Solving Systems of Linear Inequalities with Randomized Projections

Jamie Haddock

Graduate Group in Applied Mathematics,
Department of Mathematics,
University of California, Davis

Math Club
March 2, 2016

Joint work with Jesus De Loera and Deanna Needell

Optimization

I think about problems of the sort:

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } g(x) \leq 0 \end{aligned}$$

Optimization

I think about problems of the sort:

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } g(x) \leq 0 \end{aligned}$$

These sorts of problems are all around us!

Optimization

I think about problems of the sort:

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } g(x) \leq 0 \end{aligned}$$

These sorts of problems are all around us!

Today we'll consider a specific form of optimization problem...

Linear Programs

I think about problems of the sort:

$$\begin{aligned} &\min c^T x \\ &\text{s.t. } Ax \leq b \end{aligned} \qquad (LP)$$

Linear Programs

I think about problems of the sort:

$$\begin{aligned} &\min c^T x \\ &\text{s.t. } Ax \leq b \end{aligned} \qquad (LP)$$

$A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ and we are optimizing over $x \in \mathbb{R}^n$.

Linear Programs

I think about problems of the sort:

$$\begin{aligned} &\min c^T x \\ &\text{s.t. } Ax \leq b \end{aligned} \qquad (LP)$$

$A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ and we are optimizing over $x \in \mathbb{R}^n$.

But, even this can be simplified...

Linear Feasibility Problem

In fact, we'll consider the *linear feasibility problem* (LF):

Linear Feasibility Problem

In fact, we'll consider the *linear feasibility problem* (LF):

Find x such that $Ax \leq b$ or conclude one does not exist.

Linear Feasibility Problem

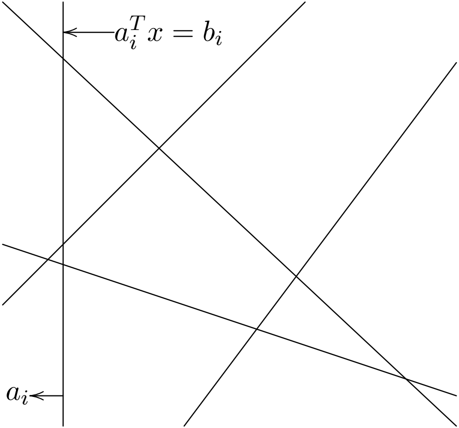
In fact, we'll consider the *linear feasibility problem* (LF):

Find x such that $Ax \leq b$ or conclude one does not exist.

It can be shown that (LP) and (LF) are equivalent.

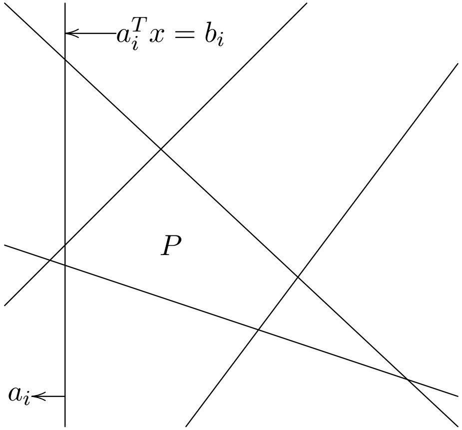
Linear Feasibility Problem

Reminder: linear equations represent a *hyperplane* (in the proper dimension), so linear inequalities define a *halfspace*.



Linear Feasibility Problem

LF can be interpreted as seeking a point within a (possibly nonempty) polyhedron $P = \{x | Ax \leq b\}$:



How to Solve LF

Isn't the linear feasibility problem easy to solve?

How to Solve LF

Isn't the linear feasibility problem easy to solve?

Answer: Sort of...

How to Solve LF

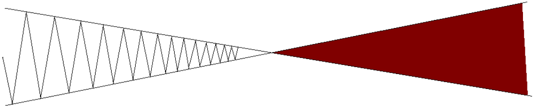
Isn't the linear feasibility problem easy to solve?

Good news: Our geometric intuition for the problem gives us a good idea for how to solve it!

How to Solve LF

Isn't the linear feasibility problem easy to solve?

Good news: Our geometric intuition for the problem gives us a good idea for how to solve it!



Motzkin

Kaczmarz

Projection Methods

If we want all of the linear inequalities to be satisfied (meaning we want our point to lie on the correct side of all the hyperplanes), then we need that each of the linear inequalities is satisfied.

Projection Methods

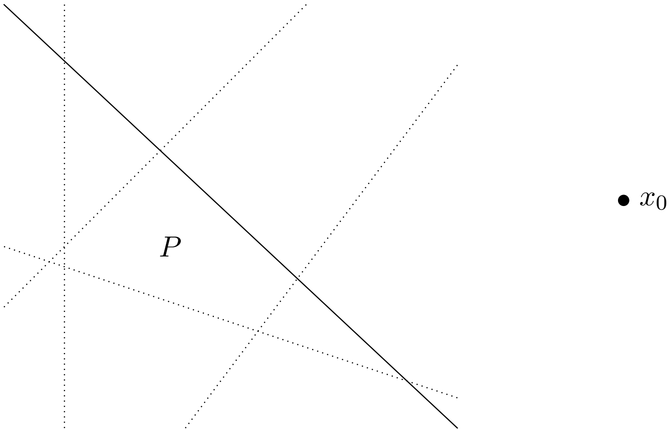
If we want all of the linear inequalities to be satisfied (meaning we want our point to lie on the correct side of all the hyperplanes), then we need that each of the linear inequalities is satisfied. Tautology.

Projection Methods

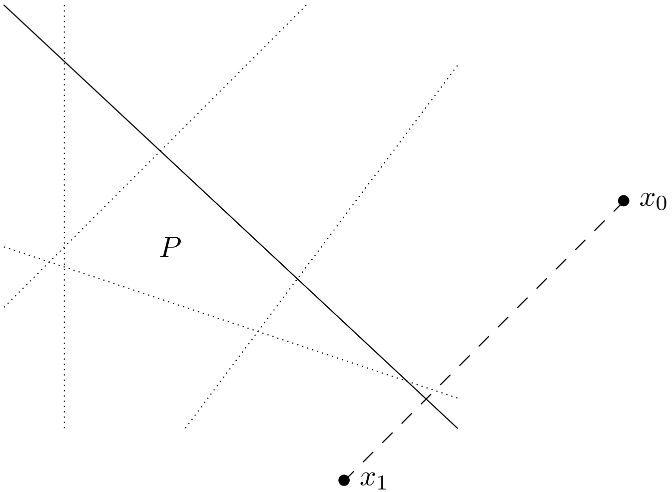
If we want all of the linear inequalities to be satisfied (meaning we want our point to lie on the correct side of all the hyperplanes), then we need that each of the linear inequalities is satisfied.

So... If we have some point that isn't satisfying one of the inequalities, we should force it to satisfy that inequality!

Projection Methods



Projection Methods



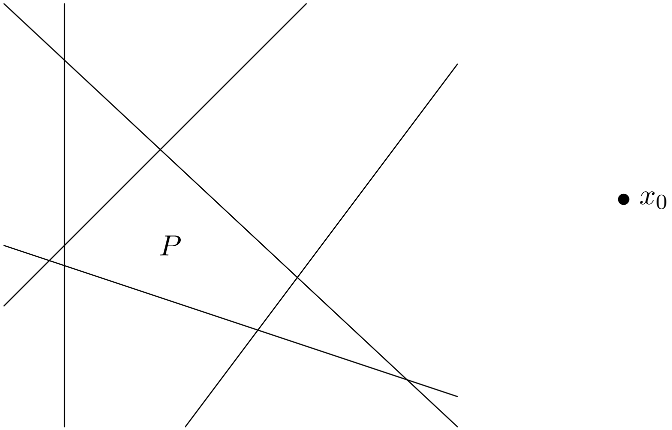
Motzkin's Relaxation Method(s)

Method

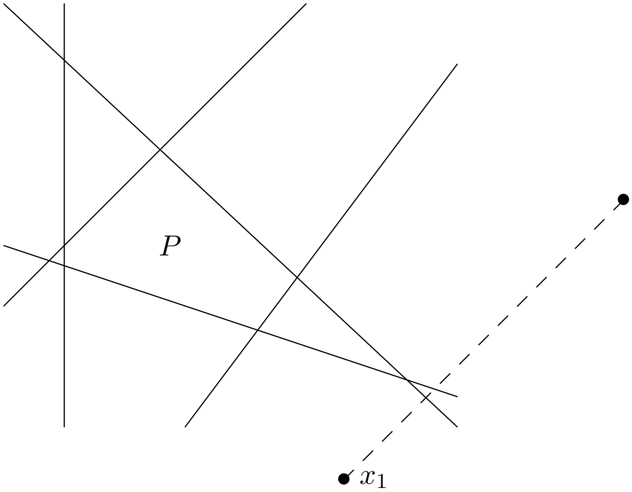
Suppose $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $P := \{x \in \mathbb{R}^n : Ax \le b\}$ is nonempty. Fix $0 < \lambda \le 2$. Given $x_0 \in \mathbb{R}^n$, iteratively construct approximations to P in the following way:

1. *If x_k is feasible, stop.*
2. *Choose $i_k \in [m]$ as $i_k := \operatorname{argmax}_{i \in [m]} a_i^T x_{k-1} - b_i$.*
3. *Define $x_k := x_{k-1} - \lambda \frac{a_{i_k}^T x_{k-1} - b_{i_k}}{||a_{i_k}||^2} a_{i_k}$.*
4. *Repeat.*

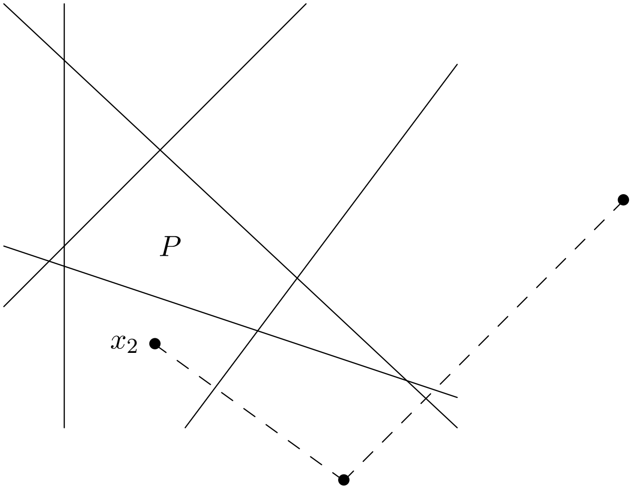
Motzkin's Method



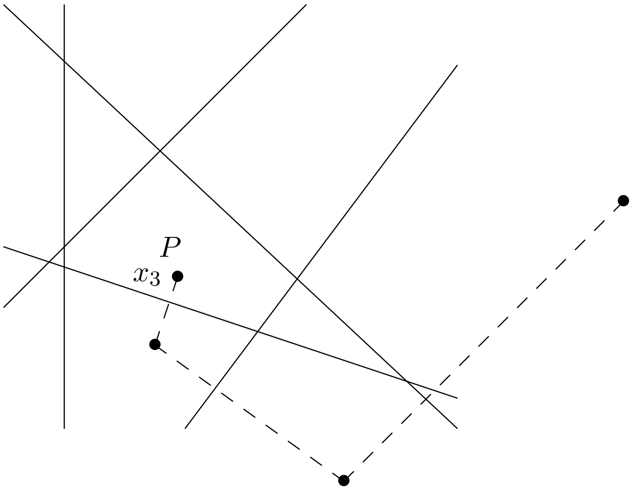
Motzkin's Method



Motzkin's Method



Motzkin's Method



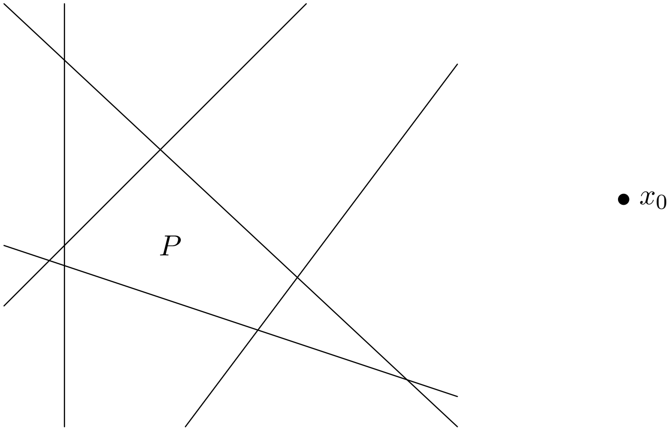
Randomized Kaczmarz Method

Method

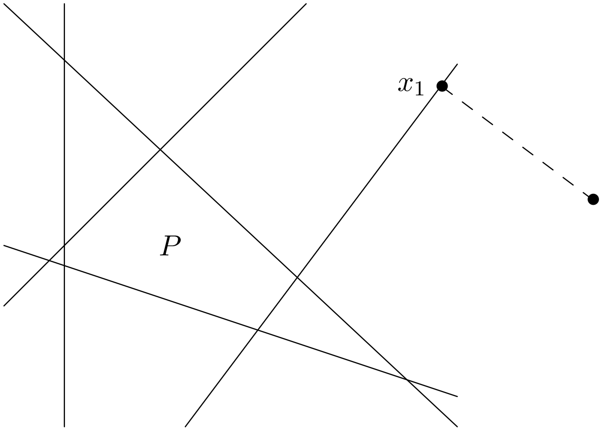
Suppose $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ *and* $P := \{x \in \mathbb{R}^n : Ax \le b\}$ *is nonempty. Given* $x_0 \in \mathbb{R}^n$, *iteratively construct approximations to* P *in the following way:*

1. *If* x_k *is feasible, stop.*
2. *Choose* $i_k \in [m]$ *with probability* $\frac{||a_{i_k}||^2}{||A||_F^2}$.
3. *Define* $x_k := x_{k-1} - \frac{(a_{i_k}^T x_{k-1} - b_{i_k})^+}{||a_{i_k}||^2} a_{i_k}$.
4. *Repeat.*

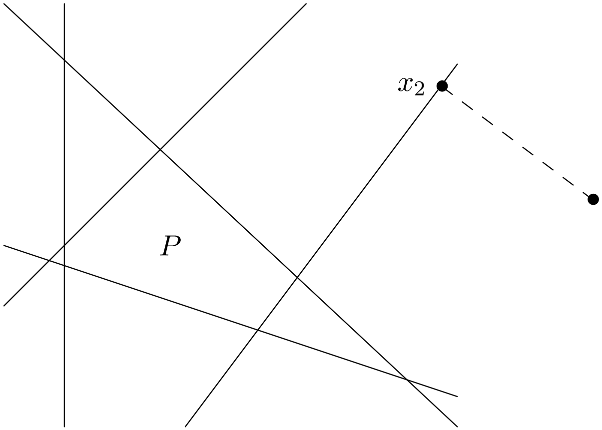
Kaczmarz Method



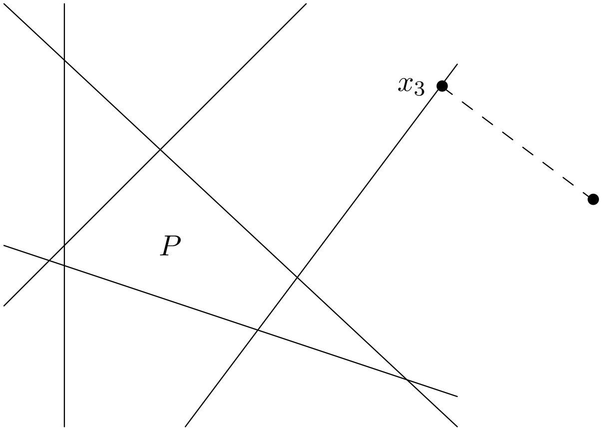
Kaczmarz Method



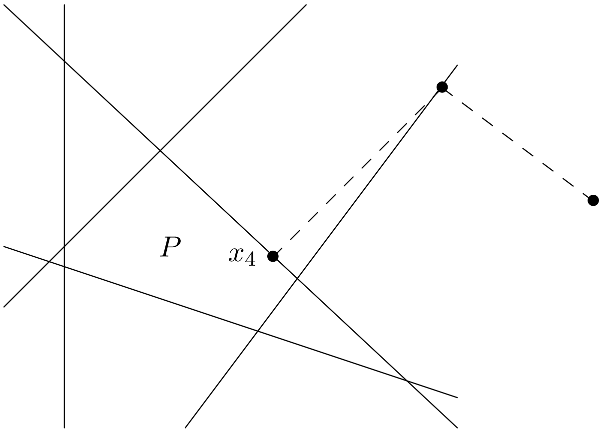
Kaczmarz Method



Kaczmarz Method



Kaczmarz Method



Motivation

Motzkin's Method
Pro: convergence produces monotone decreasing distance sequence
Con: computationally expensive for large systems

Kaczmarz Method
Pro: computationally inexpensive, able to analyze the expected convergence rate
Con: slow convergence near the polyhedral solution set

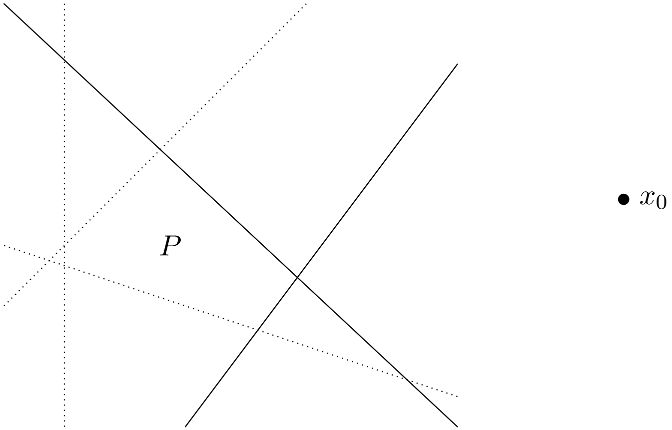
A Hybrid Method

Method (SKMM)

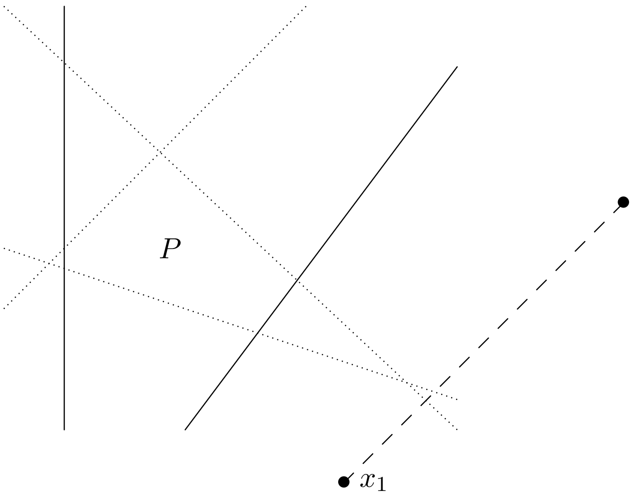
Suppose $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ *and* $P := \{x \in \mathbb{R}^n : Ax \leq b\}$ *is nonempty. Fix* $0 < \lambda \leq 2$. *Given* $x_0 \in \mathbb{R}^n$, *iteratively construct approximations to* P *in the following way:*

1. *If* x_k *is feasible, stop.*
2. *Choose* $\tau_k \subset [m]$ *to be a sample of size* β *constraints chosen uniformly at random from among the rows of* A.
3. *From among these* β *rows, choose* $i_k := \underset{i \in \tau_k}{\operatorname{argmax}}\, a_i^T x_{k-1} - b_i$.
4. *Define* $x_k := x_{k-1} - \lambda \frac{(a_{i_k}^T x_{k-1} - b_{i_k})^+}{||a_{i_k}||^2} a_{i_k}$.
5. *Repeat.*

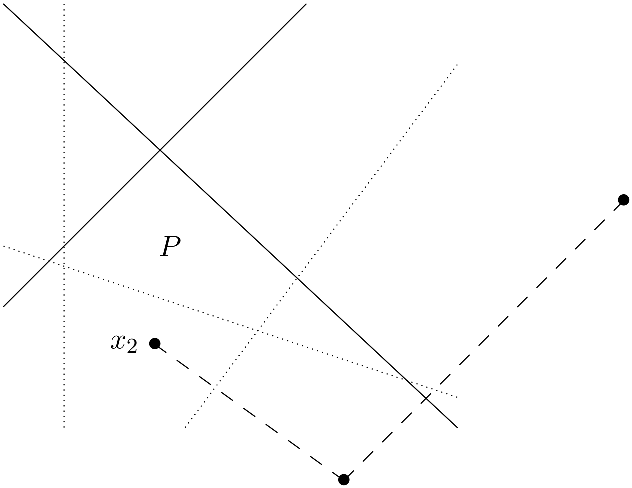
A Hybrid Method



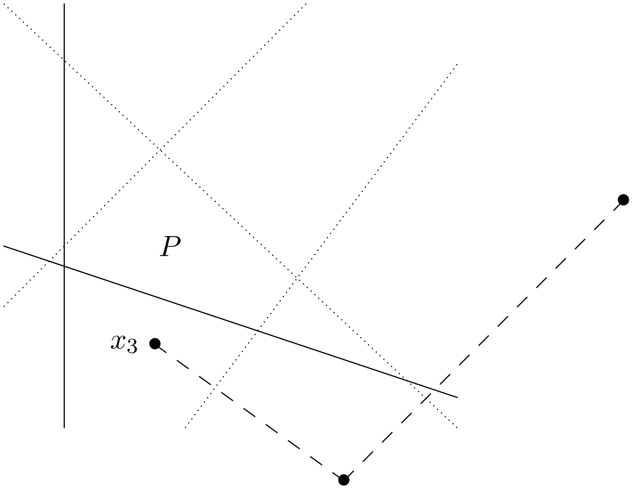
A Hybrid Method



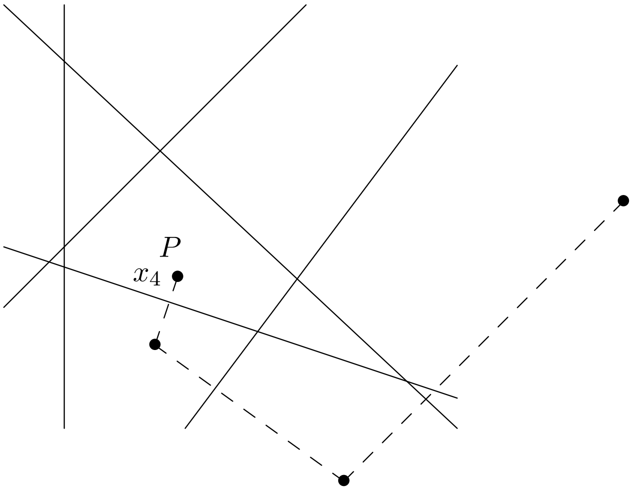
A Hybrid Method



A Hybrid Method



A Hybrid Method



Generalized Method

Note that both previous methods are captured by the class of SKMM methods:

Generalized Method

Note that both previous methods are captured by the class of SKMM methods:

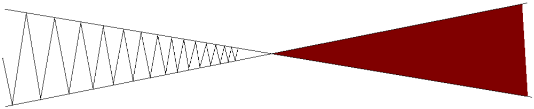
1. The Kaczmarz method is SKMM where the sample size, $\beta = 1$ and the relaxation parameter, $\lambda = 1$.

Generalized Method

Note that both previous methods are captured by the class of SKMM methods:

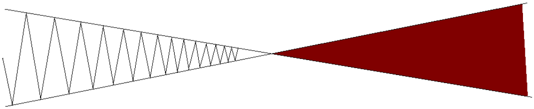
1. The Kaczmarz method is SKMM where the sample size, $\beta = 1$ and the relaxation parameter, $\lambda = 1$.
2. Motzkin's Relaxation methods are SKMM where the sample size, $\beta = m$.

An Important Reminder



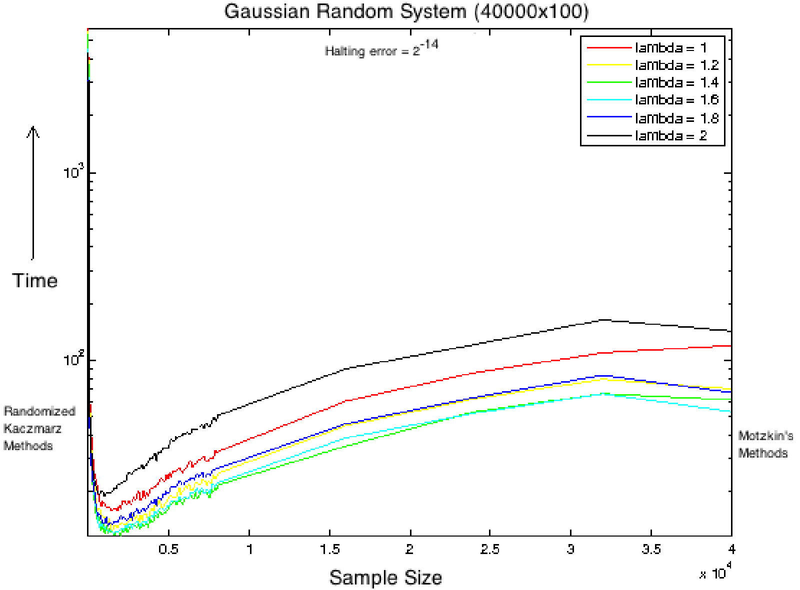
These methods may not actually stop with a solution...

An Important Reminder



These methods may not actually stop with a solution...
However, we can ensure that our iterate points get arbitrarily close to the solution set, P!

Experimental Results



Motzkin's Method Convergence Rate

Theorem (Agmon)

For a normalized system, $||a_i|| = 1$ for all $i = 1, ..., m$, if the feasible region, $P := \{x|Ax \leq b\}$, is nonempty then the relaxation methods converges linearly:

$$d(x_k, P)^2 \leq \left(1 - \frac{2\lambda - \lambda^2}{mL_2^2}\right)^k d(x_0, P)^2.$$

Random Kaczmarz Method Convergence Rate

Theorem (Lewis, Leventhal)

If the feasible region, $P := \{x | Ax \leq b\}$, is nonempty then the Randomized Kaczmarz method with relaxation parameter λ converges linearly in expectation:

$$\mathbb{E}[d(x_k, P)^2] \leq \left(1 - \frac{2\lambda - \lambda^2}{||A||_F^2 L_2^2}\right)^k d(x_0, P)^2.$$

SKM Method Convergence Rate

Theorem (De Loera, H., Needell)

If the feasible region (for normalized A) is nonempty, then the SKM methods with samples of size β converges at least linearly in expectation: In each iteration,

$$\mathbb{E}[d(x_k, P)^2] \leq \left(1 - \frac{2\lambda - \lambda^2}{S_{k-1} L_2^2}\right) d(x_{k-1}, P)^2$$

where $S_{k-1} = \max\{m - s_{k-1}, m - \beta + 1\}$ and s_{k-1} is the number of constraints satsifed by x_{k-1}. Clearly then,

$$\mathbb{E}[d(x_k, P)^2] \leq \left(1 - \frac{2\lambda - \lambda^2}{m L_2^2}\right)^k d(x_0, P)^2.$$

Improved Rate

Theorem (De Loera, H., Needell)

If the feasible region, $P = \{x | Ax \leq b\}$ is generic and nonempty (for normalized A), then an SKM method with samples of size $\beta \leq m - n$ is guaranteed an increased convergence rate after some K:

$$\mathbb{E}[d(x_k, P)^2] \leq \left(1 - \frac{2\lambda - \lambda^2}{mL_2^2}\right)^K \left(1 - \frac{2\lambda - \lambda^2}{(m - \beta + 1)L_2^2}\right)^{k-K} d(x_0, P)^2.$$

Finiteness of Motzkin's Method

Theorem (Telgen)

Either the relaxation method detects feasibility of the system, $Ax \leq b$ (with A normalized), within $k = \left\lceil \frac{2^{4L}}{n\lambda(2-\lambda)} \right\rceil$ iterations or the system is infeasible.*

*with $x_0 = 0$

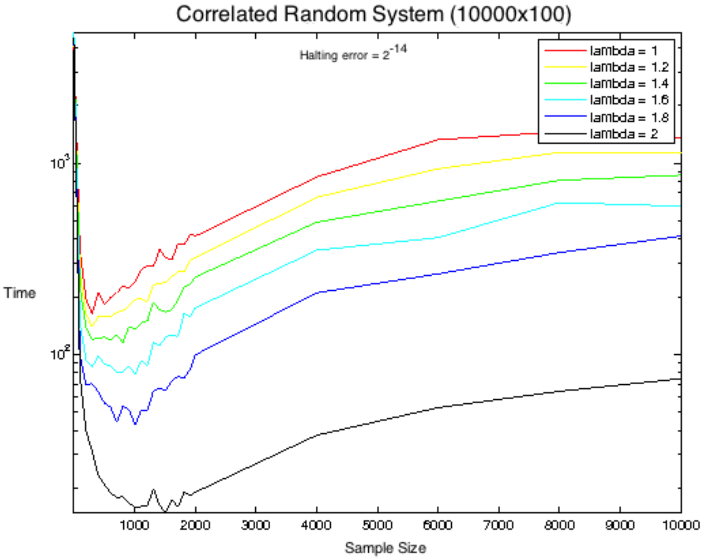
Expected Finiteness of SKM methods

Theorem (De Loera, H., Needell)

If the system, $Ax \leq b$ is feasible, then with high probability the Sampling Kaczmarz-Motzkin method with relaxation parameter* $0 < \lambda < 2$ *will detect feasibility within a given number of steps.*

*with $x_0 = 0$

Conclusions



Future work:

Future work:

1. Provide theoretical guidance for selection of the optimal sample size, β, and optimal overshooting parameter, λ for a given (class of) system(s).

Future work:

1. Provide theoretical guidance for selection of the optimal sample size, β, and optimal overshooting parameter, λ for a given (class of) system(s).
2. Describe the K after which the convergence rate is guaranteed to be improved.

Acknowledgements

Thanks to you for attending!

Are there any questions?

References I

- Kaczmarz, S. (1937).

  Angenaherte auflosung von systemen linearer gleichungen.

  Bull.Internat.Acad.Polon.Sci.Lettres A, pages 335–357.

- Leventhal, D. and Lewis, A. S. (2010).

  Randomized methods for linear constraints: convergence rates and conditioning.

  Math.Oper.Res., 35(3):641–654.

  65F10 (15A39 65K05 90C25); 2724068 (2012a:65083); Raimundo J. B. de Sampaio.

- Motzkin, T. S. and Schoenberg, I. J. (1954).

  The relaxation method for linear inequalities.

  Canadian J. Math., 6:393–404.

- Needell, D. (2010).

  Randomized kaczmarz solver for noisy linear systems.

  BIT, 50(2):395–403.

References II

- Needell, D., Sbrero, N., and Ward, R. (2013).

  Stochastic gradient descent and the randomized kaczmarz algorithm.

  submitted.

- Needell, D. and Tropp, J. A. (2013).

  Paved with good intentions: Analysis of a randomized block kaczmarz method.

  Linear Algebra Appl.

- Schrijver, A. (1986).

  Theory of linear and integer programming.

  Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons, Ltd., Chichester.

  A Wiley-Interscience Publication.

- Strohmer, T. and Vershynin, R. (2009).

  A randomized kaczmarz algorithm with exponential convergence.

  J. Fourier Anal. Appl., 15:262–278.