A Sampling Kaczmarz-Motzkin Algorithm for Linear Feasibility

Jesús De Loera, Jamie Haddock, Deanna Needell

deloera@ucdavis.edu, jhaddock@ucdavis.edu, dneedell@cmc.edu

arxiv.org/abs/1605.01418

SKM Method

Problem: Solve $Ax \leq b$ ($b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$ where $m >> n$) for a point within the convex polyhedral feasible region, P.

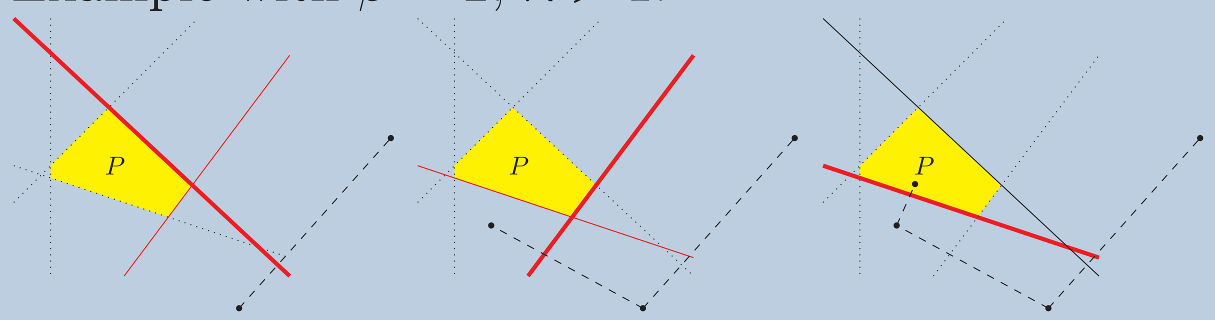
Method. Let $x_0 \in \mathbb{R}^n$ be given. Fix $0 < \lambda \leq 2$. Iteratively construct approximations to a solution lying in P in the following way:

1. If x_k is feasible, stop.
2. Choose a sample of β constraints, τ_k, uniformly at random from among the rows of A.
3. From among these β constraints, choose
$$t_k := \operatorname*{argmax}_{i\in\tau_k} a_i^T x_{k-1} - b_i.$$
4. Define $x_k := x_{k-1} + \lambda \frac{(b_{t_k} - a_{t_k}^T x_{k-1})^+}{\|a_{t_k}\|^2} a_{t_k}$.
5. Repeat.

Orthogonal Projection Method

SKM methods project towards a hyperplane defined by a row of $Ax \leq b$ in every iteration.

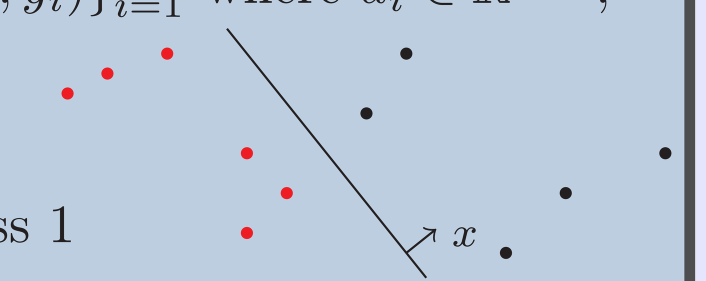
Example with $\beta = 2$, $\lambda > 1$:



SKM with $\beta = m$ recover the deterministic Motzkin relaxation methods [4], while SKM with $\beta = 1$ give variants of the randomized Kaczmarz method [2].

Motivation: SVM Classification

Support Vector Machine: Given binary classified training data, $\{(a_i, y_i)\}_{i=1}^m$ where $a_i \in \mathbb{R}^{n-1}$,



$$y_i = \begin{cases} 1 & \text{if } a_i \in \text{ class } 1 \\ -1 & \text{if } a_i \in \text{ class } 2 \end{cases}$$

find a linear classifier $F(a_i) = x^T a_i + z$ so that
$$y_i F(a_i) \geq 0 \text{ for all } i = 1, ..., m.$$

This leads to a system of linear inequalities $\tilde{A}\tilde{x} \leq 0$ where

$$\tilde{A} = \begin{bmatrix} -y_1 a_1^T & -y_1 \\ & \vdots & \\ -y_m a_m^T & -y_m \end{bmatrix} \in \mathbb{R}^{m\times n} \text{ and } \tilde{x} = \begin{bmatrix} x \\ z \end{bmatrix} \in \mathbb{R}^n.$$

*See center right plot for experimental result on SVM problem.

Theoretical Results

Our first result improves the random Kaczmarz expected rate of [3] and the relaxation method rate of [1].

Theorem 1. Let A be normalized so $\|a_i\|^2 = 1$ for all rows i. If the feasible region P is nonempty then the SKM method with samples of size β converges at least linearly in expectation and the bound on the rate depends on the number of satisfied constraints in the system $Ax \leq b$. More precisely, let s_{k-1} be the number of satisfied constraints after iteration $k-1$ and $V_{k-1} = \max\{m - s_{k-1}, m-\beta+1\}$; then, in the kth iteration,

$$\mathbb{E}[d(x_k, P)^2] \leq \left(1 - \frac{2\lambda - \lambda^2}{V_{k-1}L_2^2}\right) d(x_{k-1}, P)^2 \leq \left(1 - \frac{2\lambda-\lambda^2}{mL_2^2}\right)^k d(x_0, P)^2.$$

Our second result shows one can provide a *certificate of feasibility* after finitely many iterations of SKM. A *certificate of feasibility* is a point x_k with small maximum violation, $\theta(x_k) = \max\{0, \max\{a_i^T x_k - b_i\}\} < 2 * 2^{-\sigma}$ (σ is the binary encoding length of the rational data, A, b). By a lemma of Khachian, if such a point exists, the system $Ax \leq b$ is feasible.

Theorem 2. Suppose A, b are rational matrices with binary encoding length, σ, and that we run an SKM method on the normalized system $\tilde{A}x \leq \tilde{b}$ (where $\tilde{a}_i = \frac{1}{\|a_i\|} a_i$ and $\tilde{b}_i = \frac{1}{\|a_i\|} b_i$) with $x_0 = 0$. If the number of iterations, k is sufficiently large* and the system, $Ax \leq b$ is feasible, the probability that the iterate x_k is not a certificate of feasibility is at most

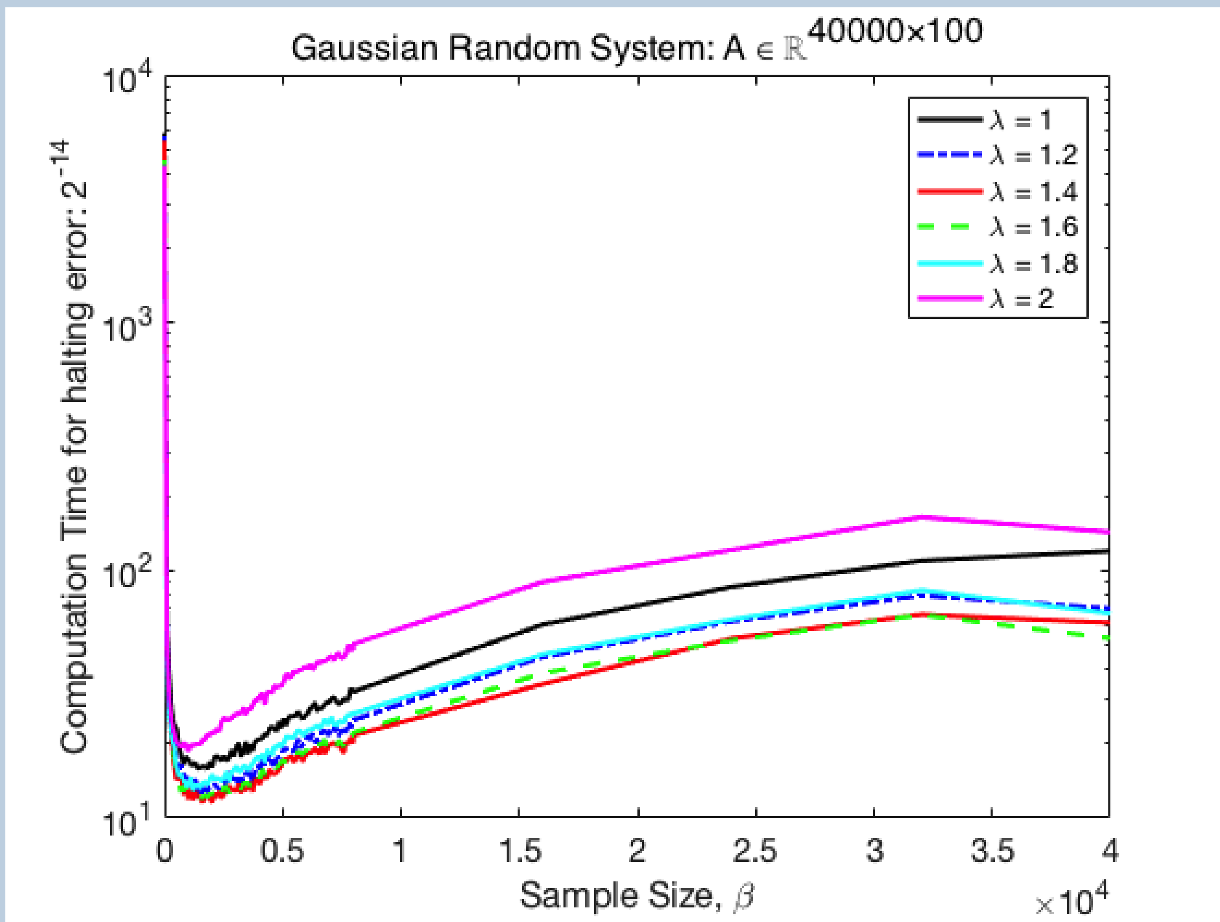
$$\frac{\max \|a_j\| 2^{2\sigma - 2}}{n^{1/2}} \left(1 - \frac{2\lambda - \lambda^2}{mL_2^2}\right)^{k/2},$$

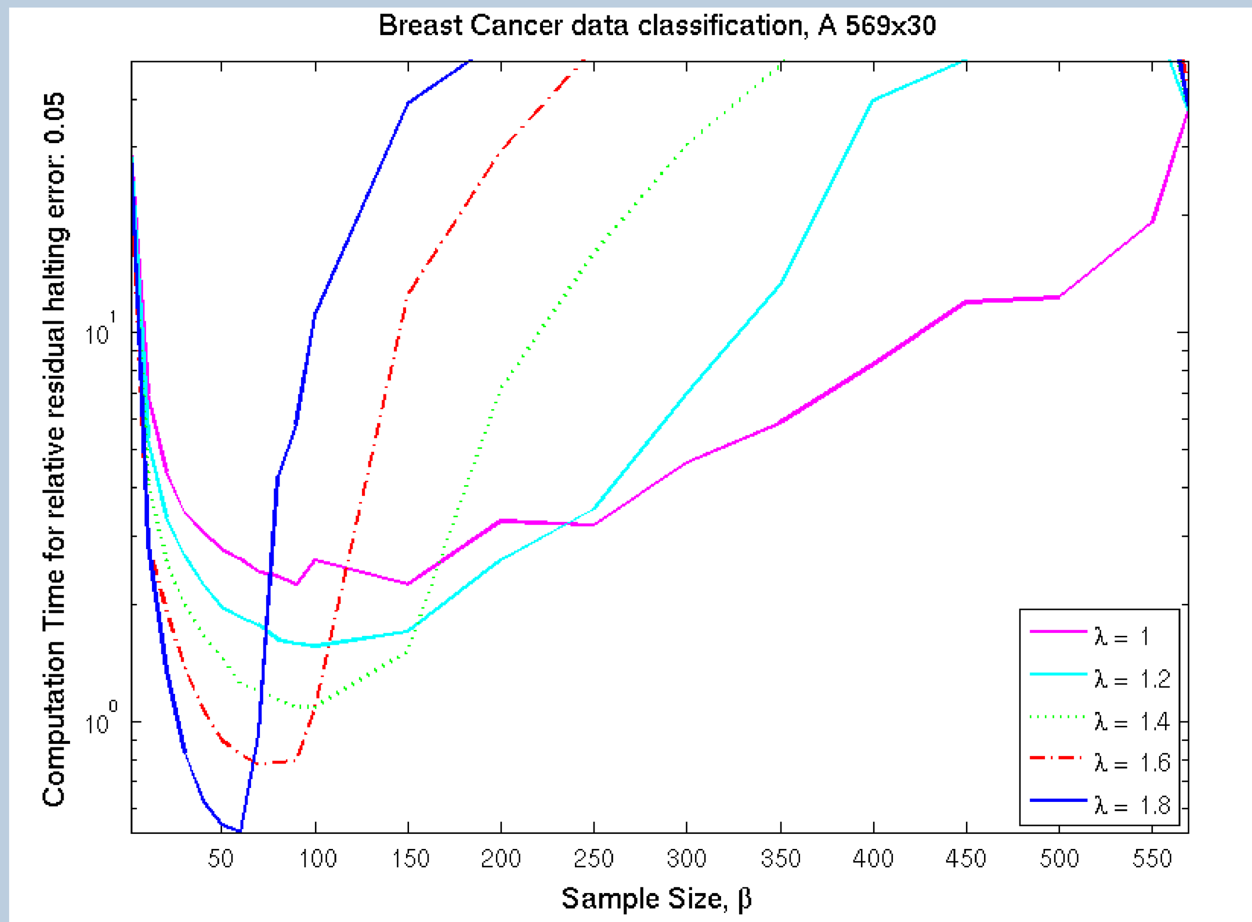
which decreases with k.

*see submitted paper for lower bound on k.

Experimental Results

We provide evidence that there is an optimal choice for the sample size, β. Regardless of choice of projection parameter, λ, we see a minimum occur for β between 1 and m.





Left: Mean computational time for SKM on $Ax \leq b$ (where $A \in \mathbb{R}^{40000\times 100}$ has Gaussian random variable entries and b chosen for non-empty interior) to reach halting residual error 2^{-14}. Right: Mean computational time required for SKM on $Ax \leq b$ (where $A \in \mathbb{R}^{569\times 30}$ and $b \in \mathbb{R}^{569}$ defines a support vector machine problem modeling the classification of breast cancer data as benign or cancerous) to reach halting residual error.

Additional Info

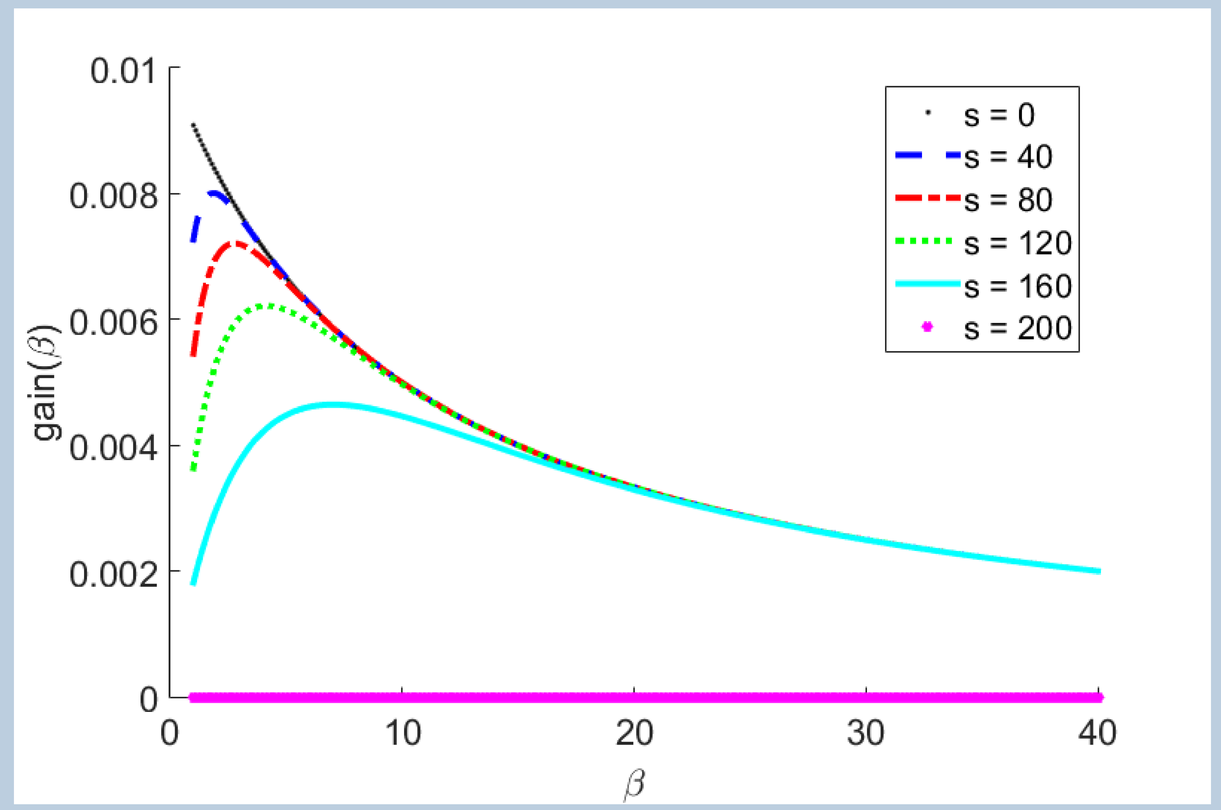
For more information scan the QR code at top right or contact the presenter at jhaddock@math.ucdavis.edu.

Choice of β: Iteration vs. Time

We seek β that maximizes the ratio of improvement made and computation cost in a fixed iteration:

$$\mathrm{gain}(\beta) := \frac{\mathbb{E}\|(A_{\tau_j}x_j - b_{\tau_j})^+\|_\infty^2}{C + cn\beta} \approx \frac{1 - (\frac{s}{m})^\beta}{C + cn\beta}.$$

Worst case improvement occurs for constant $m - s$ nonzero entries of the residual (s is number of satisfied constraints). We plot the worst case approximation as a function of β for various s.



References

[1] S. Agmon. The relaxation method for linear inequalities. *Canadian J. Math.*, 6:382–392, 1954.

[2] S. Kaczmarz. Angenäherte auflösung von systemen linearer gleichungen. *Bull.Internat.Acad.Polon.Sci.Lettres A*, pages 335–357, 1937.

[3] D. Leventhal and A. S. Lewis. Randomized methods for linear constraints: convergence rates and conditioning. *Math.Oper.Res.*, 35(3):641–654, 2010. 65F10 (15A39 65K05 90C25); 2724068 (2012a:65083); Raimundo J. B. de Sampaio.

[4] T. S. Motzkin and I. J. Schoenberg. The relaxation method for linear inequalities. *Canadian J. Math.*, 6:393–404, 1954.

[5] T. Strohmer and R. Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *J. Fourier Anal. Appl.*, 15:262–278, 2009.

Acknowledgements

The first and second authors were partially supported by the NSA, the second author was partially supported by a GAANN fellowship, and the third author was partially supported by the NSF and the Alfred P. Sloan foundation. All support is gratefully acknowledged.