Motzkin's Method and the Randomized Kaczmarz Method

Jamie Haddock

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

Davis Mathematics Conference
November 10, 2015

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:

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

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

Linear Programs

I think about problems of the sort:

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

Linear Programs

I think about problems of the sort:

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

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

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

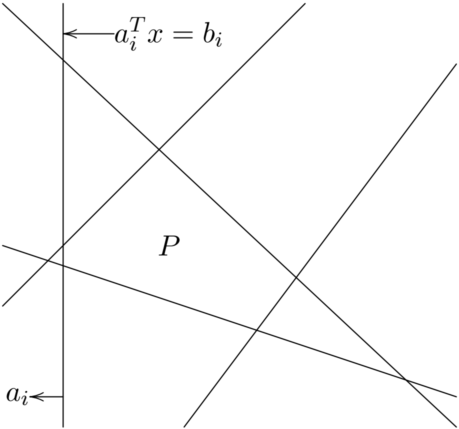
In fact, we'll consider the *linear feasibility problem* (LF):

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

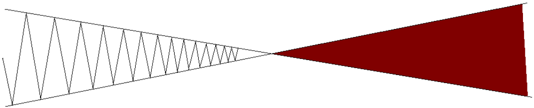
It can be shown that (LP) and (LF) are equivalent.

Linear Feasibility Problem

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



Projection Methods



Motzkin

Kaczmarz

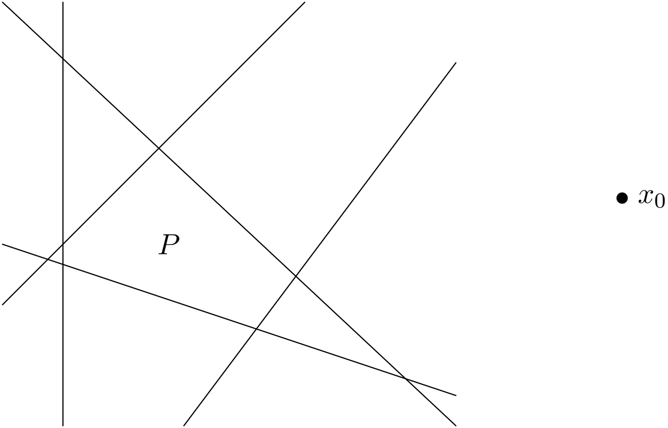
Motzkin's Relaxation Method(s)

Method

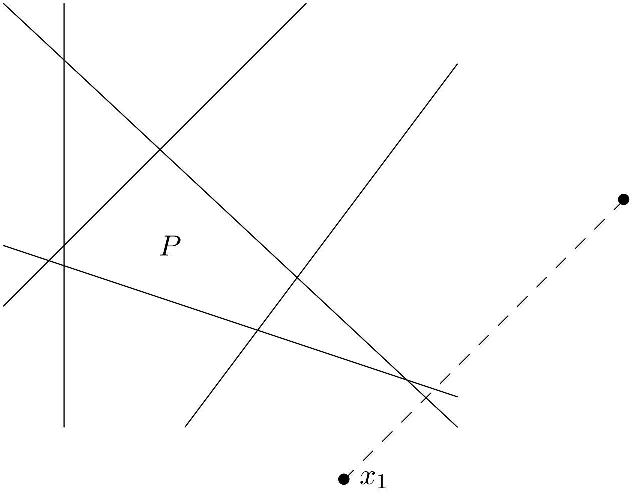
Suppose $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ *and* $P := \{x \in \mathbb{R}^n : Ax \leq b\}$ *is nonempty. Fix* $0 < \lambda \leq 2$. *Given* $x_0 \in \mathbb{R}^n$, *iteratively construct approximations to* P *in the following way:*

1. *If* x_k *is feasible, stop.*
2. *Choose* $i_k \in [m]$ *as* $i_k := \underset{i\in[m]}{\operatorname{argmax}}\, a_i^T x_{k-1} - b_i$.
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.*

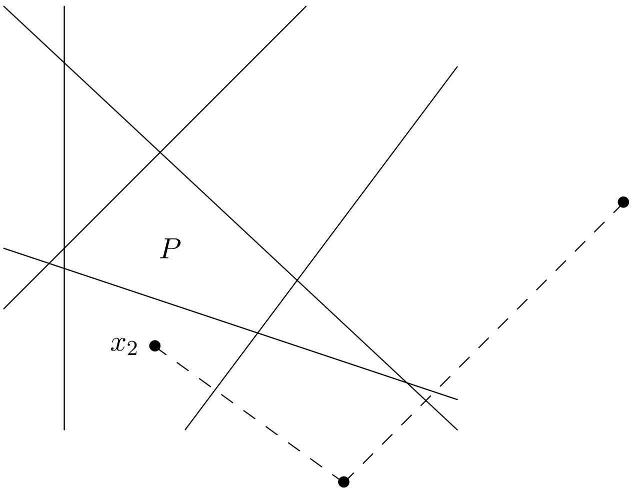
Motzkin's Method



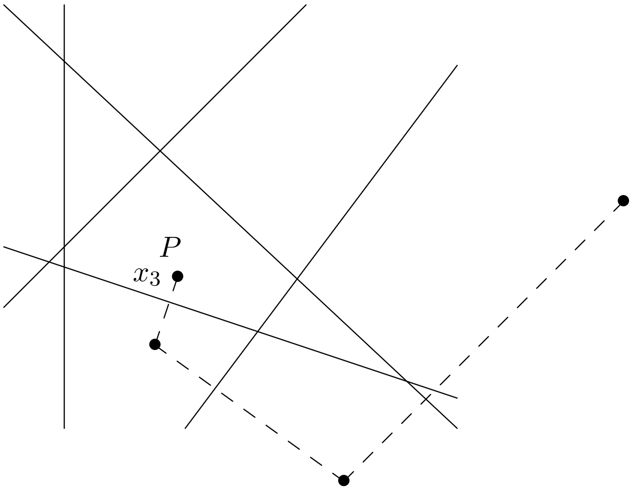
Motzkin's Method



Motzkin's Method



Motzkin's Method



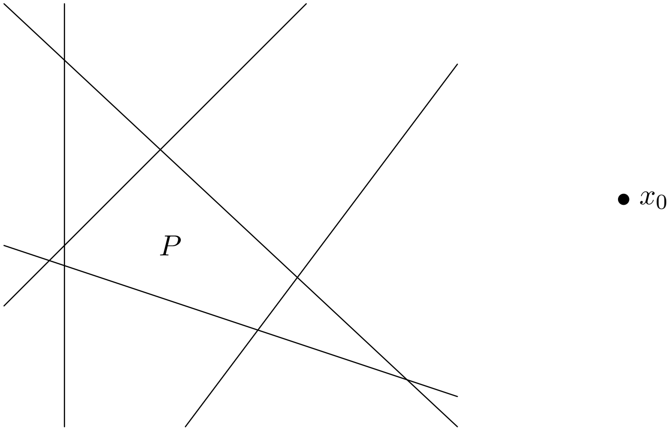
Randomized Kaczmarz Method

Method

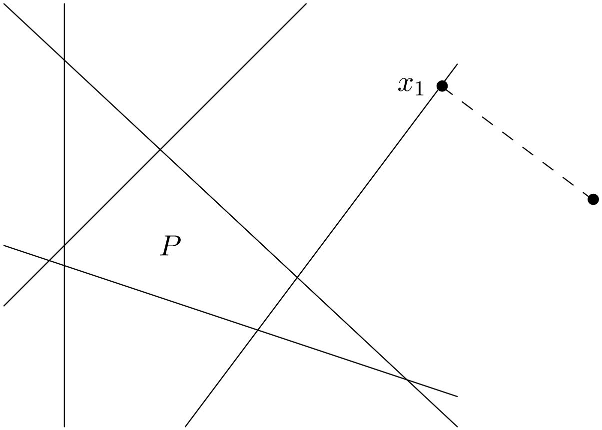
Suppose $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ *and* $P := \{x \in \mathbb{R}^n : Ax \leq b\}$ *is nonempty. Given* $x_0 \in \mathbb{R}^n$, *iteratively construct approximations to* P *in the following way:*

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.*

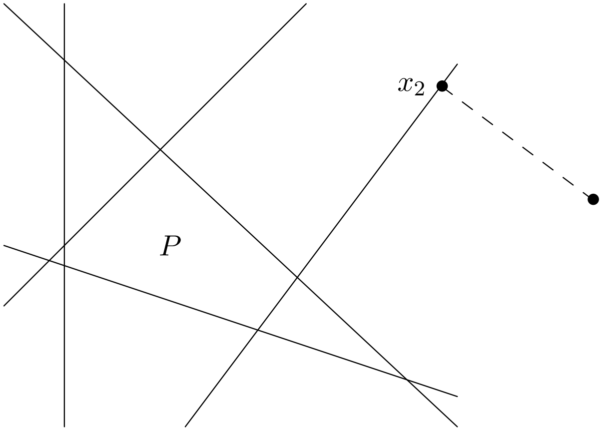
Kaczmarz Method



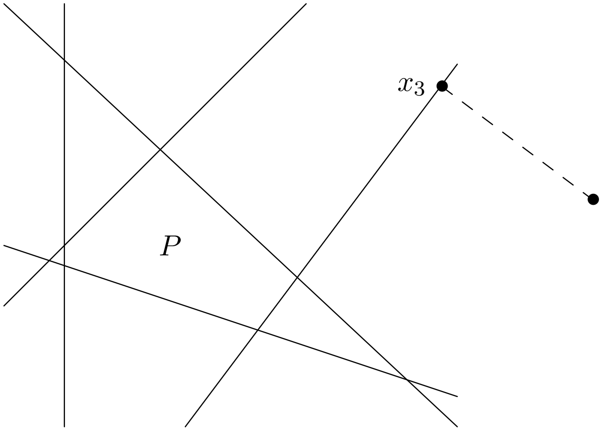
Kaczmarz Method



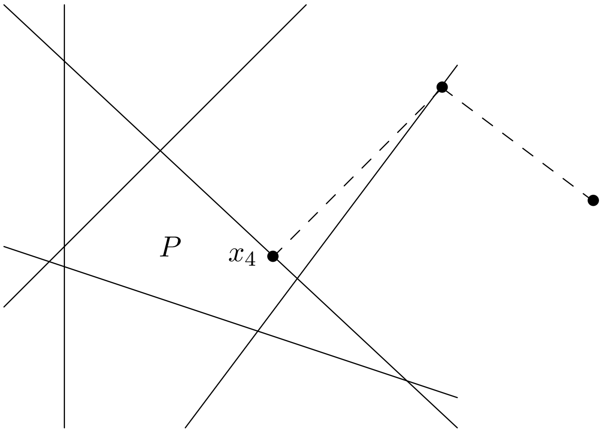
Kaczmarz Method



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* $P := \{x \in \mathbb{R}^n : Ax \le b\}$ *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* $\tau_k \subset [m]$ *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* $i_k := \underset{i\in\tau_k}{\text{argmax}}\, a_i^T x_{k-1} - b_i$.
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.*

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:

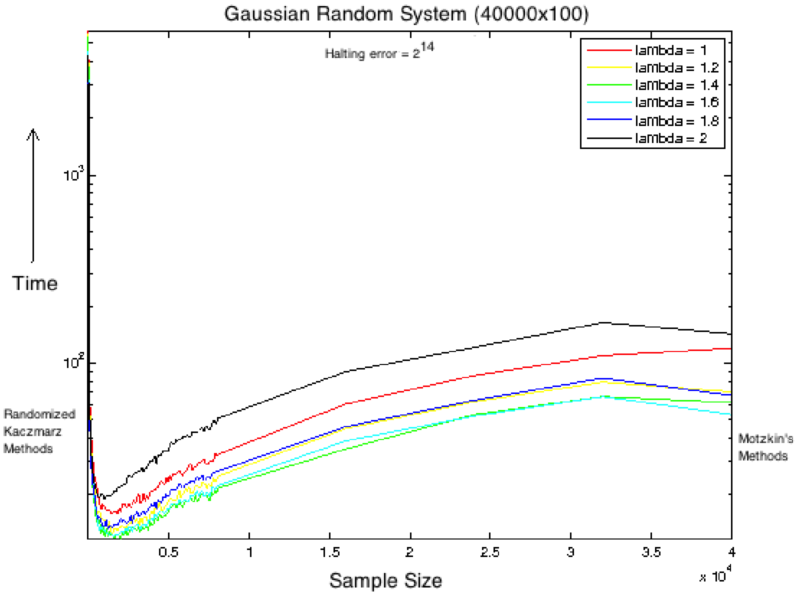
1. The Kaczmarz method is SKMM where the sample size, $\beta = 1$ and the relaxation parameter, $\lambda = 1$.

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, $\beta = m$.

Experimental Results



Future work:

Future work:

1. Provide convergence results similar to those of Strohmer & Vershynin and Lewis & Leventhal for the class of SKMM methods.

Future work:

1. Provide convergence results similar to those of Strohmer & Vershynin and Lewis & Leventhal for the class of SKMM methods.
2. It is well known that the Random Kaczmarz update step is equivalent to the coordinate descent update step applied to the dual problem. I will explore connections of SKMM to variants of randomized coordinate descent in the dual variable space.

Future work:

1. Provide convergence results similar to those of Strohmer & Vershynin and Lewis & Leventhal for the class of SKMM methods.
2. It is well known that the Random Kaczmarz update step is equivalent to the coordinate descent update step applied to the dual problem. I will explore connections of SKMM to variants of randomized coordinate descent in the dual variable space.
3. Provide theoretical guidance for selection of the optimal sample size, β, and optimal overshooting parameter, λ.

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.