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
I think about problems of the sort:

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad g(x) \leq 0
\end{align*}
\]
I think about problems of the sort:

\[
\begin{align*}
\min f(x) \\
\text{s.t. } g(x) &\leq 0
\end{align*}
\]

These sorts of problems are all around us!
I think about problems of the sort:

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad g(x) \leq 0
\end{align*}
\]

These sorts of problems are all around us!

Today we’ll consider a specific form of optimization problem...
I think about problems of the sort:

\[
\begin{align*}
\text{min } & \quad c^T x \\
\text{s.t. } & \quad Ax \leq b
\end{align*}
\] (LP)
I think about problems of the sort:

\[
\begin{align*}
\min & \quad c^T x \\
\text{s.t.} & \quad Ax \leq b
\end{align*}
\]

\((LP)\)

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

I think about problems of the sort:

\[
\begin{align*}
\text{min} & \quad c^T x \\
\text{s.t.} & \quad Ax \leq b
\end{align*}
\]

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

But, even this can be simplified...
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.
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!
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.
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 $P := \{x \in \mathbb{R}^n : Ax \leq b\}$ 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 $i_k := \arginf_{i \in [m]} a_i^T x_{k-1} - b_i$.

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

\[ P \]

\[ x_0 \]
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 $\frac{||a_{i_{k}}||^2}{||A||^2_F}$.

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 $\beta$ constraints chosen uniformly at random from among the rows of $A$.

3. From among these $\beta$ rows, choose $i_k := \arg\max_{i \in \tau_k} a_i^T x_{k-1} - b_i$.

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

\( x_0 \)

\( P \)
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

Gaussian Random System (40000x100)

Halting error = $2^{-14}$

Time

Sample Size

Randomized Kaczmarz Methods

Motzkin's Methods

- $\lambda = 1$
- $\lambda = 1.2$
- $\lambda = 1.4$
- $\lambda = 1.6$
- $\lambda = 1.8$
- $\lambda = 2$
Motzkin’s Method Convergence Rate

Theorem (Agmon)

For a normalized system, $||a_i|| = 1$ for all $i = 1, \ldots, m$, 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, \( P := \{ x \mid Ax \leq b \} \), is nonempty then the Randomized Kaczmarz method with relaxation parameter \( \lambda \) 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 $\beta$ 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 satisfied by $x_{k-1}$. Clearly then,

$$
\mathbb{E}[d(x_k, P)^2] \leq \left( 1 - \frac{2\lambda - \lambda^2}{mL_2^2} \right)^k d(x_0, P)^2.
$$
**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 $eta \leq m - n$ is guaranteed an increased convergence rate after some $K$:

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

1. Provide theoretical guidance for selection of the optimal sample size, $\beta$, and optimal overshooting parameter, $\lambda$, 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, $\beta$, and optimal overshooting parameter, $\lambda$ for a given (class of) system(s).
Future work:

1. Provide theoretical guidance for selection of the optimal sample size, $\beta$, and optimal overshooting parameter, $\lambda$ 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


Angenaherte auflösung von systemen linearer gleichungen.


Randomized methods for linear constraints: convergence rates and conditioning.


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


The relaxation method for linear inequalities.


Randomized kaczmarz solver for noisy linear systems.

REFERENCES II


Stochastic gradient descent and the randomized kaczmarz algorithm.

submitted.


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

Linear Algebra Appl.


Theory of linear and integer programming.


A Wiley-Interscience Publication.


A randomized kaczmarz algorithm with exponential convergence.