A Sampling Kaczmarz-Motzkin Algorithm for Linear Feasibility

Jesús A. De Loera, Jamie Haddock, Deanna Needell

Abstract

We combine two algorithmic techniques for solving systems of linear inequalities, the relaxation method of Agmon, Motzkin et al. and the randomized Kaczmarz method. In doing so, we obtain a family of algorithms that generalize and extend both techniques. While we prove similar convergence results, our computational experiments show our algorithms often vastly outperform the original methods.

1 Introduction

We are interested in determining the feasibility of systems of linear inequalities $Ax \leq b$ for $b \in \mathbb{R}^m$ and A an $m \times n$ matrix. We will denote the rows of A by the vectors $a_1, a_2, \ldots, a_m$. The set of all $x \in \mathbb{R}^n$ that satisfy the above constraints is a convex polyhedral region, which we denote by P. This paper merges two research trends that seem, for the most part, to have not met each other. The combination of the two algorithmic branches of thought results in an interesting new family of algorithms which outperforms its predecessors.

Motzkin's method. The first branch of research in linear feasibility is the so-called *relaxation method* or *Motzkin's method*. The original relaxation or projection algorithm by Agmon [Agm54], and Motzkin and Schoenberg [MS54] can be described as follows: Starting from any initial point x_0, a sequence of points is generated. If the current point x_i is feasible we stop, else there must be a constraint $a^T x \leq b$ that is most violated. The constraint defines a hyperplane H. If w_H is the orthogonal projection of x_i onto the hyperplane H, choose a number λ (normally chosen between 0 and 2), and the new point x_{i+1} is given by $x_{i+1} = x_i + \lambda(w_H - x_i)$. Figure 1 displays the iteration visually.

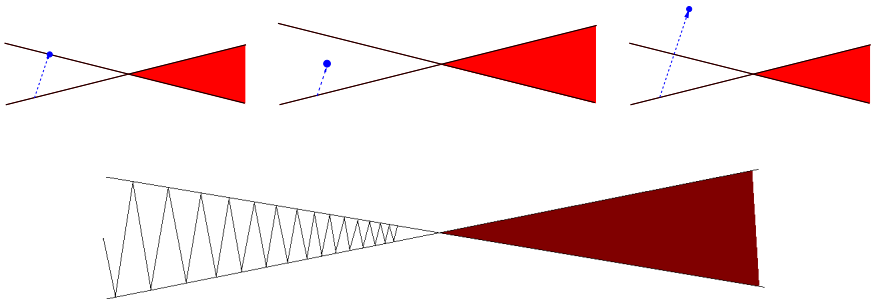

Figure 1: three projections with $\lambda = 1$, $\lambda < 1$ and $\lambda > 1$ and a visualization of several steps of the algorithm.

Many modifications and analyses of this technique have been published since the 1950s, creating an extensive bibliography. For example, versions of the relaxation method have suggested various choices of step-length multiplier, λ, and various choices of violated hyperplane. Not only does the rate of convergence

depend on λ, but on the *Hoffman constants* investigated first by Agmon [Agm54] and then later Hoffmann [Hof52]. If the system of inequalities, $Ax \leq b$ is feasible, i.e. $P \neq \emptyset$, then there exists Hoffman constants L_∞ and L_2 so that $d(x, P) \leq L_\infty \|(Ax - b)^+\|_\infty$ and $d(x, P) \leq L_2 \|(Ax - b)^+\|_2$ for all x (here and throughout, z^+ denotes the positive entries of the vector z with zeros elsewhere). The constants satisfy $L_\infty \leq \sqrt{m} L_2$. With these constants, one can prove convergence rate results like the following (a spin-off of Theorem 3 of [Agm54] which is easily proven in the style of [LL10]):

Proposition 1. Consider a normalized system with $\|a_i\| = 1$ for all $i = 1, ..., m$. If the feasible region P is nonempty then the relaxation method converges linearly:

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

A bad feature of the standard version of the relaxation method using real-valued data is that when the system $Ax \leq b$ is infeasible it cannot terminate, as there will always be a violated inequality. In the 1980's the relaxation method was revisited with interest because of its similarities to the ellipsoid method (see [AH05, Bet04, Gof80, Tel82] and references therein). One can show that the relaxation method is finite in all cases when using rational data, in that it can detect infeasible systems. In some special cases the method gives a polynomial time algorithm (e.g. for totally unimodular matrices [MTA81]), but there are also examples of exponential running times (see [Gof82, Tel82]). In late 2010, Chubanov [Chu11], announced a modification of the traditional relaxation style method, which gives a *strongly polynomial*-time algorithm in some situations [VZ14]. Unlike [Agm54, MS54], who only projected onto the original hyperplanes that describe the polyhedron P, Chubanov [Chu11] projects onto new, auxiliary inequalities which are linear combinations of the input; see Figure 2.

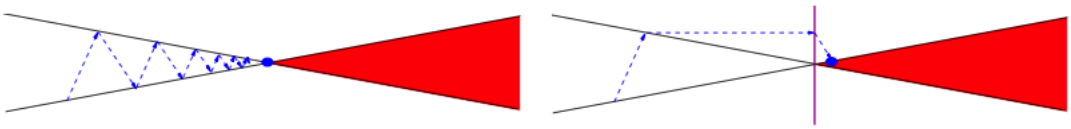

Figure 2: Left: Projecting onto original hyperplanes. Right: Projecting onto an induced hyperplane (like those in Chubanov's method).

Kaczmarz method. The second research branch is that of the Kaczmarz method [Kac37, GBH70] which is one of the most popular solvers of overdetermined systems of linear equations due to its speed and simplicity. Just like Motzkin's, it is an iterative method which consists of a series of alternating projections. The original Kaczmarz method simply cycles through the equations sequentially, so its convergence rate depends on the order of the rows. One way to overcome this is to use the equations in a *random order*, rather than sequentially [HS78, HM93, Nat01]. More precisely, we begin with $Ax \leq b$, a linear system of inequalities where A is an $m \times n$ matrix with rows a_i and x_0 an initial guess. For $k = 0, 1, 2, ...$ one defines

$$x_{k+1} = x_k - \frac{(\langle a_i, x_k \rangle - b_i)^+}{\|a_i\|_2^2} a_i$$

where i is chosen from $\{1, 2, ..., m\}$ at random with probability proportional to $\|a_i\|_2^2$. Strohmer and Vershynin [SV09] provided an elegant convergence analysis of the randomized Kaczmarz method for consistent equations. Later, Leventhal and Lewis [LL10] extended the probabilistic analysis from systems of equations

to systems of linear inequalities. They focused on giving bounds on the convergence rate that take into account the numerical conditions captured by the constants L_∞ and L_2. If one additionally makes use of a projection parameter, $\lambda \neq 1$, you can easily extend the convergence rate in [LL10] to account for this:

Proposition 2. If the feasible region, P, is nonempty then the Randomized Kaczmarz method with projection parameter λ converges linearly in expectation:

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

Note the similarities between Propositions 1 and 2: the convergence rate constants are identical for normalized systems ($\|A\|_F^2 = m$).

The work of Strohmer and Vershynin sparked a new interest in the Kaczmarz approach and there have been many recent developments in the method and its analysis. Needell [Nee10] extended this work to the case of inconsistent systems of equations, showing exponential convergence down to some fixed *convergence horizon*, see also [WAL15]. In order to break this convergence horizon, one needs to modify the Kaczmarz method since by design it projects exactly onto a given hyperplane. Built upon ideas by Popa [Pop98], Zouzias and Freris [ZF12] analyzed an extended randomized Kaczmarz method which incorporates an additional projection step to reduce the size of the residual. This was extended to the block case in [NZZ15]. The relation of these approaches to coordinate descent and gradient descent methods has also been recently studied, see e.g. [GO12, Dum14, NSW14a, OZ15a, MNR15, HNR15, OZ15a, GR15].

Other variations to the Kaczmarz method include block methods [Elf80, EHL81, NW13, NT13, BN, XZ02] which have been shown to offer acceleration for certain systems of equations with fast-multipliers. Other acceleration and convergence schemes focus on sampling selections [AWL14,EN11,NSW14b,OZ15b], projection parameters [WM67,CEG83,Tan71,HN90], adding row directions [PPKR12], parallelized implementations [LWS14, ADG14], structure exploiting approaches [LW15, LMY15], and the use of preconditioning [GPS16]. Some other references on recent work include [CP12, RM12]

For the most part, it seems that these two branches of research which address the same problems have been developing disjointly from each other. For example, the idea of taking linear combinations of the constraints was first exploited in [Chu11], but was recently reproduced independently for equations in [GR15]. Another example is the manipulation of the projection parameter λ [WM67, CEG83, Tan71, HN90]. It is a goal of this paper to bridge the separation between these two branches of research that essentially study the same iterative projection procedure. In this paper we explore a family of hybrid algorithms that use elements from both groups of research.

1.1 Our Contribution: the Sampling Kaczmarz-Motzkin Method

Our proposed family of methods, which we refer to as the *Sampling Kaczmarz-Motzkin* (SKM) methods, are intended to balance the pros and cons of the previously described methods. Namely, the relaxation method forms iterates whose distance to the polyhedral solution space are monotonically decreasing; however, the time required to choose the most violated hyperplane in each iteration is costly. Conversely, the Randomized Kaczmarz method has a very inexpensive cost per iteration; however, the method has slow convergence when many of the constraints are satisfied. Our methods will still have a probabilistic choice, like in randomized Kaczmarz, but make strong use of the maximum violation criterion within this random sample of the constraints. Our method is easily seen to interpolate between what was proposed in [LL10] and in [MS54].

Method (SKM method)**.** Suppose $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. Let $x_0 \in \mathbb{R}^n$ be given. Fix $0 < \lambda \leq 2$. We iteratively construct approximations to a solution lying in P in the following way:

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

In practice, one would want to occasionally check for feasibility of the current iterate. When we use *termination* throughout the paper, we refer to the iterates becoming feasible and the sequence becoming constant. Note that the SKM method with $\beta = m$ recovers the Motzkin relaxation methods, while the SKM method with $\beta = 1$ gives a variant of the randomized Kaczmarz method. We now state our main result.

Theorem 1. Let A be normalized so $\|a_i\|^2 = 1$ for all rows i.

(a) 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 hyperplane arrangement. More precisely, let s_{k-1} be the number of satisfied constraints after iteration $k-1$ and $S_{k-1} = \max\{m - s_{k-1}, m - \beta + 1\}$; then, in the kth iteration,

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

(b) If the feasible region P is generic and nonempty (i.e., full-dimensional and every vertex satisfies exactly n constraints with equality), then an SKM method with samples of size $\beta \leq m - n$ will converge to a single face F of P and all but the constraints defining F will eventually be satisfied. Thus, the method is guaranteed an increased convergence rate after some index K (provided the method does not terminate earlier); for $k \geq K$

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

(c) If P is nonempty and A, b are rational, then the SKM methods can detect feasibility. Moreover, the expected number of steps required for the SKM methods with projection parameter $0 < \lambda < 2$ and $x_0 = 0$ to detect feasibility is no more than

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

2 Experiments

We implemented the SKM methods on a 32GB RAM 8-node cluster, each with 12 cores of Intel Xeon E5-2640 v2 CPUs running at 2 GHz, and ran them on systems while varying the projection parameter, λ, and the sample size, β. In particular, we considered systems $Ax \leq b$ where A has entries consisting of Gaussian random variables and b is chosen to force the system to have a solution set with non-empty interior. We additionally considered systems where the rows of A are highly correlated (each row consists only of entries chosen uniformly at random from $[.9, 1]$ or only of entries chosen uniformly at random from $[-1, -.9]$) and b is chosen to force the system to have a solution set with non-empty interior. We vary the size of $A \in \mathbb{R}^{m\times n}$, which we note in each example presented below.

In Figure 3, we provide evidence that there is an optimal choice for the sample size, β. We measure the average computational time necessary for SKM with several choices of sample size β to reach halting residual error 2^{-14}. Regardless of choice of projection parameter, λ, we see a minimum occur between 1 and m.

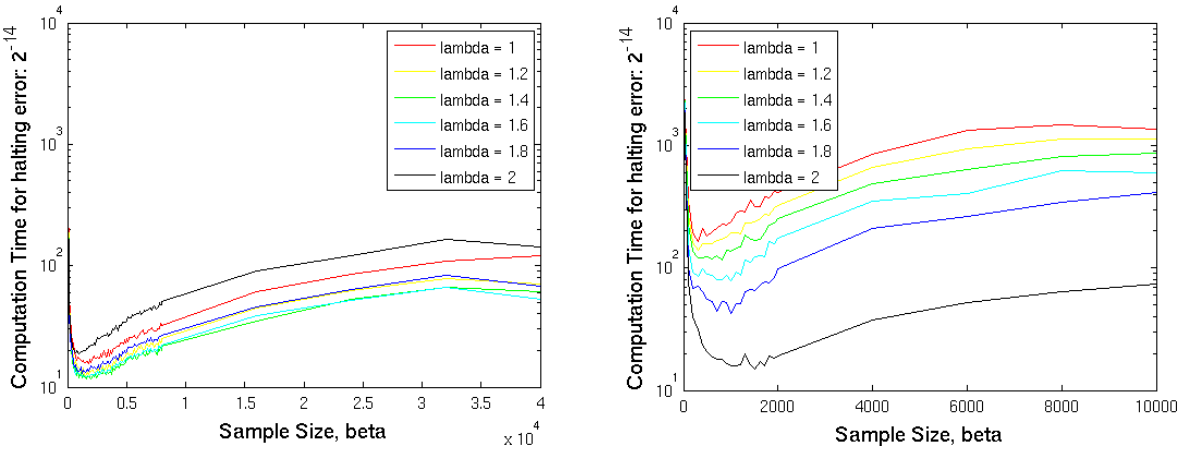

Figure 3: Left: Average computational time required for SKM on Gaussian random system of size $n = 100, m = 40000$ to reach halting residual error 2^{-14}. Right: Average computational time required for SKM on correlated random system of size $n = 100, m = 10000$ to reach halting residual error.

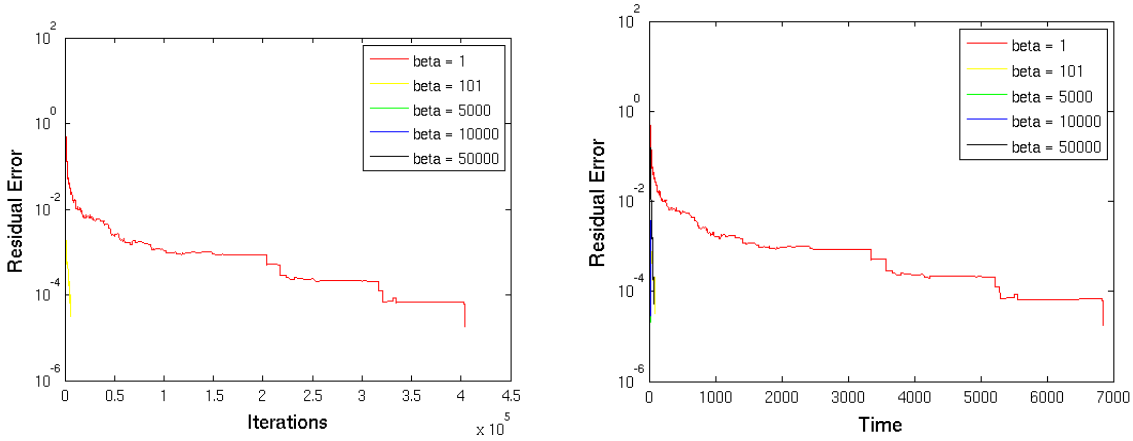

Figure 4: Left: Iterations vs. residual error for SKM on Gaussian random system of size $n = 100, m = 50000$ for sample sizes varying from 1 to m. Right: Time vs. residual error for SKM on Gaussian random system of size $n = 100, m = 50000$ for sample sizes varying from 1 to m.

For the experiments in Figures 4, 5, and 6, we fixed the projection parameter at $\lambda = 1.6$ (for reasons discussed below). Note that in Figure 4, SKM with $\beta \neq 1$ so outperforms SKM with $\beta = 1$ (randomized Kaczmarz method) that it is impossible to distinguish the performances of SKM with $\beta \approx m$ or $\beta \approx n$. Only in Figure 5, when we restricted ourselves to considering $\beta > 1$ were we able to distinguish between

the performances of SKM with these other sample sizes. On the left of Figure 5, we see the residual error decreases more quickly per iteration as the sample size, β increases. However, on the right, when measuring the computational time, SKM with $\beta \approx m/10$ performs best.

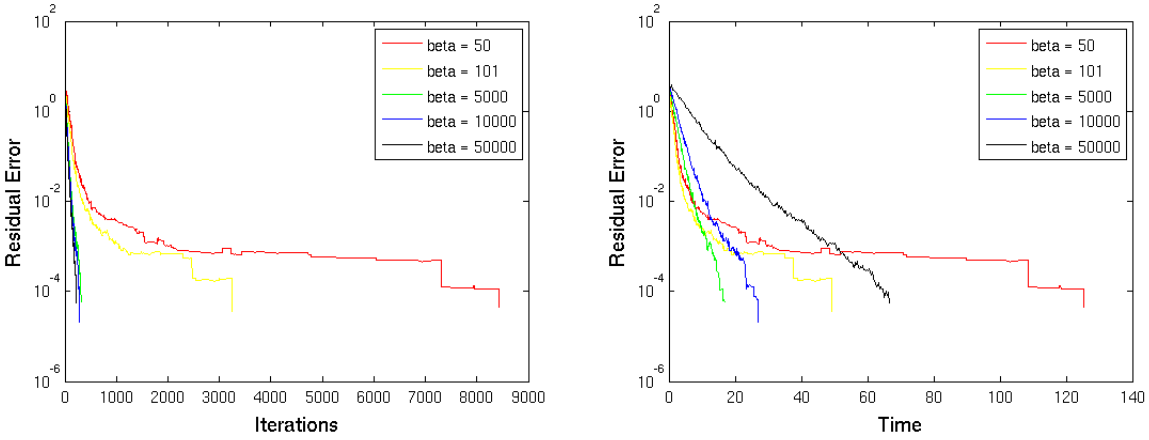

Figure 5: Left: Iterations vs. residual error for SKM on Gaussian random system of size $n = 100, m = 50000$ for sample sizes varying from 50 to m. Right: Time vs. residual error for SKM on Gaussian random system of size $n = 100, m = 50000$ for sample sizes varying from 50 to m.

In Figure 6, we ran experiments varying the halting error and see that the sample size selection, β depends additionally on the desired final distance to the feasible region, P. On the right, we attempted to pinpoint the optimal choice of β by reducing the sample sizes we were considering.

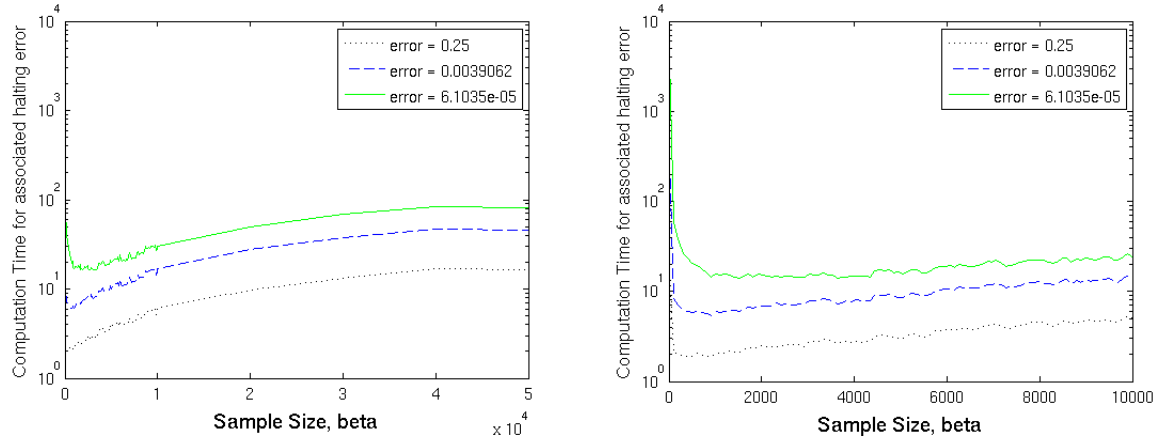

Figure 6: Left: Average computational time required for SKM on Gaussian random system of size $n = 100, m = 50000$ to reach halting residual errors for β between 1 and m. Right: Average computational time required for SKM on Gaussian random system of size $n = 100, m = 50000$ to reach halting residual errors for β between 1 and $m/5$.

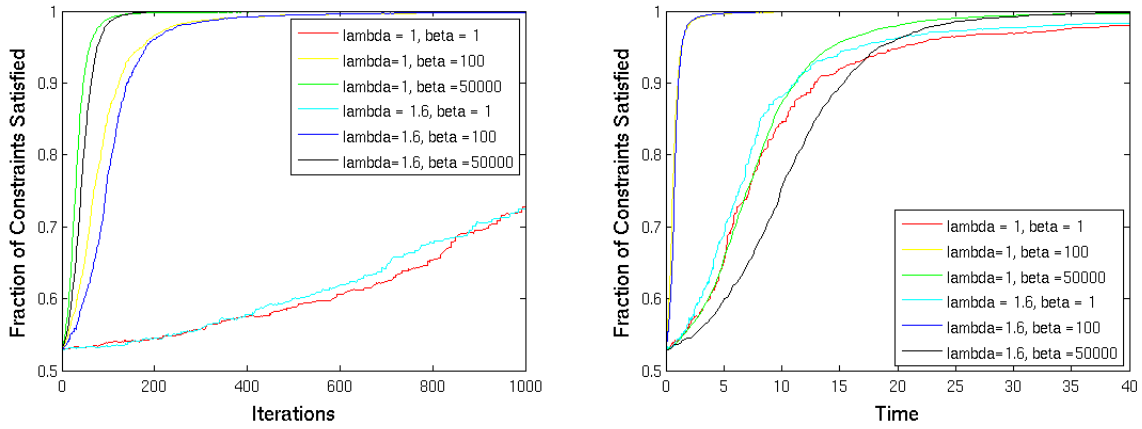

Figure 7: Left: Iterations vs. fraction of constraints satisfied for SKM methods run on Gaussian random system of size $n = 100, m = 50000$. Right: Time vs. fraction of constraints satisfied for SKM methods run on Gaussian random system of size $n = 100, m = 50000$.

Like [SV09], we observe that 'overshooting' ($\lambda > 1$) outperforms other projection parameters, $\lambda \leq 1$. In Figure 3, we see that the optimal projection parameter, λ is system dependent. For the experiments in Figure 3, we ran SKM on the same system until the iterates had residual error less than 2^{-14} and averaged the computational time taken over ten runs. The best choice of λ differed greatly between the Gaussian random systems and the correlated random systems; for Gaussian systems it was $1.4 < \lambda < 1.6$ while for correlated systems it was $\lambda = 2$.

Our bound on the distance remaining to the feasible region decreases as the number of satisfied constraints increases. In Figure 7, we see that the fraction of satisfied constraints initially increased most quickly for SKM with sample size, $1 < \beta < m$ and projection parameter, $\lambda > 1$. On the left, we show that SKM with $\beta = m$ is faster in terms of number of iterations. However, on the right, we show that SKM with $1 < \beta < m$ outperforms $\beta = m$ in terms of time because of its computational cost in each iteration.

2.1 Remarks and Future Work

While Theorem 1 suggests that the optimal choice of β could be between 1 and $m - n$, our experiments suggest the optimal choice for β could be stated exactly for certain types of systems of inequalities. One can show the optimal choice of β *in a fixed iteration* under certain assumptions on the residual and the number of satisfied constraints is a value bounded away from both 1 and $m-n$ (this follows from the proof of Theorem 1 and standard runtime analysis arguments). Exactly describing this optimal choice of β for a given system or class of systems in the general setting would be a significant contribution to this field. We believe such a result can be built upon the theory and framework we present here, but requires further analysis.

Additionally, the optimal choice of projection parameter λ is system dependent (e.g., for certain systems, one should choose $\lambda = 1$ while for certain full-dimensional systems, one should choose $\lambda > 1$). Theoretically, the convergence rate we provided in Theorem 1 Part (a) depends upon λ in a weak way; one would always choose $\lambda = 1$. However, we see experimentally that overshooting outperforms other choices of λ. Additionally, one can easily imagine that for systems whose polyhedral feasible region is full-dimensional, choosing $\lambda > 1$ will outperform $\lambda \leq 1$, as eventually, the iterates could 'hop' into the the feasible region.

The proof of Theorem 1 Part (b) suggests a possible reason why we see this in our experiments. Theorem 1 Part (b) is a consequence of the fact that if the method does not terminate then it will converge to a unique face of P. If $\lambda > 1$, then this face cannot be a facet of P, as if the method converged to such a face, it would eventually terminate, 'hopping' over the facet into P. Thus, for $\lambda > 1$ the number of faces of P the method can converge to is decreased.

Further work is needed before stating that overshooting outperforms other choices of projection parameter, λ for any particular class of systems or for defining the optimal sample size, β for a class of systems. In our future work we intend to present the optimal values of β and λ and to explore the dual versions of our family of algorithms. Moreover we will explore Chubanov's style of generation of additional linear inequalities that have proven to speed computation.

3 Acknowledgements

The first and second author were partially supported by a grant from the NSA. The second author was partially supported by a GAANN fellowship. The third author was supported by the NSF and the Alfred P. Sloan foundation.

References

[ADG14] H. Avron, A. Druinsky, and A. Gupta. Revisiting asynchronous linear solvers: Provable convergence rate through randomization. In *Parallel and Distributed Processing Symposium, 2014 IEEE 28th International*, pages 198–207. IEEE, 2014.

[Agm54] S. Agmon. The relaxation method for linear inequalities. *Canadian J. Math.*, 6:382–392, 1954.

[AH05] E. Amaldi and R. Hauser. Boundedness theorems for the relaxation method. *Math. Oper. Res.*, 30(4):939–955, 2005.

[AWL14] A. Agaskar, C. Wang, and Y. M. Lu. Randomized Kaczmarz algorithms: Exact mse analysis and optimal sampling probabilities. In *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*, pages 389–393. IEEE, 2014.

[Bet04] U. Betke. Relaxation, new combinatorial and polynomial algorithms for the linear feasibility problem. *Discrete Comput. Geom.*, 32(3):317–338, 2004.

[BN] J. Briskman and D. Needell. Block Kaczmarz method with inequalities. *Journal of Mathematical Imaging and Vision*, 52(3):385–396.

[CEG83] Y. Censor, P. P. Eggermont, and D. Gordon. Strong underrelaxation in Kaczmarz's method for inconsistent systems. *Numerische Mathematik*, 41(1):83–92, 1983.

[Chu11] S. Chubanov. A strongly polynomial algorithm for linear systems having a binary solution. *Mathematical Programming (to appear)*, pages 1–38, 2011. 10.1007/s10107-011-0445-3.

[CP12] X. Chen and A. Powell. Almost sure convergence of the Kaczmarz algorithm with random measurements. *J.Fourier Anal.Appl.*, pages 1–20, 2012. 10.1007/s00041-012-9237-2.

[Dum14] B. Dumitrescu. On the relation between the randomized extended Kaczmarz algorithm and coordinate descent. *BIT Numerical Mathematics*, pages 1–11, 2014.

[EHL81] P. P. B. Eggermont, G. T. Herman, and A. Lent. Iterative algorithms for large partitioned linear systems, with applications to image reconstruction. *Linear Algebra Appl.*, 40:37–67, 1981. LAAPAW; 65F10 (68G10); 629606 (84c:65059).

[Elf80] T. Elfving. Block-iterative methods for consistent and inconsistent linear equations. *Numer.Math.*, 35(1):1–12, 1980. NUMMA7; 65F10; 583651 (83e:65059).

[EN11] Y. C. Eldar and D. Needell. Acceleration of randomized Kaczmarz method via the Johnson-Lindenstrauss lemma. *Numer.Algorithms*, 58(2):163–177, 2011. 65F20 (65F10); 2835851; Alexander N. Malyshev.

[GBH70] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography. *J.Theoret.Biol.*, 29:471–481, 1970.

[GO12] M. Griebel and P. Oswald. Greedy and randomized versions of the multiplicative Schwarz method. *Linear Algebra and its Applications*, 437(7):1596–1610, 2012.

[Gof80] J.-L. Goffin. The relaxation method for solving systems of linear inequalities. *Math. Oper. Res.*, 5(3):388–414, 1980.

[Gof82] J.-L. Goffin. On the nonpolynomiality of the relaxation method for systems of linear inequalities. *Math. Programming*, 22(1):93–103, 1982.

[GPS16] E. Gallopoulos, B. Philippe, and A. H. Sameh. Preconditioners. In *Parallelism in Matrix Computations*, pages 311–341. Springer, 2016.

[GR15] R. M. Gower and P. Richtárik. Randomized iterative methods for linear systems. *arXiv preprint arXiv:1506.03296*, 2015.

[HM93] G. T. Herman and L. B. Meyer. Algebraic reconstruction techniques can be made computationally efficient. *IEEE Transactions on Medical Imaging*, 12(3):600–609, 1993.

[HN90] M. Hanke and W. Niethammer. On the acceleration of Kaczmarz's method for inconsistent linear systems. *Linear Alg.Appl.*, 130:83–98, 1990.

[HNR15] A. Hefny, D. Needell, and A. Ramdas. Rows vs. columns: Randomized Kaczmarz or Gauss-Seidel for ridge regression. 2015. Submitted.

[Hof52] A. J. Hoffman. On approximate solutions of systems of linear inequalities. *J. Research Nat. Bur. Standards*, 49:263–265, 1952.

[HS78] C. Hamaker and D. C. Solmon. The angles between the null spaces of x-rays. *J.Math.Anal.Appl.*, 62(1):1–23, 1978. 45L05 (65D99); 0463859 (57 #3798); Gabor T. Herman.

[Kac37] S. Kaczmarz. Angenäherte auflösung von systemen linearer gleichungen. *Bull.Internat.Acad.Polon.Sci.Lettres A*, pages 335–357, 1937.

[LL10] 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.

[LMY15] Y. Li, K. Mo, and H. Ye. Accelerating random Kaczmarz algorithm based on clustering information. *arXiv preprint arXiv:1511.05362*, 2015.

[LW15] J. Liu and S. Wright. An accelerated randomized Kaczmarz algorithm. *Mathematics of Computation*, 2015.

[LWS14] J. Liu, S. J. Wright, and S. Sridhar. An asynchronous parallel randomized Kaczmarz algorithm. *arXiv preprint arXiv:1401.4780*, 2014.

[MNR15] A. Ma, D. Needell, and A. Ramdas. Convergence properties of the randomized extended gauss-seidel and Kaczmarz methods. *SIAM J. Matrix Anal. A.*, 2015. To appear.

[MS54] T. S. Motzkin and I. J. Schoenberg. The relaxation method for linear inequalities. *Canadian J. Math.*, 6:393–404, 1954.

[MTA81] J.-F. Maurras, K. Truemper, and M. Akgül. Polynomial algorithms for a class of linear programs. *Math. Programming*, 21(2):121–136, 1981.

[Nat01] F. Natterer. *The mathematics of computerized tomography*, volume 32. Society for Industrial and Applied Mathematics, Philadelphia, PA; SIAM, 2001. 00A69 (44A12 65R10 68U99 92C55); 1847845 (2002e:00008); Fritz Keinert; Reprint of the 1986 original.

[Nee10] D. Needell. Randomized Kaczmarz solver for noisy linear systems. *BIT*, 50(2):395–403, 2010.

[NSW14a] D. Needell, N. Srebro, and R. Ward. Stochastic gradient descent and the randomized Kaczmarz algorithm. *Mathematical Programming Series A*, 2014. To appear.

[NSW14b] D. Needell, N. Srebro, and R. Ward. Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm. In *Proc. Neural Info. Proc. Systems (NIPS)*, 2014.

[NT13] D. Needell and J. A. Tropp. Paved with good intentions: Analysis of a randomized block Kaczmarz method. *Linear Algebra Appl.*, 2013.

[NW13] D. Needell and R. Ward. Two-subspace projection method for coherent overdetermined linear systems. *J.Fourier Anal.Appl.*, 19(2):256–269, 2013.

[NZZ15] D. Needell, R. Zhao, and A. Zouzias. Randomized block Kaczmarz method with projection for solving least squares. *Linear Algebra and its Applications*, 484:322–343, 2015.

[OZ15a] P. Oswald and W. Zhou. Convergence estimates for Kaczmarz-type methods. *preprint*, 2015.

[OZ15b] P. Oswald and W. Zhou. Random reordering in SOR-type methods. *arXiv preprint arXiv:1510.04727*, 2015.

[Pop98] C. Popa. Extensions of block-projections methods with relaxation parameters to inconsistent and rank-deficient least-squares problems. *BIT*, 38(1):151–176, 1998. 65F20; 1621092 (99d:65121); W. C. Rheinboldt.

[PPKR12] C. Popa, T. Preclik, H. Köstler, and U. Rüde. On Kaczmarz's projection iteration as a direct solver for linear least squares problems. *Linear Algebra and Its Applications*, 436(2):389–404, 2012.

[RM12] P. Richtárik and T. M. Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Math.Program.*, pages 1–38, 2012.

[SV09] T. Strohmer and R. Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *J. Fourier Anal. Appl.*, 15:262–278, 2009.

[Tan71] K. Tanabe. Projection method for solving a singular system of linear equations and its applications. *Numerische Mathematik*, 17(3):203–214, 1971.

[Tel82] J. Telgen. On relaxation methods for systems of linear inequalities. *European J. Oper. Res.*, 9(2):184–189, 1982.

[VZ14] L. A. Végh and G. Zambelli. A polynomial projection-type algorithm for linear programming. *Oper. Res. Lett.*, 42(1):91–96, 2014.

[WAL15] C. Wang, A. Agaskar, and Y. M. Lu. Randomized Kaczmarz algorithm for inconsistent linear systems: An exact MSE analysis. *arXiv preprint arXiv:1502.00190*, 2015.

[WM67] T. M. Whitney and R. K. Meany. Two algorithms related to the method of steepest descent. *SIAM Journal on Numerical Analysis*, 4(1):109–118, 1967.

[XZ02] J. Xu and L. Zikatanov. The method of alternating projections and the method of subspace corrections in hilbert space. *J.Amer.Math.Soc.*, 15(3):573–597, 2002. 65J05 (47A50 47N40 65F10); 1896233 (2003f:65095).

[ZF12] A. Zouzias and N. M. Freris. Randomized extended Kaczmarz for solving least-squares. *SIAM J.Matrix Anal.A.*, 34(2):773–793, 2012.