Acceleration of Randomized Kaczmarz Method

Deanna Needell [Joint work with Y. Eldar]

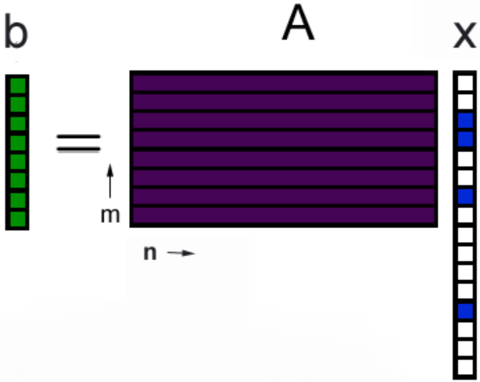
Stanford University

BIRS Banff, March 2011

Setup

Setup

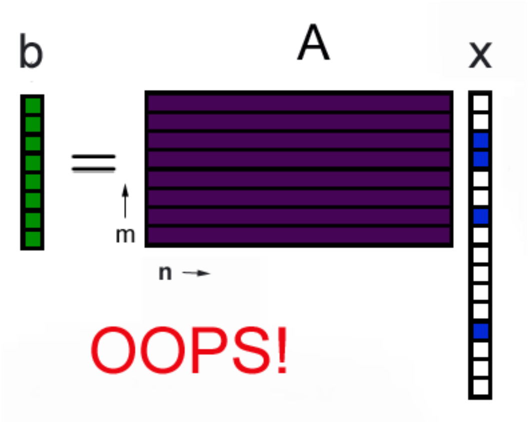
Let $Ax = b$ be an *overdetermined* consistent system of equations



Setup

Setup

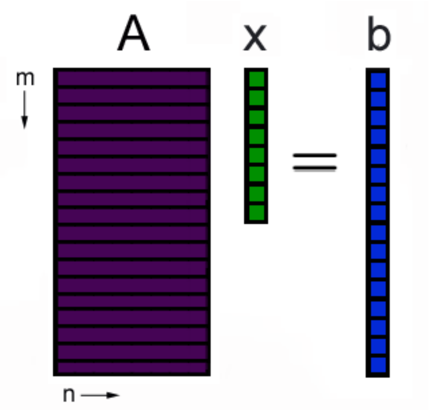
Let $Ax = b$ be an *overdetermined* consistent system of equations



Setup

Setup

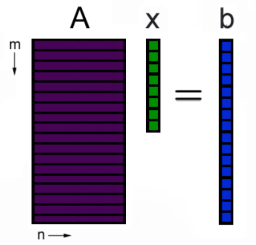
Let $Ax = b$ be an *overdetermined* consistent system of equations



Setup

Setup

Let $Ax = b$ be an *overdetermined* consistent system of equations



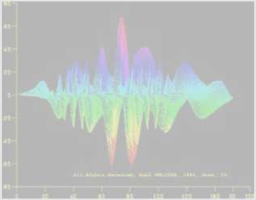
Goal

From A and b we wish to recover unknown x. Assume $m \gg n$.

Method

Kaczmarz

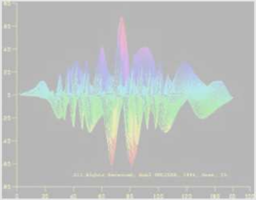
- The Kaczmarz method is an iterative method used to solve $Ax = b$.
- Due to its speed and simplicity, it's used in a variety of applications.



Method

Kaczmarz

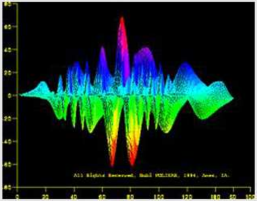
- The Kaczmarz method is an iterative method used to solve $Ax = b$.
- Due to its speed and simplicity, it's used in a variety of applications.



Method

Kaczmarz

- The Kaczmarz method is an iterative method used to solve $Ax = b$.
- Due to its speed and simplicity, it's used in a variety of applications.



Method

$$\begin{bmatrix} \text{———} \; a_1 \; \text{———} \\ \text{———} \; a_2 \; \text{———} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ \text{———} \; a_m \; \text{———} \end{bmatrix} \cdot \begin{bmatrix} x \\ \\ \\ \end{bmatrix} = \begin{bmatrix} b[1] \\ b[2] \\ \cdot \\ \cdot \\ \cdot \\ b[m] \end{bmatrix}$$

Kaczmarz

1. Start with initial guess x_0
2. $x_{k+1} = x_k + \frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2^2} a_i$ where $i = (k \bmod m) + 1$
3. Repeat (2)

Method

$$\begin{bmatrix} \text{---} & a_1 & \text{---} \\ \text{---} & a_2 & \text{---} \\ \vdots & \vdots \ \ddots & \vdots \\ \text{---} & a_m & \text{---} \end{bmatrix} \cdot \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} b[1] \\ b[2] \\ \cdot \\ \cdot \\ \cdot \\ b[m] \end{bmatrix}$$

Kaczmarz

1. Start with initial guess x_0
2. $x_{k+1} = x_k + \frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2^2} a_i$ where $i = (k \mod m) + 1$
3. Repeat (2)

Method

$$\begin{bmatrix} \text{——} & a_1 & \text{——} \\ \text{——} & a_2 & \text{——} \\ \vdots & \vdots \ \ddots & \vdots \\ \text{——} & a_m & \text{——} \end{bmatrix} \cdot \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} b[1] \\ b[2] \\ \cdot \\ \cdot \\ \cdot \\ b[m] \end{bmatrix}$$

Kaczmarz

1. Start with initial guess x_0
2. $x_{k+1} = x_k + \frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2^2} a_i$ where $i = (k \bmod m) + 1$
3. Repeat (2)

Method

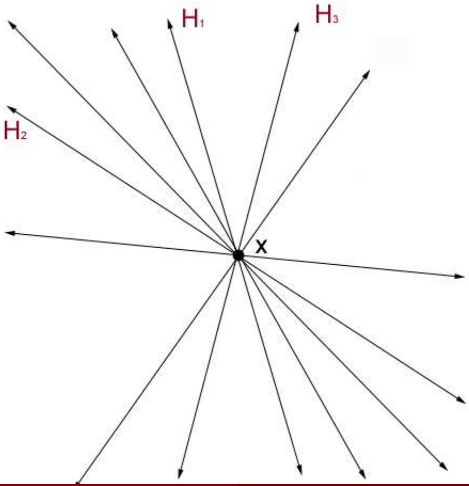
$$\begin{bmatrix} \text{---} & a_1 & \text{---} \\ \text{---} & a_2 & \text{---} \\ \vdots & \vdots & \ddots & \vdots \\ \text{---} & a_m & \text{---} \end{bmatrix} \cdot \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} b[1] \\ b[2] \\ \vdots \\ b[m] \end{bmatrix}$$

Kaczmarz

1. Start with initial guess x_0
2. $x_{k+1} = x_k + \frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2^2} a_i$ where $i = (k \bmod m) + 1$
3. Repeat (2)

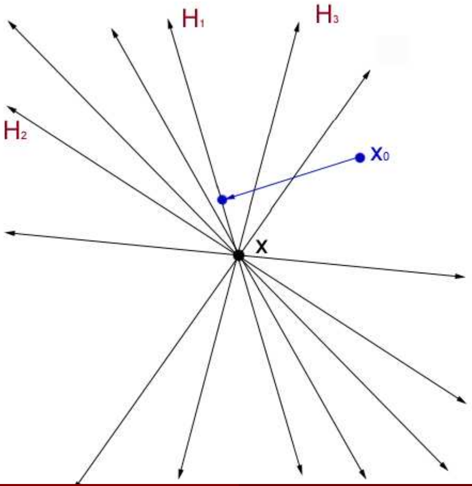
Geometrically

Denote $H_i = \{w : \langle a_i, w \rangle = b[i]\}$.



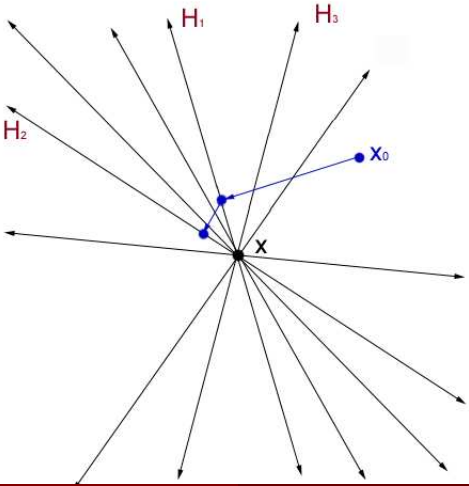
Geometrically

Denote $H_i = \{w : \langle a_i, w \rangle = b[i]\}$.



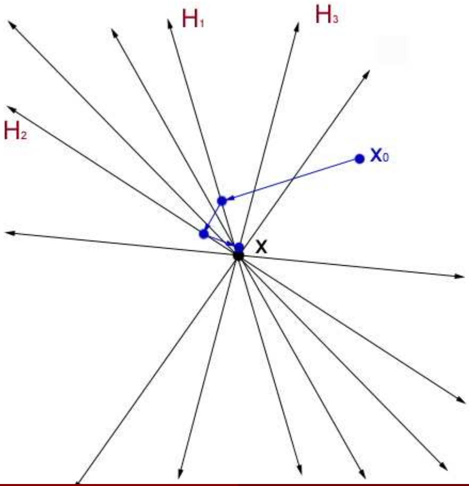
Geometrically

Denote $H_i = \{w : \langle a_i, w \rangle = b[i]\}$.



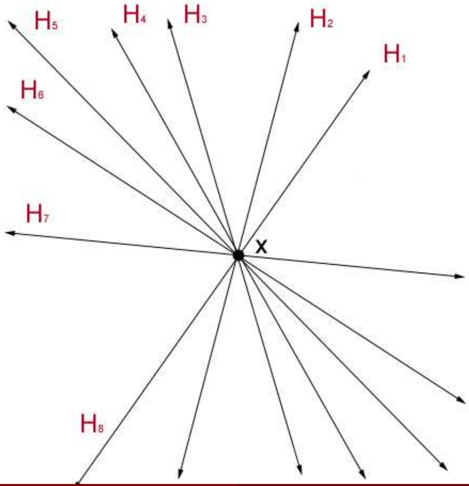
Geometrically

Denote $H_i = \{w : \langle a_i, w\rangle = b[i]\}$.



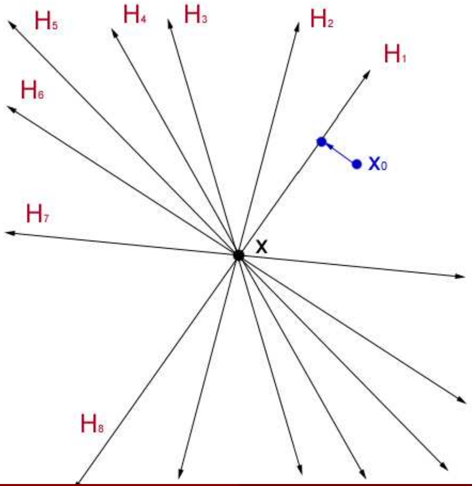
But what if...

Denote $H_i = \{w : \langle a_i, w\rangle = b[i]\}$.



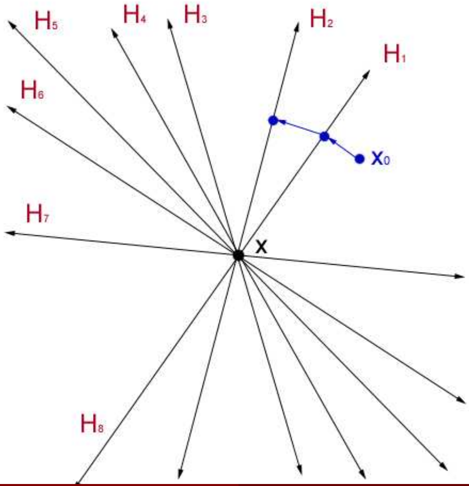
But what if...

Denote $H_i = \{w : \langle a_i, w\rangle = b[i]\}$.



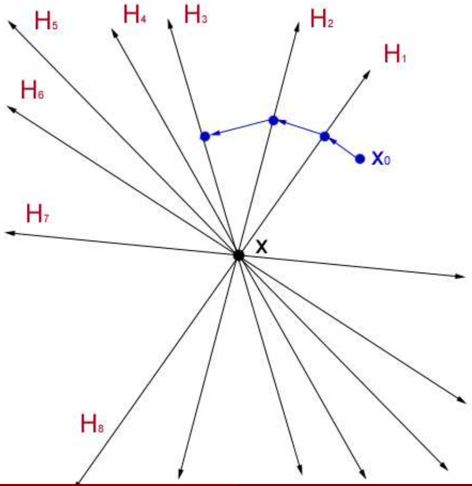
But what if...

Denote $H_i = \{w : \langle a_i, w \rangle = b[i]\}$.



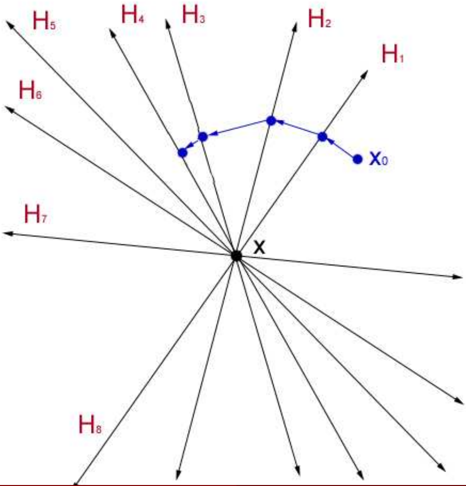
But what if...

Denote $H_i = \{w : \langle a_i, w\rangle = b[i]\}$.



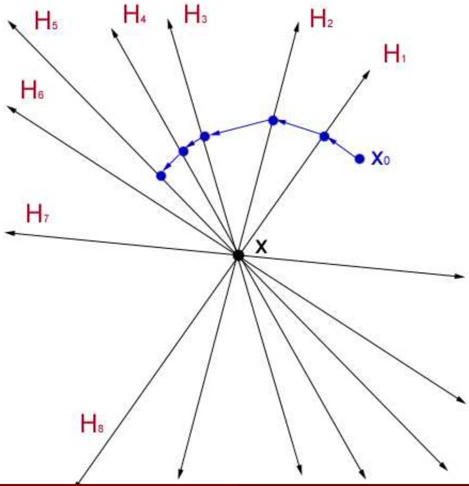
But what if...

Denote $H_i = \{w : \langle a_i, w\rangle = b[i]\}$.



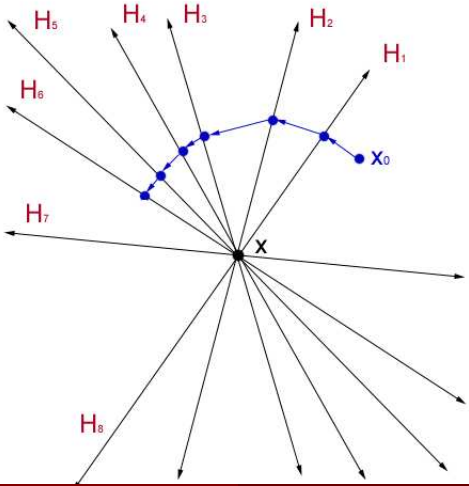
But what if...

Denote $H_i = \{w : \langle a_i, w \rangle = b[i]\}$.



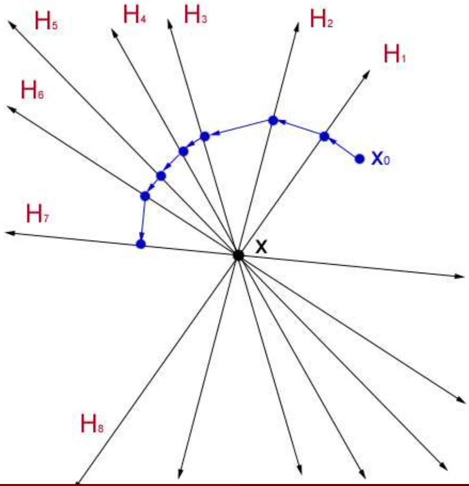
But what if...

Denote $H_i = \{w : \langle a_i, w \rangle = b[i]\}$.



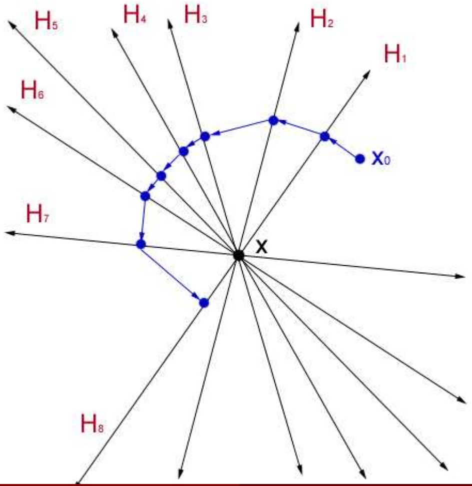
But what if...

Denote $H_i = \{w : \langle a_i, w \rangle = b[i]\}$.



But what if...

Denote $H_i = \{w : \langle a_i, w \rangle = b[i]\}$.



Randomized Kaczmarz

$$\begin{bmatrix} \text{———} & a_1 & \text{———} & \\ \text{———} & a_2 & \text{———} & \\ \vdots & \vdots & \ddots & \vdots \\ \text{———} & a_m & \text{———} & \end{bmatrix} \cdot \begin{bmatrix} x \\ \\ \\ \end{bmatrix} = \begin{bmatrix} b[1] \\ b[2] \\ \cdot \\ \cdot \\ \cdot \\ b[m] \end{bmatrix}$$

Kaczmarz

1. Start with initial guess x_0
2. $x_{k+1} = x_k + \frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2^2} a_i$ where i is chosen *randomly*
3. Repeat (2)

Randomized Kaczmarz

$$\begin{bmatrix} \text{---} \; a_1 \; \text{---} \\ \text{---} \; a_2 \; \text{---} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ \text{---} \; a_m \; \text{---} \end{bmatrix} \cdot \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} b[1] \\ b[2] \\ \vdots \\ b[m] \end{bmatrix}$$

Kaczmarz

1. Start with initial guess x_0
2. $x_{k+1} = x_k + \frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2^2} a_i$ where i is chosen *randomly*
3. Repeat (2)

Randomized Kaczmarz

Strohmer-Vershynin

1. Start with initial guess x_0
2. $x_{k+1} = x_k + \frac{b_p - \langle a_p, x_k \rangle}{\|a_p\|_2^2} a_p$ where $\mathbb{P}(p = i) = \frac{\|a_i\|_2^2}{\|A\|_F^2}$
3. Repeat (2)

Randomized Kaczmarz

Strohmer-Vershynin

1. Start with initial guess x_0
2. $x_{k+1} = x_k + \frac{b_p - \langle a_p, x_k \rangle}{\|a_p\|_2^2} a_p$ where $\mathbb{P}(p = i) = \frac{\|a_i\|_2^2}{\|A\|_F^2}$
3. Repeat (2)

Randomized Kaczmarz (RK)

Strohmer-Vershynin

- Let $R = \|A^{-1}\|^2\|A\|_F^2$ $(\|A^{-1}\| \stackrel{\text{def}}{=} \inf\{M : M\|Ax\|_2 \geq \|x\|_2 \text{ for all } x\})$
- Then $\mathbb{E}\|x_k - x\|_2^2 \leq \left(1 - \frac{1}{R}\right)^k \|x_0 - x\|_2^2$
- Well conditioned $A \rightarrow$ Convergence in $O(n)$ iterations $\rightarrow$ $O(n^2)$ total runtime.
- Better than $O(mn^2)$ runtime for Gaussian elimination and empirically often faster than Conjugate Gradient.

Randomized Kaczmarz (RK)

Strohmer-Vershynin

- Let $R = \|A^{-1}\|^2\|A\|_F^2$ ($\|A^{-1}\| \stackrel{\text{def}}{=} \inf\{M : M\|Ax\|_2 \geq \|x\|_2 \text{ for all } x\}$)
- Then $\mathbb{E}\|x_k - x\|_2^2 \leq \left(1 - \frac{1}{R}\right)^k \|x_0 - x\|_2^2$
- Well conditioned $A \to$ Convergence in $\mathrm{O}(n)$ iterations $\to \mathrm{O}(n^2)$ total runtime.
- Better than $\mathrm{O}(mn^2)$ runtime for Gaussian elimination and empirically often faster than Conjugate Gradient.

Randomized Kaczmarz (RK)

Strohmer-Vershynin

- Let $R = \|A^{-1}\|^2 \|A\|_F^2$ $(\|A^{-1}\| \stackrel{\text{def}}{=} \inf\{M : M\|Ax\|_2 \geq \|x\|_2 \text{ for all } x\})$
- Then $\mathbb{E}\|x_k - x\|_2^2 \leq \left(1 - \frac{1}{R}\right)^k \|x_0 - x\|_2^2$
- Well conditioned $A \rightarrow$ Convergence in $\mathrm{O}(n)$ iterations $\rightarrow$ $\mathrm{O}(n^2)$ total runtime.
- Better than $\mathrm{O}(mn^2)$ runtime for Gaussian elimination and empirically often faster than Conjugate Gradient.

Randomized Kaczmarz (RK)

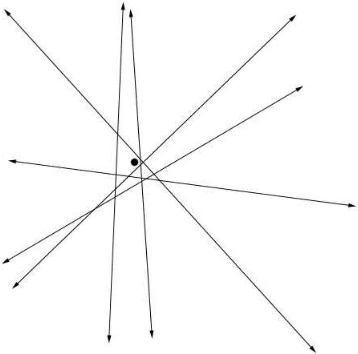
Strohmer-Vershynin

- Let $R = \|A^{-1}\|^2\|A\|_F^2$ ($\|A^{-1}\| \stackrel{\text{def}}{=} \inf\{M : M\|Ax\|_2 \geq \|x\|_2 \text{ for all } x\}$)
- Then $\mathbb{E}\|x_k - x\|_2^2 \leq \left(1 - \frac{1}{R}\right)^k \|x_0 - x\|_2^2$
- Well conditioned $A \to$ Convergence in $\mathrm{O}(n)$ iterations $\to \mathrm{O}(n^2)$ total runtime.
- Better than $\mathrm{O}(mn^2)$ runtime for Gaussian elimination and empirically often faster than Conjugate Gradient.

Randomized Kaczmarz (RK) with noise

System with noise

We now consider the consistent system $Ax = b$ corrupted by noise to form the possibly inconsistent system $Ax \approx b + z$.



Randomized Kaczmarz (RK) with noise

Theorem [N]

- Let $Ax = b$ be corrupted with noise: $Ax \approx b + z$. Then

$$\mathbb{E}\|x_k - x\|_2 \leq \left(1 - \frac{1}{R}\right)^{k/2} \|x_0 - x\|_2 + \sqrt{R}\gamma,$$

where $\gamma = \max_i \frac{|z[i]|}{\|a_i\|_2}$.

- This bound is sharp and attained in simple examples.

Randomized Kaczmarz (RK) with noise

Theorem [N]

- Let $Ax = b$ be corrupted with noise: $Ax \approx b + z$. Then

$$\mathbb{E}\|x_k - x\|_2 \leq \left(1 - \frac{1}{R}\right)^{k/2} \|x_0 - x\|_2 + \sqrt{R}\gamma,$$

where $\gamma = \max_i \frac{|z[i]|}{\|a_i\|_2}$.
- This bound is sharp and attained in simple examples.

Randomized Kaczmarz (RK) with noise

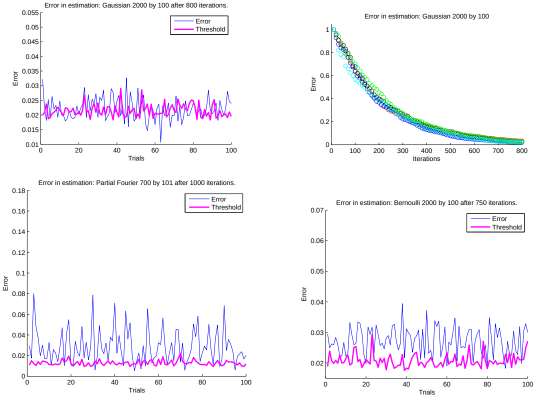


Figure: Comparison between actual error (blue) and predicted threshold (pink). Scatter plot shows exponential convergence over several trials.

Even better convergence? : Noiseless case revisited

- Recall $x_{k+1} = x_k + \frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2^2} a_i$
- Since these projections are orthogonal, the optimal projection is one that maximizes $\|x_{k+1} - x_k\|_2$.
- Therefore we choose i maximizing $|\frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2}|$.
- Too costly $\rightarrow$ Project onto low dimensional subspace.
- Use the low dimensional representations to predict the optimal projection.

Even better convergence? : Noiseless case revisited

- Recall $x_{k+1} = x_k + \frac{b[i]-\langle a_i, x_k\rangle}{\|a_i\|_2^2} a_i$
- Since these projections are orthogonal, the optimal projection is one that maximizes $\|x_{k+1} - x_k\|_2$.
- Therefore we choose i maximizing $\left|\frac{b[i]-\langle a_i, x_k\rangle}{\|a_i\|_2}\right|$.
- Too costly → Project onto low dimensional subspace.
- Use the low dimensional representations to predict the optimal projection.

Even better convergence? : Noiseless case revisited

- Recall $x_{k+1} = x_k + \frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2^2} a_i$
- Since these projections are orthogonal, the optimal projection is one that maximizes $\|x_{k+1} - x_k\|_2$.
- Therefore we choose i maximizing $\left|\frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2}\right|$.
- Too costly $\rightarrow$ Project onto low dimensional subspace.
- Use the low dimensional representations to predict the optimal projection.

Even better convergence? : Noiseless case revisited

- Recall $x_{k+1} = x_k + \frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2^2} a_i$
- Since these projections are orthogonal, the optimal projection is one that maximizes $\|x_{k+1} - x_k\|_2$.
- Therefore we choose i maximizing $\left|\frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2}\right|$.
- Too costly $\rightarrow$ Project onto low dimensional subspace.
- Use the low dimensional representations to predict the optimal projection.

Even better convergence? : Noiseless case revisited

- Recall $x_{k+1} = x_k + \frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2^2} a_i$
- Since these projections are orthogonal, the optimal projection is one that maximizes $\|x_{k+1} - x_k\|_2$.
- Therefore we choose i maximizing $\left|\frac{b[i] - \langle a_i, x_k \rangle}{\|a_i\|_2}\right|$.
- Too costly $\rightarrow$ Project onto low dimensional subspace.
- Use the low dimensional representations to predict the optimal projection.

Even better convergence? : Noiseless case revisited

- Recall $x_{k+1} = x_k + \frac{b[i]-\langle a_i, x_k \rangle}{\|a_i\|_2^2} a_i$
- Since these projections are orthogonal, the optimal projection is one that maximizes $\|x_{k+1} - x_k\|_2$.
- Therefore we choose i maximizing $\left|\frac{b[i]-\langle a_i, x_k \rangle}{\|a_i\|_2}\right|$.
- Too costly $\rightarrow$ Project onto low dimensional subspace.
- Use the low dimensional representations to predict the optimal projection.

JL Dimension Reduction

Johnson-Lindenstrauss Lemma

Let $\delta > 0$ and let S be a finite set of points in $\mathbb{R}^n$. Then for any d satisfying

$$d \geq C\frac{\log |S|}{\delta^2}, \qquad (1)$$

there exists a Lipschitz mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^d$ such that

$$(1-\delta)\|s_i - s_j\|_2^2 \leq \|\Phi(s_i) - \Phi(s_j)\|_2^2 \leq (1+\delta)\|s_i - s_j\|_2^2, \qquad (2)$$

for all $s_i, s_j \in S$.

JL Dimension Reduction

Moreover

- In the proof of the JL Lemma the map Φ is chosen as the projection onto a random d-dimensional subspace of $\mathbb{R}^n$. Now many known distributions will yield such a projection.
- Recently, transforms with fast multiplies have also been shown to satisfy the JL Lemma [Ailon-Chazelle, Hinrichs-Vybiral, Ailon-Liberty, Krahmer-Ward, ...]

Perform Reduction

Choose such a $d \times n$ projector Φ and during preprocessing set $\alpha_i = \Phi a_i$.

JL Dimension Reduction

Moreover

- In the proof of the JL Lemma the map Φ is chosen as the projection onto a random d-dimensional subspace of $\mathbb{R}^n$. Now many known distributions will yield such a projection.
- Recently, transforms with fast multiplies have also been shown to satisfy the JL Lemma [Ailon-Chazelle, Hinrichs-Vybiral, Ailon-Liberty, Krahmer-Ward, ...]

Perform Reduction

Choose such a $d \times n$ projector Φ and during preprocessing set $\alpha_i = \Phi a_i$.

JL Dimension Reduction

Moreover

- In the proof of the JL Lemma the map Φ is chosen as the projection onto a random d-dimensional subspace of $\mathbb{R}^n$. Now many known distributions will yield such a projection.
- Recently, transforms with fast multiplies have also been shown to satisfy the JL Lemma [Ailon-Chazelle, Hinrichs-Vybiral, Ailon-Liberty, Krahmer-Ward, ...]

Perform Reduction

Choose such a $d \times n$ projector Φ and during preprocessing set $\alpha_i = \Phi a_i$.

RK via Johnson-Lindenstrauss (RKJL) [N-Eldar]

Select: Select n rows so that each row a_i is chosen with probability $\|a_i\|_2^2/\|A\|_F^2$. For each set

$$\gamma_i = \frac{|b[i] - \langle \alpha_i, \Phi x_k \rangle|}{\|\alpha_i\|_2},$$

and set $j = \text{argmax}_i \gamma_i$.

Test: For a_j and the first row a_l selected set

$$\gamma_j^* = \frac{|b[j] - \langle a_j, x_k \rangle|}{\|a_j\|_2} \quad \text{and} \quad \gamma_l^* = \frac{|b[l] - \langle a_l, x_k \rangle|}{\|a_l\|_2}.$$

If $\gamma_l^* > \gamma_j^*$, set $j = l$.

Project: Set

$$x_{k+1} = x_k + \frac{b[j] - \langle a_j, x_k \rangle}{\|a_j\|_2^2} a_j.$$

Update: Set $k = k + 1$ and repeat.

RK via Johnson-Lindenstrauss (RKJL) [N-Eldar]

Select: Select n rows so that each row a_i is chosen with probability $\|a_i\|_2^2/\|A\|_F^2$. For each set

$$\gamma_i = \frac{|b[i] - \langle \alpha_i, \Phi x_k \rangle|}{\|\alpha_i\|_2},$$

and set $j = \text{argmax}_i \, \gamma_i$.

Test: For a_j and the first row a_l selected set

$$\gamma_j^* = \frac{|b[j] - \langle a_j, x_k \rangle|}{\|a_j\|_2} \quad \text{and} \quad \gamma_l^* = \frac{|b[l] - \langle a_l, x_k \rangle|}{\|a_l\|_2}.$$

If $\gamma_l^* > \gamma_j^*$, set $j = l$.

Project: Set

$$x_{k+1} = x_k + \frac{b[j] - \langle a_j, x_k \rangle}{\|a_j\|_2^2} a_j.$$

Update: Set $k = k + 1$ and repeat.

RK via Johnson-Lindenstrauss (RKJL) [N-Eldar]

Select: Select n rows so that each row a_i is chosen with probability $\|a_i\|_2^2/\|A\|_F^2$. For each set

$$\gamma_i = \frac{|b[i] - \langle \alpha_i, \Phi x_k \rangle|}{\|\alpha_i\|_2},$$

and set $j = \text{argmax}_i\, \gamma_i$.

Test: For a_j and the first row a_l selected set

$$\gamma_j^* = \frac{|b[j] - \langle a_j, x_k \rangle|}{\|a_j\|_2} \quad \text{and} \quad \gamma_l^* = \frac{|b[l] - \langle a_l, x_k \rangle|}{\|a_l\|_2}.$$

If $\gamma_l^* > \gamma_j^*$, set $j = l$.

Project: Set

$$x_{k+1} = x_k + \frac{b[j] - \langle a_j, x_k \rangle}{\|a_j\|_2^2} a_j.$$

Update: Set $k = k + 1$ and repeat.

RK via Johnson-Lindenstrauss (RKJL) [N-Eldar]

Select: Select n rows so that each row a_i is chosen with probability $\|a_i\|_2^2/\|A\|_F^2$. For each set

$$\gamma_i = \frac{|b[i] - \langle \alpha_i, \Phi x_k \rangle|}{\|\alpha_i\|_2},$$

and set $j = \text{argmax}_i\, \gamma_i$.

Test: For a_j and the first row a_l selected set

$$\gamma_j^* = \frac{|b[j] - \langle a_j, x_k \rangle|}{\|a_j\|_2} \quad \text{and} \quad \gamma_l^* = \frac{|b[l] - \langle a_l, x_k \rangle|}{\|a_l\|_2}.$$

If $\gamma_l^* > \gamma_j^*$, set $j = l$.

Project: Set

$$x_{k+1} = x_k + \frac{b[j] - \langle a_j, x_k \rangle}{\|a_j\|_2^2} a_j.$$

Update: Set $k = k + 1$ and repeat.

Runtime

Select:
- Calculate Φx_k : In general $\mathrm{O}(nd)$
- Calculate γ_i for each i (of n): $\mathrm{O}(nd)$

Test: Calculate γ_j^* and γ_l^*: $\mathrm{O}(n)$

Project: Calculate x_{k+1}: $\mathrm{O}(n)$

Overall Runtime

Since each iteration takes $\mathrm{O}(nd)$, we have convergence in $\mathrm{O}(n^2 d)$.

Choosing parameter d

Lemma: Choice of d

Let Φ be the $n \times d$ (Gaussian) matrix with $d = C\delta^{-2}\log(n)$ as in the RKJL method. Set $\gamma_i = \langle \Phi a_i, \Phi x_k \rangle$ also as in the method. Then $|\gamma_i - \langle a_i, x_k \rangle| \leq 2\delta$ for all i and k in the first $O(n)$ iterations of RKJL.

Low Risk

This shows *worst case* expected convergence in at most $O(n^2 \log n)$ time, and of course in most cases one expects far faster convergence.

Choosing parameter d

Lemma: Choice of d

Let Φ be the $n \times d$ (Gaussian) matrix with $d = C\delta^{-2}\log(n)$ as in the RKJL method. Set $\gamma_i = \langle \Phi a_i, \Phi x_k \rangle$ also as in the method. Then $|\gamma_i - \langle a_i, x_k \rangle| \leq 2\delta$ for all i and k in the first $O(n)$ iterations of RKJL.

Low Risk

This shows *worst case* expected convergence in at most $O(n^2 \log n)$ time, and of course in most cases one expects far faster convergence.

Choosing parameter d

Lemma: Choice of d

Let Φ be the $n \times d$ (Gaussian) matrix with $d = C\delta^{-2}\log(n)$ as in the RKJL method. Set $\gamma_i = \langle \Phi a_i, \Phi x_k \rangle$ also as in the method. Then $|\gamma_i - \langle a_i, x_k \rangle| \leq 2\delta$ for all i and k in the first $\mathrm{O}(n)$ iterations of RKJL.

Low Risk

This shows *worst case* expected convergence in at most $\mathrm{O}(n^2 \log n)$ time, and of course in most cases one expects far faster convergence.

Choosing parameter d

Lemma: Choice of d

Let Φ be the $n \times d$ (Gaussian) matrix with $d = C\delta^{-2}\log(n)$ as in the RKJL method. Set $\gamma_i = \langle \Phi a_i, \Phi x_k \rangle$ also as in the method. Then $|\gamma_i - \langle a_i, x_k \rangle| \leq 2\delta$ for all i and k in the first $\mathrm{O}(n)$ iterations of RKJL.

Low Risk

This shows *worst case* expected convergence in at most $\mathrm{O}(n^2 \log n)$ time, and of course in most cases one expects far faster convergence.

Analytical Justification

Theorem [Assuming row normalization]

Fix an estimation x_k and denote by x_{k+1} and x^*_{k+1} the next estimations using the RKJL and the standard RK method, respectively. Set $\gamma^*_j = |\langle a_j, x_k \rangle|^2$ and reorder these so that $\gamma^*_1 \geq \gamma^*_2 \geq \ldots \geq \gamma^*_m$. Then when $d = C\delta^{-2} \log n$,

$$\mathbb{E}\|x_{k+1} - x\|_2^2 \leq \min \left[\mathbb{E}\|x^*_{k+1} - x\|_2^2 - \sum_{j=1}^{m} \left(p_j - \frac{1}{m} \right) \gamma^*_j + 2\delta, \quad \mathbb{E}\|x^*_{k+1} - x\|_2^2 \right]$$

where

$$p_j = \begin{cases} \frac{\binom{m-j}{n-1}}{\binom{m}{n}}, & j \leq m - n + 1 \\ 0, & j > m - n + 1 \end{cases}$$

are non-negative values satisfying $\sum_{j=1}^{m} p_j = 1$ and $p_1 \geq p_2 \geq \ldots \geq p_m = 0$.

Analytical Justification

Theorem [Assuming row normalization]

Fix an estimation x_k and denote by x_{k+1} and x^*_{k+1} the next estimations using the RKJL and the standard RK method, respectively. Set $\gamma^*_j = |\langle a_j, x_k\rangle|^2$ and reorder these so that $\gamma^*_1 \geq \gamma^*_2 \geq \ldots \geq \gamma^*_m$. Then when $d = C\delta^{-2} \log n$,

$$\mathbb{E}\|x_{k+1} - x\|_2^2 \leq \min \left[\mathbb{E}\|x^*_{k+1} - x\|_2^2 - \sum_{j=1}^{m} \left(p_j - \frac{1}{m} \right) \gamma^*_j + 2\delta, \quad \mathbb{E}\|x^*_{k+1} - x\|_2^2 \right]$$

where

$$p_j = \begin{cases} \frac{\binom{m-j}{n-1}}{\binom{m}{n}}, & j \leq m - n + 1 \\ 0, & j > m - n + 1 \end{cases}$$

are non-negative values satisfying $\sum_{j=1}^{m} p_j = 1$ and $p_1 \geq p_2 \geq \ldots \geq p_m = 0$.

Analytical Justification

Theorem [Assuming row normalization]

Fix an estimation x_k and denote by x_{k+1} and x^*_{k+1} the next estimations using the RKJL and the standard RK method, respectively. Set $\gamma^*_j = |\langle a_j, x_k\rangle|^2$ and reorder these so that $\gamma^*_1 \geq \gamma^*_2 \geq \ldots \geq \gamma^*_m$. Then when $d = C\delta^{-2}\log n$,

$$\mathbb{E}\|x_{k+1} - x\|_2^2 \leq \min\left[\mathbb{E}\|x^*_{k+1} - x\|_2^2 - \sum_{j=1}^{m}\left(p_j - \frac{1}{m}\right)\gamma^*_j + 2\delta, \quad \mathbb{E}\|x^*_{k+1} - x\|_2^2\right]$$

where

$$p_j = \begin{cases} \frac{\binom{m-j}{n-1}}{\binom{m}{n}}, & j \leq m - n + 1 \\ 0, & j > m - n + 1 \end{cases}$$

are non-negative values satisfying $\sum_{j=1}^m p_j = 1$ and $p_1 \geq p_2 \geq \ldots \geq p_m = 0$.

Analytical Justification

Theorem [Assuming row normalization]

Fix an estimation x_k and denote by x_{k+1} and x^*_{k+1} the next estimations using the RKJL and the standard RK method, respectively. Set $\gamma_j^* = |\langle a_j, x_k\rangle|^2$ and reorder these so that $\gamma_1^* \geq \gamma_2^* \geq \ldots \geq \gamma_m^*$. Then when $d = C\delta^{-2}\log n$,

$$\mathbb{E}\|x_{k+1} - x\|_2^2 \leq \min\left[\mathbb{E}\|x^*_{k+1} - x\|_2^2 - \sum_{j=1}^{m}\left(p_j - \frac{1}{m}\right)\gamma_j^* + 2\delta, \quad \mathbb{E}\|x^*_{k+1} - x\|_2^2\right]$$

where

$$p_j = \begin{cases} \frac{\binom{m-j}{n-1}}{\binom{m}{n}}, & j \leq m-n+1 \\ 0, & j > m-n+1 \end{cases}$$

are non-negative values satisfying $\sum_{j=1}^{m} p_j = 1$ and $p_1 \geq p_2 \geq \ldots \geq p_m = 0$.

Analytical Justification

Corollary

Fix an estimation x_k and denote by x_{k+1} and x^*_{k+1} the next estimations using the RKJL and the standard method, respectively. Set $\gamma^*_j = |\langle a_j, x_k\rangle|^2$ and reorder these so that $\gamma^*_1 \geq \gamma^*_2 \geq \ldots \geq \gamma^*_m$. Then when exact geometry is preserved ($\delta \to 0$),

$$\mathbb{E}\|x_{k+1} - x\|_2^2 \leq \mathbb{E}\|x^*_{k+1} - x\|_2^2 - \sum_{j=1}^{m}\left(p_j - \frac{1}{m}\right)\gamma^*_j.$$

Analytical Justification

Corollary

Fix an estimation x_k and denote by x_{k+1} and x^*_{k+1} the next estimations using the RKJL and the standard method, respectively. Set $\gamma^*_j = |\langle a_j, x_k \rangle|^2$ and reorder these so that $\gamma^*_1 \geq \gamma^*_2 \geq \ldots \geq \gamma^*_m$. Then when exact geometry is preserved ($\delta \to 0$),

$$\mathbb{E}\|x_{k+1} - x\|_2^2 \leq \mathbb{E}\|x^*_{k+1} - x\|_2^2 - \sum_{j=1}^{m} \left(p_j - \frac{1}{m}\right)\gamma^*_j.$$

Analytical Justification

Corollary

Fix an estimation x_k and denote by x_{k+1} and x^*_{k+1} the next estimations using the RKJL and the standard method, respectively. Set $\gamma^*_j = |\langle a_j, x_k \rangle|^2$ and reorder these so that $\gamma^*_1 \geq \gamma^*_2 \geq \ldots \geq \gamma^*_m$. Then when exact geometry is preserved ($\delta \to 0$),

$$\mathbb{E}\|x_{k+1} - x\|_2^2 \leq \mathbb{E}\|x^*_{k+1} - x\|_2^2 - \sum_{j=1}^{m} \left(p_j - \frac{1}{m}\right)\gamma^*_j.$$

Analytical Justification

Corollary

Fix an estimation x_k and denote by x_{k+1} and x^*_{k+1} the next estimations using the RKJL and the standard method, respectively. Set $\gamma^*_j = |\langle a_j, x_k \rangle|^2$ and reorder these so that $\gamma^*_1 \geq \gamma^*_2 \geq \ldots \geq \gamma^*_m$. Then when exact geometry is preserved ($\delta \to 0$),

$$\mathbb{E}\|x_{k+1} - x\|_2^2 \leq \mathbb{E}\|x^*_{k+1} - x\|_2^2 - \sum_{j=1}^{m} \left(p_j - \frac{1}{m}\right)\gamma^*_j.$$

Empirical Evidence

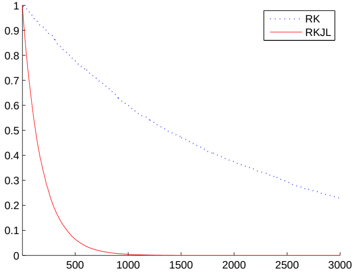


Figure: ℓ_2-Error (y-axis) as a function of the iterations (x-axis). The dashed line is standard Randomized Kaczmarz, and the solid line is the modified one, without a Johnson-Lindenstrauss projection. Instead, the best move out of the randomly chosen n rows is used. Note that we cannot afford to do this computationally.

Empirical Evidence

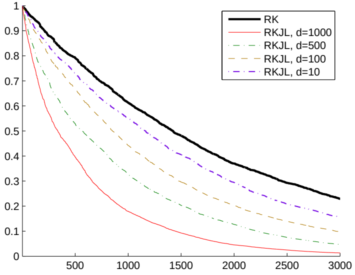


Figure: ℓ_2-Error (y-axis) as a function of the iterations (x-axis) for various values of d with $m = 60000$ and $n = 1000$.

For more information

E-mail:

- dneedell@stanford.edu

Web: http://www-stat.stanford.edu/~dneedell

References:

- Eldar, Needell, "Acceleration of Randomized Kaczmarz Method via the Johnson-Lindenstrauss Lemma", *Num. Algorithms*, to appear.
- Needell, "Randomized Kaczmarz solver for noisy linear systems", *BIT Num. Math.*, **50**(2) 395–403.
- Strohmer, Vershynin, " A randomized Kaczmarz algorithm with exponential convergence", J. Four. Ana. and App. **15** 262–278.