On Subsampled Quantile Randomized Kaczmarz

Jamie Haddock
Department of Mathematics
Harvey Mudd College
Claremont, CA, USA
jhaddock@g.hmc.edu

Anna Ma
Department of Mathematics
University of California, Irvine
Irvine, CA, USA
anna.ma@uci.edu

Elizaveta Rebrova
Department of ORFE
Princeton University
Princeton, NJ, USA
elre@princeton.edu

2023 59th Annual Allerton Conference on Communication, Control, and Computing (Allerton) | 979-8-3503-2814-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/Allerton58177.2023.10313381

***Abstract*—When solving noisy linear systems $A\mathbf{x} = \mathbf{b} + \mathbf{c}$, the theoretical and empirical performance of stochastic iterative methods, such as the Randomized Kaczmarz algorithm, depends on the noise level. However, if there are a small number of highly corrupt measurements, one can instead use quantile-based methods to guarantee convergence to the solution $\mathbf{x}$ of the system, despite the presence of noise. Such methods require the computation of the entire residual vector, which may not be desirable or even feasible in some cases. In this work, we analyze the sub-sampled quantile Randomized Kaczmarz (sQRK) algorithm for solving large-scale linear systems which utilize a sub-sampled residual to approximate the quantile threshold. We prove that this method converges to the unique solution to the linear system and provide numerical experiments that support our theoretical findings. We additionally remark on the extremely small sample size case and demonstrate the importance of interplay between the choice of quantile and subset size.**

I. Introduction

With the computational advances of recent decades, the size of datasets regularly analyzed and employed in learning pipelines has skyrocketed. However, with the growth of these available datasets has come the risk of unperceived yet devastating perturbations and alterations to the input data. The presence of corruption, outliers, adversarial noise or perturbations can be entirely disruptive to data analysis results or machine learning models [1], all while the input data is so large that end users cannot inspect for spurious results [2], [3]. The need for robust methods to corruption, outliers, and adversarial noise has only expanded in recent years and is increasingly the focus across numerous subfields of numerical linear algebra, optimization, statistics, and machine learning. Furthermore, these methods should be simple to implement, accompanied by strong theoretical guarantees, and flexible to various applications.

Simple iterative methods like those taught in introductory numerical analysis, numerical linear algebra, and numerical optimization courses are prime candidates for corruption robust methods. The information calculated in-iteration to provide the iterative update can often additionally yield information about the geometry of the problem, the trustworthiness of data, and nearness and existence of a solution. It has become common to aggregate information across multiple iterations to attempt to mitigate the effect of benign noise [4], [5]. Still, variants using this information to avoid the devastating effects of adversarial corruption in the problem-defining data are newer and less well-understood [6]–[9].

JH is grateful to and acknowledges support from NSF DMS #2211318. ER was partially supported by NSF DMS #2309685.

II. Problem setup and related literature

In this work, we consider solving linear problems where a few measurements or components have been corrupted. Problems in which a small number of untrustworthy data can have a devastating effect on a variable of interest have been considered in [10]–[13]. In particular, linear problems with a small number of outlier measurements have led to an interest in methods for robust linear regression [14]–[16]. Other relevant work includes *min-k loss SGD* [8], robust SGD [9], [17], and Byzantine approaches [18], [19].

Specifically, our setting here is as follows: consider the setting in which one is given a full rank matrix $A \in \mathbb{R}^{m\times n}$ ($m \gg n$) and a vector $\mathbf{b} + \mathbf{c} \in \mathbb{R}^m$, where $A\mathbf{x} = \mathbf{b}$ is an overdetermined consistent system with solution $\mathbf{x}$ and $\mathbf{c}$ is a sparse corruption vector. Assume the number of corruptions is no more than a fraction $\beta \in (0,1)$ of the total number of measurements, $\|\mathbf{c}\|_{\ell_0} \leq \beta m$ and define the set of corrupted equations to be $C = \text{supp}(\mathbf{c}) \subset \{1,2,\ldots,m\}$. We won't assume any structure or distribution of the corruptions apart from the sparsity.

This setting was successfully tackled in recent years via iterative methods that can integrate corruption-avoiding strategies into the iterate's design. One popular and convenient method for solving large-scale linear systems is the Randomized Kaczmarz (RK) method [20], [21] that solves the system by iterative projections into the individual solution hyperplanes, namely,

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \frac{b_i - \langle \mathbf{x}_k, \mathbf{a}_i \rangle}{\|\mathbf{a}_i\|^2}\mathbf{a}_i.$$

In the consistent overdetermined case, its expected convergence is given by

$$\mathbb{E}\|\mathbf{x}_k - \mathbf{x}\|^2 \leq r^k \|\mathbf{x}_0 - \mathbf{x}\|^2,$$

where $r = 1 - \sigma_{\min}^2(A)/\|A\|_F^2$.

A simple observation that is of utmost importance for the corrupt setting is that each iteration of the RK algorithm has the step length, $\|\mathbf{x}_{k+1} - \mathbf{x}_k\|$, of one of the *residuals*, that is, the distance from the current iterate $\mathbf{x}_k$ to one of the solution hyperplanes. In [22], it was proposed to consider the residual information to identify the "far apart" corrupted hyperplanes.

This method, however, only works when $\mathbf{x}_k$ is closer to the true solution $\mathbf{x}_*$ than all other corrupted hyperplanes, and this is achievable only with a very low corruption rate. The next idea was presented in [6] to consider the residual's stable statistics (quantiles) to identify the hyperplanes that are farther apart from $\mathbf{x}_k$ and do not use them for the current step only. This method successfully works for the systems with nearly 50% corruption rate (note that 50% is the theoretical threshold for identification of an adversarial corrupted subsystem). It is proven to converge at the per iteration rate of the same order as RK on the uncorrupted system, given that β is a small enough constant.

Some follow-up works considered the case when sparse corruptions are mixed with small noise [23], [24], and gave an alternative approach for the convergence proof [7]. A major practical hurdle of QuantileRK is a very slow exploration phase: to find the quantile estimate; one has to look at all m (recall that $m \gg n$) residuals that are updated at each next iteration $\mathbf{x}_k$. In [25], the authors propose a way to average over a significant portion of the residual thus using the information obtained on the exploration stage more efficiently. But the question *how the exploration stage can be done faster* remained. In the original QuantileRK paper, the algorithm naturally allows to compute the quantile using only a subset of the residual ([6][Algorithm 1]), but the effect of the subset T is never analyzed. Some empirical analysis of using subresidual of the size $0.2m$ is provided in [23].

In this paper, we give the first explicit analysis of the influence of residual subsampling on the convergence of the QuantileRK method. We give theoretical guarantees (Theorem 1) for the case when we subsample a fraction of the total set of equations αm and complement our study with extensive experimental data and the heuristical analysis of the convergence when the sample size is even smaller and have only constant amount of the residuals.

III. Sub-sampled Quantile Randomized Kaczmarz

The Sub-sampled Quantile Randomized Kaczmarz (sQRK) algorithm reduces the computational cost of QRK by selecting only a subset of equations to evaluate the residual and estimate the quantile. For $0 < q < 1$, the q-th quantile of a set S is defined as

$$q\text{-quant}(S) := s \in S \text{ such that } |\{r \in S : r \leq s\}| = \lfloor q|S| \rfloor.$$

At every iteration, a subset $\tau_k \subset [m]$ is uniformly selected such that $|\tau_k| \geq \alpha m$ for some $\alpha > 0$. The quantile is then evaluated using the subset τ_k and a randomly selected equation whose residual is smaller than the quantile is selected to project into. The pseudo-code for sQRK is provided in Algorithm 1.

We define two key sets associated with the algorithm:

- $B_k = \{i \in \tau_k : |\langle \mathbf{a}_i, \mathbf{x}_{k-1}\rangle - \hat{b}_i| < q\text{-quant}(S)\}$ is the set of all accepted equations in the sample, and
- $S_k := C \cap B_k$, the set of corrupted equations that can be selected after sampling and applying the quantile threshold.

Algorithm 1 Sub-Sample Quantile Randomized Kaczmarz

procedure sQRK($\mathbf{A}, \hat{\mathbf{b}}$, q, α, N)
  $\mathbf{x}_1 = \mathbf{0}$
  for $k = 1, \ldots, N$ **do**
    Sample $\tau_k \subset [m]$ uniformly such that $|\tau_k| = \lceil \alpha m \rceil$
    $\gamma_k = q\text{-quant}\left(\{|\langle \mathbf{a}_j, \mathbf{x}_{k-1}\rangle - \hat{b}_j|\}_{j \in \tau_k}\right)$
    $B_k = \{i \in \tau_k : |\langle \mathbf{a}_i, \mathbf{x}_{k-1}\rangle - \hat{b}_i| < \gamma_k\}$
    Sample $i \sim \text{Unif}(B_k)$
    $\mathbf{x}_{k+1} = \mathbf{x}_k + (\hat{b}_i - \langle \mathbf{a}_i, \mathbf{x}_k\rangle)\mathbf{a}_i^T$
  end for return $\mathbf{x}_N$
end procedure

It will be useful to bound the sizes of the sets B_k and S_k and their difference. First, we note the simple facts that $B_k \subset \tau_k$ and $S_k \subset C$. Thus,

$$|B_k| \geq \alpha q m, \qquad |S_k| \leq \beta m, \tag{1}$$

and so

$$|B_k \setminus S_k| \geq (\alpha q - \beta) m. \tag{2}$$

For the main analysis, we assume that $\alpha q - \beta \gg 0$, that is, we are guaranteed to have enough uncorrupt equations in any random sample. See Section V-D for the setting when this does not hold.

IV. Theoretical Guarantees

In this section, we provide theoretical guarantees for the sQRK algorithm. Theorem 1 presents our main results, which show that, in expectation, sQRK converges linearly to the solution of the consistent linear system $A\mathbf{x} = \mathbf{b}$, despite only having access to $\hat{\mathbf{b}} = \mathbf{b} + \mathbf{c}$. To prove Theorem 1, we first prove, in Lemma 1, an upper bound on the quantile of the residual for a subset of selected equations. Using the quantile bound, the expected error is then controlled by conditioning on the event of selecting a corrupt equation in Lemma 2 and the event of selecting a non-corrupt equation in Lemma 3.

Our main result depends on the factor

$$\sigma_{\alpha,q,\beta,\min}(A) = \min_{S \subset [m],\, |S|/m \geq \alpha q - \beta} \inf_{\mathbf{x} \neq \mathbf{0}} \frac{\|A_S \mathbf{x}\|}{\|\mathbf{x}\|}.$$

This term will govern the convergence of sQRK conditioned on sampling uncorrupted equations. With this factor, we can state our main result.

Theorem 1. *Let $A \in \mathbb{R}^{m \times n}$ with $m > n$ be a row-normalized, full rank matrix, $\mathbf{x} \in \mathbb{R}^n$ be fixed, and $\mathbf{b} = A\mathbf{x}$. Let $\mathbf{x}_k$ denote the iterates of Algorithm 1 applied to matrix A and measurements $\hat{\mathbf{b}} = \mathbf{b} + \mathbf{c}$ where the corruption vector $\mathbf{c}$ satisfies $\|\mathbf{c}\|_0 = \beta m$. Using quantile q, sampling rate α, and assuming $\alpha(1-q) > \beta$ and $\alpha q > \beta$, if*

$$r_G < \frac{1 - \frac{\beta}{\alpha q}\tilde{r}_C}{1 - \frac{\beta}{\alpha q}} \tag{3}$$

where

$$\tilde{r}_C = \left(1 + \frac{2}{\sqrt{\beta m}} \frac{\sigma_{\max}^2(A)}{\sqrt{m[\alpha(1-q) - \beta]}} + \frac{\sigma_{\max}^2(A)}{m[\alpha(1-q) - \beta]}\right)$$

and

$$r_G = 1 - \frac{\sigma^2_{\alpha,q,\beta,\min}}{\alpha q m},$$

then sQRK converges at least linearly in expectation,

$$\mathbb{E}\|\mathbf{x}_k - \mathbf{x}\|^2 \le r^k \|\mathbf{x}_0 - \mathbf{x}\|^2$$

where $r = \left(1 - \frac{\beta}{\alpha q}\right) r_G + \frac{\beta}{\alpha q} \tilde{r}_C$.

Remark 1. *Note that condition* (3) *ensures that the rate* $r < 1$ *and the method is indeed convergent. A simple calculation shows that it is equivalent to the condition*

$$\frac{\beta}{\alpha q} + \beta \frac{\sigma^2_{\max}(A)}{\sigma^2_{\alpha,q,\beta,\min}} \left(\frac{2}{\sqrt{\beta}\sqrt{d}} + \frac{1}{d} \right) < 1$$

for $d = \alpha(1-q) - \beta$. *See also Figure 1 that explores joint admissible values for the parameters* α *and* q.

The proofs of Theorem 1 and its supporting lemmas are similar to those that appear in [7] but are not immediate results of the previous work. In particular, we pay special attention to the impact of the sub-sampling rate α here. Recall that B_k denotes the "acceptable" set of equations after sampling and using the quantile at iteration k with

$$\gamma_k(\mathbf{x}_{k-1}) = q\text{-quant}\left(\{ |\langle \mathbf{a}_j, \mathbf{x}_{k-1} \rangle - \hat{b}_j| \}_{j \in \tau_k} \right),$$
$$\text{and } B_k = \{ i \in \tau_k : |\langle \mathbf{a}_i, \mathbf{x}_{k-1} \rangle - \hat{b}_i| < \gamma_k \}.$$

The set $S_k = C \cap B_k$ this is the set of corrupted equations that can be selected after sampling and using the quantile. If $S_k = \emptyset$, we can be sure that we are projecting onto an uncorrupted equation at iteration k.

Lemma 1. *Assume* $|S_k| \ge 1$. *Let* $\alpha, q > 0$ *and assume* $\alpha(1-q) > \beta$, *then for any arbitrary vector* $\mathbf{v} \in \mathbb{R}^n$ *and for all k:*

$$\gamma_k(\mathbf{v}) \le \frac{\sigma_{\max}(A)}{\sqrt{m[\alpha(1-q)-\beta]}} \|\mathbf{v} - \mathbf{x}\|.$$

Proof. We begin by considering non-corrupt equations: $i \notin C$ where

$$\langle \mathbf{a}_i, \mathbf{x} \rangle = \hat{b}_i = b_i.$$

Taking the sum of the squared residuals for non-corrupt equations, we get

$$\sum_{\substack{i=1 \\ i \notin C}}^{m} |\langle \mathbf{a}_i, \mathbf{v} \rangle - \hat{b}_i|^2 \le \|A_{\notin C}\mathbf{v} - \mathbf{b}_{\notin C}\|^2$$
$$= \|A_{\notin C}\mathbf{v} - A_{\notin C}\mathbf{x}\|^2 \le \sigma^2_{\max}(A)\|\mathbf{v} - \mathbf{x}\|^2.$$

At least $\alpha(1-q)m$ of the αm values $\{|\langle \mathbf{v}, \mathbf{a}_i \rangle - \hat{b}_i|\}_{i \in \tau_k}$ are at least $\gamma_k(\mathbf{v})$ and since $|S_k| \le |C| \le \beta m$, at least $\alpha(1-q)m - \beta m$ belong to equations that have not been corrupted. Thus,

$$\gamma_k^2(\mathbf{v}) m[\alpha(1-q)-\beta] \le \sum_{\substack{i \in \tau_k \\ i \notin C}} |\langle \mathbf{a}_i, \mathbf{v} \rangle - \hat{b}_i|^2$$
$$\le \sigma^2_{\max}(A)\|\mathbf{v} - \mathbf{x}\|^2.$$

and therefore

$$\gamma_k(\mathbf{v}) \le \frac{\sigma_{\max}(A)}{\sqrt{m[\alpha(1-q)-\beta]}} \|\mathbf{v} - \mathbf{x}\|.$$

□

Lemma 2. *The expected approximation error over the set of acceptable corrupted equations* S_k *is*

$$\mathbb{E}_{i \in S_k} \|\mathbf{x}_{k+1} - \mathbf{x}\|^2 \le r_C \|\mathbf{x}_k - \mathbf{x}\|^2,$$

where

$$r_C = 1 + \frac{2}{\sqrt{|S_k|}} \frac{\sigma^2_{\max}(A)}{\sqrt{m[\alpha(1-q)-\beta]}} + \frac{\sigma^2_{\max}(A)}{m[\alpha(1-q)-\beta]}. \quad (4)$$

Proof. Recall the definition of $\mathbf{x}_{k+1}$ from Algorithm 1 is

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \left(\hat{b}_i - \langle \mathbf{x}_k, \mathbf{a}_i \rangle\right) \mathbf{a}_i.$$

The approximation error can be written as

$$\|\mathbf{x}_{k+1} - \mathbf{x}\|^2 = \|\mathbf{x}_k - \mathbf{x}\|^2 + 2\langle \mathbf{x}_k - \mathbf{x}, \mathbf{v} \rangle + \|\mathbf{v}\|^2, \quad (5)$$

where

$$\mathbf{v} = (\hat{b}_i - \langle \mathbf{x}_k, \mathbf{a}_i \rangle)\mathbf{a}_i.$$

We proceed by bounding the last two terms of (5). For $i \in S_k$, the term $\|\mathbf{v}\|^2$ is uniformly small since

$$\|\mathbf{v}\|^2 = \|(\hat{b}_i - \langle \mathbf{x}_k, \mathbf{a}_i \rangle)\mathbf{a}_i\|^2 = |\hat{b}_i - \langle \mathbf{x}_k, \mathbf{a}_i \rangle|^2$$
$$\le \gamma_k^2(\mathbf{x}_k)$$
$$\le \frac{\sigma^2_{\max}(A)}{m[\alpha(1-q)-\beta]} \|\mathbf{x}_k - \mathbf{x}\|^2, \quad (6)$$

where the second to last inequality follows from $S_k \subseteq B_k$ and the last inequality follows from Lemma 1.

It remains to bound $\mathbb{E}_{i \in S_k} 2\langle \mathbf{x}_k - \mathbf{x}, \mathbf{v} \rangle$. We begin by showing

$$\mathbb{E}_{i \in S_k} 2\langle \mathbf{x}_k - \mathbf{x}, \mathbf{v} \rangle = \frac{2}{|S_k|} \sum_{i \in S_k} \langle \mathbf{x}_k - \mathbf{x}, (\hat{b}_i - \langle \mathbf{x}_k, \mathbf{a}_i \rangle)\mathbf{a}_i \rangle$$
$$= \frac{2}{|S_k|} \sum_{i \in S_k} (\hat{b}_i - \langle \mathbf{x}_k, \mathbf{a}_i \rangle)\langle \mathbf{x}_k - \mathbf{x}, \mathbf{a}_i \rangle$$

The Cauchy-Schwarz inequality yields

$$\sum_{i \in S_k} (\hat{b}_i - \langle \mathbf{x}_k, \mathbf{a}_i \rangle)\langle \mathbf{x}_k - \mathbf{x}, \mathbf{a}_i \rangle$$
$$\le \left(\sum_{i \in S_k} (\hat{b}_i - \langle \mathbf{x}_k, \mathbf{a}_i \rangle)^2 \sum_{i \in S_k} \langle \mathbf{x}_k - \mathbf{x}, \mathbf{a}_i \rangle^2 \right)^{\frac{1}{2}},$$

and thus

$$\mathbb{E}_{i \in S_k} 2\langle \mathbf{x}_k - \mathbf{x}, \mathbf{v} \rangle$$
$$\le \frac{2}{|S_k|} \left(\sum_{i \in S_k} (\hat{b}_i - \langle \mathbf{x}_k, \mathbf{a}_i \rangle)^2 \sum_{i \in S_k} \langle \mathbf{x}_k - \mathbf{x}, \mathbf{a}_i \rangle^2 \right)^{\frac{1}{2}}$$
$$\le \frac{2}{|S_k|} \sqrt{|S_k|} \frac{\sigma_{\max}(A)}{\sqrt{m[\alpha(1-q)-\beta]}} \|\mathbf{x}_k - \mathbf{x}\|$$
$$\times \left(\sum_{i \in S_k} \langle \mathbf{x}_k - \mathbf{x}, \mathbf{a}_i \rangle^2 \right)^{\frac{1}{2}}.$$

At this point, we estimate

$$\sum_{i\in S_k}\langle \mathbf{x}_k-\mathbf{x},\mathbf{a}_i\rangle^2 \le \sum_{i=1}^{m}\langle \mathbf{x}_k-\mathbf{x},\mathbf{a}_i\rangle^2 = \|A(\mathbf{x}_k-\mathbf{x})\|^2 \le \sigma_{\max}^2(A)\|\mathbf{x}_k-\mathbf{x}\|^2,$$

and hence

$$\mathbb{E}_{i\in S_k} 2\langle \mathbf{x}_k-\mathbf{x},v\rangle \le \frac{2}{\sqrt{|S_k|}}\frac{\sigma_{\max}^2(A)}{\sqrt{m[\alpha(1-q)-\beta]}}\|\mathbf{x}_k-\mathbf{x}\|^2. \quad (7)$$

Taking the expectation over $i\in S_k$ in (5) and using the bounds (6) and (7) obtains the final result:

$$\mathbb{E}_{i\in S_k}\|\mathbf{x}_{k+1}-\mathbf{x}\|^2 \le r_C\|\mathbf{x}_k-\mathbf{x}\|^2,$$

where r_C is as defined in (4). □

Now we consider the subset $B_k\setminus S_k$ and the event in which we project onto an uncorrupted equation.

Lemma 3. *Let $A\in\mathbb{R}^{m\times n}$ with $m>n$ be a row-normalized, full rank matrix, $\mathbf{x}\in\mathbb{R}^n$ be fixed, and $\mathbf{b}=A\mathbf{x}$. Let $\mathbf{x}_k$ denote the iterates of Algorithm 1 applied to matrix A and measurements $\hat{\mathbf{b}}=\mathbf{b}+\mathbf{c}$ where the corruption vector $\mathbf{c}$ satisfies $\|\mathbf{c}\|_0=\beta m$. Using quantile q, sampling rate α, and assuming $\alpha q>\beta$, and conditioning on the case that the sampled row index is uncorrupted yields at least linear convergence in expectation with*

$$\mathbb{E}_{i\in B_k\setminus S_k}\|\mathbf{x}_{k+1}-\mathbf{x}\|^2 \le r_G\|\mathbf{x}_k-\mathbf{x}\|^2,$$

where

$$r_G = 1-\frac{\sigma_{\alpha,q,\beta,\min}^2(A)}{\alpha q m}. \quad (8)$$

The proof of this lemma follows directly from the proof of [7, Lemma 3] where B and S are substituted by B_k and S_k, and we use the estimate $\|A_{B_k\setminus S_k}\|_F^2\le \alpha q m$.

Proof of Theorem 1. By the Law of Total Expectation and noting that $\mathbb{P}(i\notin B_k)=0$, we have

$$\begin{aligned}\mathbb{E}\|\mathbf{x}_{k+1}-\mathbf{x}\|^2 &= \mathbb{P}(i\in B_k\setminus S_k)\mathbb{E}_{i\in B_k\setminus S_k}\|\mathbf{x}_{k+1}-\mathbf{x}\|^2 \\ &\quad +\mathbb{P}(i\in S_k)\mathbb{E}_{i\in S_k}\|\mathbf{x}_{k+1}-\mathbf{x}\|^2 \\ &\le \frac{\alpha q m-|S_k|}{\alpha q m}r_G\|\mathbf{x}_k-\mathbf{x}\|^2 \\ &\quad +\frac{|S_k|}{\alpha q m}r_C\|\mathbf{x}_k-\mathbf{x}\|^2 \\ &\le \left[r_G+\frac{\beta}{\alpha q}(\tilde{r}_C-r_G)\right]\|\mathbf{x}_k-\mathbf{x}\|^2 \\ &= r\|\mathbf{x}_k-\mathbf{x}\|^2,\end{aligned}$$

where we have used the fact that $r_C-r_G\ge 0$ and $0\le|S_k|\le \beta m$ in the second inequality to bound $\frac{|S_k|}{\alpha q m}(r_C-r_G)$ and to conclude that $\frac{|S_k|}{\alpha q m}r_C\le\frac{\beta}{\alpha q}\tilde{r}_C$.

□

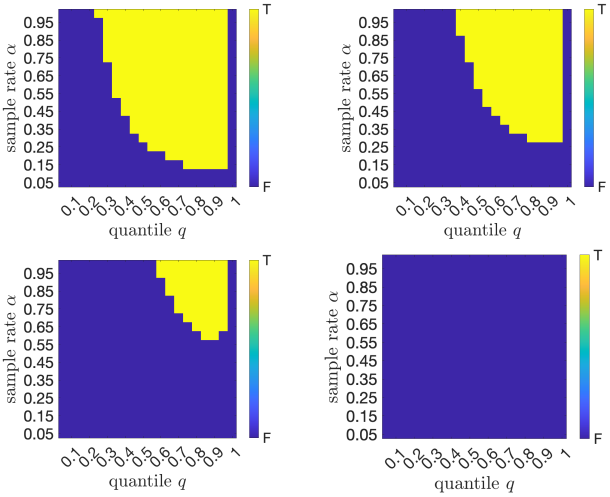

Fig. 1. Heatmap indicating which hyperparameter sets (q,α) satisfy the assumptions of Theorem 1, $\alpha(1-q)>\beta$, $\alpha q>\beta$, and (3), for $\beta=10^{-5}$ (upper left), $\beta=10^{-4}$ (upper right), $\beta=10^{-3}$ (lower left), and $\beta=10^{-2}$ (lower right) for a system defined by row-normalized $A\in\mathbb{R}^{50000\times 100}$ with i.i.d. $\mathcal{N}(0,1)$ entries.

V. Numerical Experiments

The experiments presented in this section were performed in MATLAB R2021b on a MacBook Pro 2019 with a 2.3 GHz 8-Core Intel Core i9 processor and 16 GB 2667 MHz DDR4 RAM.

A. Checking assumptions

In Figure 1, we plot a heatmap indicating which hyperparameter sets (q,α) satisfy the assumptions of Theorem 1, $\alpha(1-q)>\beta$, $\alpha q>\beta$, and

$$r_G < \frac{1-\frac{\beta}{\alpha q}\tilde{r}_C}{1-\frac{\beta}{\alpha q}}.$$

Here, our system is defined by a matrix $A\in\mathbb{R}^{50000\times 100}$ which is generated with i.i.d. $\mathcal{N}(0,1)$ entries and then row-normalized. We approximate $\sigma_{\alpha,q,\beta,\min}(A)$ by taking 100 uniform random samples of index subsets of size at least $\lceil(\alpha q-\beta)m\rceil$ and recording the minimum singular value encountered in these submatrices. The heatmap indicates "T" (true) if all three assumption hold for the indicated (q,α) pair and value of β and "F" (false) otherwise. We provide heatmaps for $\beta=10^{-5}$ (upper left), $\beta=10^{-4}$ (upper right), $\beta=10^{-3}$ (lower left), and $\beta=10^{-2}$ (lower right). We note that $|C|=0$ for $\beta=10^{-5}$, $|C|=5$ for $\beta=10^{-4}$, $|C|=50$ for $\beta=10^{-3}$, and $|C|=500$ for $\beta=10^{-2}$. As expected, the region of pairs (q,α) satisfying the assumptions is larger for smaller corruption rate β.

B. Empirical convergence

In Figures 2-5, we plot the empirical convergence of ten trials of sQRK with $q=0.9$ and a variety of β and α values. In each figure, we plot the empirical convergence of ten independent trials of 1000 iterations of sQRK with respect to

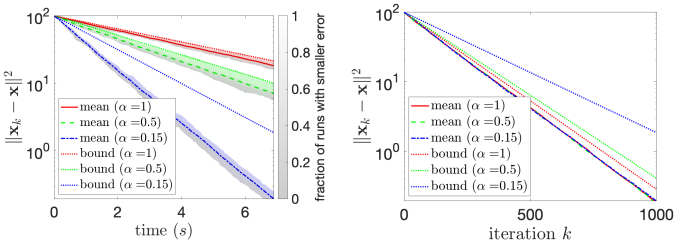

Fig. 2. Empirical convergence with respect to time (left) and iterations (right) for ten trials of sQRK with $\alpha = 1, 0.5$, or 0.15 on system defined by row-normalized $A \in \mathbb{R}^{50000\times 100}$ and $\hat{\mathbf{b}} \in \mathbb{R}^{50000}$ with $\lfloor \beta m \rfloor$ corrupted entries where $\beta = 10^{-5}$. We additionally plot the bounds provided by Theorem 1 in dotted lines if they are decreasing.

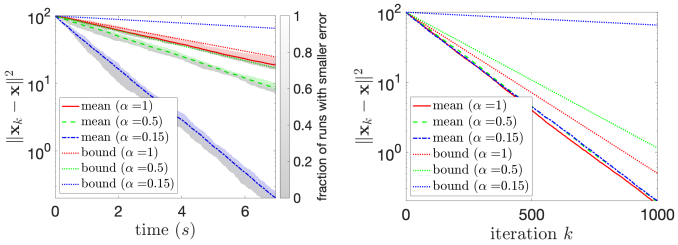

Fig. 3. Empirical convergence with respect to time (left) and iterations (right) for ten trials of sQRK with $\alpha = 1, 0.5$, or 0.15 on system defined by row-normalized $A \in \mathbb{R}^{50000\times 100}$ and $\hat{\mathbf{b}} \in \mathbb{R}^{50000}$ with $\lfloor \beta m \rfloor$ corrupted entries where $\beta = 10^{-4}$. We additionally plot the bounds provided by Theorem 1 in dotted lines if they are decreasing.

wall clock time on the left and with respect to iteration k on the right. In the error versus time plots on the left, we plot a cloud indicating the errors of the 10 independent trials (brighter color indicates more trial errors were below) and lines indicating the mean error of the ten trials. In the error versus iteration plots on the right, we only plot the mean error over the ten trials as the errors are highly similar for different values of α. We additionally plot the bounds given by Theorem 1 if they are decreasing.

For Figures 2-5, we generate a single system defined by $A \in \mathbb{R}^{50000\times 100}$ and $\mathbf{b} = \mathbf{0} \in \mathbb{R}^{50000}$. We generate A with i.i.d. $\mathcal{N}(0,1)$ entries, then row-normalize it. In each figure, we generate $\mathbf{c}$ to have $\lfloor \beta m \rfloor$ nonzero entries each with value ten. Thus $\hat{\mathbf{b}} = \mathbf{c}$ has $\lfloor \beta m \rfloor$ nonzero entries each with value ten. We approximate $\sigma_{\alpha,q,\beta,\min}(A)$ by recording the minimum singular value encountered in each of the $B_k \setminus C$ submatrices encountered during the ten trials of 1000 iterations of sQRK.

In Figure 2, we plot the empirical convergence of ten trials of 1000 iterations of sQRK with $q = 0.9$, $\beta = 10^{-5}$, and $\alpha = 1, 0.5$, and 0.15 values. We plot the empirical convergence of ten independent trials of 1000 iterations of sQRK with respect to wall clock time on the left (brighter color of cloud indicates more trial errors were below) and the mean error for each α, and we plot the mean error for each α with respect to iteration k on the right. We additionally plot the bounds given by Theorem 1 for each α (all were decreasing).

In Figure 3, we plot the empirical convergence of ten trials of 1000 iterations of sQRK with $q = 0.9$, $\beta = 10^{-4}$, and $\alpha = 1, 0.5$, and 0.15 values. We plot the empirical convergence of ten independent trials of 1000 iterations of sQRK with respect to wall clock time on the left (brighter color of cloud indicates more trial errors were below) and the mean error for each α, and we plot the mean error for each α with respect to iteration k on the right. We additionally plot the bounds given by Theorem 1 for each α (all were decreasing).

In Figure 4, we plot the empirical convergence of ten trials of 1000 iterations of sQRK with $q = 0.9$, $\beta = 10^{-3}$, and $\alpha = 1, 0.5$, and 0.15 values. We plot the empirical convergence of ten independent trials of 1000 iterations of sQRK with respect to wall clock time on the left (brighter color of cloud indicates more trial errors were below) and the mean error for each α, and we plot the mean error for each α with respect to

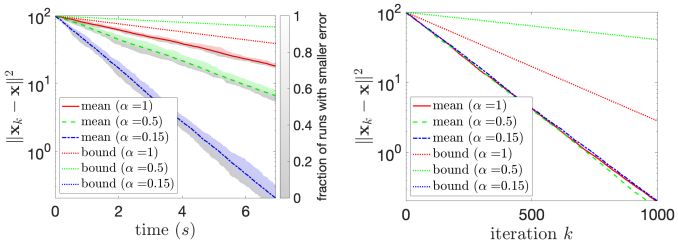

Fig. 4. Empirical convergence with respect to time (left) and iterations (right) for ten trials of sQRK with $\alpha = 1, 0.5$, or 0.15 on system defined by row-normalized $A \in \mathbb{R}^{50000\times 100}$ and $\hat{\mathbf{b}} \in \mathbb{R}^{50000}$ with $\lfloor \beta m \rfloor$ corrupted entries where $\beta = 10^{-3}$. We additionally plot the bounds provided by Theorem 1 in dotted lines if they are decreasing.

iteration k on the right. We additionally plot the bounds given by Theorem 1 for each $\alpha = 1$ and $\alpha = 0.5$.

In Figure 4, we plot the empirical convergence of ten trials of 1000 iterations of sQRK with $q = 0.9$, $\beta = 10^{-2}$, and $\alpha = 1, 0.5$, and 0.15 values. We plot the empirical convergence of ten independent trials of 1000 iterations of sQRK with respect to wall clock time on the left (brighter color of cloud indicates more trial errors were below) and the mean error for each α, and we plot the mean error for each α with respect to iteration k on the right. Note that none of the bounds given by Theorem 1 for any α were decreasing so they do not appear.

C. Varying q

In Figures 6 and 7, we plot the empirical convergence of ten trials of sQRK with and a variety of β, α, and q values. In each

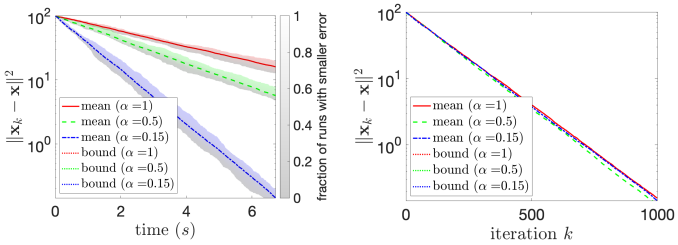

Fig. 5. Empirical convergence with respect to time (left) and iterations (right) for ten trials of sQRK with $\alpha = 1, 0.5$, or 0.15 on system defined by row-normalized $A \in \mathbb{R}^{50000\times 100}$ and $\hat{\mathbf{b}} \in \mathbb{R}^{50000}$ with $\lfloor \beta m \rfloor$ corrupted entries where $\beta = 10^{-2}$. We additionally plot the bounds provided by Theorem 1 in dotted lines if they are decreasing.

figure, we plot the empirical convergence of ten independent trials of 1000 iterations of sQRK with respect to wall clock time on the left and with respect to iteration k on the right. In the error versus time plots on the left, we plot a cloud indicating the errors of the 10 independent trials (brighter color indicates more trial errors were below) and lines indicating the mean error of the ten trials. In the error versus iteration plots on the right, we only plot the mean error over the ten trials as the errors are highly similar for different values of α. We plot the bounds given by Theorem 1 if they decrease.

For Figures 6 and 7, we generate a single system defined by $A \in \mathbb{R}^{50000\times 100}$ and $\mathbf{b} = \mathbf{0} \in \mathbb{R}^{50000}$. We generate A with i.i.d. $\mathcal{N}(0,1)$ entries, then row-normalize it. In each figure, we generate $\mathbf{c}$ to have $\lfloor \beta m \rfloor$ nonzero entries each with value ten. Thus $\hat{\mathbf{b}} = \mathbf{c}$ has $\lfloor \beta m \rfloor$ nonzero entries each with value ten. We approximate $\sigma_{\alpha,q,\beta,\min}(A)$ by recording the minimum singular value encountered in each of the $B_k \setminus C$ submatrices encountered during the ten trials of 1000 iterations of sQRK.

In Figure 6, we plot the empirical convergence of ten trials of 1000 iterations of sQRK with $\beta = 10^{-3}$, $\alpha = 1$ (top), $\alpha = 0.5$ (middle), and $\alpha = 0.15$ (bottom), and $q = 0.5, 0.7$, and 0.9. We plot the empirical convergence of ten independent trials of 1000 iterations of sQRK with respect to wall clock time on the left (brighter color of cloud indicates more trial errors were below) and the mean error for each α. We plot the mean error for each α with respect to iteration k on the right. We plot the bounds given by Theorem 1 that were decreasing.

In Figure 6, we plot the empirical convergence of ten trials of 1000 iterations of sQRK with $\beta = 10^{-4}$, $\alpha = 1$ (top), $\alpha = 0.5$ (middle), and $\alpha = 0.15$ (bottom), and $q = 0.5, 0.7$, and 0.9. We plot the empirical convergence of ten independent trials of 1000 iterations of sQRK with respect to wall clock time on the left (brighter color of cloud indicates more trial errors were below) and the mean error for each α. We plot the mean error for each α with respect to iteration k on the right. We plot the bounds given by Theorem 1 that were decreasing.

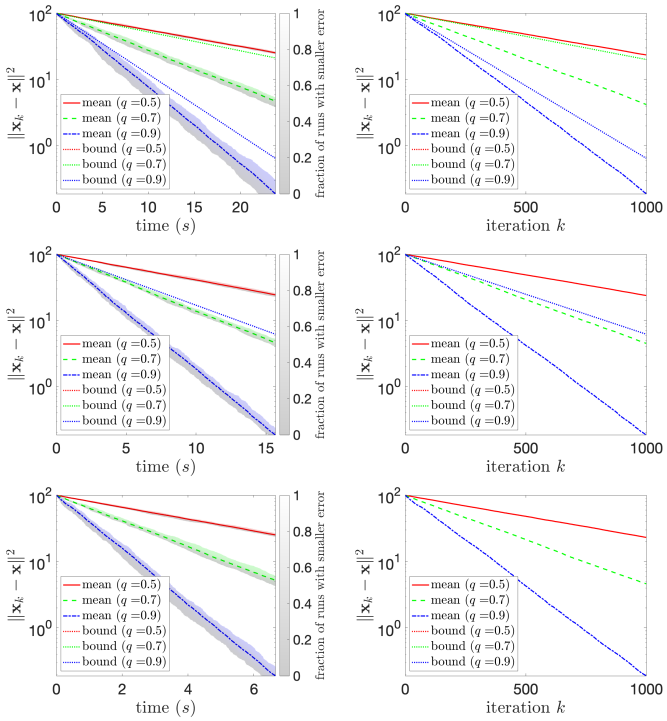

Fig. 6. Empirical convergence with respect to time (left) and iterations (right) for ten trials of sQRK with $\alpha = 1$ (top), $\alpha = 0.5$ (middle), and $\alpha = 0.15$ (bottom) and $q = 0.5, 0.7$, and 0.9 on system defined by row-normalized $A \in \mathbb{R}^{50000\times 100}$ and $\hat{\mathbf{b}} \in \mathbb{R}^{50000}$ with $\lfloor \beta m \rfloor$ corrupted entries where $\beta = 10^{-3}$. We additionally plot the bounds provided by Theorem 1 in dotted lines if they decrease.

Algorithm 2 Small Sample Quantile RK

procedure SSQRK($\mathbf{A}, \hat{\mathbf{b}}$, q, λ, N)
  $\mathbf{x}_1 = \mathbf{0}$
  for $k = 1, \ldots, N$ **do**
    Sample $\tau_k \subset [m]$ uniformly such that $|\tau_k| = \lceil \lambda \rceil$
    $\gamma_k = q$-quant $\left(\{ |\langle \mathbf{a}_j, \mathbf{x}_{k-1} \rangle - \hat{b}_j| \}_{j \in \tau_k} \right)$
    $i = \{ i \in \tau_k : |\langle \mathbf{a}_i, \mathbf{x}_{k-1} \rangle - \hat{b}_i| = \gamma_k \}$
    $\mathbf{x}_{k+1} = \mathbf{x}_k + (\hat{b}_i - \langle \mathbf{a}_i, \mathbf{x}_k \rangle) \mathbf{a}_i^T$
  end for return $\mathbf{x}_N$
end procedure

D. Small Samples

The previous section considers subsets of the size on the order of $O(m)$. While it makes the quantile learning stage α^{-1} times faster, it still presents significant computation overhead at each step compared to the classical RK method (which steps scale linearly with n and do not depend on $m \gg n$). Another approach is to approximate the quantile value from a considerably smaller, $O(1)$-size random sample $\lambda \ll m$ as outlined in Algorithm 2 below.

We first note that this approach cannot work with completely arbitrary corruption. If the sample size is less than the number of corruptions, there is a nonzero probability that all the equations in the sample are corrupted and arbitrarily far from the true solution; thus, any choice of the next equation can "undo" the earlier approach towards the solution. However, in many applications, it is natural to assume that some predetermined constant C bounds the maximal size of the corruption.

In such cases, the following heuristic can help quantify the convergence of the method. Let

$$\gamma := q'\text{-quant}\left(\{ |\langle \mathbf{a}_j, \mathbf{x}_{k-1} \rangle - \hat{b}_j| \}_{j \in [m]} \right),$$

be the q'-quantile of the full residual. Consider the following events:

- $\mathscr{E}_1$ = selected equation i is corrupted and has residual that is larger than γ
- $\mathscr{E}_2$ = selected equation i is corrupted and has residual that is smaller than γ
- $\mathscr{E}_3$ = selected equation i is uncorrupted.

For simplicity, let us consider the case of $q = 1/2$. We can

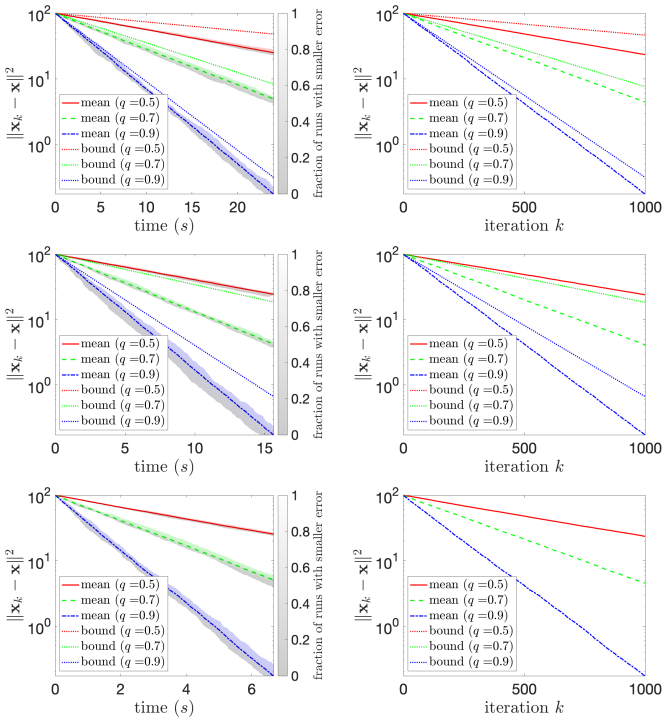

Fig. 7. Empirical convergence with respect to time (left) and iterations (right) for ten trials of sQRK with $\alpha = 1$ (top), $\alpha = 0.5$ (middle), and $\alpha = 0.15$ (bottom) and $q = 0.5, 0.7$, and 0.9 on system defined by row-normalized $A \in \mathbb{R}^{50000\times 100}$ and $\hat{\mathbf{b}} \in \mathbb{R}^{50000}$ with $\lfloor \beta m \rfloor$ corrupted entries where $\beta = 10^{-4}$. We additionally plot the bounds provided by Theorem 1 in dotted lines if they decrease.

decompose the expected error conditioned on these events:

$$\begin{aligned}\mathbb{E}\|\mathbf{x}_{k+1} - \mathbf{x}\|^2 = \mathbb{P}(i \in \mathscr{E}_1)\mathbb{E}_{i\in\mathscr{E}_1}\|\mathbf{x}_{k+1} - \mathbf{x}\|^2 \\ + \mathbb{P}(i \in \mathscr{E}_2)\mathbb{E}_{i\in\mathscr{E}_2}\|\mathbf{x}_{k+1} - \mathbf{x}\|^2 \\ + \mathbb{P}(i \in \mathscr{E}_3)\mathbb{E}_{i\in\mathscr{E}_3}\|\mathbf{x}_{k+1} - \mathbf{x}\|^2.\end{aligned}$$

As in the proof of Theorem 1, we expect that $\mathbb{P}(i \in \mathscr{E}_3)\mathbb{E}_{i\in\mathscr{E}_3}\|\mathbf{x}_{k+1} - \mathbf{x}\|^2 < r'\|\mathbf{x}_k - \mathbf{x}\|^2$ and that the other two terms are small enough. Indeed, the expected approximation error for event $\mathscr{E}_3$ depends on q. If $q \sim 0.5$, the bound on the error should be similar to that of RK. For larger q, we expect the contraction to be even stronger (i.e., r' smaller; see similar works on the Sampling Kaczmarz-Motzkin method [26], [27]). This suggests, as demonstrated in Figure 8, that taking q not too small and not too large is crucial for the successful performance of the method.

The first event has very small probability

$$\begin{aligned}\mathbb{P}(i \in \mathscr{E}_1) &\leq \mathbb{P}(\text{ chosen residual is larger than } \gamma\ ) \\ &\leq (1-q')^{\lambda/2},\end{aligned}$$

and $\|\mathbf{x}_{k+1} - \mathbf{x}_k\|$ is bounded by C. The expected approximation error in event $\mathscr{E}_2$ can be computed like in Lemma 2 with $q = q'$, and $\alpha = 1$. The main difficulty is estimating the probability that q-th residual in a random sample is (un)corrupted, which depends on the overall distribution of the corrupted residuals in the sample. If the corrupted equations tend to have larger residuals than uncorrupted ones, we will have $\mathbb{P}(i \in \mathscr{E}_2) < \beta$ and $\mathbb{P}(i \in \mathscr{E}_3) > 1 - \beta$. In Figure 8, we give empirical evidence of the convergence of the small sample Quantile RK method. We note that picking large q or too small λ results in spiking behaviour (right and middle figures), and picking too small q results in very slow convergence that can be easily overtaken by a rare event of facing a large corruption (left figure). However, $q \sim 0.5$ and $\lambda = 11 \ll 7500$ (that was the smallest sample size from Figures 2-7) demonstrate successful convergence as suggested in the discussion above.

VI. Conclusion and Future Directions

This paper considers variants of the Quantile Randomized Kaczmarz method for linear systems with corruptions which provide a computational advantage in the quantile estimation stage. We provide theoretical and empirical convergence guarantees for the case when the quantile is estimated from the sample of the size of a fraction of the total number of equations. We also consider empirically the case when the quantile is estimated from a very small sample of constant size (in particular, when it is smaller than the total number of corruptions in the system). Some interesting future directions of this work include providing theoretical analysis for the small sample size, particularly getting guidance for choosing the optimal sample size λ and quantile q. Further, extending the more efficient small sample quantile ideas to the nonlinear problems with corruptions is another valuable future direction.

Acknowledgements

We thank Jaime Pacheco (HMC) and Nestor Coria (HMC) for their assistance with the clarification of proofs and Max Collins (HMC) for the code, which helped to generate the figures in this work.

References

[1] W. Brendel, J. Rauber, and M. Bethge, "Decision-based adversarial attacks: Reliable attacks against black-box machine learning models," *arXiv preprint arXiv:1712.04248*, 2017.

[2] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial machine learning at scale," *arXiv preprint arXiv:1611.01236*, 2016.

[3] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," *arXiv preprint arXiv:1706.06083*, 2017.

[4] J. Moorman, T. Tu, D. Molitor, and D. Needell, "Randomized Kaczmarz with averaging," *BIT Numerical Mathematics*, vol. 61, no. 1, pp. 337–259, 2021.

[5] F. Bach, "Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 595–627, 2014.

[6] J. Haddock, D. Needell, E. Rebrova, and W. Swartworth, "Quantile-based iterative methods for corrupted systems of linear equations," *SIAM Journal on Matrix Analysis and Applications*, vol. 43, no. 2, pp. 605–637, 2022.

[7] S. Steinerberger, "Quantile-based random Kaczmarz for corrupted linear systems of equations," *Information and Inference: A Journal of the IMA*, vol. 12, no. 1, pp. 448–465, 2023.

[8] V. Shah, X. Wu, and S. Sanghavi, "Choosing the sample with lowest loss makes SGD robust," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (S. Chiappa and R. Calandra, eds.), vol. 108 of *Proceedings of Machine Learning Research*, pp. 2120–2130, PMLR, 26–28 Aug 2020.

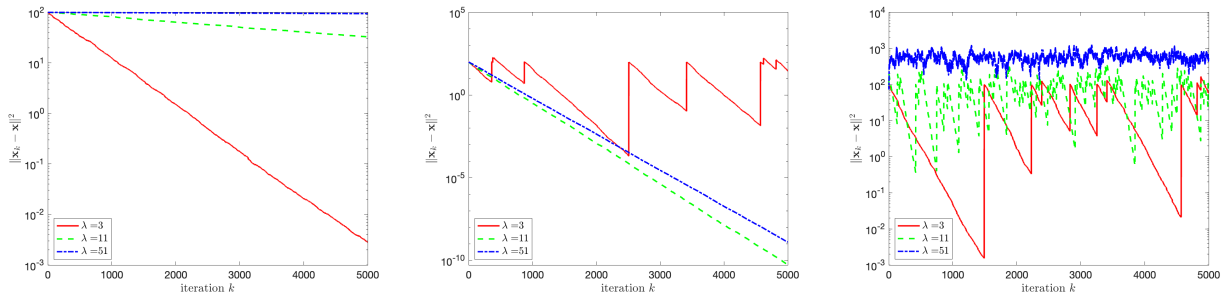

Fig. 8. We present the performance of Algorithm 2 using $\lambda = \{3, 11, 51\}$ when $m = 50,000$, $n = 100$, $\beta = 0.02$ for (left) $q = \frac{1}{\lambda}$, which corresponding to picking the smallest residual, (center) $q = \frac{1}{2}$, which corresponds to picking the median residual, and (right) $q = \frac{\lambda - 1}{\lambda}$, which corresponds to choosing the second largest residual.

[9] I. Diakonikolas, G. Kamath, D. Kane, J. Li, J. Steinhardt, and A. Stewart, "Sever: A robust meta-algorithm for stochastic optimization," in *International Conference on Machine Learning*, pp. 1596–1606, PMLR, 2019.

[10] P. Awasthi, M. F. Balcan, and P. M. Long, "The power of localization for efficiently learning linear separators with noise," in *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pp. 449–458, 2014.

[11] K. A. Lai, A. B. Rao, and S. Vempala, "Agnostic estimation of mean and covariance," in *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 665–674, IEEE, 2016.

[12] M. Charikar, J. Steinhardt, and G. Valiant, "Learning from untrusted data," in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 47–60, 2017.

[13] I. Diakonikolas, G. Kamath, D. Kane, J. Li, A. Moitra, and A. Stewart, "Robust estimators in high-dimensions without the computational intractability," *SIAM Journal on Computing*, vol. 48, no. 2, pp. 742–864, 2019.

[14] P. J. Rousseeuw, "Least median of squares regression," *Journal of the American statistical association*, vol. 79, no. 388, pp. 871–880, 1984.

[15] J. Á. Víšek, "The least trimmed squares. Part I: Consistency," *Kybernetika*, vol. 42, no. 1, pp. 1–36, 2006.

[16] K. Bhatia, P. Jain, P. Kamalaruban, and P. Kar, "Consistent robust regression," in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.

[17] A. Prasad, A. S. Suggala, S. Balakrishnan, P. Ravikumar, *et al.*, "Robust estimation via robust gradient estimation," *Journal of the Royal Statistical Society Series B*, vol. 82, no. 3, pp. 601–627, 2020.

[18] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: Byzantine tolerant gradient descent," *Advances in neural information processing systems*, vol. 30, 2017.

[19] D. Alistarh, Z. Allen-Zhu, and J. Li, "Byzantine stochastic gradient descent," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[20] S. Kaczmarz, "Angenäherte auflösung von systemen linearer gleichungen," *Bull. Int. Acad. Polon. Sci. Lett. Ser. A*, pp. 335–357, 1937.

[21] T. Strohmer and R. Vershynin, "A randomized Kaczmarz algorithm with exponential convergence," *Journal of Fourier Analysis and Applications*, vol. 15, no. 2, p. 262, 2009.

[22] J. Haddock and D. Needell, "Randomized projection methods for linear systems with arbitrarily large sparse corruptions," *SIAM Journal on Scientific Computing*, vol. 41, no. 5, pp. S19–S36, 2019.

[23] B. Jarman and D. Needell, "QuantileRK: Solving large-scale linear systems with corrupted, noisy data," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, pp. 1312–1316, IEEE, 2021.

[24] L. Zhang, H. Wang, and H. Zhang, "Quantile-based random sparse Kaczmarz for corrupted, noisy linear inverse systems," *arXiv preprint arXiv:2206.07356*, 2022.

[25] L. Cheng, B. Jarman, D. Needell, and E. Rebrova, "On block accelerations of quantile randomized Kaczmarz for corrupted systems of linear equations," *arXiv e-prints*, pp. arXiv–2206, 2022.

[26] J. Haddock and A. Ma, "Greed works: An improved analysis of sampling Kaczmarz–Motzkin," *SIAM Journal on Mathematics of Data Science*, vol. 3, no. 1, pp. 342–368, 2021.

[27] J. A. De Loera, J. Haddock, and D. Needell, "A sampling Kaczmarz–Motzkin algorithm for linear feasibility," *SIAM Journal on Scientific Computing*, vol. 39, no. 5, pp. S66–S87, 2017.