Quantizing Heavy-tailed Data in Statistical Estimation: (Near) Minimax Rates, Covariate Quantization, and Uniform Recovery

Junren Chen, Michael K. Ng, *Senior Member, IEEE*, and Di Wang, *Member, IEEE*

***Abstract*—Modern datasets often exhibit heavy-tailed behavior, while quantization is inevitable in digital signal processing and many machine learning problems. This paper studies the quantization of heavy-tailed data in several fundamental statistical estimation problems where the underlying distributions have bounded moments of some order (no greater than 4). We propose to truncate and properly dither the data prior to a uniform quantization. Our major standpoint is that (near) minimax rates of estimation error could be achieved by computationally tractable estimators based on the quantized data produced by the proposed scheme. In particular, concrete results are worked out for covariance estimation, compressed sensing (also interpreted as sparse linear regression), and matrix completion, all agreeing that the quantization only slightly worsens the multiplicative factor. Additionally, while prior results focused on the quantization of responses (i.e., measurements), we study compressed sensing where the covariates (i.e., sensing vectors) are also quantized; in this case, though our recovery program is non-convex (since our covariance matrix estimator lacks positive semi-definiteness), we prove that all local minimizers enjoy near-optimal estimation error. Moreover, by the concentration inequality of the product process and a covering argument, we establish a near minimax uniform recovery guarantee for quantized compressed sensing with heavy-tailed noise. Finally, numerical simulations are provided to corroborate our theoretical results.**

I. Introduction

Heavy-tailed distributions are ubiquitous in modern datasets, especially those arising in economy, finance, imaging, biology, see [5], [45], [57], [84], [88], [93] for instance. In the recent literature, heavy-tailed distribution is often captured by bounded l-th moment, where l is some fixed small scalar; this is essentially weaker than the sub-Gaussian assumption. As a result, outliers and extreme values appear much more frequently in data from heavy-tailed distributions (referred to as heavy-tailed data), which poses challenges for statistical analysis. In fact, many standard statistical procedures developed for sub-Gaussian data suffer from performance degradation in the heavy-tailed regime. Fortunately, the past decade has witnessed considerable progress in statistical estimation methods that are robust to heavy-tailedness, see [9], [17], [35], [36], [44], [65], [68], [73], [97] for instance.

Departing momentarily from heavy-tailed data, quantization, which maps signals to bitstreams so that they can be stored, processed, and transmitted, is an inevitable process in the era of digital signal processing. In particular, the resolution of quantization should be selected to achieve a trade-off between accuracy and various data processing costs. In some applications, relatively low resolution would be preferable. For instance, in a distributed learning setting or a MIMO system, the frequent information transmission among multiple parties often results in prohibitive communication cost [56], [69], and quantizing signals or data to fairly low resolution is an effective approach to reduce the cost [43], [96]. Under such a big picture, in recent years there has been rapidly growing literature on high-dimensional signal recovery from quantized data (see, e.g., [7], [22], [26], [32], [33], [50], [89] for 1-bit quantization, [46], [50], [89], [94] for multi-bit uniform quantization), trying to understand the interplay between quantization and signal reconstruction in some fundamental estimation problems.

Independently, a set of robustifying techniques has been developed to overcome the challenge posed by heavy-tailed data, and uniform data quantization under uniform dither was shown to cost very little in some recovery problems. Considering the ubiquitousness of heavy-tailed behavior and data quantization, a natural question is to design a quantization scheme for heavy-tailed data that only incurs minor information loss. For instance, when applied to statistical estimation problems with heavy-tailed data, an appropriate quantization scheme should enable at least one faithful estimator from the quantized data, and ideally the estimator that could nearly achieve the optimal error rate. Despite the vast literature in this field, prior results that simultaneously take heavy-tailed data and quantization into account are surprisingly rare — only the ones presented in [33] and our earlier work [22] regarding the dithered 1-bit quantizer, to the best of our knowledge. These results remain incomplete and exhibit some downsides. Specifically, [33] considered a computationally intractable program for quantized compressed sensing and used techniques hard to generalize to other problems, while the error rates in [22] are inferior to the corresponding minimax ones (under unquantized sub-Gaussian

Junren Chen is with Department of Mathematics, The University of Hong Kong. Michael K. Ng is with Department of Mathematics, Hong Kong Baptist University. Di Wang is with Division of CEMSE, King Abdullah University of Science and Technology (KAUST). (e-mail: chenjr58@connect.hku.hk; michael-ng@hkbu.edu.hk; di.wang@kaust.edu.sa)

Junren Chen was supported by the Hong Kong Ph.D. Fellowship from the Hong Kong Research Grants Council (HKRGC). Michael K. Ng was partially supported by the HKRGC GRF 17201020, 17300021, CRF C7004-21GF and Joint NSFC-RGC N-HKU76921. Di Wang was supported in part by the baseline funding BAS/1/1689-01-01, funding from the CRG grand URF/1/4663-01-01, REI/1/5232-01-01, REI/1/5332-01-01, FCC/1/1976-49-01 from CBRC of King Abdullah University of Science and Technology (KAUST), and also supported by the funding RGC/3/4816-09-01 of the SDAIA-KAUST Center of Excellence in Data Science and Artificial Intelligence (SDAIA-KAUST AI). *(Corresponding author: Junren Chen.)*

data), as will be discussed in Section I-A3. In a nutshell, a quantization scheme for heavy-tailed data arising in statistical estimation problems that allows for computationally tractable near-minimax estimators is still lacking.

This paper aims to provide a solution to the above question and narrow the gap between heavy-tailed data and data quantization in the literature. In particular, we propose a unified quantization scheme for heavy-tailed data which, when applied to the canonical estimation problems of (sparse) covariance matrix estimation, compressed sensing (or sparse linear regression) and matrix completion, allows for (near) minimax estimators that are either in closed-form or can be solved from convex programs. Additionally, we present novel developments concerning covariate (or sensing vector) quantization and uniform signal recovery in quantized compressed sensing with heavy-tailed data.

A. Related Works

This section is devoted to a review of the most relevant works. Before that we note that a heavy-tailed random variable in this work is formulated by the moment constraint $\mathbb{E}|X|^l \leq M$, where M is an absolute constant and l is some fixed small scalar (specifically, $l \leq 4$ in the present paper). In Sections I-A1 to I-A3 we focus on estimation problems from possibly heavy-tailed and/or quantized data; then, in Sections I-A4 to I-A5 we touch on two specific aspects of quantized compressed sensing, namely the lesser-known covariate quantization problem and the highly sought-after notion of uniform recovery.

1) Statistical Estimation under Heavy-Tailed Data: Developing estimation methods that are robust to heavy-tailedness has become a recent focus in the statistics literature, where heavy-tailed distributions are often only assumed to have bounded moments of some small order. In particular, significant efforts have been devoted to the fundamental problem of mean estimation for heavy-tailed distribution. For instance, effective techniques available in the literature include Catoni's mean estimator [17], [35], median of means [68], [73], and trimmed mean [28], [66]. Indeed, these methods share the same core spirit of making the outliers less influential. To this end, the trimmed method (also referred to as truncation or shrinkage) may be the most intuitive — it truncates overlarge data to some threshold so that they are more benign for the estimation procedure. For more in-depth discussions we refer to the recent survey [65]. Furthermore, these robust methods for estimating the mean have been applied to empirical risk minimization [9], [44] and various high-dimensional estimation problems [36], [97], achieving near optimal guarantees. For instance, by invoking M-estimators with truncated data, (near) minimax rates can be achieved in high-dimensional sparse linear regression, matrix completion and covariance estimation [36].

While we capture heavy-tailedness by bounded moment of some small order, there has been a line of works considering sub-exponential or more generally sub-Weibull distributions [39], [58], [80], [85], which have heavier tail than sub-Gaussian ones but still possess finite moment up to arbitrary order. Specifically, without truncation and quantization, sparse linear regression was studied under sub-exponential data in [85] and under sub-Weibull data in [58], and the obtained error rates match the ones in the sub-Gaussian case up to logarithmic factors. Additionally, under sub-exponential measurement matrix and noise, [80] established a uniform guarantee for 1-bit generative compressed sensing, while [39] analyzed generalized Lasso for a general nonlinear model. Because the tail assumptions in these works are substantially stronger than ours, there is not a common fair stage for further comparison.

2) Statistical Estimation from Quantized Data: **Quantized Compressed Sensing.** While there have been other quantization methods, we only review the most relevant memoryless quantization schemes[1] that allow for simple hardware design, with an emphasis on the benefit of dithering. An important model is 1-bit compressed sensing where only the sign of the measurement is retained [7], [48], [75], [76]; more precisely, this model concerns the recovery of sparse $\boldsymbol{\theta}^\star \in \mathbb{R}^d$ from $\mathrm{sign}(\boldsymbol{X}\boldsymbol{\theta}^\star)$ with the sensing matrix $\boldsymbol{X} \in \mathbb{R}^{n\times d}$. However, 1-bit compressed sensing associated with the direct $\mathrm{sign}(\cdot)$ quantization suffers from some frustrating limitations, e.g., the loss of signal norm information, and the identifiability issue under some regular sensing matrix (e.g., under Bernoulli sensing matrix, see [33]).[2] Fortunately, these limitations can be overcome by random dithering prior to the quantization, under which the 1-bit measurements read as $\mathrm{sign}(\boldsymbol{X}\boldsymbol{\theta}^\star + \boldsymbol{\tau})$, with $\boldsymbol{\tau} \in \mathbb{R}^n$ being some suitably chosen random dither. Specifically, under Gaussian dither $\boldsymbol{\tau} \sim \mathcal{N}(0, \boldsymbol{I}_n)$ and standard Gaussian sensing matrix $\boldsymbol{X}$, full reconstruction with norm information could be achieved [54]. More surprisingly, under a uniform random dither, recovery with norm can be achieved under a rather general sub-Gaussian sensing matrix [22], [33], [50], [89] even with near optimal error rate.

Besides the 1-bit quantizer that retains the sign, the uniform quantizer maps $a \in \mathbb{R}$ to $\mathcal{Q}_\Delta(a) = \Delta\big(\lfloor \frac{a}{\Delta} \rfloor + \frac{1}{2}\big)$ for some pre-specified $\Delta > 0$; here and hereafter, we refer to Δ as the quantization level, and note that smaller Δ represents higher resolution. While recovering $\boldsymbol{\theta}^\star$ from $\mathcal{Q}_\Delta(\boldsymbol{X}\boldsymbol{\theta}^\star)$ encounters identifiability issue,[3] it is again beneficial to use random dithering to obtain the measurements $\mathcal{Q}_\Delta(\boldsymbol{X}\boldsymbol{\theta}^\star + \boldsymbol{\tau})$. More specifically, by using uniform dither the Lasso estimator [87], [89] and projected back projection (PBP) method [94] achieve minimax rate in certain cases, and the derived error bounds for these estimators demonstrate that the dithered uniform quantization does not affect the scaling law but only slightly worsens the multiplicative factor. Although the aforementioned progress was recently made, the technique of dithering in

[1]This means that the quantization methods for different measurements are independent. For other quantization schemes, we refer to the recent survey [29].

[2]In fact, almost all existing guarantees using the 1-bit observations $\mathrm{sign}(\boldsymbol{X}\boldsymbol{\theta}^\star)$ are restricted to standard Gaussian sensing matrix consisting of i.i.d. $\mathcal{N}(0,1)$ entries, with the exceptions of [1] for sub-Gaussian sensing matrix and [30], [86] for partial Gaussian circulant matrix.

[3]For instance, if $\boldsymbol{X} \in \{-1,1\}^{n\times d}$ (typical example is the Bernoulli design where entries of $\boldsymbol{X}$ are i.i.d. zero-mean) and $\Delta = 1$, then $\boldsymbol{\theta}_1 := 1.1\boldsymbol{e}_1$ and $\boldsymbol{\theta}_2 := 1.2\boldsymbol{e}_1 + 0.1\boldsymbol{e}_2$ can never be distinguished because $\mathcal{Q}_1(\boldsymbol{X}\boldsymbol{\theta}_1) = \mathcal{Q}_1(\boldsymbol{X}\boldsymbol{\theta}_2)$ always holds.

quantization indeed has a long history and (at least) dates back to some early engineering work (e.g., [83]), see [41] for a brief introduction.

Other Estimation Problems with Quantized Data. Some other statistical estimation problems were also investigated under dithered 1-bit quantization. Specifically, [24] studied a general signal estimation problem in a traditional setting where sample size tends to infinity, showing that dithered 1-bit quantization incurs merely logarithmic rate loss. Inspired by potential application in the reduction of power consumption in a large scale massive MIMO system, [32] proposed to collect 2 bits per entry from each sub-Gaussian sample and developed an estimator that is in general near minimax optimal. Their estimator was extended to the high-dimensional sparse case in [22], and its tuning-free version was devised in [31]. More recently, a parameter-free covariance estimator with improved operator norm error rate was proposed in [19]. Next, considering the ubiquitousness of binary observations in many recommendation systems, the authors of [26] first approached the 1-bit matrix completion problem by maximum likelihood estimation with a nuclear norm constraint. Their method was developed in a series of follow-up works by using different regularizers/constraints to encourage low-rankness, or considering multi-bit quantization on a finite alphabet [3], [14], [16], [52], [59]. Using a uniformly dithered 1-bit quantizer, the 1-bit matrix completion result in [22] essentially departs from the standard likelihood approach and can tolerate pre-quantization noise with unknown distribution.

3) Quantization of Heavy-Tailed Data in Statistical Estimation: Note that the results we just reviewed are for estimation problems from either unquantized heavy-tailed data (Section I-A1) or quantized sub-Gaussian data (Section I-A2). In this part, we turn to existing results under quantized heavy-tailed data that are more closely related to this work. Despite being a natural question with practical relevance, quantization of heavy-tailed data was rarely studied in prior work; to our best knowledge, the only results were presented in [22], [33] concerning dithered 1-bit quantization. Specifically, [33, Thm. 1.11] considered heavy-tailed noise and possibly heavy-tailed covariate, implying that a sharp uniform error rate is achievable (see their Example 1.10). However, their result is for a computationally intractable program (Hamming distance minimization) and hence of limited practical value. Another limitation is that their techniques (based on random hyperplane tessellations) are specialized to 1-bit compressed sensing but do not generalize to other estimation problems. In contrast, [22] proposed a unified quantization scheme that first truncates the data and then invokes a dithered 1-bit quantizer. Although this quantization scheme could (at least) be applied to sparse covariance matrix estimation, compressed sensing, and matrix completion while still enabling practical estimators, the main drawback is that the convergence rates of estimation errors are essentially slower than the corresponding minimax optimal ones (e.g., $\tilde{O}\big(\frac{\sqrt{s}}{n^{1/3}}\big)$ for 1-bit compressed sensing under heavy-tailed noise [22, Thm. 10]), and in certain cases the rates cannot be improved without changing the quantization process (e.g., [22, Thm. 11] complements [22, Thm. 10] with a nearly matching lower bound). In a nutshell, [33] proved a sharp rate for 1-bit compressed sensing but used highly intractable program and its techniques are not extendable to other estimation regimes, while the more widely applicable scheme and practical estimators in [22] suffer from slow error rates.

4) Covariate Quantization in Compressed Sensing: In the rest of the review, we will concentrate on quantized compressed sensing, i.e., the recovery of a sparse signal $\boldsymbol{\theta}^\star \in \mathbb{R}^d$ from the quantized version of $(\boldsymbol{x}_k, y_k := \boldsymbol{x}_k^\top \boldsymbol{\theta}^\star + \epsilon_k)_{k=1}^n$ where $\boldsymbol{x}_k, y_k, \epsilon_k$ are the sensing vector, measurement and noise, respectively. Note that this formulation also models the sparse linear regression problem (e.g., [72], [81]) where one wants to learn a sparse parameter $\boldsymbol{\theta}^\star \in \mathbb{R}^d$ from the given data $(\boldsymbol{x}_k, y_k)_{k=1}^n$, which are believed to follow the linear model $y_k = \boldsymbol{x}_k^\top \boldsymbol{\theta}^\star + \epsilon_k$. In this regression setting, $\boldsymbol{x}_k, y_k$ are commonly referred to as covariate and response, respectively. We are interested in both settings in this work (as further explained in Section III-B), but for clearer presentation, we simply refer to the problem as *quantized compressed sensing*, and term $\boldsymbol{x}_k, y_k$ as *covariate* and *response*, respectively.

Note that studying "*how quantization of covariate affects the recovery/learning*" is meaningful especially when the problem is interpreted as sparse linear regression — working with low-precision data in some (distributed) learning systems could significantly reduce communication cost and power consumption [43], [96], which we will further demonstrate in Section IV-A. However, almost all of the prior works are restricted to response quantization. To the best of our knowledge, the only existing rigorous guarantees involving covariate quantization were obtained in [22, Thms. 7-8]. Nevertheless, these results require $\mathbb{E}(\boldsymbol{x}_k \boldsymbol{x}_k^\top)$ to be sparse [22, Assumption 3] (so that their sparse covariance matrix estimator is applicable). Note that this assumption is non-standard and rarely assumed in sparse linear regression and compressed sensing.[4] It is thus desired to develop theoretical guarantees without resorting to the sparsity on $\mathbb{E}(\boldsymbol{x}_k \boldsymbol{x}_k^\top)$.

5) Uniform Signal Recovery in Compressed Sensing: It is standard in compressed sensing to leverage a random sensing matrix, so a recovery guarantee can be uniform or non-uniform. More precisely, a uniform guarantee ensures the recovery of all structured signals of interest with a single draw of the sensing ensemble, while a non-uniform guarantee is only valid for a structured signal fixed before drawing the random ensemble, with the implication that a new realization of the sensing matrix is required for sensing a new signal. Uniformity is a highly desired property in compressed sensing, since in applications the measurement ensemble is typically fixed and is expected to work for all signals [21], [40]. Besides, the derivation of a uniform guarantee is often significantly harder than a non-uniform one, making uniformity an interesting theoretical problem in its own right. (As a result, in the literature of compressed sensing, theoretical results for

[4] In fact, although isotropic sensing vector (i.e., $\mathbb{E}(\boldsymbol{x}_k \boldsymbol{x}_k^\top) = \boldsymbol{I}_d$) has been conventional in compressed sensing, many results in the literature can be extended to sensing vector with general unknown covariance matrix and hence do not really rely on the sparsity of $\mathbb{E}(\boldsymbol{x}_k \boldsymbol{x}_k^\top)$.

the nonuniform setting are often a precursor to the uniform recovery guarantees.)

A classical fact in linear compressed sensing is that the restricted isometry property (RIP) of the sensing matrix implies uniform recovery of all sparse signals (e.g., [38]), but this is unfortunately not the case when it comes to nonlinear compressed sensing models. For instance, in the specific quantization model, the more general single index model $y_k = f\big(\boldsymbol{x}_k^\top \boldsymbol{\theta}^\star\big)$ with possibly unknown $f(\cdot)$, and the lesser-known phase-only model $y_k = \mathrm{sign}(\boldsymbol{x}_k^\top \boldsymbol{\theta}^\star)$ with $\boldsymbol{x}_k \in \mathbb{C}^d$, most representative results are non-uniform (e.g., [47], [75], [78], [79], [87], [89], [94]). We refer to [20], [33], [50], [75], [94] for concrete uniform guarantees, some of which remain (near) optimal (e.g., [94, Sect. 7.2A]), while others suffer from essential degradation compared to the non-uniform ones (e.g., [75, Thm. 1.3]). It is worth noting that the interesting recent work [40] provided a unified approach to uniform guarantee for a series of non-linear models; however, their uniform guarantees typically exhibit a decaying rate of $O(n^{-1/4})$ that is slower than the non-uniform one of $O(n^{-1/2})$ (Section 4 therein). As a follow-up work, [21] obtained a near optimal uniform $O(n^{-1/2})$ error rate for various nonlinear generative compressed sensing models.

The above-reviewed uniform guarantees are restricted to sub-Gaussian data. Regarding dithered 1-bit quantization of heavy-tailed data, results in [22, Sect. III] are non-uniform, while [33, Thm. 1.11] presents a sharp uniform guarantee for the intractable program of hamming distance minimization.

B. Our Contributions

We summarize our contributions in this part. Our primary contribution is a unified quantization scheme (for heavy-tailed data) that allows for near-minimax estimators. Besides, we present new developments regarding covariate quantization and uniform recovery in a heavy-tailed quantized setting. This work also provides some notable innovations that prove useful in related studies.

A Unified Quantization Scheme and Estimators with (Near) Minimax Rates. We propose a unified quantization scheme for heavy-tailed data which allows for *practical* and *(near) optimal* estimators in *multiple* estimation problems. The proposed scheme consists of three steps: 1) *truncation* that shrinks data to some threshold, 2) *dithering* that adds suitable random noise to the truncated data, and 3) *uniform quantization*. Note that the proposed scheme replaces the 1-bit quantizer in [22] with the less extreme (multi-bit) uniform quantizer $\mathcal{Q}_\Delta(\cdot)$, but the gain turns out to be significant — we are now able to derive (near) optimal rates that are essentially faster than the ones in [22], see Theorems 2-8. As a concrete example, for quantized compressed sensing with sub-Gaussian sensing vector $\boldsymbol{x}_k$ but heavy-tailed measurement y_k satisfying $\mathbb{E}|y_k|^{2+\nu} \le M$ for some $\nu > 0$, we derive the ℓ_2-norm error rate $O\big((M^{1/(2l)}+\Delta)\sqrt{\frac{s\log d}{n}}\big)$ (Theorem 5, s, d, n are respectively the sparsity, signal dimension, measurement number).

Compared to [33], our major advantages are that our estimators are computationally feasible and that our method applies to estimation problems beyond quantized compressed sensing. Concerning the effect of quantization, our results (in all the considered models) suggest a unified conclusion — *dithered uniform quantization does not affect the scaling law but only slightly worsens the multiplicative factor*, which generalizes similar findings for quantized compressed sensing in [87], [89], [94] towards two directions, i.e., to the case where heavy-tailed data present and to some other estimation problems (i.e., covariance matrix estimation, matrix completion).

New Results on Covariate Quantization. Besides the above main contributions, we establish the estimation guarantees for quantized compressed sensing under covariate quantization that are free of the non-standard assumption on the sparsity of $\mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top)$. This provides important relaxations to [22, Thms. 7-8]. More specifically, unlike [22] that relies on the sparsity of $\mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top)$ to ensure convexity, we deal with the non-convex program directly and manage to prove that all local minimizers deliver near minimax estimation errors (Theorems 9-10). Our analysis is motivated by a line of works on non-convex M-estimator [62]–[64] but also exhibits some essential differences (Remark 5). Further, we extract our techniques as a deterministic framework (Proposition 1) and then use it to establish guarantees for dithered 1-bit quantization as byproducts (Theorems 11-12), which are comparable to [22, Thms. 7-8] but free of sparsity on $\mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top)$.

Uniform Recovery Guarantee. We additionally contribute to the literature a uniform guarantee for constrained Lasso under the dithered uniform quantization of heavy-tailed response. Specifically, we upgrade our non-uniform Theorem 5 to its uniform version Theorem 13, which states that using a single realization of the sub-Gaussian sensing matrix, heavy-tailed noise and uniform dither, all s-sparse signals within an ℓ_2-ball can be uniformly recovered up to an ℓ_2-norm error of $\tilde{O}\big(\sqrt{\frac{s}{n}}\big)$, thus matching the near minimax non-uniform rate in Theorem 5 up to logarithmic factors. The proof relies on a concentration inequality for product process [67] and a careful covering argument inspired by [94]. Due to the heavy-tailed noise, a new treatment is needed before invoking the concentration result from [67].

Some Notable Innovations. An important innovation of our work is to use triangular dither when covariance estimation is necessary, which departs from the uniform dither commonly adopted in prior works (e.g., [22], [32], [89], [94]) and is novel to the literature. From a technical side, many of our analyses on the dithered quantizer are much cleaner than prior works because we make full use of the nice statistical properties of the quantization error and quantization noise (Theorem 1),[5] see Section II-B. Based on these two innovations (and thus after the first appearance of the present paper), a clean analysis on quantized low-rank multivariate regression *with possibly quantized covariates* is provided in [23], and more recently a 2-bit covariance estimator based on triangular dithering is devised in [19].

[5]Prior work that did not fully leverage these properties may incur extra technical complication, e.g., the symmetrization and contraction in [89, Lem. A.2].

C. Outline

The remainder of this paper is structured as follows. We provide the notation and preliminaries in Section II. We present the first set of main results (concerning the (near) optimal guarantees for three estimation problems under quantized heavy-tailed data) in Section III. Our second set of results (concerning covariate quantization and uniform recovery in quantized compressed sensing) is then presented in Section IV. To corroborate our theory, numerical results on synthetic data are reported in Section V. We give some remarks to conclude the paper in Section VI. All the proofs are postponed to the Appendices.

II. Preliminaries

We adopt the following conventions throughout the paper:

1) We use boldface symbols (e.g., $\boldsymbol{A}$, $\boldsymbol{x}$) to denote matrices and vectors, and regular letters (e.g., a, x) for scalars. We write $[m] = \{1, ..., m\}$ for positive integer m. We denote the complex unit by i. The i-th entry for a vector $\boldsymbol{x}$ (likewise, $\boldsymbol{y}$, $\boldsymbol{\tau}$) is denoted by x_i (likewise, y_i, τ_i).

2) Notation with "$\star$" as superscript denotes the desired underlying parameter or signal, e.g., $\boldsymbol{\Sigma}^\star$, $\boldsymbol{\theta}^\star$. Moreover, notation marked by a tilde (e.g., $\widetilde{\boldsymbol{x}}$) and a dot (e.g., $\dot{\boldsymbol{x}}$) stands for the truncated data and quantized data, respectively.

3) We reserve d and n for the problem dimension and sample size, respectively. In many cases $\widehat{\boldsymbol{\Upsilon}}$ denotes the estimation error, e.g., $\widehat{\boldsymbol{\Upsilon}} = \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star$ if $\widehat{\boldsymbol{\theta}}$ is the estimator for the desired signal $\boldsymbol{\theta}^\star$. We use Σ_s to denote the set of d-dimensional s-sparse signals.

4) For vector $\boldsymbol{x} \in \mathbb{R}^d$, we work with its transpose $\boldsymbol{x}^\top$, ℓ_p-norm $\|\boldsymbol{x}\|_p = (\sum_{i\in[d]} |x_i|^p)^{1/p}$ ($p \geq 1$), max norm $\|\boldsymbol{x}\|_\infty = \max_{i\in[d]} |x_i|$. We define the standard Euclidean sphere as $\mathbb{S}^{d-1} = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 = 1\}$.

5) For matrix $\boldsymbol{A} = [a_{ij}] \in \mathbb{R}^{m\times n}$ with singular values $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_{\min\{m,n\}}$, recall the operator norm $\|\boldsymbol{A}\|_{op} = \sup_{\boldsymbol{v}\in\mathbb{S}^{n-1}} \|\boldsymbol{A}\boldsymbol{v}\|_2 = \sigma_1$, Frobenius norm $\|\boldsymbol{A}\|_F = (\sum_{i,j} a_{ij}^2)^{1/2}$, nuclear norm $\|\boldsymbol{A}\|_{nu} = \sum_{k=1}^{\min\{m,n\}} \sigma_k$, and max norm $\|\boldsymbol{A}\|_\infty = \max_{i,j} |a_{ij}|$. $\lambda_{\min}(\boldsymbol{A})$ (resp. $\lambda_{\max}(\boldsymbol{A})$) stands for the minimum eigenvalue (resp. maximum eigenvalue) of a symmetric $\boldsymbol{A}$.

6) We denote universal constants by C, c, C_i and c_i, whose value may vary from line to line. We write $T_1 \lesssim T_2$ or $T_1 = O(T_2)$ if $T_1 \leq CT_2$. Conversely, if $T_1 \geq CT_2$ we write $T_1 \gtrsim T_2$ or $T_1 = \Omega(T_2)$. Also, we write $T_1 \asymp T_2$ if $T_1 = O(T_2)$ and $T_2 = \Omega(T_1)$ simultaneously hold.

7) We use $\mathscr{U}(\Omega)$ to denote the uniform distribution over $\Omega \subset \mathbb{R}^N$, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, $\mathtt{t}(\nu)$ to denote student's t distribution with degrees of freedom ν.

8) Our technique to handle heavy-tailedness is a data truncation step, for which we introduce the operator $\mathscr{T}_\zeta(\cdot)$ for some threshold $\zeta > 0$. It is defined as $\mathscr{T}_\zeta(a) = \mathrm{sign}(a)\min\{|a|, \zeta\}$ for some $a \in \mathbb{R}$. To truncate vectors we apply $\mathscr{T}_\zeta(\cdot)$ entry-wisely in most cases, with the exception of covariance matrix estimation under operator norm error (Theorem 3).

9) $\mathcal{Q}_\Delta(.)$ is the uniform quantizer with quantization level $\Delta > 0$. It applies to scalar a by $\mathcal{Q}_\Delta(a) = \Delta\big(\lfloor \frac{a}{\Delta} \rfloor + \frac{1}{2}\big)$, and we set $\mathcal{Q}_0(a) = a$. Given a threshold μ, the hard thresholding of scalar a is $\mathcal{T}_\mu(a) = a \cdot \mathbb{1}(|a| \geq \mu)$. Both functions element-wisely apply to vectors or matrices.

A. High-Dimensional Statistics

Let X be a real random variable, we present some basic knowledge of the sub-Gaussian and sub-exponential random variables. Then we also precisely formulate the heavy-tailed distribution.

1) The sub-Gaussian norm is defined as $\|X\|_{\psi_2} = \inf\{t > 0 : \mathbb{E}\exp(\frac{X^2}{t^2}) \leq 2\}$. A random variable X with finite $\|X\|_{\psi_2}$ is said to be sub-Gaussian. Analogously to the Gaussian variable, a sub-Gaussian random variable exhibits an exponentially decaying probability tail and satisfies a moment constraint:

$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{ct^2}{\|X\|_{\psi_2}^2}\right); \tag{1}$$

$$(\mathbb{E}|X|^p)^{1/p} \leq C\|X\|_{\psi_2}\sqrt{p},\ \forall\ p \geq 1. \tag{2}$$

Note that these two properties can also define $\|\cdot\|_{\psi_2}$ up to multiplicative constant, e.g., $\|X\|_{\psi_2} \asymp \sup_{p\geq 1} \frac{(\mathbb{E}|X|^p)^{1/p}}{\sqrt{p}}$ (see [92, Prop. 2.5.2]). For a d-dimensional random vector $\boldsymbol{x}$ we define its sub-Gaussian norm as $\|\boldsymbol{x}\|_{\psi_2} = \sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}} \|\boldsymbol{v}^\top\boldsymbol{x}\|_{\psi_2}$.

2) The sub-exponential norm is defined as $\|X\|_{\psi_1} = \inf\{t > 0 : \mathbb{E}\exp(\frac{|X|}{t}) \leq 2\}$, and X is sub-exponential if $\|X\|_{\psi_1} < \infty$. The sub-exponential X satisfies the following properties:

$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{ct}{\|X\|_{\psi_1}}\right);$$

$$(\mathbb{E}|X|^p)^{1/p} \leq C\|X\|_{\psi_1}p,\ \forall\ p \geq 1. \tag{3}$$

To relate $\|.\|_{\psi_1}$ and $\|.\|_{\psi_2}$ one has $\|XY\|_{\psi_2} \leq \|X\|_{\psi_1}\|Y\|_{\psi_1}$ [92, Lem. 2.7.7].

3) In contrast to the moment constraints in (2) and (3), heavy-tailed distributions in this work are only assumed to satisfy bounded moments of some small order no greater than 4, formulated for a random variable X as $\mathbb{E}|X|^l \leq M$ for some $M > 0$ and $l \in (0, 4]$. Following [58, Def. 2.4, 2.5], we consider the following two moment assumptions for a heavy-tailed random vector $\boldsymbol{x} \in \mathbb{R}^d$ (again, $M > 0$, $l \in (0, 4]$):

- **Marginal Moment Constraint.** The weaker assumption that constrains the moment of each coordinate is formulated by $\sup_{i\in[d]} \mathbb{E}|x_i|^l \leq M$.
- **Joint Moment Constraint.** The stronger assumption that constrains the moments "toward all directions $\boldsymbol{v} \in \mathbb{S}^{d-1}$," is formulated by $\sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}} \mathbb{E}|\boldsymbol{v}^\top\boldsymbol{x}|^l \leq M$.

B. Dithered Uniform Quantization

In this part, we describe the dithered uniform quantizer and its properties in detail. We also specify the choices of random dither in this work.

1) We first provide the detailed procedure of dithered quantization and its general property. Let $\boldsymbol{x} \in \mathbb{R}^N$ be the input signal with dimension $N \geq 1$ whose entries may be random and dependent. Independent of $\boldsymbol{x}$, we generate the random

dither $\boldsymbol{\tau} \in \mathbb{R}^N$ with i.i.d. entries from some distribution,[6] and then quantize $\boldsymbol{x}$ to $\dot{\boldsymbol{x}} = \mathcal{Q}_\Delta(\boldsymbol{x}+\boldsymbol{\tau})$. Following [41], we refer to $\boldsymbol{w} := \dot{\boldsymbol{x}} - (\boldsymbol{x}+\boldsymbol{\tau})$ as the quantization error, and $\boldsymbol{\xi} := \dot{\boldsymbol{x}} - \boldsymbol{x}$ as the quantization noise. The principal properties of dithered quantization are provided in Theorem 1.

Theorem 1. (Adapted from [41, Thms. 1-2])**.** *Consider the dithered uniform quantization described above for the input signal $\boldsymbol{x}$, with random dither $\boldsymbol{\tau} = [\tau_i]$, quantization error $\boldsymbol{w}$ and quantization noise $\boldsymbol{\xi} = [\xi_i]$. Use* i *to denote the imaginary unit, and let Y be the random variable having the same distribution as the random dither τ_i.*

(a) (Quantization Error)**.** *If $f(u) := \mathbb{E}(\exp(\mathrm{i}uY))$ satisfies $f\big(\frac{2\pi l}{\Delta}\big) = 0$ for all non-zero integer l, then $\boldsymbol{w} \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}]^N)$ is independent of $\boldsymbol{x}$.*[7]

(b) (Quantization Noise)**.** *Assume that $Z \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}])$ is independent of Y. Let $g(u) := \mathbb{E}(\exp(\mathrm{i}uY))\mathbb{E}(\exp(\mathrm{i}uZ))$. Given positive integer p, if the p-th order derivative $g^{(p)}(u)$ satisfies $g^{(p)}\big(\frac{2\pi l}{\Delta}\big) = 0$ for all non-zero integer l, then the p-th conditional moment of ξ_i does not depend on $\boldsymbol{x}$: $\mathbb{E}[\xi_i^p|\boldsymbol{x}] = \mathbb{E}(Y+Z)^p$.*

We note that Theorem 1 serves as the cornerstone for our analysis on the dithered uniform quantizer; for instance, (a) allows for applications of concentration inequalities in our analyses, and (b) inspires us to develop a covariance matrix estimator from quantized samples. The take-home message is that, adding appropriate dither before quantization can make the quantization error and quantization noise behave in a statistically nice manner. For example, the elementary form of Theorem 1(a) is that under a dither τ_i satisfying the condition there, the quantization noise $\mathcal{Q}_\Delta(x_i+\tau_i) - (x_i+\tau_i)$ follows $\mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}])$ under any given scalar x_i [41, Lem. 1].

2) We use *uniform dither* for quantization of the response in compressed sensing and matrix completion. More specifically, under $\Delta > 0$, we adopt the uniform dither $\tau_k \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}])$ for the response $y_k \in \mathbb{R}$, which is also a common choice in previous works (e.g., [33], [50], [89], [94]). For $Y \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}])$, it can be calculated that

$$\begin{aligned}\mathbb{E}(\exp(\mathrm{i}uY)) &= \int_{-\Delta/2}^{\Delta/2} \frac{1}{\Delta}\big(\cos(ux) + \mathrm{i}\sin(ux)\big)\,\mathrm{d}x \\ &= \frac{2}{\Delta u}\sin\Big(\frac{\Delta u}{2}\Big),\end{aligned} \tag{4}$$

and hence $\mathbb{E}(\exp(\mathrm{i}\frac{2\pi l}{\Delta}Y)) = 0$ holds for all non-zero integer l. Therefore, the benefit of using $\tau_k \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}])$ is that the quantization errors $w_k = \mathcal{Q}_\Delta(y_k+\tau_k) - (y_k+\tau_k)$ i.i.d. follow $\mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}])$, and are independent of $\{y_k\}$.

3) We use *triangular dither* for quantization of the covariate, i.e., the sample in covariance estimation or the covariate in compressed sensing. Particularly, when considering the uniform quantizer $\mathcal{Q}_\Delta(.)$ for the covariate $\boldsymbol{x}_k \in \mathbb{R}^d$, we adopt the dither $\boldsymbol{\tau}_k \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}]^d) + \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}]^d)$,[8] which is the sum of two independent $\mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}]^d)$ and referred to as a triangular dither [41]. Simple calculations verify that the triangular dither respects not only the condition in Theorem 1(a), but also the one in Theorem 1(b) with $p = 2$; specifically, let $Y = Y_1 + Y_2$ where Y_1 and Y_2 are independent and follow $\mathscr{U}\big([-\frac{\Delta}{2}, \frac{\Delta}{2}]\big)$, and let $Z \sim \mathscr{U}\big([-\frac{\Delta}{2}, \frac{\Delta}{2}]\big)$ be independent of Y, then based on (4), we know that

$$f(u) = \mathbb{E}(\exp(\mathrm{i}uY)) = \Big[\frac{2}{\Delta u}\sin\frac{\Delta u}{2}\Big]^2$$

satisfies $f(\frac{2\pi l}{\Delta}) = 0$, and that

$$g(u) = \mathbb{E}(\exp(\mathrm{i}uY))\mathbb{E}(\exp(\mathrm{i}uZ)) = \Big[\frac{2}{\Delta u}\sin\frac{\Delta u}{2}\Big]^3$$

satisfies $g''(\frac{2\pi l}{\Delta}) = 0$, where l is any non-zero integer. Thus, at the cost of a dithering variance larger than uniform dither, the triangular dither brings the additional nice property of signal-independent variance for the quantization noise — $\mathbb{E}(\xi_{ki}^2) = \frac{1}{4}\Delta^2$, where ξ_{ki} is the i-th entry of $\boldsymbol{\xi}_k = \mathcal{Q}_\Delta(\boldsymbol{x}_k+\boldsymbol{\tau}_k) - (\boldsymbol{x}_k+\boldsymbol{\tau}_k)$.

To the best of our knowledge, the triangular dither is new to the literature of quantized compressed sensing. We will explain its necessity if covariance estimation is involved. This is also complemented by numerical simulation (see Figure 5(a)).

III. (Near) Minimax Error Rates

In this section, we derive (near) optimal error rates for several canonical statistical estimation problems. Our novelty is that by using the proposed quantization scheme for heavy-tailed data, (near) optimal error rates could be achieved by computationally feasible estimators.

A. Quantized Covariance Matrix Estimation

Given $\mathscr{X} := \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$ as i.i.d. copies of a zero-mean random vector $\boldsymbol{x} \in \mathbb{R}^d$, one often encounters the covariance matrix estimation problem, i.e., to estimate $\boldsymbol{\Sigma}^\star = \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^\top)$. This estimation problem is of fundamental importance in multivariate analysis and has attracted much research interest (e.g., [4], [10], [12], [13]). However, the practically useful setting (e.g., in a massive MIMO system [95]) where the samples undergo certain quantization process remains underdeveloped, for which we are only aware of the 1-bit quantization results in [22], [32]. This setting poses the problem of quantized covariance matrix estimation (QCME), in which *one aims to design a quantization scheme for $\boldsymbol{x}_k$ that allows for accurate estimation of $\boldsymbol{\Sigma}^\star$ only based on the quantized samples*. We consider heavy-tailed $\boldsymbol{x}_k$ that possesses bounded fourth moments either marginally or jointly, but note that our estimation methods and theoretical results appear to be new even for sub-Gaussian $\boldsymbol{x}_k$ (Remark 1).

As introduced before, we overcome the heavy-tailedness of $\boldsymbol{x}_k$ by a data truncation step, i.e., we first truncate $\boldsymbol{x}_k$ to $\widetilde{\boldsymbol{x}}_k$ in order to make the outliers less influential. Here, we defer the

[6] Throughout this work, we suppose that a random dither is drawn independent of anything else (particularly, the signal to be quantized and other dithers), and the dither has i.i.d. entries if it is a vector.

[7] Although the statement is a bit different, it can be implied by [41, Thm. 1] and the proof therein.

[8] An equivalent statement is that entries of $\boldsymbol{\tau}_k$ are i.i.d. distributed as $\mathscr{U}\big([-\frac{\Delta}{2}, \frac{\Delta}{2}]\big) + \mathscr{U}\big([-\frac{\Delta}{2}, \frac{\Delta}{2}]\big)$. The equivalence can be clearly seen by comparing the joint probability density functions.

precise definition of $\widetilde{\boldsymbol{x}}_k$ to concrete results because it should be well suited to the error metric. After the truncation, we dither and quantize $\widetilde{\boldsymbol{x}}_k$ to $\dot{\boldsymbol{x}}_k = \mathcal{Q}_\Delta(\widetilde{\boldsymbol{x}}_k + \boldsymbol{\tau}_k)$ with the triangular dither $\boldsymbol{\tau}_k \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}]^d) + \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}]^d)$. Different from the uniform dither adopted in the literature (e.g., [22], [33], [50], [89], [94]), first let us explain our choice of triangular dither. Recall that the quantization noise and quantization error are respectively defined as $\boldsymbol{\xi}_k := \dot{\boldsymbol{x}}_k - \widetilde{\boldsymbol{x}}_k$ and $\boldsymbol{w}_k := \dot{\boldsymbol{x}}_k - \widetilde{\boldsymbol{x}}_k - \boldsymbol{\tau}_k$, thus giving $\boldsymbol{\xi}_k = \boldsymbol{\tau}_k + \boldsymbol{w}_k$. Under uniform dither or triangular dither, $\boldsymbol{w}_k$ is independent of $\widetilde{\boldsymbol{x}}_k$ and follows $\mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}]^d)$ (see Section 2.2), thus allowing us to calculate that

$$\begin{aligned}\mathbb{E}(\dot{\boldsymbol{x}}_k \dot{\boldsymbol{x}}_k^\top) &= \mathbb{E}\big((\widetilde{\boldsymbol{x}}_k + \boldsymbol{\xi}_k)(\widetilde{\boldsymbol{x}}_k + \boldsymbol{\xi}_k)^\top\big) \\ &= \mathbb{E}(\widetilde{\boldsymbol{x}}_k \widetilde{\boldsymbol{x}}_k^\top) + \mathbb{E}(\widetilde{\boldsymbol{x}}_k \boldsymbol{\xi}_k^\top) + \mathbb{E}(\boldsymbol{\xi}_k \widetilde{\boldsymbol{x}}_k^\top) + \mathbb{E}(\boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top) \\ &\overset{(i)}{=} \mathbb{E}(\widetilde{\boldsymbol{x}}_k \widetilde{\boldsymbol{x}}_k^\top) + \mathbb{E}(\boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top). \end{aligned} \tag{5}$$

Note that (i) is because

$$\begin{aligned}\mathbb{E}(\boldsymbol{\xi}_k \widetilde{\boldsymbol{x}}_k^\top) &= \mathbb{E}(\boldsymbol{\tau}_k \widetilde{\boldsymbol{x}}_k^\top) + \mathbb{E}(\boldsymbol{w}_k \widetilde{\boldsymbol{x}}_k^\top) \\ &= \mathbb{E}(\boldsymbol{\tau}_k)\mathbb{E}(\widetilde{\boldsymbol{x}}_k^\top) + \mathbb{E}(\boldsymbol{w}_k)\mathbb{E}(\widetilde{\boldsymbol{x}}_k^\top) = 0\end{aligned}$$

due to the previously noted fact that $\boldsymbol{\tau}_k$ and $\boldsymbol{w}_k$ are independent of $\widetilde{\boldsymbol{x}}_k$ and zero-mean. While with suitable choice of the truncation threshold $\mathbb{E}(\widetilde{\boldsymbol{x}}_k \widetilde{\boldsymbol{x}}_k^\top)$ is expected to well approximate $\boldsymbol{\Sigma}^\star$, the remaining $\mathbb{E}(\boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top)$ gives rise to constant bias. To address the issue, a straightforward idea is to remove the bias, which requires the full knowledge of $\mathbb{E}(\boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top)$, i.e., the covariance matrix of the quantization noise. For $i \neq j$, because $\boldsymbol{\tau}_k$, $\boldsymbol{w}_k \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}]^d)$ and

$$\mathbb{E}(w_{ki}\tau_{kj}) = \mathbb{E}_{\widetilde{x}_{ki}}(\mathbb{E}[w_{ki}\tau_{kj}|\widetilde{x}_{ki}]) = 0$$

(note that conditionally on $\widetilde{x}_{ki}$, $w_{ki} = \mathcal{Q}_\Delta(\widetilde{x}_{ki} + \tau_{ki}) - (\widetilde{x}_{ki} + \tau_{ki})$ and τ_{kj} are independent), we have

$$\begin{aligned}&\mathbb{E}(\xi_{ki}\xi_{kj}) = \mathbb{E}\big((w_{ki} + \tau_{ki})(w_{kj} + \tau_{kj})\big) \\ &= \mathbb{E}(w_{ki}w_{kj}) + \mathbb{E}(w_{ki}\tau_{kj}) + \mathbb{E}(\tau_{ki}w_{kj}) + \mathbb{E}(\tau_{ki}\tau_{kj}) \\ &= 0,\end{aligned}$$

showing that $\mathbb{E}(\boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top)$ is diagonal. Moreover, under triangular dither the i-th diagonal entry is also known as $\mathbb{E}|\xi_{ki}|^2 = \frac{\Delta^2}{4}$, see Section II-B. Taken collectively, we arrive at

$$\mathbb{E}(\boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top) = \frac{\Delta^2}{4} \boldsymbol{I}_d; \tag{6}$$

Based on (5) we thus propose the following estimator

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{k=1}^n \dot{\boldsymbol{x}}_k \dot{\boldsymbol{x}}_k^\top - \frac{\Delta^2}{4} \boldsymbol{I}_d, \tag{7}$$

which is the sample covariance of the quantized sample $\dot{\mathscr{X}} := \{\dot{\boldsymbol{x}}_1, ..., \dot{\boldsymbol{x}}_n\}$ followed by a correction step. On the other hand, the reason why the standard uniform dither is not suitable for QCME becomes self-evident — the diagonal of $\mathbb{E}(\boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top)$ remains unknown[9] and hence there is no hope to precisely remove the bias.

We are now ready to present error bounds for $\widehat{\boldsymbol{\Sigma}}$ under max-norm, operator norm. We will also investigate the high-dimensional setting by assuming some sparse structure of $\boldsymbol{\Sigma}^\star$, for which we propose a thresholding estimator. More concretely, our first result provides the error rate under $\|\cdot\|_\infty$, in which we assume $\boldsymbol{x}_k$ satisfies the marginal fourth-moment constraint and utilize an element-wise truncation $\widetilde{\boldsymbol{x}}_k = \mathscr{T}_\zeta(\boldsymbol{x}_k)$.

Theorem 2. (Element-Wise Error)**.** *Given $\Delta > 0$ and $\delta > 4$, we consider the problem of QCME described above. We suppose that $\boldsymbol{x}_k$s are i.i.d. zero-mean and satisfy the marginal moment constraint $\mathbb{E}|x_{ki}|^4 \leq M$ for any $i \in [d]$, where x_{ki} is the i-th entry of $\boldsymbol{x}_k$. We truncate $\boldsymbol{x}_k$ to $\widetilde{\boldsymbol{x}}_k = [\widetilde{x}_{ki}] = \mathscr{T}_\zeta(\boldsymbol{x}_k)$ with threshold $\zeta \asymp \big(\frac{nM}{\delta \log d}\big)^{1/4}$, then quantize $\widetilde{\boldsymbol{x}}_k$ to $\dot{\boldsymbol{x}}_k = \mathcal{Q}_\Delta(\widetilde{\boldsymbol{x}}_k + \boldsymbol{\tau}_k)$ with triangular dither $\boldsymbol{\tau}_k \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}]^d) + \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}]^d)$. If $n \gtrsim \delta \log d$, then the estimator in (7) satisfies*

$$\mathbb{P}\left(\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^\star\|_\infty \geq C\mathscr{L}\sqrt{\frac{\delta \log d}{n}}\right) \leq 2d^{2-\delta},$$

where $\mathscr{L} := \sqrt{M} + \Delta^2$.

Notably, despite the heavy-tailedness and quantization, the estimator achieves an element-wise rate $O(\sqrt{\frac{\log d}{n}})$ coincident with the one for the sub-Gaussian case. One can clearly position quantization level Δ in the multiplicative factor $\mathscr{L} = \sqrt{M} + \Delta^2$. Thus, the information loss incurred by quantization is inessential in that it does not affect the key scaling law but only slightly worsens the leading factor. These remarks on the (near) optimality and the information loss incurred by quantization remain valid in our subsequent theorems.

Our next result concerns the operator norm estimation error, under which we impose a stronger joint moment constraint on $\boldsymbol{x}_k$ and truncate $\boldsymbol{x}_k$ regarding ℓ_4-norm, i.e., $\check{\boldsymbol{x}}_{\boldsymbol{k}} = \frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|_4} \min\{\|\boldsymbol{x}_k\|_4, \zeta\}$ for some threshold ζ. After the dithered uniform quantization, we still define the estimator as (7).

Theorem 3. (Operator Norm Error)**.** *Given $\Delta > 0$ and $\delta > 0$, we consider the problem of QCME described above. Suppose that the i.i.d. zero-mean $\boldsymbol{x}_k$s satisfy $\mathbb{E}|\boldsymbol{v}^\top \boldsymbol{x}_k|^4 \leq M$ for any $\boldsymbol{v} \in \mathbb{S}^{d-1}$. We truncate $\boldsymbol{x}_k$ to $\check{\boldsymbol{x}}_k = \frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|_4} \min\{\|\boldsymbol{x}_k\|_4, \zeta\}$ with threshold $\zeta \asymp (M^{1/4} + \Delta)\big(\frac{n}{\delta \log d}\big)^{1/4}$, then quantize $\check{\boldsymbol{x}}_k$ to $\dot{\boldsymbol{x}}_k = \mathcal{Q}_\Delta(\check{\boldsymbol{x}}_k + \boldsymbol{\tau}_k)$ with triangular dither $\boldsymbol{\tau}_k \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}]^d) + \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}]^d)$. If $n \gtrsim \delta d \log d$, then the estimator in (7) satisfies*

$$\mathbb{P}\left(\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^\star\|_{op} \geq C\mathscr{L}\sqrt{\frac{\delta d \log d}{n}}\right) \leq 2d^{-\delta},$$

with $\mathscr{L} := \sqrt{M} + \Delta^2$.

The operator norm error rate in Theorem 3 is near minimax optimal, e.g., compared to the lower bound in [36, Thm. 7], which states that for any estimator $\widehat{\boldsymbol{\Sigma}}$ of the positive semi-definite matrix $\boldsymbol{\Sigma}^\star$ based on i.i.d. zero-mean $\{\boldsymbol{x}_k\}_{k=1}^n$ with covariance matrix $\boldsymbol{\Sigma}^\star$, there exists some $\boldsymbol{v}_0 \in \mathbb{S}^{d-1}$ such that $\mathbb{P}\big(\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^\star\|_{op} \geq \frac{1}{48}\sqrt{\frac{6d}{n}}\big) \geq \frac{1}{3}$, where $\boldsymbol{\Sigma}^\star = \boldsymbol{I}_d + \boldsymbol{v}_0 \boldsymbol{v}_0^\top$. Again, the quantization only affects the multiplicative factor $\mathscr{L}$. Nevertheless, one still needs (at least) $n \gtrsim d$ to achieve a small operator norm error. In fact, in a high-dimensional

[9] It depends on the input signal, see [41, Page 3].

setting where d may exceed n, even the sample covariance $\frac{1}{n}\sum_{k=1}^n \boldsymbol{x}_k\boldsymbol{x}_k^\top$ for sub-Gaussian zero-mean $\boldsymbol{x}_k$ may have extremely bad performance. To achieve small operator norm error in a high-dimensional regime, we resort to additional structure on $\boldsymbol{\Sigma}^\star$, and specifically we use column-wise sparsity as an example, which corresponds to the situations where dependencies among different coordinates are weak. Based on the estimator in Theorem 2, we further invoke a thresholding regularization [4], [12] to promote sparsity.

Theorem 4. (Sparse QCME)**.** *Under conditions and estimator $\widehat{\boldsymbol{\Sigma}}$ in Theorem 2, we additionally assume that all columns of $\boldsymbol{\Sigma}^\star = [\sigma_{ij}^\star]$ are s-sparse and consider the thresholding estimator $\widehat{\boldsymbol{\Sigma}}_s := \mathcal{T}_\mu(\widehat{\boldsymbol{\Sigma}})$ for some μ (recall that $\mathcal{T}_\mu(a) = a\cdot\mathbb{1}(|a|\geq\mu)$ for $a\in\mathbb{R}$). If $\mu = C_1(\sqrt{M}+\Delta^2)\sqrt{\frac{\delta\log d}{n}}$ with sufficiently large C_1, then $\widehat{\boldsymbol{\Sigma}}_s$ satisfies*

$$\mathbb{P}\left(\|\widehat{\boldsymbol{\Sigma}}_s - \boldsymbol{\Sigma}^\star\|_{op} \leq C\mathscr{L}s\sqrt{\frac{\delta\log d}{n}}\right) \geq 1-\exp(-0.25\delta),$$

where $\mathscr{L} := \sqrt{M}+\Delta^2$.

Notably, our estimator $\widehat{\boldsymbol{\Sigma}}_s$ achieves minimax rates $O\big(s\sqrt{\frac{\log d}{n}}\big)$ under operator norm, e.g., compared to the minimax lower bound derived in [12, Thm. 2], which states that (under some regular scaling) for any covariance estimator $\boldsymbol{\Sigma}_{es}$ based on n i.i.d. samples of $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}^\star)$ where $\boldsymbol{\Sigma}^\star$ is the true covariance matrix, there exists some covariance matrix $\boldsymbol{\Sigma}^\star$ with s-sparse columns such that $\mathbb{E}\|\boldsymbol{\Sigma}_{es}-\boldsymbol{\Sigma}^\star\|_{op}^2 \gtrsim s^2\frac{\log d}{n}$. Note that this lower bound is for a general sparse covariance matrix with column-wise sparsity, and it is possible to achieve faster rate over covariance matrices with more specific sparse structures, see [11] for instance.

To analyze the thresholding estimator, our proof resembles the ones developed in prior works (e.g., [12]) but requires more effort like bounding the additional bias terms arising from the data truncation and quantization. We also point out that the results for the full-data unquantized regime immediately follow by setting $\Delta = 0$, thus Theorems 2-3 represent the strict extension of [36, Sect. 4], and Theorem 4 complements [36] with a high-dimensional sparse setting.

As the parameter choices in Theorems 2-4 rely on M that somehow connects to $\|\boldsymbol{\Sigma}^\star\|_\infty$, our methods require certain prior estimate of the unknown $\boldsymbol{\Sigma}^\star$, or otherwise a careful tuning of the parameter. It would be interesting to investigate how to address this in practice.

Remark 1. (Sub-Gaussian Case)**.** *While we concentrate on the quantization of heavy-tailed data in this work, our results can be readily adjusted to sub-Gaussian $\boldsymbol{x}_k$, for which the truncation step is inessential and can be removed (i.e., $\zeta=\infty$). These results are also new to the literature but will not be presented here.*

B. Quantized Compressed Sensing

We consider the linear model

$$y_k = \boldsymbol{x}_k^\top\boldsymbol{\theta}^\star + \epsilon_k,\ k=1,...,n, \tag{8}$$

where $\boldsymbol{x}_k$s are the covariates, y_ks are responses, $\boldsymbol{\theta}^\star$ is the sparse signal in compressed sensing or sparse parameter vector in high-dimensional linear regression that we want to estimate. In the quantized compressed sensing (QCS) problem, we are interested in *developing quantization scheme for $(\boldsymbol{x}_k, y_k)$s (mainly for y_k in prior works) that enables accurate recovery of $\boldsymbol{\theta}^\star$ based on the quantized data*.

In spite of the same mathematical formulation, there are some important differences between compressed sensing and sparse linear regression that we should clarify first. Specifically, different from sensing vectors in compressed sensing that are generated by some analog measuring device and can oftentimes be designed, $\boldsymbol{x}_k$s in sparse linear regression represent the sample data from certain datasets that are believed to affect the responses y_ks through (8). While the sparsity of $\boldsymbol{\theta}^\star$ is arguably the most classical signal structure for compressed sensing, due to good interpretability it is also commonly adopted to achieve dimension reduction in high-dimensional statistics. In this work, we are interested in both problem settings. Thus, we do not adopt the isotropic convention (i.e., $\mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top) = \boldsymbol{I}_d$) from compressed sensing but instead deal with $\boldsymbol{x}_k$ having general unknown covariance matrix. While the study of quantization and heavy-tailed noise is meaningful in both settings, we note that some of our subsequent results are mainly of interest to the specific sensing or regression problem. For instance, the heavy-tailed covariate considered in Theorem 6 is primarily motivated by the regression setting, in which $\boldsymbol{x}_k$ may come from a dataset that exhibits a much heavier tail than sub-Gaussian data. Moreover, as will be elaborated in Section IV when appropriate, our subsequent results on covariate quantization (resp., uniform signal recovery guarantee) may prove more useful to the regression problem (resp., compressed sensing problem).

To fix idea, we assume that $\boldsymbol{x}_k$s are i.i.d. drawn from some multi-variate distribution, ϵ_ks are i.i.d. statistical noise independent of the $\boldsymbol{x}_k$s, and we truncate y_k to $\widetilde{y}_k = \mathscr{T}_{\zeta_y}(y_k)$ and then quantize it to $\dot{y}_k = \mathcal{Q}_\Delta(\widetilde{y}_k+\tau_k)$ with uniform dither $\tau_k\sim\mathscr{U}([-\frac{\Delta}{2},\frac{\Delta}{2}])$. Under these statistical assumptions and dithered quantization, near-optimal recovery guarantees have been established in [89], [94] for the regime where both $\boldsymbol{x}_k$ and ϵ_k are drawn from sub-Gaussian distributions (hence the truncation is not needed). In contrast, our focus is on the quantization of heavy-tailed data. Particularly, we always assume that the noise ϵ_ks are i.i.d. drawn from some heavy-tailed distribution, resulting in heavy-tailed responses. We will separately deal with the case of sub-Gaussian covariate and a more challenging situation where $\boldsymbol{x}_k$s are also heavy-tailed.

To estimate the sparse $\boldsymbol{\theta}^\star$, a classical approach is via the regularized M-estimator known as Lasso [70], [72], [90]

$$\arg\min_{\boldsymbol{\theta}}\ \frac{1}{2n}\sum_{k=1}^n(y_k-\boldsymbol{x}_k^\top\boldsymbol{\theta})^2+\lambda\|\boldsymbol{\theta}\|_1,$$

whose objective combines the ℓ_2-loss for data fidelity and ℓ_1-norm that encourages sparsity. Because we can only access the quantized data $(\boldsymbol{x}_k,\dot{y}_k)$ (or even $(\dot{\boldsymbol{x}}_k,\dot{y}_k)$ if covariate quantization is involved, see Section IV), the main issue lies in the ℓ_2-loss $\frac{1}{2n}\sum_{k=1}^n(y_k-\boldsymbol{x}_k^\top\boldsymbol{\theta})^2$ that requires the unquantized

data $(\boldsymbol{x}_k, y_k)$. To resolve the issue, we calculate the expected ℓ_2-loss:

$$\mathbb{E}(y_k - \boldsymbol{x}_k^\top \boldsymbol{\theta})^2 \overset{(i)}{=} \boldsymbol{\theta}^\top \mathbb{E}(\boldsymbol{x}_k \boldsymbol{x}_k^\top)\boldsymbol{\theta} - 2\mathbb{E}(y_k \boldsymbol{x}_k)^\top \boldsymbol{\theta} : \overset{(ii)}{=} \boldsymbol{\theta}^\top \boldsymbol{\Sigma}^\star \boldsymbol{\theta} - 2\boldsymbol{\Sigma}_{y\boldsymbol{x}}^\top \boldsymbol{\theta}, \tag{9}$$

where (i) holds up to an inessential constant $\mathbb{E}|y_k|^2$, and in (ii) we let $\boldsymbol{\Sigma}^\star := \mathbb{E}(\boldsymbol{x}_k \boldsymbol{x}_k^\top)$, $\boldsymbol{\Sigma}_{y\boldsymbol{x}} = \mathbb{E}(y_k \boldsymbol{x}_k)$. This inspires us to generalize the ℓ_2 loss to $\frac{1}{2}\boldsymbol{\theta}^\top \boldsymbol{Q}\boldsymbol{\theta} - \boldsymbol{b}^\top \boldsymbol{\theta}$ and consider the following program

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \mathcal{S}}{\arg\min} \ \frac{1}{2}\boldsymbol{\theta}^\top \boldsymbol{Q}\boldsymbol{\theta} - \boldsymbol{b}^\top \boldsymbol{\theta} + \lambda \|\boldsymbol{\theta}\|_1. \tag{10}$$

Compared to (9) we will use $(\boldsymbol{Q}, \boldsymbol{b})$ that well approximates $(\boldsymbol{\Sigma}^\star, \boldsymbol{\Sigma}_{y\boldsymbol{x}})$, and we also introduce the constraint $\boldsymbol{\theta} \in \mathcal{S}$ to allow more flexibility. It is important to note that this is the general strategy in this work to design estimators in different QCS settings, see more discussions in Remark 3.

The next theorem is concerned with QCS under sub-Gaussian covariate but heavy-tailed response. Note that the heavy-tailedness of y_k stems from the noise distribution assumed to have bounded $2+\nu$ moment ($\nu = 2(l-1) > 0$ in the theorem statement), but following [22], [36], [97] we directly impose the moment constraint on the response.

Theorem 5. (Sub-Gaussian Covariate, Heavy-Tailed Response)**.** *Given some $\delta > 0, \Delta > 0$, in (8) we suppose that $\boldsymbol{x}_k$s are i.i.d., zero-mean sub-Gaussian with $\|\boldsymbol{x}_k\|_{\psi_2} \le \sigma$, $\kappa_0 \le \lambda_{\min}(\boldsymbol{\Sigma}^\star) \le \lambda_{\max}(\boldsymbol{\Sigma}^\star) \le \kappa_1$ for some $\kappa_1 > \kappa_0 > 0$ where $\boldsymbol{\Sigma}^\star = \mathbb{E}(\boldsymbol{x}_k \boldsymbol{x}_k^\top)$, $\boldsymbol{\theta}^\star \in \mathbb{R}^d$ is s-sparse, the noise ϵ_ks are i.i.d. heavy-tailed and independent of $\boldsymbol{x}_k$s, and we assume $\mathbb{E}|y_k|^{2l} \le M$ for some fixed $l > 1$. In the quantization, we truncate y_k to $\widetilde{y}_k = \mathscr{T}_{\zeta_y}(y_k)$ with threshold $\zeta_y \asymp \left(\frac{nM^{1/l}}{\delta \log d}\right)^{1/2}$, then quantize $\widetilde{y}_k$ to $\dot{y}_k = \mathcal{Q}_\Delta(\widetilde{y}_k + \tau_k)$ with uniform dither $\tau_k \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}])$. For recovery, we define the estimator $\widehat{\boldsymbol{\theta}}$ as (10) with*

$$\boldsymbol{Q} = \frac{1}{n}\sum_{k=1}^n \boldsymbol{x}_k \boldsymbol{x}_k^\top, \ \boldsymbol{b} = \frac{1}{n}\sum_{k=1}^n \dot{y}_k \boldsymbol{x}_k, \ \mathcal{S} = \mathbb{R}^d.$$

We set

$$\lambda = C_1 \frac{\sigma^2}{\sqrt{\kappa_0}}(\Delta + M^{1/(2l)})\sqrt{\frac{\delta \log d}{n}}$$

with sufficiently large C_1. If $n \gtrsim \delta s \log d$ for some hidden constant only depending on (κ_0, σ), then with probability at least $1 - 9d^{1-\delta}$, the estimation error $\widehat{\boldsymbol{\Upsilon}} = \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star$ satisfies

$$\|\widehat{\boldsymbol{\Upsilon}}\|_2 \le C_3 \mathscr{L}\sqrt{\frac{\delta s \log d}{n}} \text{ and } \|\widehat{\boldsymbol{\Upsilon}}\|_1 \le C_4 \mathscr{L} s\sqrt{\frac{\delta \log d}{n}}$$

where $\mathscr{L} := \frac{\sigma^2(\Delta + M^{1/(2l)})}{\kappa_0^{3/2}}$.

The rate $O\big(\sqrt{\frac{s \log d}{n}}\big)$ for ℓ_2-norm error is minimax optimal up to logarithmic factor (e.g., compared to [81]). Note that a random noise bounded by Δ roughly contributes Δ to $(\mathbb{E}|y_k|^{2l})^{1/(2l)}$, and the latter is bounded by $M^{1/(2l)}$; because in the error bound Δ and $M^{1/(2l)}$ almost play the same role, the effect of uniform quantization can be readily interpreted as an additional bounded noise, analogously to the error rate in [87].

Next, we switch to the more challenging situation where both $\boldsymbol{x}_k$ and y_k are heavy-tailed, assuming that they both possess bounded fourth moments (a marginal moment constraint for $\boldsymbol{x}_k$). The consideration of this setting is motivated by the setting of sparse linear regression, where the covariates $\boldsymbol{x}_k$s may oftentimes exhibit heavy-tailed behavior. Specifically, we element-wisely truncate $\boldsymbol{x}_k$ to $\widetilde{\boldsymbol{x}}_k$ and set $\boldsymbol{Q} := \frac{1}{n}\sum_{k=1}^n \widetilde{\boldsymbol{x}}_k \widetilde{\boldsymbol{x}}_k^\top$ as a robust covariance matrix estimator, whose estimation performance under $\|\cdot\|_\infty$ follows immediately from Theorem 2 by setting $\Delta = 0$.

Theorem 6. (Heavy-Tailed Covariate, Heavy-Tailed Response)**.** *Given some $\delta > 0$, $\Delta > 0$, in (8) we suppose that $\boldsymbol{x}_k$s are i.i.d. zero-mean satisfying a marginal fourth moment constraint $\sup_{i\in[d]} \mathbb{E}|x_{ki}|^4 \le M$, $\kappa_0 \le \lambda_{\min}(\boldsymbol{\Sigma}^\star) \le \lambda_{\max}(\boldsymbol{\Sigma}^\star) \le \kappa_1$ for some $\kappa_1 > \kappa_0 > 0$ where $\boldsymbol{\Sigma}^\star = \mathbb{E}(\boldsymbol{x}_k \boldsymbol{x}_k^\top)$, $\boldsymbol{\theta}^\star \in \Sigma_s$ satisfies $\|\boldsymbol{\theta}^\star\|_1 \le R$, the noise ϵ_ks are i.i.d. heavy-tailed and independent of $\boldsymbol{x}_k$s, and we assume $\mathbb{E}|y_k|^4 \le M$. In the quantization, we truncate $\boldsymbol{x}_k, y_k$ respectively to $\widetilde{\boldsymbol{x}}_k = [\widetilde{x}_{ki}] = \mathscr{T}_{\zeta_x}(\boldsymbol{x}_k)$, $\widetilde{y}_k := \mathscr{T}_{\zeta_y}(y_k)$ with $\zeta_x, \zeta_y \asymp \left(\frac{nM}{\delta \log d}\right)^{1/4}$, then we quantize $\widetilde{y}_k$ to $\dot{y}_k = \mathcal{Q}_\Delta(\widetilde{y}_k + \tau_k)$ with uniform dither $\tau_k \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}])$. For recovery, we define the estimator $\widehat{\boldsymbol{\theta}}$ as (10) with*

$$\boldsymbol{Q} = \frac{1}{n}\sum_{k=1}^n \widetilde{\boldsymbol{x}}_k \widetilde{\boldsymbol{x}}_k^\top, \ \boldsymbol{b} = \frac{1}{n}\sum_{k=1}^n \dot{y}_k \widetilde{\boldsymbol{x}}_k, \ \mathcal{S} = \mathbb{R}^d.$$

We set

$$\lambda = C_1(R\sqrt{M} + \Delta^2)\sqrt{\frac{\delta \log d}{n}}$$

with sufficiently large C_1. If $n \gtrsim \delta s^2 \log d$ for some hidden constant only depending on (κ_0, M), then with probability at least $1 - 4d^{2-\delta}$, the estimation error $\widehat{\boldsymbol{\Upsilon}} := \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star$ satisfies

$$\|\widehat{\boldsymbol{\Upsilon}}\|_2 \le C_2 \mathscr{L}\sqrt{\frac{\delta s \log d}{n}} \text{ and } \|\widehat{\boldsymbol{\Upsilon}}\|_1 \le C_3 \mathscr{L} s\sqrt{\frac{\delta \log d}{n}}$$

where $\mathscr{L} := \frac{R\sqrt{M} + \Delta^2}{\kappa_0}$.

Theorem 6 generalizes [36, Thm. 2(b)] to the uniform quantization setting. Clearly, the obtained rate remains near minimax optimal if R is of minor scaling (e.g., bounded or logarithmic factors). Nevertheless, such near optimality in Theorem 6 comes at the cost of more restricted conditions and stronger scaling, as remarked in the following.

Remark 2. (Comparing Theorems 5-6)**.** *Compared with $n \gtrsim s \log d$ in Theorem 5, the first downside of Theorem 6 is the sub-optimal sample complexity $n \gtrsim s^2 \log d$, and note that $n \gtrsim s^2 \log d$ is also required in [36, Thm. 2(b)]. But indeed, it can be improved to $n \gtrsim s \log d$ by explicitly adding the constraint $\|\boldsymbol{\theta}\|_1 \le R$ to the recovery program, as will be noted as an interesting side finding in Remark 6. Secondly, following [36] we impose an ℓ_1-norm constraint $\|\boldsymbol{\theta}^\star\|_1 \le R$ that is stronger than $\|\boldsymbol{\theta}^\star\|_2 \lesssim \frac{M^{1/(2l)}}{\sigma}$ used in the proof of Theorem 5. In fact, when replacing the ℓ_1 constraint in Theorem 6 with an ℓ_2-norm bound $\|\boldsymbol{\theta}^\star\|_2 \le R$, then our proof technique leads*

to an error rate $\|\widehat{\boldsymbol{\Upsilon}}\|_2 = O\big(\sqrt{\frac{s^2\log d}{n}}\big)$ *that exhibits worse dependence on* s.

Remark 3. (Modification of ℓ_2-loss)**.** *Recall that we generalize the regular* ℓ_2*-loss* $\frac{1}{2n}\sum_{k=1}^n(y_k-\boldsymbol{x}_k^\top\boldsymbol{\theta})^2$ *to* $\frac{1}{2}\boldsymbol{\theta}^\top\boldsymbol{Q}\boldsymbol{\theta}-\boldsymbol{b}^\top\boldsymbol{\theta}$ *as loss function in (10). Note that the choice of* $(\boldsymbol{Q},\boldsymbol{b})$ *in Theorem 5 is tantamount to using the loss function* $\frac{1}{2n}\sum_{k=1}^n(\dot{y}_k-\boldsymbol{x}_k^\top\boldsymbol{\theta})^2$ *that replaces* y_k *with the quantized response* $\dot{y}_k$*; this idea is analogous to the generalized Lasso investigated for single index model [78] and dithered quantized model [89], and will be used again in quantized matrix completion, see (12) below. However, our generalized* ℓ_2*-loss provides more flexibility to deal with heavy-tailedness or quantization of* $\boldsymbol{x}_k$*, e.g.,* $(\boldsymbol{Q},\boldsymbol{b})$ *in Theorem 6 amounts to adopting* $\frac{1}{2n}\sum_{k=1}^n(\dot{y}_k-\widetilde{\boldsymbol{x}}_k^\top\boldsymbol{\theta})^2$ *as loss function, and under quantized covariate more delicate modifications are required in Theorems 9-12, which is beyond the range of prior works on generalized Lasso.*

C. Quantized Matrix Completion

Completing a low-rank matrix from only a partial observation of its entries is known as the matrix completion problem, which has found many applications including recommendation systems, image inpainting, quantum state tomography [2], [18], [27], [42], [74], to name just a few. Mathematically, let $\boldsymbol{\Theta}^\star\in\mathbb{R}^{d\times d}$ be the underlying matrix satisfying $\mathrm{rank}(\boldsymbol{\Theta}^\star)\le r$, the matrix completion problem can be formulated as

$$y_k=\big\langle\boldsymbol{X}_k,\boldsymbol{\Theta}^\star\big\rangle+\epsilon_k,\ \ k=1,2,...,n, \tag{11}$$

where $\boldsymbol{X}_k$s are distributed on $\mathcal{X}:=\{\boldsymbol{e}_i\boldsymbol{e}_j^\top:i,j\in[d]\}$ ($\boldsymbol{e}_i$ is the i-th column of $\boldsymbol{I}_d$), ϵ_k is observation noise. Note that for $\boldsymbol{X}_k=\boldsymbol{e}_{i(k)}\boldsymbol{e}_{j(k)}^\top$ one has $\big\langle\boldsymbol{X}_k,\boldsymbol{\Theta}^\star\big\rangle=\theta^\star_{i(k),j(k)}$, so each observation is a noisy entry. Our main interest is in quantized matrix completion (QMC), where our goal is to *design a quantizer for the observation* y_k *that allows for accurate estimation of* $\boldsymbol{\Theta}^\star$ *from the quantized observations*.

Unlike in compressed sensing, additional condition (besides the low-rankness) on $\boldsymbol{\Theta}^\star$ is needed to ensure the well-posedness of the matrix completion problem. More specifically, certain incoherence conditions are required if we pursue exact recovery (e.g., [15], [82]), whereas a faithful estimation can be achieved as long as the underlying matrix is not overly spiky and sufficiently diffuse (e.g., [51], [71]). The latter condition is also known as "low spikiness" and is formulated by $\frac{d\|\boldsymbol{\Theta}^\star\|_\infty}{\|\boldsymbol{\Theta}^\star\|_F}\le\alpha$ [36], [71], which has been noted to be necessary for the well-posedness of matrix completion problem [27], [71]. In subsequent works, the low-spikiness condition is often formulated as the simpler max-norm constraint $\|\boldsymbol{\Theta}^\star\|_\infty\le\alpha$ [18], [26], [37], [51], [53].

In this work, we consider the uniform sampling scheme $\boldsymbol{X}_k\sim\mathscr{U}(\mathcal{X})$, but with a little bit more work it generalizes to a more general sampling scheme [51]. We apply the proposed quantization scheme to possibly heavy-tailed y_k — we truncate y_k to $\widetilde{y}_k=\mathscr{T}_{\zeta_y}(y_k)$ with some threshold ζ_y, and then quantize $\widetilde{y}_k$ to $\dot{y}_k=\mathcal{Q}_\Delta(\widetilde{y}_k+\tau_k)$ with uniform dither $\tau_k\sim\mathscr{U}([-\frac{\Delta}{2},\frac{\Delta}{2}])$. Because we do not pursue exact recovery (which is impossible under quantization), we do not assume any incoherence condition like [82]. Instead, we only hope to accurately estimate $\boldsymbol{\Theta}^\star$, and following [18], [26], [37], [51], [53] we impose a max-norm constraint

$$\|\boldsymbol{\Theta}^\star\|_\infty\le\alpha.$$

Overall, we estimate $\boldsymbol{\Theta}^\star$ from $(\boldsymbol{X}_k,\dot{y}_k)$ by the regularized M-estimator [70], [72]

$$\widehat{\boldsymbol{\Theta}}=\underset{\|\boldsymbol{\Theta}\|_\infty\le\alpha}{\arg\min}\ \frac{1}{2n}\sum_{k=1}^n\big(\dot{y}_k-\big\langle\boldsymbol{X}_k,\boldsymbol{\Theta}\big\rangle\big)^2+\lambda\|\boldsymbol{\Theta}\|_{nu} \tag{12}$$

that combines an ℓ_2-loss and nuclear norm regularizer.

In the literature, there has been a line of works on 1-bit or multi-bit matrix completion related to our results to be presented [3], [14], [16], [52], [59]. While the referenced works commonly adopted a likelihood approach, our method is an essential departure and embraces some advantage, see a precise comparison in Remark 4. Considering such novelty, we include the result for sub-exponential ϵ_k in Theorem 7, for which the truncation of y_k becomes unnecessary and we simply set $\zeta_y=\infty$.

Theorem 7. (QMC under Sub-Exponential Noise)**.** *Given some* $\Delta>0,\delta>0$*, in (11) we suppose that* $\boldsymbol{X}_k$*s are i.i.d. uniformly distributed over* $\mathcal{X}=\{\boldsymbol{e}_i\boldsymbol{e}_j^\top:i,j\in[d]\}$*,* $\boldsymbol{\Theta}^\star\in\mathbb{R}^{d\times d}$ *satisfies* $\mathrm{rank}(\boldsymbol{\Theta}^\star)\le r$ *and* $\|\boldsymbol{\Theta}^\star\|_\infty\le\alpha$*, the noise* ϵ_k*s are i.i.d. zero-mean sub-exponential satisfying* $\|\epsilon_k\|_{\psi_1}\le\sigma$*, and are independent of* $\boldsymbol{X}_k$*s. In the quantization, we do not truncation* y_k *but directly quantize it to* $\dot{y}_k=\mathcal{Q}_\Delta(y_k+\tau_k)$ *with uniform dither* $\tau_k\sim\mathscr{U}([-\frac{\Delta}{2},\frac{\Delta}{2}])$*. We choose* $\lambda=C_1(\sigma+\Delta)\sqrt{\frac{\delta\log d}{nd}}$ *with sufficiently large* C_1*, and define* $\widehat{\boldsymbol{\Theta}}$ *as (12). If* $\delta d\log^3 d\lesssim n\lesssim\delta r^2d^2\log d$*, then with probability at least* $1-4d^{-\delta}$*, the estimation error* $\widehat{\boldsymbol{\Upsilon}}:=\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}^\star$ *satisfies*

$$\frac{\|\widehat{\boldsymbol{\Upsilon}}\|_F}{d}\le C_2\mathscr{L}\sqrt{\frac{\delta rd\log d}{n}}\ \text{ and }\ \frac{\|\widehat{\boldsymbol{\Upsilon}}\|_{nu}}{d}\le C_3\mathscr{L}r\sqrt{\frac{\delta d\log d}{n}}$$

where $\mathscr{L}:=\alpha+\sigma+\Delta$.

By contrast, under heavy-tailed noise only assumed to have bounded variance, we truncate y_k with a suitable threshold before the dithered quantization to achieve an optimal trade-off between bias and variance.

Theorem 8. (QMC under Heavy-tailed Noise)**.** *Given some* $\Delta>0,\delta>0$*, we consider (11) in the setting of Theorem 7 but with the assumption* $\|\epsilon_k\|_{\psi_1}\le\sigma$ *replaced by* $\mathbb{E}|\epsilon_k|^2\le M$*. In the quantization, we truncate* y_k *to* $\widetilde{y}_k=\mathscr{T}_{\zeta_y}(y_k)$ *with* $\zeta_y\asymp(\sqrt{M}+\alpha)\sqrt{\frac{n}{\delta d\log d}}$*, and then quantize* $\widetilde{y}_k$ *to* $\dot{y}_k=\mathcal{Q}_\Delta(\widetilde{y}_k+\tau_k)$ *with uniform dither* $\tau_k\sim\mathscr{U}([-\frac{\Delta}{2},\frac{\Delta}{2}])$*. We choose* $\lambda=C_1(\alpha+\sqrt{M}+\Delta)\sqrt{\frac{\delta\log d}{nd}}$ *with sufficiently large* C_1*, and define* $\widehat{\boldsymbol{\Theta}}$ *as (12). If* $\delta d\log d\lesssim n\lesssim\delta r^2d^2\log d$*, then with probability at least* $1-6d^{-\delta}$*, the estimation error* $\widehat{\boldsymbol{\Upsilon}}:=\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}^\star$ *satisfies*

$$\frac{\|\widehat{\boldsymbol{\Upsilon}}\|_F}{d}\le C_2\mathscr{L}\sqrt{\frac{\delta rd\log d}{n}}\ \text{ and }\ \frac{\|\widehat{\boldsymbol{\Upsilon}}\|_{nu}}{d}\le C_3\mathscr{L}r\sqrt{\frac{\delta d\log d}{n}}$$

where $\mathscr{L}:=\alpha+\sqrt{M}+\Delta$.

Compared to the information-theoretic lower bounds in [55], [71], the error rates obtained in Theorems 7-8 are minimax

optimal up to logarithmic factors. Specifically, Theorem 8 derives a near optimal guarantee for QMC with heavy-tailed observations, as the key standpoint of this paper. Note that, the 1-bit quantization counterpart of these two Theorems was derived in our previous work [22]; in sharp contrast to Theorem 8, for 1-bit QMC under heavy-tailed noise, the error rate under $\frac{\|\widehat{\boldsymbol{\Upsilon}}\|_F}{d}$ in [22, Thm. 13] reads as $O\big(\big(\frac{r^2 d\log d}{n}\big)^{1/4}\big)$ and is essentially slower; using the 1-bit observations therein, this slow error rate is indeed nearly tight due to the lower bound in [22, Thm. 14].

To close this section, we give a remark to illustrate the novelty and advantage of our QMC method by a careful comparison with prior works.

Remark 4. *QMC with 1-bit or multi-bit quantized observations has received considerable research interest [3], [14], [16], [26], [52], [59]. Adapted to our notation, these works studied the model $\dot{y}_k = \mathcal{Q}(\langle \boldsymbol{X}_k, \boldsymbol{\Theta}^\star\rangle + \tau_k)$ under general random dither τ_k and quantizer $\mathcal{Q}(.)$, and they commonly adopted regularized (or constrained) maximum likelihood estimation for estimating $\boldsymbol{\Theta}^\star$. By contrast, with the random dither and quantizer specialized to $\tau_k \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}])$ and $\mathcal{Q}_\Delta(.)$, our model is formulated as $\dot{y}_k = \mathcal{Q}_\Delta(\mathscr{T}_{\zeta_y}(\langle \boldsymbol{X}_k, \boldsymbol{\Theta}^\star\rangle + \epsilon_k) + \tau_k)$. Thus, while suffering from less generality in $(\tau_k, \mathcal{Q})$, our method embraces the advantage of robustness to pre-quantization noise ϵ_k, whose distribution is unknown and can even be heavy-tailed. Note that such unknown ϵ_k evidently forbids the likelihood approach.*

IV. Covariate Quantization and Uniform Signal Recovery in Quantized Compressed Sensing

By now we have presented near optimal results in the contexts of QCME, QCS, and QMC under heavy-tailed data that further undergo the proposed quantization scheme, which we position as the primary contribution of this work. In this section, we further provide two additional developments to enhance our results on heavy-tailed QCS.

A. Covariate Quantization

In the area of QCS, almost all prior works merely focused on the quantization of response y_k, see the recent survey [29]; here, we consider a setting of "complete quantization" — meaning that the covariate $\boldsymbol{x}_k$ is also quantized. To motivate our study of "complete quantization", we interpret compressed sensing as sparse linear regression. Indeed, to reduce the power consumption and computational cost, it is sometimes preferable to work with low-precision data in a machine learning system, e.g., the sample quantization scheme developed in [96] led to experimental success in training linear model. Also, it was shown that direct gradient quantization may not be efficient in certain distributed learning systems where the terminal nodes are connected to the server only through a very weak communication fabric and the number of parameters is extremely huge; rather, quantizing and transmitting some important samples could provably reduce communication cost [43]. In fact, the process of data collection may already appeal to quantization due to certain limits of the data acquisition device (e.g., a low-resolution analog-to-digital module used in distributed signal processing [25]). Our main goal is to understand how quantization of $(\boldsymbol{x}_k, y_k)$s affects the subsequent recovery/learning process, particularly showing that the simple dithered uniform quantization scheme still allows for an accurate estimator that may even provide near minimax error rate. To our best knowledge, the only prior rigorous estimation guarantees for QCS with covariate quantization are [22, Thms. 7-8]; these two results require a restricted and unnatural assumption, which we will also relax later.

1) Multi-bit QCS with Quantized Covariate: Since we will also consider the 1-bit quantization, we more precisely refer to the QCS under uniform quantizer as multi-bit QCS. We will generalize Theorems 5-6 to covariate quantization in the next two theorems.

Let $(\dot{\boldsymbol{x}}_k, \dot{y}_k)$ be the quantized covariate-response pair, we first quickly sketch the idea of our approach. Specifically, we stick to the framework of M-estimator in (10), which appeals to accurate surrogates for $\boldsymbol{\Sigma}^\star = \mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top)$ and $\boldsymbol{\Sigma}_{y\boldsymbol{x}} = \mathbb{E}(y_k\boldsymbol{x}_k)$ based on $(\dot{\boldsymbol{x}}_k, \dot{y}_k)$, where $\dot{\boldsymbol{x}}_{\boldsymbol{k}}$ represents the quantized covariate. Fortunately, the surrogates can be constructed analogously to our QCME estimator when triangular dither is used for quantizing $\boldsymbol{x}_k$. Let us first state our quantization scheme as follows:

- **Response Quantization.** This is the same as Theorems 5-6. We truncate y_k to $\widetilde{y}_k = \mathscr{T}_{\zeta_y}(y_k)$ with threshold ζ_y, and then quantize $\widetilde{y}_k$ to $\dot{y}_k = \mathcal{Q}_\Delta(\widetilde{y}_k + \phi_k)$ with uniform dither $\phi_k \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}])$ and quantization level $\Delta \geq 0$.
- **Covariate Quantization.** This is the same as Theorem 2. We truncate $\boldsymbol{x}_k$ to $\widetilde{\boldsymbol{x}}_k = \mathscr{T}_{\zeta_x}(\boldsymbol{x}_k)$ with threshold ζ_x, and then quantize $\widetilde{\boldsymbol{x}}_k$ to $\dot{\boldsymbol{x}}_k = \mathcal{Q}_{\bar{\Delta}}(\widetilde{\boldsymbol{x}}_k + \boldsymbol{\tau}_k)$ with triangular dither $\boldsymbol{\tau}_k \sim \mathscr{U}([-\frac{\bar{\Delta}}{2}, \frac{\bar{\Delta}}{2}]^d) + \mathscr{U}([-\frac{\bar{\Delta}}{2}, \frac{\bar{\Delta}}{2}]^d)$ and quantization level $\bar{\Delta} \geq 0$.
- **Notation.** We write the quantization noise as $\varphi_k = \dot{y}_k - \widetilde{y}_k$ and $\boldsymbol{\xi}_k = \dot{\boldsymbol{x}}_k - \widetilde{\boldsymbol{x}}_k$, the quantization error as $\vartheta_k = \dot{y}_k - (\widetilde{y}_k + \phi_k)$ and $\boldsymbol{w}_k = \dot{\boldsymbol{x}}_k - (\widetilde{\boldsymbol{x}}_k + \boldsymbol{\tau}_k)$.

We will adopt the above notation in subsequent developments. Based on the quantized covariate-response pairs $(\dot{\boldsymbol{x}}_k, \dot{y}_k)$s, we specify our estimator by setting $(\boldsymbol{Q}, \boldsymbol{b})$ in (10) as

$$\boldsymbol{Q} = \frac{1}{n}\sum_{k=1}^n \dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top - \frac{\bar{\Delta}^2}{4}\boldsymbol{I}_d \quad \text{and} \quad \boldsymbol{b} = \frac{1}{n}\sum_{k=1}^n \dot{y}_k\dot{\boldsymbol{x}}_k. \tag{13}$$

Note that the choice of $\boldsymbol{Q}$ is due to the estimator in Theorem 2, while $\boldsymbol{b}$ is inspired by the calculation

$$\begin{aligned}\mathbb{E}(\dot{y}_k\dot{\boldsymbol{x}}_k) &= \mathbb{E}\big((\widetilde{y}_k + \varphi_k)(\widetilde{\boldsymbol{x}}_k + \boldsymbol{\xi}_k)\big)\\ &= \mathbb{E}(\widetilde{y}_k\widetilde{\boldsymbol{x}}_k) + \mathbb{E}(\widetilde{y}_k\boldsymbol{\xi}_k) + \mathbb{E}(\varphi_k\widetilde{\boldsymbol{x}}_k) + \mathbb{E}(\varphi_k\boldsymbol{\xi}_k)\\ &= \mathbb{E}(\widetilde{y}_k\widetilde{\boldsymbol{x}}_k),\end{aligned}$$

where the last equality can be seen by conditioning on $\widetilde{\boldsymbol{x}}_k$ or $\widetilde{y}_k$. However, the issue is that $\boldsymbol{Q}$ is not positive semi-definite, hence the resulting program is non-convex. To explain this, note that the rank of $\frac{1}{n}\sum_{k=1}^n \dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top$ does not exceed n, so when $d > n$ at least $d - n$ eigenvalues of $\boldsymbol{Q}$ are $-\frac{\bar{\Delta}^2}{4}$. Alternatively, the non-convexity can also be seen from the

observation that setting $(\boldsymbol{Q}, \boldsymbol{b})$ as in (13) is tantamount to replacing the regular ℓ_2-loss $\frac{1}{2n}\sum_{k=1}^n (y_k - \boldsymbol{x}_k^\top \boldsymbol{\theta})^2$ with

$$\frac{1}{2n}\sum_{k=1}^n (\dot{y}_k - \dot{\boldsymbol{x}}_k^\top \boldsymbol{\theta})^2 - \frac{\bar{\Delta}}{8}\|\boldsymbol{\theta}\|_2^2.$$

We mention that the lack of positive semi-definiteness of $\boldsymbol{Q}$ is problematic in both statistics and optimization aspects: 1) Statistically, Lemma 4 used to derive the error rates in Theorems 5-6 requires $\boldsymbol{Q}$ to be positive semi-definite, and is hence no longer applicable here; 2) From the optimization side, it is, in general, unknown how to globally optimize a non-convex program.

Motivated by a line of previous works on non-convex M-estimator [62]–[64], we add an ℓ_1-norm constraint to (10) by setting $\mathcal{S} = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_1 \le R\}$, where R represents the prior estimation on $\|\boldsymbol{\theta}^\star\|_1$. Let $\partial\|\boldsymbol{\theta}_1\|_1$ be a subdifferential of $\|\boldsymbol{\theta}\|_1$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_1$,[10] we consider the local minimizer of the proposed recovery program,[11] or more generally put, $\widetilde{\boldsymbol{\theta}} \in \mathcal{S}$ that satisfies[12]

$$\big\langle \boldsymbol{Q}\widetilde{\boldsymbol{\theta}} - \boldsymbol{b} + \lambda \cdot \partial\|\widetilde{\boldsymbol{\theta}}\|_1, \boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}} \big\rangle \ge 0, \quad \forall\, \boldsymbol{\theta} \in \mathcal{S}. \tag{14}$$

We will prove a fairly strong guarantee stating that all $\widetilde{\boldsymbol{\theta}} \in \mathcal{S}$ satisfying (14) (of course including all local minimizers) enjoy near minimax error rate. While this guarantee bears resemblance to the ones in [64], we point out that, [64] only derived concrete results for the sub-Gaussian regime; because of the heavy-tailed data and quantization in our setting, some essentially different ingredients are required for the technical analysis (see Remark 5). As before, our results for sub-Gaussian $\boldsymbol{x}_k$ and heavy-tailed $\boldsymbol{x}_k$ are presented separately.

Theorem 9. (Quantized Sub-Gaussian Covariate)**.** *Given $\Delta \ge 0$, $\bar{\Delta} \ge 0$, $\delta > 0$, we consider (8) with the same assumptions on $(\boldsymbol{x}_k, y_k, \boldsymbol{\theta}^\star)$ as Theorem 5, and additionally assume that $\|\boldsymbol{\theta}^\star\|_2 \le R$. The quantization of $(\boldsymbol{x}_k, y_k)$ is described above, and we set $\zeta_x = \infty$, $\zeta_y \asymp \sqrt{\frac{nM^{1/l}}{\delta \log d}}$. For recovery, we let*

$$\boldsymbol{Q} = \frac{1}{n}\sum_{k=1}^n \dot{\boldsymbol{x}}_k \dot{\boldsymbol{x}}_k^\top - \frac{\bar{\Delta}^2}{4}\boldsymbol{I}_d,$$
$$\boldsymbol{b} = \frac{1}{n}\sum_{k=1}^n \dot{y}_k \dot{\boldsymbol{x}}_k, \ \mathcal{S} = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\|_1 \le R\sqrt{s}\},$$

and set

$$\lambda = C_1 \frac{(\sigma + \bar{\Delta})^2}{\sqrt{\kappa_0}}(\Delta + M^{1/(2l)})\sqrt{\frac{\delta \log d}{n}}$$

with sufficiently large C_1. If $n \gtrsim \delta s \log d$ for some hidden constant only depending on $(\kappa_0, \sigma, \Delta, \bar{\Delta}, M, R)$, with probability at least $1 - 8d^{1-\delta} - C_2\exp(-C_3 n)$, all $\widetilde{\boldsymbol{\theta}} \in \mathcal{S}$ satisfying (14) have estimation error $\widetilde{\boldsymbol{\Upsilon}} := \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star$ bounded by

$$\|\widetilde{\boldsymbol{\Upsilon}}\|_2 \le C\mathscr{L}\sqrt{\frac{\delta s \log d}{n}} \ \text{ and } \ \|\widetilde{\boldsymbol{\Upsilon}}\|_1 \le C'\mathscr{L}s\sqrt{\frac{\delta \log d}{n}}$$

where $\mathscr{L} := \frac{(\sigma+\bar{\Delta})^2(\boldsymbol{\Delta}+M^{1/(2l)})}{\kappa_0^{3/2}}$.

Similarly, the next result extends Theorem 6 to a setting involving covariate quantization.

Theorem 10. (Quantized Heavy-Tailed Covariate)**.** *Given $\Delta \ge 0$, $\bar{\Delta} \ge 0$, $\delta > 0$, we consider (8) with the same assumptions on $(\boldsymbol{x}_k, y_k, \boldsymbol{\theta}^\star)$ as Theorem 6. The quantization of $(\boldsymbol{x}_k, y_k)$ is described above, and we set $\zeta_x, \zeta_y \asymp \big(\frac{nM}{\delta \log d}\big)^{1/4}$. For recovery, we let*

$$\boldsymbol{Q} = \frac{1}{n}\sum_{k=1}^n \dot{\boldsymbol{x}}_k \dot{\boldsymbol{x}}_k^\top - \frac{\bar{\Delta}^2}{4}\boldsymbol{I}_d,$$
$$\boldsymbol{b} = \frac{1}{n}\sum_{k=1}^n \dot{y}_k \dot{\boldsymbol{x}}_k, \ \mathcal{S} = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\|_1 \le R\},$$

and set

$$\lambda = C_1(R\sqrt{M} + \Delta^2 + R\bar{\Delta}^2)\sqrt{\frac{\delta \log d}{n}}$$

with sufficiently large C_1. If $n \gtrsim \delta s \log d$ for some hidden constant only depending on (κ_0, M), then with probability at least $1 - 8d^{1-\delta}$, all $\widetilde{\boldsymbol{\theta}} \in \mathcal{S}$ satisfying (14) have estimation error $\widetilde{\boldsymbol{\Upsilon}} := \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star$ bounded by

$$\|\widetilde{\boldsymbol{\Upsilon}}\|_2 \le C_3\mathscr{L}\sqrt{\frac{\delta s \log d}{n}} \ \text{ and } \ \|\widetilde{\boldsymbol{\Upsilon}}\|_1 \le C_4\mathscr{L}s\sqrt{\frac{\delta \log d}{n}}.$$

where $\mathscr{L} := \frac{R\sqrt{M}+\Delta^2+R\bar{\Delta}^2}{\kappa_0}$.

Remark 5. (Comparing Our Analyses with [64])**.** *The above results are motivated by a line of works on nonconvex M-estimator [62]–[64], and our guarantee for the whole set of stationary points (14) resembles [64] most. While the main strategy for proving Theorem 9 is adjusted from [64], the proof of Theorem 10 does involve an essentially different RSC condition, see our (42). In particular, compared with [64, equation (4)], the leading factor of $\|\widetilde{\boldsymbol{\Upsilon}}\|_1^2$ in (42) degrades from $O\big(\frac{\log d}{n}\big)$ to $O\big(\sqrt{\frac{\log d}{n}}\big)$. To retain near optimal rate we need to impose a stronger scaling $\|\boldsymbol{\theta}^\star\|_1 \le R$ with proper changes in the proof. Although Theorem 10 is presented for a concrete setting, it sheds light on an extension of [64] to a weaker RSC condition that could accommodate covariate with a heavier tail. Such extension is formally presented as a deterministic framework in Proposition 1.*

Proposition 1. *Suppose that the s-sparse $\boldsymbol{\theta}^\star \in \mathbb{R}^d$ satisfies $\|\boldsymbol{\theta}^\star\|_1 \le R$, and the positive definite matrix $\boldsymbol{\Sigma}^\star \in \mathbb{R}^{d\times d}$ satisfies $\lambda_{\min}(\boldsymbol{\Sigma}^\star) \ge \kappa_0$. If for some $\boldsymbol{Q} \in \mathbb{R}^{d\times d}, \boldsymbol{b} \in \mathbb{R}^d$ we have*

$$\lambda \ge C_1 \max\big\{\|\boldsymbol{Q}\boldsymbol{\theta}^\star - \boldsymbol{b}\|_\infty, \ R \cdot \|\boldsymbol{Q} - \boldsymbol{\Sigma}^\star\|_\infty\big\} \tag{15}$$

holds for sufficiently large C_1, then all $\widetilde{\boldsymbol{\theta}} \in \mathcal{S}$ satisfying (14) with $\mathcal{S} = \{\boldsymbol{\theta} \in \mathbb{R}^d : \|\boldsymbol{\theta}\|_1 \le R\}$ have estimation error $\widetilde{\boldsymbol{\Upsilon}} := \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star$ bounded by

$$\|\widetilde{\boldsymbol{\Upsilon}}\|_2 \le C_2\frac{\sqrt{s}\lambda}{\kappa_0} \ \text{ and } \ \|\widetilde{\boldsymbol{\Upsilon}}\|_1 \le C_3\frac{s\lambda}{\kappa_0}.$$

By extracting the ingredients that guarantee (14) to be accurate, interestingly, Proposition 1 is now independent of

[10]Thus, $\partial\|\widetilde{\boldsymbol{\theta}}\|_1$ in (14) below should be understood as "there exists one element in $\partial\|\widetilde{\boldsymbol{\theta}}\|_1$ such that (14) holds."

[11]The existence of local minimizer is guaranteed because of the additional ℓ_1-constraint.

[12]To distinguish the global minimizer in (10), we denote by $\widetilde{\boldsymbol{\theta}}$ the estimator in QCS with quantized covariate.

the model assumption (8). Particularly, we could set $\boldsymbol{\Sigma}^\star = \mathbb{E}[\boldsymbol{x}_k\boldsymbol{x}_k^\top]$ when we apply Proposition 1 to (8). Compared with the framework [64, Thm. 1], the key strength of Proposition 1 is that it does not explicitly assume the RSC condition on the loss function that is hard to verify without assuming a sub-Gaussian covariate. Instead, the role of the RSC assumption in [64] is now played by $\lambda \gtrsim R\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty$, which immediately yields a kind of RSC condition by simple argument as (43). Although this RSC condition is often essentially weaker than the conventional one in terms of the leading factor of $\|\widetilde{\boldsymbol{\Upsilon}}\|_1^2$ (see Remark 5), along this line one can still derive non-trivial (or even near optimal) error rate. The gain of replacing RSC assumption with $\lambda \gtrsim R\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty$ is that the latter amounts to constructing element-wise estimator for $\boldsymbol{\Sigma}^\star$, which is often much easier for heavy-tailed covariate (e.g., due to many existing robust covariance estimators).

We conclude this part with a side interesting observation.

Remark 6. *By setting $\bar{\Delta} = 0$, Theorem 10 produces a result (with a convex program) for the setting of Theorem 6. Interestingly, with the additional ℓ_1-constraint, a notable improvement is that the sub-optimal $n \gtrsim s^2 \log d$ in Theorem 6 is sharpened to the near optimal one in Theorem 10. More concretely, this is because (ii) in (30) can be tightened to (ii) of (43). Going back to the full-data unquantized regime, Theorem 10 with $\Delta = \bar{\Delta} = 0$ recovers [36, Theorem 2(b)] with improved sample complexity requirement.*

2) 1-bit QCS with Quantized Covariate: Our consideration of covariate quantization in QCS seems fairly new to the literature. To the best of our knowledge, the only related results are [22, Thms. 7-8] for QCS with 1-bit quantized covariate and response. The assumption there, however, is quite restrictive. Specifically, it is assumed that $\boldsymbol{\Sigma}^\star = \mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top)$ has sparse columns (see [22, Assumption 3]), which is non-standard in both compressed sensing and sparse linear regression. Departing momentarily from our focus on dithered uniform quantization, we consider QCS under dithered 1-bit quantization and will apply Proposition 1 to derive results comparable to [22, Thms. 7-8] without resorting to the sparsity of $\boldsymbol{\Sigma}^\star$.

We first review the 1-bit quantization scheme developed in [22]:

- **Response Quantization.** We truncate y_k to $\widetilde{y}_k = \mathscr{T}_{\zeta_y}(y_k)$ with some threshold ζ_y, and then quantize $\widetilde{y}_k$ to $\dot{y}_k = \mathrm{sign}(\widetilde{y}_k + \phi_k)$ with uniform dither $\phi_k \sim \mathscr{U}([-\gamma_y, \gamma_y])$.
- **Covariate Quantization.** We truncate $\boldsymbol{x}_k$ to $\widetilde{\boldsymbol{x}}_k = \mathscr{T}_{\zeta_x}(\boldsymbol{x}_k)$ with some threshold ζ_x, and then quantize $\widetilde{\boldsymbol{x}}_k$ to $\dot{\boldsymbol{x}}_{k1} = \mathrm{sign}(\widetilde{\boldsymbol{x}}_k + \boldsymbol{\tau}_{k1})$ and $\dot{\boldsymbol{x}}_{k2} = \mathrm{sign}(\widetilde{\boldsymbol{x}}_k + \boldsymbol{\tau}_{k2})$, where $\boldsymbol{\tau}_{k1}, \boldsymbol{\tau}_{k2} \sim \mathscr{U}([-\gamma_x, \gamma_x]^d)$ are independent uniform dithers. (Note that we collect 2 bits per entry).

The following two results refine [22, Thms. 7-8] by deriving comparable error rates without using sparsity of $\boldsymbol{\Sigma}^\star$.

Theorem 11. (1-bit Quantized Sub-Gaussian Covariate)**.** *Given $\delta > 0$, we consider (8) where the s-sparse $\boldsymbol{\theta}^\star$ satisfies $\|\boldsymbol{\theta}^\star\|_1 \le R$, $\boldsymbol{x}_k$s are i.i.d. zero-mean sub-Gaussian with $\|\boldsymbol{x}_k\|_{\psi_2} \le \sigma$, and $\boldsymbol{\Sigma}^\star = \mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top)$ satisfies $\lambda_{\min}(\boldsymbol{\Sigma}^\star) \ge \kappa_0$ for some $\kappa_0 > 0$, the noise ϵ_ks are independent of $\boldsymbol{x}_k$s and i.i.d. sub-Gaussian, while for simplicity we assume $\|y_k\|_{\psi_2} \le \sigma$. In the quantization of $(\boldsymbol{x}_k, y_k)$ described above, we set $\zeta_x = \zeta_y = \infty$ and $\gamma_x, \gamma_y \asymp \sigma\sqrt{\log\big(\frac{n}{2\delta\log d}\big)}$. For recovery we let*

$$\boldsymbol{Q} := \frac{\gamma_x^2}{2n}\sum_{k=1}^n \big(\dot{\boldsymbol{x}}_{k1}\dot{\boldsymbol{x}}_{k2}^\top + \dot{\boldsymbol{x}}_{k2}\dot{\boldsymbol{x}}_{k1}^\top\big),$$
$$\boldsymbol{b} := \frac{\gamma_x\gamma_y}{n}\sum_{k=1}^n \dot{y}_k\dot{\boldsymbol{x}}_{k1},\ \mathcal{S} = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\|_1 \le R\}$$

and set

$$\lambda = C_1\sigma^2 R\sqrt{\frac{\delta\log d(\log n)^2}{n}}$$

with sufficiently large C_1. If $n \gtrsim \delta s\log d(\log n)^2$, then with probability at least $1-4d^{2-\delta}$, all $\widetilde{\boldsymbol{\theta}} \in \mathcal{S}$ satisfying (14) have estimation error $\widetilde{\boldsymbol{\Upsilon}} := \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star$ bounded by

$$\|\widetilde{\boldsymbol{\Upsilon}}\|_2 \le C_2\frac{\sigma^2}{\kappa_0}\cdot R\sqrt{\frac{\delta s\log d(\log n)^2}{n}},$$
$$\|\widetilde{\boldsymbol{\Upsilon}}\|_1 \le C_3\frac{\sigma^2}{\kappa_0}\cdot Rs\sqrt{\frac{\delta\log d(\log n)^2}{n}}.$$

Theorem 12. (1-bit Quantized Heavy-Tailed Covariate)**.** *Given $\delta > 0$, we consider (8) where the s-sparse $\boldsymbol{\theta}^\star$ satisfies $\|\boldsymbol{\theta}^\star\|_1 \le R$, $\boldsymbol{x}_k$s are i.i.d. zero-mean heavy-tailed satisfying the joint fourth-moment constraint $\sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}}\mathbb{E}|\boldsymbol{v}^\top\boldsymbol{x}_k|^4 \le M$, and $\boldsymbol{\Sigma}^\star = \mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top)$ satisfies $\lambda_{\min}(\boldsymbol{\Sigma}^\star) \ge \kappa_0$ for some $\kappa_0 > 0$, the noise ϵ_ks are independent of $\boldsymbol{x}_k$s and i.i.d. heavy-tailed with a bounded fourth-moment, while for simplicity we assume $\mathbb{E}|y_k|^4 \le M$. In the quantization of $(\boldsymbol{x}_k, y_k)$ described above, we set $\zeta_x, \zeta_y, \gamma_x, \gamma_y \asymp \big(\frac{nM^2}{\delta\log d}\big)^{1/8}$ and enforce $\zeta_x < \gamma_x$, $\zeta_y < \gamma_y$. For recovery we let*

$$\boldsymbol{Q} := \frac{\gamma_x^2}{2n}\sum_{k=1}^n \big(\dot{\boldsymbol{x}}_{k1}\dot{\boldsymbol{x}}_{k2}^\top + \dot{\boldsymbol{x}}_{k2}\dot{\boldsymbol{x}}_{k1}^\top\big),$$
$$\boldsymbol{b} := \frac{\gamma_x\gamma_y}{n}\sum_{k=1}^n \dot{y}_k\dot{\boldsymbol{x}}_{k1},\ \mathcal{S} = \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\|_1 \le R\}$$

and set

$$\lambda = C_1\sqrt{M}R\big(\frac{\delta\log d}{n}\big)^{1/4}$$

with sufficiently large C_1. If $n \gtrsim \delta s^2\log d$, then with probability at least $1-4d^{2-\delta}$, all $\widetilde{\boldsymbol{\theta}} \in \mathcal{S}$ satisfying (14) have estimation error $\widetilde{\boldsymbol{\Upsilon}} := \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star$ bounded by

$$\|\widetilde{\boldsymbol{\Upsilon}}\|_2 \le C_2\frac{\sqrt{M}}{\kappa_0}\cdot R\Big(\frac{\delta s^2\log d}{n}\Big)^{1/4},$$
$$\|\widetilde{\boldsymbol{\Upsilon}}\|_1 \le C_3\frac{\sqrt{M}}{\kappa_0}\cdot Rs\Big(\frac{\delta\log d}{n}\Big)^{1/4}.$$

B. Uniform Recovery Guarantee

Uniformity is a highly desired property for a compressed sensing guarantee because it allows one to use a fixed (possibly randomly drawn) measurement ensemble for all sparse signals. Unfortunately, as with many other results for nonlinear compressed sensing in the literature, our earlier recovery guarantees are non-uniform and only ensure the accurate

recovery of a sparse signal fixed before drawing the random measurement ensemble.

We provide another additional development to QCS in this part. Specifically, we establish a uniform recovery guarantee which, despite the heavy-tailed noise and nonlinear quantization scheme, notably retains a near minimax error rate. This is done by upgrading Theorem 5 to be uniform over all sparse $\boldsymbol{\theta}^\star$ by more in-depth technical tools and a careful covering argument. Part of the techniques is inspired by prior works [40], [94], but certain technical innovations are required:

1) Like the recent work [40], one crucial technical tool in our proof is a powerful concentration inequality for product process due to Mendelson [67], as adapted in the present Lemma 9. However, [40] only studied sub-Gaussian distribution, and the results produced by their unified approach typically exhibit a decaying rate of $O(n^{-1/4})$ [40, Sect. 4]. By contrast, our problem involves heavy-tailed noise only having bounded $(2+\nu)$-th moment ($\nu>0$), and we aim to establish a near minimax uniform error bound — cautiousness and new treatment are thus needed in the application of Lemma 9. More specifically, in the proof we need to bound

$$I_1=\sup_{\boldsymbol{\theta}\in\Sigma_{s,R_0}}\sup_{\boldsymbol{v}\in\mathscr{V}}\sum_{k=1}^n\big(\widetilde{y}_k\boldsymbol{x}_k^\top\boldsymbol{v}-\mathbb{E}[\widetilde{y}_k\boldsymbol{x}_k^\top\boldsymbol{v}]\big),$$

where $\mathscr{V}=\{\boldsymbol{v}:\|\boldsymbol{v}\|_2=1,\|\boldsymbol{v}\|_1\le 2\sqrt{s}\}$, and $\Sigma_{s,R_0}=\Sigma_s\cap\{\boldsymbol{\theta}\in\mathbb{R}^d:\|\boldsymbol{\theta}\|_2\le R_0\}$ is the signal space of interest, and recall that $\widetilde{y}_k=\mathscr{T}_{\zeta_y}(\boldsymbol{x}_k^\top\boldsymbol{\theta}+\epsilon_k)$ with sub-Gaussian $\boldsymbol{x}_k$. It is natural to invoke Lemma 9 to bound I_1 straightforwardly, but the issue is on lack of good bound for $\|\widetilde{y}_k\|_{\psi_2}$ due to the heavy-tailedness of ϵ_k; indeed, one only has the trivial estimate as $\|\widetilde{y}_k\|_{\psi_2}=O(\zeta_y)$, which is much worse than an $O(1)$ bound since $\zeta_y\asymp\sqrt{\frac{n}{\delta\log d}}$, and using Lemma 9 with this estimate leads to a loose bound for I_1 and finally a sub-optimal error rate. To address the issue, our main idea is to introduce the truncated heavy-tailed noise $\mathscr{T}_{\zeta_y}(\epsilon_k)$ and define $\widetilde{z}_k=\widetilde{y}_k-\mathscr{T}_{\zeta_y}(\epsilon_k)$, which enables us to decompose I_1 as

$$\begin{aligned}I_1\le&\underbrace{\sup_{\boldsymbol{\theta}\in\Sigma_{s,R_0}}\sup_{\boldsymbol{v}\in\mathscr{V}}\sum_{k=1}^n\big(\widetilde{z}_k\boldsymbol{x}_k^\top\boldsymbol{v}-\mathbb{E}[\widetilde{z}_k\boldsymbol{x}_k^\top\boldsymbol{v}]\big)}_{:=I_{11}}\\&+\underbrace{\sup_{\boldsymbol{v}\in\mathscr{V}}\sum_{k=1}^n\big(\mathscr{T}_{\zeta_y}(\epsilon_k)\boldsymbol{x}_k^\top\boldsymbol{v}-\mathbb{E}[\mathscr{T}_{\zeta_y}(\epsilon_k)\boldsymbol{x}_k^\top\boldsymbol{v}]\big)}_{:=I_{12}}.\end{aligned}$$

Now, the benefits of working with I_{11},I_{12} are that: i) We can directly invoke Lemma 9 to bound I_{11} since we have a good sub-Gaussian norm estimate $\|\widetilde{z}_k\|_{\psi_2}\le\|\boldsymbol{x}_k^\top\boldsymbol{\theta}\|_{\psi_2}\lesssim\|\boldsymbol{x}_k\|_{\psi_2}R_0$, see Step 2.1.1 in the proof; ii) I_{12} becomes the supremum of a process that is independent of $\boldsymbol{\theta}$ and only indexed by $\boldsymbol{v}$, hence Bernstein's inequality suffices for bounding I_{12} (Step 2.1.2 in the proof), analogously to the proof of the non-uniform guarantee (Theorem 5).

2) Like [94, Prop. 6.1], we invoke a covering argument with similar techniques to bound

$$I_0=\sup_{\boldsymbol{\theta}\in\Sigma_{s,R_0}}\sup_{\boldsymbol{v}\in\mathscr{V}}\sum_{k=1}^n\xi_k\boldsymbol{x}_k^\top\boldsymbol{v},$$

where $\xi_k=\mathcal{Q}_\Delta(\widetilde{y}_k+\tau_k)-\widetilde{y}_k$ is the quantization noise. Nevertheless, our Lasso estimator is different from their projected back projection estimator, and it turns out that we need to directly handle "$\sup_{\boldsymbol{v}\in\mathscr{V}}$" by Lemma 10, unlike [94, Prop. 6.2] that again used a covering argument for this purpose. See more discussions in Step 2.4 of the proof.

We are in a position to present our uniform recovery guarantee. We follow most assumptions in Theorem 5 but specify the signal space as $\boldsymbol{\theta}^\star\in\Sigma_{s,R_0}$ and impose the $(2l)$-th moment constraint on ϵ_k. Following prior works on QCS (e.g., [40], [89]), we consider constrained Lasso that utilizes an ℓ_1-constraint $\|\boldsymbol{\theta}\|_1\le\|\boldsymbol{\theta}^\star\|_1$ (rather than (10)) to pursue uniform recovery.

Theorem 13. (Uniform Version of Theorem 5)**.** *Given some $\delta>0,\Delta>0$, in (8) we suppose that $\boldsymbol{x}_k$s are i.i.d., zero-mean sub-Gaussian with $\|\boldsymbol{x}_k\|_{\psi_2}\le\sigma$, $\kappa_0\le\lambda_{\min}(\boldsymbol{\Sigma}^\star)\le\lambda_{\max}(\boldsymbol{\Sigma}^\star)\le\kappa_1$ for some $\kappa_1\ge\kappa_0>0$ where $\boldsymbol{\Sigma}^\star=\mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top)$, $\boldsymbol{\theta}^\star\in\Sigma_{s,R_0}:=\Sigma_s\cap\{\boldsymbol{\theta}:\|\boldsymbol{\theta}\|_2\le R_0\}$ for some absolute constant R_0, ϵ_ks are i.i.d. noise that are independent of $\boldsymbol{x}_k$s and satisfy $\mathbb{E}|\epsilon_k|^{2l}\le M$ for some fixed $l>1$. In quantization, we truncate y_k to $\widetilde{y}_k=\mathscr{T}_{\zeta_y}(y_k)$ with threshold $\zeta_y\asymp\big(\frac{n(M^{1/l}+\sigma^2)}{\delta\log d}\big)^{1/2}$, then quantize $\widetilde{y}_k$ to $\dot{y}_k=\mathcal{Q}_\Delta(\widetilde{y}_k+\tau_k)$ with uniform dither $\tau_k\sim\mathscr{U}([-\frac{\Delta}{2},\frac{\Delta}{2}])$. For recovery, we define the estimator $\widehat{\boldsymbol{\theta}}$ as the solution to constrained Lasso*

$$\widehat{\boldsymbol{\theta}}=\mathop{\arg\min}_{\|\boldsymbol{\theta}\|_1\le\|\boldsymbol{\theta}^\star\|_1}\ \frac{1}{2n}\sum_{k=1}^n(\dot{y}_k-\boldsymbol{x}_k^\top\boldsymbol{\theta})^2$$

If $n\gtrsim\delta s\log\mathscr{W}$ for $\mathscr{W}=\frac{\kappa_1 d^2n^3}{\Delta^2s^5\delta^3}$ and some hidden constant depending on (κ_0,σ), then with probability at least $1-Cd^{1-\delta}$ on a single random draw of $(\boldsymbol{x}_k,\epsilon_k,\tau_k)_{k=1}^n$, it holds uniformly for all $\boldsymbol{\theta}^\star\in\Sigma_{s,R_0}$ that the estimation error $\widehat{\boldsymbol{\Upsilon}}:=\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^\star$ satisfy

$$\begin{aligned}\|\widehat{\boldsymbol{\Upsilon}}\|_2&\le\frac{C_3\sigma(\sigma+M^{\frac{1}{2l}})}{\kappa_0}\sqrt{\frac{\delta s\log d}{n}}+\frac{C_3\sigma\Delta}{\kappa_0}\sqrt{\frac{\delta s\log\mathscr{W}}{n}},\\\|\widehat{\boldsymbol{\Upsilon}}\|_1&\le\frac{C_4\sigma(\sigma+M^{\frac{1}{2l}})}{\kappa_0}s\sqrt{\frac{\delta\log d}{n}}+\frac{C_4\sigma\Delta}{\kappa_0}s\sqrt{\frac{\delta\log\mathscr{W}}{n}}.\end{aligned}$$

Notably, our uniform guarantee is still minimax optimal up to some additional logarithmic factors (i.e., $\sqrt{\log\mathscr{W}}$) arising from the covering argument (Step 2.4 of the proof), whose main aim is to show that one uniform dither $\boldsymbol{\tau}=[\tau_k]$ suffices for all signals. Thus naturally, $\sqrt{\log\mathscr{W}}$ is associated with a leading factor of the quantization level Δ, meaning that the logarithmic gap between uniform recovery and non-uniform recovery closes when $\Delta\to0$. In particular, Theorem 13 implies a uniform error rate matching the non-uniform one in Theorem 5 (up to some multiplicative factors) when Δ is small enough or in an unquantized case.

To the best of our knowledge, the only existing uniform guarantee for heavy-tailed QCS is [33, Thm. 1.11], but the following distinctions make it impossible to closely compare their result with our Theorem 13: 1) [33, Thm. 1.11] is for dithered 1-bit quantization, but ours is for dithered uniform quantizer; 2) We handle heavy-tailedness by truncation, while [33, Thm. 1.11] does not involve this kind of special treatment; 3) [33, Thm. 1.11] considers a highly intractable program with

hamming distance as objective and $\boldsymbol{\theta} \in \Sigma_s$ as constraint (when specialized to sparse signal), while our Theorem 13 is for the convex program Lasso; 4) Their analysis is based on an in-depth result on random hyperplane tessellations (see also [34], [77]), while our proof follows the more standard strategy (i.e., to upgrade each piece in a non-uniform proof to be uniform) and requires certain technical innovations (e.g., the treatment to deal with the truncation step).[13] Note that [33, Thm. 1.11] is robust to adversarial bit flips, and we leave it future work to extend our results to a setting where adversarial noise presents.

V. Numerical Simulations

In this section, we provide two sets of experimental results to support and demonstrate our theoretical developments. The first set of our simulations is devoted to validating our major standpoint that near minimax rates are achievable in quantized heavy-tailed settings. Then, the second set of results is presented to illustrate the crucial role played by the appropriate dither (i.e., triangular dither for covariate, uniform dither for response) before uniform quantization. For the importance of data truncation we refer to in [36, Sect. 5], which includes three estimation problems in this work and contrasts the estimations with or without the data truncation.

A. (Near) Minimax Error Rates

Each data point in our results is set to be the mean value of 50 or 100 independent trials.

1) Quantized Covariance Matrix Estimation: We start from covariance matrix estimation, specifically, we verify the element-wise rate $\mathscr{B}_1 := O\big(\mathscr{L}\sqrt{\frac{\log d}{n}}\big)$ and operator norm rate $\mathscr{B}_2 := O\big(\mathscr{L}\sqrt{\frac{d\log d}{n}}\big)$ in Theorems 2-3.

For estimator in Theorem 2, we draw $\boldsymbol{x}_k = (x_{ki})$ such that the first two coordinates are independently drawn from $\mathtt{t}(4.5)$, $(x_{ki})_{i=3,4}$ are from $\mathtt{t}(6)$ with covariance $\mathbb{E}(x_{k3}x_{k4}) = 1.2$, and the remaining $d-4$ coordinates are i.i.d. following $\mathtt{t}(6)$. We test different choices of (d, Δ) under $n = 80 : 20 : 220$, and the log-log plots are shown in Figure 1(a). Clearly, for each (d, Δ) the experimental points roughly exhibit a straight line that is well aligned with the dashed line representing the $n^{-1/2}$ rate. As predicted by the factor $\mathscr{L} = \sqrt{M} + \Delta^2$, the curves with larger Δ are higher, but note that the error decreasing rates remain unchanged. In addition, the curves of $(d, \Delta) = (100, 1), (120, 1)$ are extremely close, which is consistent with the logarithmic dependence of $\mathscr{B}_1$ on d.

For the error bound $\mathscr{B}_2$, the coordinates of $\boldsymbol{x}_k$ are independently drawn from a scaled version of $\mathtt{t}(4.5)$ such that $\boldsymbol{\Sigma}^\star = \mathrm{diag}(2, 2, 1, ..., 1)$, and we test different settings of (d, Δ) under $n = 200 : 100 : 1000$. As shown in Figure 1(b), the operator norm error decreases with n in the optimal rate $n^{-1/2}$, and using a coarser dithered quantizer (i.e., larger Δ) only slightly lifts the curves. Indeed, the effect seems consistent with $\mathscr{L}$'s quadratic dependence on Δ. To validate the relative scaling of n and d, in addition to the setting $(d, \Delta) = (100, 1)$ under $n = 200 : 100 : 1000$, we try $(d, \Delta) = (150, 1)$ under 1.5 times the original sample size $n = 1.5 \times (200 : 100 : 1000)$ (but in Figure 1(b) we still plot the curve according to the sample size of $200 : 100 : 1000$ without the multiplicative factor of 1.5), and surprisingly the obtained curve coincides with the one for $(d, \Delta) = (100, 1)$. Thus, ignoring the logarithmic factor $\log d$, the operator norm error can be characterized by $\mathscr{B}_2$ fairly well.

Additionally, we want to compare $\mathscr{B}_1$ and $\mathscr{B}_2$ regarding the dependence on d more clearly. Specifically, we generate the samples $\boldsymbol{x}_k$s as in Figure 1(a) and test the fixed sample size $n = 180$ and varying dimension $d = 80 : 20 : 260$. The max norm estimation errors of $\widehat{\boldsymbol{\Sigma}}$ in Theorem 2 and the operator norm errors (under $d = 80 : 20 : 180$ to ensure $n \geq d$) of the estimator in Theorem 3 are reported in Figure 1(c). It is clear that the max norm error increases with d rather slowly, while the operator norm error increases much more significantly under larger d. This is consistent with the logarithmic dependence of $\mathscr{B}_1$ on d and the more essential dependence of $\mathscr{B}_2$ on d.

2) Quantized Compressed Sensing: We now switch to QCS with unquantized covariate and aim to verify the ℓ_2-norm error rate $\mathscr{B}_3 = O\big(\mathscr{L}\sqrt{\frac{s\log d}{n}}\big)$ obtained in Theorems 5-6. We let the support of the s-sparse $\boldsymbol{\theta}^\star \in \mathbb{R}^d$ be $[s]$, and then draw the non-zero entries from a uniform distribution over $\mathbb{S}^{s-1}$ (hence $\|\boldsymbol{\theta}^\star\|_2 = 1$). For the setting of Theorem 5 we adopt $\boldsymbol{x}_k \sim \mathcal{N}(0, \boldsymbol{I}_d)$ and $\epsilon_k \sim \frac{1}{\sqrt{6}}\mathtt{t}(3)$, while $\boldsymbol{x}_{ki} \overset{iid}{\sim} \frac{\sqrt{5}}{3}\mathtt{t}(4.5)$ and $\epsilon_k \sim \frac{1}{\sqrt{3}}\mathtt{t}(4.5)$ for Theorem 6. We simulate different choices of (d, s, Δ) under $n = 100 : 100 : 1000$, and the proposed convex program (10) is solved with the framework of ADMM (we refer to the review [8]). Experimental results are shown as log-log plots in Figure 2. Consistent with the theoretical bound $\mathscr{B}_3$, the errors in both cases decrease in a rate of $n^{-1/2}$, whereas the effect of uniform quantization is merely on the multiplicative factor $\mathscr{L}$. Interestingly, it seems that the gaps between $\Delta = 0, 0.5$ and $\Delta = 0.5, 1$ are in agreement with the explicit form of $\mathscr{L}$, i.e., $\mathscr{L} \asymp M^{1/(2l)} + \Delta$ for Theorem 5, and $\mathscr{L} \asymp \sqrt{M} + \Delta^2$ for Theorem 6. In addition, note that the curves of $(d, s) = (150, 5), (180, 5)$ are close, whereas increasing $s = 8$ suffers from a significantly larger error. This is consistent with the scaling law of (n, d, s) in $\mathscr{B}_3$.

Then, we simulate the complete quantization setting where both covariate and response are quantized (Theorems 9-10). The simulation details are the same as before except that $\boldsymbol{x}_k$ is also quantized with a quantization level the same as y_k. We provide the best ℓ_1-norm constraint for recovery, i.e., $\mathcal{S} := \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\|_1 \leq \|\boldsymbol{\theta}^\star\|_1\}$. Then, composite gradient descent [63], [64] is invoked to handle the non-convex estimation program. We show the log-log plots in Figure 3. Note that these results have implications similar to Figure 2, in terms of the $n^{-1/2}$ rate, the effect of quantization, and the relative scaling of (n, d, s).

3) Quantized Matrix Completion: Finally, we simulate QMC and demonstrate the error bound $\mathscr{B}_4 = O\big(\mathscr{L}\sqrt{\frac{rd\log d}{n}}\big)$ for $\|\widehat{\boldsymbol{\Upsilon}}\|_F/d$ in Theorems 7-8. We generate the rank-r $\boldsymbol{\Theta}^\star \in \mathbb{R}^{d\times d}$ as follows: we first generate $\boldsymbol{\Theta}_0 \in \mathbb{R}^{d\times r}$ with i.i.d. standard Gaussian entries to obtain the rank-r $\boldsymbol{\Theta}_1 := \boldsymbol{\Theta}_0\boldsymbol{\Theta}_0^\top$,

[13] It is possible to use such a standard strategy to upgrade Theorem 6 to a uniform result; we suspect that the error rate may exhibit worse dependence on s due to covering argument.

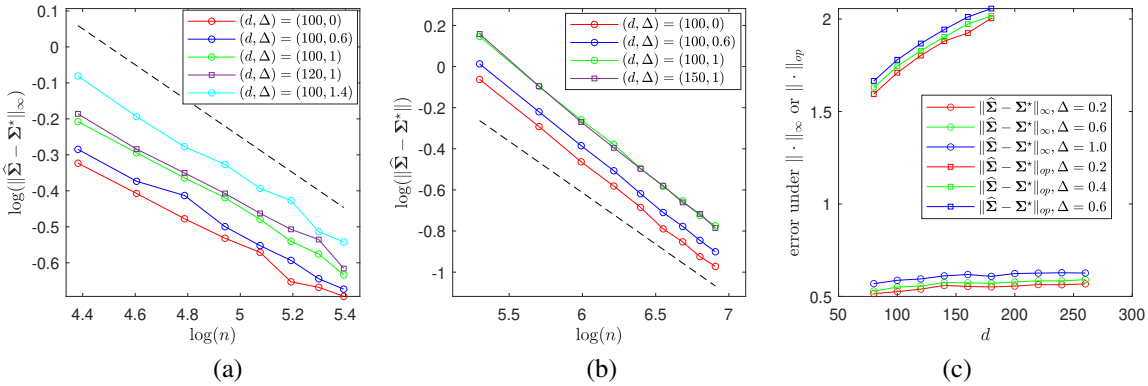

Fig. 1. (a): Element-wise error (Theorem 2); (b): operator norm error (Theorem 3); (c): the dependence on d of both error metrics.

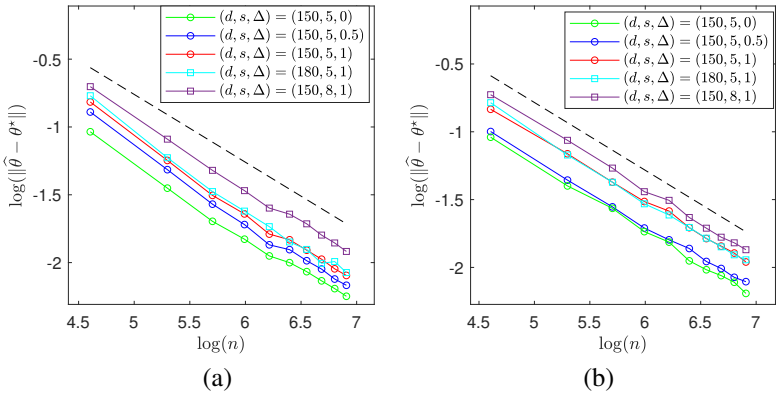

Fig. 2. (a): QCS in Theorem 5; (b): QCS in Theorem 6.

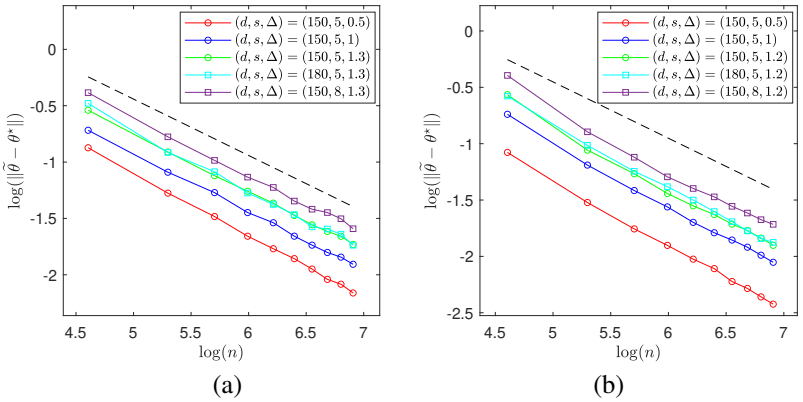

Fig. 3. (a): QCS in Theorem 9; (b): QCS in Theorem 10.

then we rescale it to $\mathbf{\Theta}^\star := k_1\mathbf{\Theta}_1$ such that $\|\mathbf{\Theta}^\star\|_F = d$. We use $\epsilon_k \sim \mathcal{N}(0,\frac{1}{4})$ to simulate the sub-exponential noise in Theorem 7, while $\epsilon_k \sim \frac{1}{\sqrt{6}}\mathsf{t}(3)$ for Theorem 8. The convex program (12) is fed with $\alpha = \|\mathbf{\Theta}^\star\|_\infty$ and optimized by the ADMM algorithm. We test different choices of (d,r,Δ) under $n = 2000:1000:8000$, with the log-log error plots displayed in Figure 4. Firstly, the experimental curves are well aligned with the dashed line that represents the optimal $n^{-1/2}$ rate. Then, comparing the results for $\Delta = 0, 0.5, 1$, we conclude that quantization only affects the multiplicative factor $\mathscr{L}$ in the estimation error. It should also be noted that, increasing either d or r leads to a significantly larger error, which is consistent with the $\mathscr{B}_4$'s essential dependence on d and r.

B. Importance of Appropriate Dithering

To demonstrate the crucial role played by the suitable dither, we provide the second set of simulations. In order to observe more significant phenomena and then conclude evidently, we may test a huge sample size but a rather simple estimation problem under coarse quantization (i.e., large Δ).

Specifically, for covariance matrix estimation we set $d = 1$ and i.i.d. draw $X_1, ..., X_n$ from $\mathcal{N}(0,1)$. Thus, the problem boils down to estimating $\mathbb{E}|X_k|^2$, for which the estimators in Theorems 2-3 coincide. Since X_k is sub-Gaussian, we

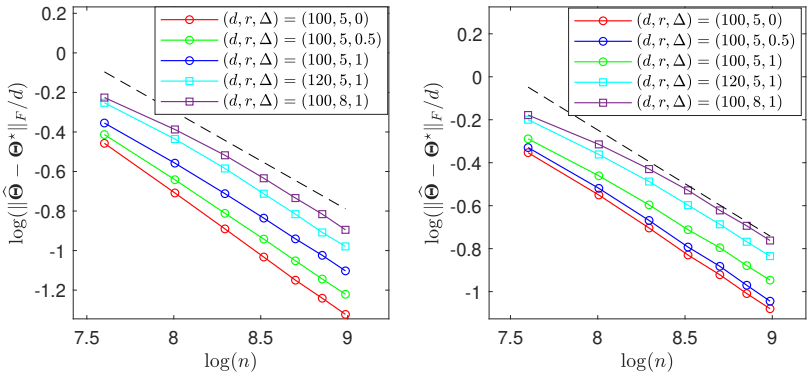

Fig. 4. (a): QMC in Theorem 7; (b): QMC in Theorem 8.

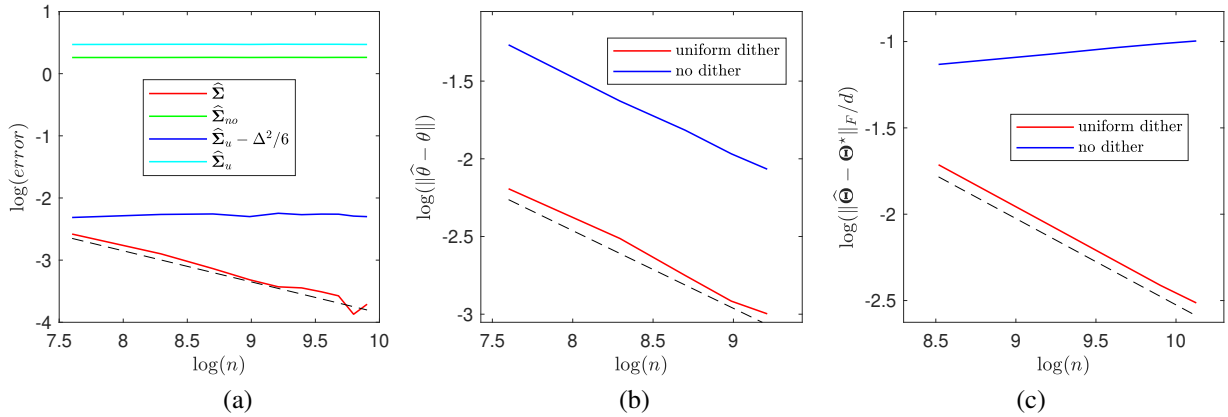

Fig. 5. (a): covariance matrix estimation; (b): QCS in Theorem 5; (c): QMC in Theorem 7.

do not perform data truncation before dithered quantization. Besides our estimator $\widehat{\boldsymbol{\Sigma}}=\frac{1}{n}\sum_{k=1}^n \dot{X}_k^2-\frac{\Delta^2}{4}$ where $\dot{X}_k=\mathcal{Q}_\Delta(X_k+\tau_k)$ and τ_k is triangular dither, we invite the following competitors:

- $\widehat{\boldsymbol{\Sigma}}_{no}=\frac{1}{n}\sum_{k=1}^n(\dot{X}_k')^2$, where $\dot{X}_k'=\mathcal{Q}_\Delta(X_k)$ is the direct quantization without dithering;
- $\widehat{\boldsymbol{\Sigma}}_u-\frac{\Delta^2}{6}$ and $\widehat{\boldsymbol{\Sigma}}_u$, where $\widehat{\boldsymbol{\Sigma}}_u=\frac{1}{n}\sum_{k=1}^n(\dot{X}_k'')^2$, and $\dot{X}_k''=\mathcal{Q}_\Delta(X_k+\tau_k'')$ is quantized under uniform dither $\tau_k''\sim\mathscr{U}([-\frac{\Delta}{2},\frac{\Delta}{2}])$.

To illustrate the choice of $\widehat{\boldsymbol{\Sigma}}_u-\frac{\Delta^2}{6}$ and $\widehat{\boldsymbol{\Sigma}}_u$, we write

$$\dot{X}_k''=X_k+\tau_k''+w_k=X_k+\xi_k$$

with quantization error $w_k\sim\mathscr{U}([-\frac{\Delta}{2},\frac{\Delta}{2}])$ (due to Theorem 1(a)) and quantization noise $\xi_k=\tau_k''+w_k$, then (5) gives

$$\mathbb{E}(\dot{X}_k'')^2=\mathbb{E}|X_k|^2+\mathbb{E}|\xi_k|^2,$$

while $\mathbb{E}|\xi_k|^2$ remains unknown. Thus, we consider $\widehat{\boldsymbol{\Sigma}}_u-\frac{\Delta^2}{6}$ because of an *unjustified* guess

$$\mathbb{E}|\xi_k|^2\approx\mathbb{E}|\tau_k''|^2+\mathbb{E}|w_k|^2=\frac{\Delta^2}{6},$$

while $\widehat{\boldsymbol{\Sigma}}_u$ simply gives up the correction of $\mathbb{E}|\xi_k|^2$. We test $\Delta=3$ under $n=(2:2:20)\cdot10^3$. From the results shown in Figure 5(a), the proposed estimator based on quantized data under triangular dither embraces the lowest estimation errors and the optimal rate of $n^{-1/2}$, whereas other competitors are not consistent, i.e., they all reach some error floors under a large sample size.

For the two remaining signal recovery problems, we simply focus on the quantization of the response y_k. In particular, we simulate QCS in the setting of Theorem 5, with $(d,s,\Delta)=(50,3,2)$ under $n:=(2:2:20)\cdot10^3$. Other experimental details are as previously stated. We compare our estimator $\widehat{\boldsymbol{\theta}}$ with its counterpart $\widehat{\boldsymbol{\theta}}'$ defined by (10) with the same $\boldsymbol{Q},\mathcal{S}$ but $\boldsymbol{b}'=\frac{1}{n}\sum_{k=1}^n\dot{y}_k'\boldsymbol{x}_k$, where $\dot{y}_k'=\mathcal{Q}_\Delta(\widetilde{y}_k)$ is a direct uniform quantization with no dither. Evidently, the simulation results in Figure 5(b) confirm that the application of a uniform dither significantly lessens the recovery errors. Without dithering, although our results under the Gaussian covariate still exhibit $n^{-1/2}$ decreasing rate, the identifiability issue unavoidably arises under the Bernoulli covariate. In that case, the simulation without dithering will evidently deviate from the $n^{-1/2}$ rate, see [87, Figure 1] for instance.

In analogy, we simulate QMC (Theorem 7) with data generated as previous experiments, and specifically we try $(d,r,\Delta)=(30,5,1.5)$ under $n=(5:5:25)\cdot10^3$. While our estimator $\widehat{\boldsymbol{\Theta}}$ is defined in (12) involving $\dot{y}_k$ from a dithered quantizer, we simulate the performance of its counterpart without dithering, i.e., $\widehat{\boldsymbol{\Theta}}'$ defined in (12) with $\dot{y}_k$ substituted by $\dot{y}_k'=\mathcal{Q}_\Delta(y_k)$. From the experimental results displayed in Figure 5(c), one shall clearly see that $\widehat{\boldsymbol{\Theta}}$ performs much

better in terms of the decreasing rate of $n^{-1/2}$ and the estimation error; while the curve without dithering even does not decrease.

VI. Concluding Remarks

In digital signal processing and many distributed machine learning systems, data quantization is an indispensable process. On the other hand, many modern datasets exhibit heavy-tailedness, and the past decade has witnessed an increasing interest in statistical estimation methods robust to heavy-tailed data. In this work, we try to bridge these two developments by studying the quantization of heavy-tailed data. We propose to truncate the heavy-tailed data prior to a uniform quantizer with random dither well suited to the problem at hand. Applying our quantization scheme to covariance matrix estimation, compressed sensing, and matrix completion, we have proposed (near) optimal estimators based on quantized data, and they are computationally feasible. These results suggest a unified conclusion that the dithered quantization does not affect the key scaling law in the error rate but only slightly worsens the multiplicative factor, which was complemented by numerical simulations. Further, in two respects, we presented additional developments for quantized compressed sensing. Firstly, we study a novel setting that involves covariate quantization. Because our quantized covariance matrix estimator is not positive semi-definite, the proposed recovery program is non-convex, but we proved that all local minimizers enjoy near minimax rates. At a higher level, this development extends a line of works on non-convex M-estimator [62]–[64] to accommodate heavy-tailed covariate, see the deterministic framework Proposition 1. As an application, we derive results for (dithered) 1-bit compressed sensing as byproducts. Secondly, we established a near minimax uniform recovery guarantee for QCS under heavy-tailed noise, which states that all sparse signals within an ℓ_2-ball can be uniformly recovered up to near optimal ℓ_2-norm error, using a single realization of the measurement ensemble. We believe the developments presented in this work will prove useful in many other estimation problems, for instance, the triangular dither and the quantization scheme apply to multi-task learning, as shown by subsequent works [23], [60].

References

[1] Albert Ai, Alex Lapanowski, Yaniv Plan, and Roman Vershynin. One-bit compressed sensing with non-gaussian measurements. *Linear Algebra and its Applications*, 441:222–239, 2014.

[2] James Bennett, Stan Lanning, et al. The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35. New York, 2007.

[3] Sonia A Bhaskar. Probabilistic low-rank matrix completion from quantized measurements. *The Journal of Machine Learning Research*, 17(1):2131–2164, 2016.

[4] Peter J Bickel and Elizaveta Levina. Covariance regularization by thresholding. *The Annals of statistics*, 36(6):2577–2604, 2008.

[5] Atanu Biswas, Sujay Datta, Jason P Fine, and Mark R Segal. *Statistical advances in the biomedical sciences: clinical trials, epidemiology, survival analysis, and bioinformatics*. John Wiley & Sons, 2007.

[6] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[7] Petros T Boufounos and Richard G Baraniuk. 1-bit compressive sensing. In *2008 42nd Annual Conference on Information Sciences and Systems*, pages 16–21. IEEE, 2008.

[8] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[9] Christian Brownlees, Emilien Joly, and Gábor Lugosi. Empirical risk minimization for heavy-tailed losses. *The Annals of Statistics*, 43(6):2507–2536, 2015.

[10] T Tony Cai, Cun-Hui Zhang, and Harrison H Zhou. Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics*, 38(4):2118–2144, 2010.

[11] T Tony Cai and Harrison H Zhou. Minimax estimation of large covariance matrices under ℓ_1-norm. *Statistica Sinica*, pages 1319–1349, 2012.

[12] T Tony Cai and Harrison H Zhou. Optimal rates of convergence for sparse covariance matrix estimation. *The Annals of Statistics*, pages 2389–2420, 2012.

[13] Tony Cai and Weidong Liu. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684, 2011.

[14] Tony Cai and Wen-Xin Zhou. A max-norm constrained minimization approach to 1-bit matrix completion. *J. Mach. Learn. Res.*, 14(1):3619–3647, 2013.

[15] Emmanuel Candes and Benjamin Recht. Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6):111–119, 2012.

[16] Yang Cao and Yao Xie. Categorical matrix completion. In *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 369–372. IEEE, 2015.

[17] Olivier Catoni. Challenging the empirical mean and empirical variance: a deviation study. In *Annales de l'IHP Probabilités et statistiques*, volume 48, pages 1148–1185, 2012.

[18] Junren Chen and Michael K Ng. Color image inpainting via robust pure quaternion matrix completion: Error bound and weighted loss. *SIAM Journal on Imaging Sciences*, 15(3):1469–1498, 2022.

[19] Junren Chen and Michael K Ng. A parameter-free two-bit covariance estimator with improved operator norm error rate. *arXiv preprint arXiv:2308.16059*, 2023.

[20] Junren Chen and Michael K. Ng. Uniform exact reconstruction of sparse signals and low-rank matrices from phase-only measurements. *IEEE Transactions on Information Theory*, 69(10):6739–6764, 2023.

[21] Junren Chen, Jonathan Scarlett, Michael K. Ng, and Zhaoqiang Liu. A unified framework for uniform signal recovery in nonlinear generative compressed sensing. *Advances in Neural Information Processing Systems*, 2023.

[22] Junren Chen, Cheng-Long Wang, Michael K. Ng, and Di Wang. High dimensional statistical estimation under uniformly dithered one-bit quantization. *IEEE Transactions on Information Theory*, 69(8):5151–5187, 2023.

[23] Junren Chen, Yueqi Wang, and Michael K. Ng. Quantized low-rank multivariate regression with random dithering. *IEEE Transactions on Signal Processing*, pages 1–16, 2023.

[24] Onkar Dabeer and Aditya Karnik. Signal parameter estimation using 1-bit dithered quantization. *IEEE Transactions on Information Theory*, 52(12):5389–5405, 2006.

[25] Alireza Danaee, Rodrigo C de Lamare, and Vitor Heloiz Nascimento. Distributed quantization-aware rls learning with bias compensation and coarsely quantized signals. *IEEE Transactions on Signal Processing*, 70:3441–3455, 2022.

[26] Mark A Davenport, Yaniv Plan, Ewout Van Den Berg, and Mary Wootters. 1-bit matrix completion. *Information and Inference: A Journal of the IMA*, 3(3):189–223, 2014.

[27] Mark A Davenport and Justin Romberg. An overview of low-rank matrix recovery from incomplete observations. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):608–622, 2016.

[28] Luc Devroye, Matthieu Lerasle, Gabor Lugosi, and Roberto I Oliveira. Sub-gaussian mean estimators. *The Annals of Statistics*, 44(6):2695–2725, 2016.

[29] Sjoerd Dirksen. Quantized compressed sensing: a survey. In *Compressed Sensing and Its Applications*, pages 67–95. Springer, 2019.

[30] Sjoerd Dirksen, Hans Christian Jung, and Holger Rauhut. One-bit compressed sensing with partial gaussian circulant matrices. *Information and Inference: A Journal of the IMA*, 9(3):601–626, 2020.

[31] Sjoerd Dirksen and Johannes Maly. Tuning-free one-bit covariance estimation using data-driven dithering. *arXiv preprint arXiv:2307.12613*, 2023.

[32] Sjoerd Dirksen, Johannes Maly, and Holger Rauhut. Covariance estimation under one-bit quantization. *The Annals of Statistics*, 50(6):3538–3562, 2022.

[33] Sjoerd Dirksen and Shahar Mendelson. Non-gaussian hyperplane tessellations and robust one-bit compressed sensing. *Journal of the European Mathematical Society*, 23(9):2913–2947, 2021.

[34] Sjoerd Dirksen, Shahar Mendelson, and Alexander Stollenwerk. Sharp estimates on random hyperplane tessellations. *SIAM Journal on Mathematics of Data Science*, 4(4):1396–1419, 2022.

[35] Jianqing Fan, Quefeng Li, and Yuyan Wang. Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1):247–265, 2017.

[36] Jianqing Fan, Weichen Wang, and Ziwei Zhu. A shrinkage principle for heavy-tailed data: High-dimensional robust low-rank matrix recovery. *Annals of statistics*, 49(3):1239, 2021.

[37] Simon Foucart, Deanna Needell, Reese Pathak, Yaniv Plan, and Mary Wootters. Weighted matrix completion from non-random, non-uniform sampling patterns. *IEEE Transactions on Information Theory*, 67(2):1264–1290, 2020.

[38] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer New York, New York, NY, 2013.

[39] Martin Genzel and Christian Kipp. Generic error bounds for the generalized lasso with sub-exponential data. *Sampling Theory, Signal Processing, and Data Analysis*, 20(2):15, 2022.

[40] Martin Genzel and Alexander Stollenwerk. A unified approach to uniform signal recovery from nonlinear observations. *Foundations of Computational Mathematics*, pages 1–74, 2022.

[41] Robert M Gray and Thomas G Stockham. Dithered quantizers. *IEEE Transactions on Information Theory*, 39(3):805–812, 1993.

[42] David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Physical review letters*, 105(15):150401, 2010.

[43] Osama A Hanna, Yahya H Ezzeldin, Christina Fragouli, and Suhas Diggavi. Quantization of distributed data for learning. *IEEE Journal on Selected Areas in Information Theory*, 2(3):987–1001, 2021.

[44] Daniel Hsu and Sivan Sabato. Loss minimization and parameter estimation with heavy tails. *The Journal of Machine Learning Research*, 17(1):543–582, 2016.

[45] Marat Ibragimov, Rustam Ibragimov, and Johan Walden. *Heavy-tailed distributions and robustness in economics and finance*, volume 214. Springer, 2015.

[46] Laurent Jacques and Valerio Cambareri. Time for dithering: fast and quantized random embeddings via the restricted isometry property. *Information and Inference: A Journal of the IMA*, 6(4):441–476, 2017.

[47] Laurent Jacques and Thomas Feuillen. The importance of phase in complex compressive sensing. *IEEE Transactions on Information Theory*, 67(6):4150–4161, 2021.

[48] Laurent Jacques, Jason N Laska, Petros T Boufounos, and Richard G Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE transactions on information theory*, 59(4):2082–2102, 2013.

[49] Halyun Jeong, Xiaowei Li, Yaniv Plan, and Ozgur Yilmaz. Sub-gaussian matrices on sets: Optimal tail dependence and applications. *Communications on Pure and Applied Mathematics*, 75(8):1713–1754, 2022.

[50] Hans Christian Jung, Johannes Maly, Lars Palzer, and Alexander Stollenwerk. Quantized compressed sensing by rectified linear units. *IEEE transactions on information theory*, 67(6):4125–4149, 2021.

[51] Olga Klopp. Noisy low-rank matrix completion with general sampling distribution. *Bernoulli*, 20(1):282–303, 2014.

[52] Olga Klopp, Jean Lafond, Éric Moulines, and Joseph Salmon. Adaptive multinomial matrix completion. *Electronic Journal of Statistics*, 9(2):2950–2975, 2015.

[53] Olga Klopp, Karim Lounici, and Alexandre B Tsybakov. Robust matrix completion. *Probability Theory and Related Fields*, 169(1):523–564, 2017.

[54] Karin Knudson, Rayan Saab, and Rachel Ward. One-bit compressive sensing with norm estimation. *IEEE Transactions on Information Theory*, 62(5):2748–2758, 2016.

[55] Vladimir Koltchinskii, Karim Lounici, and Alexandre B Tsybakov. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*, 39(5):2302–2329, 2011.

[56] Jakub Konečnỳ, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*, 2016.

[57] Piotr Kruczek, Radosław Zimroz, and Agnieszka Wyłomańska. How to detect the cyclostationarity in heavy-tailed distributed signals. *Signal Processing*, 172:107514, 2020.

[58] Arun Kumar Kuchibhotla and Abhishek Chakrabortty. Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *Information and Inference: A Journal of the IMA*, 11(4):1389–1456, 2022.

[59] Jean Lafond, Olga Klopp, Eric Moulines, and Joseph Salmon. Probabilistic low-rank matrix completion on finite alphabets. *Advances in Neural Information Processing Systems*, 27, 2014.

[60] Kangqiang Li and Yuxuan Wang. Two results on low-rank heavy-tailed multiresponse regressions. *arXiv preprint arXiv:2305.13897*, 2023.

[61] Christopher Liaw, Abbas Mehrabian, Yaniv Plan, and Roman Vershynin. A simple tool for bounding the deviation of random matrices on geometric sets. In *Geometric aspects of functional analysis*, pages 277–299. Springer, 2017.

[62] Po-Ling Loh. Statistical consistency and asymptotic normality for high-dimensional robust m-estimators. *The Annals of Statistics*, 45(2):866–896, 2017.

[63] Po-Ling Loh and Martin J. Wainwright. High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of statistics*, 40(3):1637–1664, 2012.

[64] Po-Ling Loh and Martin J. Wainwright. Regularized m-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *Journal of Machine Learning Research*, 16(19):559–616, 2015.

[65] Gábor Lugosi and Shahar Mendelson. Mean estimation and regression under heavy-tailed distributions: A survey. *Foundations of Computational Mathematics*, 19(5):1145–1190, 2019.

[66] Gabor Lugosi and Shahar Mendelson. Robust multivariate mean estimation: the optimality of trimmed mean. *The Annals of Statistics*, 49(1):393–410, 2021.

[67] Shahar Mendelson. Upper bounds on product and multiplier empirical processes. *Stochastic Processes and their Applications*, 126(12):3652–3680, 2016.

[68] Stanislav Minsker. Geometric median and robust estimation in banach spaces. *Bernoulli*, 21(4):2308–2335, 2015.

[69] Jianhua Mo and Robert W Heath. Limited feedback in single and multi-user mimo systems with finite-bit adcs. *IEEE Transactions on Wireless Communications*, 17(5):3284–3297, 2018.

[70] Sahand Negahban and Martin J Wainwright. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics*, 39(2):1069–1097, 2011.

[71] Sahand Negahban and Martin J Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *The Journal of Machine Learning Research*, 13(1):1665–1697, 2012.

[72] Sahand N Negahban, Pradeep Ravikumar, Martin J Wainwright, and Bin Yu. A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Statistical science*, 27(4):538–557, 2012.

[73] Arkadij Semenovič Nemirovskij and David Borisovich Yudin. Problem complexity and method efficiency in optimization. 1983.

[74] Luong Trung Nguyen, Junhan Kim, and Byonghyo Shim. Low-rank matrix completion: A contemporary survey. *IEEE Access*, 7:94215–94237, 2019.

[75] Yaniv Plan and Roman Vershynin. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Transactions on Information Theory*, 59(1):482–494, 2012.

[76] Yaniv Plan and Roman Vershynin. One-bit compressed sensing by linear programming. *Communications on Pure and Applied Mathematics*, 66(8):1275–1297, 2013.

[77] Yaniv Plan and Roman Vershynin. Dimension reduction by random hyperplane tessellations. *Discrete & Computational Geometry*, 51(2):438–461, 2014.

[78] Yaniv Plan and Roman Vershynin. The generalized lasso with non-linear observations. *IEEE Transactions on information theory*, 62(3):1528–1537, 2016.

[79] Yaniv Plan, Roman Vershynin, and Elena Yudovina. High-dimensional estimation with geometric constraints. *Information and Inference: A Journal of the IMA*, 6(1):1–40, 2017.

[80] Shuang Qiu, Xiaohan Wei, and Zhuoran Yang. Robust one-bit recovery via relu generative networks: Near-optimal statistical rate and global landscape analysis. In *International Conference on Machine Learning*, pages 7857–7866. PMLR, 2020.

[81] Garvesh Raskutti, Martin J Wainwright, and Bin Yu. Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *IEEE transactions on information theory*, 57(10):6976–6994, 2011.

[82] Benjamin Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12(12), 2011.

[83] Lawrence Roberts. Picture coding using pseudo-random noise. *IRE Transactions on Information Theory*, 8(2):145–154, 1962.

[84] Sima Sahu, Harsh Vikram Singh, Basant Kumar, and Amit Kumar Singh. De-noising of ultrasound image using bayesian approached heavy-tailed cauchy distribution. *Multimedia Tools and Applications*, 78(4):4089–4106, 2019.

[85] Vidyashankar Sivakumar, Arindam Banerjee, and Pradeep K Ravikumar. Beyond sub-gaussian measurements: High-dimensional structured estimation with sub-exponential designs. *Advances in neural information processing systems*, 28, 2015.

[86] Alexander Stollenwerk. *One-bit compressed sensing and fast binary embeddings*. PhD thesis, Dissertation, RWTH Aachen University, 2019, 2019.

[87] Zhongxing Sun, Wei Cui, and Yulong Liu. Quantized corrupted sensing with random dithering. *IEEE Transactions on Signal Processing*, 70:600–615, 2022.

[88] Ananthram Swami and Brian M Sadler. On some detection and estimation problems in heavy-tailed noise. *Signal Processing*, 82(12):1829–1846, 2002.

[89] Christos Thrampoulidis and Ankit Singh Rawat. The generalized lasso for sub-gaussian measurements with dithered quantization. *IEEE Transactions on Information Theory*, 66(4):2487–2500, 2020.

[90] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[91] Joel A Tropp. An introduction to matrix concentration inequalities. *arXiv preprint arXiv:1501.01571*, 2015.

[92] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[93] Robert F Woolson and William R Clarke. *Statistical methods for the analysis of biomedical data*. John Wiley & Sons, 2011.

[94] Chunlei Xu and Laurent Jacques. Quantized compressive sensing with rip matrices: The benefit of dithering. *Information and Inference: A Journal of the IMA*, 9(3):543–586, 2020.

[95] Tianyu Yang, Johannes Maly, Sjoerd Dirksen, and Giuseppe Caire. Plug-in channel estimation with dithered quantized signals in spatially non-stationary massive mimo systems. *arXiv preprint arXiv:2301.04641*, 2023.

[96] Hantian Zhang, Jerry Li, Kaan Kara, Dan Alistarh, Ji Liu, and Ce Zhang. Zipml: Training linear models with end-to-end low precision, and a little bit of deep learning. In *International Conference on Machine Learning*, pages 4035–4043. PMLR, 2017.

[97] Ziwei Zhu and Wenjing Zhou. Taming heavy-tailed features by shrinkage. In *International Conference on Artificial Intelligence and Statistics*, pages 3268–3276. PMLR, 2021.

APPENDIX

A. Proofs in Section III

1) **Quantized Covariance Matrix Estimation:** We first provide Bernstein's inequality that is recurring in our proofs. In application, we will choose the more convenient one from (16) and (17).

Lemma 1. (Bernstein's inequality, [6, Thm. 2.10, Coro. 2.11])**.** *Let $X_1, ..., X_n$ be independent random variables, and assume that there exist positive numbers v and c such that $\sum_{i=1}^n \mathbb{E}[X_i^2] \le v$ and*

$$\sum_{i=1}^n \mathbb{E}|X_i|^q \le \frac{q!}{2} v c^{q-2} \text{ for all integers } q \ge 3,$$

then for any $t > 0$ we have

$$\mathbb{P}\left(\Big| \sum_{i=1}^n (X_i - \mathbb{E}X_i) \Big| \ge \sqrt{2vt} + ct \right) \le 2\exp\big(-t\big) \quad (16)$$

$$\mathbb{P}\left(\Big| \sum_{i=1}^n (X_i - \mathbb{E}X_i) \Big| \ge t \right) \le 2\exp\left(-\frac{t^2}{2(v+ct)} \right) \quad (17)$$

We will also use the Matrix Bernstein's inequality.

Lemma 2. (Matrix Bernstein, [91, Thm. 6.1.1])**.** *Let $\boldsymbol{S}_1, ..., \boldsymbol{S}_n$ be independent zero-mean random variables with common dimension $d_1 \times d_2$. We assume that $\|\boldsymbol{S}_k\|_{op} \le L$ for $k \in [n]$ and introduce the matrix variance statistic*

$$\nu = \max\left\{ \Big\| \sum_{k=1}^n \mathbb{E}(\boldsymbol{S}_k \boldsymbol{S}_k^\top) \Big\|_{op}, \Big\| \sum_{k=1}^n \mathbb{E}(\boldsymbol{S}_k^\top \boldsymbol{S}_k) \Big\|_{op} \right\}.$$

Then for any $t \ge 0$, we have

$$\mathbb{P}\left(\Big\| \sum_{k=1}^n \boldsymbol{S}_k \Big\|_{op} \ge t \right) \le (d_1 + d_2) \exp\left(\frac{-\frac{1}{2}t^2}{\nu + \frac{Lt}{3}} \right).$$

Proof of Theorem 2:

Proof. Recall that $\boldsymbol{\xi}_k = \dot{\boldsymbol{x}}_k - \widetilde{\boldsymbol{x}}_k$ is the quantization noise, and $\mathbb{E}(\boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top) = \frac{\Delta^2}{4} \boldsymbol{I}_d$, which implies $\mathbb{E}\widehat{\boldsymbol{\Sigma}} = \mathbb{E}(\widetilde{\boldsymbol{x}}_k \widetilde{\boldsymbol{x}}_k^\top)$. Thus, by using triangle inequality we obtain

$$\begin{aligned} \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^\star\|_\infty &\le \|\widehat{\boldsymbol{\Sigma}} - \mathbb{E}\widehat{\boldsymbol{\Sigma}}\|_\infty + \|\mathbb{E}(\widetilde{\boldsymbol{x}}_k \widetilde{\boldsymbol{x}}_k^\top - \boldsymbol{x}_k \boldsymbol{x}_k^\top)\|_\infty \\ &:= I_1 + I_2. \end{aligned}$$

Step 1. Bounding I_1.

Note that

$$\|\widehat{\boldsymbol{\Sigma}} - \mathbb{E}\widehat{\boldsymbol{\Sigma}}\|_\infty = \Big\| \frac{1}{n} \sum_{k=1}^n \dot{\boldsymbol{x}}_k \dot{\boldsymbol{x}}_k^\top - \mathbb{E}(\dot{\boldsymbol{x}}_k \dot{\boldsymbol{x}}_k^\top) \Big\|_\infty,$$

so for any $(i,j) \in [d] \times [d]$ we aim to bound the (i,j)-th entry error

$$|\widehat{\sigma}_{ij} - \mathbb{E}\widehat{\sigma}_{ij}| = \left| \sum_{k=1}^n \frac{1}{n} \dot{x}_{ki} \dot{x}_{kj} - \mathbb{E}(\dot{x}_{ki} \dot{x}_{kj}) \right|$$

Observe that the quantization noise is bounded as follows

$$\|\boldsymbol{\xi}_k\|_\infty \le \|\mathcal{Q}_\Delta(\widetilde{\boldsymbol{x}}_k + \boldsymbol{\tau}_k) - (\widetilde{\boldsymbol{x}}_k + \boldsymbol{\tau}_k)\|_\infty + \|\boldsymbol{\tau}_k\|_\infty \le \frac{3\Delta}{2},$$

which implies $\mathbb{E}|\xi_{ki}|^4 \le (\frac{3\Delta}{2})^4$ and

$$\|\dot{\boldsymbol{x}}_k\|_\infty \le \|\widetilde{\boldsymbol{x}}_k\|_\infty + \|\boldsymbol{\xi}_k\|_\infty \le \zeta + \frac{3\Delta}{2}.$$

By the moment constraint on x_{ki} we have $\mathbb{E}|\widetilde{x}_{ki}|^4 \le \mathbb{E}|x_{ki}|^4 \le M$. Thus, for any positive integer $p \ge 2$ we have the following bound

$$\begin{aligned} &\sum_{k=1}^n \mathbb{E}\Big| \frac{\dot{x}_{ki} \dot{x}_{kj}}{n} \Big|^q \\ &\le \frac{(\zeta + \frac{3}{2}\Delta)^{2(q-2)}}{n^q} \sum_{k=1}^n \mathbb{E}(\dot{x}_{ki} \dot{x}_{kj})^2 \\ &\le \frac{(\zeta + \frac{3}{2}\Delta)^{2(q-2)}}{2n^q} \sum_{k=1}^n \big(\mathbb{E}|\dot{x}_{ki}|^4 + \mathbb{E}|\dot{x}_{kj}|^4 \big) \\ &\le \frac{4(\zeta + \frac{3}{2}\Delta)^{2(q-2)}}{n^q} \\ &\qquad \cdot \sum_{k=1}^n \big(\mathbb{E}|\widetilde{x}_{ki}|^4 + \mathbb{E}|\widetilde{x}_{kj}|^4 + \mathbb{E}|\xi_{ki}|^4 + \mathbb{E}|\xi_{kj}|^4 \big) \\ &\le \frac{q!}{2} v_0 c_0^{q-2}, \end{aligned} \quad (18)$$

for some $v_0 = O\big(\frac{M+\Delta^4}{n}\big)$, $c_0 = O\big(\frac{(\zeta+\Delta)^2}{n}\big)$. With these preparations, we can invoke Bernstein's inequality (Lemma 1) to obtain that, for any $t \geq 0$, with probability at least $1-2\exp(-t)$,

$$|\widehat{\sigma}_{ij} - \mathbb{E}\widehat{\sigma}_{ij}| \leq C_1\Big(\sqrt{\frac{(M+\Delta^4)t}{n}} + \frac{(\zeta^2+\Delta^2)t}{n}\Big).$$

Taking $t = \delta \log d$ and using the choice $\zeta \asymp \big(\frac{nM}{\delta \log d}\big)^{1/4}$, then applying a union bound over $(i,j) \in [d]\times[d]$, under the scaling $n \gtrsim \delta \log d$, we obtain that

$$I_1 \lesssim (\sqrt{M}+\Delta^2)\sqrt{\frac{\delta \log d}{n}}$$

holds with probability at least $1-2d^{2-\delta}$.

Step 2. Bounding I_2.

We aim to bound $|\mathbb{E}(\widetilde{x}_{ki}\widetilde{x}_{kj} - x_{ki}x_{kj})|$ for any $(i,j) \in [d]\times[d]$. First by the definition of truncation we have

$$\begin{aligned}&\big|\mathbb{E}(\widetilde{x}_{ki}\widetilde{x}_{kj} - x_{ki}x_{kj})\big| \\ \leq&\mathbb{E}\big[|x_{ki}x_{kj}|(\mathbb{1}(|x_{ki}| \geq \zeta) + \mathbb{1}(|x_{kj}| \geq \zeta))\big];\end{aligned}$$

then applying Cauchy-Schwarz to $\mathbb{E}\big[|x_{ki}x_{kj}|\mathbb{1}(|x_{ki}| \geq \zeta|)\big]$, we obtain

$$\begin{aligned}&\mathbb{E}\big[|x_{ki}x_{kj}|\mathbb{1}(|x_{ki}| \geq \zeta)\big] \\ \leq&\big[\mathbb{E}|x_{ki}x_{kj}|^2\big]^{1/2}\big[\mathbb{P}(|x_{ki}| \geq \zeta)\big]^{1/2} \\ \leq&\sqrt{M}\sqrt{\frac{M}{\zeta^4}} = \frac{M}{\zeta^2},\end{aligned}$$

where the second inequality is due to Markov's inequality. Note that this bound remains valid for $\mathbb{E}\big[|x_{ki}x_{kj}|\mathbb{1}(|x_{kj}| \geq \zeta)\big]$. Since this holds for any $(i,j) \in [d]\times[d]$, combining with $\zeta \asymp \big(\frac{nM}{\delta \log d}\big)^{1/4}$, we obtain

$$\|\mathbb{E}(\widetilde{\boldsymbol{x}}_k\widetilde{\boldsymbol{x}}_k^\top - \boldsymbol{x}_k\boldsymbol{x}_k^\top)\|_\infty \leq \frac{2M}{\zeta^2} \lesssim \sqrt{M}\sqrt{\frac{\delta \log d}{n}}.$$

By putting pieces together, we have

$$\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^\star\|_\infty \lesssim (\sqrt{M}+\Delta^2)\sqrt{\frac{\delta \log d}{n}}$$

with probability at least $1-2d^{2-\delta}$, as claimed. □

Proof of Theorem 3:

Proof. Note that the calculations in (5) and (6) remain valid (but the truncated samples are denoted by $\check{\boldsymbol{x}}_k$ rather than $\widetilde{\boldsymbol{x}}_k$), so we have $\mathbb{E}\widehat{\boldsymbol{\Sigma}} = \mathbb{E}(\check{\boldsymbol{x}}_k\check{\boldsymbol{x}}_k^\top)$. Using triangle inequality we first decompose the error as

$$\begin{aligned}\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^\star\|_{op} \leq& \|\widehat{\boldsymbol{\Sigma}} - \mathbb{E}\widehat{\boldsymbol{\Sigma}}\|_{op} + \|\mathbb{E}(\check{\boldsymbol{x}}_k\check{\boldsymbol{x}}_k^\top - \boldsymbol{x}_k\boldsymbol{x}_k^\top)\|_{op} \\ :=& I_1 + I_2.\end{aligned}$$

Step 1. Bounding I_1.

We first write that

$$\widehat{\boldsymbol{\Sigma}} - \mathbb{E}\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{k=1}^n \boldsymbol{S}_k \text{ where } \boldsymbol{S}_k = \dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top - \mathbb{E}(\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top).$$

Recall that we define quantization error as $\boldsymbol{w}_k = \dot{\boldsymbol{x}}_k - \check{\boldsymbol{x}}_k - \boldsymbol{\tau}_k$ and quantization noise as $\boldsymbol{\xi}_k = \dot{\boldsymbol{x}}_k - \check{\boldsymbol{x}}_k$, and observe that the quantization noise is bounded $\|\boldsymbol{\xi}_k\|_\infty = \|\dot{\boldsymbol{x}}_k - \check{\boldsymbol{x}}_k\|_\infty = \|\boldsymbol{\tau}_k + \boldsymbol{w}_k\|_\infty \leq \frac{3}{2}\Delta$. Thus, by $\|\boldsymbol{a}\|_2^2 \leq \sqrt{d}\|\boldsymbol{a}\|_4^2$ that holds for any $\boldsymbol{a} \in \mathbb{R}^d$, we obtain

$$\begin{aligned}\|\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top\|_{op} =& \|\dot{\boldsymbol{x}}_k\|_2^2 = \|\check{\boldsymbol{x}}_k + \boldsymbol{\xi}_k\|_2^2 \leq 2\|\check{\boldsymbol{x}}_k\|_2^2 + 2\|\boldsymbol{\xi}_k\|_2^2 \\ &\leq 2\sqrt{d}\cdot\|\check{\boldsymbol{x}}_k\|_4^2 + 2d\cdot\Big(\frac{3\Delta}{2}\Big)^2 \leq 2\sqrt{d}\zeta^2 + \frac{9}{2}d\Delta^2,\end{aligned}$$

which implies

$$\|\boldsymbol{S}_k\|_{op} \leq \|\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top\|_{op} + \mathbb{E}\|\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top\|_{op} \leq 4\sqrt{d}\zeta^2 + 9d\Delta^2.$$

Moreover, we estimate the matrix variance statistic. Since $\boldsymbol{S}_k$ is symmetric, we simply deal with $\|\mathbb{E}\boldsymbol{S}_k^2\|_{op}$ and some algebra gives $\mathbb{E}\boldsymbol{S}_k^2 = \mathbb{E}\big[\|\dot{\boldsymbol{x}}_k\|_2^2\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top\big] - \big(\mathbb{E}\big[\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top\big]\big)^2$. First let us note that

$$\begin{aligned}\Big\|\big(\mathbb{E}\big[\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top\big]\big)^2\Big\|_{op} =& \Big\|\mathbb{E}\big[\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top\big]\Big\|_{op}^2 \\ =& \Big\|\mathbb{E}\big[\check{\boldsymbol{x}}_k\check{\boldsymbol{x}}_k^\top\big] + \frac{\Delta^2}{4}\boldsymbol{I}_d\Big\|_{op}^2 \\ \leq& \Big(\Big\|\mathbb{E}\big[\check{\boldsymbol{x}}_k\check{\boldsymbol{x}}_k^\top\big]\Big\|_{op} + \frac{\Delta^2}{4}\Big)^2 \\ \leq& 2\Big\|\mathbb{E}\big[\check{\boldsymbol{x}}_k\check{\boldsymbol{x}}_k^\top\big]\Big\|_{op}^2 + \frac{\Delta^4}{8}.\end{aligned}$$

Combining with the observation that

$$\begin{aligned}\Big\|\mathbb{E}\big[\check{\boldsymbol{x}}_k\check{\boldsymbol{x}}_k^\top\big]\Big\|_{op} =& \sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}}\mathbb{E}(\boldsymbol{v}^\top\check{\boldsymbol{x}}_k)^2 \\ &\leq \sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}}\sqrt{\mathbb{E}(\boldsymbol{v}^\top\boldsymbol{x}_k)^4} \leq \sqrt{M},\end{aligned}$$

we obtain $\big\|(\mathbb{E}[\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top])^2\big\|_{op} = O(M+\Delta^4)$. Then we turn to the operator norm of $\mathbb{E}[\|\dot{\boldsymbol{x}}_k\|_2^2\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top]$. We apply Cauchy-Schwarz to estimate

$$\begin{aligned}&\big\|\mathbb{E}\big(\|\dot{\boldsymbol{x}}_k\|_2^2\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top\big)\big\|_{op} \\ =& \sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}}\mathbb{E}\big(\|\dot{\boldsymbol{x}}_k\|_2^2(\boldsymbol{v}^\top\dot{\boldsymbol{x}}_k)^2\big) \\ \leq& \sqrt{\mathbb{E}\|\dot{\boldsymbol{x}}_k\|_2^4}\sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}}\sqrt{\mathbb{E}(\boldsymbol{v}^\top\dot{\boldsymbol{x}}_k)^4}.\end{aligned} \tag{19}$$

By $\|\boldsymbol{a}\|_2^2 \leq \sqrt{d}\|\boldsymbol{a}\|_4^2$ that holds for any $\boldsymbol{a} \in \mathbb{R}^d$, $\mathbb{E}|\check{x}_{ki}|^4 \leq \mathbb{E}|x_{ki}|^4 \leq M$, $\dot{\boldsymbol{x}}_k = \check{\boldsymbol{x}}_k + \boldsymbol{\xi}_k$ and $\|\boldsymbol{\xi}_k\|_\infty \leq \frac{3\Delta}{2}$, we obtain

$$\begin{aligned}\mathbb{E}\|\dot{\boldsymbol{x}}_k\|_2^4 \leq& \mathbb{E}(\|\check{\boldsymbol{x}}_k\|_2 + \|\boldsymbol{\xi}_k\|_2)^4 \lesssim \mathbb{E}(\|\check{\boldsymbol{x}}_k\|_2^4 + \|\boldsymbol{\xi}_k\|_2^4) \\ &\leq d\mathbb{E}(\|\check{\boldsymbol{x}}_k\|_4^4 + \|\boldsymbol{\xi}_k\|_4^4) \lesssim d^2(M+\Delta^4).\end{aligned} \tag{20}$$

For any $\boldsymbol{v} \in \mathbb{S}^{d-1}$, we write $\dot{\boldsymbol{x}}_k = \check{\boldsymbol{x}}_k + \boldsymbol{\tau}_k + \boldsymbol{w}_k$ and then have the bound

$$\begin{aligned}\mathbb{E}(\boldsymbol{v}^\top\dot{\boldsymbol{x}}_k)^4 \lesssim& \mathbb{E}(\boldsymbol{v}^\top\check{\boldsymbol{x}}_k)^4 + \mathbb{E}(\boldsymbol{v}^\top\boldsymbol{\tau}_k)^4 + \mathbb{E}(\boldsymbol{v}^\top\boldsymbol{w}_k)^4 \\ \overset{(i)}{\lesssim}& M + \Delta^4,\end{aligned} \tag{21}$$

where (i) is because

$$\mathbb{E}(\boldsymbol{v}^\top\check{\boldsymbol{x}}_k)^4 \leq \mathbb{E}(\boldsymbol{v}^\top\boldsymbol{x}_k)^4 \leq M$$

and

$$\boldsymbol{\tau}_k \sim \mathscr{U}([-\frac{\Delta}{2},\frac{\Delta}{2}]^d) + \mathscr{U}([-\frac{\Delta}{2},\frac{\Delta}{2}]^d),$$

and the quantization error $\boldsymbol{w}_k$ follows $\mathscr{U}([-\frac{\Delta}{2},\frac{\Delta}{2}]^d)$; in more detail, $\|\boldsymbol{v}^\top\boldsymbol{\tau}_k\|_{\psi_2}, \|\boldsymbol{v}^\top\boldsymbol{w}_k\|_{\psi_2} = O(\Delta)$ and then the

moment constraint of sub-Gaussian random variable implies $\mathbb{E}(\boldsymbol{v}^\top \boldsymbol{\tau}_k)^4 = O(\Delta^4)$ and $\mathbb{E}(\boldsymbol{v}^\top \boldsymbol{w}_k) = O(\Delta^4)$. From (19), (20) and (21), we obtain

$$\big\|\mathbb{E}\big(\|\dot{\boldsymbol{x}}_k\|_2^2 \dot{\boldsymbol{x}}_k \dot{\boldsymbol{x}}_k^\top\big)\big\|_{op} = O\big(d(\Delta^4 + M)\big).$$

Further combining with

$$\mathbb{E}\boldsymbol{S}_k^2 = \mathbb{E}\big[\|\dot{\boldsymbol{x}}_k\|_2^2 \dot{\boldsymbol{x}}_k \dot{\boldsymbol{x}}_k^\top\big] - \big(\mathbb{E}\big[\dot{\boldsymbol{x}}_k \dot{\boldsymbol{x}}_k^\top\big]\big)^2$$

and $\big\|(\mathbb{E}[\dot{\boldsymbol{x}}_k \dot{\boldsymbol{x}}_k^\top])^2\big\|_{op} = O(M+\Delta^4)$, we arrive at $\|\mathbb{E}\boldsymbol{S}_k^2\|_{op} \lesssim d(\Delta^4+M)$ and hence $\big\|\sum_{k=1}^n \mathbb{E}\boldsymbol{S}_k^2\big\|_{op} \lesssim nd(\Delta^4+M)$. With these preparations, Matrix Bernstein's inequality (Lemma 2) yields the following inequality that holds for any $t \geq 0$

$$\mathbb{P}\Big(\|\widehat{\boldsymbol{\Sigma}} - \mathbb{E}\widehat{\boldsymbol{\Sigma}}\|_{op} \geq t\Big) \leq 2d \exp\left(-\frac{C_1 n t^2}{(M+\Delta^4)d + (\sqrt{d}\zeta^2 + d\Delta^2)t}\right).$$

Setting $t = C_2(\sqrt{M}+\Delta^2)\sqrt{\frac{\delta d \log d}{n}}$ with sufficiently large C_2, under the scaling of $n \gtrsim \delta d \log d$ and the threshold $\zeta \asymp (M^{1/4}+\Delta)\big(\frac{n}{\delta \log d}\big)^{1/4}$, we obtain that

$$I_1 = \|\widehat{\boldsymbol{\Sigma}} - \mathbb{E}\widehat{\boldsymbol{\Sigma}}\|_{op} \leq C_2(\sqrt{M}+\Delta^2)\sqrt{\frac{\delta d \log d}{n}}$$

holds with probability at least $1 - 2d^{1-\delta}$.

Step 2. Bounding I_2.

Having bounded the concentration term I_1, we now switch to the bias term

$$I_2 = \sup_{\boldsymbol{v} \in \mathbb{S}^{d-1}} \big|\boldsymbol{v}^\top \mathbb{E}(\check{\boldsymbol{x}}_k \check{\boldsymbol{x}}_k^\top - \boldsymbol{x}_k \boldsymbol{x}_k^\top)\boldsymbol{v}\big|.$$

For any $\boldsymbol{v} \in \mathbb{S}^{d-1}$, because $\check{\boldsymbol{x}}_k$ is obtained from truncating $\boldsymbol{x}_4$ regarding ℓ_4-norm, we have

$$\begin{aligned}
&\big|\boldsymbol{v}^\top \mathbb{E}(\check{\boldsymbol{x}}_k \check{\boldsymbol{x}}_k^\top - \boldsymbol{x}_k \boldsymbol{x}_k^\top)\boldsymbol{v}\big| \\
&= \Big|\mathbb{E}\Big[\big((\boldsymbol{v}^\top \check{\boldsymbol{x}}_k)^2 - (\boldsymbol{v}^\top \boldsymbol{x}_k)^2\big)\mathbb{1}(\|\boldsymbol{x}_k\|_4 \geq \zeta)\Big]\Big| \\
&\leq \mathbb{E}\big[(\boldsymbol{v}^\top \boldsymbol{x}_k)^2 \mathbb{1}(\|\boldsymbol{x}_k\|_4 \geq \zeta)\big] \\
&\overset{(i)}{\leq} \sqrt{\mathbb{E}(\boldsymbol{v}^\top \boldsymbol{x}_k)^4}\sqrt{\mathbb{P}(\|\boldsymbol{x}_k\|_4^4 \geq \zeta^4)} \\
&\overset{(ii)}{\leq} \sqrt{M\frac{\mathbb{E}\|\boldsymbol{x}_k\|_4^4}{\zeta^4}} \overset{(iii)}{\lesssim} \sqrt{\frac{M\delta d \log d}{n}},
\end{aligned}$$

where (i) and (ii) are respectively by Cauchy-Schwarz and Markov's, and in (iii) we use $\zeta \asymp (M^{1/4}+\Delta)\big(\frac{\delta d \log d}{n}\big)^{1/4}$. This leads to the bound $I_2 \lesssim \sqrt{\frac{M\delta d \log d}{n}}$. Combining the bounds of I_1, I_2 completes the proof. □

Proof of Theorem 4: This small appendix is devoted to the proof of Theorem 4, for which we need a Lemma concerning the element-wise error rate of $\widehat{\boldsymbol{\Sigma}}_s$, i.e., $|\breve{\sigma}_{ij} - \sigma_{ij}^\star|$ where we write $\widehat{\boldsymbol{\Sigma}}_s = [\breve{\sigma}_{ij}]$, $\boldsymbol{\Sigma}^\star = \mathbb{E}(\boldsymbol{x}_k \boldsymbol{x}_k^\top) = [\sigma_{ij}^\star]$. Recalling that $\widehat{\boldsymbol{\Sigma}}_s = \mathcal{T}_\mu(\widehat{\boldsymbol{\Sigma}})$, the key message from Lemma 3 is that due to the thresholding operator $\mathcal{T}_\mu(\cdot)$, $\widehat{\boldsymbol{\Sigma}}_s$ respects an element-wise bound tighter than $O\big(\sqrt{\frac{\delta \log d}{n}}\big)$ in Theorem 2, as can be seen from the additional branch $|\sigma_{ij}^\star|$ in (22).

Lemma 3. (Element-wise Error Rate of $\widehat{\boldsymbol{\Sigma}}_s$)**.** *For any $i, j \in [d]$, the thresholding estimator $\widehat{\boldsymbol{\Sigma}}_s = [\breve{\sigma}_{ij}]$ in Theorem 4 satisfies for some C that*

$$\mathbb{P}\left(|\breve{\sigma}_{ij} - \sigma_{ij}^\star| \leq C \min\left\{|\sigma_{ij}^\star|, \mathscr{L}\sqrt{\frac{\delta \log d}{n}}\right\}\right) \geq 1 - 2d^{-\delta} \tag{22}$$

where $\mathscr{L} := \sqrt{M} + \Delta^2$.

Proof. Recall that $\widehat{\boldsymbol{\Sigma}}_s = [\breve{\sigma}_{ij}] = \mathcal{T}_\mu(\widehat{\boldsymbol{\Sigma}}) = \mathcal{T}_\mu\big([\widehat{\sigma}_{ij}]\big)$ and hence $\breve{\sigma}_{ij} = \mathcal{T}_\mu(\widehat{\sigma}_{ij})$. Given (i,j), the proof of Theorem 2 delivers

$$|\widehat{\sigma}_{ij} - \sigma_{ij}^\star| \leq C_1 \mathscr{L}\sqrt{\frac{\delta \log d}{n}}$$

with probability at least $1 - 2d^{-\delta}$. Assume that we are on this event in the following analyses. As stated in Theorem 4, we set $\mu = C_2 \mathscr{L}\sqrt{\frac{\delta \log d}{n}}$ with $C_2 > C_1$, $\mathscr{L} = \sqrt{M} + \Delta^2$. Since $\breve{\sigma}_{ij} = \mathcal{T}_\mu(\widehat{\sigma}_{ij})$, we discuss whether $|\widehat{\sigma}| \geq \mu$ holds.

Case 1. when $|\widehat{\sigma}_{ij}| < \mu$ holds.

In this case we have $\breve{\sigma}_{ij} = 0$, thus $|\breve{\sigma}_{ij} - \sigma_{ij}^\star| \leq |\sigma_{ij}^\star|$. Further note that

$$\begin{aligned}
|\sigma_{ij}^\star| &\leq |\sigma_{ij}^\star - \widehat{\sigma}_{ij}| + |\widehat{\sigma}_{ij}| \\
&\leq C_1 \mathscr{L}\sqrt{\frac{\delta \log d}{n}} + \mu \lesssim \mathscr{L}\sqrt{\frac{\delta \log d}{n}},
\end{aligned}$$

so we also have $|\breve{\sigma}_{ij} - \sigma_{ij}^\star| \lesssim \mathscr{L}\sqrt{\frac{\delta \log d}{n}}$.

Case 2. when $|\widehat{\sigma}_{ij}| \geq \mu$ holds.

We consider $|\widehat{\sigma}_{ij}| \geq \mu$ that implies $\breve{\sigma}_{ij} = \widehat{\sigma}_{ij}$, then we have

$$|\breve{\sigma}_{ij} - \sigma_{ij}^\star| = |\widehat{\sigma}_{ij} - \sigma_{ij}^\star| \leq C_1 \mathscr{L}\sqrt{\frac{\delta \log d}{n}}.$$

Moreover, note that

$$\begin{aligned}
|\sigma_{ij}^\star| &\geq |\widehat{\sigma}_{ij}| - |\widehat{\sigma}_{ij} - \sigma_{ij}^\star| \geq \mu - |\widehat{\sigma}_{ij} - \sigma_{ij}^\star| \\
&\geq (C_2 - C_1)\mathscr{L}\sqrt{\frac{\delta \log d}{n}}
\end{aligned}$$

so we also have $|\breve{\sigma}_{ij} - \sigma_{ij}^\star| = O(|\sigma_{ij}^\star|)$.

Therefore, in both cases we have proved that $|\breve{\sigma}_{ij} - \sigma_{ij}^\star| \lesssim \min\big\{|\sigma_{ij}^\star|, \mathscr{L}\sqrt{\frac{\delta \log d}{n}}\big\}$, which completes the proof. □

We are now in a position to present the proof.

Proof of Theorem 4. We let $p = \frac{\delta}{4} \geq 1$ (just assume $\delta \geq 4$) and use $B_0 := \mathscr{L}\sqrt{\frac{\delta \log d}{n}}$ as shorthand. For $(i,j) \in [d] \times [d]$ we define the event $\mathscr{A}_{ij}$ as

$$\mathscr{A}_{ij} = \Big\{|\breve{\sigma}_{ij} - \sigma_{ij}^\star| \leq C_1 \min\big\{|\sigma_{ij}^\star|, B_0\big\}\Big\}.$$

By Lemma 3 we can choose C_1 to be sufficiently large such that $C_1 B_0 > 3\mu$ and $\mathbb{P}(\mathscr{A}_{ij}^\complement) \leq 2d^{-\delta}$; here, by convention we let $\mathscr{A}_{ij}^\complement$ be the complement of $\mathscr{A}_{ij}$. Our proof strategy is to first bound the p-th order moment $\mathbb{E}\|\widehat{\boldsymbol{\Sigma}}_s - \boldsymbol{\Sigma}^\star\|_{op}^p$, and then invoke Markov's inequality to derive a high probability bound. We start with the simple estimate displayed in (23), where (i) and (ii) are due to $\|\boldsymbol{A}\|_{op} \leq \sup_{j \in [d]} \sum_{i \in [d]} |a_{ij}|$ for symmetric $\boldsymbol{A}$ and $(a+b)^p \leq (2a)^p + (2b)^p$. In this proof,

$$\begin{aligned}\mathbb{E}\|\widehat{\boldsymbol{\Sigma}}_s-\boldsymbol{\Sigma}^\star\|_{op}^p &\overset{(i)}{\le} \mathbb{E}\Big(\sup_{j\in[d]}\sum_{i=1}^d|\breve{\sigma}_{ij}-\sigma_{ij}^\star|\mathbb{1}(\mathscr{A}_{ij})+\sup_{j\in[d]}\sum_{i=1}^d|\breve{\sigma}_{ij}-\sigma_{ij}^\star|\mathbb{1}(\mathscr{A}_{ij}^{\complement})\Big)^p\\ &\overset{(ii)}{\le} 2^p\mathbb{E}\sup_{j\in[d]}\Big(\sum_{i=1}^d|\breve{\sigma}_{ij}-\sigma_{ij}^\star|\mathbb{1}(\mathscr{A}_{ij})\Big)^p+2^p\mathbb{E}\sup_{j\in[d]}\Big(\sum_{i=1}^d|\breve{\sigma}_{ij}-\sigma_{ij}^\star|\mathbb{1}(\mathscr{A}_{ij}^{\complement})\Big)^p:=I_1+I_2\end{aligned} \tag{23}$$

$$\begin{aligned}W_j &\overset{(i)}{\le}\Big(\sum_{i=1}^d|\sigma_{ij}^\star|\mathbb{1}(\mathscr{A}_{ij}^{\complement})\mathbb{1}(|\widehat{\sigma}_{ij}|<\mu)+\sum_{i=1}^d|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|\mathbb{1}(\mathscr{A}_{ij}^{\complement})+\sum_{i=1}^d|\widetilde{\sigma}_{ij}-\sigma_{ij}^\star|\mathbb{1}(\mathscr{A}_{ij}^{\complement})\Big)^p\\ &\le(3d)^{p-1}\Big(\sum_{i=1}^d|\sigma_{ij}^\star|^p\mathbb{1}(\mathscr{A}_{ij}^{\complement})\mathbb{1}(|\widehat{\sigma}_{ij}|<\mu)+\sum_{i=1}^d|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|^p\mathbb{1}(\mathscr{A}_{ij}^{\complement})+\sum_{i=1}^d|\widetilde{\sigma}_{ij}-\sigma_{ij}^\star|^p\mathbb{1}(\mathscr{A}_{ij}^{\complement})\Big),\end{aligned} \tag{25}$$

the ranges of indices in summation or supremum, if omitted, are $[d]$.

Step 1. Bounding I_1.

By the definition of $\mathscr{A}_{ij}$, $|\breve{\sigma}_{ij}-\sigma_{ij}^\star|=0$ if $|\sigma_{ij}^\star|=0$. Because the columns of $\boldsymbol{\Sigma}^\star$ are s-sparse, we can straightforwardly bound I_1 as follows:

$$\begin{aligned}I_1&=2^p\mathbb{E}\sup_j\Big(\sum_{i:|\sigma_{ij}^\star|>0}|\breve{\sigma}_{ij}-\sigma_{ij}^\star|\mathbb{1}(\mathscr{A}_{ij})\Big)^p\\ &\le\big(2C_1sB_0\big)^p.\end{aligned} \tag{24}$$

Step 2. Bounding I_2.

We first write $I_2=2^p\mathbb{E}\sup_j W_j$ with

$$W_j:=\big(\sum_i|\breve{\sigma}_{ij}-\sigma_{ij}^\star|\mathbb{1}(\mathscr{A}_{ij}^{\complement})\big)^p,$$

then start from the display (25), where in (i) we define

$$\mathbb{E}\widehat{\sigma}_{ij}=\mathbb{E}(\widetilde{x}_{ki}\widetilde{x}_{kj}):=\widetilde{\sigma}_{ij}.$$

By replacing $\sup_j$ with $\sum_j$, this further gives

$$\begin{aligned}I_2&\le 6^pd^{p-1}\Big(\sum_{i,j}|\sigma_{ij}^\star|^p\mathbb{E}\big[\mathbb{1}(\mathscr{A}_{ij}^{\complement})\mathbb{1}(|\widehat{\sigma}_{ij}|<\mu)\big]\\ &\quad+\sum_{i,j}\mathbb{E}\big[|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|^p\mathbb{1}(\mathscr{A}_{ij}^{\complement})\big]\\ &\quad+\sum_{i,j}|\widetilde{\sigma}_{ij}-\sigma_{ij}^\star|^p\mathbb{P}(\mathscr{A}_{ij}^{\complement})\Big)\\ &:=6^pd^{p-1}\big(I_{21}+I_{22}+I_{23}\big).\end{aligned} \tag{26}$$

Step 2.1. Bounding I_{21}.

Note that $\mathscr{A}_{ij}^{\complement}$ means $|\breve{\sigma}_{ij}-\sigma_{ij}^\star|>C_1\min\{|\sigma_{ij}^\star|,B_0\}$, and $|\widehat{\sigma}_{ij}|<\mu$ implies $\breve{\sigma}_{ij}=0$, their combination thus allows us to proceed as the following (i) and (iii):

$$|\sigma_{ij}^\star|\overset{(i)}{>}C_1B_0\overset{(ii)}{>}3\mu\overset{(iii)}{>}3|\widehat{\sigma}_{ij}|\ge3|\sigma_{ij}^\star|-3|\widehat{\sigma}_{ij}-\sigma_{ij}^\star|,$$

where (ii) is due to our choice of C_1. Thus, $\mathscr{A}_{ij}^{\complement}\cap\{|\widehat{\sigma}_{ij}|<\mu\}$ implies $|\widehat{\sigma}_{ij}-\sigma_{ij}^\star|>\frac{2}{3}|\sigma_{ij}^\star|$ and $|\sigma_{ij}^\star|>3\mu$. Note that Step 2 in the proof of Theorem 2 gives $|\widetilde{\sigma}_{ij}-\sigma_{ij}^\star|=O\big(B_0\big)$, and hence we can assume $\mu>|\widetilde{\sigma}_{ij}-\sigma_{ij}^\star|$ and so $|\sigma_{ij}^\star|>3|\widetilde{\sigma}_{ij}-\sigma_{ij}^\star|$. Using these relations and triangle inequality, we obtain

$$\begin{aligned}\frac{2}{3}|\sigma_{ij}^\star|<|\widehat{\sigma}_{ij}-\sigma_{ij}^\star|&\le|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|+|\widetilde{\sigma}_{ij}-\sigma_{ij}^\star|\\ &<|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|+\frac{1}{3}|\sigma_{ij}^\star|,\end{aligned}$$

which implies $|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|>\frac{1}{3}|\sigma_{ij}^\star|$. Now we conclude that, $\mathscr{A}_{ij}^{\complement}\cap\{|\widehat{\sigma}_{ij}|<\mu\}$ implies $|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|>\frac{1}{3}|\sigma_{ij}^\star|$ and $|\sigma_{ij}^\star|>3\mu$, which allows us to bound I_{21} as

$$I_{21}=\sum_{i,j}|\sigma_{ij}^\star|^p\mathbb{1}(|\sigma_{ij}^\star|>3\mu)\mathbb{P}\Big(|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|>\frac{1}{3}|\sigma_{ij}^\star|\Big). \tag{27}$$

Analogously to the proof of Theorem 2, we can apply Bernstein's inequality to $\mathbb{P}\big(|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|>\frac{1}{3}|\sigma_{ij}^\star|\big)$. More specifically, by preparations as in (18), we can use (17) in Lemma 1 with

$$v=O\big(\frac{M+\Delta^4}{n}\big),\ c=O\big(\frac{\zeta^2+\Delta^2}{n}\big)=O\big(\frac{\Delta^2}{n}+\sqrt{\frac{M}{n\delta\log d}}\big)$$

(recall that $\zeta\asymp\big(\frac{nM}{\Delta\log d}\big)^{1/4}$). For some absolute constants C_2,C_3, it gives the display (28), where in (i) we use

$$\min\Big\{\frac{|\sigma_{ij}^\star|}{M+\Delta^4},\Delta^{-2},\sqrt{\frac{\delta\log d}{nM}}\Big\}\gtrsim\frac{1}{\sqrt{M}+\Delta^2}\sqrt{\frac{\delta\log d}{n}}$$

that holds because $|\sigma_{ij}^\star|>3\mu$ and $n\gtrsim\delta\log d$. We substitute (28) into (27) and perform some estimates as in the display (29), where in (i) we substitute $|\sigma_{ij}^\star|>3\mu$ from the indicator function into the exponent, (ii) is because $\sup_{t\ge0}t^p\exp\big(-\frac{t}{2}\big)\le p^p$, $p=\frac{\delta}{4}$, and we consider $\mu=C_4(\sqrt{M}+\Delta^2)\sqrt{\frac{\delta\log d}{n}}$ with C_4 large enough.

Step 2.2. Bounding I_{22}.

Then, we deal with I_{22} by Cauchy-Schwarz

$$I_{22}\le\sum_{i,j}\sqrt{\mathbb{E}|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|^{2p}}\sqrt{\mathbb{P}(\mathscr{A}_{ij}^{\complement})}.$$

As in (28), we can use (17) in Lemma 1 with $v=O\big(\frac{M+\Delta^4}{n}\big)$ and $c=O\big(\frac{\Delta^2}{n}+\sqrt{\frac{M}{n\delta\log d}}\big)$, yielding that for any $t\ge0$, $\mathbb{P}\big(|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|\ge t\big)\le2\exp\big(-\frac{t^2}{2(v+ct)}\big)\le2\exp\big(-\frac{t^2}{4v}\big)+$

$$\begin{aligned}&\mathbb{P}\Big(|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|>\frac{1}{3}|\sigma^\star_{ij}|\Big)\le 2\exp\left(-\frac{|\sigma^\star_{ij}|^2}{C_2\big\{\frac{M+\Delta^4}{n}+\frac{\Delta^2|\sigma^\star_{ij}|}{n}+\sqrt{\frac{M}{n\delta\log d}}|\sigma^\star_{ij}|\big\}}\right)\\ \le&2\exp\left(-\frac{3n|\sigma^\star_{ij}|}{C_2}\min\Big\{\frac{|\sigma^\star_{ij}|}{M+\Delta^4},\frac{1}{\Delta^2},\sqrt{\frac{\delta\log d}{nM}}\Big\}\right)\overset{(i)}{\le}2\exp\left(-\frac{C_3|\sigma^\star_{ij}|\sqrt{n\delta\log d}}{\sqrt{M}+\Delta^2}\right),\end{aligned}\tag{28}$$

$$\begin{aligned}I_{21}&\le 2\sum_{i,j}|\sigma^\star_{ij}|^p\mathbb{1}\big(|\sigma^\star_{ij}|>3\mu\big)\exp\left(-\frac{C_3|\sigma^\star_{ij}|\sqrt{n\delta\log d}}{\sqrt{M}+\Delta^2}\right)\\ &=2\sum_{i,j}\left(\frac{\sqrt{M}+\Delta^2}{C_3\sqrt{n\delta\log d}}\right)^p\cdot\left(\frac{C_3|\sigma^\star_{ij}|\sqrt{n\delta\log d}}{\sqrt{M}+\Delta^2}\right)^p\exp\left(-\frac{0.5C_3|\sigma^\star_{ij}|\sqrt{n\delta\log d}}{\sqrt{M}+\Delta^2}\right)\\ &\qquad\cdot\exp\left(-\frac{0.5C_3|\sigma^\star_{ij}|\sqrt{n\delta\log d}}{\sqrt{M}+\Delta^2}\right)\mathbb{1}\big(|\sigma^\star_{ij}|>3\mu\big)\\ &\overset{(i)}{\le}2\sum_{i,j}\left(\frac{\sqrt{M}+\Delta^2}{C_3\sqrt{n\delta\log d}}\right)^p\cdot\Big(\sup_{t\ge0}\ t^p\exp\Big(-\frac{t}{2}\Big)\Big)\cdot\exp\left(-\frac{3C_3}{2}\frac{\sqrt{n\delta\log d}\cdot\mu}{\sqrt{M}+\Delta^2}\right)\\ &\overset{(ii)}{\le}2d^{2-10\delta}\left(\frac{\sqrt{M}+\Delta^2}{C_3}\sqrt{\frac{\delta}{n\log d}}\right)^p\le 2d^{2-10\delta}(C_3^{-1}B_0)^p,\end{aligned}\tag{29}$$

$2\exp\big(-\frac{t}{4c}\big)$. Based on this probability tail bound, we can bound the moment via integral as follows

$$\begin{aligned}&\mathbb{E}|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|^{2p}\\ &=2p\int_0^\infty t^{2p-1}\mathbb{P}(|\widehat{\sigma}_{ij}-\mathbb{E}\widehat{\sigma}_{ij}|>t)\ \mathrm{d}t\\ &\le 4p\int_0^\infty t^{2p-1}\big(\exp(-\frac{t^2}{4v})+\exp(-\frac{t}{4c})\big)\ \mathrm{d}t\\ &=2\big[(4v)^p\Gamma(p+1)+(4c)^{2p}\Gamma(2p+1)\big]\\ &\overset{(i)}{\le}2\big[(4vp)^p+(8cp)^{2p}\big],\end{aligned}$$

where we use $\Gamma(p+1)\le p^p$, $\Gamma(2p+1)\le(2p)^{2p}$ in (i) under suitably large p. Thus, it follows that

$$\begin{aligned}I_{22}&\le\sum_{i,j}2d^{-\frac{\delta}{2}}\sqrt{(4vp)^p+(8cp)^{2p}}\\ &\le 2d^{2-\frac{\delta}{2}}\big[(2\sqrt{pv})^p+(8cp)^p\big]\overset{(i)}{\le}2d^{2-\frac{\delta}{2}}(C_4B_0)^p,\end{aligned}$$

where (i) is due to $2\sqrt{pv}\le(\sqrt{M}+\Delta^2)\sqrt{\frac{\delta}{n}}$ and $8cp=\frac{2\Delta^2\delta}{n}+2\sqrt{\frac{\delta M}{n\log d}}$ (recall that $p=\frac{\delta}{4}$).

Step 2.3. Bounding I_{23}.

From Step 2 in the proof of Theorem 2 we have $|\widetilde{\sigma}_{ij}-\sigma^\star_{ij}|\le C_5B_0$. This directly leads to

$$I_{23}\le d^2\cdot 2d^{-\delta}\cdot(C_5B_0)^p=2d^{2-\delta}(C_5B_0)^p.$$

We are in a position to combine everything and conclude the proof. Putting all pieces into (26), it follows that $I_2\le d^{1-\frac{\delta}{4}}(C_6B_0)^p$. Assuming $\delta\ge4$, such upper bound is dominated by (24) for I_1, we can hence conclude that $\mathbb{E}\|\widehat{\mathbf{\Sigma}}_s-\mathbf{\Sigma}^\star\|_{op}^p\le(C_6sB_0)^p$. Therefore, by Markov's inequality,

$$\begin{aligned}&\mathbb{P}(\|\widehat{\mathbf{\Sigma}}_s-\mathbf{\Sigma}^\star\|_{op}\ge C_6esB_0)\\ \le&\frac{\mathbb{E}\|\widehat{\mathbf{\Sigma}}_s-\mathbf{\Sigma}^\star\|_{op}^p}{(C_6esB_0)^p}\le\exp(-p)=\exp\Big(-\frac{\delta}{4}\Big),\end{aligned}$$

which completes the proof. □

2) **Quantized Compressed Sensing:** Note that our estimation procedure in QCS, QMC falls in the framework of regularized M-estimator, see [22], [36], [70] for instance. Particularly, we introduce the following deterministic result for analysing the estimator (10).

Lemma 4. (Adapted from [22, Coro. 2])**.** *Consider (8) and the estimator $\widehat{\boldsymbol{\theta}}$ defined in (10), let $\widehat{\mathbf{\Upsilon}}:=\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^\star$ be the estimation error. If $\boldsymbol{Q}$ is positive semi-definite, and $\lambda\ge2\|\boldsymbol{Q}\boldsymbol{\theta}^\star-\boldsymbol{b}\|_\infty$, then it holds that $\|\widehat{\mathbf{\Upsilon}}\|_1\le10\sqrt{s}\|\widehat{\mathbf{\Upsilon}}\|_2$. Moreover, if for some $\kappa>0$ we have the restricted strong convexity (RSC) $\widehat{\mathbf{\Upsilon}}^\top\boldsymbol{Q}\widehat{\mathbf{\Upsilon}}\ge\kappa\|\widehat{\mathbf{\Upsilon}}\|_2^2$, then we have the error bounds $\|\widehat{\mathbf{\Upsilon}}\|_2\le30\sqrt{s}\big(\frac{\lambda}{\kappa}\big)$ and $\|\widehat{\mathbf{\Upsilon}}\|_1\le300s\big(\frac{\lambda}{\kappa}\big)$.*[14]

To establish the RSC condition, a convenient way is to use the matrix deviation inequality. The following Lemma is adapted from [61], by combining Theorem 3 and Remark 1 therein.[15]

Lemma 5. (Adapted from [61, Thm. 3])**.** *Assume $\boldsymbol{A}\in\mathbb{R}^{n\times d}$ has independent zero-mean sub-Gaussian rows $\boldsymbol{\alpha}_k^\top$s satisfying $\|\boldsymbol{\alpha}_k\|_{\psi_2}\le K$, and the eigenvalues of $\mathbf{\Sigma}:=\mathbb{E}(\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^\top)$ are between $[\kappa_0,\kappa_1]$ for some $\kappa_1\ge\kappa_0>0$. For $\mathcal{T}\subset\mathbb{R}^d$ we let*

[14]We do not optimize the constants in Lemmas 4, 6 for easy reference.

[15]The dependence on K can be further refined [49], while this is not pursued in the present paper.

$\mathrm{rad}(\mathcal{T}) = \sup_{\boldsymbol{x}\in\mathcal{T}} \|\boldsymbol{x}\|_2$ *be its radius. Then with probability at least* $1-\exp(-u^2)$*, it holds that*

$$\sup_{\boldsymbol{x}\in\mathcal{T}} \Big| \|\boldsymbol{A}\boldsymbol{x}\|_2 - \sqrt{n}\|\sqrt{\boldsymbol{\Sigma}}\boldsymbol{x}\|_2 \Big| \le \frac{C\sqrt{\kappa_1}K^2}{\kappa_0}\Big(\omega(\mathcal{T}) + u\cdot \mathrm{rad}(\mathcal{T})\Big),$$

where $\omega(\mathcal{T}) = \mathbb{E}\sup_{\boldsymbol{v}\in\mathcal{T}}[\boldsymbol{g}^\top \boldsymbol{v}]$ *with* $\boldsymbol{g}\sim\mathcal{N}(0,\boldsymbol{I}_d)$ *is the Gaussian width of* $\mathcal{T}$*.*

Based on Lemma 4, the proofs of Theorems 5-6 are divided into two steps, i.e., showing $\lambda \ge 2\|\boldsymbol{Q}\boldsymbol{\theta}^\star - \boldsymbol{b}\|_\infty$ and verifying the RSC. While we still have full $\boldsymbol{x}_k$ in Theorems 5-6, we will study the more challenging settings where the covariates $\boldsymbol{x}_k$s are also quantized via $\mathcal{Q}_{\bar{\Delta}}(\cdot)$ in Theorems 9-10, in which we can take $\bar{\Delta}=0$ to return the settings of Theorems 5-6. Using such perspective, for most technical ingredients (e.g., the verification of $\lambda \ge 2\|\boldsymbol{Q}\boldsymbol{\theta}^\star - \boldsymbol{b}\|_\infty$) in the proofs of Theorems 5-6 we can simply refer to the counterparts established in the proofs of Theorems 9-10. This avoids repetition and will be explained in the proofs more clearly.

Proof of Theorem 5: *Proof.* We divide the proofs into two steps.

Step 1. Proving $\lambda \ge 2\|\boldsymbol{Q}\boldsymbol{\theta}^\star - \boldsymbol{b}\|_\infty$

Recall that we choose $\boldsymbol{Q} = \frac{1}{n}\sum_{k=1}^n \boldsymbol{x}_k\boldsymbol{x}_k^\top$ and $\boldsymbol{b} = \frac{1}{n}\sum_{k=1}^n \dot{y}_k\boldsymbol{x}_k$. In the setting of Theorem 9, the process of obtaining $\dot{y}_k$ remains the same, while the covariates $\boldsymbol{x}_k$s are further quantized to $\dot{\boldsymbol{x}}_k = \mathcal{Q}_{\bar{\Delta}}(\boldsymbol{x}_k + \boldsymbol{\tau_k})$ for some $\bar{\Delta} > 0$ under triangular dither $\boldsymbol{\tau}_k \sim \mathscr{U}([-\frac{\bar{\Delta}}{2},\frac{\bar{\Delta}}{2}]^d) + \mathscr{U}([-\frac{\bar{\Delta}}{2},\frac{\bar{\Delta}}{2}]^d)$, and we choose

$$\boldsymbol{Q} = \frac{1}{n}\sum_{k=1}^n \dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top - \frac{\bar{\Delta}^2}{4}\boldsymbol{I}_d,\ \boldsymbol{b} = \frac{1}{n}\sum_{k=1}^n \dot{y}_k\dot{\boldsymbol{x}}_k$$

there. As a result, by considering $\bar{\Delta}=0$, it can be implied by Step 1 in the proof of Theorem 9 that under the choice

$$\lambda = C_1\frac{\sigma^2}{\sqrt{\kappa_0}}(\Delta + M^{1/(2l)})\sqrt{\frac{\delta\log d}{n}}$$

with sufficiently large C_1,

$$\lambda \ge 2\Big\|\frac{1}{n}\sum_{k=1}^n \boldsymbol{x}_k\boldsymbol{x}_k^\top\boldsymbol{\theta}^\star - \frac{1}{n}\sum_{k=1}^n \dot{y}_k\boldsymbol{x}_k\Big\|_\infty$$

holds with probability at least $1-8d^{1-\delta}$. Then, by using Lemma 4 we obtain $\|\widehat{\boldsymbol{\Upsilon}}\|_1 \le 10\sqrt{s}\|\widehat{\boldsymbol{\Upsilon}}\|_2$.

Step 2. Verifying the RSC $\widehat{\boldsymbol{\Upsilon}}^\top\boldsymbol{Q}\widehat{\boldsymbol{\Upsilon}} \ge \kappa\|\widehat{\boldsymbol{\Upsilon}}\|_2^2$

We refer to Step 2 in the proof of Theorem 9. In particular, with the choices $\bar{\Delta}=0$ and $\boldsymbol{v}=\widehat{\boldsymbol{\Upsilon}}$ in (38), combined with $\|\widehat{\boldsymbol{\Upsilon}}\|_1 \le 10\sqrt{s}\|\widehat{\boldsymbol{\Upsilon}}\|_2$, we obtain

$$\begin{aligned}\frac{1}{\sqrt{n}}\|\boldsymbol{X}\widehat{\boldsymbol{\Delta}}\|_2 &\ge \sqrt{\kappa_0}\|\widehat{\boldsymbol{\Upsilon}}\|_2 - \frac{C_2\sqrt{\kappa_1}\sigma^2}{\kappa_0}\sqrt{\frac{\delta s\log d}{n}}\|\widehat{\boldsymbol{\Upsilon}}\|_2 \\ &\ge \frac{1}{2}\sqrt{\kappa_0}\|\widehat{\boldsymbol{\Upsilon}}\|_2,\end{aligned}$$

where the last inequality is due to the assumed scaling $n \gtrsim \delta s\log d$. With these preparations, a direct application of Lemma 4 completes the proof. □

Proof of Theorem 6: *Proof.* The proof is similarly based on Lemma 4.

Step 1. Proving $\lambda \ge 2\|\boldsymbol{Q}\boldsymbol{\theta}^\star - \boldsymbol{b}\|_\infty$

Recall that we choose $\boldsymbol{Q} = \frac{1}{n}\sum_{k=1}^n \widetilde{\boldsymbol{x}}_k\widetilde{\boldsymbol{x}}_k^\top$ and $\boldsymbol{b} = \frac{1}{n}\sum_{k=1}^n \dot{y}_k\widetilde{\boldsymbol{x}}_k$. In the setting of Theorem 10, the process of obtaining $\dot{y}_k$ remains the same, while the truncated covariates $\widetilde{\boldsymbol{x}}_k$s are further quantized to $\dot{\boldsymbol{x}}_k = \mathcal{Q}_{\bar{\Delta}}(\widetilde{\boldsymbol{x}}_k + \boldsymbol{\tau}_k)$ for some $\bar{\Delta} \ge 0$ under triangular dither $\boldsymbol{\tau}_k \sim \mathscr{U}([-\frac{\bar{\Delta}}{2},\frac{\bar{\Delta}}{2}]^d) + \mathscr{U}([-\frac{\bar{\Delta}}{2},\frac{\bar{\Delta}}{2}]^d)$, and we choose

$$\boldsymbol{Q} = \frac{1}{n}\sum_{k=1}^n \dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top - \frac{\bar{\Delta}^2}{4}\boldsymbol{I}_d,\ \boldsymbol{b} = \frac{1}{n}\sum_{k=1}^n \dot{y}_k\dot{\boldsymbol{x}}_k$$

there. As a result, by considering $\bar{\Delta}=0$, it can be implied by step 1 in the proof of Theorem 10 that, our choice

$$\lambda = C_1(R\sqrt{M} + \Delta^2)\sqrt{\frac{\delta\log d}{n}}$$

with sufficiently large C_1 ensures $\lambda \ge 2\|\boldsymbol{Q}\boldsymbol{\theta}^\star - \boldsymbol{b}\|_\infty$ with the promised probability. By Lemma 4 we obtain $\|\widehat{\boldsymbol{\Upsilon}}\|_1 \le 10\sqrt{s}\|\widehat{\boldsymbol{\Upsilon}}\|_2$.

Step 2. Verifying the RSC $\widehat{\boldsymbol{\Upsilon}}^\top\boldsymbol{Q}\widehat{\boldsymbol{\Upsilon}} \ge \kappa\|\widehat{\boldsymbol{\Upsilon}}\|_2^2$

Unlike the case of sub-Gaussian covariate that is based on matrix deviation inequality (Lemma 5), here we establish a lower bound for $\widehat{\boldsymbol{\Upsilon}}^\top\boldsymbol{Q}\widehat{\boldsymbol{\Upsilon}}$ using the bound on $\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty$ (Theorem 2). Specifically, setting $\Delta=0$ in Theorem 2 yields that, $\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty \lesssim \sqrt{\frac{\delta M\log d}{n}}$ holds with probability at least $1-2d^{2-\delta}$, which allows us to proceed as follows:

$$\begin{aligned}\widehat{\boldsymbol{\Upsilon}}^\top\boldsymbol{Q}\widehat{\boldsymbol{\Upsilon}} &= \widehat{\boldsymbol{\Upsilon}}^\top\boldsymbol{\Sigma}^\star\widehat{\boldsymbol{\Upsilon}} - \widehat{\boldsymbol{\Upsilon}}^\top(\boldsymbol{\Sigma}^\star - \boldsymbol{Q})\widehat{\boldsymbol{\Upsilon}} \\ &\overset{(i)}{\ge} \kappa_0\|\widehat{\boldsymbol{\Upsilon}}\|_2^2 - \sqrt{\frac{\delta M\log d}{n}}\|\widehat{\boldsymbol{\Upsilon}}\|_1^2 \\ &\overset{(ii)}{\ge} \Big(\kappa_0 - C_6 s\sqrt{\frac{\delta M\log d}{n}}\Big)\|\widehat{\boldsymbol{\Upsilon}}\|_2^2 \\ &\overset{(iii)}{\ge} \frac{\kappa_0}{2}\|\widehat{\boldsymbol{\Upsilon}}\|_2^2,\end{aligned} \tag{30}$$

where (i) is because $\widehat{\boldsymbol{\Upsilon}}^\top(\boldsymbol{\Sigma}^\star - \boldsymbol{Q})\widehat{\boldsymbol{\Upsilon}} \le \|\widehat{\boldsymbol{\Upsilon}}\|_1^2\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty$, (ii) is due to $\|\widehat{\boldsymbol{\Upsilon}}\|_1 \le 10\sqrt{s}\|\widehat{\boldsymbol{\Upsilon}}\|_2$, (iii) is due to the the assumed scaling $n \gtrsim \delta s^2\log d$. Now the desired results follow immediately from Lemma 4. □

3) **Quantized Matrix Completion:** Under the observation model (11), we first provide a deterministic framework for analysing the estimator (12).

Lemma 6. (Adapted from [22, Coro. 3])**.** *Let* $\widehat{\boldsymbol{\Upsilon}} := \widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^\star$*. If*

$$\lambda \ge 2\left\|\frac{1}{n}\sum_{k=1}^n(\langle\boldsymbol{X}_k,\boldsymbol{\Theta}^\star\rangle - \dot{y}_k)\boldsymbol{X}_k\right\|_{op}, \tag{31}$$

then it holds that $\|\widehat{\boldsymbol{\Upsilon}}\|_{nu} \le 10\sqrt{r}\|\widehat{\boldsymbol{\Upsilon}}\|_F$*. Moreover, if for some* $\kappa > 0$ *we have the restricted strong convexity (RSC)* $\frac{1}{n}\sum_{k=1}^n\big|\langle\boldsymbol{X}_k,\widehat{\boldsymbol{\Upsilon}}\rangle\big|^2 \ge \kappa\|\widehat{\boldsymbol{\Upsilon}}\|_F^2$*, then we have the error bounds* $\|\widehat{\boldsymbol{\Upsilon}}\|_F \le 30\sqrt{r}\big(\frac{\lambda}{\kappa}\big)$ *and* $\|\widehat{\boldsymbol{\Upsilon}}\|_{nu} \le 300r\big(\frac{\lambda}{\kappa}\big)$*.*

Clearly, to derive statistical error rate of $\widehat{\boldsymbol{\Theta}}$ from Lemma 6, the key ingredients are (31) and the RSC. Specialized to the

covariate $\boldsymbol{X}_k \sim \mathscr{U}\big(\{\boldsymbol{e}_i\boldsymbol{e}_j^\top : i,j\in[d]\}\big)$ in matrix completion, we will use the following lemma to establish RSC.

Lemma 7. (Adapted from [22, Lem. 4] with $q=0$)**.** *Given some $\alpha>0, \delta>0$, we define the constraint set ψ with sufficiently large ψ as*

$$\begin{aligned}\mathcal{C}(\psi)=\Big\{\boldsymbol{\Theta}\in\mathbb{R}^{d\times d}:&\|\boldsymbol{\Theta}\|_\infty\le 2\alpha,\\ &\|\boldsymbol{\Theta}\|_{nu}\le 10\sqrt{r}\|\boldsymbol{\Theta}\|_F,\\ &\|\boldsymbol{\Theta}\|_F^2\ge(\alpha d)^2\sqrt{\frac{\psi\delta\log d}{n}}\Big\}.\end{aligned}\tag{32}$$

Let $\boldsymbol{X}_1,...,\boldsymbol{X}_n$ be i.i.d. uniformly distributed on $\{\boldsymbol{e}_i\boldsymbol{e}_j^\top : i,j\in[d]\}$, then there exist absolute constants $\kappa\in(0,1)$ and C, such that with probability at least $1-d^{-\delta}$ we have

$$\frac{1}{n}\sum_{k=1}^n\big|\langle\boldsymbol{X}_k,\boldsymbol{\Theta}\rangle\big|^2\ge\frac{\kappa\|\boldsymbol{\Theta}\|_F^2}{d^2}-\frac{C\alpha^2 rd\log d}{n},\ \forall\ \boldsymbol{\Theta}\in\mathcal{C}(\psi).\tag{33}$$

Matrix completion with sub-exponential noise was studied in [51], and we make use of the following Lemma in the sub-exponential case.

Lemma 8. (Adapted from [51, Lem. 5])**.** *Given some $\delta>0$. Let $\boldsymbol{X}_1,...,\boldsymbol{X}_n$ be i.i.d. uniformly distributed on $\{\boldsymbol{e}_i\boldsymbol{e}_j^\top : i,j\in[d]\}$, independent of $\boldsymbol{X}_k$s, $\epsilon_1,...,\epsilon_n$ are i.i.d. zero-mean and satisfy $\|\epsilon_k\|_{\psi_1}\le\sigma$. If $n\gtrsim\delta d\log^3 d$, with probability at least $1-d^{-\delta}$ we have*

$$\Big\|\frac{1}{n}\sum_{k=1}^n\epsilon_k\boldsymbol{X}_k\Big\|_{op}\le\sigma\sqrt{\frac{\delta\log d}{nd}}.$$

Proof of Theorem 7: *Proof.* We divide the proof into two steps.

Step 1. Proving (31)

Defining $w_k:=\dot{y}_k-y_k-\tau_k$ as the quantization error, from Theorem 1(a) we know that w_ks are independent of $\boldsymbol{X}_k$ and i.i.d. uniformly distributed on $[-\frac{\Delta}{2},\frac{\Delta}{2}]$. Thus, we can further write that

$$\dot{y}_k=y_k+\tau_k+w_k=\big\langle\boldsymbol{X}_k,\boldsymbol{\Theta}^\star\big\rangle+\epsilon_k+\tau_k+w_k,$$

which allows us to decompose I into

$$\begin{aligned}I\le&\Big\|\frac{1}{n}\sum_{k=1}^n\epsilon_k\boldsymbol{X}_k\Big\|_{op}+\Big\|\frac{1}{n}\sum_{k=1}^n\tau_k\boldsymbol{X}_k\Big\|_{op}+\Big\|\frac{1}{n}\sum_{k=1}^n w_k\boldsymbol{X}_k\Big\|_{op}\\ :=&I_1+I_2+I_3.\end{aligned}$$

Because ϵ_ks are independent of $\boldsymbol{X}_k$s and i.i.d. sub-exponential noise satisfying $\|\epsilon_k\|_{\psi_1}\le\sigma$, under the scaling $n\gtrsim\delta d\log^3 d$, Lemma 8 implies that $I_1\lesssim\sigma\sqrt{\frac{\delta\log d}{nd}}$ holds with probability at least $1-d^{-\delta}$. Analogously, τ_ks and w_ks are independent of $\{\boldsymbol{X}_k : k\in[n]\}$ and are i.i.d. uniformly distributed on $[-\frac{\Delta}{2},\frac{\Delta}{2}]$, Lemma 8 also applies to I_2 and I_3, yielding that with the promised probability $I_2+I_3\lesssim\Delta\sqrt{\frac{\delta\log d}{nd}}$. Taken collectively, $I\lesssim(\sigma+\Delta)\sqrt{\frac{\delta\log d}{nd}}$, so setting $\lambda=C_1(\sigma+\Delta)\sqrt{\frac{\delta\log d}{nd}}$ with sufficiently large C_1 ensures $\lambda\ge 2I$, with probability at least $1-3d^{-\delta}$. Further, Lemma 6 gives $\|\widehat{\boldsymbol{\Upsilon}}\|_{nu}\le 10\sqrt{r}\|\widehat{\boldsymbol{\Upsilon}}\|_F$.

Step 2. Verifying RSC

First note that $\|\widehat{\boldsymbol{\Upsilon}}\|_\infty\le\|\widehat{\boldsymbol{\Theta}}\|_\infty+\|\boldsymbol{\Theta}^\star\|_\infty\le 2\alpha$; and as proved before, $\|\widehat{\boldsymbol{\Upsilon}}\|_{nu}\le 10\sqrt{r}\|\widehat{\boldsymbol{\Upsilon}}\|_F$. To proceed we define the constraint set $\mathcal{C}(\psi)$ as in (32) with some properly chosen constant ψ. Then using Lemma 7, for some absolute constants κ, C, (33) holds with probability at least $1-d^{-\delta}$. Then we discuss several cases.

1) If $\widehat{\boldsymbol{\Upsilon}}\notin\mathcal{C}(\psi)$, because $\widehat{\boldsymbol{\Upsilon}}$ satisfies the first two constraints in the definition of $\mathcal{C}(\psi)$, it must violate the third constraint and satisfy $\|\widehat{\boldsymbol{\Upsilon}}\|_F^2\le(\alpha d)^2\sqrt{\frac{\psi\delta\log d}{n}}$, which gives

$$\|\widehat{\boldsymbol{\Upsilon}}\|_F\lesssim\alpha d\big(\frac{\delta\log d}{n}\big)^{1/4}\overset{(i)}{\lesssim}\alpha d\sqrt{\frac{\delta rd\log d}{n}},$$

as desired. Note that (i) is due to the scaling $n\lesssim\delta r^2d^2\log d$.

2) If $\widehat{\boldsymbol{\Upsilon}}\in\mathcal{C}(\psi)$, (33) implies that $\frac{1}{n}\sum_{k=1}^n\big|\langle\boldsymbol{X}_k,\widehat{\boldsymbol{\Upsilon}}\rangle\big|^2\ge\frac{\kappa\|\widehat{\boldsymbol{\Upsilon}}\|_F^2}{d^2}-C\frac{\alpha^2rd\log d}{n}$, and we further consider the following two cases.

2.1) If $C\frac{\alpha^2rd\log d}{n}\ge\frac{\kappa\|\widehat{\boldsymbol{\Upsilon}}\|_F^2}{2d^2}$, we have $\|\widehat{\boldsymbol{\Upsilon}}\|_F\lesssim\alpha d\sqrt{\frac{rd\log d}{n}}$, as desired.

2.2) If $C\frac{\alpha^2rd\log d}{n}<\frac{\kappa\|\widehat{\boldsymbol{\Upsilon}}\|_F^2}{2d^2}$, then the RSC condition holds: $\frac{1}{n}\sum_{k=1}^n\big|\langle\boldsymbol{X}_k,\widehat{\boldsymbol{\Upsilon}}\rangle\big|^2\ge\frac{\kappa\|\widehat{\boldsymbol{\Upsilon}}\|_F^2}{2d^2}$. This allows us to apply Lemma 6 to obtain $\|\widehat{\boldsymbol{\Upsilon}}\|_F\lesssim(\sigma+\Delta)d\sqrt{\frac{\delta rd\log d}{n}}$.

Thus, in any case, we have shown $\|\widehat{\boldsymbol{\Upsilon}}\|_F=O\big((\alpha+\sigma+\Delta)d\sqrt{\frac{\delta rd\log d}{n}}\big)$. The bound on $\|\widehat{\boldsymbol{\Upsilon}}\|_{nu}$ follows immediately from $\|\widehat{\boldsymbol{\Upsilon}}\|_{nu}\le 10\sqrt{r}\|\widehat{\boldsymbol{\Upsilon}}\|_F$. The proof is complete. □

Proof of Theorem 8: *Proof.* The proof is based on Lemma 6 and divided into two steps.

Step 1. Proving (31)

Recall that the quantization error $w_k:=\dot{y}_k-\widetilde{y}_k-\tau_k$ is zero-mean and independent of $\boldsymbol{X}_k$ (Theorem 1(a)), thus we have $\mathbb{E}(\dot{y}_k\boldsymbol{X}_k)=\mathbb{E}(\widetilde{y}_k\boldsymbol{X}_k)+\mathbb{E}(\tau_k\boldsymbol{X}_k)+\mathbb{E}(w_k\boldsymbol{X}_k)=\mathbb{E}(\widetilde{y}_k\boldsymbol{X}_k)$. Combining with $\mathbb{E}\big(\langle\boldsymbol{X}_k,\boldsymbol{\Theta}^\star\rangle\boldsymbol{X}_k\big)=\mathbb{E}(y_k\boldsymbol{X}_k)$, triangle inequality can first decompose the target term into

$$\begin{aligned}&\Big\|\frac{1}{n}\sum_{k=1}^n\big(\dot{y}_k-\langle\boldsymbol{X}_k,\boldsymbol{\Theta}^\star\rangle\big)\boldsymbol{X}_k\Big\|_{op}\\ \le&\Big\|\frac{1}{n}\sum_{k=1}^n\dot{y}_k\boldsymbol{X}_k-\mathbb{E}(\dot{y}_k\boldsymbol{X}_k)\Big\|_{op}+\Big\|\mathbb{E}\big(y_k\boldsymbol{X}_k-\widetilde{y}_k\boldsymbol{X}_k\big)\Big\|_{op}\\ &+\Big\|\frac{1}{n}\sum_{k=1}^n\langle\boldsymbol{X}_k,\boldsymbol{\Theta}^\star\rangle\boldsymbol{X}_k-\mathbb{E}\big(\langle\boldsymbol{X}_k,\boldsymbol{\Theta}^\star\rangle\boldsymbol{X}_k\big)\Big\|_{op}\\ :=&I_1+I_2+I_3.\end{aligned}$$

Step 1.1. Bounding I_1 and I_3

We write $I_1=\|\sum_{k=1}^n\boldsymbol{S}_k\|_{op}$ and $I_3=\|\sum_{k=1}^n\boldsymbol{W}_k\|_{op}$ by defining

$$\begin{aligned}\boldsymbol{S}_k&=\frac{1}{n}\big(\dot{y}_k\boldsymbol{X}_k-\mathbb{E}(\dot{y}_k\boldsymbol{X}_k)\big),\\ \boldsymbol{W}_k&=\frac{1}{n}\big(\langle\boldsymbol{X}_k,\boldsymbol{\Theta}^\star\rangle\boldsymbol{X}_k-\mathbb{E}\big[\langle\boldsymbol{X}_k,\boldsymbol{\Theta}^\star\rangle\boldsymbol{X}_k\big]\big).\end{aligned}$$

By $|\dot{y}_k| \le |\widetilde{y}_k| + |\tau_k| + |w_k| \le \zeta_y + \Delta$ we have

$$\begin{aligned}\|\boldsymbol{S}_k\|_{op} &\le \frac{1}{n}\|\dot{y}_k\boldsymbol{X}_k\|_{op} + \frac{1}{n}\|\mathbb{E}(\dot{y}_k\boldsymbol{X}_k)\|_{op}\\ &\le \frac{1}{n}\|\dot{y}_k\boldsymbol{X}_k\|_{op} + \frac{1}{n}\mathbb{E}\|\dot{y}_k\boldsymbol{X}_k\|_{op} \le \frac{2(\zeta_y+\Delta)}{n}.\end{aligned}$$

Analogously, we have $\|\boldsymbol{W}_k\|_{op} \le \frac{2\alpha}{n}$ since $\big|\langle \boldsymbol{X}_k, \boldsymbol{\Theta}^\star\rangle\big| \le \|\boldsymbol{\Theta}^\star\|_\infty \le \alpha$. In addition, by

$$\|\mathbb{E}\{(\boldsymbol{A}-\mathbb{E}\boldsymbol{A})^\top(\boldsymbol{A}-\mathbb{E}\boldsymbol{A})\}\|_{op} \le \|\mathbb{E}(\boldsymbol{A}^\top\boldsymbol{A})\|_{op}\ (\forall \boldsymbol{A})$$

and the simple fact $\mathbb{E}(\boldsymbol{X}_k\boldsymbol{X}_k^\top) = \mathbb{E}(\boldsymbol{X}_k^\top\boldsymbol{X}_k) = \boldsymbol{I}_d/d$, we estimate the matrix variance statistic as follows

$$\begin{aligned}&\Big\|\sum_{k=1}^n \mathbb{E}(\boldsymbol{S}_k\boldsymbol{S}_k^\top)\Big\|_{op} = n\big\|\mathbb{E}(\boldsymbol{S}_k\boldsymbol{S}_k^\top)\big\|_{op}\\ \le&\frac{1}{n}\big\|\mathbb{E}\big(\dot{y}_k^2\boldsymbol{X}_k\boldsymbol{X}_k^\top\big)\big\|_{op} = \frac{1}{n}\sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}}\mathbb{E}\big(\dot{y}_k^2\cdot\|\boldsymbol{X}_k^\top\boldsymbol{v}\|_2^2\big)\\ =&\frac{1}{n}\sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}}\mathbb{E}_{\boldsymbol{X}_k}\Big(\big[\mathbb{E}_{\dot{y}_k|\boldsymbol{X}_k}(\dot{y}_k^2)\big]\|\boldsymbol{X}_k^\top\boldsymbol{v}\|_2^2\Big)\\ \overset{(i)}{\le}&\frac{4}{n}(\alpha^2+M+\Delta^2)\sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}}\mathbb{E}_{\boldsymbol{X}_k}\|\boldsymbol{X}_k^\top\boldsymbol{v}\|_2^2\\ \le&\frac{4(\alpha^2+M+\Delta^2)}{nd},\end{aligned}$$

where (i) is because given $\boldsymbol{X}_k$ we can estimate $\mathbb{E}_{\dot{y}_k|\boldsymbol{X}_k}(\dot{y}_k^2) \le 2\big(\mathbb{E}_{\dot{y}_k|\boldsymbol{X}_k}(\widetilde{y}_k^2)+\Delta^2\big)$ since $|\dot{y}_k - \widetilde{y}_k| \le \Delta$, and moreover we have

$$\begin{aligned}\mathbb{E}_{\dot{y}_k|\boldsymbol{X}_k}(\widetilde{y}_k^2) \le&\mathbb{E}_{\dot{y}_k|\boldsymbol{X}_k}(y_k^2)\\ \le&2\big(\mathbb{E}_{\dot{y}_k|\boldsymbol{X}_k}(\langle\boldsymbol{X}_k,\boldsymbol{\Theta}^\star\rangle^2) + \mathbb{E}_{\dot{y}_k|\boldsymbol{X}_k}(\epsilon_k^2)\big)\\ \le&2(\alpha^2+M).\end{aligned}$$

It is not hard to see that this bound remains valid for $\|\sum_{k=1}^n \mathbb{E}(\boldsymbol{S}_k^\top\boldsymbol{S}_k)\|_{op}$. Also, by similar arguments one can prove

$$\max\left\{\Big\|\sum_{k=1}^n\mathbb{E}(\boldsymbol{W}_k^\top\boldsymbol{W}_k)\Big\|_{op}, \Big\|\sum_{k=1}^n\mathbb{E}(\boldsymbol{W}_k\boldsymbol{W}_k^\top)\Big\|_{op}\right\} \le \frac{\alpha^2}{nd}.$$

Thus, Matrix Bernstein's inequality (Lemma 2) gives

$$\mathbb{P}\big(I_1 \ge t\big) \le 2d\cdot\exp\Big(-\frac{C_4ndt^2}{(\alpha^2+M+\Delta^2)+(\zeta_y+\Delta)dt}\Big)$$
$$\mathbb{P}\big(I_3 \ge t\big) \le 2d\cdot\exp\Big(-\frac{C_5ndt^2}{\alpha^2+\alpha dt}\Big)$$

Thus, setting $t = C_6(\alpha+\sqrt{M}+\Delta)\sqrt{\frac{\delta\log d}{nd}}$ in the two inequalities above with sufficiently large C_6, combining with the scaling that $\sqrt{\frac{\delta d\log d}{n}} = O(1)$, we obtain

$$I_1 + I_3 \lesssim (\alpha+\sqrt{M}+\Delta)\sqrt{\frac{\delta\log d}{nd}}$$

with probability at least $1-4d^{1-\delta}$.

Step 1.2. Bounding I_2

Let us bound $\|\mathbb{E}\big((y_k-\widetilde{y}_k)\boldsymbol{X}_k\big)\|_\infty$ first. Write (i,j)-th entry of $\boldsymbol{X}_k$ as $x_{k,ij}$, then for given (i,j), $\mathbb{P}(x_{k,ij}=1) = d^{-2}$, $x_{k,ij}=0$ otherwise. We can thus proceed by the following estimations:

$$\begin{aligned}&\big|\mathbb{E}\big((y_k-\widetilde{y}_k)x_{k,ij}\big)\big|\\ =&\big|\mathbb{E}\big((y_k-\widetilde{y}_k)x_{k,ij}\mathbb{1}(|y_k|\ge\zeta_y)\big)\big|\\ \le&\mathbb{E}\big(|y_k|x_{k,ij}\mathbb{1}(|y_k|\ge\zeta_y)\big)\\ =&\mathbb{E}_{x_{k,ij}}\big(\big\{\mathbb{E}_{y_k|x_{k,ij}}|y_k|\mathbb{1}(|y_k|\ge\zeta_y)\big\}x_{k,ij}\big)\\ =&d^{-2}\mathbb{E}_{y_k|x_{k,ij}=1}\big(|y_k|\mathbb{1}(|y_k|\ge\zeta_y)\big)\\ \overset{(i)}{\le}&d^{-2}\sqrt{\mathbb{E}_{y_k\sim\theta^\star_{ij}+\epsilon_k}(y_k^2)}\sqrt{\mathbb{P}_{y_k\sim\theta^\star_{ij}+\epsilon_k}(y_k^2\ge\zeta_y^2)}\\ \overset{(ii)}{\le}&d^{-2}\frac{\alpha^2+M}{\zeta_y} \lesssim \frac{\alpha+\sqrt{M}}{d^2}\sqrt{\frac{\delta d\log d}{n}},\end{aligned}$$

where (i), (ii) is by Cauchy-Schwarz and Markov's, respectively. Since this holds for any (i,j), we obtain

$$\|\mathbb{E}\big((y_k-\widetilde{y}_k)\boldsymbol{X}_k\big)\|_\infty = O\big((\alpha+\sqrt{M})d^{-2}\sqrt{\frac{\delta d\log d}{n}}\big).$$

which further gives

$$I_2 = O\big((\alpha+\sqrt{M})\sqrt{\frac{\delta\log d}{nd}}\big)$$

by using $\|\boldsymbol{A}\|_{op} \le d\|\boldsymbol{A}\|_\infty$ ($\forall \boldsymbol{A}\in\mathbb{R}^{d\times d}$). Putting pieces together, with probability at least $1-4d^{1-\delta}$ we have

$$\|\frac{1}{n}\sum_k\big(\dot{y}_k - \langle\boldsymbol{X}_k,\boldsymbol{\Theta}^\star\rangle\big)\boldsymbol{X}_k\|_{op} \lesssim (\alpha+\sqrt{M}+\Delta)\sqrt{\frac{\delta\log d}{nd}},$$

hence $\lambda = C_1(\alpha+\sqrt{M}+\Delta)\sqrt{\frac{\delta\log d}{nd}}$ ensures (31) under the same probability. Further, Lemma 6 gives $\|\widehat{\boldsymbol{\Upsilon}}\|_{nu} \le 10\sqrt{r}\|\widehat{\boldsymbol{\Upsilon}}\|_F$.

Step 2. Verifying RSC

The remaining part is almost the same as Step 2 in the proof of Theorem 7 — defining the constraint set $\mathcal{C}(\psi)$ as (32) and then discussing several cases based on whether $\widehat{\boldsymbol{\Upsilon}} \in \mathcal{C}(\psi)$ holds. Thus, we conclude the proof without providing the details. □

B. Proofs in Section IV

This appendix collects the proofs in Section IV concerning covariate quantization and uniform signal recovery in QCS.

1) **Covariate Quantization:** Because of the non-convexity, the proofs in this part can no longer be based on Lemma 4. Indeed, bounding the estimation errors of $\widetilde{\boldsymbol{\theta}}$s satisfying (14) require more tedious manipulations essentially due to the additional ℓ_1 constraint (induced by the constraint $\mathcal{S}$ in (14)).

Proof of Theorem 9: *Proof.* The proof is divided into three steps — the first two steps resemble the previous proofs that are based on Lemma 4, while we bound the estimation errors in the last step.

Step 1. Proving $\lambda \ge \beta\|\boldsymbol{Q}\boldsymbol{\theta}^\star - \boldsymbol{b}\|_\infty$ for some pre-specified $\beta > 2$

Recall that $(\boldsymbol{Q},\boldsymbol{b})$ are constructed from the quantized data as $\boldsymbol{Q} = \frac{1}{n}\sum_{k=1}^n \dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top - \frac{\bar{\Delta}^2}{4}\boldsymbol{I}_d$ and $\boldsymbol{b} = \frac{1}{n}\sum_{k=1}^n \dot{y}_k\dot{\boldsymbol{x}}_k$. We will show that,

$$\lambda = C_1\frac{(\sigma+\bar{\Delta})^2}{\sqrt{\kappa_0}}(\Delta + M^{1/(2l)})\sqrt{\frac{\log d}{n}}$$

guarantees

$$\lambda \geq \beta\Big\|\frac{1}{n}\sum_{k=1}^{n}\Big(\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top - \frac{\bar{\Delta}^2}{4}\boldsymbol{I}_d\Big)\boldsymbol{\theta}^\star - \frac{1}{n}\sum_{k=1}^{n}\dot{y}_k\dot{\boldsymbol{x}}_k\Big\|_\infty$$

holds with the promised probability, where $\beta > 2$ is any pre-specified constant. Recall the notation we introduced: $\dot{y}_k = \widetilde{y}_k + \phi_k + \vartheta_k$ with the quantization error $\vartheta_k \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}])$ being independent of $\widetilde{y}_k$, $\dot{\boldsymbol{x}}_k = \boldsymbol{x}_k + \boldsymbol{\tau}_k + \boldsymbol{w}_k$ with the quantization error $\boldsymbol{w}_k \sim \mathscr{U}([-\frac{\Delta}{2}, \frac{\Delta}{2}]^d)$ being independent of $\boldsymbol{x}_k$. Combining with the assumptions that the dithers are independent of $(\boldsymbol{x}_k, y_k)$ and that ϕ_ks and $\boldsymbol{\tau}_k$s are independent, we have

$$\begin{aligned}\mathbb{E}(\dot{y}_k\dot{\boldsymbol{x}}_k) &= \mathbb{E}\big((\widetilde{y}_k + \phi_k + \vartheta_k)(\boldsymbol{x}_k + \boldsymbol{\tau}_k + \boldsymbol{w}_k)\big) = \mathbb{E}(\widetilde{y}_k\boldsymbol{x}_k),\\ \mathbb{E}\Big(\Big[\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top - \frac{\bar{\Delta}^2}{4}\boldsymbol{I}_d\Big]\boldsymbol{\theta}^\star\Big) &= \mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top\boldsymbol{\theta}^\star) = \mathbb{E}(y_k\boldsymbol{x}_k),\end{aligned} \tag{34}$$

which allows us to decompose the target term as two concentration terms (I_1, I_3) and a bias term (I_2)

$$\begin{aligned}&\Big\|\frac{1}{n}\sum_{k=1}^{n}\Big[\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top - \frac{\bar{\Delta}^2}{4}\boldsymbol{I}_d\Big]\boldsymbol{\theta}^\star - \frac{1}{n}\sum_{k=1}^{n}\dot{y}_k\dot{\boldsymbol{x}}_k\Big\|_\infty\\ &\leq \Big\|\frac{1}{n}\sum_{k=1}^{n}\dot{y}_k\dot{\boldsymbol{x}}_k - \mathbb{E}(\dot{y}_k\dot{\boldsymbol{x}}_k)\Big\|_\infty + \Big\|\mathbb{E}\big(y_k\boldsymbol{x}_k - \widetilde{y}_k\boldsymbol{x}_k\big)\Big\|_\infty\\ &+ \Big\|\frac{1}{n}\sum_{k=1}^{n}\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top\boldsymbol{\theta}^\star - \mathbb{E}(\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top\boldsymbol{\theta}^\star)\Big\|_\infty := I_1 + I_2 + I_3.\end{aligned}$$

Step 1.1. Bounding I_1

Denote the i-th entry of $\boldsymbol{x}_k, \dot{\boldsymbol{x}}_k, \boldsymbol{\tau}_k, \boldsymbol{w}_k$ by $x_{ki}, \dot{x}_{ki}, \tau_{ki}, w_{ki}$, respectively. For I_1, the i-th entry reads $\frac{1}{n}\sum_{k=1}^{n}\dot{y}_k\dot{x}_{ki} - \mathbb{E}(\dot{y}_k\dot{x}_{ki})$. By using the relations

$$|\dot{y}_k| \leq |\widetilde{y}_k| + |\phi_k| + |\vartheta_k| \leq \zeta_y + \Delta,$$

and

$$\|\dot{\boldsymbol{x}}_k\|_{\psi_2} \leq \|\boldsymbol{x}_k\|_{\psi_2} + \|\boldsymbol{\tau}_k\|_{\psi_2} + \|\boldsymbol{w}_k\|_{\psi_2} \lesssim \sigma + \bar{\Delta}$$

and

$$\mathbb{E}|\dot{y}_k|^{2l} \lesssim \mathbb{E}|\widetilde{y}_k|^{2l} + \mathbb{E}|\phi_k + \vartheta_k|^{2l} \lesssim M + \Delta^{2l},$$

for any integer $q \geq 2$ we can bound that

$$\begin{aligned}&\sum_{k=1}^{n}\mathbb{E}\Big|\frac{\dot{y}_k\dot{x}_{ki}}{n}\Big|^q\\ &\leq \frac{(\zeta_y+\Delta)^{q-2}}{n^q}\sum_{k=1}^{n}\mathbb{E}|\dot{y}_k^2\dot{x}_{ki}^q|\\ &\overset{(i)}{\leq} \frac{(\zeta_y+\Delta)^{q-2}}{n^q}\sum_{k=1}^{n}\big\{\mathbb{E}|\dot{y}_k|^{2l}\big\}^{\frac{1}{l}}\big\{\mathbb{E}|\dot{x}_{ki}|^{\frac{lq}{l-1}}\big\}^{1-\frac{1}{l}}\\ &\overset{(ii)}{\lesssim} \Big(\frac{(\sigma+\bar{\Delta})(\zeta_y+\Delta)}{n}\Big)^{q-2}\\ &\qquad\cdot\Big(\frac{(\sigma+\bar{\Delta})^2(M^{\frac{1}{l}}+\Delta^2)}{n}\Big)\Big(\sqrt{\frac{lq}{l-1}}\Big)^q;\end{aligned} \tag{35}$$

combining with Stirling's approximation and treating l as absolute constant, this provides

$$\begin{aligned}\sum_{k=1}^{n}\mathbb{E}\Big|\frac{\dot{y}_k\dot{x}_{ki}}{n}\Big|^q &\leq \frac{q!}{2}v_0c_0^{q-2}\\ \text{where } v_0 = O\Big(\frac{(\sigma+\bar{\Delta})^2(M^{1/l}+\Delta^2)}{n}\Big),&\\ c_0 = O\Big(\frac{(\sigma+\bar{\Delta})(\zeta_y+\Delta)}{n}\Big).&\end{aligned}$$

In (35), (i) is due to Holder's, and in (ii) we use the moment constraint of sub-Gaussian variable (2). With these preparations, we invoke Bernstein's inequality (Lemma 1) and then a union bound over $i \in [d]$ to obtain

$$\begin{aligned}\mathbb{P}\Big(I_1 \lesssim (\sigma+\bar{\Delta})(M^{\frac{1}{2l}}+\Delta)\sqrt{\frac{t}{n}} + \frac{(\sigma+\bar{\Delta})(\zeta_y+\Delta)t}{n}\Big)&\\ \geq 1 - 2d\cdot\exp(-t),&\end{aligned}$$

Thus, taking $t = \delta\log d$ and plug in $\zeta_y \asymp \sqrt{\frac{nM^{1/l}}{\delta\log d}}$, we obtain

$$\mathbb{P}\Big(I_1 \lesssim (\sigma+\bar{\Delta})(M^{1/(2l)}+\Delta)\sqrt{\frac{\delta\log d}{n}}\Big) \geq 1 - 2d^{1-\delta}.$$

Step 1.2. Bounding I_2

Moreover, we estimate the i-th entry of I_2 by

$$\begin{aligned}&|\mathbb{E}\big((y_k - \widetilde{y}_k)x_{ki}\big)| \leq \mathbb{E}|y_kx_{ki}\mathbb{1}(|y_k| \geq \zeta_y)|\\ &\overset{(i)}{\leq} \big(\mathbb{E}|y_k|^{\frac{2l}{2l-1}}|x_{ki}|^{\frac{2l}{2l-1}}\big)^{1-\frac{1}{2l}}\big(\mathbb{P}(|y_k| \geq \zeta_y)\big)^{\frac{1}{2l}}\\ &\overset{(ii)}{\leq} \Big(\big[\mathbb{E}|y_k|^{2l}\big]^{\frac{1}{2l-1}}\big[\mathbb{E}|x_{ki}|^{\frac{l}{l-1}}\big]^{\frac{2l-2}{2l-1}}\Big)^{1-\frac{1}{2l}}\Big(\mathbb{P}(|y_k|^{2l} \geq \zeta_y^{2l})\Big)^{\frac{1}{2l}}\\ &\overset{(iii)}{\leq} \frac{\sigma M^{1/l}}{\zeta_y} \lesssim \sigma M^{\frac{1}{2l}}\sqrt{\frac{\delta\log d}{n}},\end{aligned} \tag{36}$$

where (i), (ii) are due to Holder's, (iii) is due to Markov's. Since this holds for $i \in [d]$, it gives $I_2 \lesssim \sigma M^{1/(2l)}\sqrt{\frac{\delta\log d}{n}}$.

Step 1.3. Bounding I_3

We first derive a bound for $\|\boldsymbol{\theta}^\star\|_2$ that is implicitly implied by other conditions:

$$\begin{aligned}&M^{1/l} \geq \mathbb{E}|y_k|^2 \geq \mathbb{E}(\boldsymbol{x}_k^\top\boldsymbol{\theta}^\star)^2 = (\boldsymbol{\theta}^\star)^\top\boldsymbol{\Sigma}^\star\boldsymbol{\theta}^\star \geq \kappa_0\|\boldsymbol{\theta}^\star\|_2^2\\ &\Longrightarrow \|\boldsymbol{\theta}^\star\|_2 = O\Big(\frac{M^{1/(2l)}}{\sqrt{\kappa_0}}\Big).\end{aligned}$$

Hence, we can estimate

$$\begin{aligned}\|(\dot{\boldsymbol{x}}_k^\top\boldsymbol{\theta}^\star)\dot{x}_{ki}\|_{\psi_1} &\leq \|\dot{\boldsymbol{x}}_k^\top\boldsymbol{\theta}^\star\|_{\psi_2}\|\dot{x}_{ki}\|_{\psi_2}\\ &\leq \|\dot{\boldsymbol{x}}_k\|_{\psi_2}^2\|\boldsymbol{\theta}^\star\|_2 \lesssim (\sigma+\bar{\Delta})^2\frac{M^{1/(2l)}}{\sqrt{\kappa_0}}.\end{aligned}$$

Then, we invoke Bernstein's inequality regarding the independent sum of sub-exponential random variables (e.g., [92, Thm.

2.8.1])[16] to obtain that for any $t \geq 0$ we have

$$\mathbb{P}\left(\Big|\frac{1}{n}\sum_{k=1}^{n}(\dot{\boldsymbol{x}}_k^\top\boldsymbol{\theta}^\star)\dot{x}_{ki} - \mathbb{E}\{(\dot{\boldsymbol{x}}_k^\top\boldsymbol{\theta}^\star)\dot{x}_{ki}\}\Big| \geq t\right)$$
$$\leq 2\exp\left(-C_5 n\min\left\{\frac{\sqrt{\kappa_0}t}{(\sigma+\bar{\Delta})^2 M^{\frac{1}{2l}}}, \Big(\frac{\sqrt{\kappa_0}t}{(\sigma+\bar{\Delta})^2 M^{\frac{1}{2l}}}\Big)^2\right\}\right)$$

Hence, we can set $t = C_6\frac{(\sigma+\bar{\Delta})^2}{\sqrt{\kappa_0}}M^{1/(2l)}\sqrt{\frac{\delta\log d}{n}}$ with sufficiently large C_6, and further apply union bound over $i \in [d]$, under the scaling that $\frac{\delta\log d}{n}$ is small enough, we obtain

$$I_3 \lesssim \frac{(\sigma+\bar{\Delta})^2}{\sqrt{\kappa_0}}M^{1/(2l)}\sqrt{\frac{\delta\log d}{n}}$$

holds with probability at least $1-2d^{1-\delta}$.

Putting pieces together, since $\kappa_0 \lesssim \sigma^2$, it is immediate that

$$I \lesssim \frac{(\sigma+\bar{\Delta})^2}{\sqrt{\kappa_0}}\big(\Delta + M^{1/(2l)}\big)\sqrt{\frac{\delta\log d}{n}}$$

holds with probability at least $1-8d^{1-\delta}$. Since

$$\lambda = C_1\frac{(\sigma+\bar{\Delta})^2}{\sqrt{\kappa_0}}(\Delta + M^{1/(2l)})\sqrt{\frac{\delta\log d}{n}}$$

with sufficiently large C_1, $\lambda \geq \beta\cdot\|\boldsymbol{Q}\boldsymbol{\theta}^\star - \boldsymbol{b}\|_\infty$ holds w.h.p.

Step 2. Verifying RSC

We provide a lower bound for $\boldsymbol{v}^\top\boldsymbol{Q}\boldsymbol{v} = \frac{1}{n}\|\dot{\boldsymbol{X}}\boldsymbol{v}\|_2^2 - \frac{\bar{\Delta}^2}{4}\|\boldsymbol{v}\|_2^2$ by using the matrix deviation inequality (Lemma 5). First note that the rows of $\dot{\boldsymbol{X}}$ are sub-Gaussian $\|\dot{\boldsymbol{x}}_k\|_{\psi_2} \lesssim \sigma + \bar{\Delta}$. Since $\mathbb{E}(\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top) = \boldsymbol{\Sigma}^\star + \frac{\bar{\Delta}^2}{4}\boldsymbol{I}_d$, all eigenvalues of $\dot{\boldsymbol{\Sigma}} := \mathbb{E}(\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top)$ are between $[\kappa_0 + \frac{1}{4}\bar{\Delta}^2, \kappa_1 + \frac{1}{4}\bar{\Delta}^2]$. Thus, we invoke Lemma 5 for $\mathcal{T} := \{\boldsymbol{v}\in\mathbb{R}^d : \|\boldsymbol{v}\|_1 = 1\}$ with $u = \sqrt{\delta\log d}$; due to the well-known Gaussian width estimate $\omega(\mathcal{T}) \lesssim \sqrt{\log d}$ [92, Example 7.5.9], with probability at least $1-d^{-\delta}$ the following event holds

$$\sup_{\|\boldsymbol{v}\|_1=1}\Big|\|\dot{\boldsymbol{X}}\boldsymbol{v}\|_2 - \sqrt{n}\|\dot{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{v}\|_2\Big|$$
$$\leq \frac{c_1\sqrt{\kappa_1 + \frac{1}{4}\bar{\Delta}^2}(\sigma+\bar{\Delta})^2}{\kappa_0 + \frac{1}{4}\bar{\Delta}^2}\sqrt{\delta\log d} := c_1\mathscr{L}_1\sqrt{\delta\log d}.$$

Under the same probability, a simple rescaling then provides

$$\Big|\frac{1}{\sqrt{n}}\|\dot{\boldsymbol{X}}\boldsymbol{v}\|_2 - \|\dot{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{v}\|_2\Big| \leq c_1\mathscr{L}_1\sqrt{\frac{\delta\log d}{n}}\|\boldsymbol{v}\|_1,\ \forall\ \boldsymbol{v}\in\mathbb{R}^d, \tag{37}$$

which implies

$$\frac{1}{\sqrt{n}}\|\dot{\boldsymbol{X}}\boldsymbol{v}\|_2 \geq \|\dot{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{v}\|_2 - c_1\mathscr{L}_1\Big(\frac{\delta\log d}{n}\Big)^{1/2}\|\boldsymbol{v}\|_1$$
$$\geq \Big(\kappa_0 + \frac{1}{4}\bar{\Delta}^2\Big)^{1/2}\|\boldsymbol{v}\|_2 - c_1\mathscr{L}_1\Big(\frac{\delta\log d}{n}\Big)^{1/2}\|\boldsymbol{v}\|_1,\ \forall\ \boldsymbol{v}\in\mathbb{R}^d. \tag{38}$$

[16]The application of Bernstein's inequality leads to the σ^2 (σ is the upper bound on $\|\boldsymbol{x}_k\|_{\psi_2}$) dependence in the multiplicative factor $\mathscr{L}$. It is possible to refine this quadratic dependence by using a new Bernstein's inequality developed in [49, Thm. 1.3], but we do not pursue this in the present paper.

Based on (38), we let $\hat{c} := \frac{2\kappa_0+\bar{\Delta}^2}{4\kappa_0+\bar{\Delta}^2}$ and use the inequality $(a-b)^2 \geq \hat{c}a^2 - \frac{\hat{c}}{1-\hat{c}}b^2$ to obtain

$$\boldsymbol{v}^\top\boldsymbol{Q}\boldsymbol{v} = \frac{1}{n}\|\dot{\boldsymbol{X}}\boldsymbol{v}\|_2^2 - \frac{\bar{\Delta}^2}{4}\|\boldsymbol{v}\|_2^2$$
$$\geq \hat{c}\big(\kappa_0 + \frac{1}{4}\bar{\Delta}^2\big)\|\boldsymbol{v}\|_2^2 - c_1^2\mathscr{L}_1^2\frac{\hat{c}}{1-\hat{c}}\frac{\delta\log d}{n}\|\boldsymbol{v}\|_1^2 - \frac{\bar{\Delta}^2}{4}\|\boldsymbol{v}\|_2^2$$
$$\geq \frac{\kappa_0}{2}\|\boldsymbol{v}\|_2^2 - c_1^2\mathscr{L}_1^2\big(1+\frac{\bar{\Delta}^2}{2\kappa_0}\big)\frac{\delta\log d}{n}\|\boldsymbol{v}\|_1^2$$
$$:= \frac{\kappa_0}{2}\|\boldsymbol{v}\|_2^2 - c_2(\kappa_0,\sigma,\bar{\Delta})\cdot\frac{\delta\log d}{n}\|\boldsymbol{v}\|_1^2,$$

which holds for all $\boldsymbol{v}\in\mathbb{R}^d$ and $\hat{c}_2 := c_2(\kappa_0,\sigma,\bar{\Delta})$ is a constant depending on $\kappa_0,\sigma,\bar{\Delta}$ (we remove the dependence on κ_1 by $\kappa_1 \lesssim \sigma^2$).

Step 3. Bounding the Estimation Error

We are in a position to bound the estimation error of any $\widetilde{\boldsymbol{\theta}}$ satisfying (14). Note that definition of $\partial\|\widetilde{\boldsymbol{\theta}}\|$ gives $\lambda\|\boldsymbol{\theta}^\star\|_1 - \lambda\|\widetilde{\boldsymbol{\theta}}\|_1 \geq \langle\lambda\cdot\partial\|\widetilde{\boldsymbol{\theta}}\|_1, -\widetilde{\boldsymbol{\Upsilon}}\rangle$. Thus, we set $\boldsymbol{\theta} = \boldsymbol{\theta}^\star$ in (14) and proceed as follows

$$\begin{aligned}
0 &\geq \langle\boldsymbol{Q}\widetilde{\boldsymbol{\theta}} - \boldsymbol{b} + \lambda\cdot\partial\|\widetilde{\boldsymbol{\theta}}\|_1, \widetilde{\boldsymbol{\Upsilon}}\rangle \\
&= \widetilde{\boldsymbol{\Upsilon}}^\top\boldsymbol{Q}\widetilde{\boldsymbol{\Upsilon}} + \langle\boldsymbol{Q}\boldsymbol{\theta}^\star - \boldsymbol{b}, \widetilde{\boldsymbol{\Upsilon}}\rangle + \langle\lambda\cdot\partial\|\widetilde{\boldsymbol{\theta}}\|_1, \widetilde{\boldsymbol{\Upsilon}}\rangle \\
&\overset{(i)}{\geq} \frac{\kappa_0}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_2^2 - \frac{2\hat{c}_2 R\sqrt{s}\cdot\delta\log d}{n}\|\widetilde{\boldsymbol{\Upsilon}}\|_1 \\
&\qquad - \frac{\lambda}{\beta}\|\widetilde{\boldsymbol{\Upsilon}}\|_1 + \lambda\big(\|\widetilde{\boldsymbol{\theta}}\|_1 - \|\boldsymbol{\theta}^\star\|_1\big) \\
&\overset{(ii)}{\geq} \frac{\kappa_0}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_2^2 - \frac{\lambda}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_1 + \lambda\big(\|\widetilde{\boldsymbol{\theta}}\|_1 - \|\boldsymbol{\theta}^\star\|_1\big),
\end{aligned} \tag{39}$$

where we use $\lambda \geq \beta\|\boldsymbol{Q}\boldsymbol{\theta}^\star - \boldsymbol{b}\|_\infty$ ($\beta > 2$) and $\|\widetilde{\boldsymbol{\Upsilon}}\|_1 \leq \|\widetilde{\boldsymbol{\theta}}\|_1 + \|\boldsymbol{\theta}^\star\|_1 \leq 2R\sqrt{s}$ in (i), and (ii) is due to the scaling

$$2\hat{c}_2 R\delta\sqrt{s}\frac{\log d}{n} \leq \Big(\frac{1}{2} - \frac{1}{\beta}\Big)\lambda$$

that holds under the assumed $n \gtrsim \delta\log d$ for some hidden constant depending on $(\kappa_0,\sigma,\bar{\Delta},\Delta,M,R)$. Thus, we arrive at $\frac{\lambda}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_1 \geq \lambda\big(\|\widetilde{\boldsymbol{\theta}}\|_1 - \|\boldsymbol{\theta}^\star\|_1\big)$.

For $\boldsymbol{a}\in\mathbb{R}^d$, $\mathcal{J}\subset[d]$ we obtain $\boldsymbol{a}_\mathcal{J}\in\mathbb{R}^d$ by keeping entries of $\boldsymbol{a}$ in $\mathcal{J}$ while setting others to zero. Let $\mathcal{A}$ be the support of $\boldsymbol{\theta}^\star$, $\mathcal{A}^c = [d]\setminus\mathcal{A}$, then we have

$$\begin{aligned}
\frac{1}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_1 &\geq \|\boldsymbol{\theta}^\star + \widetilde{\boldsymbol{\Upsilon}}\|_1 - \|\boldsymbol{\theta}^\star\|_1 \\
&= \|\boldsymbol{\theta}^\star + \widetilde{\boldsymbol{\Upsilon}}_\mathcal{A} + \widetilde{\boldsymbol{\Upsilon}}_{\mathcal{A}^c}\|_1 - \|\boldsymbol{\theta}^\star\|_1 \\
&\geq \|\boldsymbol{\theta}^\star\|_1 + \|\widetilde{\boldsymbol{\Upsilon}}_{\mathcal{A}^c}\|_1 - \|\widetilde{\boldsymbol{\Upsilon}}_\mathcal{A}\|_1 - \|\boldsymbol{\theta}^\star\|_1 \\
&= \|\widetilde{\boldsymbol{\Upsilon}}_{\mathcal{A}^c}\|_1 - \|\widetilde{\boldsymbol{\Upsilon}}_\mathcal{A}\|_1.
\end{aligned} \tag{40}$$

Further use $\frac{1}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_1 \leq \frac{1}{2}\|\widetilde{\boldsymbol{\Upsilon}}_\mathcal{A}\|_1 + \frac{1}{2}\|\widetilde{\boldsymbol{\Upsilon}}_{\mathcal{A}^c}\|_1$, we obtain $\|\widetilde{\boldsymbol{\Upsilon}}_{\mathcal{A}^c}\|_1 \leq 3\|\widetilde{\boldsymbol{\Upsilon}}_\mathcal{A}\|_1$. Hence, we have $\|\widetilde{\boldsymbol{\Upsilon}}\|_1 \leq \|\widetilde{\boldsymbol{\Upsilon}}_\mathcal{A}\|_1 + \|\widetilde{\boldsymbol{\Upsilon}}_{\mathcal{A}^c}\|_1 \leq 4\|\widetilde{\boldsymbol{\Upsilon}}_\mathcal{A}\|_1 \leq 4\sqrt{s}\|\widetilde{\boldsymbol{\Upsilon}}\|_2$. Now, we further substitute this into (39) and obtain

$$\frac{1}{2}\kappa_0\|\widetilde{\boldsymbol{\Upsilon}}\|_2^2 \leq \frac{\lambda}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_1 + \lambda\big(\|\boldsymbol{\theta}^\star\|_1 - \|\widetilde{\boldsymbol{\theta}}\|_1\big)$$
$$\leq \frac{3\lambda}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_1 \leq 6\lambda\sqrt{s}\|\widetilde{\boldsymbol{\Upsilon}}\|_2.$$

Thus, we arrive at the desired error bound for ℓ_2-norm

$$\|\widetilde{\boldsymbol{\Upsilon}}\|_2 \lesssim \mathscr{L}\sqrt{\frac{\delta s\log d}{n}},\ \text{with}\ \mathscr{L}:=\frac{(\sigma+\bar{\Delta})^2(\Delta+M^{1/(2l)})}{\kappa_0^{3/2}}.$$

We simply use $\|\widetilde{\boldsymbol{\Upsilon}}\|_1\le 4\sqrt{s}\|\widetilde{\boldsymbol{\Upsilon}}\|_2$ again to establish the bound for $\|\widetilde{\boldsymbol{\Upsilon}}\|_1$. The proof is complete. □

Proof of Theorem 10: *Proof.* The proof is divided into two steps. Compared to the last proof, due to the heavy-tailedness of $\boldsymbol{x}_k$, the step of "verifying RSC" reduces to the simpler argument in (42).

Step 1. Proving $\lambda \ge \beta\|\boldsymbol{Q}\boldsymbol{\theta}^\star-\boldsymbol{b}\|_\infty$ for some pre-specified $\beta>2$

Recall that $(\boldsymbol{Q},\boldsymbol{b})$ are constructed from the quantized data as $\boldsymbol{Q}=\frac{1}{n}\sum_{k=1}^n \dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top-\frac{\bar{\Delta}}{4}\boldsymbol{I}_d$ and $\boldsymbol{b}=\frac{1}{n}\sum_{k=1}^n \dot{y}_k\dot{\boldsymbol{x}}_k$. Thus, our main aim in this step is to prove that

$$\lambda = C_1(R\sqrt{M}+\Delta^2+R\bar{\Delta}^2)\sqrt{\frac{\delta\log d}{n}}$$

suffices to ensure

$$\lambda\ge\beta\Big\|\frac{1}{n}\sum_{k=1}^n\Big(\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top-\frac{\bar{\Delta}^2}{4}\boldsymbol{I}_d\Big)\boldsymbol{\theta}^\star-\frac{1}{n}\sum_{k=1}^n\dot{y}_k\dot{\boldsymbol{x}}_k\Big\|_\infty$$

with the promised probability and any pre-specified $\beta>2$. We let $\dot{\boldsymbol{x}}_k=\widetilde{\boldsymbol{x}}_k+\boldsymbol{\tau}_k+\boldsymbol{w}_k$, $\dot{y}_k=\widetilde{y}_k+\phi_k+\vartheta_k$ with quantization errors $\boldsymbol{w}_k$ and ϑ_k. Analogously to (34), we have $\mathbb{E}[\dot{y}_k\dot{\boldsymbol{x}}_k]=\mathbb{E}(\widetilde{y}_k\widetilde{\boldsymbol{x}}_k)$ and $\mathbb{E}(\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top\boldsymbol{\theta}^\star)=\frac{\bar{\Delta}^2}{4}\boldsymbol{\theta}^\star+\mathbb{E}(\widetilde{\boldsymbol{x}}_k\widetilde{\boldsymbol{x}}_k^\top\boldsymbol{\theta}^\star)$. Thus, the term we want to bound can be first decomposed into two concentration terms (I_1,I_3) and one bias term (I_2):

$$\begin{aligned}
&\Big\|\Big(\frac{1}{n}\sum_{k=1}^n\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top-\frac{\bar{\Delta}^2}{4}\boldsymbol{I}_d\Big)\boldsymbol{\theta}^\star-\frac{1}{n}\sum_{k=1}^n\dot{y}_k\dot{\boldsymbol{x}}_k\Big\|_\infty\\
\le&\Big\|\frac{1}{n}\sum_{k=1}^n\dot{y}_k\dot{\boldsymbol{x}}_k-\mathbb{E}(\dot{y}_k\dot{\boldsymbol{x}}_k)\Big\|_\infty+\Big\|\mathbb{E}\big(\widetilde{\boldsymbol{x}}_k\widetilde{\boldsymbol{x}}_k^\top\boldsymbol{\theta}^\star\big)-\mathbb{E}\big(\widetilde{y}_k\widetilde{\boldsymbol{x}}_k\big)\Big\|_\infty\\
+&\Big\|\Big(\frac{1}{n}\sum_{k=1}^n\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top-\mathbb{E}(\dot{\boldsymbol{x}}_k\dot{\boldsymbol{x}}_k^\top)\Big)\boldsymbol{\theta}^\star\Big\|_\infty:=I_1+I_2+I_3,\quad(41)
\end{aligned}$$

Step 1.1. Bounding I_1

Denote the i-th entry of $\boldsymbol{x}_k,\widetilde{\boldsymbol{x}}_k,\dot{\boldsymbol{x}}_k,\boldsymbol{\tau}_k,\boldsymbol{w}_k$ by $x_{ki},\widetilde{x}_{ki},\dot{x}_{ki},\tau_{ki},w_{ki}$, respectively. Since $|\dot{y}_k|\le|\widetilde{y}_k|+|\phi_k|+|\vartheta_k|\le|\widetilde{y}_k|+\Delta$, $\dot{x}_{ki}\le|\widetilde{x}_{ki}|+|\tau_{ki}|+|w_{ki}|\le|\widetilde{x}_{ki}|+\frac{3}{2}\bar{\Delta}$, for any integer $q\ge2$ we can bound the moments as

$$\begin{aligned}
&\sum_{k=1}^n\mathbb{E}\Big|\frac{\dot{y}_k\dot{x}_{ki}}{n}\Big|^q\le\frac{(\zeta_x+\frac{3}{2}\bar{\Delta})^{q-2}(\zeta_y+\Delta)^{q-2}}{n^q}\sum_{k=1}^n\mathbb{E}|\dot{y}_k\dot{x}_{ki}|^2\\
\le&\frac{[(\zeta_x+\frac{3}{2}\bar{\Delta})(\zeta_y+\Delta)]^{q-2}}{n^q}\sum_{k=1}^n\sqrt{\mathbb{E}|\dot{y}_k|^4\mathbb{E}|\dot{x}_{ki}|^4}\\
\lesssim&\Big(\frac{(\zeta_x+\frac{3}{2}\bar{\Delta})(\zeta_y+\Delta)}{n}\Big)^{q-2}\Big(\frac{M+\Delta^4+\bar{\Delta}^4}{n}\Big),
\end{aligned}$$

and in the last inequality we use $\mathbb{E}|\widetilde{y}_k|\le\mathbb{E}|y_k|^4\le M$ and $\mathbb{E}|\widetilde{x}_{ki}|^4\le\mathbb{E}|x_{ki}|^4\le M$. With these preparations, we apply Bernstein's inequality (Lemma 1) and a union bound, yielding that

$$\begin{aligned}
&\mathbb{P}\left(I_1\ge C_5\left\{\sqrt{\frac{(M+\Delta^4+\bar{\Delta}^4)t}{n}}+\frac{(\zeta_x+\frac{3}{2}\bar{\Delta})(\zeta_y+\Delta)t}{n}\right\}\right)\\
&\qquad\le 2d\cdot\exp(-t).
\end{aligned}$$

Set $t=\delta\log d$, we obtain that

$$I_1\lesssim(\sqrt{M}+\Delta^2+\bar{\Delta}^2)\sqrt{\frac{\delta\log d}{n}}$$

holds with probability at least $1-2d^{1-\delta}$.

Step 1.2. Bounding I_2

Noting that $\mathbb{E}(y_k\boldsymbol{x}_k)=\mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top\boldsymbol{\theta}^\star)+\mathbb{E}(\epsilon_k\boldsymbol{x}_k)=\mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top\boldsymbol{\theta}^\star)$, we could further decompose I_2 as

$$\begin{aligned}
I_2\le&\|\mathbb{E}(\widetilde{\boldsymbol{x}}_k\widetilde{\boldsymbol{x}}_k^\top\boldsymbol{\theta}^\star)-\mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top\boldsymbol{\theta}^\star)\|_\infty\\
&+\|\mathbb{E}(y_k\boldsymbol{x}_k)-\mathbb{E}(\widetilde{y}_k\widetilde{\boldsymbol{x}}_k)\|_\infty:=I_{21}+I_{22}
\end{aligned}$$

To bound I_{21}, we note that the assumption and truncation procedure for $\boldsymbol{x}_k$ are the same as in Theorem 2; thus, Step 2 in the proof of Theorem 2 can yield that

$$\big\|\mathbb{E}(\widetilde{\boldsymbol{x}}_k\widetilde{\boldsymbol{x}}_k^\top-\boldsymbol{x}_k\boldsymbol{x}_k^\top)\big\|_\infty\le\sqrt{\frac{\delta M\log d}{n}}.$$

Thus, we have

$$I_{21}\le\|\mathbb{E}(\widetilde{\boldsymbol{x}}_k\widetilde{\boldsymbol{x}}_k^\top-\boldsymbol{x}_k\boldsymbol{x}_k^\top)\|_\infty\|\boldsymbol{\theta}^\star\|_1\le R\sqrt{M}\sqrt{\frac{\delta\log d}{n}}.$$

To bound I_{22}, we estimate the i-th entry

$$\begin{aligned}
&\big|\mathbb{E}(y_kx_{ki}-\widetilde{y}_k\widetilde{x}_{ki})\big|\\
=&\big|\mathbb{E}(y_kx_{ki}-\widetilde{y}_k\widetilde{x}_{ki})\big(\mathbb{1}(|y_k|>\zeta_y)+\mathbb{1}(|x_{ki}|\ge\zeta_x)\big)\big|\\
\le&\mathbb{E}\big(|y_kx_{ki}|\mathbb{1}(|y_k|>\zeta_y)\big)+\mathbb{E}\big(|y_kx_{ki}|\mathbb{1}(|x_{ki}|\ge\zeta_x)\big)\\
\overset{(i)}{\le}&M\Big(\frac{1}{\zeta_x^2}+\frac{1}{\zeta_y^2}\Big)\lesssim\sqrt{\frac{\delta M\log d}{n}},
\end{aligned}$$

where (i) is because

$$\begin{aligned}
&\mathbb{E}\big(|y_kx_{ki}|\mathbb{1}(|y_k|>\zeta_y)\big)\\
\le&\big[\mathbb{E}|y_k^2x_{ki}^2|\big]^{1/2}\sqrt{\mathbb{P}(|y_k|>\zeta_y)}\\
\le&(\mathbb{E}y_k^4)^{1/4}(\mathbb{E}x_{ki}^4)^{1/4}\sqrt{\frac{\mathbb{E}y_k^4}{\zeta_y^4}}\le\frac{M}{\zeta_y^2}
\end{aligned}$$

and applying similar treatment to $\mathbb{E}\big(|y_kx_{ki}|\mathbb{1}(|x_{ki}|>\zeta_x)\big)$. Overall, we have

$$I_2\lesssim R\sqrt{M}\sqrt{\frac{\delta\log d}{n}}.$$

Step 1.3. Bounding I_3

We first note that

$$I_3=\|(\boldsymbol{Q}-\boldsymbol{\Sigma}^\star)\boldsymbol{\theta}^\star\|_\infty\le\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty\cdot\|\boldsymbol{\theta}^\star\|_1\le R\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty.$$

By Theorem 2, we know that

$$\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty\lesssim(\sqrt{M}+\bar{\Delta}^2)\sqrt{\frac{\delta\log d}{n}}$$

holds with probability at least $1-2d^{2-\delta}$, which leads to

$$I_3\le\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty\|\boldsymbol{\theta}^\star\|_1\lesssim R(\sqrt{M}+\bar{\Delta}^2)\sqrt{\frac{\delta\log d}{n}}.$$

Thus, by combining everything, we obtain that

$$\|\boldsymbol{Q}\boldsymbol{\theta}^\star-\boldsymbol{b}\|_\infty\lesssim(R\sqrt{M}+\Delta^2+R\bar{\Delta}^2)\sqrt{\frac{\delta\log d}{n}}$$

holds with probability at least $1-4d^{2-\delta}$. Compared to our choice of λ, the claim of this step follows.

Step 2. Bounding the Estimation Error

We are now ready to bound the estimation error of any $\widetilde{\boldsymbol{\theta}}$ satisfying (14). Set $\boldsymbol{\theta}=\boldsymbol{\theta}^\star$ in (41), it gives

$$\big\langle \boldsymbol{Q}\widetilde{\boldsymbol{\theta}}-\boldsymbol{b}+\lambda\cdot\partial\|\widetilde{\boldsymbol{\theta}}\|_1,\widetilde{\boldsymbol{\Upsilon}}\big\rangle\le 0.$$

Recall that we can assume

$$\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty\le C_6(\sqrt{M}+\bar{\Delta}^2)\sqrt{\frac{\delta\log d}{n}}$$

with probability at least $1-2d^{2-\delta}$, which leads to

$$\begin{aligned}\widetilde{\boldsymbol{\Upsilon}}^\top\boldsymbol{Q}\widetilde{\boldsymbol{\Upsilon}}&=\widetilde{\boldsymbol{\Upsilon}}^\top\boldsymbol{\Sigma}^\star\widetilde{\boldsymbol{\Upsilon}}-\widetilde{\boldsymbol{\Upsilon}}^\top(\boldsymbol{\Sigma}^\star-\boldsymbol{Q})\widetilde{\boldsymbol{\Upsilon}}\\&\ge\kappa_0\|\widetilde{\boldsymbol{\Upsilon}}\|_2^2-C_6(\sqrt{M}+\bar{\Delta}^2)\sqrt{\frac{\delta\log d}{n}}\|\widetilde{\boldsymbol{\Upsilon}}\|_1^2.\end{aligned}\tag{42}$$

Thus, it follows that

$$\begin{aligned}0&\ge\big\langle\boldsymbol{Q}\widetilde{\boldsymbol{\theta}}-\boldsymbol{b}+\lambda\cdot\partial\|\widetilde{\boldsymbol{\theta}}\|_1,\widetilde{\boldsymbol{\Upsilon}}\big\rangle\\&=\big\langle\boldsymbol{Q}\boldsymbol{\theta}^\star-\boldsymbol{b},\widetilde{\boldsymbol{\Upsilon}}\big\rangle+\widetilde{\boldsymbol{\Upsilon}}^\top\boldsymbol{Q}\widetilde{\boldsymbol{\Upsilon}}+\lambda\big\langle\partial\|\widetilde{\boldsymbol{\theta}}\|_1,\widetilde{\boldsymbol{\Upsilon}}\big\rangle\\&\overset{(i)}{\ge}C_0\sqrt{M}\|\widetilde{\boldsymbol{\Upsilon}}\|_2^2-C_6(\sqrt{M}+\bar{\Delta}^2)\sqrt{\frac{\delta\log d}{n}}\|\widetilde{\boldsymbol{\Upsilon}}\|_1^2\\&\quad-\|\boldsymbol{Q}\boldsymbol{\theta}^\star-\boldsymbol{b}\|_\infty\|\widetilde{\boldsymbol{\Upsilon}}\|_1+\lambda\big(\|\widetilde{\boldsymbol{\theta}}\|_1-\|\boldsymbol{\theta}^\star\|_1\big)\\&\overset{(ii)}{\ge}C_0\sqrt{M}\|\widetilde{\boldsymbol{\Upsilon}}\|_2^2+\lambda\big(\|\widetilde{\boldsymbol{\theta}}\|_1-\|\boldsymbol{\theta}^\star\|_1\big)\\&\quad-\Big(2C_6R(\sqrt{M}+\bar{\Delta}^2)\sqrt{\frac{\delta\log d}{n}}+\|\boldsymbol{Q}\boldsymbol{\theta}^\star-\boldsymbol{b}\|_\infty\Big)\|\widetilde{\boldsymbol{\Upsilon}}\|_1\\&\overset{(iii)}{\ge}C_0\sqrt{M}\|\widetilde{\boldsymbol{\Upsilon}}\|_2^2-\frac{\lambda}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_1+\lambda\big(\|\widetilde{\boldsymbol{\theta}}\|_1-\|\boldsymbol{\theta}^\star\|_1\big).\end{aligned}\tag{43}$$

Note that (i) is due to (42) and $\|\boldsymbol{\theta}^\star\|_1-\|\widetilde{\boldsymbol{\theta}}\|_1\ge\big\langle\partial\|\widetilde{\boldsymbol{\theta}}\|_1,-\widetilde{\boldsymbol{\Upsilon}}\big\rangle$, in (ii) we use $\|\widetilde{\boldsymbol{\Upsilon}}\|_1\le\|\widetilde{\boldsymbol{\theta}}\|_1+\|\boldsymbol{\theta}^\star\|_1\le 2R$, and from Step 1 (iii) holds when

$$\lambda=C_2(R\sqrt{M}+\Delta^2+R\bar{\Delta}^2)\sqrt{\frac{\delta\log d}{n}}$$

with sufficiently large C_2. Therefore, we arrive at $\frac{1}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_1\ge\|\widetilde{\boldsymbol{\theta}}\|_1-\|\boldsymbol{\theta}^\star\|_1$. Similar to Step 3 in the proof of Theorem 9, we can show $\|\widetilde{\boldsymbol{\Upsilon}}\|_1\le 4\sqrt{s}\|\widetilde{\boldsymbol{\Upsilon}}\|_2$. Applying (43) again, it yields

$$\kappa_0\|\widetilde{\boldsymbol{\Upsilon}}\|_2^2\le\frac{\lambda}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_1+\lambda\|\widetilde{\boldsymbol{\Upsilon}}\|_1\le\frac{3\lambda}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_1\le 6\sqrt{s}\lambda\|\widetilde{\boldsymbol{\Upsilon}}\|_2.$$

Thus, we obtain $\|\widetilde{\boldsymbol{\Upsilon}}\|_2\lesssim\mathscr{L}\sqrt{\frac{\delta s\log d}{n}}$ with $\mathscr{L}:=\frac{R\sqrt{M}+\Delta^2+R\bar{\Delta}^2}{\kappa_0}$. The proof can be concluded by further applying $\|\widetilde{\boldsymbol{\Upsilon}}\|_1\le 4\sqrt{s}\|\widetilde{\boldsymbol{\Upsilon}}\|_2$ to derive the bound for ℓ_1-norm. □

Proof of Proposition 1:

Proof. We let $\boldsymbol{\theta}=\boldsymbol{\theta}^\star$ in (14), then proceeds as the proof of Theorem 10:

$$\begin{aligned}0&\ge\big\langle\boldsymbol{Q}\widetilde{\boldsymbol{\theta}}-\boldsymbol{b}+\lambda\cdot\partial\|\widetilde{\boldsymbol{\theta}}\|_1,\widetilde{\boldsymbol{\Upsilon}}\big\rangle\\&=\widetilde{\boldsymbol{\Upsilon}}^\top\boldsymbol{\Sigma}^\star\widetilde{\boldsymbol{\Upsilon}}+\big\langle\boldsymbol{Q}\boldsymbol{\theta}^\star-\boldsymbol{b},\widetilde{\boldsymbol{\Upsilon}}\big\rangle\\&\quad+\widetilde{\boldsymbol{\Upsilon}}^\top(\boldsymbol{Q}-\boldsymbol{\Sigma}^\star)\widetilde{\boldsymbol{\Upsilon}}+\lambda\big\langle\partial\|\widetilde{\boldsymbol{\theta}}\|_1,\widetilde{\boldsymbol{\Upsilon}}\big\rangle\\&\overset{(i)}{\ge}\kappa_0\|\widetilde{\boldsymbol{\Upsilon}}\|_2^2-\|\boldsymbol{Q}\boldsymbol{\theta}^\star-\boldsymbol{b}\|_\infty\|\widetilde{\boldsymbol{\Upsilon}}\|_1\\&\quad-\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty\|\widetilde{\boldsymbol{\Upsilon}}\|_1^2+\lambda\big(\|\widetilde{\boldsymbol{\theta}}\|_1-\|\boldsymbol{\theta}^\star\|_1\big)\\&\overset{(ii)}{\ge}\kappa_0\|\widetilde{\boldsymbol{\Upsilon}}\|_2^2+\lambda\big(\|\widetilde{\boldsymbol{\theta}}\|_1-\|\boldsymbol{\theta}^\star\|_1\big)\\&\quad-\big(\|\boldsymbol{Q}\boldsymbol{\theta}^\star-\boldsymbol{b}\|_\infty+2R\cdot\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty\big)\|\widetilde{\boldsymbol{\Upsilon}}\|_1\\&\overset{(iii)}{\ge}\kappa_0\|\widetilde{\boldsymbol{\Upsilon}}\|_2^2-\frac{\lambda}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_1+\lambda\big(\|\widetilde{\boldsymbol{\theta}}\|_1-\|\boldsymbol{\theta}^\star\|_1\big),\end{aligned}\tag{44}$$

where in (i) we use $\lambda_{\min}(\boldsymbol{\Sigma}^\star)\ge\kappa_0$ and $\|\boldsymbol{\theta}^\star\|_1-\|\widetilde{\boldsymbol{\theta}}\|_1\ge\big\langle\partial\|\widetilde{\boldsymbol{\theta}}\|_1,-\widetilde{\boldsymbol{\Upsilon}}\big\rangle$, (ii) is by $\|\widetilde{\boldsymbol{\Upsilon}}\|_1\le\|\widetilde{\boldsymbol{\theta}}\|_1+\|\boldsymbol{\theta}^\star\|_1\le 2R$, in (iii) we use the assumption (15). Thus, by $\kappa_0\|\widetilde{\boldsymbol{\Upsilon}}\|_2^2\ge 0$ we obtain $2\big(\|\widetilde{\boldsymbol{\theta}}\|_1-\|\boldsymbol{\theta}^\star\|_1\big)\le\|\widetilde{\boldsymbol{\Upsilon}}\|_1$. Similarly to Step 3 in the proof of Theorem 9, we can show $\|\widetilde{\boldsymbol{\Upsilon}}\|_1\le 4\sqrt{s}\|\widetilde{\boldsymbol{\Upsilon}}\|_2$. Again we use (44), it gives

$$\begin{aligned}\kappa_0\|\widetilde{\boldsymbol{\Upsilon}}\|_2^2&\le\frac{\lambda}{2}\|\widetilde{\boldsymbol{\Upsilon}}\|_1+\lambda\big(\|\boldsymbol{\theta}^\star\|_1-\|\widetilde{\boldsymbol{\theta}}\|_1\big)\\&\le\frac{3}{2}\lambda\|\widetilde{\boldsymbol{\Upsilon}}\|_1\le 6\lambda\sqrt{s}\|\widetilde{\boldsymbol{\Upsilon}}\|_2,\end{aligned}$$

thus displaying $\|\widetilde{\boldsymbol{\Upsilon}}\|_2\le\frac{6\lambda\sqrt{s}}{\kappa_0}$. The proof can be concluded by using $\|\widetilde{\boldsymbol{\Upsilon}}\|_1\le 4\sqrt{s}\|\widetilde{\boldsymbol{\Upsilon}}\|_2$. □

Proof of Theorem 11:

Proof. To invoke Proposition 1, it is enough to verify (15). Recalling $\boldsymbol{\Sigma}^\star=\mathbb{E}(\boldsymbol{x}_k\boldsymbol{x}_k^\top)$, we first invoke [22, Thm. 1] and obtain $\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty=O\big(\sigma^2\sqrt{\frac{\delta\log d(\log n)^2}{n}}\big)$ holds with probability at least $1-2d^{2-\delta}$. This confirms $\lambda\gtrsim R\cdot\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty$. Then, it remains to upper bound $\|\boldsymbol{Q}\boldsymbol{\theta}^\star-\boldsymbol{b}\|_\infty$:

$$\begin{aligned}\|\boldsymbol{Q}\boldsymbol{\theta}^\star-\boldsymbol{b}\|_\infty&\le\|(\boldsymbol{Q}-\boldsymbol{\Sigma}^\star)\boldsymbol{\theta}^\star\|_\infty+\|\boldsymbol{\Sigma}^\star\boldsymbol{\theta}^\star-\boldsymbol{b}\|_\infty\\&\le\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty\|\boldsymbol{\theta}^\star\|_1+\Big\|\frac{\gamma_x\gamma_y}{n}\sum_{k=1}^n\dot{y}_k\dot{\boldsymbol{x}}_{k1}-\mathbb{E}(y_k\boldsymbol{x}_k)\Big\|_\infty\\&\overset{(i)}{\lesssim}\sigma^2(R+1)\sqrt{\frac{\delta\log d(\log n)^2}{n}},\end{aligned}$$

where in (i) we use a known estimate from [22, Eq. (A.31)]:

$$\begin{aligned}\mathbb{P}\left(\Big\|\frac{\gamma_x\gamma_y}{n}\sum_{k=1}^n\dot{y}_k\dot{\boldsymbol{x}}_{k1}-\mathbb{E}(y_k\boldsymbol{x}_k)\Big\|_\infty\lesssim\sigma^2\sqrt{\frac{\delta\log d(\log n)^2}{n}}\right)\\\ge 1-2d^{1-\delta}.\end{aligned}$$

Thus, by setting

$$\lambda=C_1\sigma^2R\sqrt{\frac{\delta\log d(\log n)^2}{n}},$$

(15) can be satisfied with probability at least $1-4d^{2-\delta}$, hence using Proposition 1 concludes the proof. □

Proof of Theorem 12:

Proof. The proof is again based on Proposition 1 and some ingredients from [22]. From [22, Thm. 4],

$$\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty \lesssim \Big(\frac{M^2\delta\log d}{n}\Big)^{1/4}$$

holds with probability at least $1-2d^{2-\delta}$, thus confirming $\lambda \gtrsim R\cdot\|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty$ with the same probability. Moreover,

$$\begin{aligned}\|\boldsymbol{Q}\boldsymbol{\theta}^\star-\boldsymbol{b}\|_\infty &\le \|(\boldsymbol{Q}-\boldsymbol{\Sigma}^\star)\boldsymbol{\theta}^\star\|_\infty+\|\boldsymbol{\Sigma}^\star\boldsymbol{\theta}^\star-\boldsymbol{b}\|_\infty\\ &\le \|\boldsymbol{Q}-\boldsymbol{\Sigma}^\star\|_\infty\|\boldsymbol{\theta}^\star\|_1+\Big\|\frac{\gamma_x\gamma_y}{n}\sum_{k=1}^n \dot{y}_k\dot{\boldsymbol{x}}_{k1}-\mathbb{E}(y_k\boldsymbol{x}_k)\Big\|_\infty\\ &\overset{(i)}{\lesssim} \sqrt{M}(R+1)\Big(\frac{\delta\log d}{n}\Big)^{1/4},\end{aligned}$$

where (i) is due to a known estimate from [22, Eq. (A.34)]:

$$\begin{aligned}\mathbb{P}\left(\Big\|\frac{\gamma_x\gamma_y}{n}\sum_{k=1}^n \dot{y}_k\dot{\boldsymbol{x}}_{k1}-\mathbb{E}(y_k\boldsymbol{x}_k)\Big\|_\infty \lesssim \sqrt{M}\Big(\frac{\delta\log d}{n}\Big)^{1/4}\right)\\ \ge 1-2d^{1-\delta}\end{aligned}$$

Thus, with probability at least $1-4d^{2-\delta}$, (15) holds if

$$\lambda = C_1\sqrt{M}R\Big(\frac{\delta\log d}{n}\Big)^{1/4}$$

with sufficiently large C_1. The proof can be concluded by invoking Proposition 1. □

2) **Uniform Recovery Guarantee:** We need some auxiliary results to support the proof. The first one is a concentration inequality for product process due to Mendelson [67]; the following statement can be directly adapted from [40, Thm. 8] by specifying the pseudo-metrics as ℓ_2-distance.

Lemma 9. (Concentration of Product Process)**.** *Let $\{g_{\boldsymbol{a}}\}_{\boldsymbol{a}\in\mathcal{A}}$ and $\{h_{\boldsymbol{b}}\}_{\boldsymbol{b}\in\mathcal{B}}$ be stochastic processes indexed by two sets $\mathcal{A}\subset\mathbb{R}^p$, $\mathcal{B}\subset\mathbb{R}^q$, both defined on a common probability space $(\Omega,\mathrm{A},\mathbb{P})$. We assume that there exist $K_{\mathcal{A}},K_{\mathcal{B}},r_{\mathcal{A}},r_{\mathcal{B}}\ge 0$ such that*

$$\begin{aligned}&\|g_{\boldsymbol{a}}-g_{\boldsymbol{a}'}\|_{\psi_2}\le K_{\mathcal{A}}\|\boldsymbol{a}-\boldsymbol{a}'\|_2,\ \|g_{\boldsymbol{a}}\|_{\psi_2}\le r_{\mathcal{A}},\ \forall\ \boldsymbol{a},\boldsymbol{a}'\in\mathcal{A};\\ &\|h_{\boldsymbol{b}}-h_{\boldsymbol{b}'}\|_{\psi_2}\le K_{\mathcal{B}}\|\boldsymbol{b}-\boldsymbol{b}'\|_2,\ \|h_{\boldsymbol{b}}\|_{\psi_2}\le r_{\mathcal{B}},\ \forall\ \boldsymbol{b},\boldsymbol{b}'\in\mathcal{B}.\end{aligned}$$

Finally, let $X_1,...,X_m$ be independent copies of a random variable $X\sim\mathbb{P}$, then for every $u\ge 1$ the following holds with probability at least $1-2\exp(-cu^2)$

$$\begin{aligned}&\sup_{\substack{\boldsymbol{a}\in\mathcal{A}\\ \boldsymbol{b}\in\mathcal{B}}}\frac{1}{n}\left|\sum_{i=1}^n g_{\boldsymbol{a}}(X_i)h_{\boldsymbol{b}}(X_i)-\mathbb{E}\big[g_{\boldsymbol{a}}(X_i)h_{\boldsymbol{b}}(X_i)\big]\right|\\ &\quad\le C\Big(\frac{(K_{\mathcal{A}}\cdot\omega(\mathcal{A})+u\cdot r_{\mathcal{A}})\cdot(K_{\mathcal{B}}\cdot\omega(\mathcal{B})+u\cdot r_{\mathcal{B}})}{n}\\ &\qquad+\frac{r_{\mathcal{A}}\cdot K_{\mathcal{B}}\cdot\omega(\mathcal{B})+r_{\mathcal{B}}\cdot K_{\mathcal{A}}\cdot\omega(\mathcal{A})+u\cdot r_{\mathcal{A}}r_{\mathcal{B}}}{\sqrt{n}}\Big),\end{aligned}$$

where $\omega(\mathcal{A})=\mathbb{E}\sup_{\boldsymbol{a}\in\mathcal{A}}(\boldsymbol{g}^\top\boldsymbol{a})$ with $\boldsymbol{g}\sim\mathcal{N}(0,\boldsymbol{I}_p)$ is the Gaussian width of $\mathcal{A}\subset\mathbb{R}^p$, and similarly, $\omega(\mathcal{B})$ is the Gaussian width of $\mathcal{B}$.

We will use the following result that can be found in [61, Thm. 8].

Lemma 10. *Let $(X_{\boldsymbol{u}})_{\boldsymbol{u}\in\mathcal{T}}$ be a random process indexed by points in a bounded set $\mathcal{T}\subset\mathbb{R}^n$. Assume that the process has sub-Gaussian increments, i.e., there exists $M>0$ such that $\|X_{\boldsymbol{u}}-X_{\boldsymbol{v}}\|_{\psi_2}\le M\|\boldsymbol{u}-\boldsymbol{v}\|_2$ holds for any $\boldsymbol{u},\boldsymbol{v}\in\mathcal{T}$. Then for every $t>0$, the event*

$$\sup_{\boldsymbol{u},\boldsymbol{v}\in\mathcal{T}}|X_{\boldsymbol{u}}-X_{\boldsymbol{v}}|\le CM\cdot\big(\omega(\mathcal{T})+t\cdot\mathrm{diam}(\mathcal{T})\big)$$

holds with probability at least $1-\exp(-t^2)$, where $\mathrm{diam}(\mathcal{T}):=\sup_{\boldsymbol{x},\boldsymbol{y}\in\mathcal{T}}\|\boldsymbol{x}-\boldsymbol{y}\|_2$ denotes the diameter of $\mathcal{T}$.

Proof of Theorem 13:

Proof. We start from the optimality

$$\sum_{k=1}^n(\dot{y}_k-\boldsymbol{x}_k^\top\widehat{\boldsymbol{\theta}})^2\le\sum_{k=1}^n(\dot{y}_k-\boldsymbol{x}_k^\top\boldsymbol{\theta}^\star)^2.$$

By substituting $\widehat{\boldsymbol{\theta}}=\boldsymbol{\theta}^\star+\widehat{\boldsymbol{\Upsilon}}$ and performing some algebra, we obtain

$$\sum_{k=1}^n(\boldsymbol{x}_k^\top\widehat{\boldsymbol{\Upsilon}})^2\le 2\sum_{k=1}^n\big(\dot{y}_k-\boldsymbol{x}_k^\top\boldsymbol{\theta}^\star\big)\boldsymbol{x}_k^\top\widehat{\boldsymbol{\Upsilon}}.$$

Due to the constraint we have $\|\boldsymbol{\theta}^\star+\widehat{\boldsymbol{\Upsilon}}\|_1\le\|\boldsymbol{\theta}^\star\|_1$, then similar to (40) we can show $\|\widehat{\boldsymbol{\Upsilon}}\|_1\le 2\sqrt{s}\|\widehat{\boldsymbol{\Upsilon}}\|_2$ holds. Thus, we let $\mathscr{V}=\{\boldsymbol{v}:\|\boldsymbol{v}\|_2=1,\|\boldsymbol{v}\|_1\le 2\sqrt{s}\}$, then the following holds uniformly for all $\boldsymbol{\theta}^\star\in\Sigma_{s,R_0}$

$$\begin{aligned}&\|\widehat{\boldsymbol{\Upsilon}}\|_2^2\cdot\inf_{\boldsymbol{v}\in\mathscr{V}}\sum_{k=1}^n(\boldsymbol{x}_k^\top\boldsymbol{v})^2\\ &\le 2\|\widehat{\boldsymbol{\Upsilon}}\|_2\cdot\sup_{\boldsymbol{v}\in\mathscr{V}}\sum_{k=1}^n(\dot{y}_k-\boldsymbol{x}_k^\top\boldsymbol{\theta}^\star)\boldsymbol{x}_k^\top\boldsymbol{v}.\end{aligned}\tag{45}$$

Similarly to previous developments, our strategy is to lower bound the left-hand side while upper bound the right hand side, but with the difference that the bounds must be valid uniformly for all $\boldsymbol{\theta}^\star\in\Sigma_{s,R_0}$.

Step 1. Bounding the Left-Hand Side From Below

Letting $\bar{\Delta}=0$ and restricting $\boldsymbol{v}$ to $\mathscr{V}$, we use (38) in the proof of Theorem 9, then for some constant $c(\kappa_0,\sigma)$ depending on κ_0,σ, with probability at least $1-d^{-\delta}$

$$\inf_{\boldsymbol{v}\in\mathscr{V}}\frac{1}{\sqrt{n}}\left[\sum_{k=1}^n(\boldsymbol{x}_k^\top\boldsymbol{v})^2\right]^{1/2}\ge\sqrt{\kappa_0}-c(\kappa_0,\sigma)\cdot\sqrt{\frac{s\log d}{n}}.$$

Thus, if $n\ge\frac{4c^2(\kappa_0,\sigma)}{\kappa_0}s\log d$, then it holds that

$$\inf_{\boldsymbol{v}\in\mathscr{V}}\sum_{k=1}^n(\boldsymbol{x}_k^\top\boldsymbol{v})^2\ge\frac{1}{4}\kappa_0 n.$$

Step 2. Bounding the Right-Hand Side Uniformly

To pursue the uniformity over $\boldsymbol{\theta}^\star\in\Sigma_{s,R_0}$, we take a supremum by replacing specific $\boldsymbol{\theta}^\star$ with $\sup_{\boldsymbol{\theta}\in\Sigma_{s,R_0}}$, then we consider the upper bound on

$$I:=\sup_{\boldsymbol{\theta}\in\Sigma_{s,R_0}}\sup_{\boldsymbol{v}\in\mathscr{V}}\sum_{k=1}^n(\dot{y}_k-\boldsymbol{x}_k^\top\boldsymbol{\theta})\boldsymbol{x}_k^\top\boldsymbol{v},\tag{46}$$

where $\dot{y}_k = \mathcal{Q}_\Delta(\widetilde{y}_k + \tau_k)$, $\widetilde{y}_k = \mathscr{T}_{\zeta_y}(y_k) := \mathrm{sign}(y_k)\min\{|y_k|, \zeta_y\}$, $y_k = \boldsymbol{x}_k^\top\boldsymbol{\theta} + \epsilon_k$; note that $\dot{y}_k, \widetilde{y}_k, y_k$ depend on $\boldsymbol{\theta}$, and we will use notation $\dot{y}_{\boldsymbol{\theta},k}, \widetilde{y}_{\boldsymbol{\theta},k}, y_{\boldsymbol{\theta},k}$ to indicate such dependence when necessary. In this proof, the ranges of $\boldsymbol{\theta}$ and $\boldsymbol{v}$ (e.g., in supremum), if omitted, are respectively $\boldsymbol{\theta} \in \Sigma_{s,R_0}$ and $\boldsymbol{v} \in \mathscr{V}$. Now let the quantization noise be $\xi_k = \dot{y}_k - \widetilde{y}_k$, observing that $\mathbb{E}(y_k\boldsymbol{x}_k^\top\boldsymbol{v}) = \mathbb{E}(\boldsymbol{\theta}^\top\boldsymbol{x}_k\boldsymbol{x}_k^\top\boldsymbol{v}) + \mathbb{E}(\epsilon_k\boldsymbol{x}_k^\top\boldsymbol{v}) = \mathbb{E}(\boldsymbol{\theta}^\top\boldsymbol{x}_k\boldsymbol{x}_k^\top\boldsymbol{v})$, then we can first decompose I as

$$\begin{aligned} I &\le \sup_{\boldsymbol{\theta},\boldsymbol{v}} \sum_{k=1}^n \xi_k\boldsymbol{x}_k^\top\boldsymbol{v} + \sup_{\boldsymbol{\theta},\boldsymbol{v}} \sum_{k=1}^n \big(\widetilde{y}_k\boldsymbol{x}_k^\top\boldsymbol{v} - \mathbb{E}[\widetilde{y}_k\boldsymbol{x}_k^\top\boldsymbol{v}]\big) \\ &\quad + \sup_{\boldsymbol{\theta},\boldsymbol{v}} \sum_{k=1}^n \mathbb{E}\big((\widetilde{y}_k - y_k)\boldsymbol{x}_k^\top\boldsymbol{v}\big) \\ &\quad + \sup_{\boldsymbol{\theta},\boldsymbol{v}} \sum_{k=1}^n \big(\boldsymbol{\theta}^\top\boldsymbol{x}_k\boldsymbol{x}_k^\top\boldsymbol{v} - \mathbb{E}[\boldsymbol{\theta}^\top\boldsymbol{x}_k\boldsymbol{x}_k^\top\boldsymbol{v}]\big) \\ &:= I_0 + I_1 + I_2 + I_3, \end{aligned} \tag{47}$$

where I_0 is the term arising from quantization, I_1 is the concentration term involving truncation of heavy-tailed data for which we develop some new machinery to bound it, I_2 is the bias term, I_3 is a more regular concentration term that can be bounded via Lemma 9. In the remainder of the proof, we will bound I_1, I_2, I_3 separately and finally deal with I_0.

Step 2.1. Bounding I_1

Using $\widetilde{y}_k = \mathscr{T}_{\zeta_y}(\boldsymbol{x}_k^\top\boldsymbol{\theta} + \epsilon_k)$, I_1 is concerned with the concentration of the product process

$$\Big\{\sum_{k=1}^n \mathscr{T}_{\zeta_y}(\boldsymbol{\theta}^\top\boldsymbol{x}_k + \epsilon_k)\boldsymbol{x}_k^\top\boldsymbol{v}\Big\}_{\boldsymbol{\theta},\boldsymbol{v}}$$

about its mean. It is natural to apply Lemma 9 towards this end, but we lack good bound on $\|\mathscr{T}_{\zeta_y}(\boldsymbol{\theta}^\top\boldsymbol{x}_k + \epsilon_k)\|_{\psi_2}$ because of the heavy-tailedness of ϵ_k (on the other hand, the bound $O(\zeta_y)$ is just too crude to yield a sharp rate). Our strategy is already introduced in the mainbody — we introduce $\widetilde{z}_k := \widetilde{y}_k - \mathscr{T}_{\zeta_y}(\epsilon_k)$ and decompose I_1 as

$$\begin{aligned} I_1 &\le \underbrace{\sup_{\boldsymbol{v},\boldsymbol{\theta}} \sum_{k=1}^n \big(\widetilde{z}_k\boldsymbol{x}_k^\top\boldsymbol{v} - \mathbb{E}[\widetilde{z}_k\boldsymbol{x}_k^\top\boldsymbol{v}]\big)}_{:=I_{11}} \\ &\quad + \underbrace{\sup_{\boldsymbol{v}} \sum_{k=1}^n \big(\mathscr{T}_{\zeta_y}(\epsilon_k)\boldsymbol{x}_k^\top\boldsymbol{v} - \mathbb{E}[\mathscr{T}_{\zeta_y}(\epsilon_k)\boldsymbol{x}_k^\top\boldsymbol{v}]\big)}_{:=I_{12}}. \end{aligned}$$

Thus, it suffices to bound I_{11} and I_{12}.

Step 2.1.1. Bounding I_{11}

We use Lemma 9 to bound I_{11}. For any $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathscr{V}$, it is evident that we have $\|\boldsymbol{x}_k^\top\boldsymbol{v}\|_{\psi_2} \le \|\boldsymbol{x}_k\|_{\psi_2} \le \sigma$ and $\|\boldsymbol{x}_k^\top\boldsymbol{v}_1 - \boldsymbol{x}_k^\top\boldsymbol{v}_2\|_{\psi_2} \le \sigma\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_2$. Regarding

$$\widetilde{z}_k = \widetilde{z}_{\boldsymbol{\theta},k} := \mathscr{T}_{\zeta_y}(\boldsymbol{x}_k^\top\boldsymbol{\theta} + \epsilon_k) - \mathscr{T}_{\zeta_y}(\epsilon_k)$$

indexed by $\boldsymbol{\theta} \in \Sigma_{s,R_0}$, the 1-Lipschitzness of $\mathscr{T}_{\zeta_y}(\cdot)$ gives $|\widetilde{z}_k| \le |\boldsymbol{x}_k^\top\boldsymbol{\theta}|$, and then the definition of sub-Gaussian norm yields

$$\|\widetilde{z}_k\|_{\psi_2} \le \|\boldsymbol{x}_k^\top\boldsymbol{\theta}\|_{\psi_2} \le \|\boldsymbol{x}_k\|_{\psi_2}\|\boldsymbol{\theta}\|_2 \le R_0\sigma$$

(this addresses the aforementioned issue). Further, for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Sigma_{s,R_0}$ we verify the sub-Gaussian increments

$$\begin{aligned} &|\widetilde{z}_{\boldsymbol{\theta}_1,k} - \widetilde{z}_{\boldsymbol{\theta}_2,k}| \\ =&\big|\mathscr{T}_{\zeta_y}(\boldsymbol{x}_k^\top\boldsymbol{\theta}_1 + \epsilon_k) - \mathscr{T}_{\zeta_y}(\epsilon_k) \\ &\qquad - \mathscr{T}_{\zeta_y}(\boldsymbol{x}_k^\top\boldsymbol{\theta}_2 + \epsilon_k) + \mathscr{T}_{\zeta_y}(\epsilon_k)\big| \\ =&\big|\mathscr{T}_{\zeta_y}(\boldsymbol{x}_k^\top\boldsymbol{\theta}_1 + \epsilon_k) - \mathscr{T}_{\zeta_y}(\boldsymbol{x}_k^\top\boldsymbol{\theta}_2 + \epsilon_k)\big| \\ \le&\big|\boldsymbol{x}_k^\top(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)\big|, \end{aligned} \tag{48}$$

which leads to

$$\begin{aligned} &\|\widetilde{z}_{\boldsymbol{\theta}_1,k} - \widetilde{z}_{\boldsymbol{\theta}_2,k}\|_{\psi_2} \le \|\boldsymbol{x}_k^\top(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)\|_{\psi_2} \\ \le&\|\boldsymbol{x}_k\|_{\psi_2}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2 \le \sigma\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2 \end{aligned}$$

With these preparations, we can invoke Lemma 9 use the well-known estimates $\omega(\Sigma_{s,R_0}), \omega(\mathscr{V}) = O(\sqrt{s\log d})$[17] to obtain that, with probability at least $1 - 2\exp(-cu^2)$ we have

$$\begin{aligned} I_{11} &\lesssim \sigma^2\Big[\sqrt{n}\big(\omega(\Sigma_{s,R_0}) + \omega(\mathscr{V}) + u\big) \\ &\qquad + \big(\omega(\Sigma_{s,R_0}) + u\big)\cdot\big(\omega(\mathscr{V}) + u\big)\Big] \\ &\lesssim \sigma^2\Big[\sqrt{n}\big(\sqrt{s\log d} + u\big) \\ &\qquad + \big(\sqrt{s\log d} + u\big)\cdot\big(\sqrt{s\log d} + u\big)\Big]. \end{aligned} \tag{49}$$

Therefore, we can set $u = \sqrt{\delta s\log d}$ in (49), under the scaling of $n \gtrsim s\delta\log d$ it provides

$$\mathbb{P}\big(I_{11} \lesssim \sigma^2\sqrt{n\delta s\log d}\big) \ge 1 - 2d^{-\delta\Omega(s)}. \tag{50}$$

Step 2.1.2. Bounding I_{12}

By $\|\boldsymbol{v}\|_1 \le 2\sqrt{s}$ we have

$$I_{12} \le 2\sqrt{s}\|\sum_{k=1}^n(\mathscr{T}_{\zeta_y}(\epsilon_k)\boldsymbol{x}_k - \mathbb{E}[\mathscr{T}_{\zeta_y}(\epsilon_k)\boldsymbol{x}_k])\|_\infty.$$

Then to apply Bernstein's inequality, for integer $q \ge 2$ and $i \in [d]$, analogously to (35) in the proof of Theorem 9, we can bound that

$$\begin{aligned} &\sum_{k=1}^n \mathbb{E}\left|\frac{\mathscr{T}_{\zeta_y}(\epsilon_k)x_{ki}}{n}\right|^q \\ &\le \Big(\frac{\zeta_y}{n}\Big)^{q-2}\frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\big|\mathscr{T}_{\zeta_y}^2(\epsilon_k)x_{ki}^q\big| \\ &\le \Big(\frac{\sigma\zeta_y}{n}\Big)^{q-2}\Big(\frac{\sigma^2M^{\frac{1}{l}}}{n}\Big)(Cq)^{\frac{q}{2}} \le \frac{q!}{2}v_0c_0^{q-2}, \end{aligned}$$

for some $v_0 = O\big(\frac{\sigma^2M^{1/l}}{n}\big)$, $c_0 = O\big(\frac{\sigma\zeta_y}{n}\big)$. Then we use Lemma 1 to obtain that, with probability at least $1-2\exp(-t)$ we have

$$\begin{aligned} &\left|\frac{1}{n}\sum_{k=1}^n(\mathscr{T}_{\zeta_y}(\epsilon_k)x_{ki} - \mathbb{E}[\mathscr{T}_{\zeta_y}(\epsilon_k)x_{ki}])\right| \\ &\qquad \le C\sigma\Big(M^{\frac{1}{2l}}\sqrt{\frac{t}{n}} + \frac{\zeta_y t}{n}\Big) \end{aligned}$$

[17] In fact, we have the tighter estimate $\omega(\Sigma_{s,R_0}), \omega(\mathscr{V}) = O\Big(\sqrt{s\log\big(\frac{ed}{s}\big)}\Big)$ (e.g., [76]) but we simply put $\sqrt{s\log d}$ to be consistent with earlier results concerning unconstrained Lasso.

Then we use $\zeta_y \asymp \big(\sigma + M^{\frac{1}{2l}}\big)\sqrt{\frac{n}{\delta \log d}}$, set $t \asymp \delta \log d$, and take a union bound over $i \in [d]$ to obtain that,

$$\|\sum_{k=1}^{n} \frac{1}{n}(\mathscr{T}_{\zeta_y}(\epsilon_k)\boldsymbol{x}_k - \mathbb{E}[\mathscr{T}_{\zeta_y}(\epsilon_k)\boldsymbol{x}_k])\|_\infty \lesssim \sigma(M^{1/(2l)} + \sigma)\sqrt{\frac{\delta \log d}{n}}$$

holds with probability at least $1 - 2d^{1-\delta}$, which implies the following under the same probability

$$I_{12} \lesssim \sigma\big(M^{\frac{1}{2l}} + \sigma\big)\sqrt{ns\delta \log d}. \tag{51}$$

Therefore, combining (50) and (51), we obtain that

$$I_1 \lesssim \sigma\big(M^{\frac{1}{2l}} + \sigma\big)\sqrt{ns\delta \log d}$$

with the promised probability.

Step 2.2. Bounding I_2

For this bias term the supremum does not make things harder. We begin with

$$I_2 = n \cdot \sup_{\boldsymbol{\theta},\boldsymbol{v}} \mathbb{E}\big((\widetilde{y}_k - y_k)\boldsymbol{x}_k^\top \boldsymbol{v}\big) \le 2n\sqrt{s} \cdot \sup_{\boldsymbol{\theta}} \|\mathbb{E}(\widetilde{y}_k - y_k)\boldsymbol{x}_k\|_\infty.$$

Fix any $\boldsymbol{\theta} \in \Sigma_{s,R_0}$, we have

$$\mathbb{E}|y_k|^{2l} \lesssim \mathbb{E}|\boldsymbol{x}_k^\top \boldsymbol{\theta}|^{2l} + \mathbb{E}|\epsilon_k|^{2l} \lesssim M + \sigma^{2l}.$$

Then following arguments similarly to (36) we obtain

$$\|\mathbb{E}(\widetilde{y}_k - y_k)\boldsymbol{x}_k\|_\infty \lesssim \frac{\sigma M^{1/l}}{\zeta_y} \lesssim \sigma(M^{1/(2l)} + \sigma)\sqrt{\frac{\delta \log d}{n}},$$

which implies $I_2 \lesssim \sigma M^{\frac{1}{2l}}\sqrt{n\delta s \log d}$.

Step 2.3. Bounding I_3

It is evident that we can apply Lemma 9 with $(g_{\boldsymbol{\theta}}(\boldsymbol{x}_k), h_{\boldsymbol{v}}(\boldsymbol{x}_k)) = (\boldsymbol{\theta}^\top \boldsymbol{x}_k, \boldsymbol{v}^\top \boldsymbol{x}_k)$, $(\mathcal{A}, \mathcal{B}) = (\Sigma_{s,R_0}, \mathscr{V})$, and hence with $K_{\mathcal{A}}, r_{\mathcal{A}}, K_{\mathcal{B}}, r_{\mathcal{B}} = O(\sigma)$. Along with $\omega(\Sigma_{s,R_0}), \omega(\mathscr{V}) \lesssim \sqrt{s \log d}$, we obtain that the following holds with probability at least $1 - 2\exp(-cu^2)$:

$$I_3 \le \sigma^2\Big[\big(\sqrt{s \log d} + u\big)^2 + \sqrt{n}\big(\sqrt{s \log d} + u\big)\Big].$$

By taking $u \asymp \sqrt{s\delta \log d}$, under the scaling $n \gtrsim \delta s \log d$, it follows that $I_3 \lesssim \sigma^2\sqrt{n\delta s \log d}$ with probability at least $1 - 2d^{-\delta\Omega(s)}$.

Step 2.4. Bounding I_0

It remains to bound $I_0 = \sup_{\boldsymbol{\theta},\boldsymbol{v}} \sum_{k=1}^n \xi_k \boldsymbol{x}_k^\top \boldsymbol{v}$. Bounding I_0 is similar to establishing the "limited projection distortion (LPD)" property in [94], but the key distinction is that $\boldsymbol{\theta}$ and $\boldsymbol{v}$ in I_0 take value in different spaces.

The main difficulty associated with "$\sup_{\boldsymbol{\theta}}$" lies in the discontinuity of the quantization noise $\xi_k := \mathcal{Q}_\Delta(\widetilde{y}_k + \tau_k) - \widetilde{y}_k$, which we overcome by a covering argument and some machinery developed in [94, Prop. 6.1]. However, the essential difference from [94] is that we use Lemma 10 to handle "$\sup_{\boldsymbol{v}}$", while [94] again used covering argument for $\boldsymbol{v}$ to strengthen their Proposition 6.1 to their Proposition 6.2, which is unfortunately insufficient in our setting because the covering number of $\mathscr{V}$ significantly increases under smaller covering radius (on the other hand, using covering argument for $\boldsymbol{v}$ suffices for the analyses in [94] regarding a different estimator named *projected back projection*).

Let us first construct a ρ-net of Σ_{s,R_0} denoted by $\mathcal{G} = \{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_N\}$, so that for any $\boldsymbol{\theta} \in \Sigma_{s,R_0}$ we can pick $\boldsymbol{\theta}' \in \mathcal{G}$ satisfying $\|\boldsymbol{\theta}' - \boldsymbol{\theta}\|_2 \le \rho$; here, the covering radius ρ is to be chosen later, and we assume that $N \le \big(\frac{9d}{\rho s}\big)^s$ [76, Lemma 3.3]. As is standard in a covering argument, we first control I_0 over the net $\mathcal{G}$ (by replacing "$\sup_{\boldsymbol{\theta} \in \Sigma_{s,R_0}}$" with "$\sup_{\boldsymbol{\theta} \in \mathcal{G}}$"), and then bound the approximation error induced by such replacement.

Step 2.4.1. Bounding I_0 over $\mathcal{G}$

In this step, we want to bound $I_{0,\mathcal{G}} := \sup_{\boldsymbol{\theta} \in \mathcal{G}} \sup_{\boldsymbol{v} \in \mathscr{V}} \sum_{k=1}^n \xi_k \boldsymbol{x}_k^\top \boldsymbol{v}$. First let us consider a fixed $\boldsymbol{\theta} \in \Sigma_{s,R_0}$. Then since $|\xi_k| \le \Delta$, we have $\|\xi_k \boldsymbol{x}_k\|_{\psi_2} \lesssim \Delta\sigma$. Because $\{\xi_k \boldsymbol{x}_k : k \in [n]\}$ are independent zero mean, by [92, Prop. 2.6.1] we have $\big\|\sum_{k=1}^n \xi_k \boldsymbol{x}_k\big\|_{\psi_2} \lesssim \sqrt{n}\Delta\sigma$. Define $\mathscr{V}' = \mathscr{V} \cup \{0\}$, then for any $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathscr{V}'$ we have

$$\Big\|\Big(\sum_{k=1}^n \xi_k \boldsymbol{x}_k\Big)^\top \boldsymbol{v}_1 - \Big(\sum_{k=1}^n \xi_k \boldsymbol{x}_k\Big)^\top \boldsymbol{v}_2\Big\|_{\psi_2} \le C\sqrt{n}\Delta\sigma\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_2.$$

Thus, by Lemma 10, it holds with probability at least $1 - \exp(-u^2)$ that

$$\begin{aligned}
&\sup_{\boldsymbol{v} \in \mathscr{V}} \Big(\sum_{k=1}^n \xi_k \boldsymbol{x}_k\Big)^\top \boldsymbol{v} \\
&\le \sup_{\boldsymbol{v}, \boldsymbol{v}' \in \mathscr{V}'} \Big|\Big(\sum_{k=1}^n \xi_k \boldsymbol{x}_k\Big)^\top \boldsymbol{v} - \Big(\sum_{k=1}^n \xi_k \boldsymbol{x}_k\Big)^\top \boldsymbol{v}'\Big| \\
&\le C\sqrt{n}\Delta\sigma\big(\omega(\mathscr{V}') + u\big) \le C_1\sqrt{n}\Delta\sigma\Big(\sqrt{s \log \frac{9d}{s}} + u\Big).
\end{aligned}$$

Moreover, by a union bound over $\mathcal{G}$, we obtain that

$$I_{0,\mathcal{G}} \lesssim \sigma\Delta\sqrt{n}\Big(\sqrt{s \log \frac{9d}{s}} + u\Big)$$

holds with probability at least $1 - \exp\big(s \log \frac{9d}{\rho s} - u^2\big)$. We set $u \asymp \sqrt{s\delta \log\big(\frac{9d}{\rho s}\big)}$ and arrive at

$$\mathbb{P}\left(I_{0,\mathcal{G}} \lesssim \sigma\Delta\sqrt{ns\delta \log \frac{9d}{\rho s}}\right) \ge 1 - \Big(\frac{9d}{\rho s}\Big)^{-\Omega(\delta s)}. \tag{52}$$

Step 2.4.2. Bounding the Approximation Error

From now on we indicate the dependence of ξ_k on $\boldsymbol{\theta}$ by using the notation $\xi_{\boldsymbol{\theta},k} := \mathcal{Q}_\Delta(\widetilde{y}_{\boldsymbol{\theta},k} + \tau_k) - \widetilde{y}_{\boldsymbol{\theta},k}$ where $\widetilde{y}_{\boldsymbol{\theta},k} = \mathscr{T}_{\zeta_y}(\boldsymbol{x}_k^\top \boldsymbol{\theta} + \epsilon_k)$. For any $\boldsymbol{\theta} \in \Sigma_{s,R_0}$ we pick one $\boldsymbol{\theta}' \in \mathcal{G}$ such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_2 \le \rho$; we fix such correspondence and remember that from now on every $\boldsymbol{\theta} \in \Sigma_{s,R_0}$ is associated with some $\boldsymbol{\theta}' \in \mathcal{G}$, (which of course depends on $\boldsymbol{\theta}$ but our notation omits such dependence). Thus, we can bound $I_0 = \sup_{\boldsymbol{\theta},\boldsymbol{v}} \sum_{k=1}^n \xi_{\boldsymbol{\theta},k} \boldsymbol{x}_k^\top \boldsymbol{v}$ as

$$\begin{aligned}
I_0 &\le \sup_{\boldsymbol{\theta},\boldsymbol{v}} \sum_{k=1}^n \xi_{\boldsymbol{\theta}',k} \boldsymbol{x}_k^\top \boldsymbol{v} + \sup_{\boldsymbol{\theta},\boldsymbol{v}} \sum_{k=1}^n (\xi_{\boldsymbol{\theta},k} - \xi_{\boldsymbol{\theta}',k}) \boldsymbol{x}_k^\top \boldsymbol{v} \\
&\le I_{0,\mathcal{G}} + I_{01}.
\end{aligned} \tag{53}$$

Note that the bound on $I_{0,\mathcal{G}}$ is available in (52), so it remains to bound $I_{01} := \sup_{\boldsymbol{\theta},\boldsymbol{v}} \sum_{k=1}^n (\xi_{\boldsymbol{\theta},k} - \xi_{\boldsymbol{\theta}',k}) \boldsymbol{x}_k^\top \boldsymbol{v}$, which can be understood as the approximation error of the net $\mathcal{G}$ regarding the empirical process of interest. To facilitate the presentation we switch to the more compact notations — let $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ with rows $\boldsymbol{x}_k^\top$ be the sensing matrix, $\boldsymbol{\xi}_{\boldsymbol{\theta}} = [\xi_{\boldsymbol{\theta},k}] \in \mathbb{R}^n$ be the quantization error indexed by $\boldsymbol{\theta}$, $\boldsymbol{\tau} = [\tau_k] \in \mathbb{R}^n$ be the random dither vector, $\boldsymbol{\epsilon} = [\epsilon_k] \in \mathbb{R}^n$ be the heavy-tailed noise vector, $\boldsymbol{y}_{\boldsymbol{\theta}} = [y_{\boldsymbol{\theta},k}] = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon} \in \mathbb{R}^n$ and $\widetilde{\boldsymbol{y}}_{\boldsymbol{\theta}} = [\widetilde{y}_{\boldsymbol{\theta},k}] = \mathscr{T}_{\zeta_y}(\boldsymbol{y}_{\boldsymbol{\theta}})$ be the measurement vector and truncated measurement vector, respectively. With these conventions we can write $I_{01} = \sup_{\boldsymbol{\theta},\boldsymbol{v}} (\boldsymbol{\xi}_{\boldsymbol{\theta}} - \boldsymbol{\xi}_{\boldsymbol{\theta}'})^\top \boldsymbol{X}\boldsymbol{v}$. Recall that a specific $\boldsymbol{\theta}'$ has been specified for each $\boldsymbol{\theta} \in \Sigma_{s,R_0}$, so defining $\boldsymbol{\Psi}_{\boldsymbol{\theta}} := \boldsymbol{\xi}_{\boldsymbol{\theta}} - \boldsymbol{\xi}_{\boldsymbol{\theta}'}$ allows us to write $I_{01} = \sup_{\boldsymbol{\theta},\boldsymbol{v}} \boldsymbol{\Psi}_{\boldsymbol{\theta}}^\top \boldsymbol{X}\boldsymbol{v}$. Further, we define $\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\theta}} := \widetilde{\boldsymbol{y}}_{\boldsymbol{\theta}} - \widetilde{\boldsymbol{y}}_{\boldsymbol{\theta}'}, \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}} := \mathcal{Q}_\Delta(\widetilde{\boldsymbol{y}}_{\boldsymbol{\theta}} + \boldsymbol{\tau}) - \mathcal{Q}_\Delta(\widetilde{\boldsymbol{y}}_{\boldsymbol{\theta}'} + \boldsymbol{\tau})$ and make the following observation

$$\boldsymbol{\Psi}_{\boldsymbol{\theta}} = \boldsymbol{\xi}_{\boldsymbol{\theta}} - \boldsymbol{\xi}_{\boldsymbol{\theta}'} = \widehat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\theta}}. \tag{54}$$

We pause to establish a property of $\boldsymbol{X}$ that holds w.h.p.. Specifically, we restrict (37) to $\boldsymbol{v} \in \mathscr{V}$ (recall that $\bar{\Delta} = 0$ and so $\dot{\boldsymbol{X}} = \boldsymbol{X}$ and $\dot{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^\star$), then under the promised probability it holds for some $c(\kappa_0, \sigma)$ that

$$\sup_{\boldsymbol{v}\in\mathscr{V}} \left| \frac{\|\boldsymbol{X}\boldsymbol{v}\|_2}{\sqrt{n}} - \|\sqrt{\boldsymbol{\Sigma}^\star}\boldsymbol{v}\|_2 \right| \le c(\kappa_0,\sigma)\sqrt{\frac{\delta s \log d}{n}}.$$

Thus, when $n \gtrsim \delta s \log d$ with for large enough hidden constant depending on (κ_0, σ), it holds that

$$\sup_{\boldsymbol{v}\in\mathscr{V}} \frac{\|\boldsymbol{X}\boldsymbol{v}\|_2}{\sqrt{n}} \le \sup_{\boldsymbol{v}\in\mathscr{V}} \|\sqrt{\boldsymbol{\Sigma}^\star}\boldsymbol{v}\|_2 + \sqrt{\kappa_1} \le 2\sqrt{\kappa_1}. \tag{55}$$

We proceed by assuming we are on this event, which allows us to bound I_{01} as

$$\begin{aligned} I_{01} = \sup_{\theta,\boldsymbol{v}} \boldsymbol{\Psi}_{\boldsymbol{\theta}}^\top \boldsymbol{X}\boldsymbol{v} &\le \sup_{\boldsymbol{\theta}} \|\boldsymbol{\Psi}_{\boldsymbol{\theta}}\|_2 \sup_{\boldsymbol{v}\in\mathscr{V}} \|\boldsymbol{X}\boldsymbol{v}\|_2 \\ &\le 2\sqrt{\kappa_1 n} \cdot \sup_{\boldsymbol{\theta}} \|\boldsymbol{\Psi}_{\boldsymbol{\theta}}\|_2. \end{aligned} \tag{56}$$

To bound $\sup_{\boldsymbol{\theta}} \|\boldsymbol{\Psi}_{\boldsymbol{\theta}}\|_2$, motivated by (54), we will investigate $\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}}$ more carefully. We pick a threshold $\eta \in (0, \frac{\Delta}{2})$ (that is to be chosen later), and by the 1-Lipschitzness of $\mathscr{T}_{\zeta_y}(\cdot)$ we have

$$\begin{aligned} \sup_{\boldsymbol{\theta}} \|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\theta}}\|_2 = \sup_{\boldsymbol{\theta}} \|\widetilde{\boldsymbol{y}}_{\boldsymbol{\theta}} - \widetilde{\boldsymbol{y}}_{\boldsymbol{\theta}'}\|_2 &\le \sup_{\boldsymbol{\theta}} \|\boldsymbol{y}_{\boldsymbol{\theta}} - \boldsymbol{y}_{\boldsymbol{\theta}'}\|_2 \\ = \sup_{\boldsymbol{\theta}} \|\boldsymbol{X}(\boldsymbol{\theta} - \boldsymbol{\theta}')\|_2 &\le 2\sqrt{\kappa_1 n}\rho, \end{aligned} \tag{57}$$

where the last inequality is because $\boldsymbol{\theta} - \boldsymbol{\theta}'$ is $2s$-sparse, hence (55) implies $\|\boldsymbol{X}(\boldsymbol{\theta} - \boldsymbol{\theta}')\|_2 \le 2\sqrt{\kappa_1 n}\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_2 \le 2\sqrt{\kappa_1 n}\rho$.

To proceed, we will define for specific $\boldsymbol{\theta}$ the index vectors $\boldsymbol{J}_{\boldsymbol{\theta},1}, \boldsymbol{J}_{\boldsymbol{\theta},2} \in \{0,1\}^n$ and use $|\boldsymbol{J}_{\boldsymbol{\theta},1}|$ to denote the number of 1s in $\boldsymbol{J}_{\boldsymbol{\theta},1}$ (similar meaning for $|\boldsymbol{J}_{\boldsymbol{\theta},2}|$). Specifically, using the entry-wise notation $\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\theta}} = [\widetilde{\Psi}_{\boldsymbol{\theta},k}]$ we define $\boldsymbol{J}_{\boldsymbol{\theta},1} = [\mathbb{1}(|\widetilde{\Psi}_{\boldsymbol{\theta},k}| \ge \eta)]$. Recall that (57) gives $\sup_{\boldsymbol{\theta}} \|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\theta}}\|_2^2 \le 4\kappa_1 n\rho^2$; combined with the simple observation $\sup_{\boldsymbol{\theta}} \|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\theta}}\|_2^2 \ge \sup_{\boldsymbol{\theta}} \eta^2 |\boldsymbol{J}_{\boldsymbol{\theta},1}|$, we obtain a uniform bound on $|\boldsymbol{J}_{\boldsymbol{\theta},1}|$ as

$$\sup_{\boldsymbol{\theta}\in\Sigma_{s,R_0}} |\boldsymbol{J}_{\boldsymbol{\theta},1}| \le \frac{4\kappa_1 n\rho^2}{\eta^2}. \tag{58}$$

Next, we define the index vector $\boldsymbol{J}_{\boldsymbol{\theta},2}$ for $\boldsymbol{\theta} \in \mathcal{G}$: first let $\mathscr{E}_{\boldsymbol{\theta},i} = \{\mathcal{Q}_\Delta(\widetilde{y}_{\boldsymbol{\theta},i} + \tau_i + t) \text{ is discontinuous in } t \in [-\eta,\eta]\}$, and then we define $\boldsymbol{J}_{\boldsymbol{\theta},2} := [\mathbb{1}(\mathscr{E}_{\boldsymbol{\theta},i})]$. Then by Lemma 11 that we prove later, we have

$$\begin{aligned} &\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\mathcal{G}} |\boldsymbol{J}_{\boldsymbol{\theta},2}| \le \frac{Cn\eta}{\Delta}\Big) \\ &\qquad \ge 1 - \exp\Big(-\frac{cn\eta}{\Delta} + s\log\frac{9d}{\rho s}\Big) := 1 - \mathscr{P}_1. \end{aligned} \tag{59}$$

Note that $\mathscr{E}_{\boldsymbol{\theta},i}$ does not happen (i.e., $\mathbb{1}(\mathscr{E}_{\boldsymbol{\theta},i}) = 0$) means that $\mathcal{Q}_\Delta(\widetilde{y}_{\boldsymbol{\theta},i} + \tau_i + t)$ is continuous in $t \in [-\eta,\eta]$; combined with the definition of $\mathcal{Q}_\Delta(\cdot)$, this is also equivalent to the statement that "$\mathcal{Q}_\Delta(\widetilde{y}_{\boldsymbol{\theta},i} + \tau_i + t)$ remains constant in $t \in [-\eta,\eta]$." Thus, given a fixed $\boldsymbol{\theta} \in \Sigma_{s,R_0}$ and its associated $\boldsymbol{\theta}'$, suppose that the i-th entry of $\boldsymbol{J}_{\boldsymbol{\theta}',2}$ is zero (meaning that "$\mathcal{Q}_\Delta(\widetilde{y}_{\boldsymbol{\theta}',i} + \tau_i + t)$ remains constant in $t \in [-\eta,\eta]$"), if additionally i-th entry of $\boldsymbol{J}_{\boldsymbol{\theta},1}$ is zero (i.e., $|\widetilde{\Psi}_{\boldsymbol{\theta},i}| < \eta$), then the i-th entry of $\widehat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}} = \mathcal{Q}_\Delta(\widetilde{\boldsymbol{y}}_{\boldsymbol{\theta}} + \boldsymbol{\tau}) - \mathcal{Q}_\Delta(\widetilde{\boldsymbol{y}}_{\boldsymbol{\theta}'} + \boldsymbol{\tau})$ vanishes:

$$\begin{aligned} \widehat{\Psi}_{\boldsymbol{\theta},i} &= \mathcal{Q}_\Delta(\widetilde{y}_{\boldsymbol{\theta},i} + \tau_i) - \mathcal{Q}_\Delta(\widetilde{y}_{\boldsymbol{\theta}',i} + \tau_i) \\ &= \mathcal{Q}_\Delta(\widetilde{y}_{\boldsymbol{\theta}',i} + \widetilde{\Psi}_{\boldsymbol{\theta},i} + \tau_i) - \mathcal{Q}_\Delta(\widetilde{y}_{\boldsymbol{\theta}',i} + \tau_i) = 0; \end{aligned}$$

combining with (54), this implies $\Psi_{\boldsymbol{\theta},i} = -\widetilde{\Psi}_{\boldsymbol{\theta},i}$. Recall from (56) that we want to bound $\sup_{\boldsymbol{\theta}} \|\boldsymbol{\Psi}_{\boldsymbol{\theta}}\|_2$. Write $\boldsymbol{J}^c_{\boldsymbol{\theta},1} = \mathbf{1} - \boldsymbol{J}_{\boldsymbol{\theta},1}$ and $\boldsymbol{J}^c_{\boldsymbol{\theta}',2} = \mathbf{1} - \boldsymbol{J}_{\boldsymbol{\theta}',2}$, then denoting hadamard product by $\odot$ and using the decomposition $\mathbf{1} = \max\{\boldsymbol{J}_{\boldsymbol{\theta},1}, \boldsymbol{J}_{\boldsymbol{\theta}',2}\} + \min\{\boldsymbol{J}^c_{\boldsymbol{\theta},1}, \boldsymbol{J}^c_{\boldsymbol{\theta}',2}\}$ and (54) we have the display (60), where (i) is because entries of $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$ equal to those of $-\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\theta}}$ if the index corresponds to $\min\{\boldsymbol{J}^c_{\boldsymbol{\theta},1}, \boldsymbol{J}^c_{\boldsymbol{\theta}',2}\} = 1$, and in (ii) we use the simple bound $\|\boldsymbol{\Psi}_{\boldsymbol{\theta}}\|_\infty \le \|\boldsymbol{\xi}_{\boldsymbol{\theta}}\|_\infty + \|\boldsymbol{\xi}_{\boldsymbol{\theta}'}\|_\infty \le 2\Delta$ and the derived bounds on $\sup_{\boldsymbol{\theta}} |\boldsymbol{J}_{\boldsymbol{\theta},1}|, \sup_{\boldsymbol{\theta}\in\mathcal{G}} |\boldsymbol{J}_{\boldsymbol{\theta},2}|, \sup_{\boldsymbol{\theta}} \|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\theta}}\|_2$ in (58), (59) and (57), respectively.

Step 2.4.3. Concluding the Bound on I_0

We are ready to put pieces together, specify ρ, η, and conclude the bound on I_0. Overall, with probability at least $1 - \mathscr{P}_1 - \mathscr{P}_2$ for $\mathscr{P}_1$ defined in (59) and some $\mathscr{P}_2$ within the promised probability, combining (52), (53), (56) and (60) we obtain

$$I_0 \lesssim \sigma\Delta\sqrt{ns\delta\log\frac{9d}{\rho s}} + \frac{\kappa_1 n\Delta\rho}{\eta} + n\sqrt{\kappa_1\eta\Delta} + \kappa_1\rho n.$$

Thus, we take the (near-optimal) choice of (ρ, η) as $\rho \asymp \frac{\Delta}{\sqrt{\kappa_1}}\big(\frac{s\delta}{n}\big)^{3/2}$ and $\eta \asymp \frac{\delta\Delta s}{n}\log\frac{9d}{\rho s}$, under which we obtain that, with the promised probability (as $\mathscr{P}_1$ is also sufficiently small), we obtain the bound on I_0 as

$$I_0 \lesssim \sigma\Delta\sqrt{ns\delta\log\Big(\frac{\kappa_1 d^2 n^3}{\Delta^2 s^5\delta^3}\Big)} \tag{61}$$

We can conclude the proof with all the works above. Substituting $\inf_{\boldsymbol{v}\in\mathscr{V}} \sum_{k=1}^n (\boldsymbol{x}_k^\top \boldsymbol{v})^2 \ge \frac{1}{4}\kappa_0 n$ and the definition of I in (46) into (45), then we obtain $\frac{1}{4}\kappa_0 n\|\widehat{\boldsymbol{\Upsilon}}\|_2^2 \le 2\|\widehat{\boldsymbol{\Upsilon}}\|_2 I$ that holds uniformly for all $\boldsymbol{\theta} \in \Sigma_{s,R_0}$, which implies $\sup_{\boldsymbol{\theta}} \|\widehat{\boldsymbol{\Upsilon}}\|_2 \le \frac{8I}{\kappa_0 n}$. Substituting the derived bounds on I_1, I_2, I_3, I_0 into (47), with the promised probability we have

$$I \lesssim \sigma\big(\sigma + M^{\frac{1}{2l}}\big)\sqrt{ns\delta\log d} + \sigma\Delta\sqrt{ns\delta\log\Big(\frac{\kappa_1 d^2 n^3}{\Delta^2 s^5\delta^3}\Big)},$$

$$\begin{aligned}\sup_{\boldsymbol{\theta}}\big\|\boldsymbol{\Psi}_{\boldsymbol{\theta}}\big\|_2 &= \sup_{\boldsymbol{\theta}}\big\|\boldsymbol{\Psi}_{\boldsymbol{\theta}}\odot\max\{\boldsymbol{J}_{\boldsymbol{\theta},1},\boldsymbol{J}_{\boldsymbol{\theta}',2}\}+\boldsymbol{\Psi}_{\boldsymbol{\theta}}\odot\min\{\boldsymbol{J}^c_{\boldsymbol{\theta},1},\boldsymbol{J}^c_{\boldsymbol{\theta}',2}\}\big\|_2\\ &\le \sup_{\boldsymbol{\theta}}\big\|\boldsymbol{\Psi}_{\boldsymbol{\theta}}\odot\max\{\boldsymbol{J}_{\boldsymbol{\theta},1},\boldsymbol{J}_{\boldsymbol{\theta}',2}\}\big\|_2+\sup_{\boldsymbol{\theta}}\big\|\boldsymbol{\Psi}_{\boldsymbol{\theta}}\odot\min\{\boldsymbol{J}^c_{\boldsymbol{\theta},1},\boldsymbol{J}^c_{\boldsymbol{\theta}',2}\}\big\|_2\\ &\overset{(i)}{\le} \sup_{\boldsymbol{\theta}}\big\|\boldsymbol{\Psi}_{\boldsymbol{\theta}}\big\|_\infty\cdot\sqrt{|\boldsymbol{J}_{\boldsymbol{\theta},1}|+|\boldsymbol{J}_{\boldsymbol{\theta}',2}|}\ +\sup_{\boldsymbol{\theta}}\big\|\widetilde{\boldsymbol{\Psi}}_{\boldsymbol{\theta}}\odot\min\{\boldsymbol{J}^c_{\boldsymbol{\theta},1},\boldsymbol{J}^c_{\boldsymbol{\theta}',2}\}\big\|_2\\ &\overset{(ii)}{\le} C\Big\{\Delta\sqrt{n}\Big(\frac{\sqrt{\kappa_1}\rho}{\eta}+\sqrt{\frac{\eta}{\Delta}}\Big)+\rho\sqrt{\kappa_1 n}\Big\},\end{aligned}\tag{60}$$

so the uniform bound on $\|\widehat{\boldsymbol{\Upsilon}}\|_2$ follows immediately. Further using $\|\widehat{\boldsymbol{\Upsilon}}\|_1\le 2\sqrt{s}\|\widehat{\boldsymbol{\Upsilon}}\|_2$ completes the proof. □

Lemma 11. (Bounding $\sup_{\boldsymbol{\theta}\in\mathcal{G}}|\boldsymbol{J}_{\boldsymbol{\theta},2}|$)**.** *Along the proof of Theorem 13, it holds that*

$$\mathbb{P}\Big(\sup_{\boldsymbol{\theta}\in\mathcal{G}}|\boldsymbol{J}_{\boldsymbol{\theta},2}|\ge\frac{Cn\eta}{\Delta}\Big)\le\exp\Big(-\frac{cn\eta}{\Delta}+s\log\frac{9d}{\rho s}\Big).\tag{62}$$

Proof. Notation and details in the proof of Theorem 13 will be used. We first consider a fixed $\boldsymbol{\theta}\in\mathcal{G}$, and by a simple shifting $\mathscr{E}_{\boldsymbol{\theta},i}$ happens if and only if $\mathcal{Q}_\Delta(\cdot)$ is discontinuous in $[\widetilde{y}_{\boldsymbol{\theta},i}+\tau_i-\eta,\widetilde{y}_{\boldsymbol{\theta},i}+\tau_i+\eta]$, which is also equivalently to

$$[\widetilde{y}_{\boldsymbol{\theta},i}+\tau_i-\eta,\widetilde{y}_{\boldsymbol{\theta},i}+\tau_i+\eta]\cap(\Delta\cdot\mathbb{Z})=\varnothing.$$

Because $\tau_i\sim\mathscr{U}\big([-\frac{\Delta}{2},\frac{\Delta}{2}]\big)$ and $\eta<\frac{\Delta}{2}$, $\mathbb{P}(\mathscr{E}_{\boldsymbol{\theta},i})=\frac{2\eta}{\Delta}$ is valid independent of the location of $[\widetilde{y}_{\boldsymbol{\theta},i}-\eta,\widetilde{y}_{\boldsymbol{\theta},i}+\eta]$. Thus, for fixed $\boldsymbol{\theta}$, by conditioning on $(\boldsymbol{X},\boldsymbol{\epsilon})$, $|\boldsymbol{J}_{\boldsymbol{\theta},2}|$ follows the binomial distribution with n trials and probability of success $p:=\frac{2\eta}{\Delta}$. This allows us to write $|\boldsymbol{J}_{\boldsymbol{\theta},2}|=\sum_{k=1}^n J_k$ with J_k i.i.d. following Bernoulli distribution with success probability $\mathbb{E}J_k=p$. Then for any integer $q\ge 2$ we have

$$\begin{aligned}\sum_{k=1}^n\mathbb{E}|J_k-\mathbb{E}J_k|^q&\le\sum_{k=1}^n\mathbb{E}|J_k-\mathbb{E}J_k|^2\\ &\le np(1-p)\le\frac{q!}{2}np.\end{aligned}$$

Now we invoke Bernstein's inequality (Lemma 1) to obtain that for any $t>0$,

$$\mathbb{P}\big(|\boldsymbol{J}_{\boldsymbol{\theta},2}|-np\ge\sqrt{2npt}+t\big)\le\exp(-t).$$

We let $t=cnp$ and take a union bound over $\boldsymbol{\theta}\in\mathcal{G}$; this yields the desired claim since $|\mathcal{G}|\le\big(\frac{9d}{\rho s}\big)^s$. □

Junren Chen received the B.Sc. degree in Mathematics and Applied Mathematics from Sun Yat-sen University. He is currently pursuing the Ph.D. degree with the Department of Mathematics, The University of Hong Kong. His research interests include compressed sensing, high-dimensional statistics, signal and image processing, quantization, and optimization. He received the Hong Kong Ph.D. Fellowship from the Hong Kong Research Grants Council for supporting his Ph.D. study.

Michael K. Ng (Senior Member, IEEE) received the B.Sc. and M.Phil. degrees from The University of Hong Kong, Hong Kong, in 1990 and 1992, respectively, and the Ph.D. degree from The Chinese University of Hong Kong, Hong Kong, in 1995. From 1995 to 1997, he was a Research Fellow with the Computer Sciences Laboratory, The Australian National University, Canberra, ACT, Australia. He was an Assistant Professor/Associate Professor with The University of Hong Kong from 1997 to 2005. He was a Professor/Chair Professor (2005-2019) with the Department of Mathematics, Hong Kong Baptist University, Hong Kong, Chair Professor (2019-2023) with the Department of Mathematics, The University of Hong. He is currently a Chair Professor in Mathematics and Chair Professor in Data Science at Hong Kong Baptist University. His research interests include applied and computational mathematics, machine learning and artificial intelligence, and data science. Dr. Ng serves as an editorial board member of several international journals. He was selected for the 2017 Class of Fellows of the Society for Industrial and Applied Mathematics. He received the Feng Kang Prize for his significant contributions to scientific computing.

Di Wang (Member, IEEE) received the Ph.D. degree in computer science from The State University of New York (SUNY) at Buffalo. He is currently an Assistant Professor of computer science and a Faculty Member of statistics with the Division of Computer, Electrical and Mathematical Sciences and Engineering (CEMSE), King Abdullah University of Science and Technology (KAUST). He is also the Principal Investigator (PI) of the Provable Responsible AI and Data Analytics (PRADA) Laboratory, a member of the Computational Bioscience Research Center (CBRC), and an affiliated Faculty Member with the SDAIA-KAUST Center of Excellence in Data Science and Artificial Intelligence (SDAIA-KAUST AI). His research interests include trustworthy machine learning, machine learning theory, and AI for science.