including deterministic ones like the Discrete Cosine Transform (DCT). Additionally, we delve into the scenario of sample abundance, a common situation in one-bit sensing, where we reformulate the relaxed optimization problem (??) into a linear feasibility problem. Strong linear feasibility solvers only provide a solution that satisfies all inequalities. Nevertheless, we require a solution that not only satisfies the given inequalities but also fulfills the constraints of the original problem (??). Using our theorems, we aim to address the question of how many samples are necessary for a linear inequality system solver to obtain a solution within the ball with a specific radius, around the original signal. Moreover, we endeavor to harness the advantages of randomized algorithms for one-bit sensing, a novel endeavor where we explore their theoretical guarantees and convergence analysis in this context for the first time. It is important to note that all the theoretical guarantees derived in this paper are uniform reconstruction guarantees. This means that they hold true for all desired signals in the given space, ensuring consistent and reliable reconstruction across the entire signal set.

B. Contributions of the Paper

This paper principally contributes to the following areas:

1) RKA-based recovery for the dithered one-bit sensing. In this paper, we consider the deployment of one-bit sampling with time-varying thresholds,leading to an increased sample size and a highly overdetermined system as a result. The proposed One-bit aided Randomized Kaczmarz Algorithm, which we refer to as ORKA, can find the desired signal $\mathbf{X}_{\star}$ in (??) by (i) generating abundant one-bit measurements, in order to define a large scale overdetermined system where a finite volume feasible set is created for (??), and (ii) solving this obtained linear feasibility problem by leveraging one of the efficient solver families of overdetermined systems, Kaczmarz algorithms. In [?], we showed that the sheer number of measurements acquired in one bit sampling facilitates recovering the signal of interest in a less costly manner by making costly constraints such as semidefiniteness and rank redundant. Then, a simple RKA was utilized to solve the obtained linear feasibility problem. This idea is generalized in this paper to (??) where we generate the abundant samples and eventually introduce a one-bit linear feasibility region named the one-bit polyhedron. In other words, by using this technique, we make (??) into a large-scale overdetermined system, an ideal application setting for Kaczmarz algorithms. To solve our highly overdetermined system,

variance σ^2. The sub-Gaussian norm of a random variable X characterized by

$$\|X\|_{\psi_2} = \inf\left\{t > 0 : \mathbb{E}\left\{e^{X^2/t^2}\right\} \leq 2\right\}. \tag{2}$$

II. Projections On Convex Sets: Dealing With Costly Constraints

To tackle (??), many non-convex and local optimization algorithms have been developed over the years. Nevertheless, in recent decades, convex programming formulations via relaxation have come to the fore to approximate global solutions. Within the convex framework, various iterative methods have been proposed to tackle the problem with a Lagrangian formulation such as Uzawa's algorithm and the proximal forward-backward splitting method (PFBS) [?, ?, ?]. Moreover, to keep the problem solution inside the constraint set Ω_c, the orthogonal projection P_{Ω_c} is applied to solutions at each iteration. The Lagrangian for (??) is written as $\mathcal{L}(\mathbf{X}, \boldsymbol{\lambda}) = f(\mathbf{X}) + \langle \boldsymbol{\lambda}, \mathbf{y} - \mathcal{A}(\mathbf{X})\rangle$, where $\boldsymbol{\lambda} \in \mathbb{R}^n$ [?]. Uzawa's algorithm aims to find a saddle point $(\mathbf{X}_\star, \boldsymbol{\lambda}_\star)$, where $\sup_{\boldsymbol{\lambda}} \inf_{\mathbf{X}} \mathcal{L}(\mathbf{X}, \boldsymbol{\lambda}) = \inf_{\mathbf{X}} \sup_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{X}, \boldsymbol{\lambda})$, with the iterative procedure:

$$\begin{cases} \mathbf{X}_k = P_{\Omega_c}\left(\mathcal{A}^\star\left(\boldsymbol{\lambda}_{k-1}\right)\right), \\ \boldsymbol{\lambda}_k = \boldsymbol{\lambda}_{k-1} + \alpha_k\left(\mathbf{y} - \mathcal{A}\left(\mathbf{X}_k\right)\right), \end{cases} \tag{3}$$

where α_k is the step size and $\mathcal{A}^\star$ is the adjoint of $\mathcal{A}$. Since every linear equation can be reformulated in standard form, we recast $\mathcal{A}(\mathbf{X}) = \mathbf{y}$ as $\mathbf{A}\mathbf{x} = \mathbf{y}$, where $\mathbf{A} \in \mathbb{C}^{n \times n_1 n_2}$ is a matrix version of the operator $\mathcal{A}$, and $\mathbf{x} = \mathrm{vec}(\mathbf{X})$ [?]. The optimization problem (??) is equivalently given by [?, ?]

$$\underset{\mathbf{X} \in \Omega_c}{\text{minimize}} \quad g(\mathbf{X}) = \frac{1}{2}\left\|\mathbf{y} - \mathbf{A}\,\mathrm{vec}(\mathbf{X})\right\|_2^2 + \lambda f(\mathbf{X}) \tag{4}$$

To solve this problem, instead of using proximal methods, a projected gradient method such as Nesterov iterative approach may be utilized, i.e., $\mathbf{X}_k = P_{\Omega_c}\left(\mathbf{X}_{k-1} - \alpha_k \nabla g\left(\mathbf{X}_{k-1}\right)\right)$.

Famous examples for P_{Ω_c} are the singular value thresholding operator (SVT) and the semi-definite orthogonal projector, respectively. SVT is useful when $f(\mathbf{X}) = \|\mathbf{X}\|_\star$, mathematically defined as: $P_{\Omega_c} = \mathbf{U}\,\mathrm{Diag}\left(\{\sigma_k\}^+\right)\mathbf{V}^T$ [?], where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices from singular value decomposition (SVD), and $\{\sigma_k\}$ are the singular values. The approximate solution should be projected onto a feasible convex set at each iteration via recovering all singular values and eigenvalues and comparing their smaller elements with a threshold, which is quite expensive [?].

where, at each iteration i, the index j is drawn from the set $\mathcal{J}$ independently at random following the distribution $\Pr\{j=k\}=\frac{\|\mathbf{c}_k\|_2^2}{\|\mathbf{C}\|_{\mathrm{F}}^2}$. Assuming that the linear system is consistent with nonempty feasible set $\mathcal{P}_{\mathbf{x}}$ created by the intersection of hyperplanes around the desired point $\mathbf{x}_\star$, RKA converges linearly in expectation to the solution $\widehat{\mathbf{x}}\in\mathcal{P}_{\mathbf{x}}$ [?, ?]:

$$\mathbb{E}\left\{\hbar\left(\mathbf{x}_i,\widehat{\mathbf{x}}\right)\right\}\leq\left(1-q_{\mathrm{RKA}}\right)^i\ \hbar\left(\mathbf{x}_0,\widehat{\mathbf{x}}\right), \tag{14}$$

where $\hbar\left(\mathbf{x}_i,\widehat{\mathbf{x}}\right)=\|\mathbf{x}_i-\widehat{\mathbf{x}}\|_2^2$, is the euclidean distance between two points in the space, i is the number of required iterations for RKA, and $q_{\mathrm{RKA}}\in(0,1)$ is given by $q_{\mathrm{RKA}}=\frac{1}{\kappa^2(\mathbf{C})}$, with $\kappa\left(\mathbf{C}\right)=\|\mathbf{C}\|_{\mathrm{F}}\|\mathbf{C}^{\dagger}\|_2$ denoting the scaled condition number of a matrix $\mathbf{C}$.

C. Sampling Kaczmarz Motzkin Method

The SKM combines the ideas of both the RKA and the Motzkin method. The generalized convergence analysis of the SKM with sketch matrix which has been formulated based on the convergence analysis of RKA, and sampling Motzkin method for solving linear feasibility problem has been comprehensively explored in [?]. The central contribution of SKM lies in its innovative way of projection plane selection. The hyperplane selection is done as follows: At iteration i, the SKM algorithm selects a collection of γ (denoted by the set $\mathcal{T}_i$) rows, uniformly at random out of m rows of the constraint matrix $\mathbf{C}$. Then, out of these γ rows, the row $j^\star$ with the maximum positive residual is selected; i.e.

$$j^\star=\operatorname{argmax}\ \left\{\left(b_j-\mathbf{c}_j\mathbf{x}_i\right)^+\right\},\ j\in\mathcal{T}_i. \tag{15}$$

Finally, the solution is updated as [?, ?] $\mathbf{x}_{i+1}=\mathbf{x}_i+\lambda_i\frac{\beta_i}{\|\mathbf{c}_{j^\star}\|_2^2}\mathbf{c}_{j^\star}^{\mathrm{H}}$, where λ_i is a relaxation parameter which for consistent systems must satisfy $0\leq\lim_{i\to\infty}\inf\lambda_i\leq\lim_{i\to\infty}\sup\lambda_i<2$ [?], to ensure convergence.

D. Preconditioned SKM

Assume $\mathcal{P}_{\mathbf{x}}$ as the space created by the intersection of hyperplanes in a linear feasibility problem. According to the convergence rate of RKA, reducing the value of the scaled condition number $\kappa\left(\mathbf{C}\right)$ or equivalently increasing the value of q_{RKA} in (??) leads to an accelerated convergence of the RKA to $\mathcal{P}_{\mathbf{x}}$. As a result, the upper bound of the recovery error $\mathbb{E}\left\{\|\mathbf{x}_i-\widehat{\mathbf{x}}\|_2^2\right\}$ with $\widehat{\mathbf{x}}\in\mathcal{P}_{\mathbf{x}}$, decreases, as well. From another perspective, this property (lower value of $\kappa\left(\mathbf{C}\right)$) provides the RKA or its variant, SKM, enjoying a lower number of

Algorithm 1 PrSKM Algorithm

Input: A matrix $\mathbf{C}$ and the measurement vector $\mathbf{b}$.

Output: A solution $\bar{\mathbf{x}}$ in a nonempty feasible set of $\mathbf{C}\mathbf{x} \succeq \mathbf{b}$.

1: $[\mathbf{Q}_c, \mathbf{R}_c] = \mathrm{QR}(\mathbf{C})$ ▷ $\mathrm{QR}(\cdot)$ computes the QR-decomposition of a matrix $\mathbf{C}$.
2: $\mathbf{Q}_c\mathbf{z} \succeq \mathbf{b}$ ▷ The new problem that we should solve respect to $\mathbf{z}$.
3: Choose a sample set of γ constraints (denoted as $\mathcal{T}_i$) uniformly at random from the rows of $\mathbf{Q}_c$.
4: $j^{\star} \leftarrow \operatorname{argmax} \left\{(b_j - \mathbf{q}_j\mathbf{z}_i)^{+}\right\},\ j \in \mathcal{T}_i$ ▷ $\mathbf{q}_j$ is the j-th row of $\mathbf{Q}_c$.
5: $\mathbf{z}_{i+1} \leftarrow \mathbf{z}_i + \lambda_i \frac{(b_{j^\star} - \mathbf{q}_{j^\star}\mathbf{z}_i)^{+}}{\|\mathbf{q}_{j^\star}\|_2^2} \mathbf{q}_{j^\star}^{\mathrm{H}}$.
6: Repeat steps (4)-(6) until convergence and obtain $\bar{\mathbf{z}}$.
7: $\mathbf{R}_c\bar{\mathbf{x}} = \bar{\mathbf{z}}$ ▷ Obtain $\bar{\mathbf{x}}$ via the Gaussian elimination algorithm.
8: return $\bar{\mathbf{x}}$

upper triangular matrix, leading to[1] $\mathbf{Q}_c = \mathbf{C}\mathbf{R}_c^{-1}$. Thus, based on Theorem ??, a good choice for the preconditioner is $\mathbf{M} = \mathbf{R}_c^{-1}$. To find a point $\bar{\mathbf{z}}$ in a nonempty feasible set, the SKM method described in Section ?? is employed. Finally, one may approach the solution of the original linear feasibility $\mathbf{C}\mathbf{x} \succeq \mathbf{b}$ by computing $\bar{\mathbf{x}} = \mathbf{R}_c^{-1}\bar{\mathbf{z}}$. We refer to this method Preconditioned SKM (PrSKM) which is summarized in Algorithm ??. As shown in Theorem ??, the scaled condition number of the matrix $\mathbf{Q}_c$ is $\kappa(\mathbf{Q}_c) = \sqrt{n}$. Therefore, step 5 of Algorithm ?? converges linearly in expectation to a nonempty feasible set of $\mathbf{Q}_c\mathbf{z} \succeq \mathbf{b}$ as follows,

$$\mathbb{E}\left\{\hbar\left(\mathbf{x}_i, \widehat{\mathbf{x}}\right)\right\} \leq \left(1 - \frac{1}{n}\right)^i \hbar\left(\mathbf{x}_0, \widehat{\mathbf{x}}\right). \tag{18}$$

Note that to run the Algorithm ?? for solving the one-bit polyhedron (??), it is enough to set $\mathbf{C} = \mathbf{P}$ and $\mathbf{b} = \mathrm{vec}(\mathbf{R}) \odot \mathrm{vec}(\mathbf{\Gamma})$. Later, in Theorem ??, we will show that it is only required to set $\mathbf{C} = \mathbf{A}$ in Algorithm ?? which is more computationally and storagelly efficient compared to that of setting $\mathbf{C} = \mathbf{P}$.

[1] For a matrix $\mathbf{C} \in \mathbb{R}^{m\times n}$, since we have assumed $m > n$, we can obtain a unitary matrix $\mathbf{Q}_c \in \mathbb{R}^{m\times n}$ and an upper triangular matrix $\mathbf{R}_c \in \mathbb{R}^{n\times n}$ such that the QR-decomposition holds; i.e. $\mathbf{C} = \mathbf{Q}_c\mathbf{R}_c$.

of $\mathbf{C}\mathbf{R}_s^{-1}\mathbf{z} \succeq \mathbf{b}$ as follows:

$$\mathbb{E}\left\{\hbar\left(\mathbf{x}_i, \widehat{\mathbf{x}}\right)\right\} \leq \left(1-\frac{1}{3n}\right)^i \hbar\left(\mathbf{x}_0, \widehat{\mathbf{x}}\right). \tag{19}$$

The proof of Theorem ?? is presented in Appendix ??.

F. Block SKM

In two previous sections, we have proposed PrSKM and storage-friendly PrSKM algorithms in order to solve the one-bit polyhedron (??) in an asymptotic sample size scenario. It is worth noting that both methods are row-based approaches, where at each iteration, the row index of the matrix $\mathbf{P}$ is chosen independently at random. However, the matrix $\mathbf{P}$ in (??) has a block structure as formulated in (??). This fact motivates us to investigate the block-based RKA methods to find the desired signal in the one-bit polyhedron $\mathcal{P}_{\mathbf{x}}$ for further efficiency enhancement. Our proposed algorithm for block systems, Block SKM, is motivated by (i) random selection of one block at each iteration, (ii) choosing a subset of rows using the idea of Motzkin sampling, and (iii) updating the solution using the randomized block Kaczmarz method [?, ?], which takes advantage of the efficient matrix-vector multiplication, thus giving the method a significant reduction in computational cost [?]. Algorithm ?? shows the implementation of our proposed Block SKM method to solve the linear feasibility $\mathbf{B}\mathbf{x} \succeq \mathbf{b}$ with $\mathbf{B} = \left[\begin{array}{c|c|c} \mathbf{B}_1^{\top} & \cdots & \mathbf{B}_m^{\top} \end{array}\right]^{\top}$ and $\mathbf{b} = \left[\begin{array}{c|c|c} \mathbf{b}_1^{\top} & \cdots & \mathbf{b}_m^{\top} \end{array}\right]^{\top}$ where $\mathbf{B}_j \in \mathbb{R}^{n\times d}$ and $\mathbf{b}_j \in \mathbb{R}^n$ for all $j \in \{1, \cdots, m\}$. Note that in step 4 of Algorithm ??, the reason behind choosing $k' < d$ is due to the computation of $\left(\mathbf{B}'_j\mathbf{B}'^{\top}_j\right)^{-1}$ in the next step (step 5). For $k' > d$, the matrix $\mathbf{B}'_j\mathbf{B}'^{\top}_j$ is rank-deficient and its inverse is not available. The Block SKM algorithm can be considered to be a special case of the more general sketch-and-project method with a sparse block sketch matrix as defined in [?][?]. The convergence of Block SKM algorithm with the sparse Gaussian sketch in the case of $k' = 1$ is presented in the following lemma:

Lemma 1. The Block SKM algorithm with the sparse Gaussian sketch in the case of $k' = 1$, converges linearly in expectation to the nonempty feasible set of $\mathbf{B}\mathbf{x} \succeq \mathbf{b}$, $\mathbf{B} \in \mathbb{R}^{mn\times d}$, as follows,

$$\mathbb{E}\left\{\hbar\left(\mathbf{x}_i, \widehat{\mathbf{x}}\right)\right\} \leq \left(1-\frac{c\sigma^2_{\min}(\hat{\mathbf{B}})\log n}{\|\hat{\mathbf{B}}\|_{\mathrm{F}}^2}\right)^K \hbar\left(\mathbf{x}_0, \widehat{\mathbf{x}}\right), \tag{20}$$

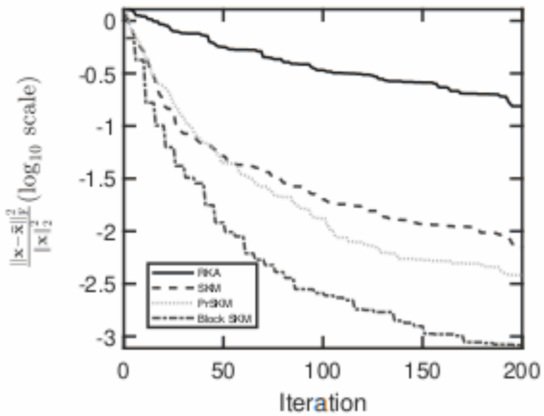


Figure 1. Comparing the recovery performance of the two proposed Kaczmarz algorithms, namely the PrSKM and the Block SKM, with that of SKM and RKA for a linear inequality system.

Also, the desired signal $\mathbf{x}$ is generated as $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{10})$. All time-varying sampling threshold sequences are generated according to $\left\{\boldsymbol{\tau}^{(\ell)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{10})\right\}_{\ell=1}^{m}$. The performance of the RKA, SKM, PrSKM, and Block SKM is illustrated in Fig. ??. Similar to the linear feasibility of equalities, it can be seen that the Block SKM has a better accuracy in recovering the desired signal $\mathbf{x}$ in the one-bit polyhedron (??) compared to the other three approaches. The NMSE results in Fig. ?? are averaged over 1000 experiments.

IV. Probabilistic Effect of Sample Abundance In One-Bit Sensing

In Section ??, we will introduce the concept of FVP and subsequently obtain the required number of one-bit samples m' to accurately capture the solution in the case of sample abundance, one-bit CS, and one-bit low-rank matrix recovery. In Section ??, we will provide the convergence of ORKA based on the theoretical results obtained in Section ??.

A. Finite Volume Property

Define the distance between the original signal $\mathbf{x}_\star$ and the j-th hyperplane presented in (??) as

$$d_j^{(\ell)}\left(\mathbf{x}_\star, \tau_j^{(\ell)}\right) = \left|r_j^{(\ell)}\mathbf{a}_j\mathbf{x}_\star - r_j^{(\ell)}\tau_j^{(\ell)}\right|, \quad j \in [n], \quad \ell \in [m], \tag{21}$$

It is essential to clarify that in our analysis, we adopt the worst-case scenario for the distance between the desired point and the solutions. This approach considers the solution to lie precisely on the hyperplane. However, it is crucial to note that in reality, the solution to a

$$\mathbf{P}\mathbf{x} + \boldsymbol{\nu} \succeq \operatorname{vec}(\mathbf{R}) \odot \operatorname{vec}(\boldsymbol{\Gamma}), \tag{40}$$

where $\mathbf{P}$ is defined in (??) and $\boldsymbol{\nu} = \tilde{\boldsymbol{\Omega}}\mathbf{z}$ is the noise of our system with $\tilde{\boldsymbol{\Omega}}$ defined in (??). For instance, assuming a zero-mean Gaussian noise vector $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_z)$ with the covariance matrix $\boldsymbol{\Sigma}_z$, the distribution of $\boldsymbol{\nu}$ will be $\mathcal{N}\left(\mathbf{0}, \tilde{\boldsymbol{\Omega}}\boldsymbol{\Sigma}_z\tilde{\boldsymbol{\Omega}}^{\mathrm{H}}\right)$.

The robustness of the RKA against noise has been demonstrated in [?] and [?]. Furthermore, the authors of [?] specifically explored the performance of the RKA in the presence of Gaussian and Poisson noise, highlighting its robustness even when dealing with Poisson noisy measurements. In our discussion, in Section ?? we will explore how the inconsistency of a linear system in a noisy scenario manifests itself in the recovery error of the RKA. Next, in Section ?? we will propose a novel algorithm to have a robust recovery performance in the presence of impulsive noise.

A. Robustness of ORKA Against Noise

Given a linear system of equations $\mathbf{U}\mathbf{x} = \mathbf{b}$ that is highly over-determined and subject to a noise vector $\mathbf{n} = [n_j]$ resulting in a corrupted system of equations $\mathbf{U}\mathbf{x} \approx \mathbf{b} + \mathbf{n}$. The convergence rate of the noisy RKA was comprehensively discussed in [?, Theorem 2.1] for the case of $\mathbf{U}\mathbf{x} \approx \mathbf{b} + \mathbf{n}$. The primary contrast between the convergence rates of RKA and noisy RKA, as demonstrated in [?, Theorem 2.1], lies in the second term of convergence rate $\kappa^2 \max_j \frac{n_j^2}{\|\mathbf{u}_j\|_2^2}$. This term indicates the degree to which the error in the corrupted system $\mathbf{U}\mathbf{x} \approx \mathbf{b} + \mathbf{n}$ deviates from the main solution.

Drawing inspiration from the convergence rate of the noisy RKA, we can similarly derive the convergence rate of noisy RKA in the case of noisy linear system of inequalities $\mathbf{C}\mathbf{x}+\mathbf{n} \succeq \mathbf{b}$ using the following proposition:

Proposition 2. Let $\mathbf{C} \in \mathbb{R}^{m\times n}$ have full column rank and assume $\widehat{\mathbf{x}}$ is the solution of the noisy linear feasibility problem $\mathbf{C}\mathbf{x} + \mathbf{n} \succeq \mathbf{b}$. Let $\bar{\mathbf{x}}_i$ be the i-th iterate of the noisy RKA run with $\mathbf{C}\mathbf{x} \succeq \mathbf{b}$, and let n_j denote the j-th element of $\mathbf{n}$, respectively. Then we have

$$\mathbb{E}\left\{\hbar\left(\bar{\mathbf{x}}_i, \widehat{\mathbf{x}}\right)\right\} \leq \left(1 - \frac{1}{\kappa^2(\mathbf{C})}\right)^i \hbar\left(\mathbf{x}_0, \widehat{\mathbf{x}}\right) + \kappa^2 \max_j \gamma_j, \tag{41}$$

where $\gamma_j = \frac{\left((n_j)^+\right)^2}{\|\mathbf{c}_j\|_2^2}$.

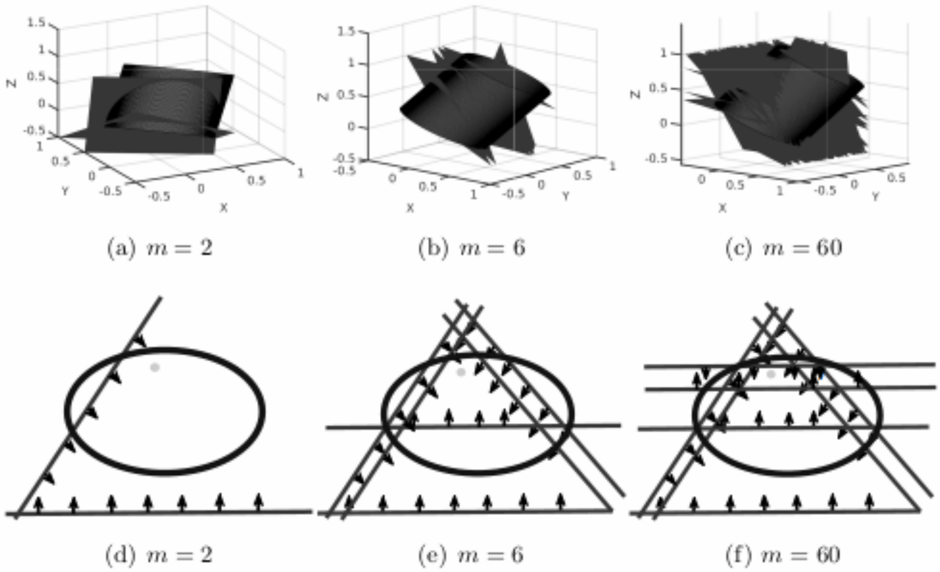


Figure 2. Shrinkage of the one-bit polyhedron (??) in blue, ultimately placed within the unit ball of the nuclear norm $\|\mathbf{X}\|_\star \leq 1$ shown with black cylindrical region and its red contours, when the number constraints (samples) grows large. The arrows point to the half-space associated with each inequality constraint. The evolution of the feasible regime is depicted with increasing samples in three cases: (a) and (d) small sample-size regime, constraints not forming a finite-value polyhedron; (b) and (e) medium sample-size regime, constraints forming a finite-volume polyhedron, parts of which are outside the cylinder; (c) and (f) large sample-size regime, constraints forming a finite-volume polyhedron inside the nuclear norm cylinder, making its constraint redundant. The original signal representing the signal to be recovered is shown by yellow.

rank matrix factorization with ORKA and the alternating minimization (AltMin) algorithm discussed in [?].

The one-bit polyhedron $\mathcal{P}^{(M)}$ associated with the low-rank matrix factorization approach is written as

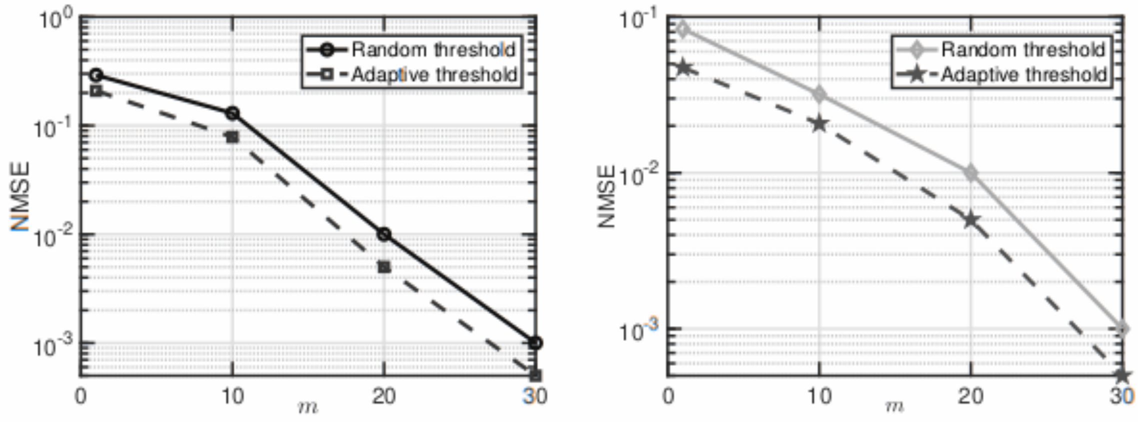
$$\mathcal{P}_0^{(M)} = \left\{\mathbf{L}, \mathbf{W} \mid r_j^{(\ell)} \operatorname{Tr}\left(\mathbf{A}_j^\top \mathbf{L}\mathbf{W}^\top\right) \geq r_j^{(\ell)} \tau_j^{(\ell)},\ j \in [n],\ \ell \in [m]\right\}. \tag{52}$$

To find the solution from $\mathcal{P}_0^{(M)}$, we use the idea of AltMin algorithm. We split $\mathcal{P}_0^{(M)}$ into two linear feasibility sub-problems with respect to $\mathbf{L}$ and $\mathbf{W}$, respectively. Specifically, with respect to $\mathbf{L}$ when $\mathbf{W}$ is fixed we have:

$$\mathcal{P}_{\mathbf{L}}^{(M)} = \left\{\mathbf{L} \mid r_j^{(\ell)} \operatorname{Tr}\left(\mathbf{W}^\top \mathbf{A}_j^\top \mathbf{L}\right) \geq r_j^{(\ell)} \tau_j^{(\ell)},\ j \in [n],\ \ell \in [m]\right\}, \tag{53}$$

and with respect to $\mathbf{W}$ when $\mathbf{L}$ is fixed we have:

$$\mathcal{P}_{\mathbf{W}}^{(M)} = \left\{\mathbf{W} \mid r_j^{(\ell)} \operatorname{Tr}\left(\mathbf{A}_j^\top \mathbf{L}\mathbf{W}^\top\right) \geq r_j^{(\ell)} \tau_j^{(\ell)},\ j \in [n],\ \ell \in [m]\right\}. \tag{54}$$



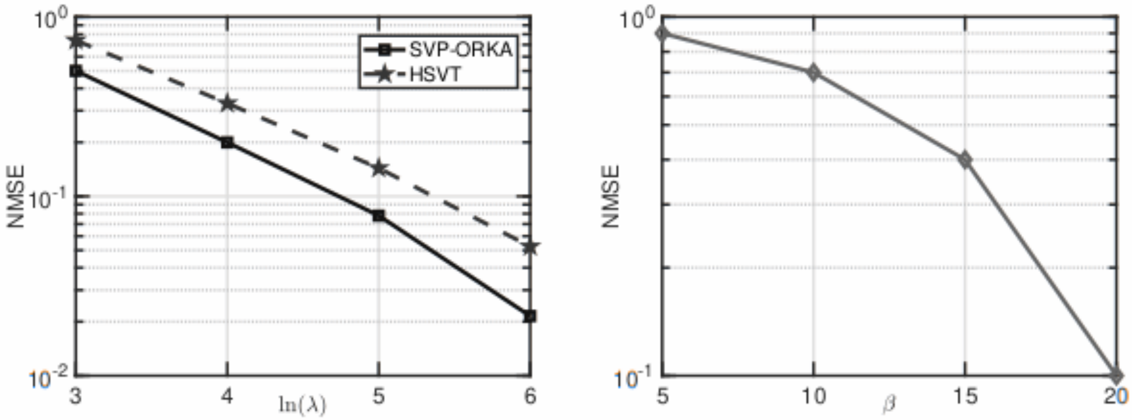
(a) (b)

Figure 3. Comparison between the recovery performance of Block SKM using random thresholds and adaptive thresholds in the sample abundance scenario for (a) one-bit low-rank matrix sensing, and (ii) one-bit CS.

of sparsity $s = 10$. The settings for time-varying sampling thresholds were considered to be the same as the one-bit low-rank matrix sensing case. Fig. ??(b) displays the recovery performance of Block SKM using random thresholds in comparison with adaptive thresholds. Consistent with previous observations, the utilization of adaptive thresholds improves the recovery performance.

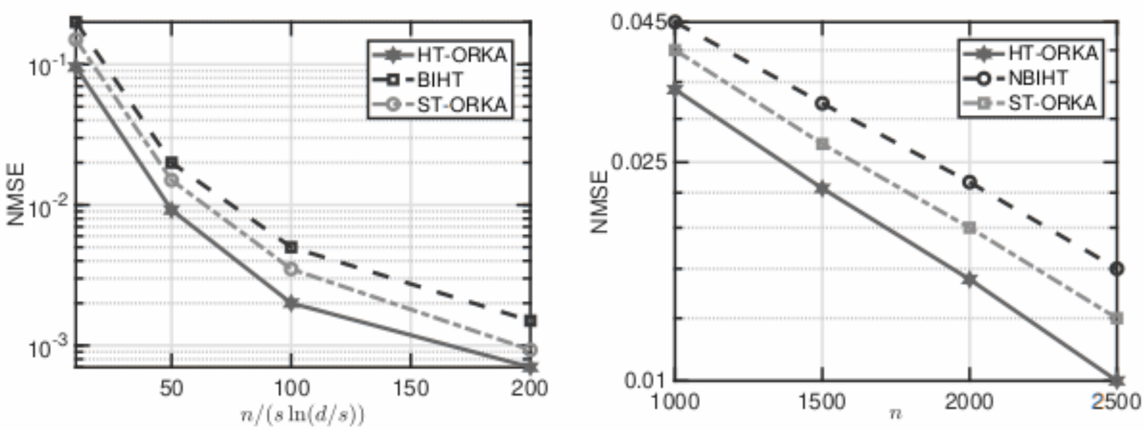
B. Sample Scarcity

Similar to sample abundance, herein we investigate the performance of our proposed methods for one-bit low-rank matrix sensing and one-bit CS when we have a limited number of samples. Note that in all experiments, the high-resolution measurements were contaminated by the additive Gaussian noise with the standard deviation 0.1 except the case related to low-rank matrix sensing by SVP-ORKA which was considered to be noiseless. One-bit low-rank matrix sensing. We generated a collection of sampling matrices $\{\mathbf{A}_j\}_{j=1}^{n}$, where each entry is independently sampled from a standard normal distribution. The desired matrix $\mathbf{X}_\star \in \mathbb{R}^{30\times 30}$ was generated with $\operatorname{rank}(\mathbf{X}_\star) = 2$. Define the oversampling factor as $\lambda = \frac{n}{n_1 r} = \frac{n}{60}$. In our experiments, we have set $\log(\lambda) \in \{3, 4, 5, 6\}$. The number of time-varying sampling threshold sequences was fixed at $m = 1$. The generation of time-varying sampling thresholds followed the same procedure as in the previous cases. Fig. ??(a) compares the recovery performance of SVP-ORKA with hard singular value thresholding (HSVT) algorithm [?] in the noiseless scenario. As can be observed, SVP-



(a) (b)

Figure 4. (a) Comparison between the recovery performance of SVP-ORKA and HSVT algorithm over different values of oversampling factor λ. (b) Recovery performance of Algorithm ?? (ORKA with low-rank matrix factorization) over different values of sampling factor β.



(a) (b)

Figure 5. Comparison between the recovery performance of HT-ORKA, ST-ORKA and (a) BIHT with random thresholds, and (b) NBIHT in ditherless scenario.

ORKA outperforms HSVT over different values of the oversampling factor. In another experimental setting aimed at assessing the performance of Algorithm ?? (ORKA with low-rank matrix factorization), we generated the desired matrix $\mathbf{X}_\star \in \mathbb{R}^{30\times 30}$ with $\operatorname{rank}(\mathbf{X}_\star) = 1$. The remaining parameter settings were identical to the previous example. Note that in this case we define the oversampling factor as $\beta = \frac{n}{n_1^2 r} = \frac{n}{900}$. In our simulations, we have set $\beta = \{5, 10, 15, 20\}$. As can be seen in Fig. ??(b), the recovery performance of Algorithm ?? enhances as the value of the oversampling factor β grows large.

One-bit CS. Each element of the sensing matrix $\mathbf{A} \in \mathbb{R}^{n\times 100}$ was independently drawn from a standard normal distribution. The desired signal $\mathbf{x}_\star \in \mathbb{R}^{100}$ was assumed to have a sparsity level of $s = 15$. Define the oversampling factor as $n/(s\log(d/s))$. In our experiments, we have set the oversampling factor to the values $\{10, 50, 100, 200\}$. The number of time-varying sampling threshold sequences was fixed at $m = 1$, and the generation of time-varying sampling thresholds followed the same procedure as in the previous cases. In Fig. ??(a), we compare the recovery performance of HT-ORKA, ST-ORKA, and BIHT algorithm with random time-varying sampling thresholds [?]. It is evident that HT-ORKA outperforms both ST-ORKA and BIHT in recovering the s-sparse signal in the one-bit CS problem. In the next example, we examine the effectiveness of our proposed algorithms, HT-ORKA and ST-ORKA, in recovering a sparse signal in a ditherless scenario. The sensing matrix $\mathbf{A} \in \mathbb{R}^{n\times 256}$ was generated in the same manner as in the previous example. The number of high-resolution samples was considered to be $n \in \{1000, 1500, 2000, 2500\}$. The desired signal $\mathbf{x}_\star \in \mathbb{R}^{256}$ was assumed to have a sparsity level of $s = 25$. In Fig. ??(b), we compare the recovery performance of HT-ORKA, ST-ORKA and the NBIHT algorithm [?]. Once again, similar to the previous example, HT-ORKA exhibits superior recovery performance compared to ST-ORKA and the NBIHT method.

X. Discussion

In this paper, we have established the theoretical guarantees for uniform perfect reconstruction in dithered one-bit sensing. Our approach involves transforming the one-bit signal reconstruction problem into a linear feasibility problem. We then introduced the FVP theorem to analyze the possibility of creating a finite volume formed by the hyperplanes around the original signal. The FVP theorem allows us to determine the minimum probability of achieving perfect reconstruction and the required number of samples to attain uniform perfect reconstruction. What sets this theorem apart from others, such as the random hyperplane tessellations theorem, is that it approaches one-bit sensing from the perspective of a linear feasibility problem. It investigates the number of samples needed to capture the original signal within the space of hyperplanes. An intriguing aspect of the FVP theorem is its capability to provide guarantees not only for uniform dithering and restricted sampling matrices considered in previous efforts. Leveraging this theorem, we were able to derive theoretical guarantees for deterministic matrices like DCT as well

The only restriction in the FVP theorem is that the distances between the original signals and the surrounding hyperplanes defined in (??) should be considered as sub-Gaussian. It remains an open problem for other distributions, which may be explored in future research.

This work represents a pioneering effort in the literature, as it explores the performance of randomized algorithms in one-bit sensing for the first time. We introduced two novel variations of the Kaczmarz algorithm, PrSKM and Block SKM, which served as the foundation for our proposed algorithm, ORKA. In our investigation of ORKA, we analyzed its upper recovery bound and demonstrated that it decays concerning the number of measurements. Specifically, for both compressed sensing and low-rank matrix recovery, the decay rate is $\mathcal{O}\left(m^{-\frac{2}{3}}\right)$. These findings contribute valuable insights into the potential of randomized algorithms in one-bit sensing applications. To the best of our knowledge, we are the first to derive the convergence rate of RKA for a noisy linear inequality system. This novel finding highlights the robustness of the algorithm, even in the presence of noise.

We have introduced an improved update process for designing the thresholding process. Unlike the sigma-delta thresholding design discussed in references [?, ?], our approach does not require updating the one-bit data, eliminating the need for feedback in the sampling scheme. Through numerical demonstrations, we showed that this adaptive thresholding process enhances signal reconstruction performance from one-bit data. Through numerical experiments, we experimentally demonstrated that the proposed algorithm exhibits superior reconstruction performance in one-bit CS compared to the NBIHT for the ditherless scenario and BIHT adapted with dithering for the dithered scenario. Furthermore, our results show that the proposed SVP-ORKA outperforms the HSVT algorithm in terms of recovery accuracy.

In addition to the common situation of sample abundance in dithered one-bit quantization, we also address scenarios with sample restrictions. For these scenarios, we further develop the proposed randomized algorithms into storage-friendly approaches, such as random sketching and low-rank matrix factorization. This extension allows for efficient handling of limited samples, broadening the applicability of the algorithms to a wider range of practical settings. The convergence rate of ORKA with low-rank matrix factorization remains an open problem. This is because it employs the idea of cyclic algorithms or AltMin, whose convergence is still an ongoing topic of research in the literature [?].

general Hoeffding's inequality [?, Theorem 2.6.2] to the event $T_{\text{ave}} \leq C\rho$ as follows:

$$\Pr\left(T_{\text{ave}} \geq C\rho\right) \leq e^{\frac{-c_1(C\rho-\mu)^2}{K}m'}. \tag{87}$$

Consider the following lemma:

Lemma 9. Define $T_{\text{ave}}(\mathbf{x}_\star)$ as in (??). Then for any $\mathbf{x}_\star, \bar{\mathbf{x}} \in \mathcal{T}$ we have

$$T_{\text{ave}}(\mathbf{x}_\star) \leq \frac{1}{m'}\sum_{\ell=1}^{m}\left\|\boldsymbol{\tau}^{(\ell)}\right\|_1 + c_{\mathbf{A}}\left(\|\mathbf{x}_\star-\bar{\mathbf{x}}\|_2+1\right), \tag{88}$$

where $c_{\mathbf{A}} = \frac{1}{n}\sum_{j=1}^{n}\|\mathbf{a}_j\|_2$.

Proof: For any $j \in [n], \ell \in [m]$ and any $\mathbf{x}_\star, \bar{\mathbf{x}} \in \mathcal{T}$ we have

$$\begin{aligned}\left|\mathbf{a}_j\left(\mathbf{x}_\star-\bar{\mathbf{x}}\right)-\tau_j^{(\ell)}\right| &\leq \left|\tau_j^{(\ell)}\right| + \left|\mathbf{a}_j\left(\mathbf{x}_\star-\bar{\mathbf{x}}\right)\right| \\ &\leq \left|\tau_j^{(\ell)}\right| + \|\mathbf{a}_j\|_2\|\mathbf{x}_\star-\bar{\mathbf{x}}\|_2,\end{aligned} \tag{89}$$

where the last step is derived based on the Cauchy–Schwarz inequality. By averaging the left and the right-hand sides of (??) over all $j \in [n], \ell \in [m]$ we have

$$\begin{aligned}\frac{1}{m'}\sum_{j,\ell=1}^{m'}\left|\mathbf{a}_j\left(\mathbf{x}_\star-\bar{\mathbf{x}}\right)-\tau_j^{(\ell)}\right| &\leq \frac{1}{m'}\sum_{\ell=1}^{m}\left\|\boldsymbol{\tau}^{(\ell)}\right\|_1 + \|\mathbf{x}_\star-\bar{\mathbf{x}}\|_2\left(\frac{1}{n}\sum_{j=1}^{n}\|\mathbf{a}_j\|_2\right) \\ &= \frac{1}{m'}\sum_{\ell=1}^{m}\left\|\boldsymbol{\tau}^{(\ell)}\right\|_1 + c_{\mathbf{A}}\|\mathbf{x}_\star-\bar{\mathbf{x}}\|_2,\end{aligned} \tag{90}$$

where $T_{\text{ave}}(\mathbf{x}_\star-\bar{\mathbf{x}}) = \frac{1}{m'}\sum_{j,\ell=1}^{m'}\left|\mathbf{a}_j\left(\mathbf{x}_\star-\bar{\mathbf{x}}\right)-\tau_j^{(\ell)}\right|$. Note that $T_{\text{ave}}(\mathbf{x}_\star-\bar{\mathbf{x}})$ and $T_{\text{ave}}(\mathbf{x}_\star)$ are related because

$$\begin{aligned}\left|\mathbf{a}_j\left(\mathbf{x}_\star-\bar{\mathbf{x}}\right)-\tau_j^{(\ell)}\right| &= \left|\mathbf{a}_j\mathbf{x}_\star-\tau_j^{(\ell)}-\mathbf{a}_j\bar{\mathbf{x}}\right| \\ &\geq \left|\mathbf{a}_j\mathbf{x}_\star-\tau_j^{(\ell)}\right| - \left|\mathbf{a}_j\bar{\mathbf{x}}\right| \\ &\geq \left|\mathbf{a}_j\mathbf{x}_\star-\tau_j^{(\ell)}\right| - \|\mathbf{a}_j\|_2\|\bar{\mathbf{x}}\|_2.\end{aligned} \tag{91}$$

Averaging both sides of (??) over all $j \in [n], \ell \in [m]$ leads to

$$\frac{1}{m'}\sum_{j,\ell=1}^{m'}\left|\mathbf{a}_j\left(\mathbf{x}_\star-\bar{\mathbf{x}}\right)-\tau_j^{(\ell)}\right| \geq \frac{1}{m'}\sum_{j,\ell=1}^{m'}\left|\mathbf{a}_j\mathbf{x}_\star-\tau_j^{(\ell)}\right| - c_{\mathbf{A}}\|\bar{\mathbf{x}}\|_2, \tag{92}$$

which informs $T_{\text{ave}}(\mathbf{x}_\star) \leq T_{\text{ave}}(\mathbf{x}_\star-\bar{\mathbf{x}}) + c_{\mathbf{A}}\|\bar{\mathbf{x}}\|_2$. Combining this result with (??) completes the proof. ■

Based on Lemma ??, to ensure that the event $T_{\text{ave}} \leq C\rho = \mu+\delta$ implies $\|\bar{\mathbf{x}}-\mathbf{x}_\star\|_2 \leq \rho$ with a failure probability at most $e^{\frac{-c_1\delta^2}{K}m'}$, we should have

$$\delta \ll \frac{1}{m'}\sum_{\ell=1}^{m}\left\|\boldsymbol{\tau}^{(\ell)}\right\|_1 - \mu + c_{\mathbf{A}}\left(\rho+1\right). \tag{93}$$

Now to include all possible $\mathbf{x}_\star, \bar{\mathbf{x}} \in \mathcal{T}$ that satisfy the uniform perfect reconstruction criterion with $\bar{\mathbf{x}} \in \mathcal{B}_\rho(\mathbf{x}_\star)$, we consider a ρ-net $\{\bar{\mathbf{x}}_1, \cdots, \bar{\mathbf{x}}_K\}$ for the set $\mathcal{T}$ which means that, for any $\mathbf{x}_\star \in \mathcal{T}$, one can find $k \in [1:K]$ such that $\|\mathbf{x}_\star - \bar{\mathbf{x}}_k\|_2 \leq \rho$. We can take $K \leq \left(1+\frac{3}{\rho}\right)^d$ [?] and employ the union bound to obtain the general case of (??) as follows:

$$\begin{aligned}
&\Pr\left(T_{\text{ave}} \geq C\rho \text{ for some } k \in [1:K]\right) \\
&\leq K e^{\frac{-c_1(C\rho-\mu)^2}{K} m'} \\
&\leq \left(1+\frac{3}{\rho}\right)^d e^{\frac{-c_1(C\rho-\mu)^2}{K} m'} \\
&\leq e^{d\log\left(1+\frac{3}{\rho}\right) - \frac{c_1(C\rho-\mu)^2}{K} m'} \\
&\leq e^{\frac{3d}{\rho} - \frac{c_1(C\rho-\mu)^2}{K} m'}.
\end{aligned} \tag{94}$$

Considering the bound on δ in (??), to achieve a minimum probability of $1-\eta$, it is sufficient to ensure that the upper bound of (??) is lower than η which results in

$$m' \geq \frac{K}{c_1}\delta^{-2}\left(\frac{3d}{\rho} + \log\left(\frac{1}{\eta}\right)\right), \tag{95}$$

Therefore, the constant C_1 in Theorem ?? is $C_1 = \frac{K}{c_1}$.

Appendix D
Proof of Corollary ??

To prove Corollary ??, we alternatively define the distances in (??) as

$$d_j^{(\ell)}\left(\mathbf{x}_\star, \tau_j^{(\ell)}\right) = \left|\mathbf{a}_j \mathbf{x}_\star - \tau_j^{(\ell)}\right|, \quad j \in [n],\ \ell \in [m]. \tag{96}$$

Note that the definition presented in (??) is exactly the same with the one in (??) owing to the fact that each $r_j^{(\ell)}$ is either $+1$ or -1. The dct coefficients $\{f_j = \mathbf{a}_j \mathbf{x}_\star\}$ are defined as

$$f_j = \sqrt{\frac{2}{d}} \sum_i^d \Lambda_i \cos\left(\frac{\pi j}{2d}(2i+1)\right)(\mathbf{x}_\star)_i, \quad j \in [n], \tag{97}$$

where Λ_i is

$$\Lambda_i = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } i = 0, \\ 1 & \text{otherwise.} \end{cases} \tag{98}$$

Note that when the distances defined in (??) are bounded random variables, $0 \leq d_j^{(\ell)}\left(\mathbf{x}_\star, \tau_j^{(\ell)}\right) \leq b$, we will have $\Pr\left(T_{\text{ave}} \geq C\rho\right) \leq e^{\frac{-2(C\rho-\mu)^2}{b^2} m'}$ [?, Theorem 2]. Therefore, to prove the Corollary ??, we only need to obtain the parameters μ and b. It is easy to verify that

$$b = \max\left\{\left|\sup_{j\in[n]} f_j + \tilde{b}\right|, \left|\inf_{j\in[n]} f_j - \tilde{b}\right|\right\}. \tag{99}$$

Now to include all possible $\mathbf{x}_\star, \mathbf{x}' \in \mathbf{\Sigma}_S$ that satisfy the uniform perfect reconstruction criterion, we consider a ρ_s-net $\{\mathbf{x}'_1, \cdots, \mathbf{x}'_K\}$ for the set $\mathbf{\Sigma}_S$ which means that, for any $\mathbf{x}_\star \in \mathbf{\Sigma}_S$, one can find $k \in [1:K]$ such that $\|\mathbf{x}_\star - \mathbf{x}'_k\|_2 \leq \rho_s$. We can take $K \leq \left(1+\frac{3}{\rho_s}\right)^s$ and employ the union bound to derive the general case of (??) as follows:

$$\begin{aligned}
&\Pr\left(T_{\text{ave}^s} \geq C\rho_s \text{ for some } k \in [1:K]\right)\\
&\leq K e^{\frac{-c_1(C\rho_s-\mu_s)^2}{K_s}m'}\\
&\leq \left(1+\frac{3}{\rho}\right)^s e^{\frac{-c_1(C\rho_s-\mu_s)^2}{K_s}m'}\\
&\leq e^{s\log\left(1+\frac{3}{\rho}\right)-\frac{c_1(C\rho_s-\mu_s)^2}{K_s}m'}\\
&\leq e^{\frac{3s}{\rho_s}-\frac{c_1(C\rho_s-\mu_s)^2}{K_s}m'}.
\end{aligned} \tag{115}$$

The probability mentioned earlier corresponds to a single possibility of a fixed index set S. However, to account for all possible s-sparse solutions, we unfix the index set S and employ the union bound and express the general case of (??) as follows:

$$\begin{aligned}
&\Pr\left(T^s_{\text{ave}} \geq C\rho_s \text{ for some } S \subseteq [1:n] \text{ with } \operatorname{card}(S)=s\right)\\
&= \binom{d}{s} e^{\frac{3s}{\rho_s}-\frac{c_1(C\rho_s-\mu_s)^2}{K_s}m'}\\
&\leq e^{s\log\left(\frac{ed}{s}\right)+\frac{3}{\rho_s}s-\frac{c_1(C\rho_s-\mu_s)^2}{K_s}m'}\\
&= e^{s\left(\log\left(\frac{ed}{s}\right)+\frac{3}{\rho_s}\right)-\frac{c_1(C\rho_s-\mu_s)^2}{K_s}m'},
\end{aligned} \tag{116}$$

where we have used the inequality $\binom{d}{s} \leq \left(\frac{ed}{s}\right)^s$. Considering the bound on δ in (??), to achieve a minimum probability of $1-\eta$, it is sufficient to ensure that the upper bound of (??) is lower than η which results in

$$m' \geq \frac{K_s}{c_1}\delta^{-2}\left(\log\left(\frac{1}{\eta}\right)+s\left(\log\left(\frac{ed}{s}\right)+\frac{3}{\rho_s}\right)\right). \tag{117}$$

Therefore, the constant δ_s in Theorem ?? is $\delta_s = \frac{K_s}{c_1}\delta^{-2}$.

Appendix G
Proof of Theorem ??

Define the set $\mathcal{K}_{n_1,r}$ as

$$\mathcal{K}_{n_1,r} = \left\{\mathbf{X} \in \mathbb{R}^{n_1\times n_1} \mid \operatorname{rank}(\mathbf{X}) \leq r, \|\mathbf{X}\|_{\mathrm{F}} \leq 1\right\}. \tag{118}$$