Stochastic Iterative Greedy Algorithms for Sparse Reconstruction

Deanna Needell

Joint work with N. Nguyen, T. Woolf

Support from Alfred P. Sloan Foundation and NSF CAREER #1348721

CLAREMONT McKENNA COLLEGE

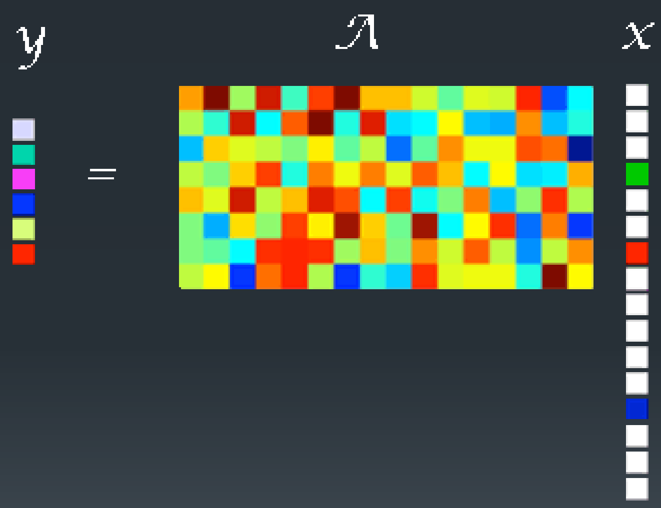
Sparse reconstruction

$$\min_x F(x) \qquad \text{subject to} \quad \|x\|_0 \leq k.$$

Sparse reconstruction

$$\min_{x} F(x) \qquad \text{subject to} \quad \|x\|_0 \leq k.$$

- Compressed sensing:



$$\min_{x \in \mathbb{R}^n} \frac{1}{2m} \|y - Ax\|_2^2 \qquad \text{subject to} \quad \|x\|_0 \leq k,$$

Sparse reconstruction

$$\min_x F(x) \qquad \text{subject to} \quad \|x\|_0 \leq k.$$

- Matrix recovery:

$$y_i = \langle A_i, X^\star \rangle + e_i$$

$$\min_{X \in \mathbb{R}^{n_1 \times n_2}} \frac{1}{2m} \|y - \mathcal{A}(X)\|_2^2 \quad \text{subject to} \quad \text{rank}(X) \leq k,$$

Sparse reconstruction

- Generalized sparse recovery:

$$x = \sum_{i=1}^{k} \alpha_i d_i, \qquad d_i \in \mathcal{D},$$

$$\|x\|_{0,\mathcal{D}} = \min_k \{k : x = \sum_{i \in T} \alpha_i d_i \quad \text{with} \quad |T| = k\}$$

$$\min_x \underbrace{\frac{1}{M} \sum_{i=1}^{M} f_i(x)}_{F(x)} \quad \text{subject to} \quad \|x\|_{0,\mathcal{D}} \leq k,$$

Methods

- Compressed sensing / matrix recovery
  - L1-minimization & nuclear norm minimization
  - Iterative methods (IHT, CoSaMP, OMP, …)
- Optimization
  - (Stochastic) gradient descent
  - (Stochastic) coordinate descent
  - …

Assumptions

Definition 1 ($\mathcal{D}$-restricted strong convexity ($\mathcal{D}$-RSC)). *The function $F(x)$ satisfies the $\mathcal{D}$-RSC if there exists a positive constant ρ_k^- such that*

$$F(x') - F(x) - \langle \nabla F(x), x' - x \rangle \geq \frac{\rho_k^-}{2} \|x' - x\|_2^2, \tag{16}$$

for all vectors x and x' of size n such that $|\mathrm{supp}_{\mathcal{D}}(x) \cup \mathrm{supp}_{\mathcal{D}}(x')| \leq k$.

Definition 2 ($\mathcal{D}$-restricted strong smoothness ($\mathcal{D}$-RSS)). *The function $f_i(x)$ satisfies the $\mathcal{D}$-RSS if there exists a positive constant $\rho_k^+(i)$ such that*

$$\|\nabla f_i(x') - \nabla f_i(x)\|_2 \leq \rho_k^+(i) \|x' - x\|_2 \tag{17}$$

for all vectors x and x' of size n such that $|\mathrm{supp}_{\mathcal{D}}(x) \cup \mathrm{supp}_{\mathcal{D}}(x')| \leq k$.

We may wish to consider blocks of the matrix A and break F(x) into functionals corresponding to each block i. Call the number of blocks b.

Restricted Isometry Property

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2 \quad \text{for all } k\text{-sparse vectors } x,$$

In particular, we require that the design matrix A satisfies

$$\frac{1}{m}\|Ax\|_2^2 \geq (1-\delta_k)\|x\|_2^2 \quad \frac{1}{b}\|A_{b_i}x\|_2^2 \leq (1+\delta_k)\|x\|_2^2$$

Here, $(1+\delta_k)$ and $(1-\delta_k)$ with $\delta_k \in (0,1]$ play the role of $\rho_k^+(i)$ and ρ_k^-

Stochastic greedy methods

Define $\text{approx}_k(x, \eta)$ as the operator that constructs a set Γ of cardinality k such that

$$\|\mathcal{P}_\Gamma x - x\|_2 \leq \eta \, \|x - x_k\|_2 \,,$$

Algorithm 1 StoIHT algorithm

input: k, γ, η, $p(i)$, and stopping criterion

initialize: x^0 and $t = 0$

repeat

- **randomize:** select an index i_t from $[M]$ with probability $p(i_t)$
- **proxy:** $b^t = x^t - \frac{\gamma}{Mp(i_t)} \nabla f_{i_t}(x^t)$
- **identify:** $\Gamma^t = \text{approx}_k(b^t, \eta)$
- **estimate:** $x^{t+1} = \mathcal{P}_{\Gamma^t}(b^t)$
- $t = t + 1$

until halting criterion *true*

output: $\hat{x} = x^t$

Stochastic greedy methods

Define $\text{approx}_k(x, \eta)$ as the operator that constructs a set Γ of cardinality k such that

$$\|\mathcal{P}_\Gamma x - x\|_2 \leq \eta \, \|x - x_k\|_2 \, ,$$

Algorithm 2 StoGradMP algorithm

input: k, η_1, η_2, $p(i)$, and stopping criterion
initialize: x^0, $\Lambda = 0$, and $t = 0$
repeat
  randomize: select an index i_t from $[M]$ with probability $p(i_t)$
  proxy: $r^t = \nabla f_{i_t}(x^t)$
  identify: $\Gamma = \text{approx}_{2k}(r^t, \eta_1)$
  merge: $\widehat{\Gamma} = \Gamma \cup \Lambda$
  estimate: $b^t = \text{argmin}_w \, F(x) \quad w \in \text{span}(D_{\widehat{\Gamma}})$
  prune: $\Lambda = \text{approx}_k(b^t, \eta_2)$
  update: $x^{t+1} = \mathcal{P}_\Lambda(b^t)$
  t = t+1
until halting criterion *true*
output: $\hat{x} = x^t$

Theoretical guarantees : StoIHT

Theorem 1. *Let $x^\star$ be a feasible solution of (5) and x^0 be the initial solution. At the $(t+1)$-th iteration of Algorithm 1, the expectation of the recovery error is bounded by*

$$\mathbb{E}\left\|x^{t+1} - x^\star\right\|_2 \leq \kappa^{t+1}\left\|x^0 - x^\star\right\|_2 + \frac{\sigma_{x^\star}}{(1-\kappa)} \tag{21}$$

Sparse signal recovery:

$$\mathbb{E}\left\|x^{t+1} - x^\star\right\|_2 \leq (3/4)^{t+1}\left\|x^\star\right\|_2 + c\sqrt{\frac{\sigma^2 k_0 \log n}{b}}.$$

Low-rank matrix recovery:

$$\mathbb{E}\left\|X^{t+1} - X^\star\right\|_F \leq (3/4)^{t+1}\left\|X^\star\right\|_F + c\left(\sqrt{\frac{\sigma^2 kn}{b}} + \sqrt{\frac{(\eta^2-1)\sigma^2 n^2}{b}}\right).$$

Empirical results

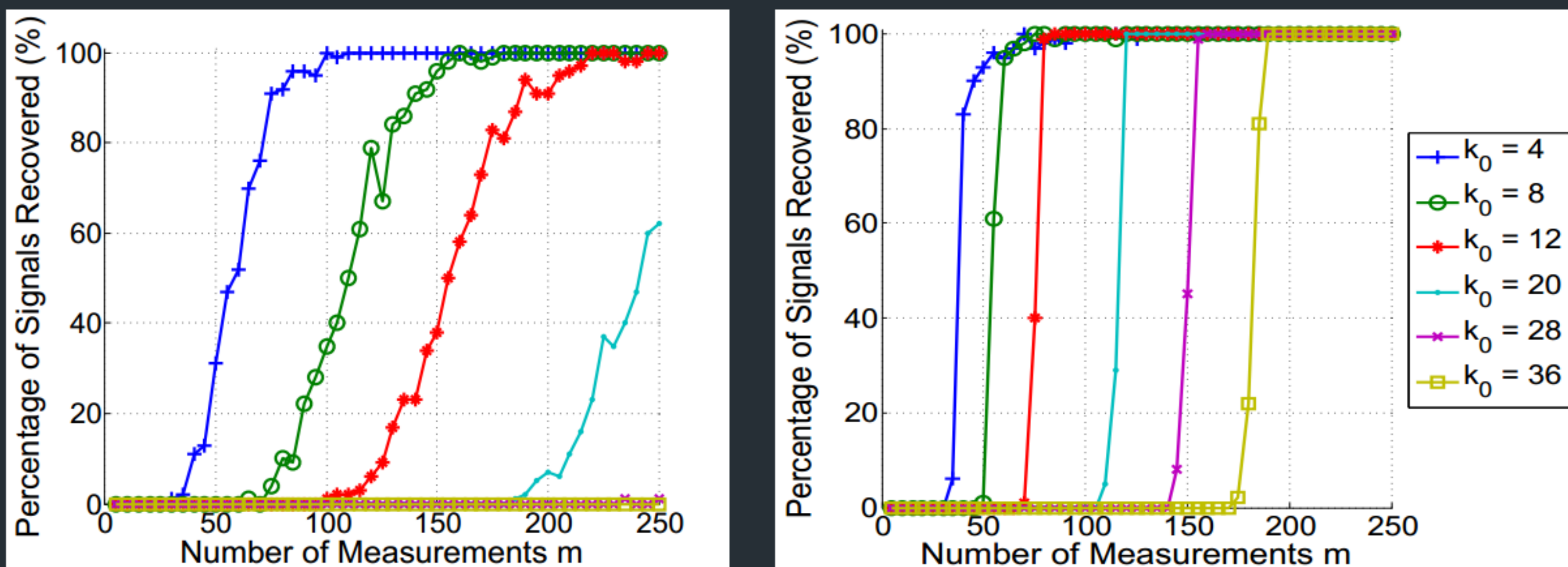


Figure 1: Sparse Vector Recovery: Percent recovery as a function of the number of measurements for IHT (left) and StoIHT (right) for various sparsity levels k_0.

Empirical results

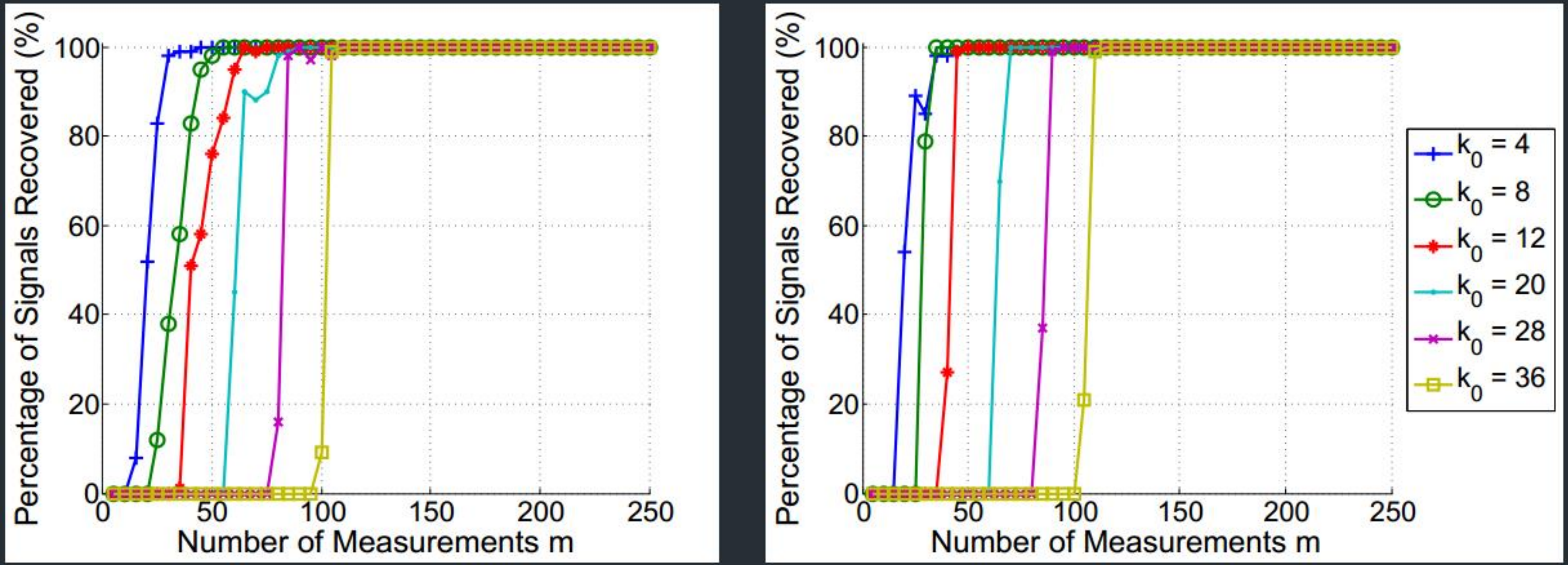


Figure 2: Sparse Vector Recovery: Percent recovery as a function of the number of measurements for GradMP (left) and StoGradMP (right) for various sparsity levels k_0.

Empirical results

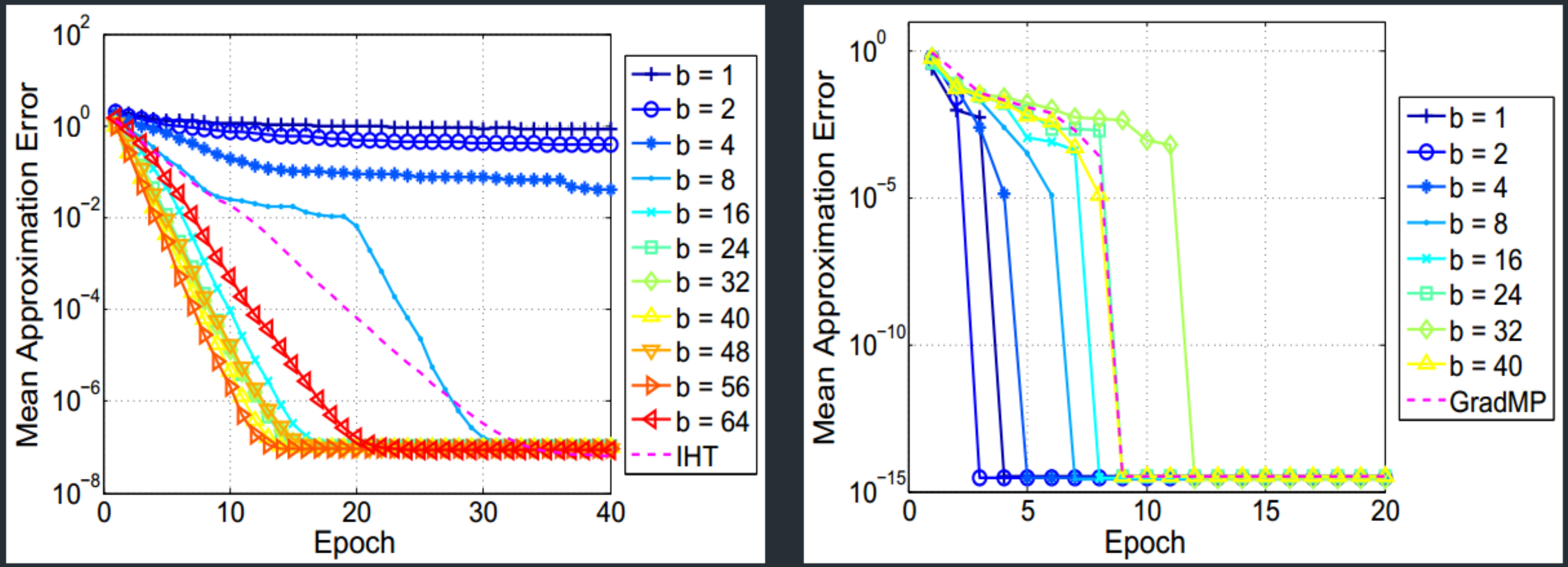


Figure 3: Sparse Vector Recovery: Recovery error as a function of epochs and various block sizes b for IHT methods (left) and GradMP methods (right).

Empirical results

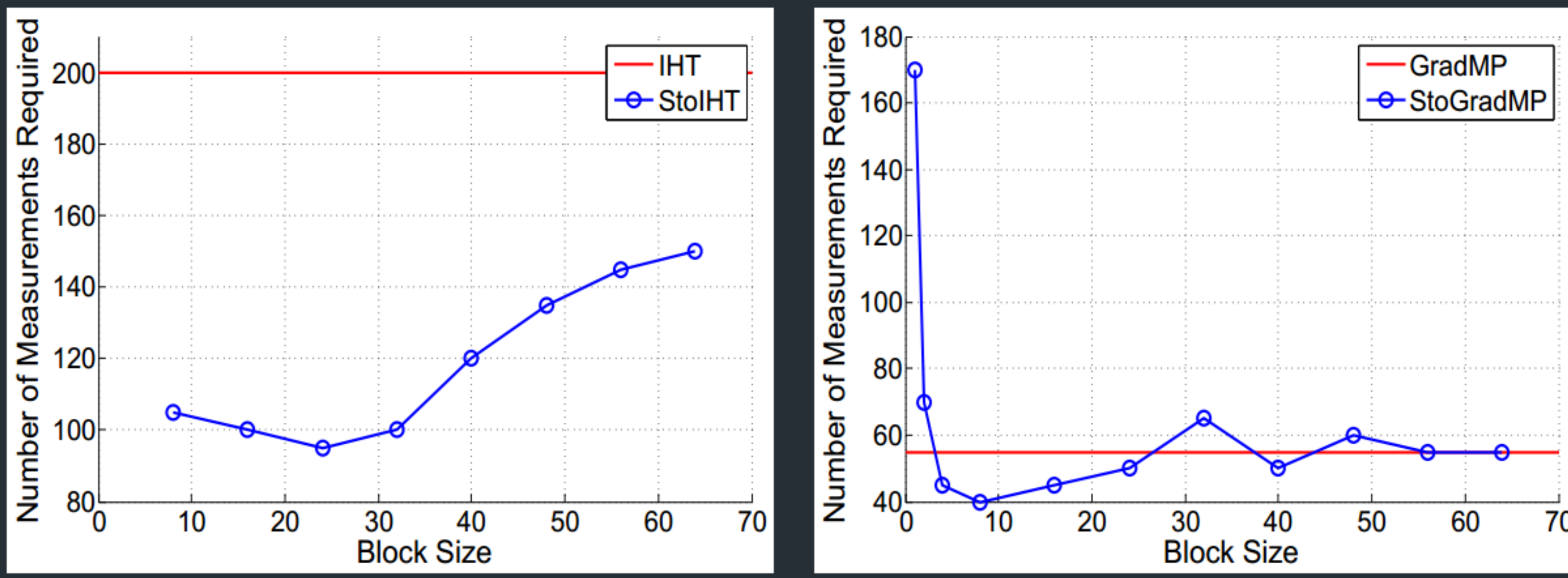
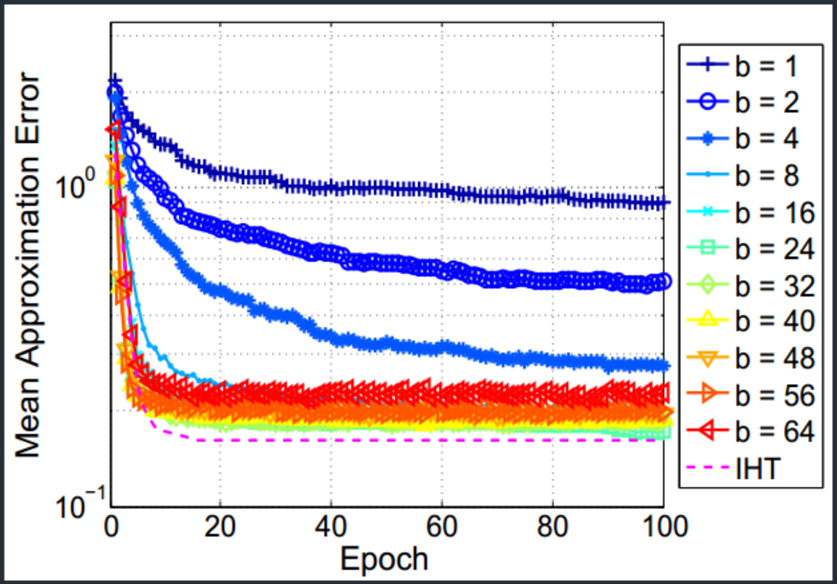


Figure 4: Sparse Vector Recovery: Number of measurements required for signal recovery as a function of block size (blue marker) for StoIHT (left) and StoGradMP (right). Number of measurements required for deterministic method shown as red solid line.

Empirical results



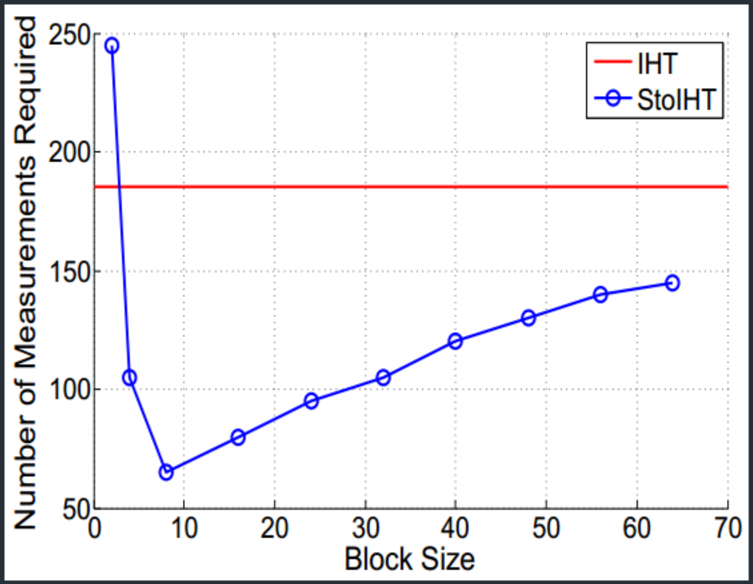


Figure 5: Sparse Vector Recovery: A comparison of IHT and StoIHT in the presence of noise. Recovery error versus epoch (left) and measurements required versus block size (right).

Empirical results

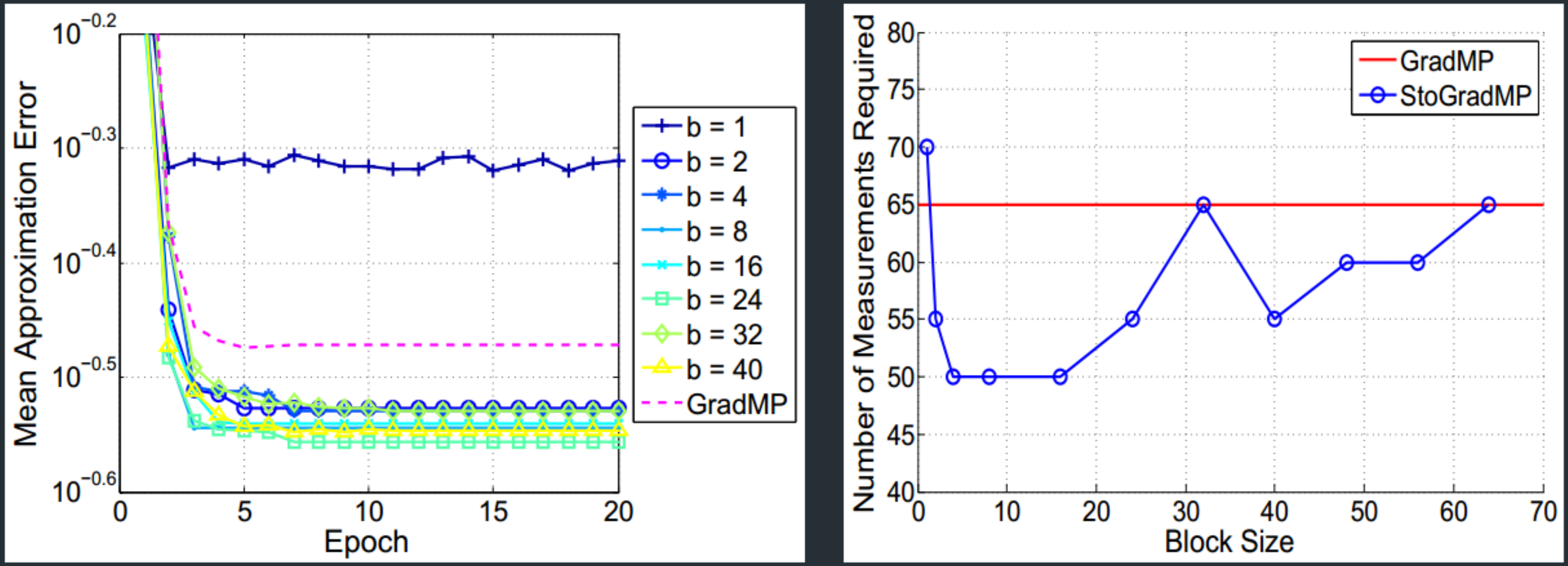


Figure 6: Sparse Vector Recovery: A comparison of GradMP and StoGradMP in the presence of noise. Recovery error versus epoch (left) and measurements required versus block size (right).

Empirical results

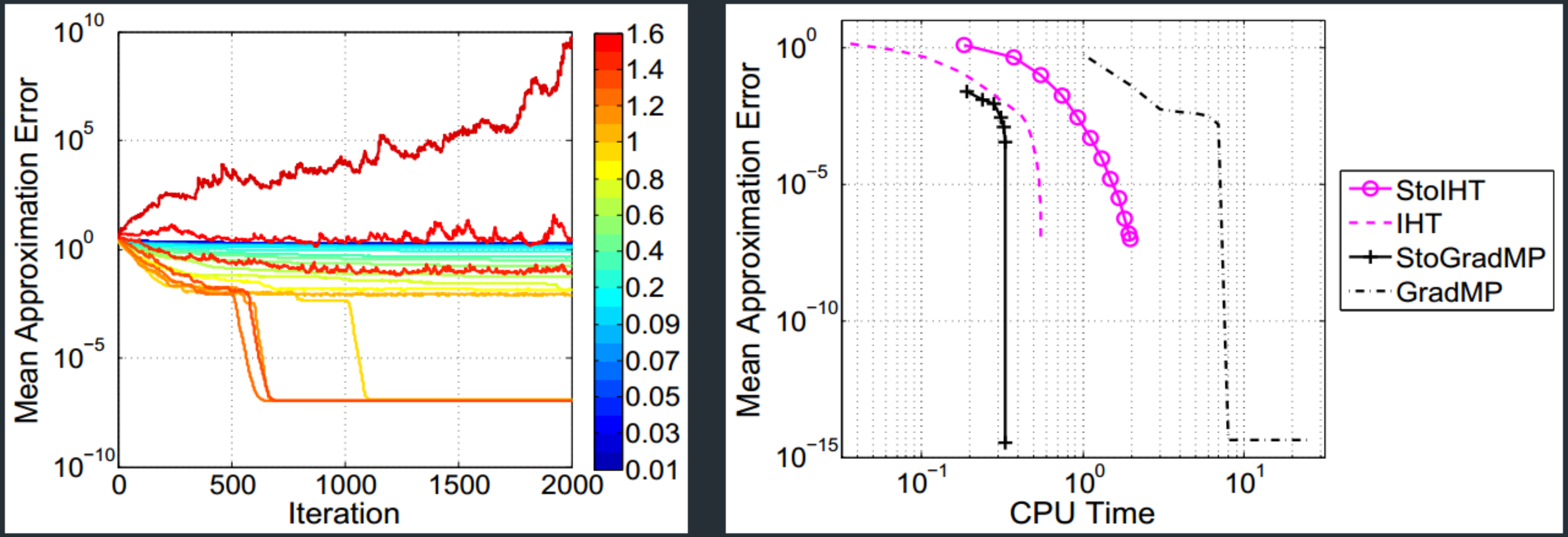


Figure 7: Sparse Vector Recovery: (Left) A comparison of StoIHT for various values of the step size γ (shown in the colorbar). (Right) A comparison of the mean recovery error as a function of runtime for IHT, StoIHT, GradMP, and StoGradMP.

Empirical results

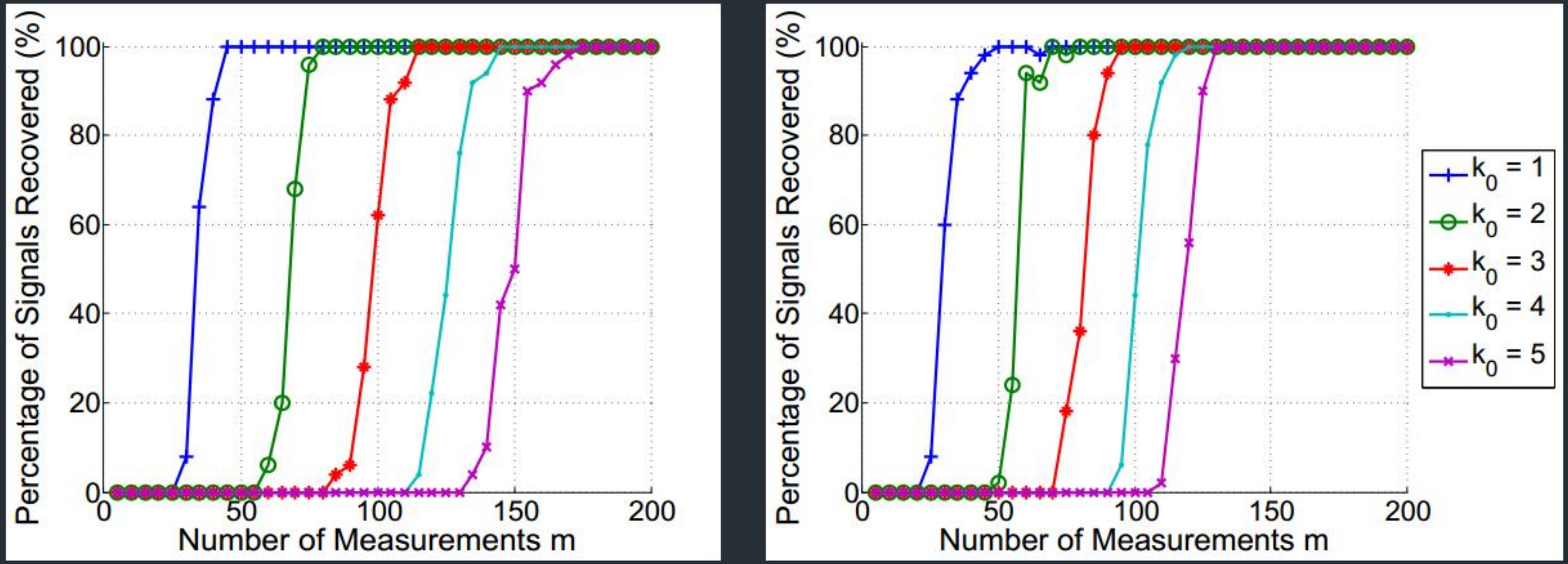


Figure 8: Low-Rank Matrix Recovery: Percent recovery as a function of the number of measurements for IHT (left) and StoIHT (right) for various rank levels k_0.

Empirical results

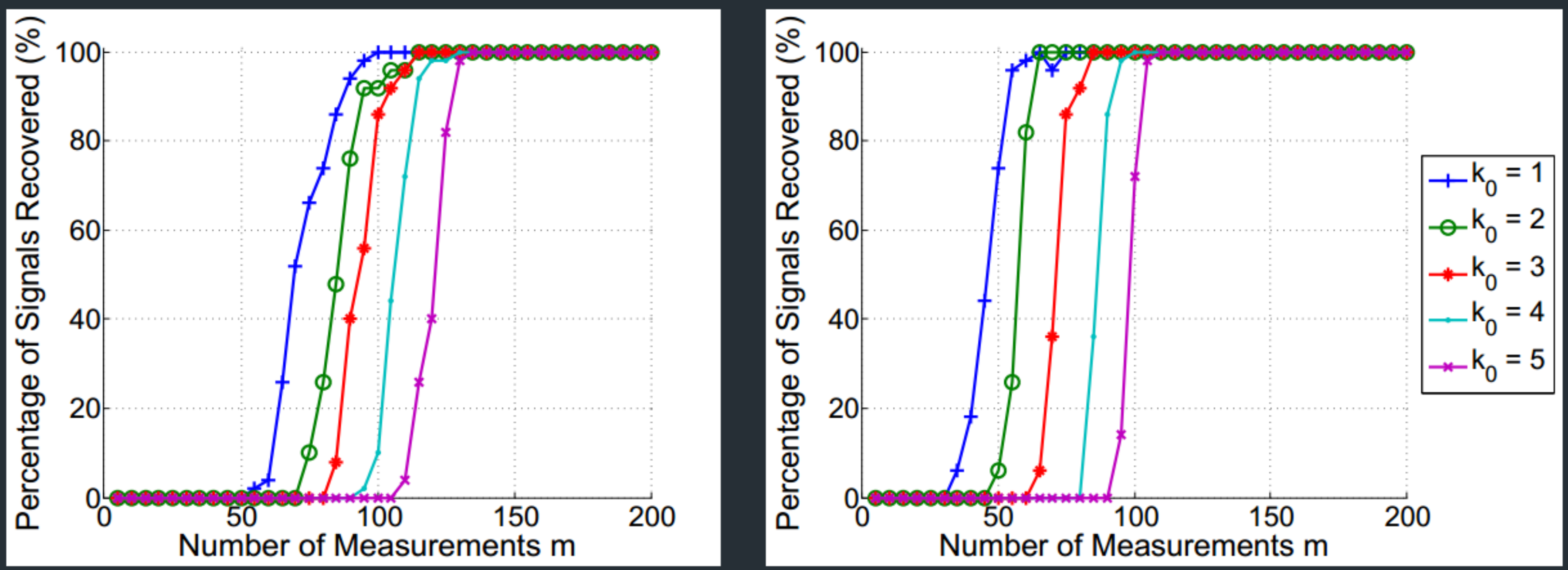


Figure 9: Low-Rank Matrix Recovery: Percent recovery as a function of the number of measurements for GradMP (left) and StoGradMP (right) for various rank levels k_0.

Empirical results

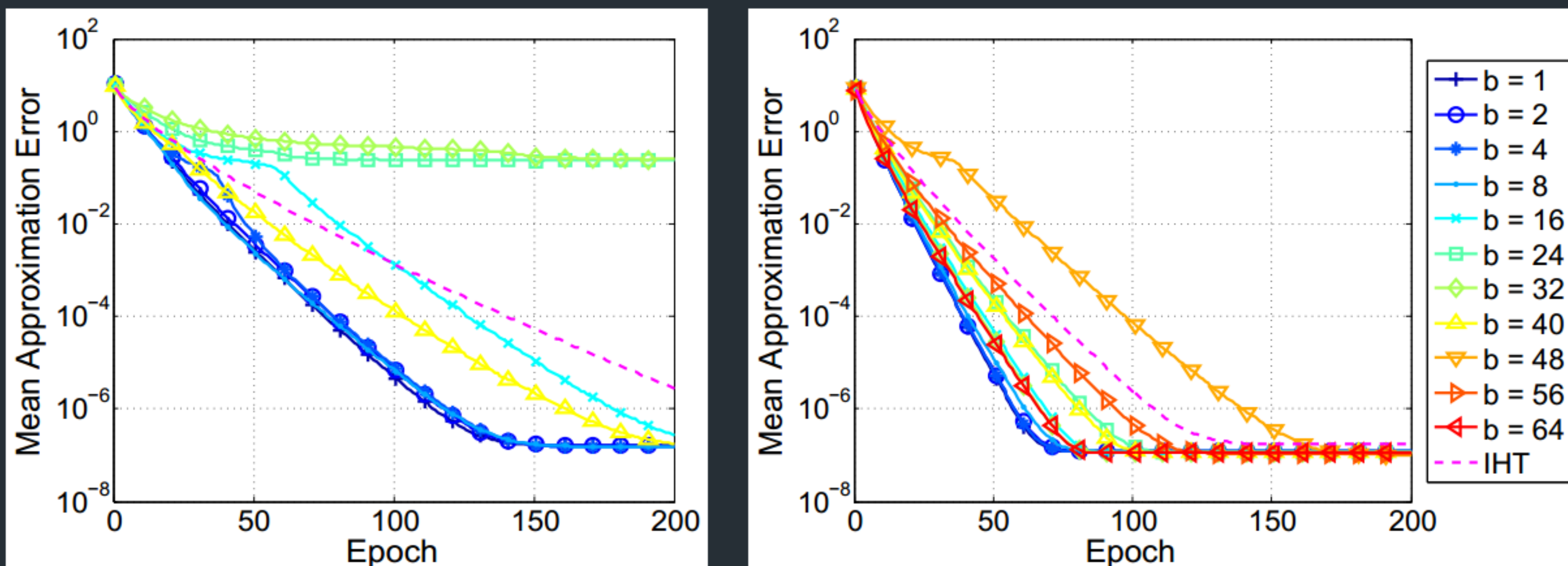


Figure 10: Low-Rank Matrix Recovery: Recovery error as a function of the number of epochs for the IHT methods using $m = 90$ (left) and $m = 140$ (right), shown for various block sizes b used in StoIHT.

Empirical results

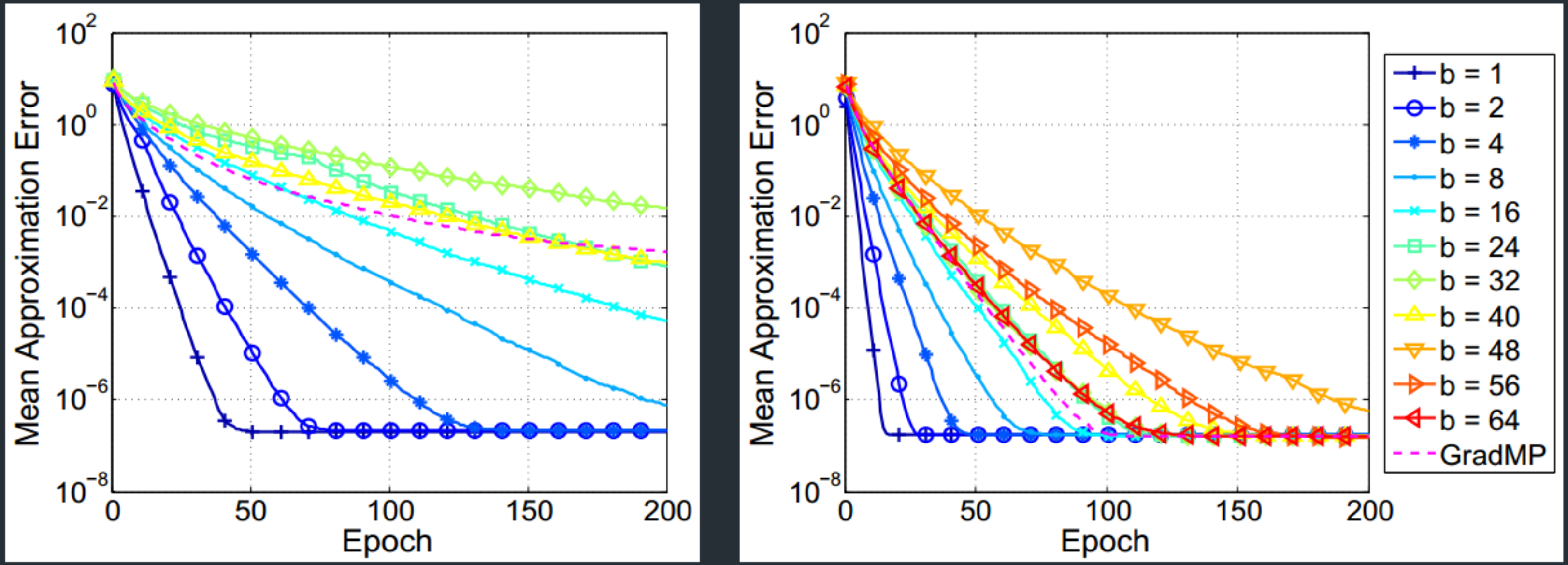


Figure 11: Low-Rank Matrix Recovery: Recovery error as a function of the number of epochs for the GradMP methods using $m = 90$ (left) and $m = 140$ (right), shown for various block sizes b used in StoGradMP.

Empirical results

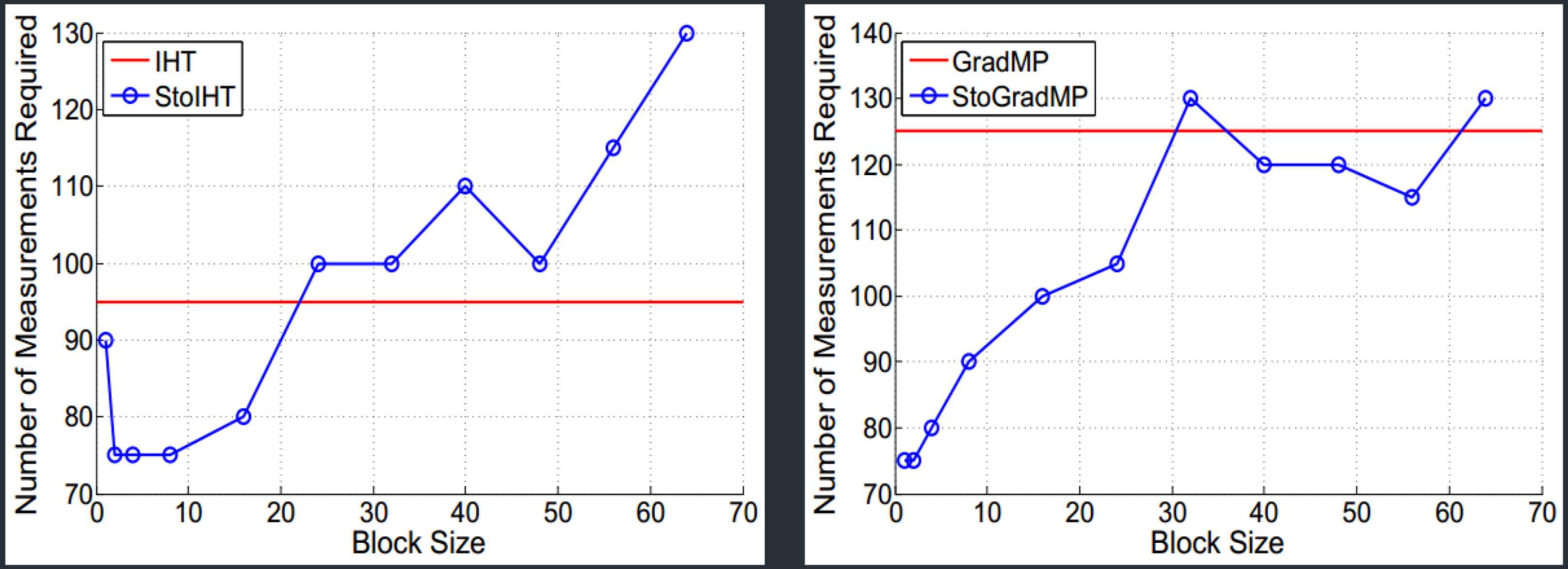


Figure 12: Low-Rank Matrix Recovery: Number of measurements required for signal recovery as a function of block size (blue marker) for StoIHT (left) and StoGradMP (right). Number of measurements required for deterministic method shown as red solid line.

Empirical results

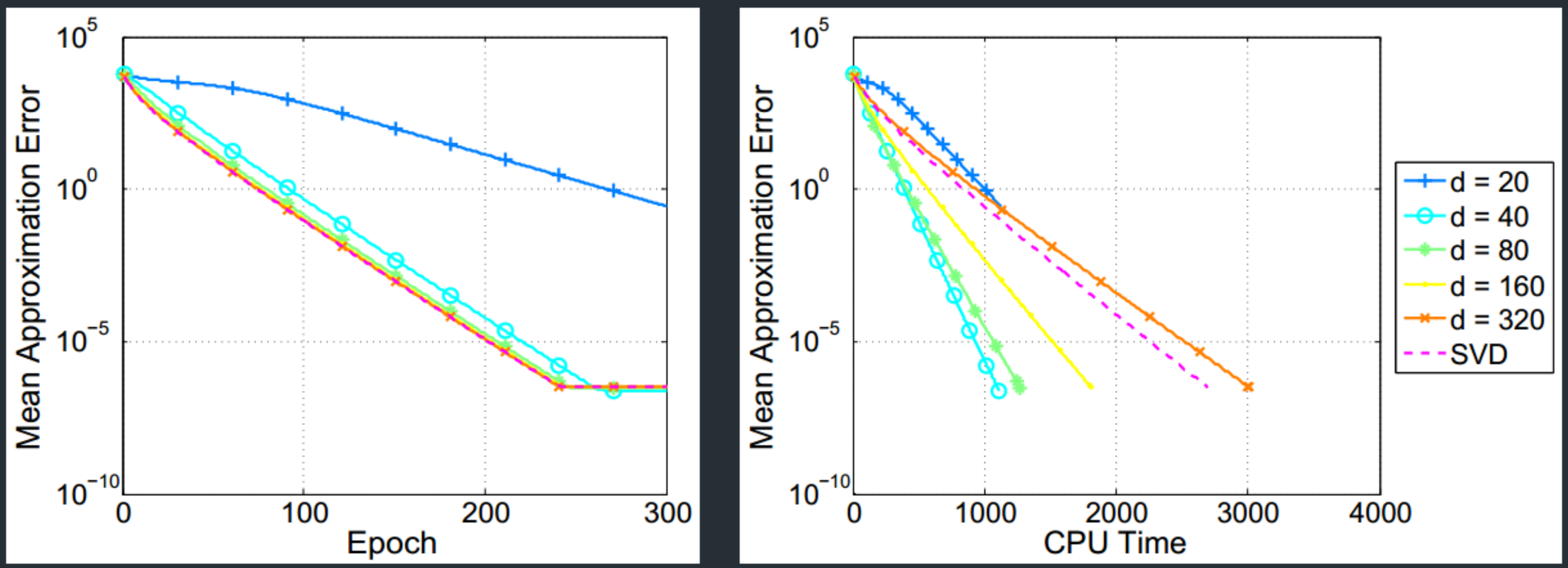


Figure 13: Low-Rank Matrix Recovery with Approximations: Mean recovery error as a function of epochs (left) and runtime (right) for various over-sampling factors d using the StoIHT algorithm. Performance using full SVD computation shown as dashed line.

Empirical results

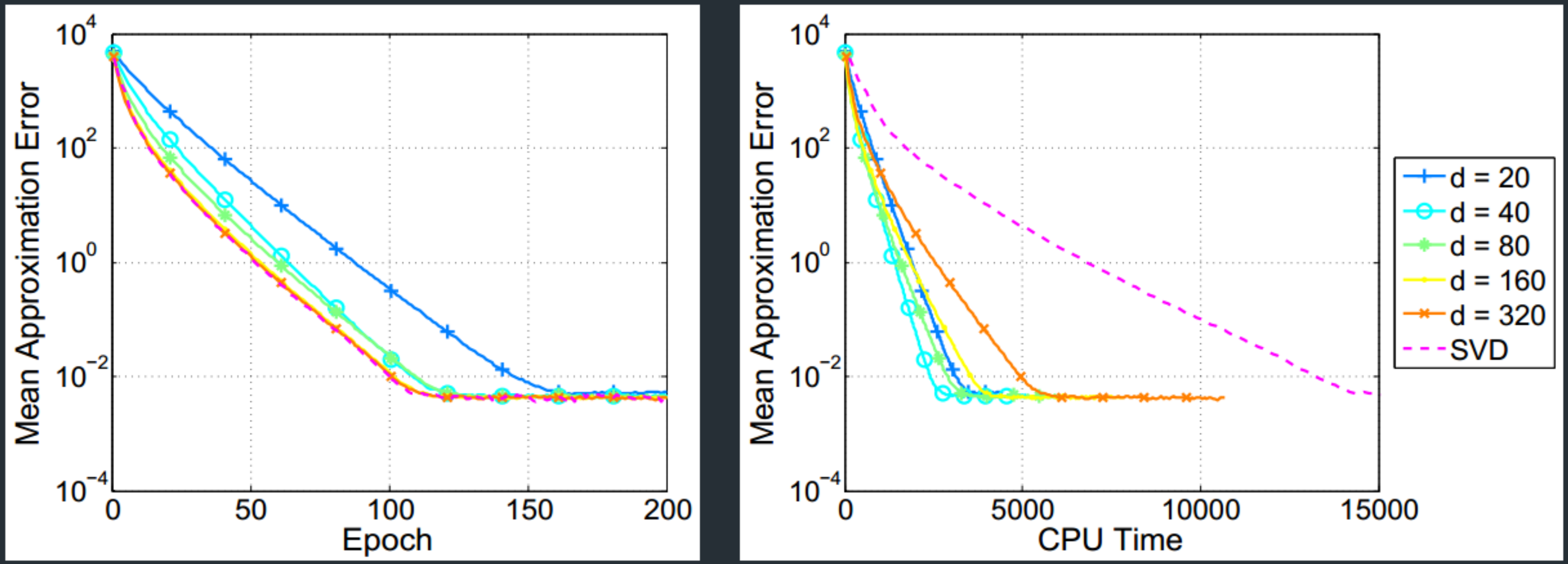


Figure 14: Low-Rank Matrix Recovery with Approximations: Mean recovery error as a function of epochs (left) and runtime (right) for various over-sampling factors d using the StoGradMP algorithm. Performance using full SVD computation shown as dashed line.

Thank you!

For more info:

- dneedell@cmc.edu
- www.cmc.edu/pages/faculty/DNeedell
- *Linear Convergence of Stochastic Iterative Greedy Algorithms with Sparse Constraints*
  by N. Nguyen, D. Needell, and T. Woolf.
  Submitted (Arxiv arxiv.org/abs/1407.0088)