A Sampling Kaczmarz-Motzkin Algorithm for Linear Feasibility

Jamie Haddock

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

Copper Mountain Conference on Iterative Methods
March 21, 2016

Joint work with Jesus De Loera and Deanna Needell

Linear Feasibility Problem

We are interested in solving the *linear feasibility problem* (LF):

Linear Feasibility Problem

We are interested in solving the *linear feasibility problem* (LF):

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

Linear Feasibility Problem

We are interested in solving the *linear feasibility problem* (LF):

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

We consider large-scale problems in which $A \in \mathbb{R}^{m \times n}$, $m >> n$.

Linear Feasibility Problem

We are interested in solving the *linear feasibility problem* (LF):

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

We consider large-scale problems in which $A \in \mathbb{R}^{m \times n}$, $m >> n$.

These problems arise in machine learning classification, *support-vector machines* (Boser, Guyon, Vapnik 1992), (Cortes, Vapnik 1995).

Projection Methods

If $P := \{x \in \mathbb{R}^n : Ax \leq b\}$ is nonempty, these methods construct an approximation to an element of P:

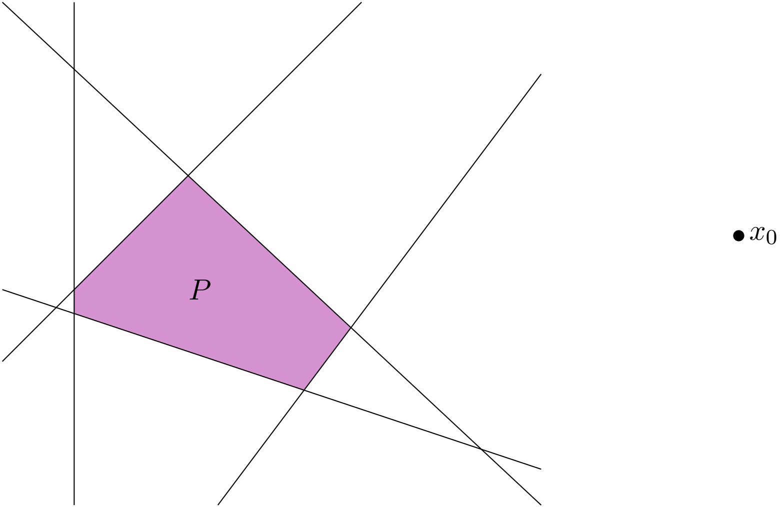
1. Motzkin's Relaxation Method(s)
2. Randomized Kaczmarz Method
3. Sampling Kaczmarz-Motzkin Method (SKM)

Motzkin's Relaxation Method(s)

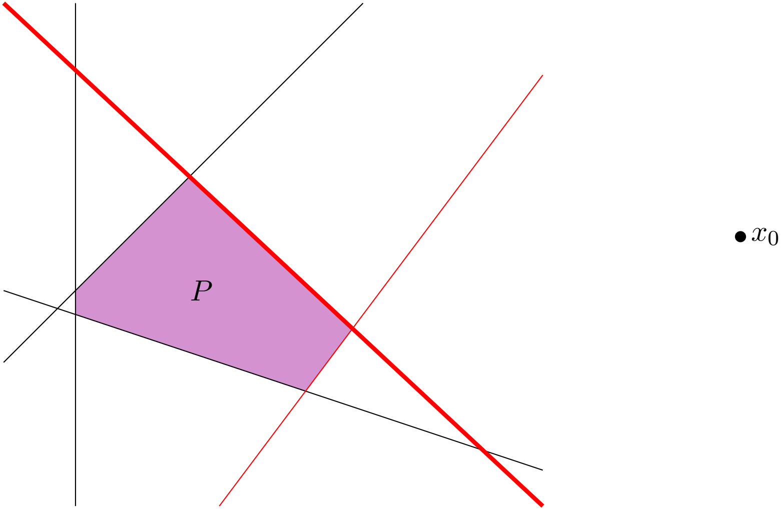
Given $x_0 \in \mathbb{R}^n$, fix $0 < \lambda \leq 2$ and iteratively construct approximations to P:

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.

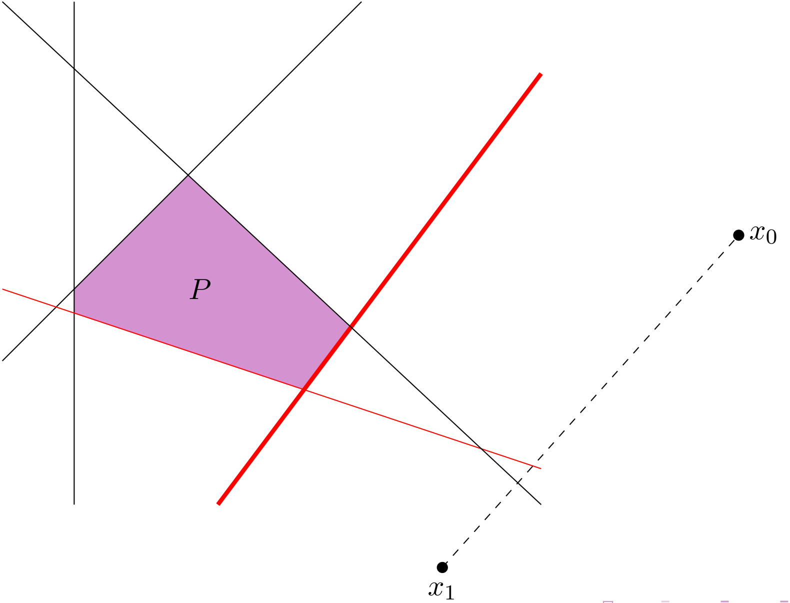
Motzkin's Method



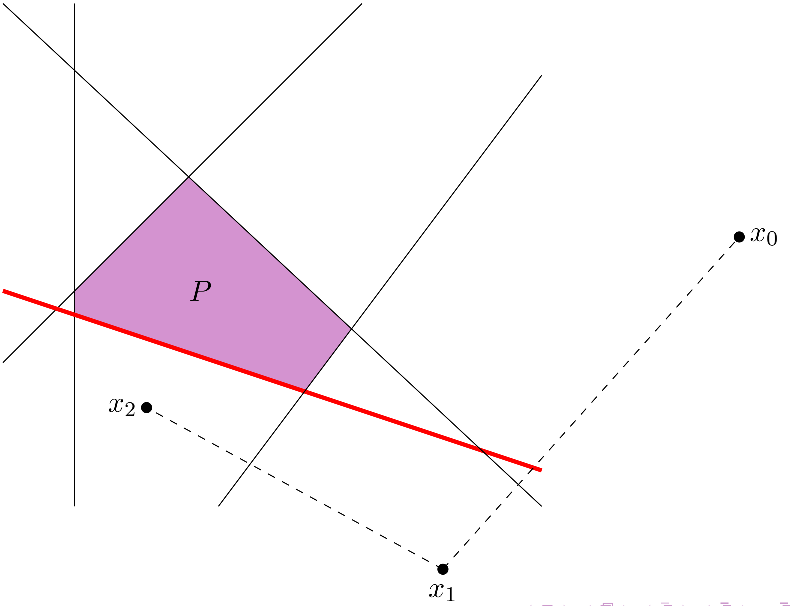
Motzkin's Method



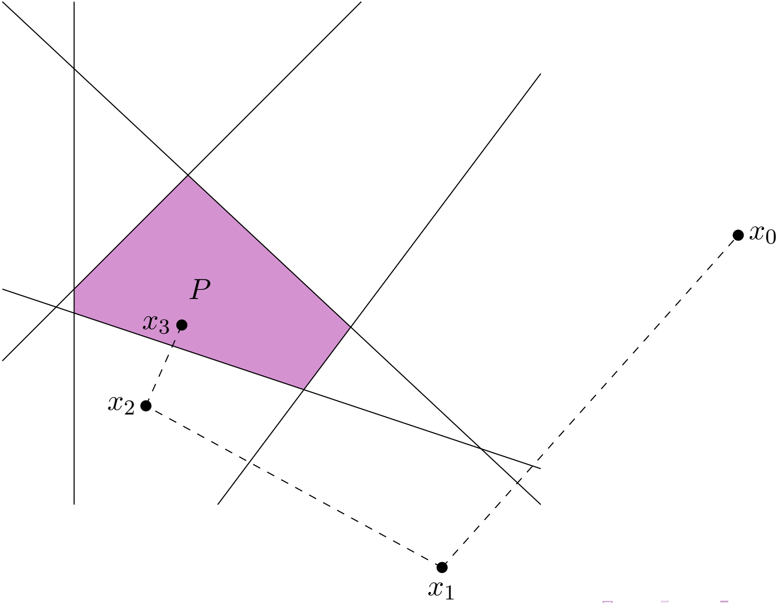
Motzkin's Method



Motzkin's Method



Motzkin's Method

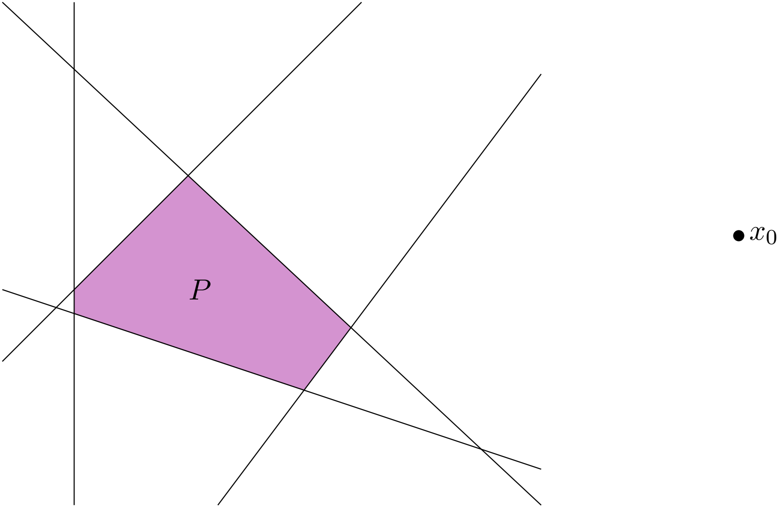


Randomized Kaczmarz Method

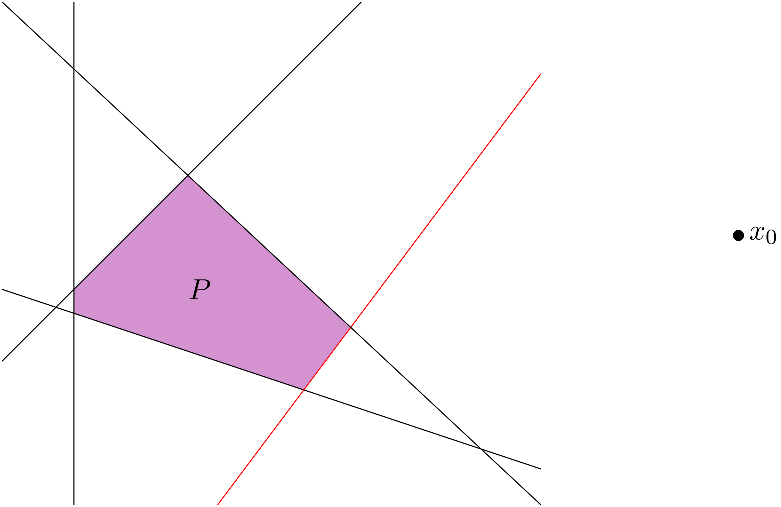
Given $x_0 \in \mathbb{R}^n$, iteratively construct approximations to P:

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.

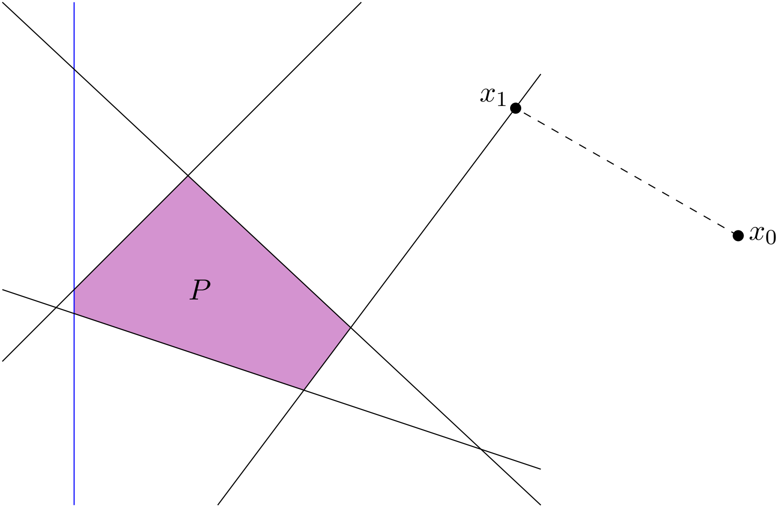
Kaczmarz Method



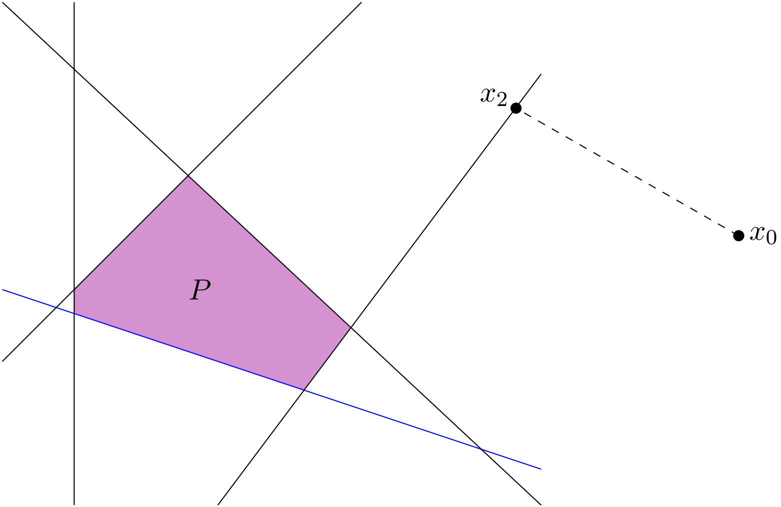
Kaczmarz Method



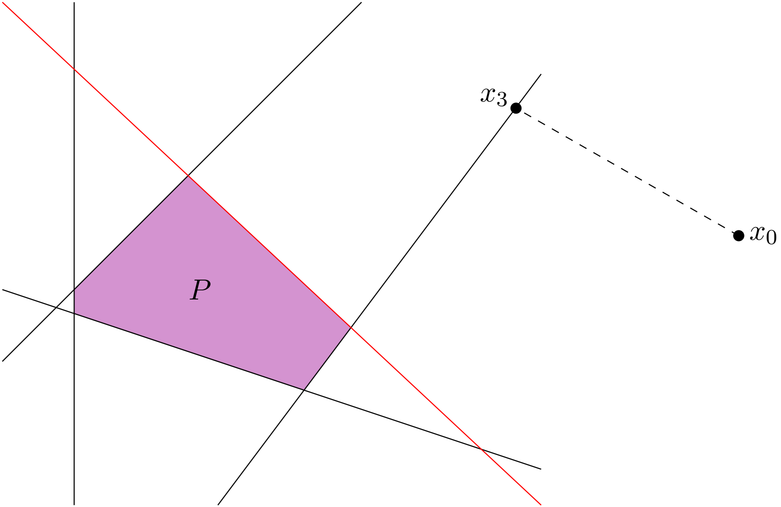
Kaczmarz Method



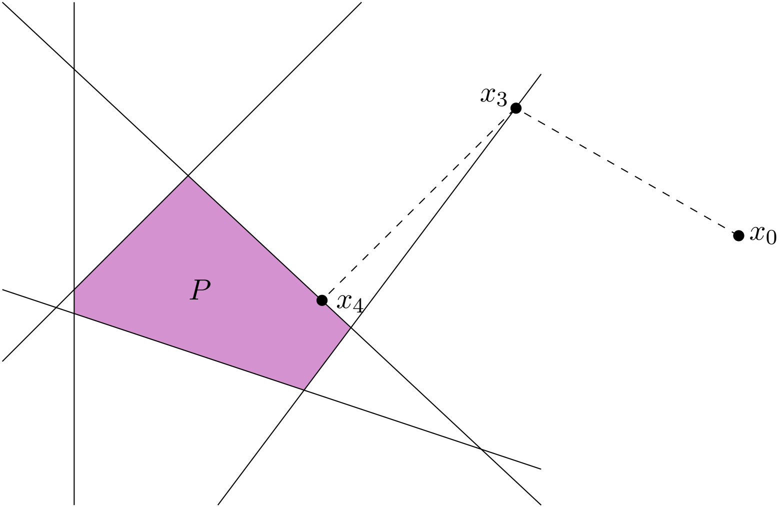
Kaczmarz Method



Kaczmarz Method



Kaczmarz Method

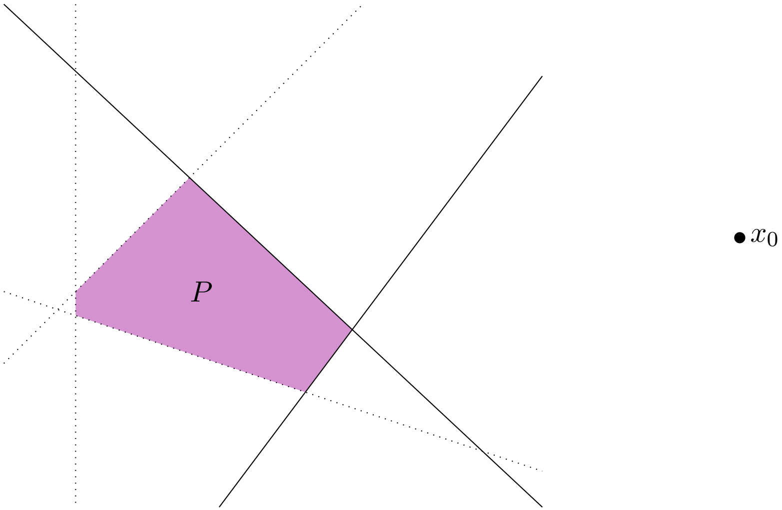


A Hybrid Method (SKM)

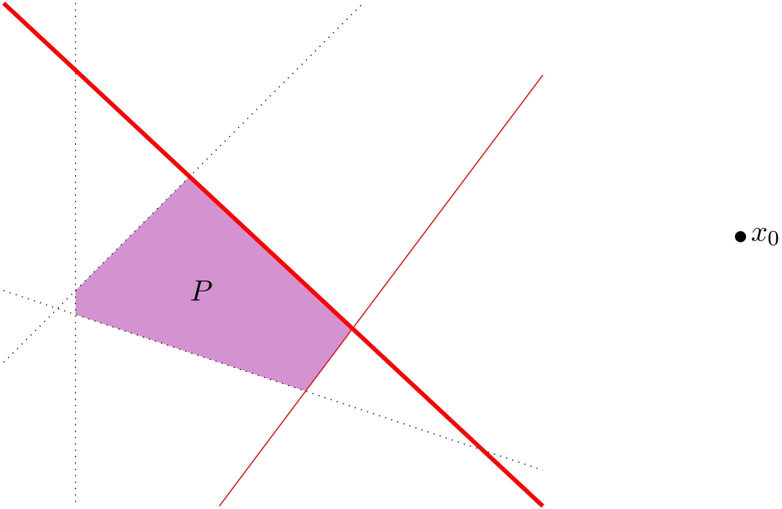
Given $x_0 \in \mathbb{R}^n$, fix $0 < \lambda \leq 2$ and iteratively construct approximations to P in the following way:

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.

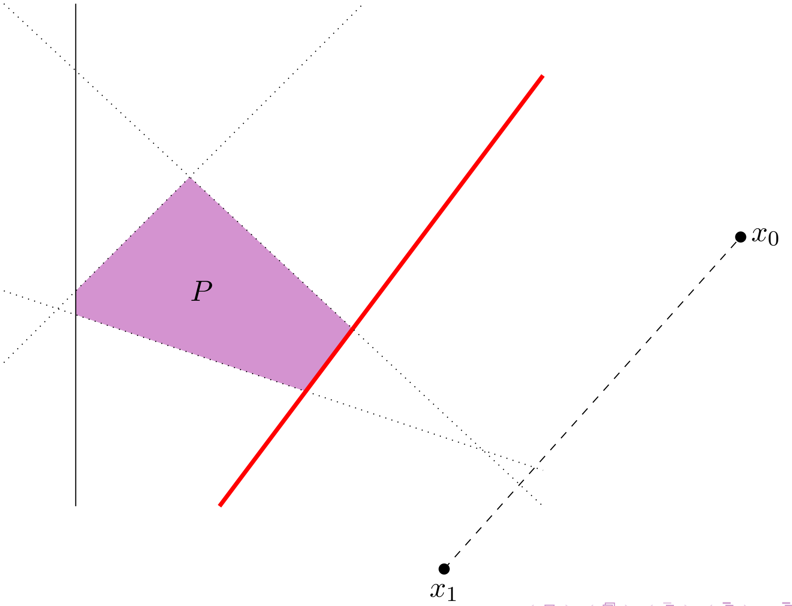
A Hybrid Method



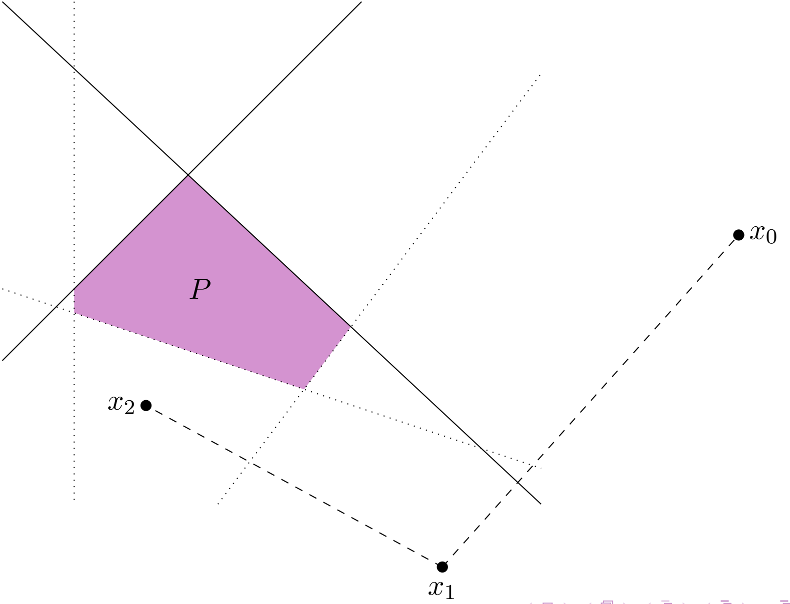
A Hybrid Method



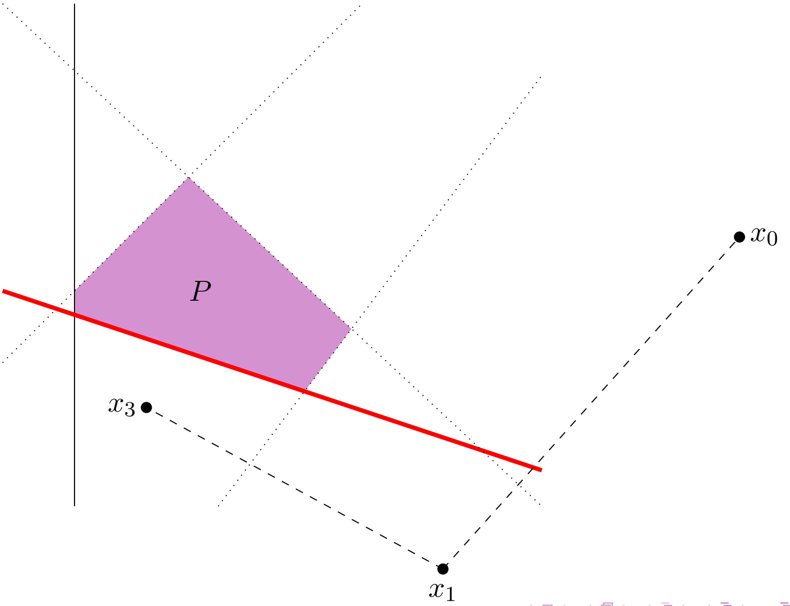
A Hybrid Method



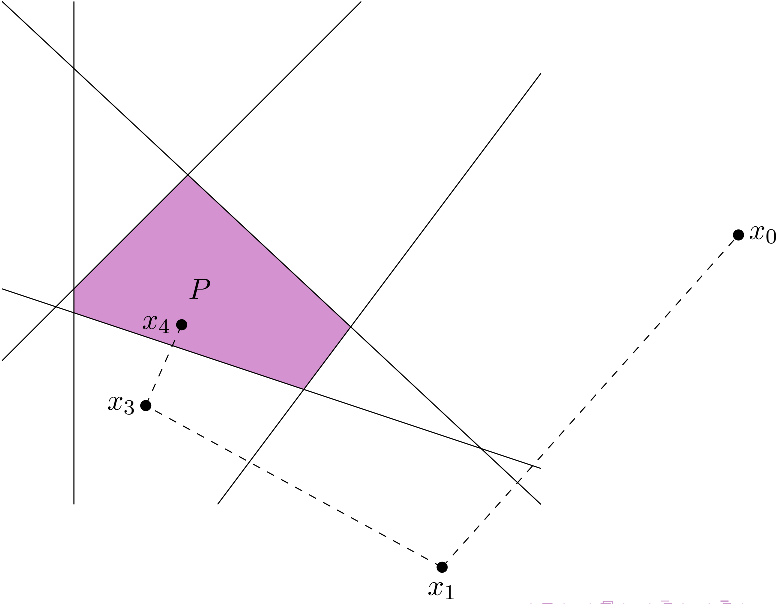
A Hybrid Method



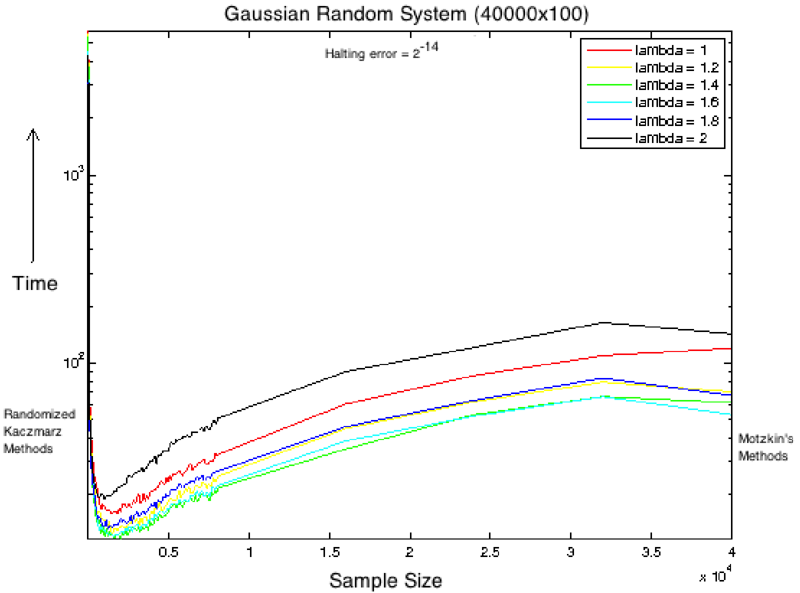
A Hybrid Method



A Hybrid Method



Experimental Results



Previously Known Convergence Rates

Theorem (Relaxation Method, Agmon 1954)

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

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

Previously Known Convergence Rates

Theorem (Relaxation Method, Agmon 1954)

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

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

The *Hoffman constant*, L_2 is an error bound defined as the minimum constant that satisfies

$$d(x, P) \leq L_2 ||(Ax - b)^+||_2.$$

Previously Known Convergence Rates

Theorem (Relaxation Method, Agmon 1954)

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

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

Theorem (Random Kaczmarz Method, Lewis, Leventhal 2008)

If $P := \{x|Ax \leq b\}$ 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 converge 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 satisfed by x_{k-1}. 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 nondegenerate (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 (Goffin 1980, Telgen 1982)

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

*with $x_0 = 0$

Expected Finiteness of SKM methods

Theorem (De Loera, H., Needell)

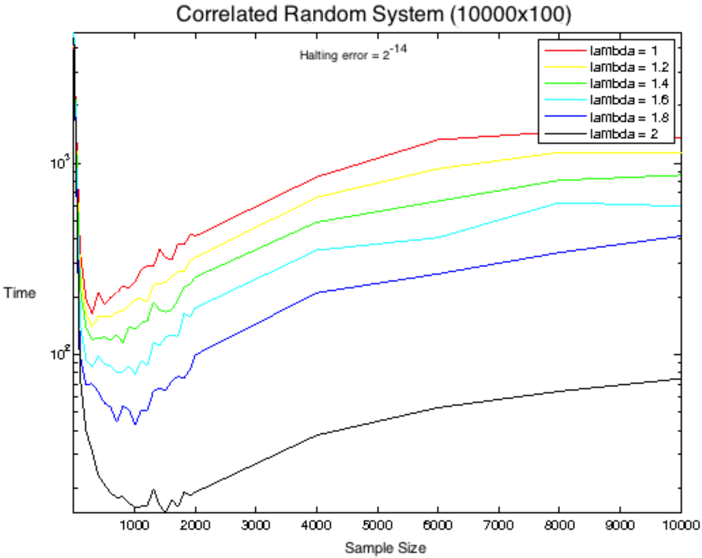
If the rational, normalized system $Ax \leq b$ is feasible (A, b with binary encoding size Σ), then the SKM methods can detect feasibility.

The expected number of steps required for the SKM methods with projection parameter $0 < \lambda < 2$ to detect feasibility is no more than*

$$\left\lceil \frac{4\Sigma - 4 - \log n}{\log\left(\frac{m^2 L_2^2}{m^2 L_2^2 - (2\lambda - \lambda^2)}\right)} \right\rceil.$$

*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.

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.
3. Explore connections of SKM to variants of randomized coordinate descent in the dual variable space.

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.