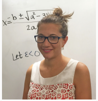
Iterative Projection Methods

for noisy and corrupted systems of linear equations

Jamie Haddock

Tulane Probability and Statistics Seminar, November 7, 2018

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 2018+)
https://arxiv.org/abs/1803.08114
https://arxiv.org/abs/1605.01418 (SISC 2017)

BIG Data

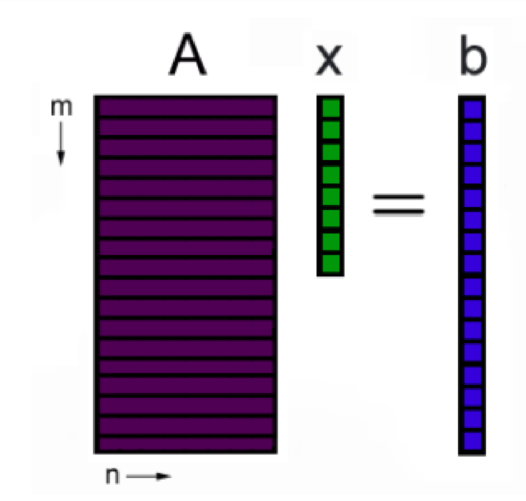

The big data opportunity
LET'S SOLVE THIS PROBLEM BY USING THE BIG DATA NONE OF US HAVE THE SLIGHTEST IDEA WHAT TO DO WITH
TOM FISH BURNE
© marketoonist.com
DellWorld'15

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), $A\mathbf{x} = \mathbf{b}$ ($A\mathbf{x} \le \mathbf{b}$), 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 an element:

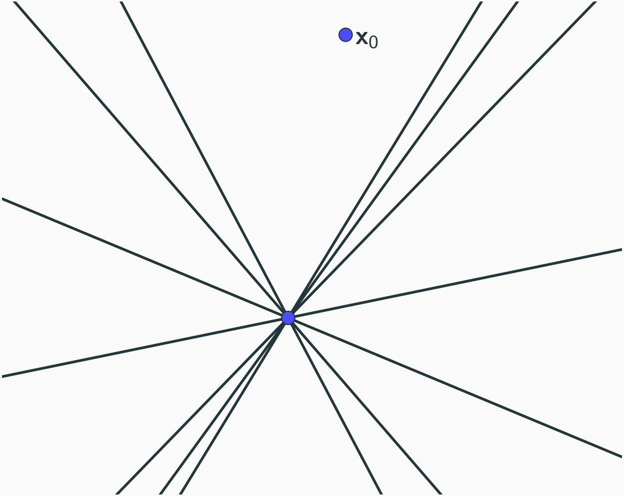
1. Randomized Kaczmarz Method
2. Motzkin's Method
3. Sampling Kaczmarz-Motzkin Methods (SKM)

Randomized Kaczmarz Method

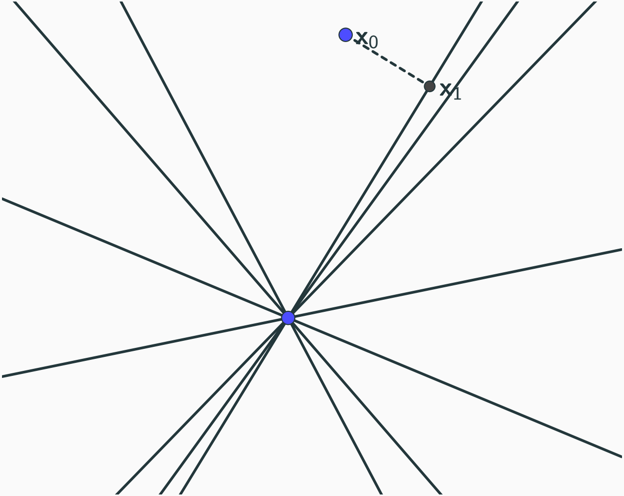
Given $\mathbf{x}_0 \in \mathbb{R}^n$:

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.

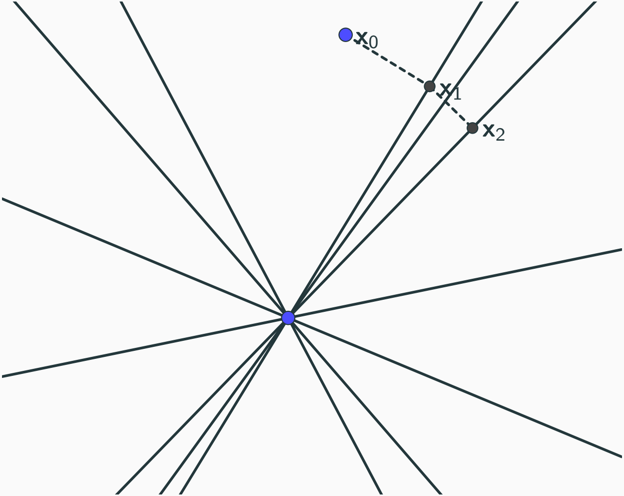
Kaczmarz Method



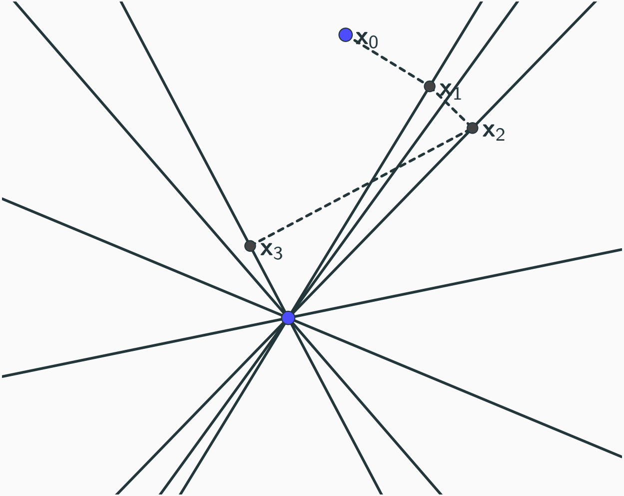
Kaczmarz Method



Kaczmarz Method



Kaczmarz Method



Convergence Rate

Theorem (Strohmer - Vershynin 2009)
Let $\mathbf{x}$ *be the solution to the consistent system of linear equations* $A\mathbf{x} = \mathbf{b}$*. Then the Random Kaczmarz method converges to* $\mathbf{x}$ *linearly in expectation:*

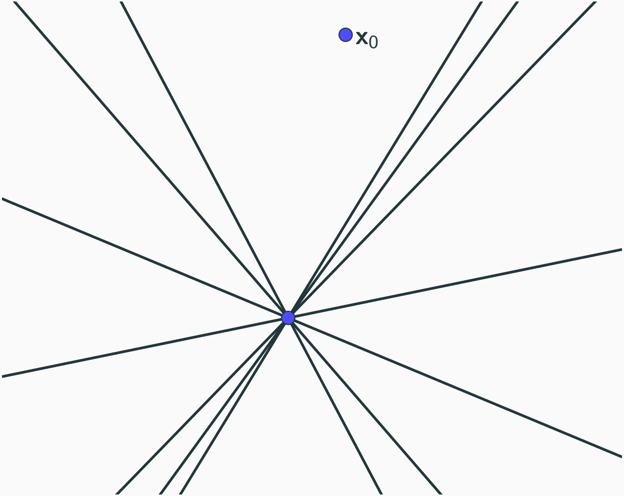
$$\mathbb{E}||\mathbf{x}_k - \mathbf{x}||_2^2 \leq \left(1 - \frac{1}{||A||_F^2 ||A^{-1}||_2^2}\right)^k ||\mathbf{x}_0 - \mathbf{x}||_2^2.$$

Motzkin's Relaxation Method

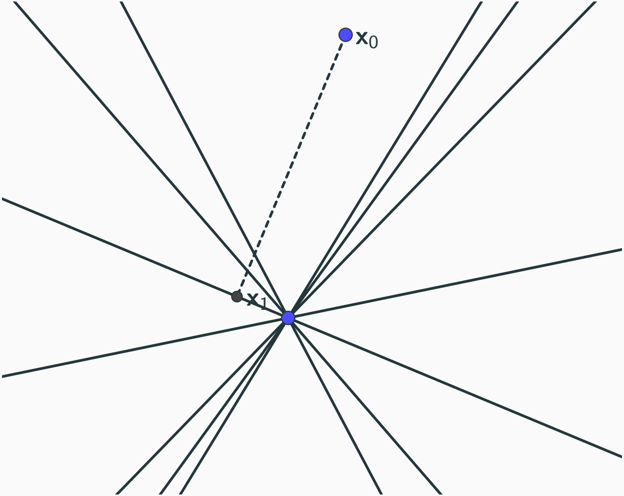
Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. If $\mathbf{x}_k$ is feasible, stop.
2. Choose $i_k \in [m]$ as $i_k := \underset{i \in [m]}{\operatorname{argmax}} |\mathbf{a}_i^T \mathbf{x}_{k-1} - b_i|$.
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.

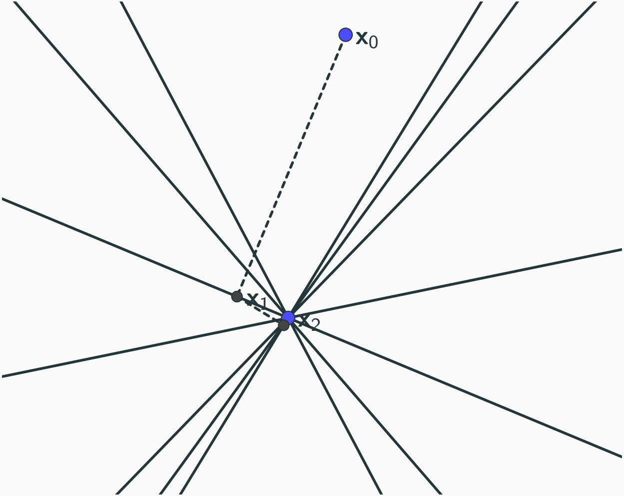
Motzkin's Method



Motzkin's Method



Motzkin's Method



Convergence Rate

Theorem (Agmon 1954)

For a consistent, normalized system, $\|\mathbf{a}_i\| = 1$ *for all* $i = 1, ..., m$, *Motzkin's method converges linearly to the solution* $\mathbf{x}$*:*

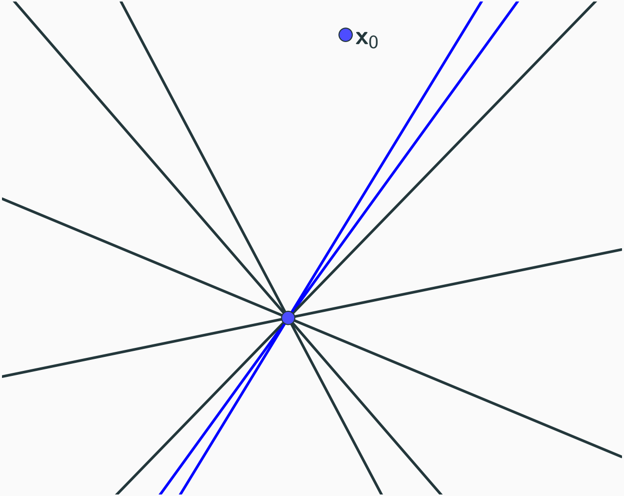
$$\|\mathbf{x}_k - \mathbf{x}\|^2 \leq \left(1 - \frac{1}{m\|A^{-1}\|^2}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|^2.$$

Our Hybrid Method (SKM)

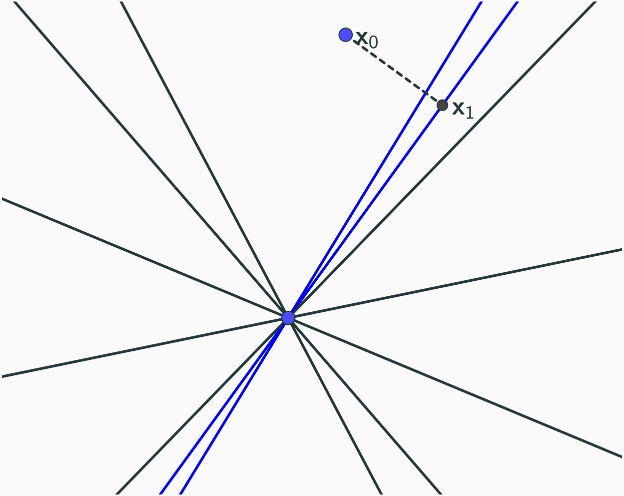
Given $\mathbf{x}_0 \in \mathbb{R}^n$:

1. Choose $\tau_k \subset [m]$ to be a sample of size β constraints chosen uniformly at random from among the rows of A.
2. From among these 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.

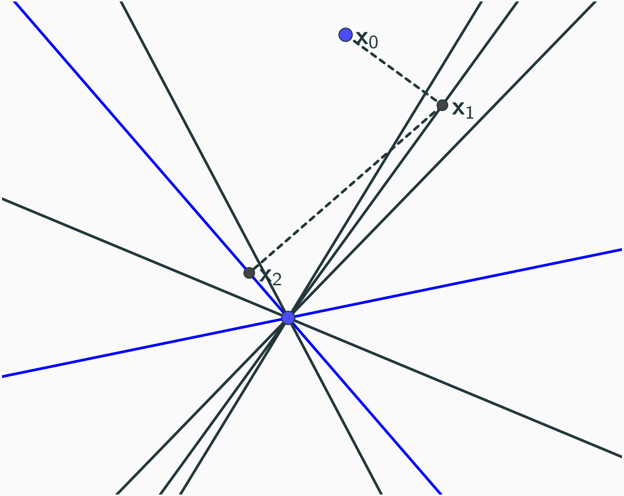
SKM



SKM



SKM



SKM Method Convergence Rate

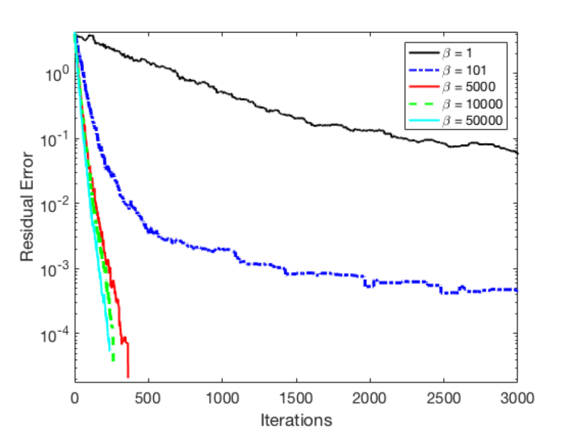
Theorem (De Loera - H. - Needell 2017)

For a consistent, normalized system the SKM method with samples of size β *converges to the solution* x *at least linearly in expectation: If* s_{k-1} *is the number of constraints satisfied by* x_{k-1} *and* $V_{k-1} = \max\{m - s_{k-1}, m - \beta + 1\}$ *then*

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

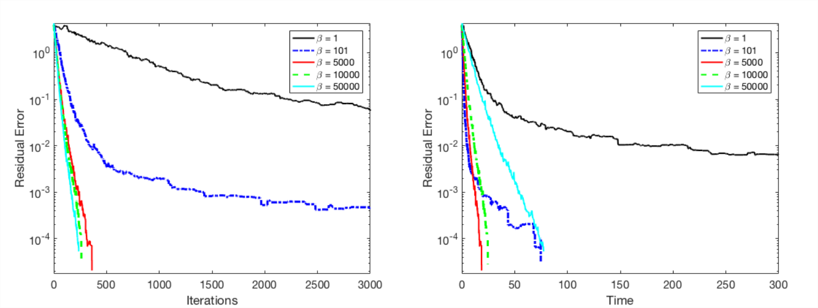
$$\leq \left(1 - \frac{1}{m\|A^{-1}\|^2}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|^2.$$

Experimental Convergence



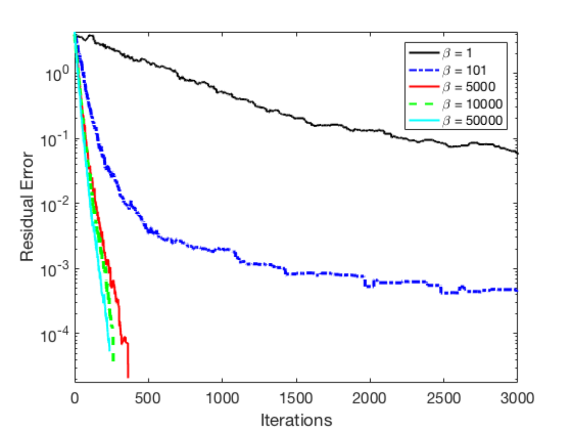
- β: sample size
- A is 50000×100 Gaussian matrix, consistent system
- 'faster' convergence for larger sample size

Experimental Convergence



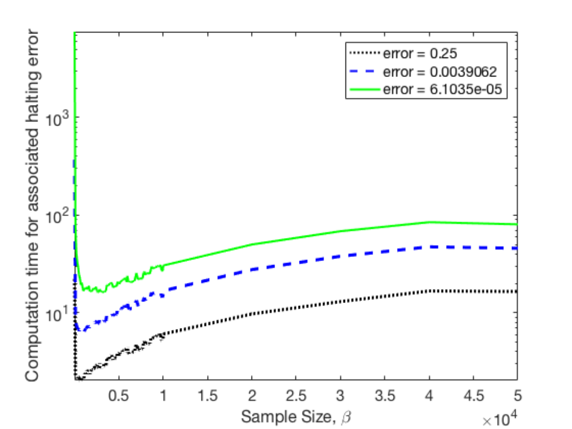
- β: sample size
- A is 50000×100 Gaussian matrix, consistent system
- 'faster' convergence for larger sample size

Experimental Convergence



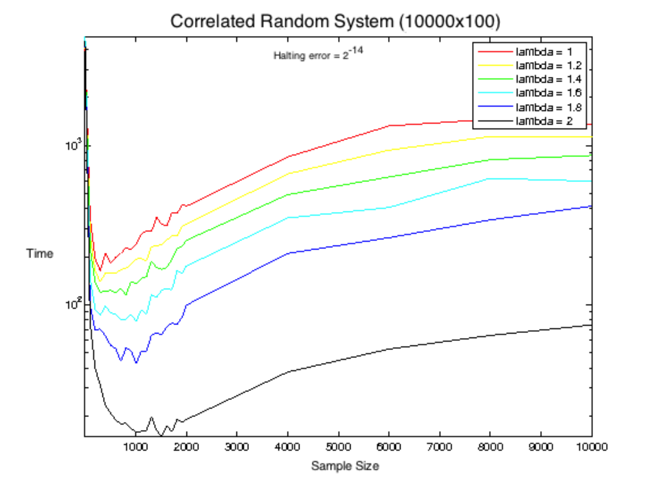
- β: sample size
- A is 50000×100 Gaussian matrix, consistent system
- 'faster' convergence for larger sample size

Experimental Convergence



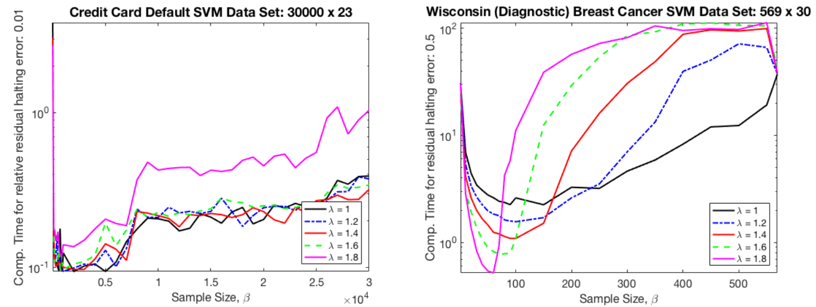
- β: sample size
- A is 50000×100 Gaussian matrix, consistent system
- 'faster' convergence for larger sample size

Experimental Convergence



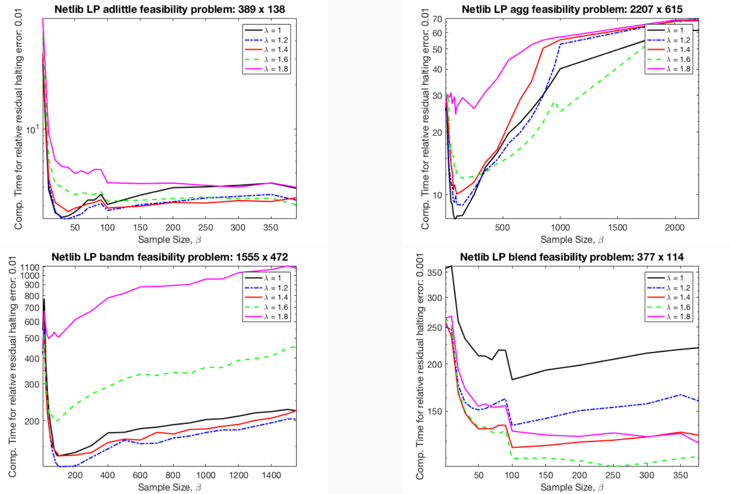
▷ A is 10000×100 "correlated" matrix, consistent system

Experimental Convergence



- SVM linear feasibility problem $A\mathbf{x} \leq \mathbf{b}$

Experimental Convergence



- ▷ LP linear feasibility problem $A\mathbf{x} \leq \mathbf{b}$

Convergence Rates

- RK: $\mathbb{E}||\mathbf{x}_k - \mathbf{x}||_2^2 \leq \left(1 - \frac{1}{||A||_F^2 ||A^{-1}||_2^2}\right)^k ||\mathbf{x}_0 - \mathbf{x}||_2^2.$

Convergence Rates

- RK: $\mathbb{E}||\mathbf{x}_k - \mathbf{x}||_2^2 \leq \left(1 - \frac{1}{||A||_F^2||A^{-1}||_2^2}\right)^k ||\mathbf{x}_0 - \mathbf{x}||_2^2.$
- MM: $\|\mathbf{x}_k - \mathbf{x}\|^2 \leq \left(1 - \frac{1}{m\|A^{-1}\|^2}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|^2.$

Convergence Rates

- RK: $\mathbb{E}||\mathbf{x}_k - \mathbf{x}||_2^2 \leq \left(1 - \frac{1}{||A||_F^2||A^{-1}||_2^2}\right)^k ||\mathbf{x}_0 - \mathbf{x}||_2^2.$
- MM: $\|\mathbf{x}_k - \mathbf{x}\|^2 \leq \left(1 - \frac{1}{m\|A^{-1}\|^2}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|^2.$
- SKM: $\mathbb{E}\|\mathbf{x}_k - \mathbf{x}\|^2 \leq \left(1 - \frac{1}{m\|A^{-1}\|^2}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|^2.$

Convergence Rates

- RK: $\mathbb{E}||\mathbf{x}_k - \mathbf{x}||_2^2 \leq \left(1 - \frac{1}{||A||_F^2||A^{-1}||_2^2}\right)^k ||\mathbf{x}_0 - \mathbf{x}||_2^2.$
- MM: $\|\mathbf{x}_k - \mathbf{x}\|^2 \leq \left(1 - \frac{1}{m\|A^{-1}\|^2}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|^2.$
- SKM: $\mathbb{E}\|\mathbf{x}_k - \mathbf{x}\|^2 \leq \left(1 - \frac{1}{m\|A^{-1}\|^2}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|^2.$

- Why are these all the same?

An Accelerated Convergence Rate

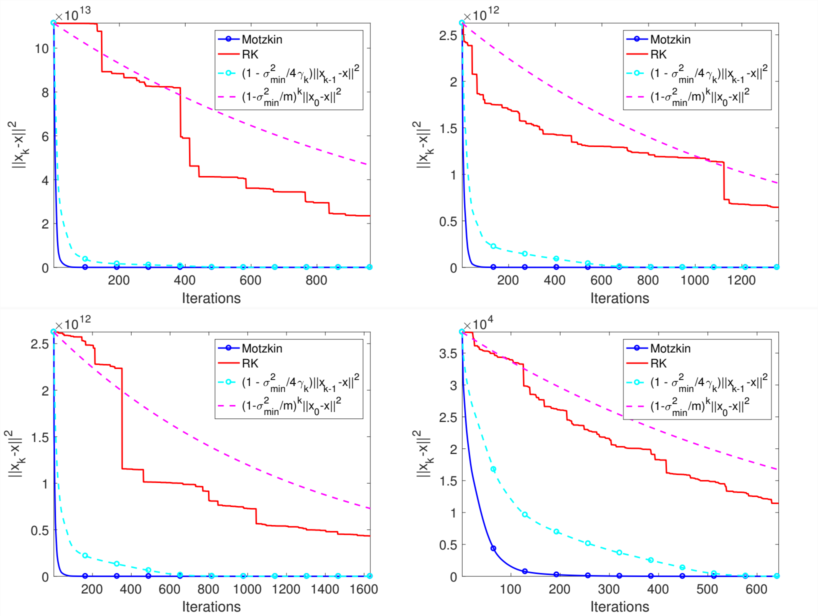
Theorem (H. - Needell 2018+)
Let $\mathbf{x}$ denote the solution of the consistent, normalized system $A\mathbf{x} = \mathbf{b}$. Motzkin's method exhibits the (possibly highly accelerated) convergence rate:

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

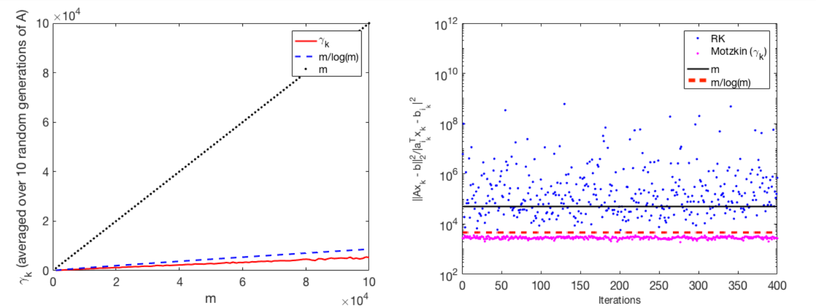
Here γ_k bounds the dynamic range of the kth residual, $\gamma_k := \frac{\|A\mathbf{x}_k - A\mathbf{x}\|^2}{\|A\mathbf{x}_k - A\mathbf{x}\|_\infty^2}$.

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

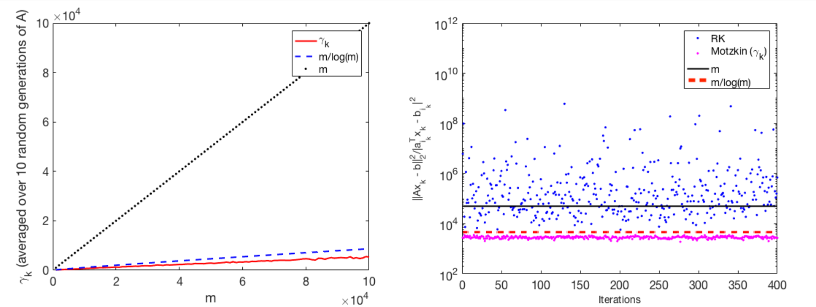
Netlib LP Systems



γ_k: Gaussian systems

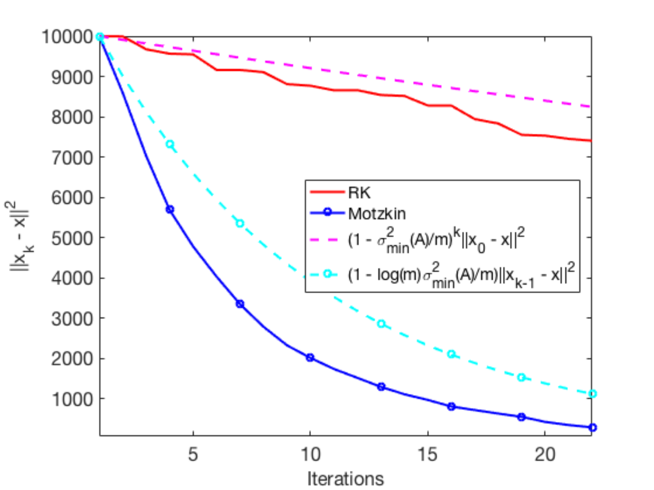


γ_k: Gaussian systems



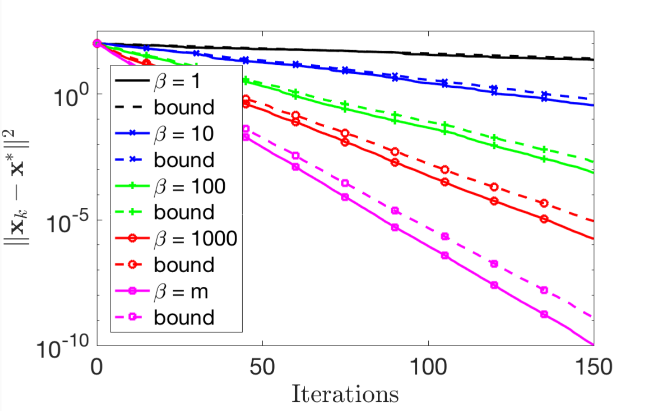
$$\gamma_k \lesssim \frac{nm}{\log m}$$

Gaussian Convergence



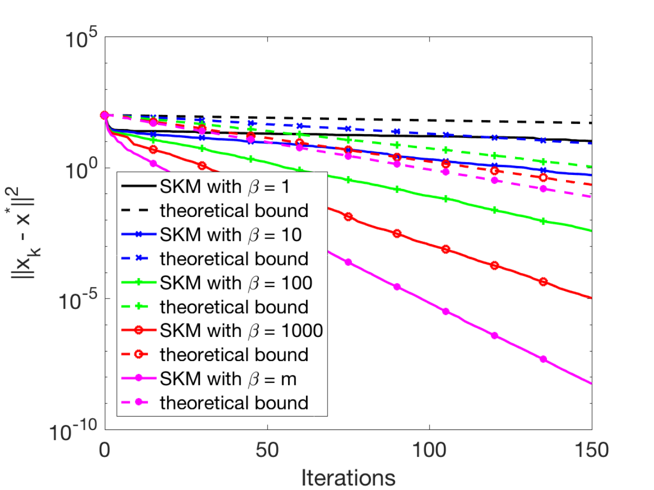
▷ A is 50000×100 Gaussian matrix, consistent system

Extending to SKM



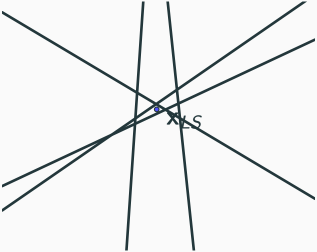
- A is 50000×100 Gaussian matrix, consistent system
- bound uses dynamic range of sample of β rows
- use this bound to design methods which identify optimal β?

Extending to SKM



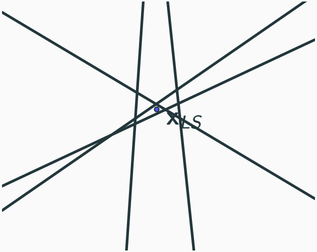
- A is 50000×100 "correlated" matrix, consistent system
- bound uses dynamic range of sample of β rows

Is this the right problem?



◁ noisy

Is this the right problem?



◁ noisy

▷ corrupted

Noisy Convergence Results

Theorem (Needell 2010)

Let A have full column rank, denote the desired solution to the system $A\mathbf{x} = \mathbf{b}$ by $\mathbf{x}$, and define the error term $\mathbf{e} = A\mathbf{x} - \mathbf{b}$. Then RK iterates satisfy

$$\mathbb{E}\|\mathbf{x}_k - \mathbf{x}\|^2 \leq \left(1 - \frac{1}{\|A\|_F^2 \|A^{-1}\|^2}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|^2 + \|A\|_F^2 \|A^{-1}\|^2 \|\mathbf{e}\|_\infty^2$$

Noisy Convergence Results

Theorem (Needell 2010)

Let A have full column rank, denote the desired solution to the system $A\mathbf{x} = \mathbf{b}$ by $\mathbf{x}$, and define the error term $\mathbf{e} = A\mathbf{x} - \mathbf{b}$. Then RK iterates satisfy

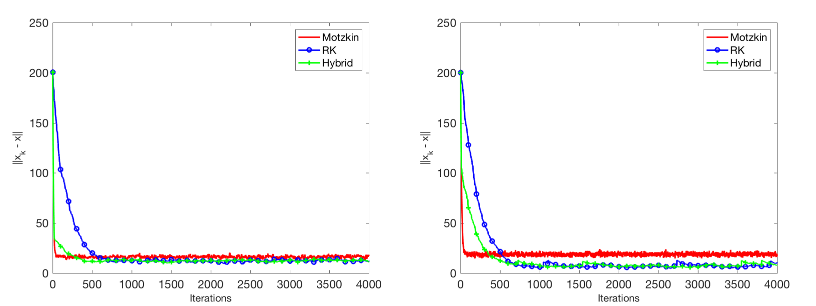
$$\mathbb{E}\|\mathbf{x}_k - \mathbf{x}\|^2 \leq \left(1 - \frac{1}{\|A\|_F^2\|A^{-1}\|^2}\right)^k \|\mathbf{x}_0 - \mathbf{x}\|^2 + \|A\|_F^2\|A^{-1}\|^2\|\mathbf{e}\|_\infty^2$$

Theorem (H. - Needell 2018+)

Let $\mathbf{x}$ denote the desired solution of the system $A\mathbf{x} = \mathbf{b}$ and define the error term $\mathbf{e} = \mathbf{b} - A\mathbf{x}$. If Motzkin's method is run with stopping criterion $\|A\mathbf{x}_k - \mathbf{b}\|_\infty \leq 4\|\mathbf{e}\|_\infty$, then the iterates satisfy

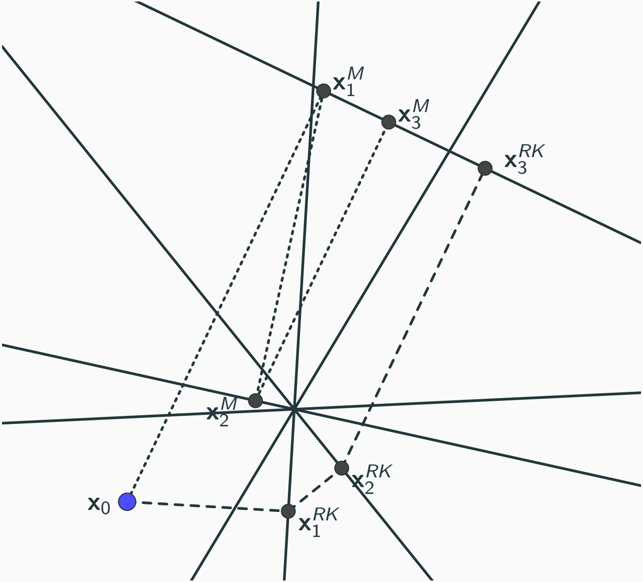
$$\|\mathbf{x}_T - \mathbf{x}\|^2 \leq \prod_{k=0}^{T-1}\left(1 - \frac{1}{4\gamma_k\|A^{-1}\|^2}\right) \cdot \|\mathbf{x}_0 - \mathbf{x}\|^2 + 2m\|A^{-1}\|^2\|\mathbf{e}\|_\infty^2$$

Noisy Convergence



- A is 50000×100 Gaussian matrix, inconsistent system ($A\mathbf{x} = \mathbf{b} + \mathbf{e}$)
- Left: Gaussian error $\mathbf{e}$
- Right: sparse, 'spiky' error $\mathbf{e}$
- Motzkin suffers from a worse 'convergence horizon' if $\mathbf{e}$ is sparse

What about corruption?



Problem

	Problem:	$A\mathbf{x} = \mathbf{b} + \mathbf{e}$
(Corrupted)	**Error (e):**	sparse, arbitrarily large entries
	Solution ($\mathbf{x}^*$):	$\mathbf{x}^* \in \{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$

Problem

	Problem:	$A\mathbf{x} = \mathbf{b} + \mathbf{e}$
(Corrupted)	**Error (e):**	sparse, arbitrarily large entries
	Solution ($\mathbf{x}^*$):	$\mathbf{x}^* \in \{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$

Applications: logic programming, error correction in telecommunications

Problem

	Problem:	$A\mathbf{x} = \mathbf{b} + \mathbf{e}$
(Corrupted)	**Error (e):**	sparse, arbitrarily large entries
	Solution ($\mathbf{x}^*$):	$\mathbf{x}^* \in \{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$

Applications: logic programming, error correction in telecommunications

	Problem:	$A\mathbf{x} = \mathbf{b} + \mathbf{e}$
~~**(Noisy)**~~	**Error (e):**	small, evenly distributed entries
	Solution ($\mathbf{x}_{\mathbf{LS}}$):	$\mathbf{x}_{\text{LS}} \in \operatorname{argmin} \|A\mathbf{x} - \mathbf{b} - \mathbf{e}\|^2$

Why not least-squares?

MAX-FS: Given $A\mathbf{x} = \mathbf{b}$, determine the largest feasible subsystem.

`MAX-FS`: Given $A\mathbf{x} = \mathbf{b}$, determine the largest feasible subsystem.

- `MAX-FS` is NP-hard even when restricted to homogenous systems with coefficients in $\{-1, 0, 1\}$ (Amaldi - Kann 1995)

MAX-FS

MAX-FS: Given $A\mathbf{x} = \mathbf{b}$, determine the largest feasible subsystem.

- MAX-FS is NP-hard even when restricted to homogenous systems with coefficients in $\{-1, 0, 1\}$ (Amaldi - Kann 1995)
- no PTAS unless $P = NP$

Proposed Method

Goal: Use RK to detect the corrupted equations with high probability.

Proposed Method

Goal: Use RK to detect the corrupted equations with high probability.

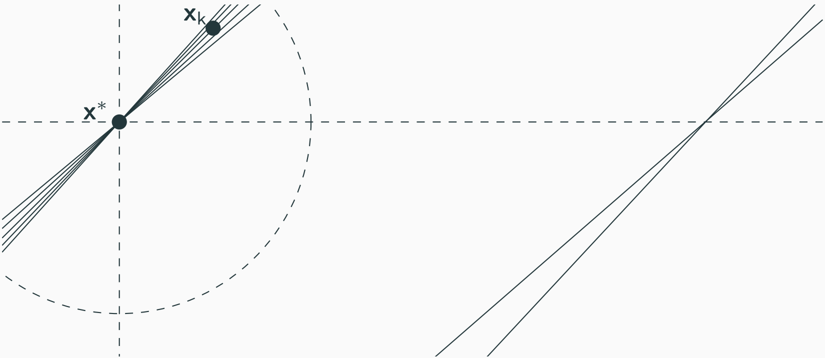
Lemma

Let $\epsilon^* = \min_{i \in supp(\mathbf{e})} |A\mathbf{x}^* - \mathbf{b}|_i = |e_i|$ *and suppose* $|supp(\mathbf{e})| = s$*. If* $||\mathbf{a}_i|| = 1$ *for* $i \in [m]$ *and* $||\mathbf{x} - \mathbf{x}^*|| < \frac{1}{2}\epsilon^*$ *we have that the* $d \leq s$ *indices of largest magnitude residual entries are contained in* $supp(\mathbf{e})$*. That is, we have* $D \subset supp(\mathbf{e})$*, where*

$$D = \underset{D \subset [A], |D| = d}{argmax} \sum_{i \in D} |A\mathbf{x} - \mathbf{b}|_i.$$

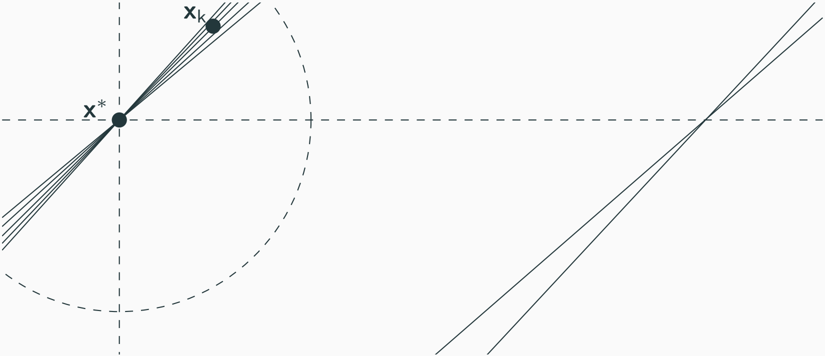
Proposed Method

Goal: Use RK to detect the corrupted equations with high probability.



Proposed Method

Goal: Use RK to detect the corrupted equations with high probability.



We call $\epsilon^*/2$ the *detection horizon*.

Proposed Method

Method 1 Multiple Round Kaczmarz

procedure $\mathrm{MRK}(A, \mathbf{b}, k, W, d)$

  $S = \emptyset$

  for $i = 1, 2, \ldots W$ **do**

    $\mathbf{x}_k^i = k$th iterate produced by RK with $\mathbf{x}_0 = \mathbf{0}$, A, $\mathbf{b}$.

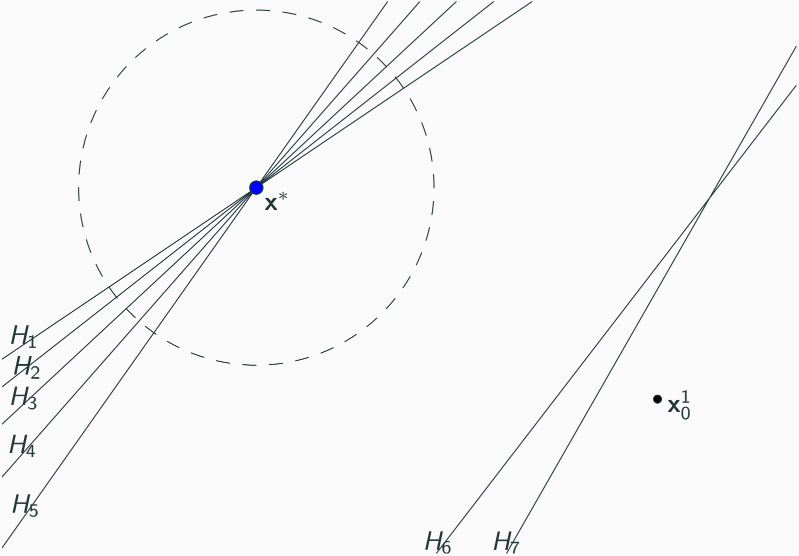
    $D = d$ indices of the largest entries of the residual, $|A\mathbf{x}_k^i - \mathbf{b}|$.

    $S = S \cup D$

  return $\mathbf{x}$, where $A_{S^C}\mathbf{x} = \mathbf{b}_{S^C}$

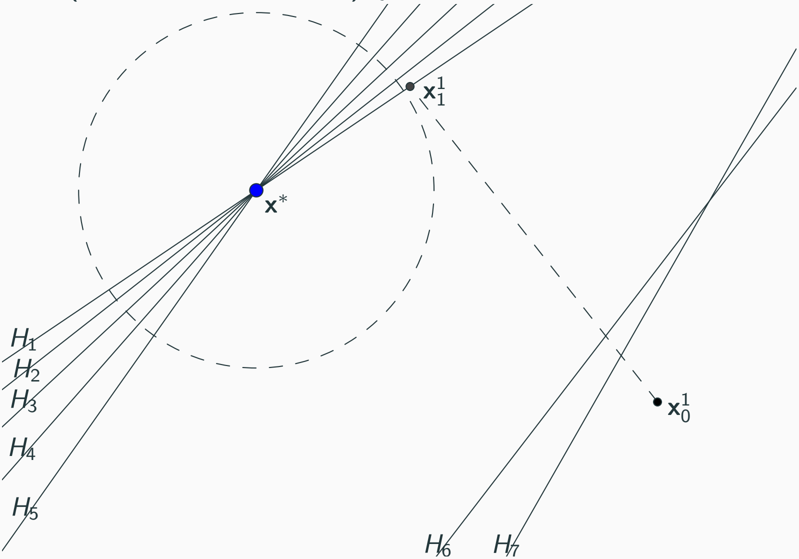
Example

MRK(A,b,$k=2$,$W=3$,$d=1$): $j=1$, $i=1$, $S=\emptyset$



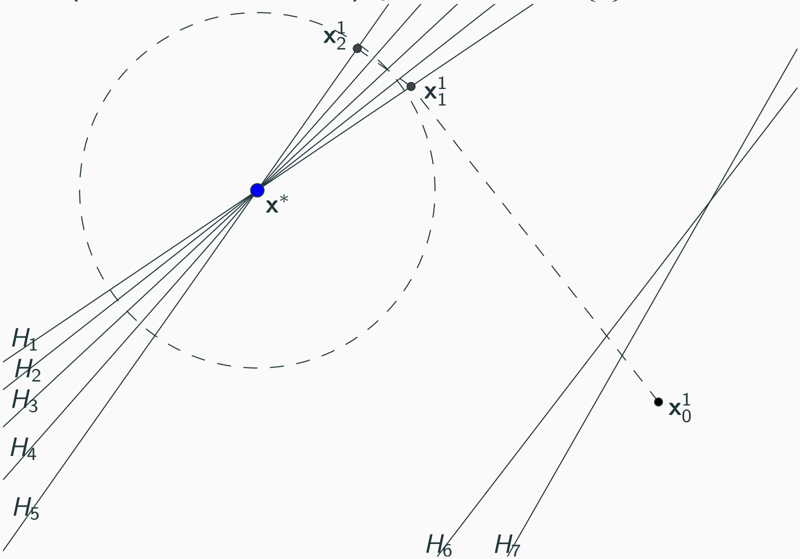
Example

MRK(A,b,$k=2$,$W=3$,$d=1$): $j=1$, $i=1$, $S=\emptyset$

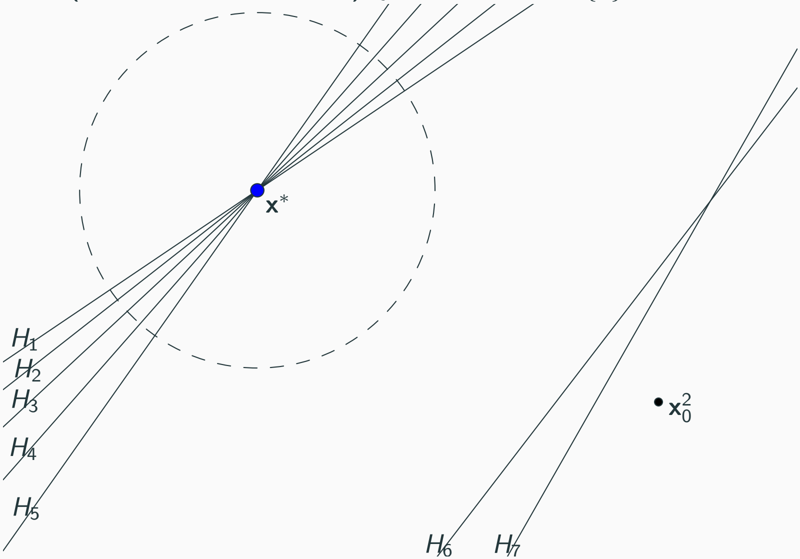


Example

MRK(A**,b,**$k=2$**,**$W=3$**,**$d=1$**):** $j = 2,\ i = 1,\ S = \{7\}$

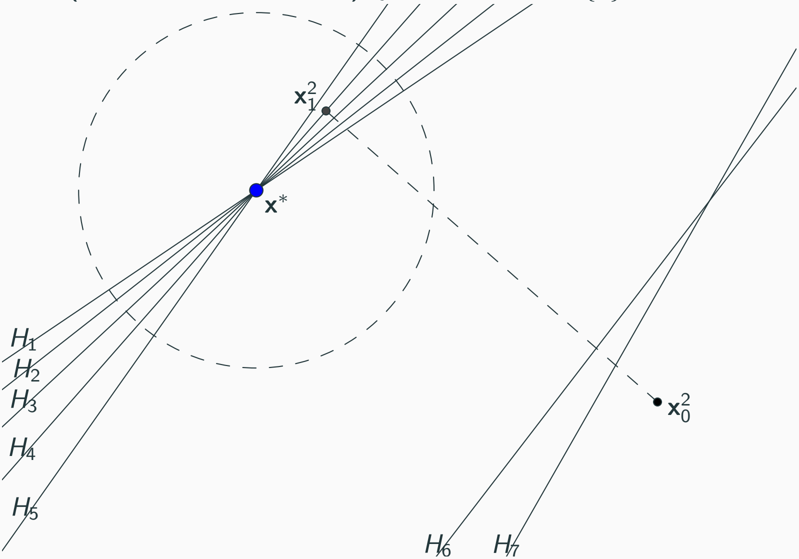


MRK(A**,b,**$k=2$**,**$W=3$**,**$d=1$**):** $j=1$, $i=2$, $S=\{7\}$



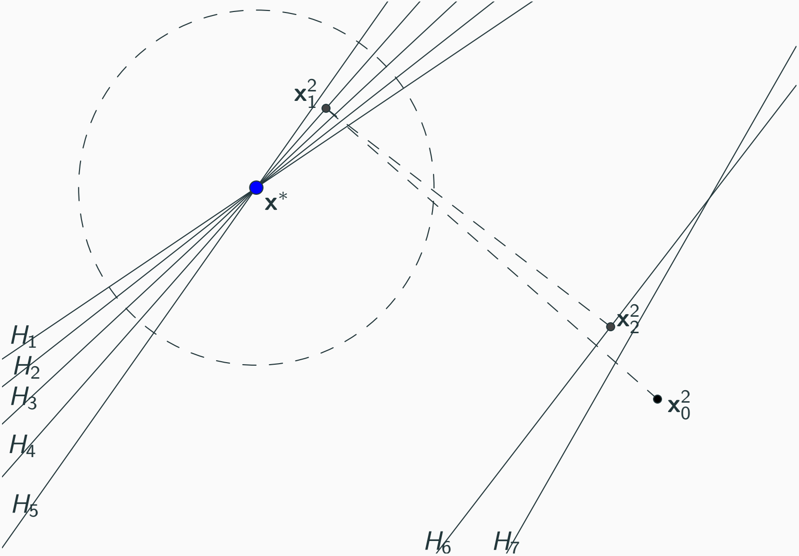
Example

MRK(A,b,$k = 2$,$W = 3$,$d = 1$): $j = 1$, $i = 2$, $S = \{7\}$



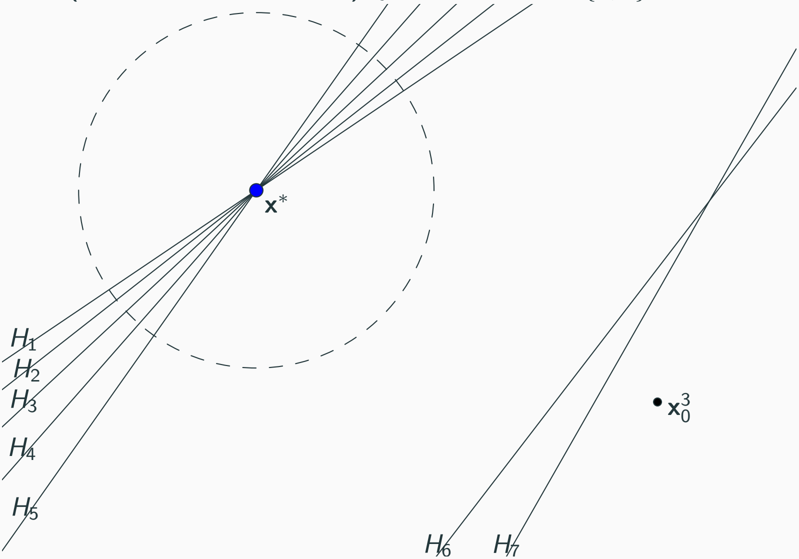
Example

MRK(A,$\mathbf{b}$,$k=2$,$W=3$,$d=1$): $j=2$, $i=2$, $S=\{7,5\}$



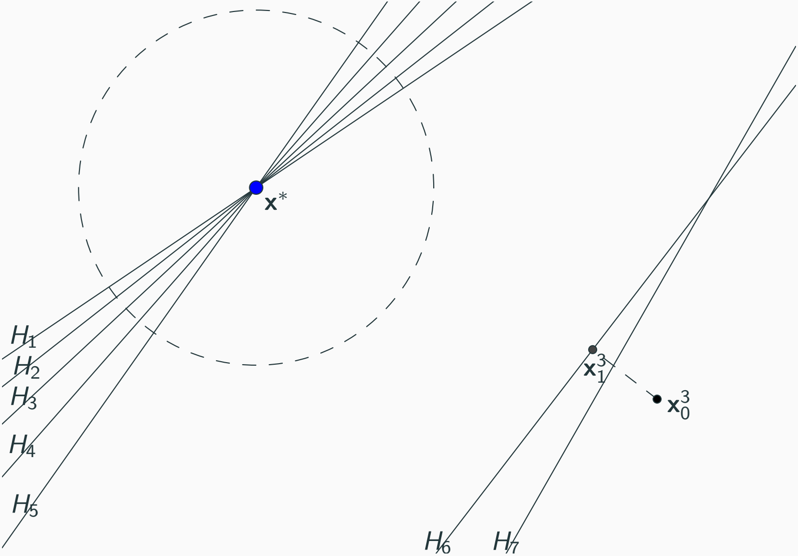
Example

MRK(A**,b,**$k=2$**,**$W=3$**,**$d=1$**):** $j=1$, $i=3$, $S=\{7,5\}$



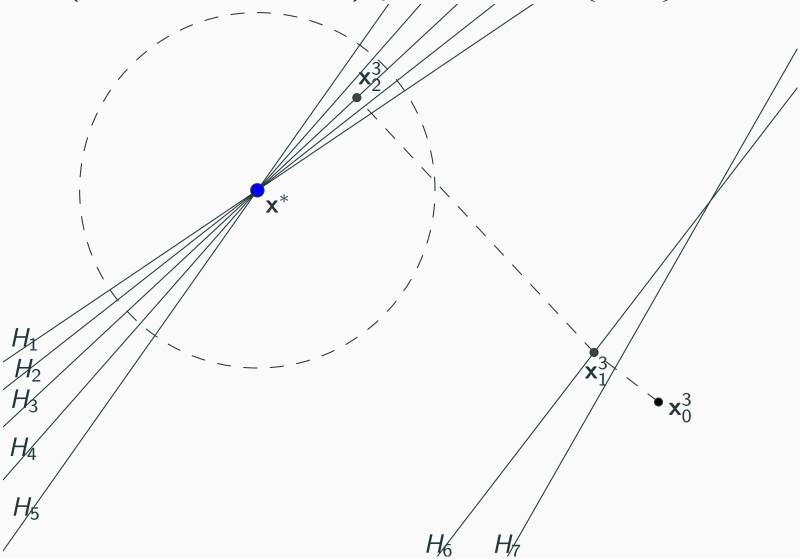
Example

MRK(A**,**b**,**$k = 2$**,**$W = 3$**,**$d = 1$**):** $j = 1$, $i = 3$, $S = \{7, 5\}$



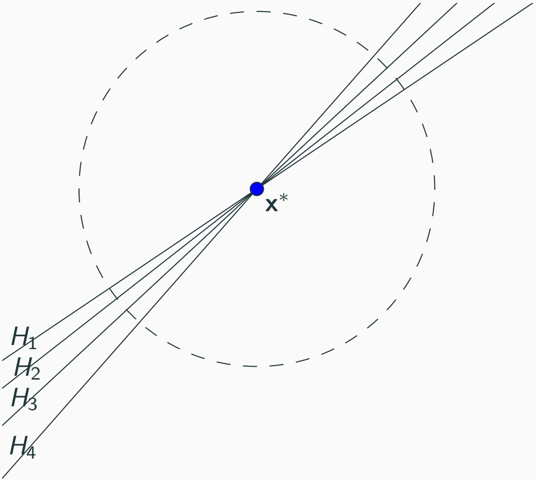
Example

MRK(A,b,$k = 2$,$W = 3$,$d = 1$): $j = 2$, $i = 3$, $S = \{7, 5, 6\}$



Example

Solve $A_{S^C}\mathbf{x} = \mathbf{b}_{S^C}$.



Theoretical Guarantees

Lemma

Let $\epsilon^* = \min_{i \in supp(\mathbf{e})} |A\mathbf{x}^* - \mathbf{b}|_i = |e_i|$ *and suppose* $|supp(\mathbf{e})| = s$*. Assume that* $||\mathbf{a}_i|| = 1$ *for all* $i \in [m]$ *and let* $0 < \delta < 1$*. Define*

$$k^* = \left\lceil \frac{\log\left(\frac{\delta(\epsilon^*)^2}{4||\mathbf{x}^*||^2}\right)}{\log\left(1 - \frac{\sigma^2_{\min}(A_{supp(\mathbf{e})^C})}{m-s}\right)} \right\rceil.$$

Then in window i *of the Windowed Kaczmarz method, the iterate produced by the RK iterations,* $\mathbf{x}^i_{k^*}$ *satisfies*

$$\mathbb{P}\left[||\mathbf{x}^i_{k^*} - \mathbf{x}^*|| \leq \frac{1}{2}\epsilon^*\right] \geq p := (1-\delta)\left(\frac{m-s}{m}\right)^{k^*}.$$

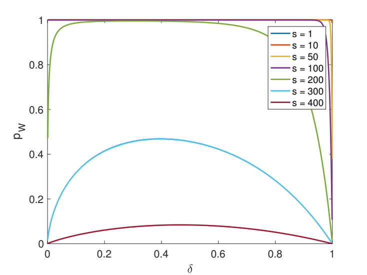
Theoretical Guarantees

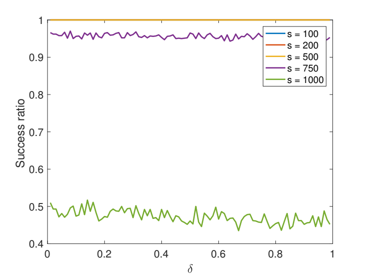
Theorem (H. - Needell 2018+)

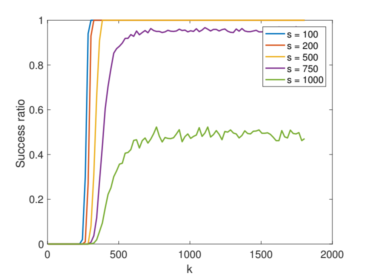
Assume that $\|\mathbf{a}_i\| = 1$ *for all* $i \in [m]$ *and let* $0 < \delta < 1$. *Suppose* $d \geq s = |supp(\mathbf{e})|$, $W \leq \lfloor \frac{m-n}{d} \rfloor$ *and* k^* *is as given in the previous lemma. Then the Windowed Kaczmarz method on* $A, \mathbf{b}$ *will detect the corrupted equations (*$supp(\mathbf{e}) \subset S$*) and the remaining equations given by* $A_{[m]-S}, \mathbf{b}_{[m]-S}$ *will have solution* $\mathbf{x}^*$ *with probability at least*

$$p_W := 1 - \left[1 - (1-\delta)\left(\frac{m-s}{m}\right)^{k^*}\right]^W.$$

Success Rates (Gaussian $A \in \mathbb{R}^{50000\times 100}$)



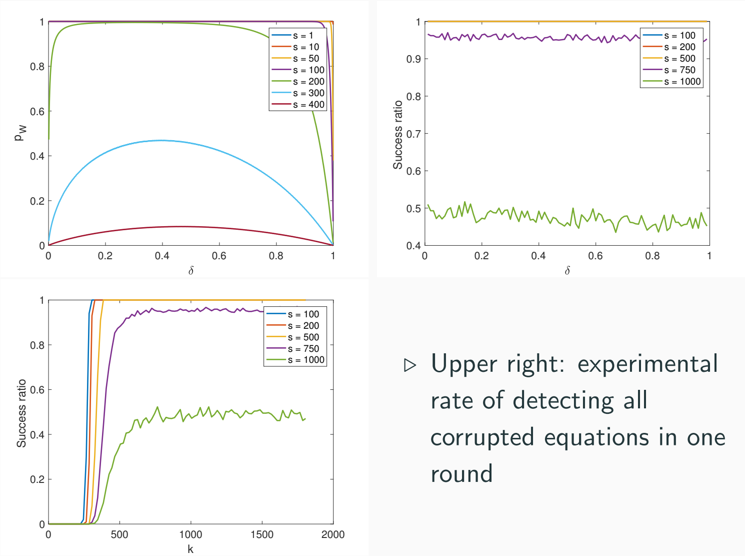




- Upper left: probability of detecting all corrupted equations in one round,

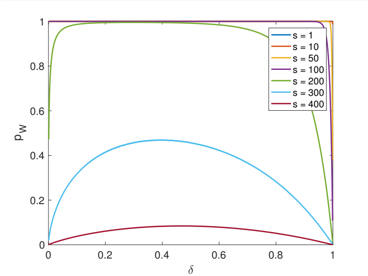
$$p_W := 1 - \left[1 - (1-\delta)\left(\frac{m-s}{m}\right)^{k^*}\right]^W$$

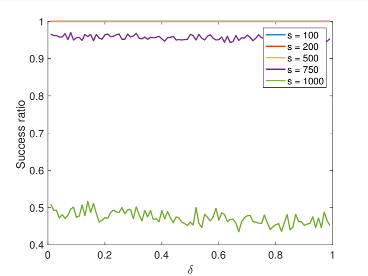
Success Rates (Gaussian $A \in \mathbb{R}^{50000 \times 100}$)

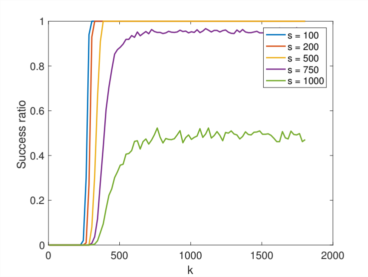


- Upper right: experimental rate of detecting all corrupted equations in one round

Success Rates (Gaussian $A \in \mathbb{R}^{50000\times 100}$)

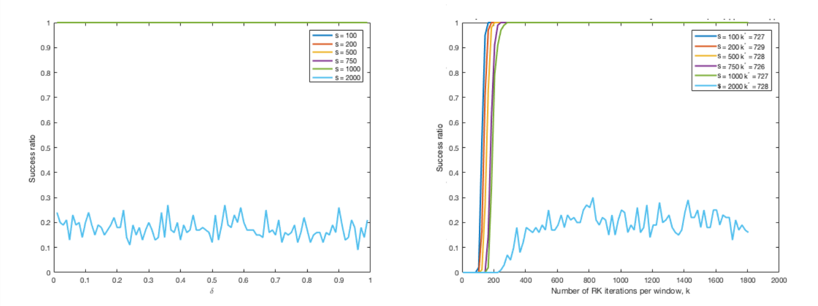






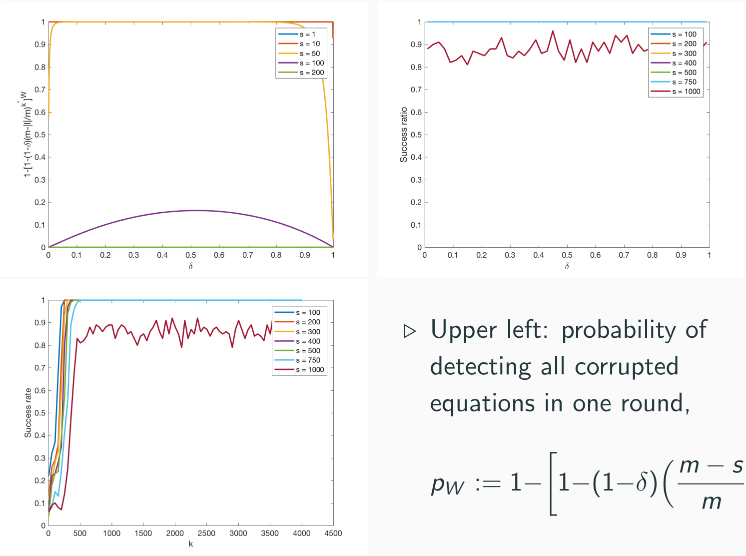
- ▷ Lower left: experimental rate of detecting all corrupted equations in one round for varying number of RK iterations k

Total Experimental Success Rates (Gaussian $A \in \mathbb{R}^{50000\times 100}$)



- ▷ experimental rate of success of detecting all corrupted equations over all $W = \lfloor \frac{m-n}{d} \rfloor$ windows

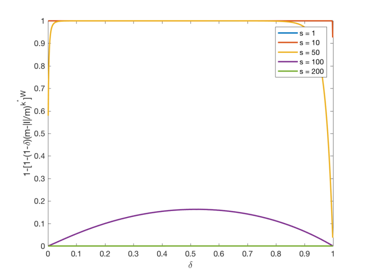
Success Rates ("correlated" $A \in \mathbb{R}^{50000 \times 100}$)

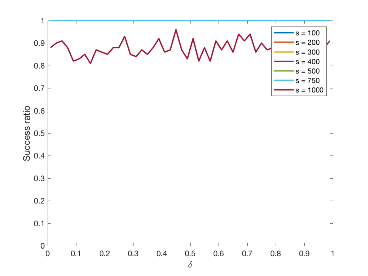


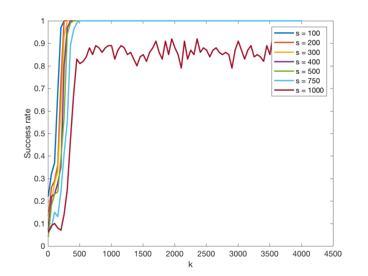
- ▷ Upper left: probability of detecting all corrupted equations in one round,

$$p_W := 1 - \left[1 - (1-\delta)\left(\frac{m-s}{m}\right)^{k^*}\right]^W$$

Success Rates (“correlated” $A \in \mathbb{R}^{50000\times 100}$)

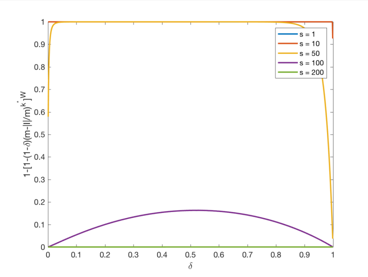


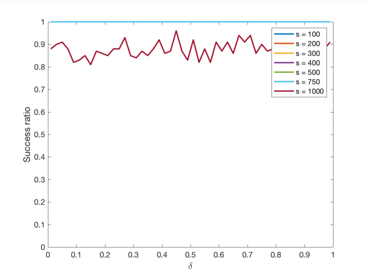


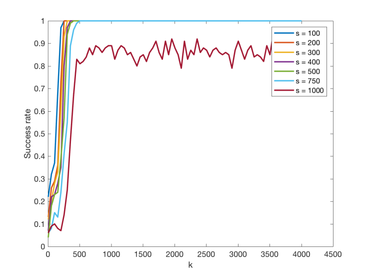


- ▷ Upper right: experimental rate of detecting all corrupted equations in one round

Success Rates (“correlated” $A \in \mathbb{R}^{50000\times 100}$)







- ▷ Lower left: experimental rate of detecting all corrupted equations in one round for varying number of RK iterations k

Conclusions

- ▷ Motzkin's method is accelerated even in the presence of noise
  - γ_k, the parameter governing this acceleration, governs the acceleration of SKM
- ▷ γ_k can be bounded for some systems

- ▷ RK methods may be used to detect corruption
- ▷ theoretical bounds do not reflect empirical results

Future Work

- identify useful bounds on γ_k for other useful systems
- design dynamic sampling algorithms which use the optimal sample size β

- reduce dependence on artificial parameters in corruption detection bounds
- introduce a Bayesian framework into MRK

Thanks for attending!

Questions?

[DLHN17] 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.

[HN18a] J. Haddock and D. Needell. **On Motzkin's method for inconsistent linear systems.** BIT Numerical Mathematics, 2018. To appear.

[HN18b] J. Haddock and D. Needell. **Randomized projection methods for linear systems with arbitrarily large sparse corruptions.** 2018. Submitted.

[MS54] T. S. Motzkin and I. J. Schoenberg. **The relaxation method for linear inequalities.** Canadian J. Math., 6:393–404, 1954.

[Nee10] D. Needell. **Randomized Kaczmarz solver for noisy linear systems.** BIT, 50(2):395–403, 2010.

[SV09] T. Strohmer and R. Vershynin. **A randomized Kaczmarz algorithm with exponential convergence.** J. Fourier Anal. Appl., 15:262–278, 2009.