RANDOMIZED KACZMARZ CONVERGES ALONG SMALL SINGULAR VECTORS*

STEFAN STEINERBERGER†

Abstract. Randomized Kaczmarz is a simple iterative method for finding solutions of linear systems $Ax = b$. We point out that the arising sequence $(x_k)_{k=1}^{\infty}$ tends to converge to the solution x in an interesting way: generically, as $k \to \infty$, $x_k - x$ tends to the singular vector of A corresponding to the smallest singular value. This has interesting consequences: in particular, the error analysis of Strohmer and Vershynin is optimal. It also quantifies the "preconvergence" phenomenon where the method initially seems to converge faster. This fact also allows for a fast computation of vectors x for which the Rayleigh quotient $\|Ax\|/\|x\|$ is small: solve $Ax = 0$ via randomized Kaczmarz.

Key words. Kaczmarz method, algebraic reconstruction technique, ART, projection onto convex sets, POCS, randomized Kaczmarz method, singular vector

AMS subject classifications. 52C35, 65F10

DOI. 10.1137/20M1350947

1. Introduction. Let $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ and full rank. Suppose $x \in \mathbb{R}^n$ and

$$Ax = b;$$

how would one go about finding x? A popular method for solving these types of problems is the Kaczmarz method [9] (also known as the algebraic reconstruction technique in computer tomography [4, 6, 14] or as the projection onto convex sets method [2, 20]). Denoting the rows of A by $a_1, \ldots, a_m$, the linear system can be written as a collection of inner products

$$\forall \, 1 \leq i \leq m : \langle a_i, x \rangle = b_i.$$

The Kaczmarz method is an iterative scheme: given x_k, let us pick an equation, say, the ith equation, and modify x_k the smallest possible amount necessary to make it correct: set $x_{k+1} = x_k + \delta a_i$, where δ is such that $\langle a_i, x_{k+1} \rangle = b_i$. Formally,

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

The Kaczmarz method is then given by cycling through the indices. The method has a geometric interpretation (see Figure 1): x_{k+1} is obtained by projecting x_k onto the hyperplane given by $\langle a_i, x \rangle = b_i$. This interpretation was given in the original 1937 paper of Kaczmarz, who ends his paper by saying that "The convergence of this method is completely obvious from a geometric perspective" [9]. However, given the intricate underlying geometry, convergence rates are difficult to obtain. In a seminal paper, Strohmer and Vershynin [21] showed that by choosing the equations in a random order (where the ith equation is chosen with a likelihood proportional to $\|a_i\|^2$), the method converges exponentially: more precisely, they show that

---

†Department of Mathematics, University of Washington, Seattle, WA 98195 USA (steinerb@uw.edu).
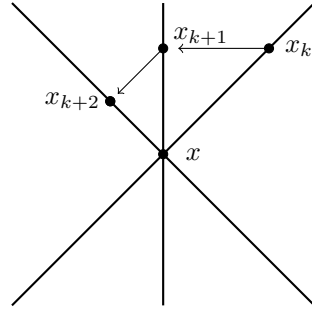
FIG. 1. *A geometric interpretation: iterative projections onto the hyperplanes given by $\langle a_i, x \rangle = b_i$.*

$$\mathbb{E} \, \|x_k - x\|_2^2 \leq \left(1 - \frac{1}{\|A\|_F^2 \|A^{-1}\|_2^2} \right)^k \|x_0 - x\|_2^2,$$

where $\|A^{-1}\|_2$ is the operator norm of the inverse and $\|A\|_F$ is the Frobenius norm. This is a remarkable result in many ways. The analysis of Kaczmarz and Kaczmarz-type algorithms has become quite popular: we refer to [1, 3, 5, 10, 11, 12, 13, 15, 16, 17, 18, 22, 23] and references therein.

2. Results.

2.1. Summary. A priori, one might be inclined to believe that the very nature of the algorithm produces an essentially random sequence in the sense that the normalized vectors $(x_k - x)/\|x_k - x\|$ behave mostly randomly or, alternatively put, that the approximations converge to $x_k \to x$ "from all directions." It is the goal of this short note to show that this is decidedly *not* the case: the singular vectors $v_1, \ldots, v_n$ of the matrix A describe directions of distinguished dynamics. We describe three main observations:

(1) Section 2.1: While $\|x_k - x\|$ decreases exponentially, this is happening at different rates in different subspaces. We show (Theorem 1) that for any (right) singular vector v_ℓ of A associated to the singular value σ_ℓ

$$\mathbb{E} \, \langle x_k - x, v_\ell \rangle = \left(1 - \frac{\sigma_\ell^2}{\|A\|_F^2} \right)^k \langle x_0 - x, v_\ell \rangle .$$

Contributions from the smallest singular vectors (σ_ℓ small) decay the slowest: hence we expect them to be dominant for k large. We recall that for the smallest singular vector σ_n, we have the identity

$$\frac{1}{\sigma_n} = \|A^{-1}\|_2 \qquad \text{leading to the rate} \qquad 1 - \frac{1}{\|A\|_F^2 \|A^{-1}\|_2^2},$$

which recovers the convergence rate of Strohmer and Vershynin [21]. Thus in the generic setting one cannot hope to improve the rate.

(2) Section 2.2: This phenomenon also impacts convergence rates: if x_k is not mainly the linear combination of singular vectors corresponding to small singular values, randomized Kaczmarz converges *faster*, and we have

$$\mathbb{E} \, \|x_{k+1} - x\|_2^2 \leq \left(1 - \frac{1}{\|A\|_F^2} \left\| A \frac{x_k - x}{\|x_k - x\|} \right\|^2 \right) \|x_k - x\|_2^2.$$

We observe that this factor may actually be quite large since

$$\|A^{-1}\|^2 \le \left\| A \frac{x_k - x}{\|x_k - x\|} \right\|^2 \le \|A\|^2.$$

The lower bound is only close to being attained when $x_k - x$ is close to the space spanned by singular vectors given by small singular values of A.

(3) Section 2.3: Once x_k is mainly the combination of singular vectors corresponding to small singular values, the randomized Kaczmarz method is also less likely to explore adjacent subspaces. We show that the normalized vector $(x_k - x)/\|x_k - x\|$ then moves only very slowly on $\mathbb{S}^{n-1}$ since

$$\mathbb{E} \left\langle \frac{x_k - x}{\|x_k - x\|}, \frac{x_{k+1} - x}{\|x_{k+1} - x\|} \right\rangle^2 = 1 - \frac{1}{\|A\|_F^2} \left\| A \frac{x_k - x}{\|x_k - x\|} \right\|^2.$$

2.2. Transfer to small singular vectors. We start by establishing the main phenomenon which can be phrased as a simple algebraic identity.

THEOREM 1. *Let v_ℓ be a (right) singular vector of A associated to the singular value σ_ℓ. Then*

$$\mathbb{E} \langle x_k - x, v_\ell \rangle = \left(1 - \frac{\sigma_\ell^2}{\|A\|_F^2} \right)^k \langle x_0 - x, v_\ell \rangle.$$

We recall that

$$\sum_{i=1}^n \sigma_i^2 = \|A\|_F^2,$$

so the arising factor multiplicative factor is always between 0 and 1. This shows that the randomized Kaczmarz method has exponential decay at different exponential rates that depend on the singular values. In particular, the slowest rate of decay is given by the smallest singular value σ_n. Since

$$\frac{1}{\sigma_n} = \|A^{-1}\|_2 \qquad \text{the rate is} \qquad 1 - \frac{1}{\|A\|_F^2 \|A^{-1}\|_2^2},$$

which recovers the convergence rate of Strohmer and Vershynin [21]. In particular, if x_0 has contributions from singular vectors corresponding to large singular values, the sequence initially decays faster than expected since components corresponding to larger singular values decay quicker. This has been known for a long time: Randomized Kaczmarz initially acts stronger on large singular values, and in particular it serves as a regularizer; we refer to [8, 19]. A previous result in that spirit was obtained by Jiao, Jin, and Lu [7] who obtained an "averaged" version of Theorem 1: by decomposing $\mathbb{R}^n = L + H$ into subspaces spanned by low singular vectors L (corresponding to small singular values) and high singular vectors H (corresponding to large singular values), they obtain the estimate

$$\mathbb{E}\|\pi_H x_{k+1}\|^2 \le (1 - c_1)\|\pi_H x_k\|^2 + c_2\|\pi_L x_k\|^2,$$

where π is the projection and $c_1, c_2 > 0$ are given in terms of singular values.

2.3. Speed of convergence. The phenomenon described in Theorem 1 obviously has an effect on the speed of the algorithm: whenever $x_k - x$ is not close to singular vectors of A that correspond to small singular vector, the randomized Kaczmarz method actually converges faster. Theorem 2 is already implicit in Strohmer and Vershynin [21], where a formulation of it appears as an intermediate step.

THEOREM 2 (see also Strohmer and Vershynin [21]). *We have*

$$\mathbb{E} \|x_{k+1} - x\|_2^2 \leq \left(1 - \frac{1}{\|A\|_F^2} \left\|A \frac{x_k - x}{\|x_k - x\|}\right\|^2\right) \|x_k - x\|_2^2.$$

Observe that for any vector $\|v\| = 1$, we have

$$\|Av\|^2 \geq \sigma_n^2 \|v\|^2 = \|A^{-1}\|_2^{-2} \cdot \|v\|_2,$$

where σ_n is the smallest singular value. But clearly this last inequality is only sharp or close to sharp if v is close to a linear combination of singular vectors corresponding to small singular values; otherwise, the quantity is indeed larger.

2.4. Stability of small singular vectors. We conclude with a result of a geometric flavor. A natural object is the sequence of vectors

$$\frac{x_k - x}{\|x_k - x\|} \in \mathbb{S}^{n-1}.$$

It essentially describes from which direction the sequence $(x_k)_{k=0}^{\infty}$ approaches the solution x. The result in the prior sections showed that we expect this vector will eventually be concentrated on subspaces corresponding to singular vectors associated to small singular values. A different question would be how much we expect the normalized vector to "move around."

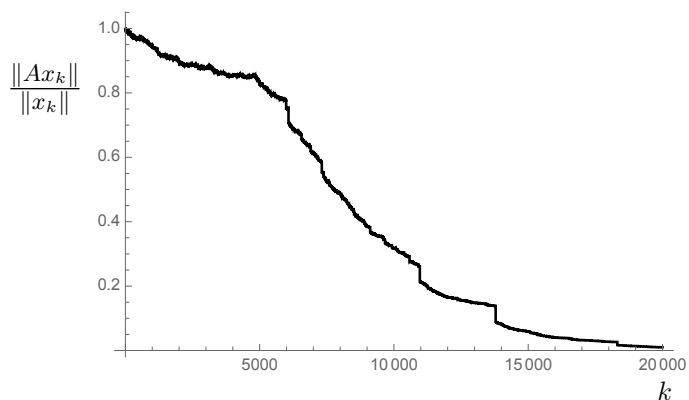THEOREM 3. *If $x_k \neq x$ and $\mathbb{P}(x_{k+1} = x) = 0$, then*

$$\mathbb{E} \left\langle \frac{x_k - x}{\|x_k - x\|}, \frac{x_{k+1} - x}{\|x_{k+1} - x\|} \right\rangle^2 = 1 - \frac{1}{\|A\|_F^2} \left\|A \frac{x_k - x}{\|x_k - x\|}\right\|^2.$$

We recall that for two unit vectors $\|u\| = 1 = \|v\|$, the inner quantity $\langle u, v \rangle^2$ can be interpreted as a notion of distance between the vectors: it is close to 1 only if $u \sim \pm v$. Since we only care about subspaces, the ambiguity in the sign does not play a role. We observe that the direction $(x_k - x)/\|x_k - x\|$ undergoes large fluctuations when $\|A(x_k - x)\|$ is large and is quite stable when $\|A(x_k - x)\|$ is small. Once x_k is mainly composed of small singular vectors, its direction does not change very much anymore. Randomized Kaczmarz spends more time close to these subspaces than it does in other regions of space: the smaller the smallest singular value, the more pronounced the phenomenon. Could this property be used for convergence acceleration?

2.5. Finding small Rayleigh quotients. We discuss an amusing application: finding vectors $\|x\| = 1$ such that $\|Ax\|$ is small. In light of our results, such a sequence of vectors can be obtained from the randomized Kaczmarz method applied to the equation

$$Ax = 0.$$

We illustrate this with a simple example: we take $A \in \mathbb{R}^{1000 \times 1000}$ by picking each entry to be independent and identically distributed $\sim \mathcal{N}(0, 1)$, we add $100 \cdot \text{Id}_{1000}$, and then set the last row a_{1000} to be a tiny perturbation of a_{999} by adding 0.01 to each entry. After that, we normalize all rows to have size $\|a_i\| = 1$. This generates a matrix having singular values between ~ 0.5 and ~ 1.5 and one tiny singular value at scale $\sim 10^{-4}$. We start with $x_0 = (1, \ldots, 1)$ and plot $\|Ax_k\|/\|x_k\|$ (see Figure 2). We observe that within $5000 \sim 5n$ iterations, the ratio is actually smaller than any singular value but the last one—this implies that x_k is actually already mostly given by v_{1000}. After 10,000 iterations, we observe $|\langle x_{10^4}/\|x_{10^4}\|, v_{1000}\rangle| \sim 0.9$.

FIG. 2. *A sample evolution of* $\|Ax_k\|/\|x_k\|$.

3. Proofs.

3.1. Reduction to $Ax = 0$. We start with the ansatz

$$x_k = x + y_k.$$

Plugging in, we obtain

$$
\begin{aligned}
x + y_{k+1} = x_{k+1} &= x_k + \frac{b_i - \langle a_i, x_k \rangle}{\|a_i\|^2} a_i \\
&= x + y_k + \frac{b_i - \langle a_i, x + y_k \rangle}{\|a_i\|^2} a_i \\
&= x + y_k - \frac{\langle a_i, y_k \rangle}{\|a_i\|^2} a_i + \left(\frac{b_i - \langle a_i, x \rangle}{\|a_i\|^2} a_i \right).
\end{aligned}
$$

We note that the final term in parentheses vanishes because x is the exact solution and thus $b_i - \langle a_i, x \rangle = 0$. This results in

$$y_{k+1} = y_k - \frac{\langle a_i, y_k \rangle}{\|a_i\|^2} a_i,$$

which is exactly the randomized Kaczmarz method for $Ay = 0$. Linearity of all the underlying objects (or, alternatively, the geometric interpretation in terms of hyperplane intersections) explains why this is perhaps not too surprising. In any case, this allows us to restrict ourselves to the case $Ax = 0$.

3.2. Proof of Theorem 1.

Proof. As discussed above, we can assume without loss of generality that $b = 0$. It remains to understand the behavior of the randomized Kaczmarz method when applied to $Ax = 0$. We have

$$
\begin{aligned}
\mathbb{E} \langle x_{k+1}, v_\ell \rangle &= \sum_{i=1}^{m} \frac{\|a_i\|^2}{\|A\|_F^2} \left\langle x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|^2} a_i, v_\ell \right\rangle \\
&= \frac{1}{\|A\|_F^2} \sum_{i=1}^{m} \left(\|a_i\|^2 \langle x_k, v_\ell \rangle - \langle a_i, x_k \rangle \langle a_i, v_\ell \rangle \right) \\
&= \langle x_k, v_\ell \rangle - \frac{1}{\|A\|_F^2} \sum_{i=1}^{m} \langle a_i, x_k \rangle \langle a_i, v_\ell \rangle.
\end{aligned}
$$

This sum has a special structure: $\langle a_i, x_k \rangle$ is simply the ith entry of Ax_k and, likewise, $\langle a_i, v_\ell \rangle$ is simply the ith entry of Av_ℓ. This simplifies the sum

$$\sum_{i=1}^{m} \langle a_i, x_k \rangle \langle a_i, v_\ell \rangle = \langle Ax_k, Av_\ell \rangle.$$

We then expand x_k into right singular vectors v_i and make use of $Av_i = \sigma_i u_i$ and the orthogonality of the left singular vectors u_i to obtain

$$\langle x_k, v_\ell \rangle - \frac{1}{\|A\|_F^2} \sum_{i=1}^{m} \langle a_i, x_k \rangle \langle a_i, v_\ell \rangle = \langle x_k, v_\ell \rangle - \frac{1}{\|A\|_F^2} \langle Ax_k, Av_\ell \rangle$$

$$= \langle x_k, v_\ell \rangle - \frac{1}{\|A\|_F^2} \left\langle A \left(\sum_{i=1}^{n} \langle x_k, v_i \rangle v_i \right), Av_\ell \right\rangle$$

$$= \langle x_k, v_\ell \rangle - \frac{1}{\|A\|_F^2} \left\langle \sum_{i=1}^{n} \sigma_i \langle x_k, v_i \rangle u_i, \sigma_\ell u_\ell \right\rangle$$

$$= \langle x_k, v_\ell \rangle \left(1 - \frac{\sigma_\ell^2}{\|A\|_F^2} \right),$$

which is the desired result. $\square$

3.3. Proof of Theorem 2.

Proof. This proof follows the argument of Strohmer and Vershynin [21]. If Z is the (vector-valued) random variable assuming the value $Z = a_j/\|a_j\|^2$ with probability $\|a_j\|^2/\|A\|_F^2$, then Strohmer and Vershynin have established the inequality

$$\|x_{k+1} - x\|_2^2 \leq \left(1 - \left| \left\langle \frac{x_k - x}{\|x_k - x\|}, Z \right\rangle \right|^2 \right) \|x_k - x\|_2^2.$$

At this point we simply compute

$$\mathbb{E} \left| \left\langle \frac{x_k - x}{\|x_k - x\|}, Z \right\rangle \right|^2 = \sum_{i=1}^{m} \frac{\|a_j\|_2^2}{\|A\|_F^2} \left\langle \frac{x_k - x}{\|x_k - x\|}, \frac{a_j}{\|a_j\|_2} \right\rangle^2$$

$$= \frac{1}{\|A\|_F^2} \sum_{i=1}^{m} \left\langle \frac{x_k - x}{\|x_k - x\|}, a_j \right\rangle^2$$

$$= \frac{1}{\|A\|_F^2} \left\| A \frac{x_k - x}{\|x_k - x\|} \right\|^2. \qquad \square$$

3.4. Proof of Theorem 3.

Proof. As above, we can assume without loss of generality that $b = 0$. Moreover, since the entire expression is invariant under scaling, we can assume without loss of generality that $\|x_k\| = 1$. If $\mathbb{P}(x_{k+1} = 0) = 0$, we have

$$\mathbb{E} \left\langle x_k, \frac{x_{k+1}}{\|x_{k+1}\|} \right\rangle^2 = \sum_{i=1}^{m} \frac{\|a_i\|^2}{\|A\|_F^2} \left\langle x_k, \frac{x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|^2} a_i}{\left\| x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|^2} a_i \right\|} \right\rangle^2$$

$$= \sum_{i=1}^{m} \frac{\|a_i\|^2}{\|A\|_F^2} \frac{\left\langle x_k, x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|^2} a_i \right\rangle^2}{\left\| x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|^2} a_i \right\|^2}.$$

We observe that

$$x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|^2} a_i$$

is the projection of x_k onto the subspace $\langle x, a_i \rangle = 0$ and thus

$$\left\langle x_k, x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|^2} a_i \right\rangle = \left\| x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|^2} a_i \right\|^2.$$

Thus

$$\sum_{i=1}^m \frac{\|a_i\|^2}{\|A\|_F^2} \frac{\left\langle x_k, x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|^2} a_i \right\rangle^2}{\left\| x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|^2} a_i \right\|^2} = \sum_{i=1}^m \frac{\|a_i\|^2}{\|A\|_F^2} \frac{\left\| x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|^2} a_i \right\|^4}{\left\| x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|^2} a_i \right\|^2}$$

$$= \sum_{i=1}^m \frac{\|a_i\|^2}{\|A\|_F^2} \left\| x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|} \frac{a_i}{\|a_i\|} \right\|^2.$$

We recall that $\|x_k\| = 1$ and thus

$$\left\| x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|} \frac{a_i}{\|a_i\|} \right\|^2 = \|x_k\|^2 - 2\left\langle \frac{\langle a_i, x_k \rangle}{\|a_i\|} \frac{a_i}{\|a_i\|}, x_k \right\rangle + \left\| \frac{\langle a_i, x_k \rangle}{\|a_i\|} \frac{a_i}{\|a_i\|} \right\|^2$$

$$= 1 - 2\frac{\langle a_i, x_k \rangle^2}{\|a_i\|^2} + \frac{\langle a_i, x_k \rangle^2}{\|a_i\|^2} = 1 - \frac{\langle a_i, x_k \rangle^2}{\|a_i\|^2}.$$

Altogether, we have

$$\sum_{i=1}^m \frac{\|a_i\|^2}{\|A\|_F^2} \left\| x_k - \frac{\langle a_i, x_k \rangle}{\|a_i\|} \frac{a_i}{\|a_i\|} \right\|^2 = \sum_{i=1}^m \frac{\|a_i\|^2}{\|A\|_F^2} \left(1 - \frac{\langle a_i, x_k \rangle^2}{\|a_i\|^2} \right)$$

$$= \frac{1}{\|A\|_F^2} \sum_{i=1}^m \left(\|a_i\|^2 - \langle a_i, x_k \rangle^2 \right)$$

$$= 1 - \frac{1}{\|A\|_F^2} \sum_{i=1}^m \langle a_i, x_k \rangle^2 = 1 - \frac{\|Ax_k\|^2}{\|A\|_F^2}. \qquad \square$$

REFERENCES

[1] Z.-Z. BAI AND W.-T. WU, *On convergence rate of the randomized Kaczmarz method*, Linear Algebra Appl., 553 (2018), pp. 252–269.

[2] C. CENKER, H. G. FEICHTINGER, M. MAYER, H. STEIER, AND T. STROHMER, *New variants of the POCS method using affine subspaces of finite codimension, with applications to irregular sampling*, in Proceedings of SPIE: Visual Communications and Image Processing, 1992, pp. 299–310.

[3] Y. C. ELDAR AND D. NEEDELL, *Acceleration of randomized Kaczmarz method via the Johnson-Lindenstrauss lemma*, Numer. Algorithms, 58 (2011), pp. 163–177.

[4] R. GORDON, R. BENDER, AND G. HERMAN, *Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography*, J. Theoret. Biol., 29 (1970), pp. 471–481.

[5] R. M. GOWER AND P. RICHTARIK, *Randomized iterative methods for linear systems*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 1660–1690.

[6] G. T. HERMAN, Academic Press, New York, 1980. *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*.

[7] Y. Jiao, B. Jin, and X. Lu, *Preasymptotic convergence of randomized Kaczmarz method*, Inverse Problems, 33 (2017), 125012.

[8] B. Jin and X. Lu, *On the regularizing property of stochastic gradient descent*, Inverse Problems, 35 (2018), 015004.

[9] S. Kaczmarz, *Angenaherte Auflosung von Systemen linearer Gleichungen*, Bulletin International de l'Academie Polonaise des Sciences et des Lettres. Classe des Sciences Mathematiques et Naturelles. Serie A, Sciences Mathematiques, 35 (1937), pp. 355–357.

[10] Y.-T. Lee and A. Sidford, *Efficient accelerated coordinate descent methods and faster algorithms for solving linear systems*, in Proceedings of FOCS 2013.

[11] D. Leventhal and A. S. Lewis, *Randomized methods for linear constraints: Convergence rates and conditioning*, Math. Oper. Res., 35 (2010), pp. 641–654.

[12] J. Liu and S. Wright, *An accelerated randomized Kaczmarz algorithm*, Math. Comp., 85 (2016), pp. 153–178.

[13] A. Ma, D. Needell, and A. Ramdas, *Convergence properties of the randomized extended Gauss–Seidel and Kaczmarz methods*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 1590–1604.

[14] F. Natterer, *The Mathematics of Computerized Tomography*, Wiley, New York, 1986.

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

[16] D. Needell and J. Tropp, *Paved with good intentions: Analysis of a randomized block Kaczmarz method*, Linear Algebra Appl., 441 (2014), pp. 199–221.

[17] D. Needell, R. Ward, and N. Srebro, *Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm*, in Proceedings of Advances in Neural Information Processing Systems, 2014, pp. 1017–1025.

[18] D. Needell, R. Zhao, and A. Zouzias, *Randomized block Kaczmarz method with projection for solving least squares*, Linear Algebra Appl., 484 (2015), pp. 322–343.

[19] E. L. Piccolomini and F. Zama, *The conjugate gradient regularization method in computed tomography problems*, Appl. Math. Comput., 102 (1999), pp. 87–99.

[20] K. M. Sezan and H. Stark, *Applications of convex projection theory to image recovery in tomography and related areas*, in Image Recovery: Theory and Application, H. Stark, ed., Academic Press, New York, 1987, pp. 415–462.

[21] T. Strohmer and R. Vershynin, *A randomized Kaczmarz algorithm for linear systems with exponential convergence*, J. Fourier Anal. Appl., 15 (2009), pp. 262–278.

[22] Y. Tan and R. Vershynin, *Phase retrieval via randomized Kaczmarz: Theoretical guarantees*, Inf. Inference, 8 (2019), pp. 97–123.

[23] A. Zouzias and N. M. Freris, *Randomized extended Kaczmarz for solving least squares*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 773–793.