MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

A Kaczmarz Method for Low Rank Matrix Recovery

Mansour, H.; Kamilov, U.; Yilmaz, O.

TR2017-073 June 2017

Abstract

The Kaczmarz method [1], [2], [3] was initially proposed as a row-based technique for reconstructing signals by finding the solutions to overdetermined linear systems. Its usefulness has seen wide application in irregular sampling and tomography [4], [5], [6]. In recent years, several modifications to the Kaczmarz update iterations have improved the recovery capabilities [7], [8], [9], [10], [11]. In particular, signal sparsity was exploited in [12], [13] and low-rankness in [14] to improve the rate of convergence in the overdetermined case while also enabling recovery from underdetermined linear systems.

Signal Processing with Adaptive Sparse Structural Representation (SPARS)

A Kaczmarz Method for Low Rank Matrix Recovery

Hassan Mansour, Ulugbek S. Kamilov
Mitsubishi Electric Research Laboratories (MERL)
Cambridge, MA 02139, USA
Email: {mansour, kamilov}@merl.com

Özgür Yılmaz
Mathematics Department
University of British Columbia
Vancouver, BC V6T1Z4, Canada
Email: oyilmaz@math.ubc.ca

I. Introduction

The Kaczmarz method [1], [2], [3] was initially proposed as a row-based technique for reconstructing signals by finding the solutions to overdetermined linear systems. Its usefulness has seen wide application in irregular sampling and tomography [4], [5], [6]. In recent years, several modifications to the Kaczmarz update iterations have improved the recovery capabilities [7], [8], [9], [10], [11]. In particular, signal sparsity was exploited in [12], [13] and low-rankness in [14] to improve the rate of convergence in the overdetermined case while also enabling recovery from underdetermined linear systems.

Consider the linear system of equations $\mathbf{b} = \mathcal{A}(\mathbf{X})$, where $\mathbf{X} \in \mathbb{R}^{m\times n}$ is a rank r matrix of size $m \times n$ with $r \ll \min\{m, n\}$, $\mathcal{A} : \mathbb{R}^{m\times n} \to p$ is a linear operator that samples p measurements from $\mathbf{X}$, and $\mathbf{b} \in \mathbb{R}^p$ is a measurement vector. The linear system above can be written in vector form as $\mathbf{b} = \mathbf{A}\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^{mn}$ is a vectorization of $\mathbf{X}$, and $\mathbf{A} = [\mathbf{a}_1\mathbf{a}_2 \dots \mathbf{a}_p]^T : \mathbb{R}^{mn} \to \mathbb{R}^p$ is the corresponding measurement matrix with rows $\mathbf{a}_i^T$ for $i \in \{1 \dots p\}$.

Given an estimate $\mathbf{x}_t$ of the signal $\mathbf{x}$ at iteration t, the Kaczmarz method proceeds by projecting $\mathbf{x}_t$ orthogonally onto the solution space defined by row i of $\mathbf{A}$, i.e.,

$$\begin{aligned}\mathbf{x}_{t+1} &= \arg\min_{\mathbf{z}} \tfrac{1}{2}\|\mathbf{z} - \mathbf{x}_t\|_2^2 \text{ subject to } \mathbf{a}_i^T\mathbf{z} = \mathbf{b}(i) \\ &= \mathbf{x}_t + \mathbf{a}_i \tfrac{\mathbf{b}(i) - \mathbf{a}_i^T\mathbf{x}_t}{\|\mathbf{a}_i\|_2^2}.\end{aligned} \tag{1}$$

In what follows, we denote by $\mathcal{R}(\cdot)$ the reshaping operator that maps between the vector and matrix forms of a signal, i.e., $\mathbf{z} := \mathcal{R}(\mathbf{Z})$ and $\mathbf{Z} := \mathcal{R}^*(\mathbf{z})$, where $\mathcal{R}^*(\cdot)$ is the adjoint operator.

II. A Kaczmarz method for low rank matrix recovery

We propose a weighted Kaczmarz method that can recover low rank matrices from linear measurements both in the overdetermined and underdetermined regimes. Our method is inspired by the support identification technique used in [12] for the recovery of k-sparse vectors using weighted Kaczmarz iterations. The method in [12], projects the iterate $\mathbf{x}_t$ onto a weighted hyperplane from the constraint set that favors the support S_t of the largest k entries of $\mathbf{x}_t$. More precisely, the following update rule is used

$$\begin{aligned}\mathbf{x}_{t+1} &= \arg\min_{\mathbf{z}} \tfrac{1}{2}\|\mathbf{z} - \mathbf{x}_t\|_2^2 \text{ subject to } \mathbf{a}_i^T\mathbf{W}_t\mathbf{z} = \mathbf{b}(i) \\ &= \mathbf{x}_t + \mathbf{W}_t\mathbf{a}_i \tfrac{\mathbf{b}(i) - \mathbf{a}_i^T\mathbf{W}_t\mathbf{x}_t}{\|\mathbf{W}_t\mathbf{a}_i\|_2^2},\end{aligned} \tag{2}$$

where $\mathbf{W}_t$ is a diagonal weighting matrix with $\mathbf{W}_t(j,j) = 1,\ j \in S_t$ and $\mathbf{W}_t(j,j) = \omega,\ 0 < \omega < 1,\ j \notin S_t$.

In this work, we propose to project a weighted version $\tilde{\mathbf{x}}_t$ of the iterate $\mathbf{x}_t$ onto the weighted hyperplane that favors the subspace of the rank-r approximation of the matricization of $\mathbf{x}_t$. Let $\mathbf{X}_t \in \mathbb{R}^{m\times n}$ be the matricization of the iterate $\mathbf{x}_t$, and denote by $\mathbf{U}_t \in \mathbb{R}^{m\times r}$ and $\mathbf{V}_t \in \mathbb{R}^{n\times r}$ the singular vector matrices whose columns span the row and column subspaces of the best rank-r approximation of $\mathbf{X}_t$, respectively. Also, let $\widetilde{\mathbf{X}}_t$ be the rank-r subspace weighted iterate given by

$$\widetilde{\mathbf{X}}_t = (1 - \widetilde{\omega})\mathcal{P}_t(\mathbf{X}_t) + \widetilde{\omega}\mathbf{X}_t \tag{3}$$

for some positive weight $\widetilde{\omega} < 1$, where $\mathcal{P}_t(\mathbf{X}_t) = \mathbf{U}_t\mathbf{U}_t^T\mathbf{X}_t\mathbf{V}_t\mathbf{V}_t^T$ is the projection operator onto the subspace spanned by $\mathbf{U}_t$ and $\mathbf{V}_t$. Our low rank seeking Kaczmarz method then uses the following projection

$$\mathbf{X}_{t+1} = \arg\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{Z} - \widetilde{\mathbf{X}}_t\|_F^2 \text{ s.t. } \mathbf{a}_i^T\mathcal{R}((1-\omega)\mathcal{P}_t(\mathbf{Z}) + \omega\mathbf{Z}) = \mathbf{b}(i), \tag{4}$$

where $0 < \omega \leq \widetilde{\omega} < 1$. The solution to (4) can be found by forming the unconstrained Lagrangian and finding its minimizer using

$$\mathbf{X}_{t+1} = \widetilde{\mathbf{X}}_t + (1 - \omega)\mathcal{P}_t(\mathcal{R}^*(\mathbf{a}_i\lambda)) + \omega\mathcal{R}^*(\mathbf{a}_i\lambda), \tag{5}$$

where λ is the Lagrange dual variable given by

$$\begin{aligned}\lambda = &\left((1 - \omega^2)\mathbf{a}_i^T\mathcal{R}(\mathcal{P}_t(\mathcal{R}^*(\mathbf{a}_i))) + \omega^2\|\mathbf{a}_i\|_2^2\right)^{-1} \\ &\times \left(\mathbf{b}(i) - \mathbf{a}_i^T\mathcal{R}((1 - \omega)\mathcal{P}_t(\widetilde{\mathbf{X}}_t) + \omega\widetilde{\mathbf{X}}_t)\right).\end{aligned} \tag{6}$$

We note that a major computational bottleneck is the singular value decomposition (SVD) of the iterates $\mathbf{X}_t$ to compute $\mathbf{U}_t$ and $\mathbf{V}_t$. However, this cost could be mitigated using warm starts and incremental SVD techniques [15]. Finally, we observed that applying the weighting step (3) with $\widetilde{\omega} \approx 0.95$ in the sparse case of [12] results in a significant improvement in performance compared to (2) alone.

III. Numerical evaluation

We test the recovery performance of our proposed low rank Kaczmarz method by recovering an $m \times n$ matrix $\mathbf{X}$ with $m = 100, n = 200$ and rank 5 from standard Gaussian measurements both in the overdetermined and underdetermined regimes. Under both regimes, we decay the weights according to $\widetilde{\omega} = \omega = 1 - \sqrt{\frac{t}{t_{\max}+m+n}}$. We observed that nearly identical performance is achieved if $\widetilde{\omega}$ is kept constant at any value in $[0.1, 0.9]$. In the overdetermined case, we acquire $p = 5mn$ measurements and compare the convergence rate in terms of the relative error $\frac{\|\mathbf{X}_t - \mathbf{X}\|_F}{\|\mathbf{X}\|_F}$ with that of standard Kaczmarz [3] and the adaptive step-size Bregman approach (SVT) of [14]. Fig. 1 illustrates the faster convergence observed by our proposed method (Low Rank Kaczmarz) compared to the reference methods for $t_{\max} = 2p$. In the underdetermined case, we acquire $p = mn/5$ measurements. Fig. 2 compares the performance of the above schemes using $t_{\max} = 40p$ with the batch singular value projection (SVP) algorithm [16], where every iteration uses p times the number of measurements relative to the Kaczmarz methods. The circles indicate the reconstruction performance of SVP over its 40 iterations. We also evaluated the recovery performance in the presence of additive white Gaussian noise $\mathbf{e}$ such that $\frac{\|\mathbf{e}\|_2}{\|\mathbf{A}\mathbf{x}\|_2} = 10^{-2}$ and illustrate the recovery performance in Figs. 3 and 4. We observed that bounding $\widetilde{\omega}$ and ω above 0.5 results in a more robust reconstruction and helps prevent overfitting the noise.

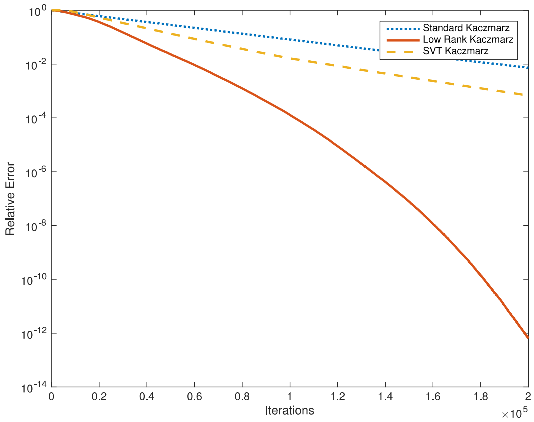

Fig. 1: Reconstruction performance in the overdetermined regime $p = 5mn$ with no noise.

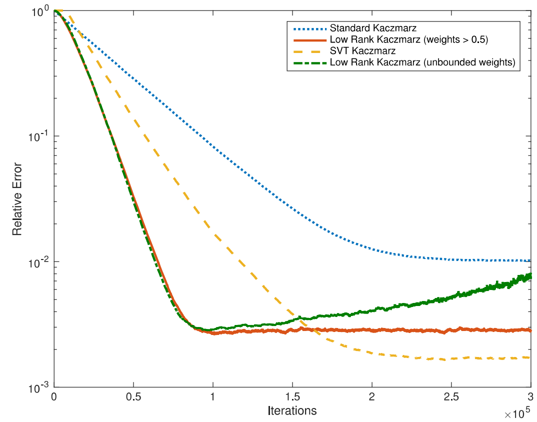

Fig. 3: Reconstruction performance in the overdetermined regime $p = 5mn$ with relative noise level at 10^{-2}.

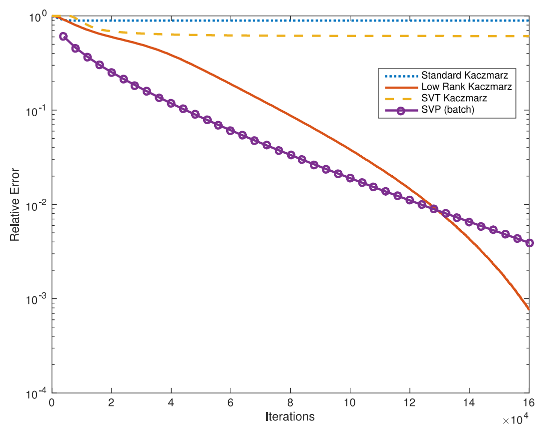

Fig. 2: Reconstruction performance in the underdetermined regime $p = mn/5$ with no noise.

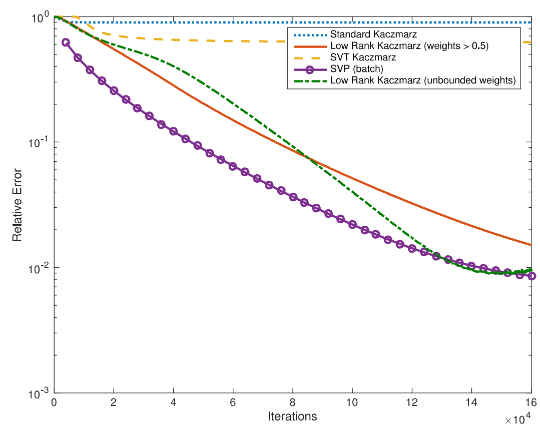

Fig. 4: Reconstruction performance in the underdetermined regime $p = mn/5$ with relative noise level at 10^{-2}.

References

[1] S. Kaczmarz, "Angenäherte Auflösung von Systemen linearer Gleichungen." *Bull. Int. Acad. Polon. Sci. A*, vol. 1937, pp. 355–357, 1937.

[2] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography." *Journal of theoretical biology*, vol. 29, no. 3, pp. 471–481, Dec. 1970.

[3] T. Strohmer and R. Vershynin, "A randomized Kaczmarz algorithm with exponential convergence," *Journal of Fourier Analysis and Applications*, vol. 15, pp. 262 – 278, 2009.

[4] K. 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, 1987, pp. 415–462.

[5] 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 *Conf. SPIE 92 Boston*, 1992, pp. 299–310.

[6] F. Natterer, *The Mathematics of Computerized Tomography*, ser. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 2001.

[7] S. Sra and J. Tropp, "Row-action methods for compressed sensing," in *Proc. IEEE International Conf. on Acoustics, Speech and Signal Processing*, vol. 3, may 2006, p. III.

[8] Y. Eldar and D. Needell, "Acceleration of randomized Kaczmarz method via the Johnson-Lindenstrauss lemma." *Numer. Algorithms*, vol. 58, no. 2, pp. 163–177, 2011.

[9] D. Needell and R. Ward, "Two-subspace projection method for coherent overdetermined systems," 2012.

[10] D. Needell and J. A. Tropp, "Paved with good intentions: Analysis of a randomized block Kaczmarz method," *CoRR*, vol. abs/1208.3805, 2012.

[11] D. Needell, N. Srebro, and R. Ward, "Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm," *http://arxiv.org/abs/1310.5715*, 2014.

[12] H. Mansour and O. Yilmaz, "A sparse randomized Kaczmarz algorithm," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2013, pp. 621–621.

[13] D. A. Lorenz, F. Schpfer, and S. Wenger, "The linearized Bregman method via split feasibility problems: Analysis and generalizations," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 1237–1262, 2014.

[14] F. Schopfer and D. Lorenz, "A sparse Kaczmarz solver and a linearized Bregman method for online compressed sensing," *https://arxiv.org/abs/1610.02889*, 2016.

[15] M. Brand, "Fast low-rank modifications of the thin singular value decomposition," *Linear Algebra and its Applications*, vol. 415, no. 1, pp. 20 – 30, 2006, special Issue on Large Scale Linear and Nonlinear Eigenvalue Problems.

[16] P. Jain, R. Meka, and I. Dhillon, "Guaranteed rank minimization via singular value projection," in *Proceedings of the 23rd International Conference on Neural Information Processing Systems*, ser. NIPS'10. USA: Curran Associates Inc., 2010, pp. 937–945.