Randomized projection algorithms for overdetermined linear systems

Deanna Needell

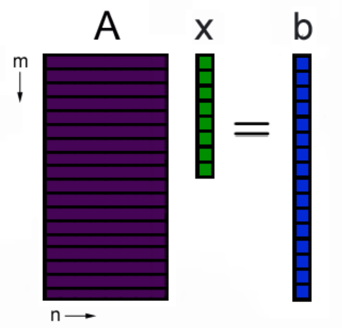
Claremont McKenna College

ISMP, Berlin 2012

Setup

Setup

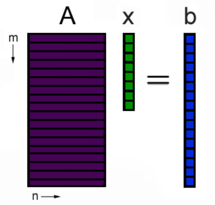
Let $Ax = b$ be an *overdetermined*, standardized, full rank system of equations.



Setup

Setup

Let $Ax = b$ be an *overdetermined*, standardized, full rank system of equations.

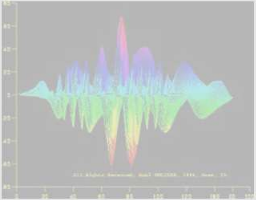


Goal

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

Kaczmarz

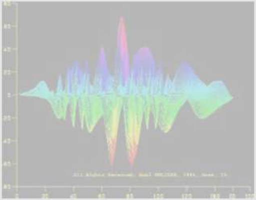
- The Kaczmarz method is an iterative method used to solve $Ax = b$.
- Due to its speed and simplicity, it's used in a variety of applications.



Method

Kaczmarz

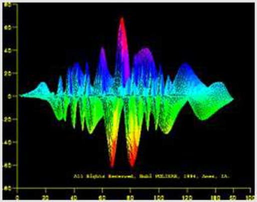
- The Kaczmarz method is an iterative method used to solve $Ax = b$.
- Due to its speed and simplicity, it's used in a variety of applications.



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

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

Kaczmarz

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

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. $x_{k+1} = x_k + (b[i] - \langle a_i, x_k \rangle) a_i$ where $i = (k \bmod m) + 1$
3. Repeat (2)

Method

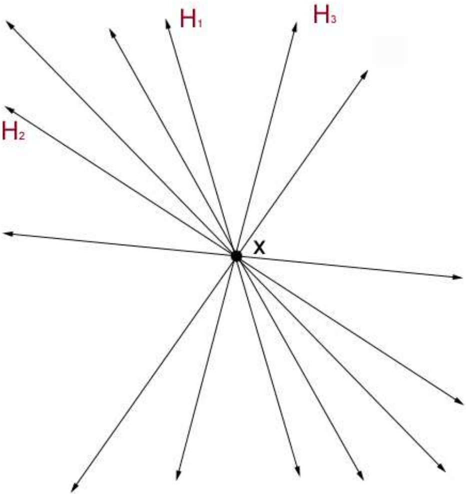
$$\begin{bmatrix} \text{———} \; a_1 \; \text{———} \\ \text{———} \; a_2 \; \text{———} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ \text{———} \; a_m \; \text{———} \end{bmatrix} \cdot \begin{bmatrix} x \\ \\ \\ \end{bmatrix} = \begin{bmatrix} b[1] \\ b[2] \\ \cdot \\ \cdot \\ \cdot \\ b[m] \end{bmatrix}$$

Kaczmarz

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

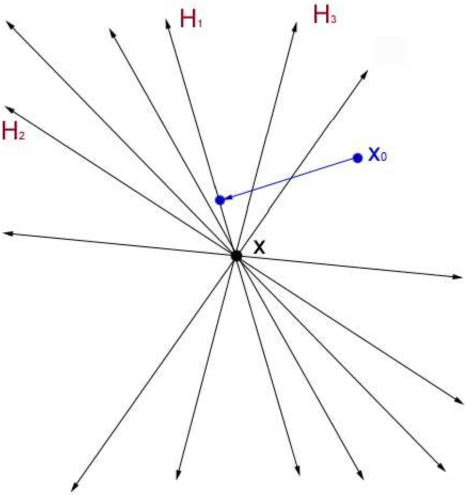
Geometrically

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



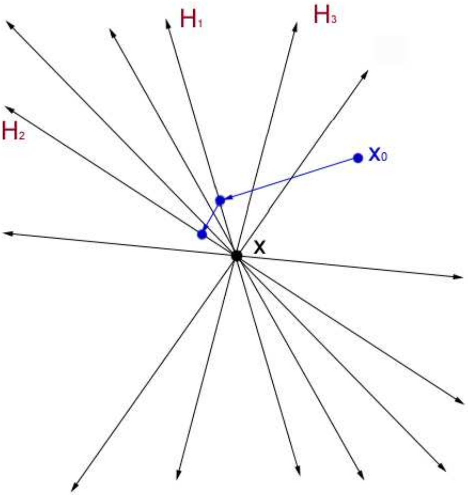
Geometrically

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



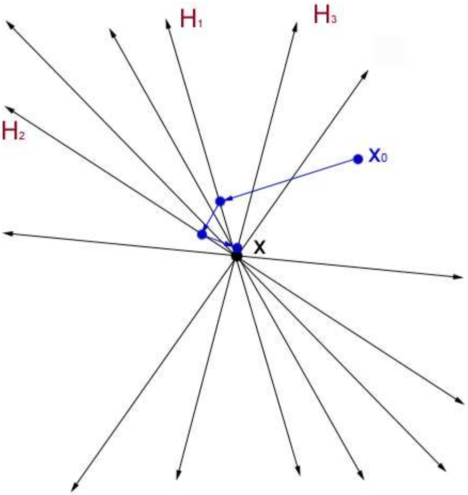
Geometrically

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



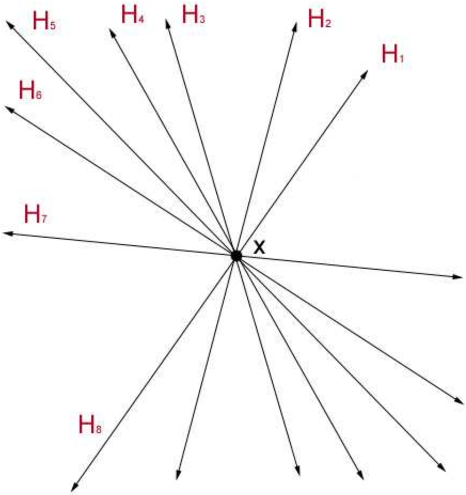
Geometrically

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



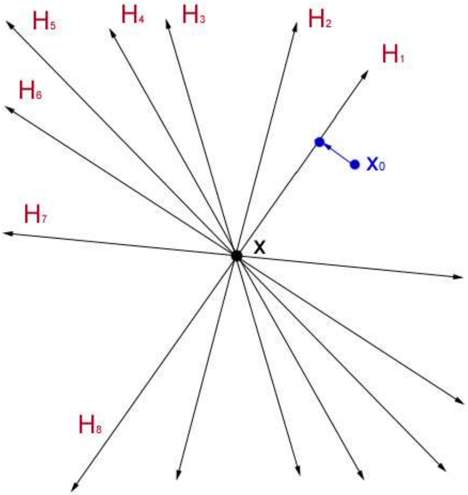
But what if...

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



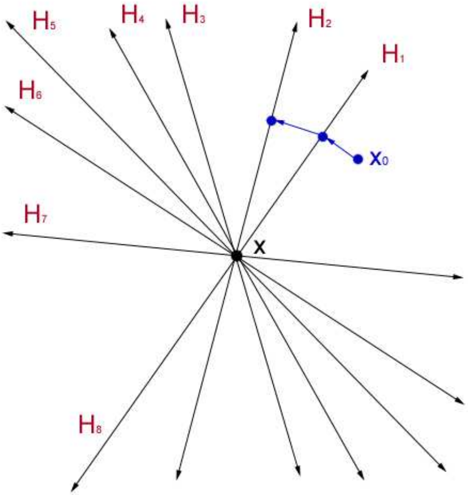
But what if...

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



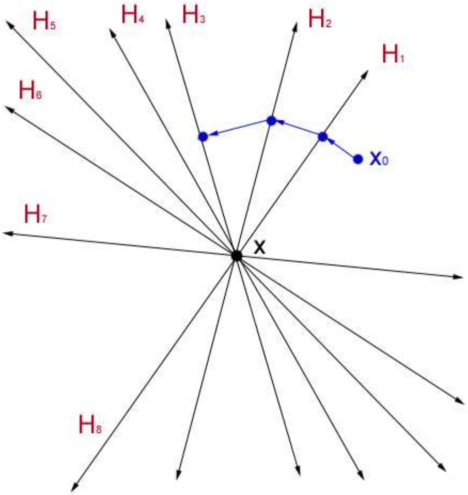
But what if...

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



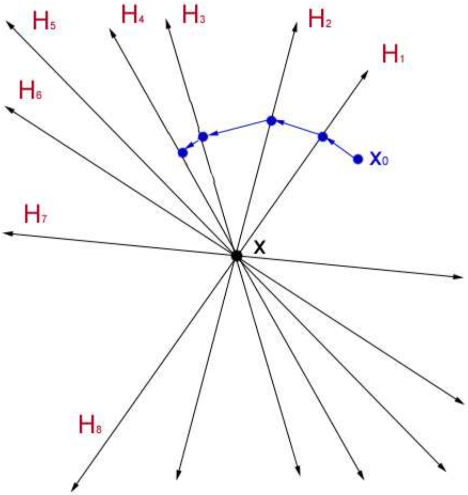
But what if...

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



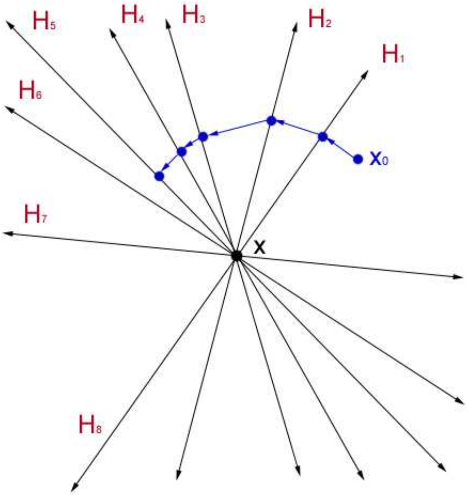
But what if...

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



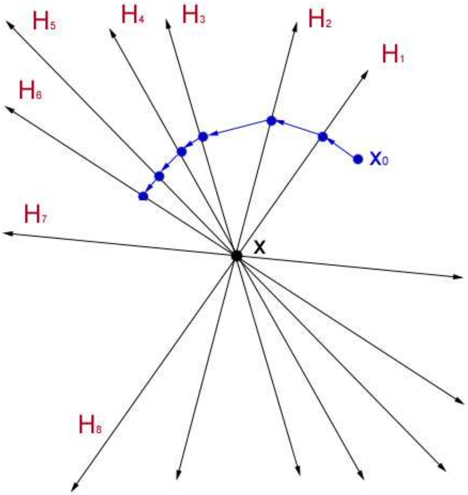
But what if...

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



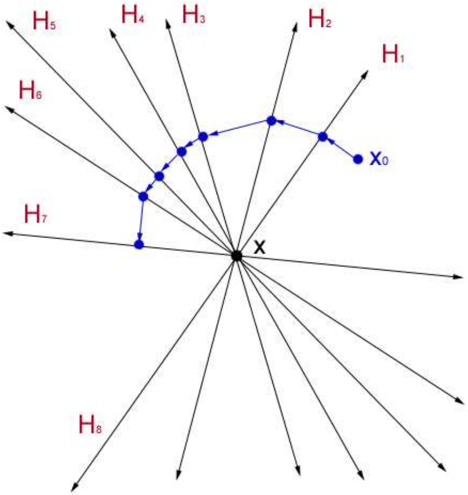
But what if...

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



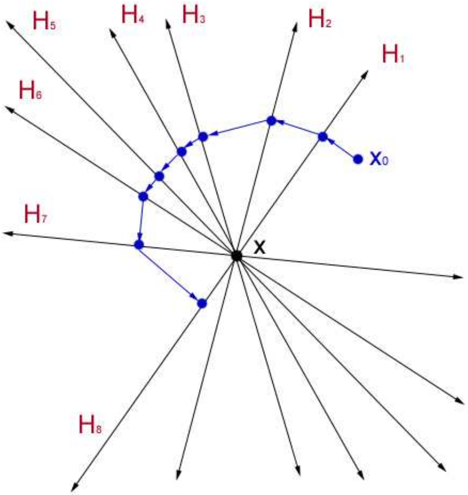
But what if...

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

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

Kaczmarz

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

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. $x_{k+1} = x_k + (b[i] - \langle a_i, x_k \rangle) a_i$ where i is chosen *randomly*
3. Repeat (2)

Randomized Kaczmarz

Theorem [Strohmer-Vershynin]: Consistent case $Ax = b$

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

Randomized Kaczmarz

Theorem [Strohmer-Vershynin]: Consistent case $Ax = b$

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

Randomized Kaczmarz (RK)

Theorem [Strohmer-Vershynin]

- Let $R = m\|A^{-1}\|^2$ ($\|A^{-1}\| \stackrel{\text{def}}{=} \inf\{M : M\|Ax\|_2 \geq \|x\|_2 \text{ for all } x\}$)
- 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)

Theorem [Strohmer-Vershynin]

- Let $R = m\|A^{-1}\|^2$ ($\|A^{-1}\| \stackrel{\text{def}}{=} \inf\{M : M\|Ax\|_2 \geq \|x\|_2 \text{ for all } x\}$)
- 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)

Theorem [Strohmer-Vershynin]

- Let $R = m\|A^{-1}\|^2$ $(\|A^{-1}\| \stackrel{\text{def}}{=} \inf\{M : M\|Ax\|_2 \geq \|x\|_2 \text{ for all } x\})$
- 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)

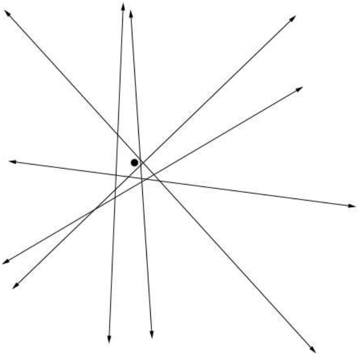
Theorem [Strohmer-Vershynin]

- Let $R = m\|A^{-1}\|^2$ $(\|A^{-1}\| \stackrel{\text{def}}{=} \inf\{M : M\|Ax\|_2 \geq \|x\|_2 \text{ for all } x\})$
- 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) with noise

System with noise

We now consider the system $Ax = b + e$.



Randomized Kaczmarz (RK) with noise

Theorem [N]

- Let $Ax = b + e$. Then

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

.

- This bound is sharp and attained in simple examples.

Randomized Kaczmarz (RK) with noise

Theorem [N]

- Let $Ax = b + e$. Then

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

.

- This bound is sharp and attained in simple examples.

Randomized Kaczmarz (RK) with noise

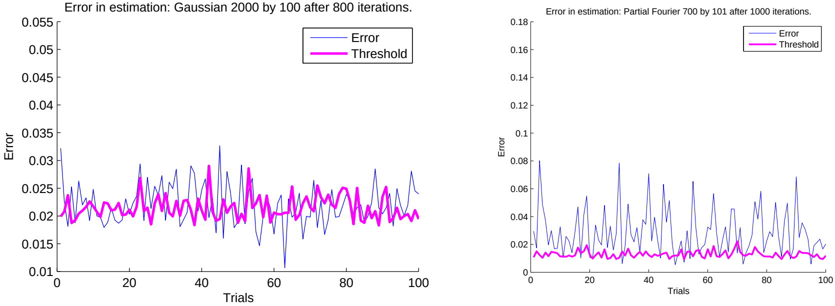


Figure: Comparison between actual error (blue) and predicted threshold (pink). Scatter plot shows exponential convergence over several trials.

Even better convergence?

- Recall $x_{k+1} = x_k + (b[i] - \langle a_i, x_k \rangle)a_i$
- Since these projections are orthogonal, the optimal projection is one that maximizes $\|x_{k+1} - x_k\|_2$.
- What if we relax: $x_{k+1} = x_k + \gamma(b[i] - \langle a_i, x_k \rangle)a_i$
- Can we choose γ optimally?
- Idea: In each "iteration," project once with relaxation optimally and then project normally.

Even better convergence?

- Recall $x_{k+1} = x_k + (b[i] - \langle a_i, x_k \rangle) a_i$
- Since these projections are orthogonal, the optimal projection is one that maximizes $\|x_{k+1} - x_k\|_2$.
- What if we relax: $x_{k+1} = x_k + \gamma(b[i] - \langle a_i, x_k \rangle) a_i$
- Can we choose γ optimally?
- Idea: In each "iteration," project once with relaxation optimally and then project normally.

Even better convergence?

- Recall $x_{k+1} = x_k + (b[i] - \langle a_i, x_k \rangle) a_i$
- Since these projections are orthogonal, the optimal projection is one that maximizes $\|x_{k+1} - x_k\|_2$.
- What if we relax: $x_{k+1} = x_k + \gamma (b[i] - \langle a_i, x_k \rangle) a_i$
- Can we choose γ optimally?
- Idea: In each "iteration," project once with relaxation optimally and then project normally.

Even better convergence?

- Recall $x_{k+1} = x_k + (b[i] - \langle a_i, x_k \rangle) a_i$
- Since these projections are orthogonal, the optimal projection is one that maximizes $\|x_{k+1} - x_k\|_2$.
- What if we relax: $x_{k+1} = x_k + \gamma(b[i] - \langle a_i, x_k \rangle) a_i$
- Can we choose γ optimally?
- Idea: In each "iteration," project once with relaxation optimally and then project normally.

Even better convergence?

- Recall $x_{k+1} = x_k + (b[i] - \langle a_i, x_k \rangle) a_i$
- Since these projections are orthogonal, the optimal projection is one that maximizes $\|x_{k+1} - x_k\|_2$.
- What if we relax: $x_{k+1} = x_k + \gamma (b[i] - \langle a_i, x_k \rangle) a_i$
- Can we choose γ optimally?
- Idea: In each "iteration," project once with relaxation optimally and then project normally.

Even better convergence?

Two-subspace Kaczmarz

- Randomly select two rows, a_s and a_r
- Perform initial projection: $y = x_k + \gamma(b[i] - \langle a_i, x_k \rangle)a_i$ with γ optimal
- Peform second projection: $x_{k+1} = y + (b[i] - \langle a_i, y \rangle)a_i$
- Repeat

Even better convergence?

Two-subspace Kaczmarz

- Randomly select two rows, a_s and a_r
- Perform initial projection: $y = x_k + \gamma(b[i] - \langle a_i, x_k \rangle)a_i$ with γ optimal
- Peform second projection: $x_{k+1} = y + (b[i] - \langle a_i, y \rangle)a_i$
- Repeat

Even better convergence?

Two-subspace Kaczmarz

- Randomly select two rows, a_s and a_r
- Perform initial projection: $y = x_k + \gamma(b[i] - \langle a_i, x_k \rangle)a_i$ with γ optimal
- Peform second projection: $x_{k+1} = y + (b[i] - \langle a_i, y \rangle)a_i$
- Repeat

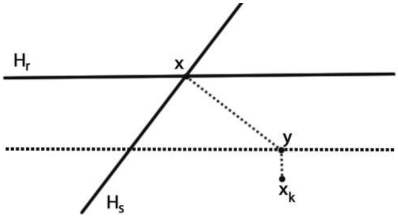
Even better convergence?

Two-subspace Kaczmarz

- Randomly select two rows, a_s and a_r
- Perform initial projection: $y = x_k + \gamma(b[i] - \langle a_i, x_k \rangle)a_i$ with γ optimal
- Peform second projection: $x_{k+1} = y + (b[i] - \langle a_i, y \rangle)a_i$
- Repeat

Two-subspace Kaczmarz

Geometrically, we choose γ in such a way:



Two-subspace Kaczmarz

The optimal choice of γ in a single iteration is

$$\gamma = \frac{-\langle a_r - \langle a_s, a_r\rangle a_s, x_k - x + (b_s - \langle x_k, a_s\rangle)a_s\rangle}{(b_r - \langle x_k, a_r\rangle)\|a_r - \langle a_s, a_r\rangle a_s\|_2^2}.$$

Two-Subspace Kaczmarz method

- Select two distinct rows of A uniformly at random
- $\mu_k \leftarrow \langle a_r, a_s\rangle$
- $y_k \leftarrow x_{k-1} + (b_s - \langle x_{k-1}, a_s\rangle)a_s$
- $v_k \leftarrow \frac{a_r - \mu_k a_s}{\sqrt{1-|\mu_k|^2}}$
- $\beta_k \leftarrow \frac{b_r - b_s\mu_k}{\sqrt{1-|\mu_k|^2}}$
- $x_k \leftarrow y_k + (\beta_k - \langle y_k, v_k\rangle)v_k$

Two-subspace Kaczmarz

The optimal choice of γ in a single iteration is

$$\gamma = \frac{-\langle a_r - \langle a_s, a_r\rangle a_s, x_k - x + (b_s - \langle x_k, a_s\rangle)a_s\rangle}{(b_r - \langle x_k, a_r\rangle)\|a_r - \langle a_s, a_r\rangle a_s\|_2^2}.$$

Two-Subspace Kaczmarz method

- Select two distinct rows of A uniformly at random
- $\mu_k \leftarrow \langle a_r, a_s\rangle$
- $y_k \leftarrow x_{k-1} + (b_s - \langle x_{k-1}, a_s\rangle)a_s$
- $v_k \leftarrow \frac{a_r - \mu_k a_s}{\sqrt{1-|\mu_k|^2}}$
- $\beta_k \leftarrow \frac{b_r - b_s\mu_k}{\sqrt{1-|\mu_k|^2}}$
- $x_k \leftarrow y_k + (\beta_k - \langle y_k, v_k\rangle)v_k$

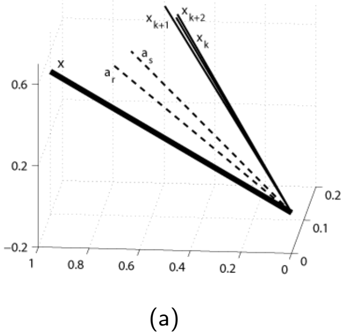


Figure: For coherent systems, the one-subspace randomized Kaczmarz algorithm (a) converges more slowly than the two-subspace Kaczmarz algorithm (b).

Define the coherence parameters:

$$\Delta = \Delta(A) = \max_{j \neq k} |\langle a_j, a_k \rangle| \quad \textit{and} \quad \delta = \delta(A) = \min_{j \neq k} |\langle a_j, a_k \rangle|. \quad (1)$$

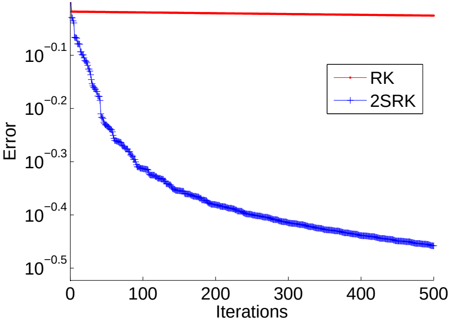


Figure: Randomized Kaczmarz (RK) versus two-subspace RK (2SRK). A has highly coherent rows with $\delta = 0.992$ and $\Delta = 0.998$.

Two-Subspace Kaczmarz

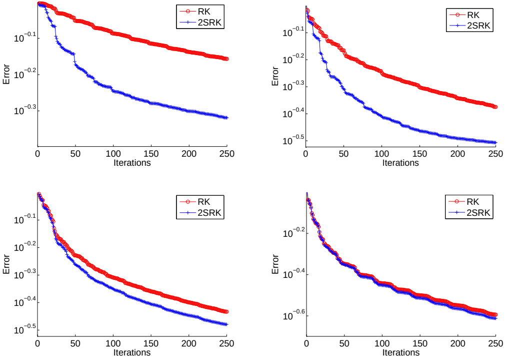


Figure: Randomized Kaczmarz (RK) versus two-subspace RK (2SRK). A has highly coherent rows with coherence parameters (a) $\delta = 0.837$ and $\Delta = 0.967$, (b) $\delta = 0.534$ and $\Delta = 0.904$, (c) $\delta = 0.018$ and $\Delta = 0.819$, and (d) $\delta = 0$ and $\Delta = 0.610$.

Results

Recall the coherence parameters:

$$\Delta = \Delta(A) = \max_{j \neq k} |\langle a_j, a_k \rangle| \quad \textit{and} \quad \delta = \delta(A) = \min_{j \neq k} |\langle a_j, a_k \rangle|. \quad (2)$$

Theorem [N-Ward]

Let $b = Ax + e$, then the two-subspace Kaczmarz method yields

$$\mathbb{E}\|x - x_k\|_2 \leq \eta^{k/2}\|x - x_0\|_2 + \frac{3}{1 - \sqrt{\eta}} \cdot \frac{\|e\|_\infty}{\sqrt{1 - \Delta^2}},$$

where $D = \min\left\{\frac{\delta^2(1-\delta)}{1+\delta}, \frac{\Delta^2(1-\Delta)}{1+\Delta}\right\}$, $R = m\|A^{-1}\|^2$ denotes the scaled condition number, and $\eta = \left(1 - \frac{1}{R}\right)^2 - \frac{D}{R}$.

Results

Recall the coherence parameters:

$$\Delta = \Delta(A) = \max_{j \neq k} |\langle a_j, a_k \rangle| \quad \textit{and} \quad \delta = \delta(A) = \min_{j \neq k} |\langle a_j, a_k \rangle|. \quad (2)$$

Theorem [N-Ward]

Let $b = Ax + e$, then the two-subspace Kaczmarz method yields

$$\mathbb{E}\|x - x_k\|_2 \leq \eta^{k/2}\|x - x_0\|_2 + \frac{3}{1 - \sqrt{\eta}} \cdot \frac{\|e\|_\infty}{\sqrt{1 - \Delta^2}},$$

where $D = \min\left\{\frac{\delta^2(1-\delta)}{1+\delta}, \frac{\Delta^2(1-\Delta)}{1+\Delta}\right\}$, $R = m\|A^{-1}\|^2$ denotes the scaled condition number, and $\eta = \left(1 - \frac{1}{R}\right)^2 - \frac{D}{R}$.

Remarks

1. When $\Delta = 1$ or $\delta = 0$ we recover the same convergence rate as provided for the standard Kaczmarz method since the two-subspace method utilizes two projections per iteration.

2. The bound presented in the theorem is a pessimistic bound. Even when $\Delta = 1$ or $\delta = 0$, the two-subspace method improves on the standard method if any rows of A are highly correlated (but not equal).

Results

Remarks

1. When $\Delta = 1$ or $\delta = 0$ we recover the same convergence rate as provided for the standard Kaczmarz method since the two-subspace method utilizes two projections per iteration.
2. The bound presented in the theorem is a pessimistic bound. Even when $\Delta = 1$ or $\delta = 0$, the two-subspace method improves on the standard method if any rows of A are highly correlated (but not equal).

The parameter D

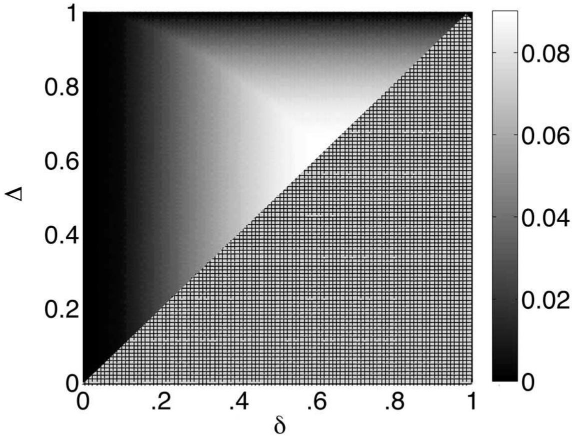


Figure: A plot of the improved convergence factor D as a function of the coherence parameters δ and $\Delta \geq \delta$.

Generalization to more than two rows?

Block Kaczmarz [N-Tropp]

Randomized Block Kaczmarz method

Given a partition of the rows, T:

- Select a block τ of the partition at random
- $x_k \leftarrow x_{k-1} + A_\tau^{\dagger}(b_\tau - A_\tau x_{k-1})$

The convergence rate heavily depends on the conditioning of the blocks A_τ $\rightarrow$ need to control geometric properties of the partition.

Block Kaczmarz [N-Tropp]

Randomized Block Kaczmarz method

Given a partition of the rows, T:

- Select a block τ of the partition at random
- $x_k \leftarrow x_{k-1} + A_\tau{}^\dagger(b_\tau - A_\tau x_{k-1})$

The convergence rate heavily depends on the conditioning of the blocks A_τ $\rightarrow$ need to control geometric properties of the partition.

Block Kaczmarz [N-Tropp]

Randomized Block Kaczmarz method

Given a partition of the rows, T:

- Select a block τ of the partition at random
- $x_k \leftarrow x_{k-1} + {A_\tau}^\dagger(b_\tau - A_\tau x_{k-1})$

The convergence rate heavily depends on the conditioning of the blocks A_τ $\rightarrow$ need to control geometric properties of the partition.

Block Kaczmarz [N-Tropp]

Randomized Block Kaczmarz method

Given a partition of the rows, T:

- Select a block τ of the partition at random
- $x_k \leftarrow x_{k-1} + A_\tau{}^\dagger (b_\tau - A_\tau x_{k-1})$

The convergence rate heavily depends on the conditioning of the blocks A_τ $\rightarrow$ need to control geometric properties of the partition.

Block Kaczmarz

Row paving

An (d, α, β) *row paving* of a matrix A is a partition $T = \{\tau_1, \ldots, \tau_d\}$ of the row indices that verifies

$$\alpha \le \lambda_{\min}(A_\tau A_\tau^*) \quad \text{and} \quad \lambda_{\max}(A_\tau A_\tau^*) \le \beta \quad \text{for each } \tau \in T.$$

Theorem [N-Tropp]

Suppose A admits an (d, α, β) row paving T and that $b = Ax + e$. The convergence of the block Kaczmarz method satisfies

$$\mathbb{E}\|x_k - x\|_2^2 \le \left[1 - \frac{\sigma_{\min}^2(A)}{\beta d}\right]^k \|x_0 - x\|_2^2 + \frac{\beta}{\alpha} \cdot \frac{\|e\|_2^2}{\sigma_{\min}^2(A)}. \quad (3)$$

Block Kaczmarz

Row paving

An (d, α, β) *row paving* of a matrix A is a partition $T = \{\tau_1, \dots, \tau_d\}$ of the row indices that verifies

$$\alpha \le \lambda_{\min}(A_\tau A_\tau^*) \quad \text{and} \quad \lambda_{\max}(A_\tau A_\tau^*) \le \beta \quad \text{for each } \tau \in T.$$

Theorem [N-Tropp]

Suppose A admits an (d, α, β) row paving T and that $b = Ax + e$. The convergence of the block Kaczmarz method satisfies

$$\mathbb{E}\|x_k - x\|_2^2 \le \left[1 - \frac{\sigma_{\min}^2(A)}{\beta d}\right]^k \|x_0 - x\|_2^2 + \frac{\beta}{\alpha} \cdot \frac{\|e\|_2^2}{\sigma_{\min}^2(A)}. \quad (3)$$

Block Kaczmarz

Good row pavings [Bougain-Tzafriri, Tropp]

For any $\delta \in (0,1)$, A admits a row paving with

$$d \le C \cdot \delta^{-2} \|A\|^2 \log(1+n) \quad \text{and} \quad 1-\delta \le \alpha \le \beta \le 1+\delta.$$

Theorem [N-Tropp]

Let A have row paving above with $\delta = 1/2$. The block Kaczmarz method yields

$$\mathbb{E}\|x_k - x\|_2^2 \le \left[1 - \frac{1}{C\kappa^2(A)\log(1+n)}\right]^k \|x_0 - x\|_2^2 + \frac{3\|e\|_2^2}{\sigma_{\min}^2(A)}.$$

Block Kaczmarz

Good row pavings [Bougain-Tzafriri, Tropp]

For any $\delta \in (0,1)$, A admits a row paving with

$$d \le C \cdot \delta^{-2} \|A\|^2 \log(1+n) \quad \text{and} \quad 1-\delta \le \alpha \le \beta \le 1+\delta.$$

Theorem [N-Tropp]

Let A have row paving above with $\delta = 1/2$. The block Kaczmarz method yields

$$\mathbb{E}\|x_k - x\|_2^2 \le \left[1 - \frac{1}{C\kappa^2(A)\log(1+n)}\right]^k \|x_0 - x\|_2^2 \; + \; \frac{3\|e\|_2^2}{\sigma_{\min}^2(A)}.$$

Block Kaczmarz

Theorem [Bougain-Tzafriri, Tropp]

A random partition of the row indices with $m \geq \|A\|^2$ blocks is a row paving with upper bound $\beta \leq 6\log(1+n)$, with probability at least $1 - n^{-1}$.

Theorem [Bourgain-Tzafriri, Tropp]

Suppose that A is incoherent. A random partition of the row indices into m blocks where $m \geq C \cdot \delta^{-2}\|A\|^2 \log(1+n)$ is a row paving of A whose paving bounds satisfy $1 - \delta \leq \alpha \leq \beta \leq 1 + \delta$, with probability at least $1 - n^{-1}$.

Block Kaczmarz

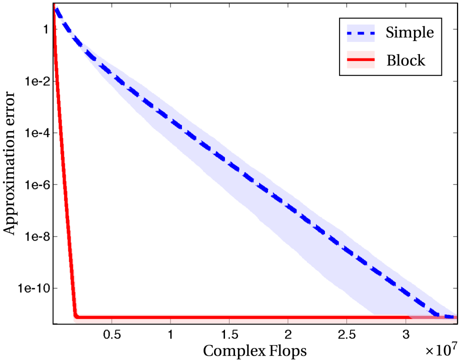


Figure: The matrix A is a fixed 300×100 matrix consisting of 15 partial circulant blocks. Error $\|x_k - x\|_2$ per flop count.

Block Kaczmarz

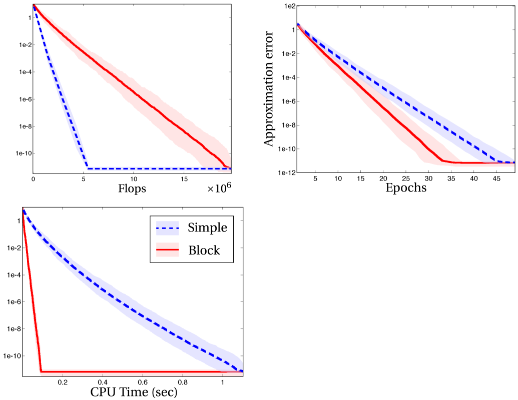


Figure: The matrix A is a fixed 300×100 matrix with rows drawn randomly from the unit sphere, with $d = 10$ blocks. Error $\|x_k - x\|_2$ over various computational resources.

For more information

E-mail:

- dneedell@cmc.edu

Web: www.cmc.edu/pages/faculty/DNeedell

References:

- Strohmer, Vershynin, "A randomized Kaczmarz algorithm with exponential convergence", J. Four. Ana. and App. **15** 262–278.
- Tropp, "Column subset selection, matrix factorization, and eigenvalue optimization", Proc. ACM-SIAM Symposium on Discrete Algorithms, 2009.
- Needell, "Randomized Kaczmarz solver for noisy linear systems", *BIT Num. Math.*, **50**(2) 395–403.
- Needell, Ward, "Two-subspace Projection Method for Coherent Overdetermined Systems", submitted.
- Needell, Tropp , "Paved with Good Intentions: Analysis of a Randomized Block Kaczmarz Method", Submitted.