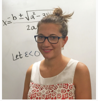
Greedy and Randomized Projection Methods

Jamie Haddock

UC Irvine Probability Seminar,

October 1, 2019

Computational and Applied Mathematics
UCLA



joint with Jesús A. De Loera, Deanna Needell, and Anna Ma

https://arxiv.org/abs/1802.03126 (BIT Numerical Mathematics 2019)
https://arxiv.org/abs/1605.01418 (SISC 2017)

BIG Data

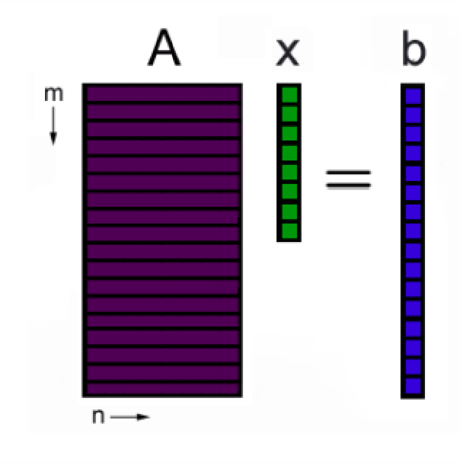

The big data opportunity
LET'S SOLVE THIS PROBLEM BY USING THE BIG DATA NONE OF US HAVE THE SLIGHTEST IDEA WHAT TO DO WITH
TOM FISH BURNE
© marketoonist.com
DellWorld'15

BIG Data

The Economist
Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
The right to eat cats and dogs
The data deluge
AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

Setup

We are interested in solving highly overdetermined systems of equations (or inequalities), $A\mathbf{x} = \mathbf{b}$ ($A\mathbf{x} \leq \mathbf{b}$), where $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$ and $m \gg n$. Rows are denoted $\mathbf{a}_i^T$.



Iterative Projection Methods

If $\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}\}$ is nonempty, these methods construct an approximation to a solution:

1. Randomized Kaczmarz Method

Applications:

1. Tomography (Algebraic Reconstruction Technique)

Iterative Projection Methods

If $\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}\}$ is nonempty, these methods construct an approximation to a solution:

1. Randomized Kaczmarz Method
2. Motzkin's Method

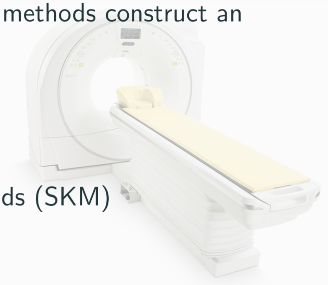
Applications:

1. Tomography (Algebraic Reconstruction Technique)
2. Linear programming

Iterative Projection Methods

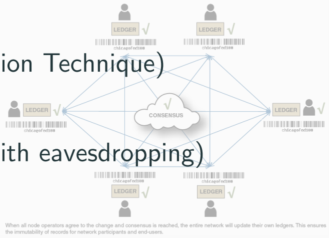
If $\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}\}$ is nonempty, these methods construct an approximation to a solution:

1. Randomized Kaczmarz Method
2. Motzkin's Method
3. Sampling Kaczmarz-Motzkin Methods (SKM)

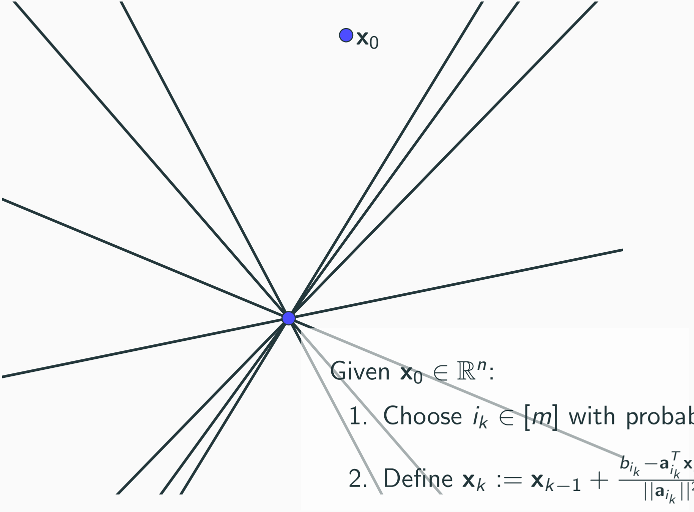


Applications:

1. Tomography (Algebraic Reconstruction Technique)
2. Linear programming
3. Average consensus (greedy gossip with eavesdropping)



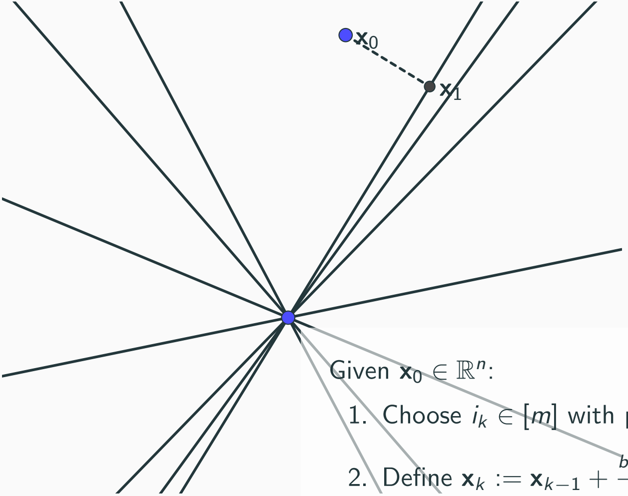
Kaczmarz Method



Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $i_k \in [m]$ with probability $\frac{\|\mathbf{a}_{i_k}\|^2}{\|A\|_F^2}$.
2. Define $\mathbf{x}_k := \mathbf{x}_{k-1} + \frac{b_{i_k} - \mathbf{a}_{i_k}^T \mathbf{x}_{k-1}}{||\mathbf{a}_{i_k}||^2} \mathbf{a}_{i_k}$.
3. Repeat.

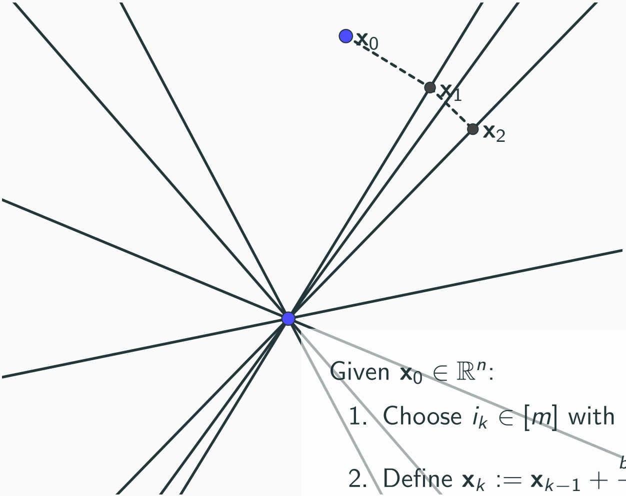
Kaczmarz Method



Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $i_k \in [m]$ with probability $\frac{\|\mathbf{a}_{i_k}\|^2}{\|A\|_F^2}$.
2. Define $\mathbf{x}_k := \mathbf{x}_{k-1} + \frac{b_{i_k} - \mathbf{a}_{i_k}^T \mathbf{x}_{k-1}}{||\mathbf{a}_{i_k}||^2} \mathbf{a}_{i_k}$.
3. Repeat.

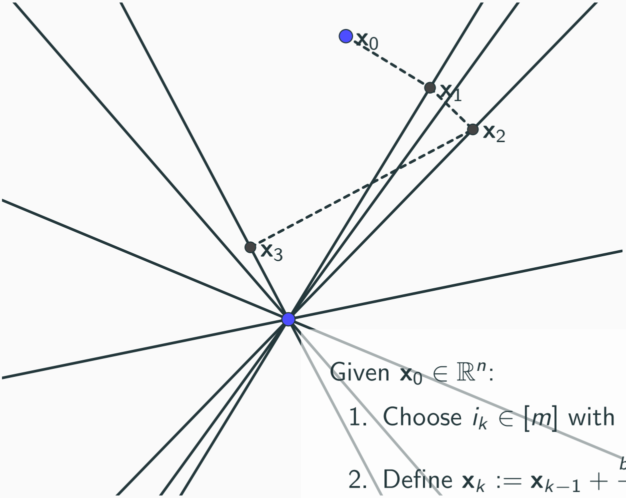
Kaczmarz Method



Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $i_k \in [m]$ with probability $\frac{\|\mathbf{a}_{i_k}\|^2}{\|A\|_F^2}$.
2. Define $\mathbf{x}_k := \mathbf{x}_{k-1} + \frac{b_{i_k} - \mathbf{a}_{i_k}^T \mathbf{x}_{k-1}}{||\mathbf{a}_{i_k}||^2} \mathbf{a}_{i_k}$.
3. Repeat.

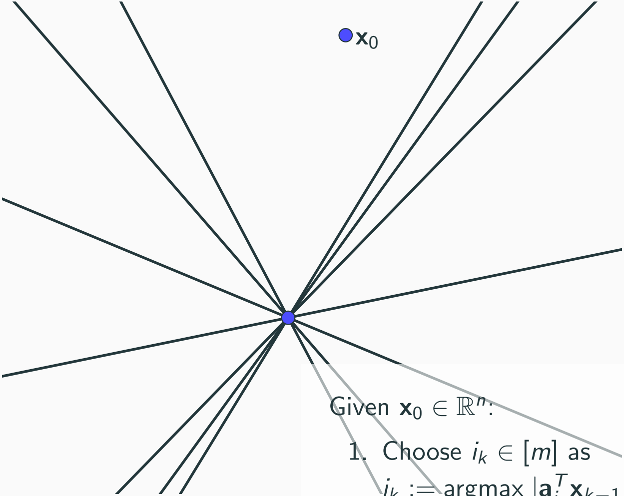
Kaczmarz Method



Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $i_k \in [m]$ with probability $\frac{\|\mathbf{a}_{i_k}\|^2}{\|A\|_F^2}$.
2. Define $\mathbf{x}_k := \mathbf{x}_{k-1} + \frac{b_{i_k} - \mathbf{a}_{i_k}^T \mathbf{x}_{k-1}}{\|\mathbf{a}_{i_k}\|^2} \mathbf{a}_{i_k}$.
3. Repeat.

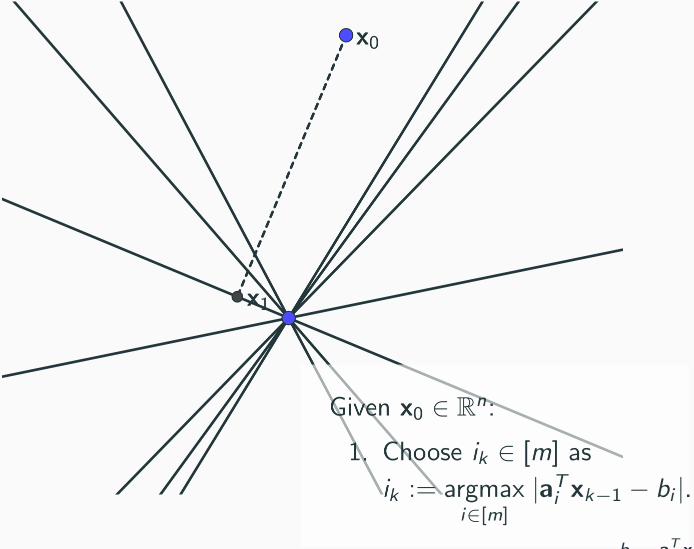
Motzkin's Method



Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $i_k \in [m]$ as
   $i_k := \underset{i \in [m]}{\operatorname{argmax}} |\mathbf{a}_i^T \mathbf{x}_{k-1} - b_i|$.
2. Define $\mathbf{x}_k := \mathbf{x}_{k-1} + \frac{b_{i_k} - \mathbf{a}_{i_k}^T \mathbf{x}_{k-1}}{||\mathbf{a}_{i_k}||^2} \mathbf{a}_{i_k}$.
3. Repeat.

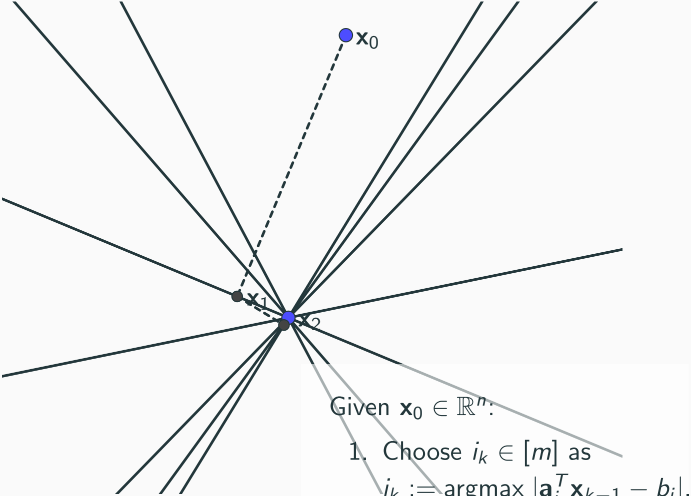
Motzkin's Method



Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $i_k \in [m]$ as
   $i_k := \underset{i \in [m]}{\operatorname{argmax}} |\mathbf{a}_i^T \mathbf{x}_{k-1} - b_i|$.
2. Define $\mathbf{x}_k := \mathbf{x}_{k-1} + \frac{b_{i_k} - \mathbf{a}_{i_k}^T \mathbf{x}_{k-1}}{||\mathbf{a}_{i_k}||^2} \mathbf{a}_{i_k}$.
3. Repeat.

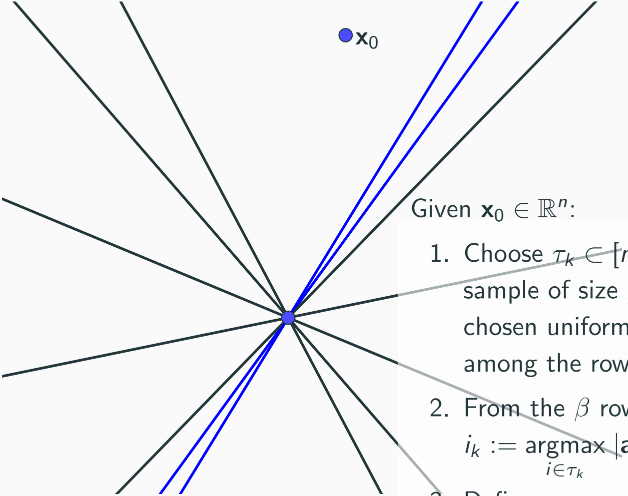
Motzkin's Method



Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $i_k \in [m]$ as
$$i_k := \underset{i \in [m]}{\operatorname{argmax}} |\mathbf{a}_i^T \mathbf{x}_{k-1} - b_i|.$$
2. Define $\mathbf{x}_k := \mathbf{x}_{k-1} + \frac{b_{i_k} - \mathbf{a}_{i_k}^T \mathbf{x}_{k-1}}{||\mathbf{a}_{i_k}||^2} \mathbf{a}_{i_k}$.
3. Repeat.

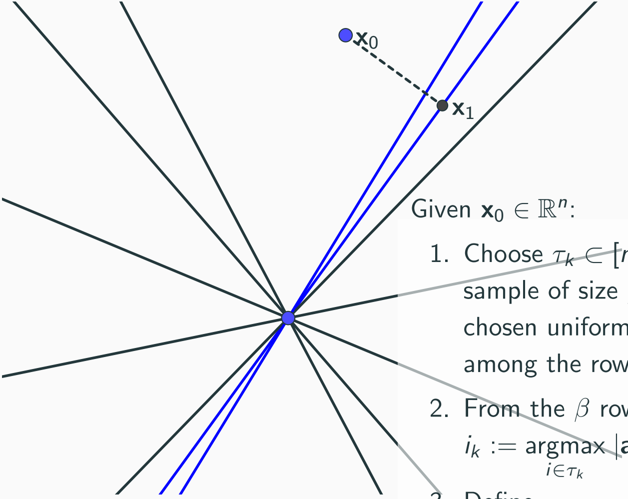
Our Hybrid Method (SKM)



Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $\tau_k \subset [m]$ to be a sample of size β constraints chosen uniformly at random among the rows of A.
2. From the β rows, choose $i_k := \underset{i \in \tau_k}{\text{argmax}} |\mathbf{a}_i^T \mathbf{x}_{k-1} - b_i|$.
3. Define $\mathbf{x}_k := \mathbf{x}_{k-1} + \frac{b_{i_k} - \mathbf{a}_{i_k}^T \mathbf{x}_{k-1}}{||\mathbf{a}_{i_k}||^2} \mathbf{a}_{i_k}$.
4. Repeat.

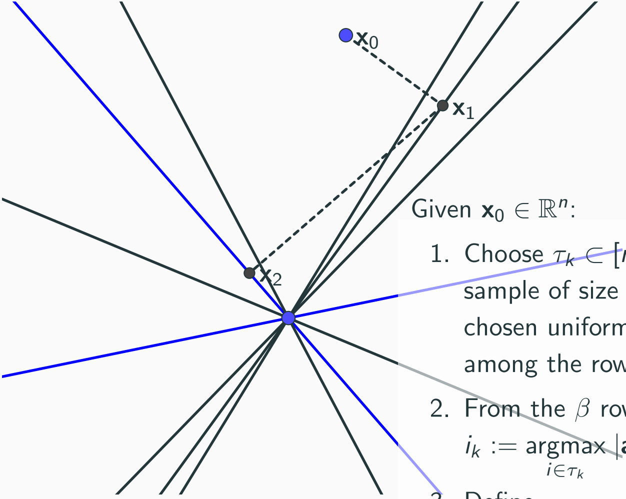
Our Hybrid Method (SKM)



Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $\tau_k \subset [m]$ to be a sample of size β constraints chosen uniformly at random among the rows of A.
2. From the β rows, choose $i_k := \operatorname{argmax}_{i \in \tau_k} |\mathbf{a}_i^T \mathbf{x}_{k-1} - b_i|$.
3. Define $\mathbf{x}_k := \mathbf{x}_{k-1} + \frac{b_{i_k} - \mathbf{a}_{i_k}^T \mathbf{x}_{k-1}}{||\mathbf{a}_{i_k}||^2} \mathbf{a}_{i_k}$.
4. Repeat.

Our Hybrid Method (SKM)



Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $\tau_k \subset [m]$ to be a sample of size β constraints chosen uniformly at random among the rows of A.
2. From the β rows, choose $i_k := \underset{i \in \tau_k}{\operatorname{argmax}} |\mathbf{a}_i^T \mathbf{x}_{k-1} - b_i|$.
3. Define $\mathbf{x}_k := \mathbf{x}_{k-1} + \frac{b_{i_k} - \mathbf{a}_{i_k}^T \mathbf{x}_{k-1}}{||\mathbf{a}_{i_k}||^2} \mathbf{a}_{i_k}$.
4. Repeat.

Glimpse of HUGE Body of Literature

RK: [Strohmer-Vershynin '09], [Needell-Srebro-Ward '16]

Greedy: [Censor '81], [Nutini et al '16], [Bai-Wu '18], [Du-Gao '19]

Accel.: [Hanke-Niethammer '90], [Liu-Wright '16], [Morshed-Islam '19]

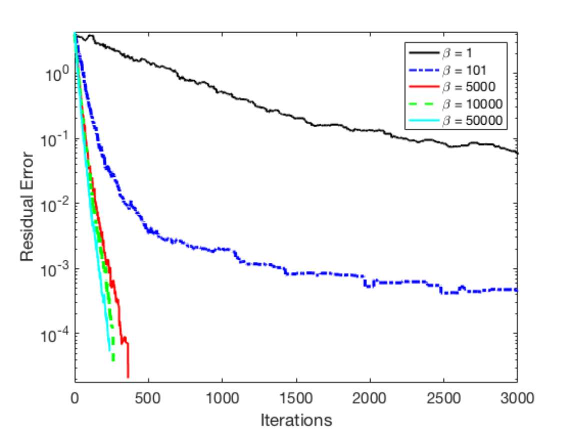
Block: [Popa et al '12], [Needell-Tropp '14], [Needell-Zhao-Zouzias '15],

Sketching: [Gower-Richtarik '15], [Needell-Rebrova '19]

Phase retrieval: [Tan-Vershynin '17], [Jeong-Güntürk '17]

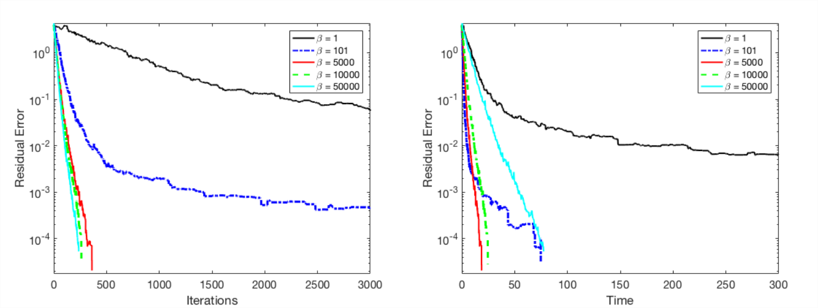
LP: [Motzkin-Schoenberg '54], [Agmon '54], [Goffin '80], [Chubanov '12]

Experimental Convergence



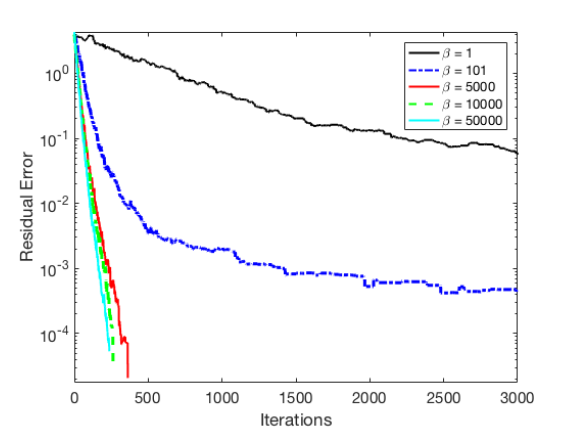
- β: sample size
- A is 50000×100 Gaussian matrix, consistent system
- 'faster' convergence for larger sample size

Experimental Convergence



- β: sample size
- A is 50000×100 Gaussian matrix, consistent system
- 'faster' convergence for larger sample size

Experimental Convergence



- β: sample size
- A is 50000×100 Gaussian matrix, consistent system
- 'faster' convergence for larger sample size

Convergence Rates

Below are the convergence rates for the methods on a system, $A\mathbf{x} = \mathbf{b}$, which is consistent with unique solution $\mathbf{x}$, whose rows have been normalized to have unit norm.

- ▷ RK (Strohmer - Vershynin '09):

$$\mathbb{E}||\mathbf{x}_k - \mathbf{x}||_2^2 \leq \left(1 - \frac{\sigma_{\min}^2(A)}{m}\right)^k ||\mathbf{x}_0 - \mathbf{x}||_2^2$$

Convergence Rates

Below are the convergence rates for the methods on a system, $A\mathbf{x} = \mathbf{b}$, which is consistent with unique solution $\mathbf{x}$, whose rows have been normalized to have unit norm.

- RK (Strohmer - Vershynin '09):

$$\mathbb{E}||\mathbf{x}_k - \mathbf{x}||_2^2 \leq \left(1 - \frac{\sigma_{\min}^2(A)}{m}\right)^k ||\mathbf{x}_0 - \mathbf{x}||_2^2$$

- MM (Agmon '54):

$$\|\mathbf{x}_k - \mathbf{x}\|_2^2 \leq \left(1 - \frac{\sigma_{\min}^2(A)}{m}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|_2^2$$

Convergence Rates

Below are the convergence rates for the methods on a system, $A\mathbf{x} = \mathbf{b}$, which is consistent with unique solution $\mathbf{x}$, whose rows have been normalized to have unit norm.

- RK (Strohmer - Vershynin '09):

$$\mathbb{E}||\mathbf{x}_k - \mathbf{x}||_2^2 \leq \left(1 - \frac{\sigma_{\min}^2(A)}{m}\right)^k ||\mathbf{x}_0 - \mathbf{x}||_2^2$$

- MM (Agmon '54):

$$\|\mathbf{x}_k - \mathbf{x}\|_2^2 \leq \left(1 - \frac{\sigma_{\min}^2(A)}{m}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|_2^2$$

- SKM (DeLoera - H. - Needell '17):

$$\mathbb{E}\|\mathbf{x}_k - \mathbf{x}\|_2^2 \leq \left(1 - \frac{\sigma_{\min}^2(A)}{m}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|_2^2$$

Convergence Rates

Below are the convergence rates for the methods on a system, $A\mathbf{x} = \mathbf{b}$, which is consistent with unique solution $\mathbf{x}$, whose rows have been normalized to have unit norm.

- RK (Strohmer - Vershynin '09):

$$\mathbb{E}||\mathbf{x}_k - \mathbf{x}||_2^2 \leq \left(1 - \frac{\sigma_{\min}^2(A)}{m}\right)^k ||\mathbf{x}_0 - \mathbf{x}||_2^2$$

- MM (Agmon '54):

$$\|\mathbf{x}_k - \mathbf{x}\|_2^2 \leq \left(1 - \frac{\sigma_{\min}^2(A)}{m}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|_2^2$$

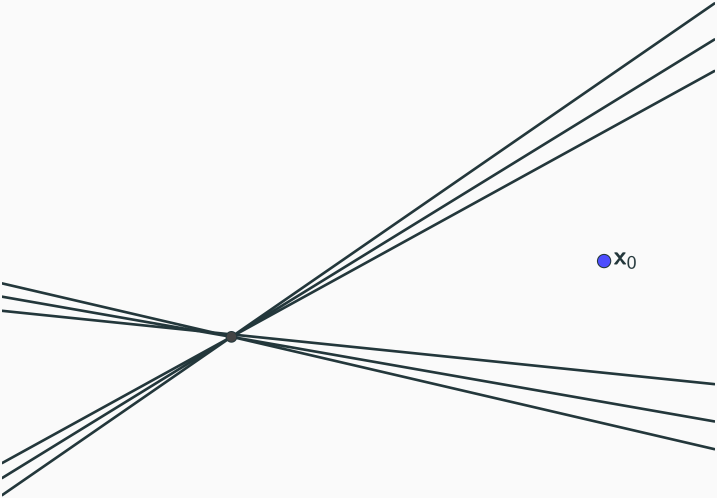
- SKM (DeLoera - H. - Needell '17):

$$\mathbb{E}\|\mathbf{x}_k - \mathbf{x}\|_2^2 \leq \left(1 - \frac{\sigma_{\min}^2(A)}{m}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|_2^2$$

`Why are these all the same?`

A Pathological Example

Because.



Structure of the Residual

Several works have used sparsity of the residual to improve the convergence rate of greedy methods.

[De Loera, H., Needell '17], [Bai, Wu '18], [Du, Gao '19]

Structure of the Residual

Several works have used sparsity of the residual to improve the convergence rate of greedy methods.

[De Loera, H., Needell '17], [Bai, Wu '18], [Du, Gao '19]

However, not much sparsity can be expected in most cases. Instead, we'd like to use dynamic range of the residual to guarantee faster convergence.

$$\gamma_k := \frac{\|A\mathbf{x}_k - A\mathbf{x}\|^2}{\|A\mathbf{x}_k - A\mathbf{x}\|_\infty^2}$$

An Accelerated Convergence Rate

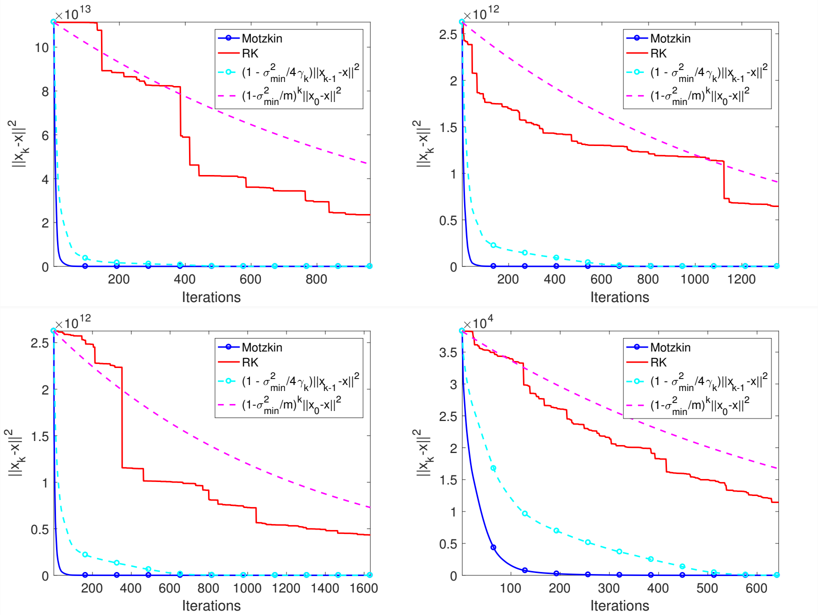
Theorem (H. - Needell '19)
Let $\mathbf{x}$ *denote the solution of the consistent, normalized system* $A\mathbf{x} = \mathbf{b}$*. Motzkin's method exhibits the (possibly highly accelerated) convergence rate:*

$$\|\mathbf{x}_T - \mathbf{x}\|^2 \leq \prod_{k=0}^{T-1} \left(1 - \frac{\sigma_{\min}^2(A)}{4\gamma_k}\right) \cdot \|\mathbf{x}_0 - \mathbf{x}\|^2$$

Here γ_k *bounds the dynamic range of the* k*th residual,* $\gamma_k := \frac{\|A\mathbf{x}_k - A\mathbf{x}\|^2}{\|A\mathbf{x}_k - A\mathbf{x}\|_\infty^2}$.

- improvement over previous result when $4\gamma_k < m$

Netlib LP Systems



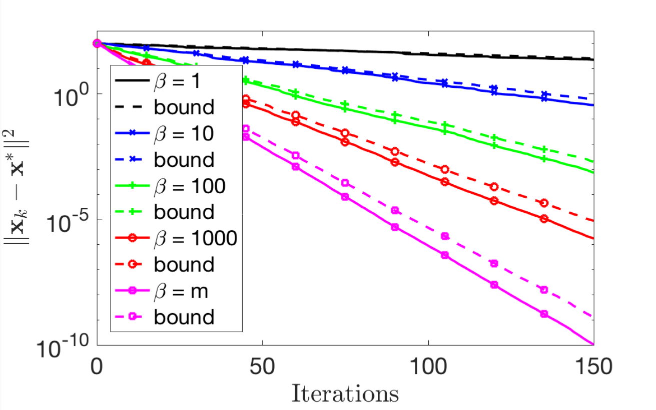
Extending to SKM

Corollary (H. - Ma 2019+)
Let A be normalized so $\|\mathbf{a}_i\|_2 = 1$ for all rows $i = 1, ..., m$. If the system $A\mathbf{x} = \mathbf{b}$ is consistent with the unique solution $\mathbf{x}^*$ then the SKM method converges at least linearly in expectation and the rate depends on the dynamic range of the random sample of rows of A, τ_j. Precisely, in the $j+1$st iteration of SKM, we have

$$\mathbb{E}_{\tau_j}\|\mathbf{x}_{j+1} - \mathbf{x}^*\|_2^2 \leq \left(1 - \frac{\beta\sigma_{\min}^2(A)}{\gamma_j m}\right)\|\mathbf{x}_j - \mathbf{x}^*\|_2^2$$

where $\gamma_j = \frac{\sum_{\tau_j \in \binom{[m]}{\beta}} \|A_{\tau_j}\mathbf{x}_j - \mathbf{b}_{\tau_j}\|_2^2}{\sum_{\tau_j \in \binom{[m]}{\beta}} \|A_{\tau_j}\mathbf{x}_j - \mathbf{b}_{\tau_j}\|_\infty^2}$.

Extending to SKM



- A is 50000×100 Gaussian matrix, consistent system
- bound uses dynamic range of sample of β rows

What can we say about γ_j?

$$1 \leq \gamma_j \leq \beta$$

What can we say about γ_j?

$$1 \leq \gamma_j \leq \beta$$

$$\nearrow \qquad\qquad\qquad \nwarrow$$

$$A\mathbf{x}_k - \mathbf{b} = \mathbf{e}_i \qquad\qquad A\mathbf{x}_k - \mathbf{b} = c\mathbf{1}$$

What can we say about γ_j?

$$1 \le \gamma_j \le \beta$$

$$\nearrow \qquad\qquad \nwarrow$$

$A\mathbf{x}_k - \mathbf{b} = \mathbf{e}_i$ — "Best" case

$A\mathbf{x}_k - \mathbf{b} = c\mathbf{1}$ — "Worst" case

What can we say about γ_j?

$$1 \leq \gamma_j \leq \beta$$

$\nearrow$ $\nwarrow$

$A\mathbf{x}_k - \mathbf{b} = \mathbf{e}_i$ — "Best" case

$A\mathbf{x}_k - \mathbf{b} = c\mathbf{1}$ — "Worst" case

	Best Case	Worst Case	Previous Best	Previous Worst
MM	$1 - \sigma^2_{\min}(A)$	$1 - \frac{\sigma^2_{\min}(A)}{m}$	$1 - \frac{\sigma^2_{\min}(A)}{4}$	$1 - \frac{\sigma^2_{\min}(A)}{m}$
SKM	$1 - \frac{\beta\sigma^2_{\min}(A)}{m}$		$1 - \frac{\sigma^2_{\min}(A)}{m}$	
RK	$1 - \frac{\sigma^2_{\min}(A)}{m}$			

Table 1: Contraction terms α such that $\mathbb{E}_{\tau_k}\|\mathbf{e}_k\|^2 \leq \alpha\|\mathbf{e}_{k-1}\|^2$.

What can we say about γ_j?

$$1 \leq \gamma_j \leq \beta$$

$$\nearrow \qquad\qquad \nwarrow$$

$A\mathbf{x}_k - \mathbf{b} = \mathbf{e}_i$ — "Best" case

$A\mathbf{x}_k - \mathbf{b} = c\mathbf{1}$ — "Worst" case

<table>
<tr><th></th><th>Best Case</th><th>Worst Case</th><th>Previous Best</th><th>Previous Worst</th></tr>
<tr><td>MM</td><td>$1 - \sigma^2_{\min}(A)$</td><td rowspan="3">$1 - \frac{\sigma^2_{\min}(A)}{m}$</td><td>$1 - \frac{\sigma^2_{\min}(A)}{4}$</td><td rowspan="3">$1 - \frac{\sigma^2_{\min}(A)}{m}$</td></tr>
<tr><td>SKM</td><td>$1 - \frac{\beta\sigma^2_{\min}(A)}{m}$</td><td rowspan="2">$1 - \frac{\sigma^2_{\min}(A)}{m}$</td></tr>
<tr><td>RK</td><td>$1 - \frac{\sigma^2_{\min}(A)}{m}$</td></tr>
</table>

Table 1: Contraction terms α such that $\mathbb{E}_{\tau_k}\|\mathbf{e}_k\|^2 \leq \alpha\|\mathbf{e}_{k-1}\|^2$.

`Nervous?`

What can we say about γ_j?

$$1 \leq \gamma_j \leq \beta$$

$\nearrow$ $\nwarrow$

$A\mathbf{x}_k - \mathbf{b} = \mathbf{e}_i$ — "Best" case

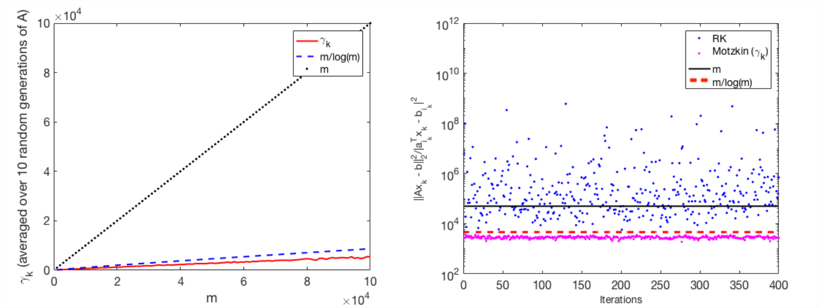
$A\mathbf{x}_k - \mathbf{b} = c\mathbf{1}$ — "Worst" case

	Best Case	Worst Case	Previous Best	Previous Worst
MM	$1 - \sigma^2_{\min}(A)$	$1 - \frac{\sigma^2_{\min}(A)}{m}$	$1 - \frac{\sigma^2_{\min}(A)}{4}$	$1 - \frac{\sigma^2_{\min}(A)}{m}$
SKM	$1 - \frac{\beta\sigma^2_{\min}(A)}{m}$		$1 - \frac{\sigma^2_{\min}(A)}{m}$	
RK	$1 - \frac{\sigma^2_{\min}(A)}{m}$			

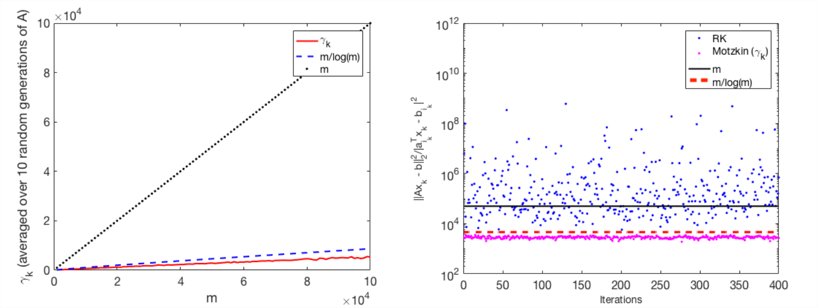
Table 1: Contraction terms α such that $\mathbb{E}_{\tau_k}\|\mathbf{e}_k\|^2 \leq \alpha\|\mathbf{e}_{k-1}\|^2$.

Nervous? $\gamma_k \geq \frac{\beta}{m}\sigma^2_{\min}(A)$ when A is row-normalized

γ_k: Gaussian systems

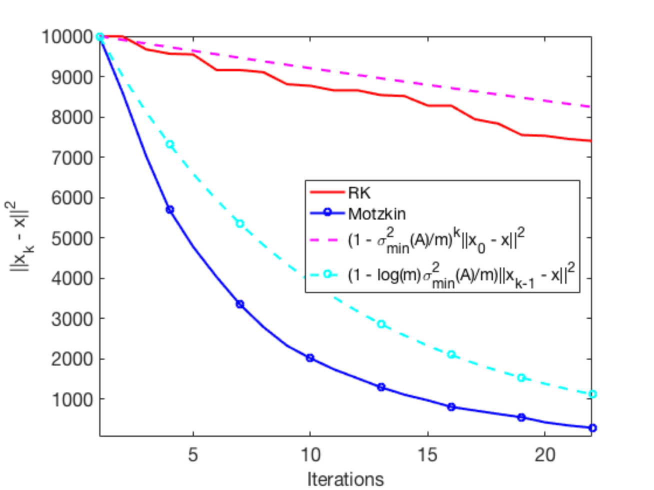


γ_k: Gaussian systems



$$\gamma_k \lesssim \frac{n\beta}{\log \beta}$$

Gaussian Convergence



▷ A is 50000×100 Gaussian matrix, consistent system

Generalizing the Convergence Result

- ▷ can immediately generalize to varying β (SKM with β_k)

Generalizing the Convergence Result

- can immediately generalize to varying β (SKM with β_k)
- to generalize to non-normalized A, we need

Generalizing the Convergence Result

- ▷ can immediately generalize to varying β (SKM with β_k)
- ▷ to generalize to non-normalized A, we need
  - a sampling distribution that depends upon $\|\mathbf{a}_i\|^2$

Generalizing the Convergence Result

- ▷ can immediately generalize to varying β (SKM with β_k)
- ▷ to generalize to non-normalized A, we need
  - a sampling distribution that depends upon $\|\mathbf{a}_i\|^2$
  - a sampling distribution that depends upon $\mathbf{x}_k$

Generalizing the Convergence Result

- ▷ can immediately generalize to varying β (SKM with β_k)
- ▷ to generalize to non-normalized A, we need
  - a sampling distribution that depends upon $\|\mathbf{a}_i\|^2$
  - a sampling distribution that depends upon $\mathbf{x}_k$

Generalized SKM sampling distribution: $p_{\mathbf{x}} : \binom{[m]}{\beta_k} \to [0, 1)$

Generalizing the Convergence Result

- ▷ can immediately generalize to varying β (SKM with β_k)
- ▷ to generalize to non-normalized A, we need
  - a sampling distribution that depends upon $\|\mathbf{a}_i\|^2$
  - a sampling distribution that depends upon $\mathbf{x}_k$

Generalized SKM sampling distribution: $p_{\mathbf{x}} : \binom{[m]}{\beta_k} \to [0, 1)$

- $t(\tau, \mathbf{x}) = \text{argmax}_{t \in \tau}(\mathbf{a}_t \mathbf{x} - b_t)^2$

greedy subresidual choice

Generalizing the Convergence Result

- ▷ can immediately generalize to varying β (SKM with β_k)
- ▷ to generalize to non-normalized A, we need
  - a sampling distribution that depends upon $\|\mathbf{a}_i\|^2$
  - a sampling distribution that depends upon $\mathbf{x}_k$

Generalized SKM sampling distribution: $p_{\mathbf{x}} : \binom{[m]}{\beta_k} \to [0, 1)$

- $t(\tau, \mathbf{x}) = \text{argmax}_{t \in \tau}(\mathbf{a}_t \mathbf{x} - b_t)^2$

greedy subresidual choice

- $p_{\mathbf{x}}(\tau_k) = \frac{\|\mathbf{a}_{t(\tau_k, \mathbf{x})}\|^2}{\sum_{\tau \in \binom{[m]}{\beta_k}} \|\mathbf{a}_{t(\tau, \mathbf{x})}\|^2}$

proportional to norm of selected row

Generalized SKM

Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $\tau_k \in \binom{[m]}{\beta_k}$ according to $p_{\mathbf{x}_{k-1}}$.
2. Choose $i_k := t(\tau_k, \mathbf{x}_{k-1})$.
3. Define $\mathbf{x}_k := \mathbf{x}_{k-1} + \frac{b_{i_k} - \mathbf{a}_{i_k}^T \mathbf{x}_{k-1}}{||\mathbf{a}_{i_k}||^2} \mathbf{a}_{i_k}$.
4. Repeat.

Generalized SKM

Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $\tau_k \in \binom{[m]}{\beta_k}$ according to $p_{\mathbf{x}_{k-1}}$.
2. Choose $i_k := t(\tau_k, \mathbf{x}_{k-1})$.
3. Define $\mathbf{x}_k := \mathbf{x}_{k-1} + \frac{b_{i_k} - \mathbf{a}_{i_k}^T \mathbf{x}_{k-1}}{||\mathbf{a}_{i_k}||^2} \mathbf{a}_{i_k}$.
4. Repeat.

▷ If $\beta_k = 1$, this is the distribution in [Strohmer, Vershynin '09].

Generalized SKM

Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $\tau_k \in \binom{[m]}{\beta_k}$ according to $p_{\mathbf{x}_{k-1}}$.
2. Choose $i_k := t(\tau_k, \mathbf{x}_{k-1})$.
3. Define $\mathbf{x}_k := \mathbf{x}_{k-1} + \frac{b_{i_k} - \mathbf{a}_{i_k}^T \mathbf{x}_{k-1}}{||\mathbf{a}_{i_k}||^2} \mathbf{a}_{i_k}$.
4. Repeat.

▷ If $\beta_k = 1$, this is the distribution in [Strohmer, Vershynin '09].

▷ If $\|\mathbf{a}_i\|^2 = 1$, this is the uniform distribution over $\binom{[m]}{\beta_k}$.

Generalized Result

Theorem (H. - Ma 2019+)

Let $\mathbf{x}^*$ denote the unique solution to the system of equations $A\mathbf{x} = \mathbf{b}$. Then generalized SKM converges at least linearly in expectation and the bound on the rate depends on the dynamic range, γ_k of the random sample of β_k rows of A, τ_k. Precisely, in the kth iteration of generalized SKM, we have

$$\mathbb{E}_{\tau_k}\|\mathbf{x}_k - \mathbf{x}^*\|^2 \leq \left(1 - \frac{\beta_k \binom{m}{\beta_k} \sigma^2_{\min}(A)}{\gamma_k m \sum_{\tau \in \binom{[m]}{\beta_k}} \|\mathbf{a}_{t(\tau, \mathbf{x}_{k-1})}\|^2}\right) \|\mathbf{x}_{k-1} - \mathbf{x}^*\|^2.$$

Generalized Result

Theorem (H. - Ma 2019+)

Let $\mathbf{x}^*$ denote the unique solution to the system of equations $A\mathbf{x} = \mathbf{b}$. Then generalized SKM converges at least linearly in expectation and the bound on the rate depends on the dynamic range, γ_k of the random sample of β_k rows of A, τ_k. Precisely, in the kth iteration of generalized SKM, we have

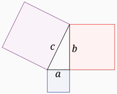
$$\mathbb{E}_{\tau_k}\|\mathbf{x}_k - \mathbf{x}^*\|^2 \leq \left(1 - \frac{\beta_k \binom{m}{\beta_k} \sigma_{\min}^2(A)}{\gamma_k m \sum_{\tau \in \binom{[m]}{\beta_k}} \|\mathbf{a}_{t(\tau, \mathbf{x}_{k-1})}\|^2}\right) \|\mathbf{x}_{k-1} - \mathbf{x}^*\|^2.$$

$\triangleright$ If all rows of A have the same norm, then

$$\mathbb{E}_{\tau_k}\|\mathbf{x}_k - \mathbf{x}^*\|^2 \leq \left(1 - \frac{\beta_k \sigma_{\min}^2(A)}{\gamma_k \|A\|_F^2}\right) \|\mathbf{x}_{k-1} - \mathbf{x}^*\|^2.$$

Sketch of Proof

$$\left\|\mathbf{x}_k - \mathbf{x}^*\right\|^2 = \left\|\mathbf{x}_{k-1} - \mathbf{x}^*\right\|^2 - \frac{\left\|A_{\tau_k}\mathbf{x}_{k-1} - \mathbf{b}_{\tau_k}\right\|_\infty^2}{\left\|\mathbf{a}_{t(\tau_k, \mathbf{x}_{k-1})}\right\|^2}$$

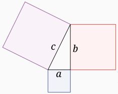


[1]

[1] Originally created by en:User:Michael Hardy, then scaled, with colour and labels being added by en:User:Wapcaplet, transformed in svg format by fr:Utilisateur:Steff, changed colors and font by de:Leo2004. (https://commons.wikimedia.org/wiki/File:Pythagorean.svg)

Sketch of Proof

$$\left\|\mathbf{x}_k - \mathbf{x}^*\right\|^2 = \left\|\mathbf{x}_{k-1} - \mathbf{x}^*\right\|^2 - \frac{\left\|A_{\tau_k}\mathbf{x}_{k-1} - \mathbf{b}_{\tau_k}\right\|_\infty^2}{\left\|\mathbf{a}_{t(\tau_k,\mathbf{x}_{k-1})}\right\|^2}$$



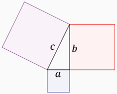
[1]

$$\mathbb{E}_{\tau_k}\frac{\left\|A_{\tau_k}\mathbf{x}_{k-1} - \mathbf{b}_{\tau_k}\right\|_\infty^2}{\left\|\mathbf{a}_{t(\tau_k,\mathbf{x}_{k-1})}\right\|^2} = \sum_{\tau\in\binom{[m]}{\beta_k}} \frac{\left\|\mathbf{a}_{t(\tau,\mathbf{x}_{k-1})}\right\|^2}{\sum_{\pi\in\binom{[m]}{\beta_k}}\left\|\mathbf{a}_{t(\pi,\mathbf{x}_{k-1})}\right\|^2}\frac{\left\|A_{\tau}\mathbf{x}_{k-1} - \mathbf{b}_{\tau}\right\|_\infty^2}{\left\|\mathbf{a}_{t(\tau,\mathbf{x}_{k-1})}\right\|^2}$$

[1] Originally created by en:User:Michael Hardy, then scaled, with colour and labels being added by en:User:Wapcaplet, transformed in svg format by fr:Utilisateur:Steff, changed colors and font by de:Leo2004. (https://commons.wikimedia.org/wiki/File:Pythagorean.svg)

Sketch of Proof

$$\left\|\mathbf{x}_k - \mathbf{x}^*\right\|^2 = \left\|\mathbf{x}_{k-1} - \mathbf{x}^*\right\|^2 - \frac{\left\|A_{\tau_k}\mathbf{x}_{k-1} - \mathbf{b}_{\tau_k}\right\|_\infty^2}{\left\|\mathbf{a}_{t(\tau_k,\mathbf{x}_{k-1})}\right\|^2}$$



[1]

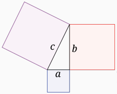
$$\mathbb{E}_{\tau_k}\frac{\left\|A_{\tau_k}\mathbf{x}_{k-1} - \mathbf{b}_{\tau_k}\right\|_\infty^2}{\left\|\mathbf{a}_{t(\tau_k,\mathbf{x}_{k-1})}\right\|^2} = \sum_{\tau\in\binom{[m]}{\beta_k}} \frac{\left\|\mathbf{a}_{t(\tau,\mathbf{x}_{k-1})}\right\|^2}{\sum_{\pi\in\binom{[m]}{\beta_k}}\left\|\mathbf{a}_{t(\pi,\mathbf{x}_{k-1})}\right\|^2} \frac{\left\|A_{\tau}\mathbf{x}_{k-1} - \mathbf{b}_{\tau}\right\|_\infty^2}{\left\|\mathbf{a}_{t(\tau,\mathbf{x}_{k-1})}\right\|^2}$$

$$= \frac{1}{\sum_{\pi\in\binom{[m]}{\beta_k}}\left\|\mathbf{a}_{t(\pi,\mathbf{x}_{k-1})}\right\|^2} \sum_{\tau\in\binom{[m]}{\beta_k}} \left\|A_{\tau}\mathbf{x}_{k-1} - \mathbf{b}_{\tau}\right\|_\infty^2$$

[1] Originally created by en:User:Michael Hardy, then scaled, with colour and labels being added by en:User:Wapcaplet, transformed in svg format by fr:Utilisateur:Steff, changed colors and font by de:Leo2004. (https://commons.wikimedia.org/wiki/File:Pythagorean.svg)

Sketch of Proof

$$\left\|\mathbf{x}_k - \mathbf{x}^*\right\|^2 = \left\|\mathbf{x}_{k-1} - \mathbf{x}^*\right\|^2 - \frac{\left\|A_{\tau_k}\mathbf{x}_{k-1} - \mathbf{b}_{\tau_k}\right\|_\infty^2}{\left\|\mathbf{a}_{t(\tau_k,\mathbf{x}_{k-1})}\right\|^2}$$



[1]

$$\begin{aligned}
\mathbb{E}_{\tau_k}\frac{\left\|A_{\tau_k}\mathbf{x}_{k-1} - \mathbf{b}_{\tau_k}\right\|_\infty^2}{\left\|\mathbf{a}_{t(\tau_k,\mathbf{x}_{k-1})}\right\|^2} &= \sum_{\tau\in\binom{[m]}{\beta_k}} \frac{\left\|\mathbf{a}_{t(\tau,\mathbf{x}_{k-1})}\right\|^2}{\sum_{\pi\in\binom{[m]}{\beta_k}}\left\|\mathbf{a}_{t(\pi,\mathbf{x}_{k-1})}\right\|^2} \frac{\left\|A_{\tau}\mathbf{x}_{k-1} - \mathbf{b}_{\tau}\right\|_\infty^2}{\left\|\mathbf{a}_{t(\tau,\mathbf{x}_{k-1})}\right\|^2} \\
&= \frac{1}{\sum_{\pi\in\binom{[m]}{\beta_k}}\left\|\mathbf{a}_{t(\pi,\mathbf{x}_{k-1})}\right\|^2} \sum_{\tau\in\binom{[m]}{\beta_k}} \left\|A_{\tau}\mathbf{x}_{k-1} - \mathbf{b}_{\tau}\right\|_\infty^2 \\
&= \frac{\binom{m}{\beta_k}\beta_k}{\gamma_k m \sum_{\pi\in\binom{[m]}{\beta_k}}\left\|\mathbf{a}_{t(\pi,\mathbf{x}_{k-1})}\right\|^2}\left\|A\mathbf{x}_{k-1} - \mathbf{b}\right\|^2
\end{aligned}$$

[1] Originally created by en:User:Michael Hardy, then scaled, with colour and labels being added by en:User:Wapcaplet, transformed in svg format by fr:Utilisateur:Steff, changed colors and font by de:Leo2004. (https://commons.wikimedia.org/wiki/File:Pythagorean.svg)

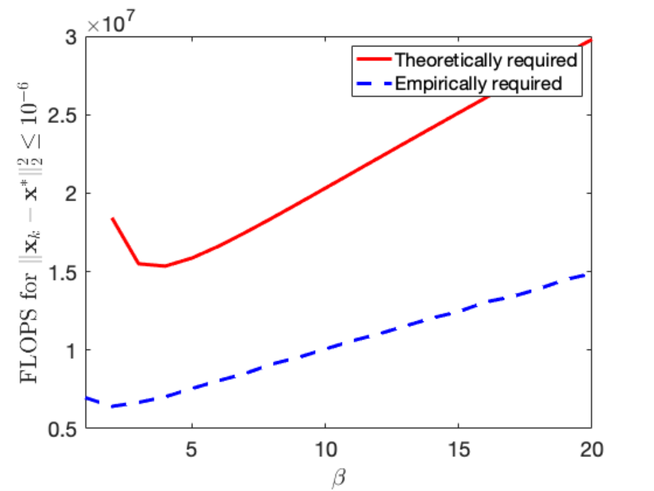
Now can we determine the optimal β?

Now can we determine the optimal β?

Roughly, if we know the value of γ_j.

Now can we determine the optimal β?

Roughly, if we know the value of γ_j.



Conclusions and Future Work

Conclusions and Future Work

- presented a convergence rate for SKM which generalizes (and improves) results for RK and Motzkin

$$\mathbb{E}_{\tau_k}\|\mathbf{x}_k - \mathbf{x}^*\|^2 \leq \left(1 - \frac{\beta_k \binom{m}{\beta_k} \sigma^2_{\min}(A)}{\gamma_k m \sum_{\tau \in \binom{[m]}{\beta_k}} \|\mathbf{a}_{t(\tau, \mathbf{x}_{k-1})}\|^2}\right) \|\mathbf{x}_{k-1} - \mathbf{x}^*\|^2$$

Conclusions and Future Work

- presented a convergence rate for SKM which generalizes (and improves) results for RK and Motzkin

$$\mathbb{E}_{\tau_k}\|\mathbf{x}_k - \mathbf{x}^*\|^2 \leq \left(1 - \frac{\beta_k \binom{m}{\beta_k} \sigma_{\min}^2(A)}{\gamma_k m \sum_{\tau \in \binom{[m]}{\beta_k}} \|\mathbf{a}_{t(\tau,\mathbf{x}_{k-1})}\|^2}\right)\|\mathbf{x}_{k-1} - \mathbf{x}^*\|^2$$

- proved a result which “easily” yields results which illustrate SKM improvement for specific types of measurement matrices

Conclusions and Future Work

- presented a convergence rate for SKM which generalizes (and improves) results for RK and Motzkin

$$\mathbb{E}_{\tau_k}\|\mathbf{x}_k - \mathbf{x}^*\|^2 \leq \left(1 - \frac{\beta_k \binom{m}{\beta_k} \sigma^2_{\min}(A)}{\gamma_k m \sum_{\tau \in \binom{[m]}{\beta_k}} \|\mathbf{a}_{t(\tau, \mathbf{x}_{k-1})}\|^2}\right) \|\mathbf{x}_{k-1} - \mathbf{x}^*\|^2$$

- proved a result which “easily” yields results which illustrate SKM improvement for specific types of measurement matrices
- specialized our result to Gaussian matrices

Conclusions and Future Work

- presented a convergence rate for SKM which generalizes (and improves) results for RK and Motzkin

$$\mathbb{E}_{\tau_k}\|\mathbf{x}_k - \mathbf{x}^*\|^2 \leq \left(1 - \frac{\beta_k \binom{m}{\beta_k} \sigma_{\min}^2(A)}{\gamma_k m \sum_{\tau \in \binom{[m]}{\beta_k}} \|\mathbf{a}_{t(\tau,\mathbf{x}_{k-1})}\|^2}\right)\|\mathbf{x}_{k-1} - \mathbf{x}^*\|^2$$

- proved a result which “easily” yields results which illustrate SKM improvement for specific types of measurement matrices
- specialized our result to Gaussian matrices

- identify useful bounds on γ_k for other useful systems

Conclusions and Future Work

- ▷ presented a convergence rate for SKM which generalizes (and improves) results for RK and Motzkin

$$\mathbb{E}_{\tau_k}\|\mathbf{x}_k - \mathbf{x}^*\|^2 \leq \left(1 - \frac{\beta_k \binom{m}{\beta_k} \sigma^2_{\min}(A)}{\gamma_k m \sum_{\tau \in \binom{[m]}{\beta_k}} \|\mathbf{a}_{t(\tau,\mathbf{x}_{k-1})}\|^2}\right) \|\mathbf{x}_{k-1} - \mathbf{x}^*\|^2$$

- ▷ proved a result which "easily" yields results which illustrate SKM improvement for specific types of measurement matrices
- ▷ specialized our result to Gaussian matrices

- ▷ identify useful bounds on γ_k for other useful systems
- ▷ identify optimal β of systems for which γ_k is known

Thanks for listening!

Questions?

[1] J. A. De Loera, J. Haddock, and D. Needell. **A sampling Kaczmarz-Motzkin algorithm for linear feasibility.** SIAM Journal on Scientific Computing, 39(5):S66–S87, 2017.

[2] J. Haddock and D. Needell. **On Motzkins method for inconsistent linear systems.** BIT Numerical Mathematics, 59(2):387–401, 2019.

[3] Nicolas Loizou and Peter Richtárik. **Revisiting randomized gossip algorithms: General framework, convergence rates and novel block and accelerated protocols.** arXiv preprint arXiv:1905.08645, 2019.

[4] T. S. Motzkin and I. J. Schoenberg. **The relaxation method for linear inequalities.** Canadian J. Math., 6:393–404, 1954.

[5] T. Strohmer and R. Vershynin. **A randomized Kaczmarz algorithm with exponential convergence.** J. Fourier Anal. Appl., 15:262–278, 2009.