High-dimensional Sparse FFT

Bosu Choi, Andrew Christlieb[a], Yang Wang[b]

(a) Department of Computational Mathematics, Science and Engineering (CMSE),
Michigan State Univ,
(b) Department of Mathematics, The Hong Kong Univ. of Science and Technology

March 1, 2019

Sparse Fourier Transforms (SFTs)

- We consider
$$f(\mathbf{x}) = \sum_{\mathbf{n}\in\mathbb{N}^D} c_{\mathbf{n}} \mathbb{e}^{2\pi \mathbb{i} \mathbf{n}\cdot\mathbf{x}} \approx \sum_{\mathbf{n}\in\mathcal{S}} c_{\mathbf{n}} \mathbb{e}^{2\pi \mathbb{i} \mathbf{n}\cdot\mathbf{x}}$$
where $\mathbf{n} \in [-\frac{N}{2}, \frac{N}{2})^D \cap \mathbb{Z}^D$, nonzero $c_{\mathbf{n}} \in \mathbb{C}$, $\mathbf{x} \in [0,1)^D$ or $\mathbb{R}^D$, $|\mathcal{S}| = s \ll N^D$.

- The Goal: Approximate $f : [0,1)^D \mapsto \mathbb{C}$ using as few evaluations as possible, as quickly as possible.

- There exist 1D sparse Fourier transforms utilizing the ideas of phase-shift and isolation of the Fourier frequencies by taking modulo p, a prime number. Various transformations, projections, and rotations in the physical domain are taken to isolate the frequency vectors and approximate them entry-wise.

Sparse Fourier Transforms (SFTs)

- We consider

$$f(\mathbf{x}) = \sum_{\mathbf{n}\in\mathbb{N}^D} c_{\mathbf{n}} \mathrm{e}^{2\pi \mathrm{i} \mathbf{n}\cdot\mathbf{x}} \approx \sum_{\mathbf{n}\in\mathcal{S}} c_{\mathbf{n}} \mathrm{e}^{2\pi \mathrm{i} \mathbf{n}\cdot\mathbf{x}}$$

  where $\mathbf{n} \in [-\frac{N}{2}, \frac{N}{2})^D \cap \mathbb{Z}^D$, nonzero $c_{\mathbf{n}} \in \mathbb{C}$, $\mathbf{x} \in [0,1)^D$ or $\mathbb{R}^D$, $|\mathcal{S}| = s \ll N^D$.

- The Goal: Approximate $f : [0,1)^D \mapsto \mathbb{C}$ using as few evaluations as possible, as quickly as possible.

- There exist 1D sparse Fourier transforms utilizing the ideas of phase-shift and isolation of the Fourier frequencies by taking modulo p, a prime number. Various transformations, projections, and rotations in the physical domain are taken to isolate the frequency vectors and approximate them entry-wise.

Sparse Fourier Transforms (SFTs)

- We consider

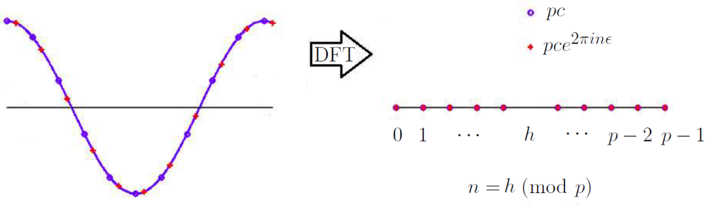
$$f(\mathbf{x}) = \sum_{\mathbf{n}\in\mathbb{N}^D} c_{\mathbf{n}} \mathbb{e}^{2\pi \mathbb{i} \mathbf{n}\cdot\mathbf{x}} \approx \sum_{\mathbf{n}\in\mathcal{S}} c_{\mathbf{n}} \mathbb{e}^{2\pi \mathbb{i} \mathbf{n}\cdot\mathbf{x}}$$

  where $\mathbf{n} \in [-\frac{N}{2}, \frac{N}{2})^D \cap \mathbb{Z}^D$, nonzero $c_{\mathbf{n}} \in \mathbb{C}$, $\mathbf{x} \in [0,1)^D$ or $\mathbb{R}^D$, $|\mathcal{S}| = s \ll N^D$.

- The Goal: Approximate $f : [0,1)^D \mapsto \mathbb{C}$ using as few evaluations as possible, as quickly as possible.

- There exist 1D sparse Fourier transforms utilizing the ideas of phase-shift and isolation of the Fourier frequencies by taking modulo p, a prime number. Various transformations, projections, and rotations in the physical domain are taken to isolate the frequency vectors and approximate them entry-wise.

Description of 1D Sublinear Sparse Fourier Transform (D=1)



- $f(x) = ce^{2\pi i n x}$, $(s = 1)$
- $\boldsymbol{f}_{p,0} = \{f(\frac{\ell}{p})\ ,\ \ell = 0, 1, \cdots, p-1\}$
- $\boldsymbol{f}_{p,\epsilon} = \{f(\frac{\ell}{p} + \epsilon)\ ,\ \ell = 0, 1, \cdots, p-1\}$

- p : prime number s.t. $s < p \ll N$, $\epsilon \leq 1/N$
- $n = \frac{1}{2\pi\epsilon}\mathrm{Arg}\left(\frac{pce^{2\pi i \epsilon n}}{pc}\right)$, $c = \frac{1}{p}pc$

Description of 1D Sublinear Sparse Fourier Transform

- Get two sets of $p > s$ samples from f where p is a prime number and $\epsilon \leq 1/N$.
- $\mathbf{f}_{p,0} = \left(f(0), f(\frac{1}{p}), f(\frac{2}{p}), f(\frac{3}{p}), \cdots, f(\frac{p-1}{p})\right)$
- $\mathbf{f}_{p,\epsilon} = \left(f(0+\epsilon), f(\frac{1}{p}+\epsilon), f(\frac{2}{p}+\epsilon), f(\frac{3}{p}+\epsilon), \cdots, f(\frac{p-1}{p}+\epsilon)\right)$
- $\mathcal{F}(\mathbf{f}_{p,0})[h] = p\sum_{n_j=h(\bmod p)} c_j,\ h = 0, 1, 2, \cdots, p-1$
- $\mathcal{F}(\mathbf{f}_{p,\epsilon})[h] = p\sum_{n_j=h(\bmod p)} c_j e^{2\pi i \epsilon n_j}$
- $n_j = \frac{1}{2\pi\epsilon} Arg\left(\frac{\mathcal{F}(\mathbf{f}_{p,\epsilon})[h]}{\mathcal{F}(\mathbf{f}_{p,0})[h]}\right) = \frac{1}{2\pi\epsilon} Arg\left(\frac{pc_j e^{2\pi i \epsilon n_j}}{pc_j}\right)$, $c_j = \frac{1}{p}\mathcal{F}(\mathbf{f}_{p,0})[h]$ if there is no collision of frequencies.

Test for Collision

- If there is only one frequency n congruent to h mod p, then the following equation holds,
- $\frac{|\mathcal{F}(\mathbf{f}_{p,\epsilon})[h]|}{|\mathcal{F}(\mathbf{f}_{p,0})[h]|} = 1.$
- Otherwise, the equation does not hold for certain ϵ.

- Average-case time complexity : $\Theta(s \log s)$
- Average-case sampling complxity : $\Theta(s)$

2D Sparse Fourier Transforms

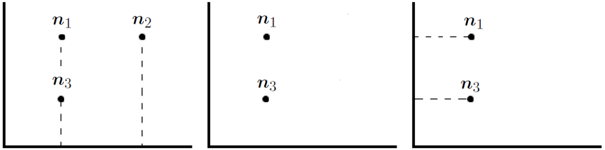


Figure: Process of the parallel projection method in 2D

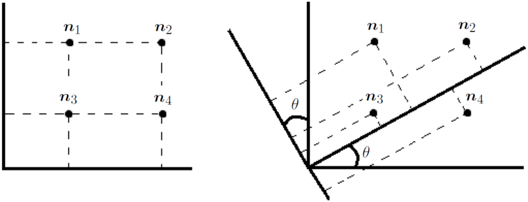


Figure: Worst case scenario in 2D and solving it through the tilting method

Analysis of SFTs

Performance of the tilting method in 2D

Let $\mathbf{n} = (n_1, n_2) \in \mathcal{S} \subset \left[-\frac{N}{2}, \frac{N}{2}\right)^2 \cap \mathbb{Z}^2$. If $\tan\theta = \frac{a}{b}$ such that $c > b > a$ are Pythagorean triples where $b > N$ and a are relative primes, then all $(cn_1\cos\theta - cn_2\sin\theta, cn_1\sin\theta + cn_2\cos\theta)$ rotated by θ does not collide with any other pair through the parallel projection. Thus, all rotated pairs can be identified by the parallel projection method.

- A finite series of rotations in 2D subspaces can be utilized to extend the tilting method to the general higher dimensional setting.
- Furthermore, if the number of dimensions, D, gets larger, the probability that the worst-case scenario happens converges to 0.

Partial Unwrapping Method for $4D$

- $f(\mathbf{x}) = \sum_{\mathbf{n}} ce^{2\pi i\mathbf{n}\cdot\mathbf{x}}$
- $c \in \mathbb{C}$, $\mathbf{n} \in \left[-\frac{N}{2}, \frac{N}{2}\right)^4 \cap \mathbb{Z}^4 = \left(\left[-\frac{N}{2}, \frac{N}{2}\right)^2 \cap \mathbb{Z}^2\right)^2$
- $(n_1, n_2, n_3, n_4) \to (n_1 + Nn_2, n_3 + Nn_4) =: (\widetilde{n}_1, \widetilde{n}_2)$
- The bandwidth of each entry is increased so that the chance of collision from projection decreased.
- Shifting size ϵ should be smaller $\leq \frac{1}{N^2}$

Analysis of SFTs

Performance of the tilting method in 2D

Let $\mathbf{n} = (n_1, n_2) \in \mathcal{S} \subset \left[-\frac{N}{2}, \frac{N}{2}\right)^2 \cap \mathbb{Z}^2$. If $\tan\theta = \frac{a}{b}$ such that $c > b > a$ are Pythagorean triples where $b > N$ and a are relative primes, then all $(cn_1 \cos\theta - cn_2 \sin\theta, cn_1 \sin\theta + cn_2 \cos\theta)$ rotated by θ does not collide with any other pair through the parallel projection. Thus, all rotated pairs can be identified by the parallel projection method.

- A finite series of rotations in 2D subspaces can be utilized to extend the tilting method to the general higher dimensional setting.
- Furthermore, if the number of dimensions, D, gets larger, the probability that the worst-case scenario happens converges to 0.

Partial Unwrapping Method for $4D$

- $f(\mathbf{x}) = \sum_{\mathbf{n}} ce^{2\pi i \mathbf{n}\cdot\mathbf{x}}$
- $c \in \mathbb{C}$, $\mathbf{n} \in \left[-\frac{N}{2}, \frac{N}{2}\right)^4 \cap \mathbb{Z}^4 = \left(\left[-\frac{N}{2}, \frac{N}{2}\right)^2 \cap \mathbb{Z}^2\right)^2$
- $(n_1, n_2, n_3, n_4) \to (n_1 + Nn_2, n_3 + Nn_4) =: (\widetilde{n}_1, \widetilde{n}_2)$
- The bandwidth of each entry is increased so that the chance of collision from projection decreased.
- Shifting size ϵ should be smaller$\leq \frac{1}{N^2}$

Analysis of SFTs

Performance of the tilting method in 2D

Let $\mathbf{n} = (n_1, n_2) \in \mathcal{S} \subset \left[-\frac{N}{2}, \frac{N}{2}\right)^2 \cap \mathbb{Z}^2$. If $\tan\theta = \frac{a}{b}$ such that $c > b > a$ are Pythagorean triples where $b > N$ and a are relative primes, then all $(cn_1 \cos\theta - cn_2 \sin\theta, cn_1 \sin\theta + cn_2 \cos\theta)$ rotated by θ does not collide with any other pair through the parallel projection. Thus, all rotated pairs can be identified by the parallel projection method.

- A finite series of rotations in 2D subspaces can be utilized to extend the tilting method to the general higher dimensional setting.
- Furthermore, if the number of dimensions, D, gets larger, the probability that the worst-case scenario happens converges to 0.

Partial Unwrapping Method for $4D$

- $f(\mathbf{x}) = \sum_{\mathbf{n}} ce^{2\pi i \mathbf{n}\cdot\mathbf{x}}$
- $c \in \mathbb{C}$, $\mathbf{n} \in \left[-\frac{N}{2}, \frac{N}{2}\right)^4 \cap \mathbb{Z}^4 = \left(\left[-\frac{N}{2}, \frac{N}{2}\right)^2 \cap \mathbb{Z}^2\right)^2$
- $(n_1, n_2, n_3, n_4) \to (n_1 + Nn_2, n_3 + Nn_4) =: (\widetilde{n}_1, \widetilde{n}_2)$
- The bandwidth of each entry is increased so that the chance of collision from projection decreased.
- Shifting size ϵ should be smaller$\leq \frac{1}{N^2}$

Analysis of SFTs

Performance of the tilting method in 2D

Let $\mathbf{n} = (n_1, n_2) \in \mathcal{S} \subset \left[-\frac{N}{2}, \frac{N}{2}\right)^2 \cap \mathbb{Z}^2$. If $\tan\theta = \frac{a}{b}$ such that $c > b > a$ are Pythagorean triples where $b > N$ and a are relative primes, then all $(cn_1 \cos\theta - cn_2 \sin\theta, cn_1 \sin\theta + cn_2 \cos\theta)$ rotated by θ does not collide with any other pair through the parallel projection. Thus, all rotated pairs can be identified by the parallel projection method.

- A finite series of rotations in 2D subspaces can be utilized to extend the tilting method to the general higher dimensional setting.
- Furthermore, if the number of dimensions, D, gets larger, the probability that the worst-case scenario happens converges to 0.

Partial Unwrapping Method for $4D$

- $f(\mathbf{x}) = \sum_{\mathbf{n}} c e^{2\pi i \mathbf{n}\cdot\mathbf{x}}$
- $c \in \mathbb{C}$, $\mathbf{n} \in \left[-\frac{N}{2}, \frac{N}{2}\right)^4 \cap \mathbb{Z}^4 = \left(\left[-\frac{N}{2}, \frac{N}{2}\right)^2 \cap \mathbb{Z}^2\right)^2$
- $(n_1, n_2, n_3, n_4) \to (n_1 + Nn_2, n_3 + Nn_4) =: (\widetilde{n}_1, \widetilde{n}_2)$
- The bandwidth of each entry is increased so that the chance of collision from projection decreased.
- Shifting size ϵ should be smaller $\leq \frac{1}{N^2}$

Analysis of SFTs

Average-case runtime complexity

Assume $N \geq 5s$ and there is no worst-case scenario. Let $T(s)$ denote the runtime of the parallel projection method on a random signal setting. Then $\mathbb{E}[T(s)] = \Theta(Ds \log s)$ and

$$\mathbb{P}[T(s) > \Theta(Ds \log s) + tDs \log s] \leq 5^{-t}.$$

Average-case sampling complexity

Assume $N \geq 5s$ and there is no worst-case scenario. Let $S(s)$ denote the number of samples used in the parallel projection method on a random signal setting. Then $\mathbb{E}[S(s)] = \Theta(Ds)$ and

$$\mathbb{P}[S(s) > \Theta(Ds) + tDs] \leq 5^{-t}.$$

Multiscale Algorithm for Noisy Data

- $\widetilde{\mathbf{f}}^{w,y}_{p,\epsilon}[j] = \mathbf{f}^{w,y}_{p,\epsilon}[j] + z_j$ where z_j's are i.i.d. complex Gaussian variables with mean 0 and variance σ^2.
- $\widehat{\widetilde{\mathbf{f}}}^{w,y}_{p,\epsilon}[h] = \widehat{\mathbf{f}}^{w,y}_{p,\epsilon}[h] + \widehat{z}[h]$ where $\widehat{z}[h] = \sum_{j=0}^{p-1} z_j e^{-2\pi i hj/p}$
- $\mathbf{E}\left[\widehat{\widetilde{\mathbf{f}}}^{w,y}_{p,0}[h]\right] = \widehat{\mathbf{f}}^{w,y}_{p,0}[h]$ and $\mathbf{E}\left[|\widehat{\widetilde{\mathbf{f}}}^{w,y}_{p,0}[h] - \widehat{\mathbf{f}}^{w,y}_{p,0}[h]|^2\right] = p\sigma^2$
- $\mathbf{E}\left[\widehat{\widetilde{\mathbf{f}}}^{w,y}_{p,\epsilon}[h]\right] = \widehat{\mathbf{f}}^{w,y}_{p,\epsilon}[h]$ and $\mathbf{E}\left[|\widehat{\widetilde{\mathbf{f}}}^{w,y}_{p,\epsilon}[h] - \widehat{\mathbf{f}}^{w,y}_{p,\epsilon}[h]|^2\right] = p\sigma^2$

- For a non-collision n_y,

$$\frac{\widehat{\widetilde{\mathbf{f}}}^{w,y}_{p,\epsilon}[h]}{\widehat{\widetilde{\mathbf{f}}}^{w,y}_{p,0}[h]} = \frac{\widehat{\mathbf{f}}^{w,y}_{p,0}[h] e^{2\pi i n_y \epsilon} + \mathcal{O}(\sigma\sqrt{p})}{\widehat{\mathbf{f}}^{w,y}_{p,0}[h] + \mathcal{O}(\sigma\sqrt{p})}$$
$$= e^{2\pi i n_y \epsilon} + \mathcal{O}(\sigma / c_{\mathbf{n}}\sqrt{p})$$

- $\left\| \frac{1}{2\pi} \mathrm{Arg}\left(\frac{\widehat{\widetilde{\mathbf{f}}}^{w,y}_{p,\epsilon}[h]}{\widehat{\widetilde{\mathbf{f}}}^{w,y}_{p,0}[h]}\right) - n_y\epsilon \right\|_{\mathbb{Z}} \leq \mathcal{O}\left(\frac{\sigma}{|c_{\mathbf{n}}|\sqrt{p}}\right)$

Multiscale Frequency Estimation

- $\epsilon_0 < 1/N, \epsilon_0 \widetilde{n}_y =_{\mathbb{Z}} \frac{1}{2\pi} \mathrm{Arg} \left(\frac{\widetilde{\widehat{\mathbf{f}}}^{w,y}_{p,\epsilon_0}[h]}{\widetilde{\widehat{\mathbf{f}}}^{w,y}_{p,0}[h]} \right)$
- $\widetilde{n}_y = n_y (\text{mod } p)$
- $\epsilon_1 > 1/N > \epsilon_0, b_1 = \frac{1}{2\pi} \mathrm{Arg} \left(\frac{\widetilde{\widehat{\mathbf{f}}}^{w,y}_{p,\epsilon_1}[h]}{\widetilde{\widehat{\mathbf{f}}}^{w,y}_{p,0}[h]} \right)$
- $b_1 \approx \epsilon_1 n_y (\text{mod } [-\frac{1}{2}, \frac{1}{2}))$
- $\epsilon_1 (n_y - \widetilde{n}_y) \approx (b_1 - \epsilon_1 \widetilde{n}_y)(\text{mod } [-\frac{1}{2}, \frac{1}{2}))$
- $n_y - (\widetilde{n}_y + (b_1 - \epsilon_1 \widetilde{n}_y)(\text{mod } [-\frac{1}{2}, \frac{1}{2}))/\epsilon_1) = \mathcal{O}(\frac{\sigma}{\epsilon_1 \sqrt{p}})$
- This error correction process is iterated with progressively larger shifts ϵ_j.

Multiscale Frequency Estimation

Let $n \in [-\frac{N}{2}, \frac{N}{2})$. Let $0 < \epsilon_0 < \epsilon_1 < \cdots < \epsilon_m$ and $b_0, b_1, \cdots, b_m \in \mathbb{R}$ such that

$$\|\epsilon_j n - b_j\|_{\mathbb{Z}} < \delta, 0 \leq j \leq m$$

where $0 < \delta \leq \frac{1}{4}$. Assume that $\epsilon_0 \leq \frac{1-2\delta}{N}$ and $\beta_j := \epsilon_j / \epsilon_{j-1} \leq (1-2\delta)/(2\delta)$. Then there exist $d_0, d_1, \cdots, d_m \in \mathbb{R}$, each computable from $\{\epsilon_j\}$ and $\{b_j\}$, such that

$$|\widetilde{n} - n| \leq \frac{\delta}{\epsilon_0} \prod_{j=1}^{m} \beta_j^{-1} \text{ where } \widetilde{n} := \sum_{j=0}^{m} \frac{d_j}{\epsilon_j}.$$

Corollary 1

Assume that in the above theorem we have $\beta_j = \beta$ where $\beta \leq (1-2\delta)/(2\delta)$, i.e., $\epsilon_j = \beta^j \epsilon_0$ for all j. Let $m \geq \lfloor \log_\beta \frac{2\delta}{\epsilon_0} \rfloor + 1$. Then

$$|\widetilde{n} - n| \leq \frac{\delta}{\epsilon_0} \beta^{-m} < \frac{1}{2}.$$

Multiscale Approach for Noisy Samples

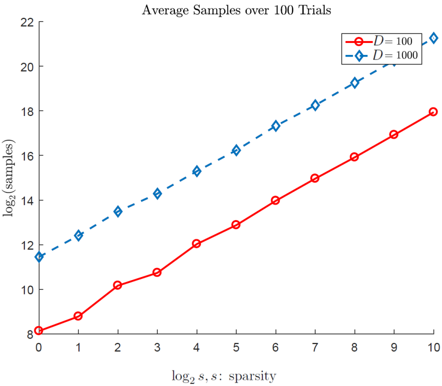
Average-case analysis of the multiscale approach

Let $f^z(\mathbf{x}) = f(\mathbf{x}) + z(\mathbf{x})$, where $\hat{f}(\mathbf{n})$ is s-sparse with all frequencies satisfying $\mathbf{n} \in \mathcal{S} \subset [-N/2, N/2)^D \cap \mathbb{Z}^D$ and not forming any worst case scenario, and z is complex i.i.d. Gaussian noise of variance σ^2. Moreover, suppose that $s > C(\beta(\beta+1)c_{\min}\sigma)^2$ for some constant C (chosen carefully so that some technical assumptions are satisfied). The multiscale parallel projection method, given N, D, s, β with $N > 5s$ and access to $f^z(\mathbf{x})$ returns a list of s pairs $(\hat{\mathbf{n}}, c_{\hat{\mathbf{n}}})$ such that (i) each $\hat{\mathbf{n}} \in \mathcal{S}$ and (ii) for each $\hat{\mathbf{n}}$, there is an $\mathbf{n} \in \mathcal{S}$ such that $\mathbf{n} \in \mathcal{S}$ with $|c_{\mathbf{n}} - c_{\hat{\mathbf{n}}}| \leq C\sigma/\sqrt{s}$. The average-case runtime and sampling complexity are

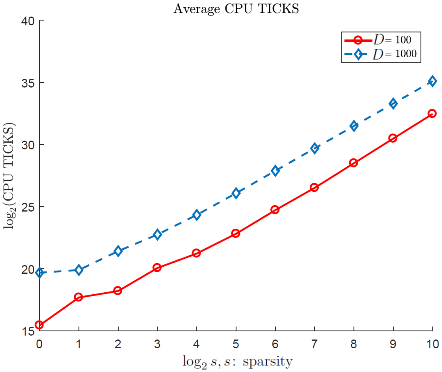
$$\Theta(sD \log s \log N) \qquad \text{and} \qquad \Theta(sD \log N),$$

respectively, over the class of random signals.

Numerics



(a)



(b)

Figure: Average samples and average runtime (nosieless)

Numerics

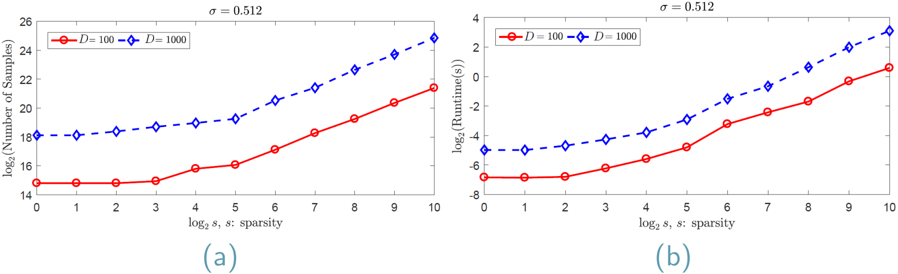
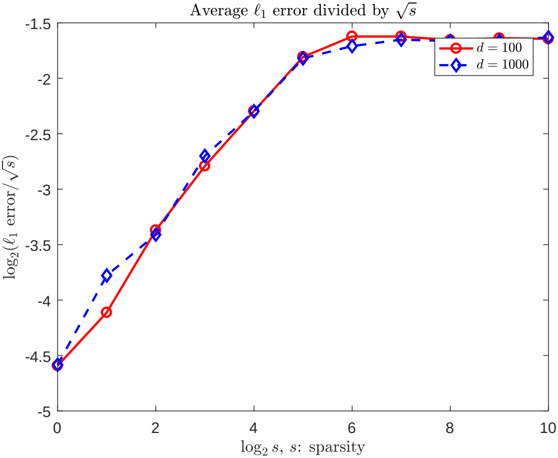


Figure: Average samples and average runtime (noisy)

Numerics



(a)

Figure: Average ℓ^1 error divided by s

Thanks for Listening!

Questions?