & \text{on } \Gamma \\ 0 & \text{elsewhere} \end{cases}$$

$\lambda^* = -\hat{\partial x}$

set $x^*_\epsilon = x_0 + \epsilon\lambda^*$

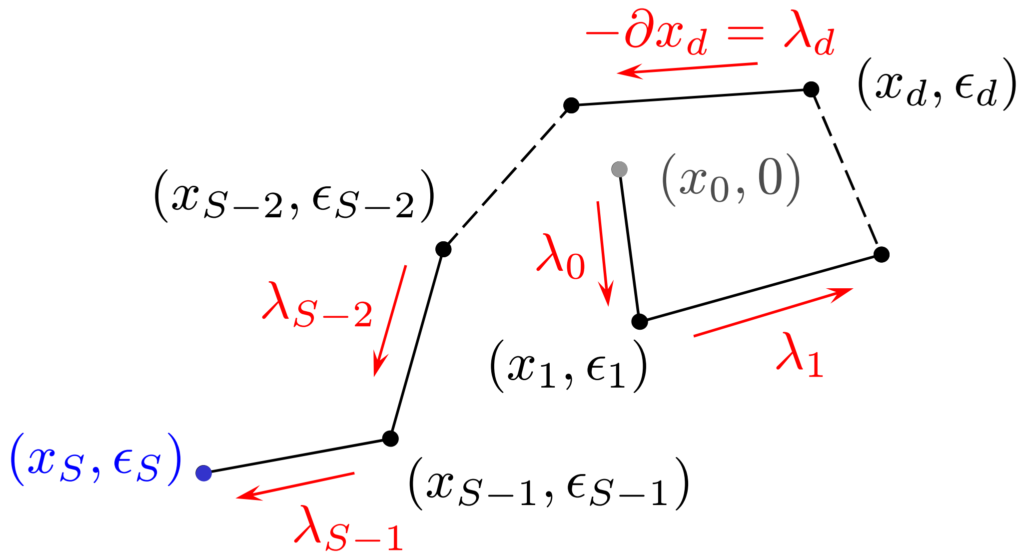
H1. A_Γ is full rank.

H2. $\|A^T_{\Gamma^c} A_\Gamma (A^T_\Gamma A_\Gamma)^{-1} z\|_\infty < 1$

H3. sign$[(A^T_\Gamma A_\Gamma)^{-1} z] = z$

$\Gamma_\lambda = \Gamma_x$
$z_\lambda = -z_x$

S step Solution



- Trace the path backwards, starting from exact solution x_0.
- S step solution property holds if $x_S = 0$. Means all the elements are removed from the support in S steps.
- Only if conditions (H1-H3) hold at every step.

Dantzig Shrinkability

1. $k = 0$, $\Gamma_0 =$ supp x_0, and $z_0 = \text{sign}(x_0\big|_{\Gamma_0})$.
2. If $x_k = 0$, return `Success`.
3. Check that

$$\|A^T_{\Gamma_k^c} A_{\Gamma_k} (A^T_{\Gamma_k} A_{\Gamma_k})^{-1} z\|_\infty < 1$$
$$\text{sign}[(A^T_{\Gamma_k} A_{\Gamma_k})^{-1} z] = z$$

   If either condition fails, break and return `Failure`.

4. Set $\lambda_k = \begin{cases} -(A^T_{\Gamma_k} A_{\Gamma_k})^{-1} z_k & \text{on} \quad \Gamma_k \\ 0 & \text{on} \quad \Gamma_k^c \end{cases},$

$$\epsilon_{k+1} = \min_{\gamma \in \Gamma_k} \left(\frac{x_k(\gamma)}{-\lambda_k(\gamma)} \right),$$
$$x_{k+1} = x_k + \epsilon_{k+1} \lambda_k,$$
$$\gamma'_{k+1} = \text{arg min}_{\gamma \in \Gamma_k} \left(\frac{x_k(\gamma)}{-\lambda_k(\gamma)} \right),$$
$$\Gamma_{k+1} = \Gamma_k \backslash \gamma'_{k+1},$$
$$z_{k+1} = z_k \text{ restricted to } \Gamma_{k+1}.$$

5. Set $k \leftarrow k + 1$, and return to step 2.

Sufficient Conditions for S step Solution

- A_Γ be full rank. (H1)
- Let $G = I - A_\Gamma^T A_\Gamma$, then (H2-H3) will be satisfied if $\|G\| < 1$ and

$$\max_{\gamma \in \{1,\ldots,n\}} |\langle (A_\Gamma^T A_\Gamma)^{-1} Y_\gamma, z \rangle| < 1, \qquad (2)$$

with

$$Y_\gamma = \begin{cases} A_\Gamma^T a_\gamma & \gamma \in \Gamma^c \\ A_\Gamma^T a_\gamma - \mathbf{1}_\gamma & \gamma \in \Gamma \end{cases},$$

Condition H3 !

If $\|G\| < 1$, we can write $(A_\Gamma^T A_\Gamma)^{-1} z$ in the following way

$$(A_\Gamma^T A_\Gamma)^{-1} z = (I - G)^{-1} z = \sum_{\ell=0}^{\infty} G^\ell z = \left(z + \sum_{\ell=1}^{\infty} G^\ell z \right),$$

Neumann series

condition (H3) will be satisfied if

$$\left\| \sum_{\ell=1}^{\infty} G^\ell z \right\|_\infty < 1.$$

$$\begin{aligned}
\max_{\gamma \in \Gamma} \left| \langle \mathbf{1}_\gamma, \sum_{\ell=1}^{\infty} G^\ell z \rangle \right| &= \max_{\gamma \in \Gamma} \left| \langle \sum_{\ell=1}^{\infty} G^\ell \mathbf{1}_\gamma, z \rangle \right| \\
&= \max_{\gamma \in \Gamma} \left| \langle \sum_{\ell=1}^{\infty} G^{\ell-1} g_\gamma, z \rangle \right| \\
&= \max_{\gamma \in \Gamma} \left| \langle (A_\Gamma^T A_\Gamma)^{-1} g_\gamma, z \rangle \right|
\end{aligned}$$

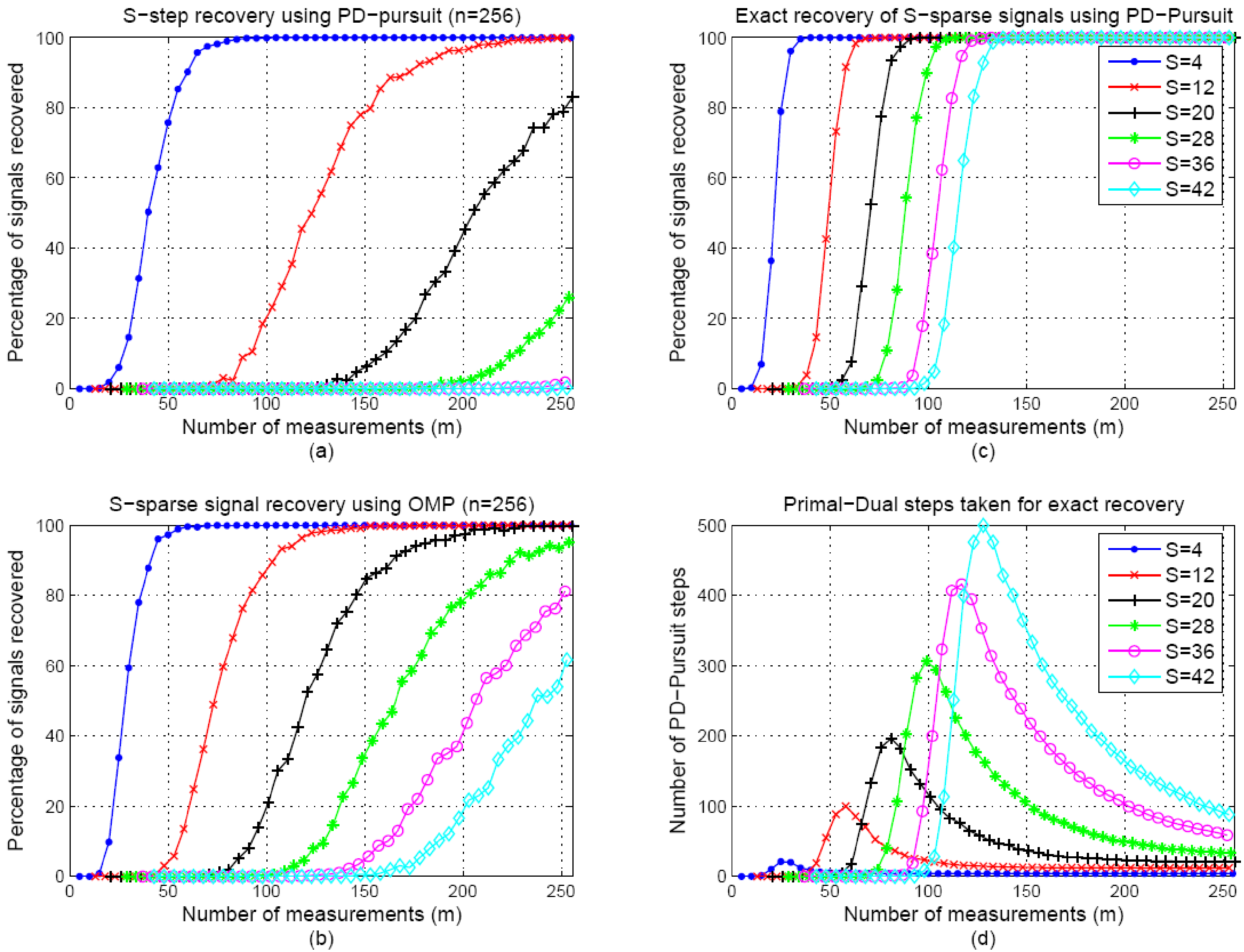
Outline of Proof for S-step property

- Bound the norm of Y_γ for all $\gamma \in \{1, \dots, n\}$
- Bound the norm of $w_\gamma := (A_\Gamma^T A_\Gamma^T)^{-1} Y_\gamma$
- Use Cauchy-Schwarz inequality to satisfy $|\langle w_\gamma, z \rangle| < 1$ for all γ

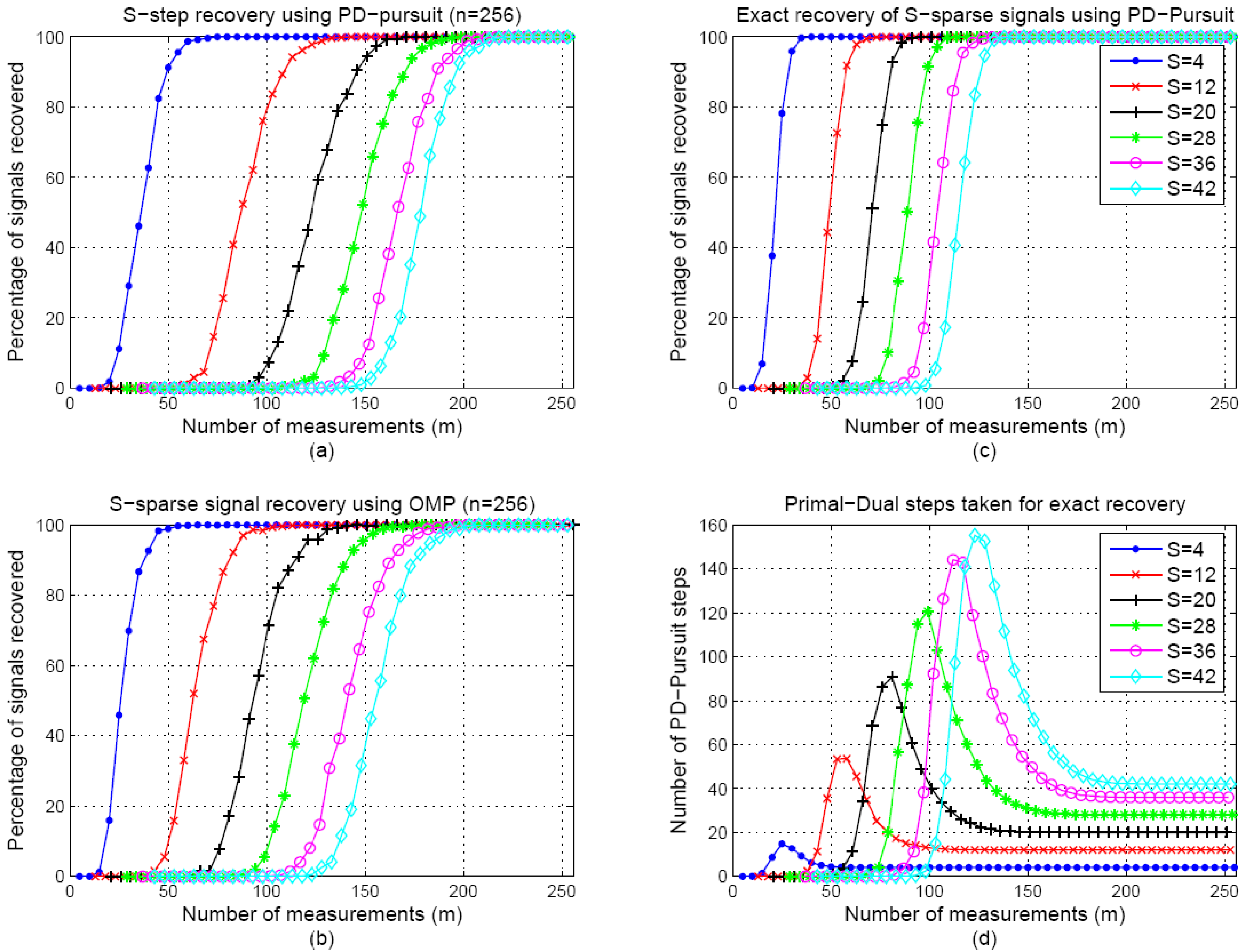
Random Matrices : Gaussian or Bernoulli

- Each entry of Y_γ is bounded by $C_\beta \sqrt{\frac{\log n}{m}}$ with probability exceeding $1 - O(n^{-\beta})$, for some constant $\beta > 0$. So $\|Y_\gamma\| < C_\beta \sqrt{\frac{S \log n}{m}}$ with same probability.
- Uniform uncertainty principle tells us that $\|(A_\Gamma^T A_\Gamma)^{-1}\| < 2$ with overwhelming high probability.
- Using Cauchy-Schwarz inequality, (2) is satisfied with probability exceeding $1 - O(n^{-\beta})$ if $m \geq C_\beta \cdot S^2 \log n$

Experimental Results (Gaussian)



Experimental Results (Ortho-Gaussian)



Lasso and Dantzig Selector

- Lasso

$$\underset{\tilde{x}}{\text{minimize}} \ \frac{1}{2}\|y - A\tilde{x}\|_2^2 + \epsilon\|\tilde{x}\|_1 \qquad \text{(Lasso)}$$

- Optimality conditions

L1. $A_\Gamma^T(Ax^* - y) = -\epsilon z$

L2. $|a_\gamma^T(Ax^* - y)| < \epsilon$ for all $\gamma \in \Gamma^c$

$$\partial x^{\text{Lasso}}\big|_\Gamma = (A_\Gamma^T A_\Gamma)^{-1} z \qquad \text{(Lasso update)}$$

$$\partial x^{\text{DS}}\big|_{\Gamma_x} = -(A_{\Gamma_\lambda}^T A_{\Gamma_x})^{-1} z_\lambda \qquad \text{(DS update)}$$

Future Work

- Better bound on required number of measurements!

$$S^2 \cdot \log n \quad \xrightarrow{?} \quad S \cdot \log^{\alpha} n,$$

  for some small $\alpha > 0$.

- Investigate the effect of orthogonal rows in the S-step recovery.
- Dynamic update of measurements.
- Implementation for largescale problems.

Questions

Thankyou !

Muhammad Salman Asif
sasif@gatech.edu