Journal of Machine Learning Research 16 (2015) 1757-1792 Submitted 7/15; Published 9/15

Optimal Estimation of Low Rank Density Matrices

Vladimir Koltchinskii[*] VLAD@MATH.GATECH.EDU
Dong Xia [†] DXIA7@MATH.GATECH.EDU
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332, USA.

Editor: Alex Gammerman and Vladimir Vovk

Abstract

The density matrices are positively semi-definite Hermitian matrices of unit trace that describe the state of a quantum system. The goal of the paper is to develop minimax lower bounds on error rates of estimation of low rank density matrices in trace regression models used in quantum state tomography (in particular, in the case of Pauli measurements) with explicit dependence of the bounds on the rank and other complexity parameters. Such bounds are established for several statistically relevant distances, including quantum versions of Kullback-Leibler divergence (relative entropy distance) and of Hellinger distance (so called Bures distance), and Schatten p-norm distances. Sharp upper bounds and oracle inequalities for least squares estimator with von Neumann entropy penalization are obtained showing that minimax lower bounds are attained (up to logarithmic factors) for these distances.

Keywords: quantum state tomography, low rank density matrix, minimax lower bounds

1. Introduction

This paper deals with optimality properties of estimators of density matrices, describing states of quantum systems, that are based on penalized empirical risk minimization with specially designed complexity penalties such as von Neumann entropy of the state. Alexey Chervonenkis was a co-founder of the theory of empirical risk minimization that is of crucial importance in machine learning, but he also had very broad interests that included, in particular, quantum mechanics. By the choice of the topic, we would like to honor the memory of this great man and great scientist.

Let $\mathbb{M}_m(\mathbb{C})$ be the set of all $m\times m$ matrices with complex entries and let $\mathbb{H}_m = \mathbb{H}_m(\mathbb{C}) \subset \mathbb{M}_m(\mathbb{C})$ be the set of all Hermitian matrices: $\mathbb{H}_m = \{A \in \mathbb{M}_m(\mathbb{C}) : A = A^*\}$, A^* denoting the adjoint matrix of A. For $A \in \mathbb{H}_m$, $\mathrm{tr}(A)$ denotes the trace of A and $A \succcurlyeq 0$ means that A is positively semi-definite. Let $\mathcal{S}_m := \{S \in \mathbb{H}_m : S \succcurlyeq 0, \mathrm{tr}(S) = 1\}$ be the set of all positively semi-definite Hermitian matrices of unit trace called *density matrices*. In quantum mechanics, the state of a quantum system is usually characterized by a density matrix $\rho \in \mathcal{S}_m$ (or, more generally, by a self-adjoint positively semi-definite operator of unit trace acting in an infinite-dimensional Hilbert space, called a density operator). Often, very

[*]. Supported in part by NSF Grants DMS-1509739, DMS-1207808, CCF-1523768 and CCF-1415498
[†]. Supported in part by NSF Grant DMS-1207808

large density matrices are needed to represent or to approximate the density operator of the state. For instance, for a quantum system consisting of b qubits, the density matrices are of the size $m \times m$ with $m = 2^b$, so the dimension of the density matrix grows exponentially with b. For instance, for a 10 qubit system, one has to deal with matrices that have 2^{20} entries. Thus, it becomes natural in the problems of statistical estimation of density matrix ρ to take an advantage of the fact that it might be low rank, or nearly low rank (that is, it could be well approximated by low rank matrices) which reduces the complexity of the estimation problem.

In *quantum state tomography (QST)*, the goal is to estimate an unknown state $\rho \in \mathcal{S}_m$ based on a number of specially designed measurements for the system prepared in state ρ (see Gross et al. 2010, Gross 2011, Koltchinskii 2011a, Cai et al. 2015 and references therein). Given an observable $A \in \mathbb{H}_m$ with spectral representation $A = \sum_{j=1}^{m'} \lambda_j P_j$, where $m' \leq m$, λ_j being the eigenvalues of A and P_j being the corresponding mutually orthogonal eigenprojectors, the outcome of a measurement of A for the system prepared in state ρ is a random variable Y taking values λ_j with probabilities $\mathrm{tr}(\rho P_j)$. The expectation of Y is then $\mathbb{E}_\rho Y = \mathrm{tr}(\rho A)$, so, Y could be viewed as a noisy observation of the value of linear functional $\mathrm{tr}(\rho A)$ of the unknown density matrix ρ. A common approach is to choose an observable A at random, assuming that it is the value of a random variable X with some design distribution Π in the space $\mathbb{H}_m$. More precisely, given a sample of n i.i.d. copies $X_1, \ldots, X_n$ of X, n measurements are being performed for the system identically prepared n times in state ρ resulting in outcomes $Y_1, \ldots, Y_n$. Based on the data $(X_1, Y_1), \ldots, (X_n, Y_n)$, the goal is to estimate the target density matrix ρ. Clearly, the observations satisfy the following model

$$Y_j = \mathrm{tr}(\rho X_j) + \xi_j, \ j = 1, \ldots, n, \tag{1}$$

where $\{\xi_j\}$ is a random noise consisting of n i.i.d. random variables satisfying the condition $\mathbb{E}_\rho(\xi_j | X_j) = 0, j = 1, \ldots, n$. This is a special case of so called *trace regression model* intensively studied in the recent literature (see, e.g., Koltchinskii et al. 2011, Koltchinskii 2011b and references therein).

1.1 Assumptions

A common choice of design distribution in this type of problems is so called *uniform sampling from an orthonormal basis* described in the following assumptions.

Assumption 1 *Let $\mathcal{E} = \{E_1, \ldots, E_{m^2}\} \subset \mathbb{H}_m$ be an orthonormal basis of $\mathbb{H}_m$ with respect to the Hilbert–Schmidt inner product: $\langle A, B \rangle = tr(AB)$. Moreover, suppose that, for some $U > 0$,*

$$\|E_j\|_\infty \leq U, j = 1, \ldots, n,$$

where $\| \cdot \|_\infty$ denotes the operator norm (the spectral norm).

Since $\|E_j\|_2 = 1$, where $\| \cdot \|_2$ denotes the Hilbert–Schmidt (or Frobenius) norm, we can assume that $U \leq 1$. Moreover, $U \geq m^{-1/2}$ since $1 = \|E_j\|_2 \leq m^{1/2} \|E_j\|_\infty \leq m^{1/2} U$.

Assumption 2 *Let Π be the uniform distribution in the finite set $\mathcal{E}$ (see Assumption 1), let X be a random variable sampled from Π and let $X_1, \ldots, X_n$ be i.i.d. copies of X.*

It will be assumed in what follows that assumptions 1 and 2 hold (unless it is stated otherwise). Under these assumptions, $Y_1, \dots, Y_n$ could be viewed as noisy observations of a random sample of Fourier coefficients $\langle \rho, X_1 \rangle, \dots, \langle \rho, X_n \rangle$ of the target density matrix ρ in the basis $\mathcal{E}$. The above model (in which $X_1, \dots, X_n$ are uniformly sampled from an orthonormal basis and $Y_1, \dots, Y_n$ are the outcomes of measurements of the observables $X_1, \dots, X_n$ for the system being identically prepared n times in the same state ρ) will be called in what follows the *standard QST model.* It is a special case of *trace regression model with bounded response*:

Assumption 3 (Trace regression with bounded responce) *Suppose that Assumption 1 holds and let (X, Y) be a random couple such that X is sampled from the uniform distribution Π in an orthonormal basis $\mathcal{E} \subset \mathbb{H}_m$. Suppose also that, for some $\rho \in \mathcal{S}_m$, $\mathbb{E}(Y|X) = \langle \rho, X \rangle$ a.s. and, for some $\bar{U} > 0$, $|Y| \le \bar{U}$ a.s.. The data $(X_1, Y_1), \dots (X_n, Y_n)$ consists of n i.i.d. copies of (X, Y).*

We are also interested in the *trace regression model with Gaussian noise*:

Assumption 4 (Trace regression with Gaussian noise) *Suppose Assumption 1 holds and let (X, Y) be a random couple such that X is sampled from the uniform distribution Π in an orthonormal basis $\mathcal{E} \subset \mathbb{H}_m$ and, for some $\rho \in \mathcal{S}_m$, $Y = \langle \rho, X \rangle + \xi$, where ξ is a normal random variable with mean 0 and variance σ_ξ^2, ξ and X being independent. The data $(X_1, Y_1), \dots (X_n, Y_n)$ consists of n i.i.d. copies of (X, Y).*

Note that this model is not directly applicable to the “standard QST problem” described above, where the response variable Y is discrete. However, if the measurements are repeated multiple times for each observable X_j and the resulting outcomes are averaged to reduce the variance, the noise of such averaged measurements becomes approximately Gaussian and it is of interest to characterize the estimation error in terms of the variance of the noise.

An important example of an orthonormal basis used in quantum state tomography is so called *Pauli basis*, see, e.g., Gross et al. (2010), Gross (2011). The Pauli basis in the space $\mathbb{H}_2$ of 2×2 Hermitian matrices (observables in a single qubit system) consists of four matrices W_1, W_2, W_3, W_4 defined as $W_i = \frac{1}{\sqrt{2}}\sigma_i,\ i = 1, \dots, 4$, where

$$\sigma_1 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_2 := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_4 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

It is easy to check that $\{W_0, W_1, W_2, W_3\}$ indeed forms an orthonormal basis in $\mathbb{H}_2$. The Pauli basis in the space $\mathbb{H}_m$ for $m = 2^b$ (the space of observables for a b qubits system) is defined by tensorisation, namely, it consists of 4^b tensor products $W_{i_1} \otimes \dots \otimes W_{i_b}, (i_1, \dots, i_b) \in \{1, 2, 3, 4\}^b$. Let us write these matrices as $E_1, \dots, E_{m^2}$ with $E_1 = W_1 \otimes \dots \otimes W_1$. It is easy to see that each of them has eigenvalues $\pm \frac{1}{\sqrt{m}}$ and $\|E_j\|_\infty = m^{-1/2}$, so, for this basis, $U = m^{-1/2}$. The fact that, for the Pauli basis, the operator norms of basis matrices are as small as possible plays an important role in quantum state tomography (Gross et al., 2010; Gross, 2011; Liu, 2011). Let $E_j = \frac{1}{\sqrt{m}} Q_j^+ - \frac{1}{\sqrt{m}} Q_j^-$ be the spectral representation of E_j. Then, an outcome of a measurement of E_j in state ρ is a random variable τ_j taking values

$\pm\frac{1}{\sqrt{m}}$ with probabilities $\langle \rho, Q_t^{\pm}\rangle$. Its expectation is $\mathbb{E}_\rho \tau_j = \langle \rho, E_j\rangle$. Of course, there exists a unique representation of density matrix ρ in the Pauli basis that can be written as follows: $\rho = \sum_{j=1}^{m^2} \frac{\alpha_j}{\sqrt{m}} E_j$ with $\alpha_1 = 1$. Then, we clearly have $\mathbb{E}_\rho \tau_j = \frac{\alpha_j}{\sqrt{m}}$ and $\mathbb{P}_\rho\Big\{\tau_j = \pm\frac{1}{\sqrt{m}}\Big\} = \frac{1\pm\alpha_j}{2}$ (for $j = 1$, this gives $\mathbb{P}_\rho\Big\{\tau_1 = \frac{1}{\sqrt{m}}\Big\} = 1$). As a consequence, $\mathrm{Var}_\rho(\tau_j) = \frac{1-\alpha_j^2}{m}$. Note that $\sum_{j=1}^{m^2} \frac{\alpha_j^2}{m} = \|\rho\|_2^2 \le \mathrm{tr}^2(\rho) = 1$. This implies that there exists j such that $\alpha_j^2 \le \frac{1}{2}$ and $\mathrm{Var}_\rho(\tau_j) \ge \frac{1}{2m}$. In fact, the number of such j must be large, say, at least $\frac{m^2}{2}$ (provided that $m > 4$). Thus, for "most" of the values of j, $\mathrm{Var}_\rho(\tau_j) \asymp \frac{1}{m}$. A way to reduce the variance is to repeat the measurement of each observable X_j K times (for a system identically prepared in state ρ) and to average the outcomes of such K measurements. The resulting response variable is $Y_j = \langle \rho, X_j\rangle + \xi_j$, where $\mathbb{E}_\rho(\xi_j|X_j) = 0$ and $\mathbb{E}_\rho(\xi_j^2|X_j) = \mathrm{Var}_\rho(Y_j|X_j) = \frac{1-\alpha_{\nu_j}^2}{Km}$, ν_j being defined by the relationship $X_j = E_{\nu_j}$.

1.2 Preliminaries and Notations

Some notations will be used throughout the paper. The Euclidean norm in $\mathbb{C}^m$ will be denoted by $\|\cdot\|$ and the notation $\langle\cdot,\cdot\rangle$ will be used for both the Euclidean inner product in $\mathbb{C}^m$ and for the Hilbert–Schmidt inner product in $\mathbb{H}_m$. $\|\cdot\|_p, p \ge 1$ will be used to denote the *Schatten* p-norm in $\mathbb{H}_m$, namely $\|A\|_p^p = \sum_j^m |\lambda_j(A)|^p$, $A \in \mathbb{H}_m$, $\lambda_1(A) \ge \ldots \ge \lambda_m(A)$ being the eigenvalues of A. In particular, $\|\cdot\|_2$ denotes the Hilbert–Schmidt (or Frobenius) norm, $\|\cdot\|_1$ denotes the nuclear (or trace) norm and $\|\cdot\|_\infty$ denotes the operator (or spectral) norm: $\|A\|_\infty = \max_{1\le j\le m} |\lambda_j(A)| = |\lambda_1(A)|$. The following well known *interpolation inequality* for Schatten p-norms will be used to extend the bounds proved for some values of p to the whole range of its values. It easily follows from similar bounds for ℓ_p-spaces.

Lemma 1 (Interpolation inequality) *For $1 \le p < q < r \le \infty$, and let $\mu \in [0,1]$ be such that*

$$\frac{\mu}{p} + \frac{1-\mu}{r} = \frac{1}{q}.$$

Then, for all $A \in \mathbb{H}_m$,

$$\|A\|_q \le \|A\|_p^{\mu}\|A\|_r^{1-\mu}.$$

Given $A \in \mathbb{H}_m$, define a function $f_A : \mathbb{H}_m \mapsto \mathbb{R} : f_A(x) := \langle A, x\rangle, x \in \mathbb{H}_m$. For a given random variable X in $\mathbb{H}_m$ with a distribution Π, we have $\|f_A\|_{L_2(\Pi)}^2 = \mathbb{E} f_A^2(X) = \mathbb{E}\langle A, X\rangle^2$. Sometimes, with a minor abuse of notation, we might write $\|A\|_{L_2(\Pi)}^2 = \int_{\mathbb{H}_m} \langle A, x\rangle^2 \Pi(dx) = \|f_A\|_{L_2(\Pi)}^2$. In what follows, Π will be typically the uniform distribution in an orthonormal basis $\mathcal{E} = \{E_1, \ldots, E_{m^2}\} \subset \mathbb{H}_m$, implying that

$$\|f_A\|_{L_2(\Pi)}^2 = \|A\|_{L_2(\Pi)}^2 = m^{-2}\|A\|_2^2,$$

so, the $L_2(\Pi)$-norm is just a rescaled Hilbert–Schmidt norm.

Consider $A \in \mathbb{H}_m$ with spectral representation $A = \sum_{j=1}^{m'} \lambda_j P_j$, $m' \le m$ with distinct non-zero eigenvalues λ_j. Denote by $\mathrm{sign}(A) := \sum_{j=1}^{m'} \mathrm{sign}(\lambda_j) P_j$ and by $\mathrm{supp}(A)$ the linear span of the images of projectors $P_j, j = 1, \dots, m'$ (the subspace $\mathrm{supp}(A) \subset \mathbb{C}^m$ will be called *the support* of A).

Given a subspace $L \subset \mathbb{C}^m$, $L^\perp$ denotes the orthogonal complement of L and P_L denotes the orthogonal projection onto L. Let $\mathcal{P}_L, \mathcal{P}_L^\perp$ be orthogonal projection operators in the space $\mathbb{H}_m$ (equipped with the Hilbert–Schmidt inner product), defined as follows:

$$\mathcal{P}_L^\perp(A) = P_{L^\perp} A P_{L^\perp}, \ \ \mathcal{P}_L(A) = A - P_{L^\perp} A P_{L^\perp}.$$

These two operators split any Hermitian matrix A into two orthogonal parts, $\mathcal{P}_L(A)$ and $\mathcal{P}_L^\perp(A)$, the first one being of rank at most $2\mathrm{dim}(L)$.

For a convex function $f : \mathbb{H}_m \mapsto \mathbb{R}$, $\partial f(A)$ denotes the subdifferential of f at the point $A \in \mathbb{H}_m$. It is well known that

$$\partial \|A\|_1 = \Big\{ \mathrm{sign}(A) + \mathcal{P}_L^\perp(M) : M \in \mathbb{H}_m, \|M\|_\infty \le 1 \Big\}, \tag{2}$$

where $L = \mathrm{supp}(A)$ (see Koltchinskii 2011b, p. 240 and references therein).

C, C_1, C', c, c', etc will denote constants (that do not depend on parameters of interest such as m, n, etc) whose values could change from line to line (or, even, within the same line) without further notice. For nonnegative A and B, $A \lesssim B$ (equivalently, $B \gtrsim A$) means that $A \le CB$ for some absolute constant $C > 0$, and $A \asymp B$ means that $A \lesssim B$ and $B \lesssim A$. Sometimes, symbols $\lesssim, \gtrsim$ and $\asymp$ could be provided with subscripts (say, $A \lesssim_\gamma B$) to indicate that constant C may depend on a parameter (say, γ).

In what follows, P denotes the distribution of (X, Y) and P_n denotes the corresponding empirical distribution based on the sample $(X_1, Y_1), \dots, (X_n, Y_n)$ of n i.i.d. observations. Similarly, Π is the distribution of X (typically, uniform in an orthonormal basis) and Π_n is the corresponding empirical distribution based on the sample $(X_1, \dots, X_n)$. We will use standard notations $Pf = \mathbb{E} f(X, Y), P_n f = n^{-1} \sum_{j=1}^n f(X_j, Y_j)$ and $\Pi g = \mathbb{E} g(X), P_n g = n^{-1} \sum_{j=1}^n g(X_j)$.

1.3 Estimation Methods

Recall that the central problem in quantum state tomography is to estimate a large density matrix ρ based on the data $(X_1, Y_1), \dots, (X_n, Y_n)$ satisfying the trace regression model. Often, the goal is to develop adaptive estimators with optimal dependence of the estimation error (measured by various statistically relevant distances) on the unknown rank of the target matrix ρ under the assumption that ρ is low rank, or on other complexity parameters in the case when the target matrix ρ can be well approximated by low rank matrices.

The simplest estimation procedure for density matrix ρ is the least squares estimator defined by the following convex optimization problem:

$$\hat{\rho} := \operatorname*{arg\,min}_{S \in \mathcal{S}_m} \frac{1}{n} \sum_{j=1}^n \left(Y_j - \langle S, X_j \rangle \right)^2. \tag{3}$$

Since, for all $S \in \mathcal{S}_m$, $\|S\|_1 = \mathrm{tr}(S) = 1$, we have that

$$\hat{\rho} = \hat{\rho}^\varepsilon := \operatorname*{arg\,min}_{S \in \mathcal{S}_m} \left[\frac{1}{n} \sum_{j=1}^n \left(Y_j - \langle S, X_j \rangle \right)^2 + \varepsilon \|S\|_1 \right], \ \ \varepsilon \ge 0. \tag{4}$$

Thus, in the case of density matrices, the least squares estimator $\hat{\rho}$ coincides with the *matrix LASSO* estimator $\hat{\rho}^{\varepsilon}$ with nuclear norm penalty and arbitrary value of regularization parameter ε. The nuclear norm penalty is used as a proxy of the rank that provides a convex relaxation for rank penalized least squares method. Matrix LASSO is a standard method of low rank estimation in trace regression models that has been intensively studied in the recent years, see, for instance, Candés and Plan (2011), Rohde and Tsybakov (2011), Koltchinskii (2011b), Koltchinskii et al. (2011), Negahban and Wainwright (2010) and references therein. In the case of estimation of density matrices, due to their positive semidefiniteness and trace constraint, the nuclear norm penalization is present implicitly even in the case of a non-penalized least squares estimator $\hat{\rho}$ (see also Koltchinskii 2013a, Kalev et al. 2015 where similar ideas were used).

Note that the estimator $\hat{\rho}$ can be also rewritten as

$$\hat{\rho} := \operatorname*{arg\,min}_{S\in\mathcal{S}_m}\Big[\|S\|^2_{L_2(\Pi_n)} - \frac{2}{n}\sum_{j=1}^{n} Y_j\langle S, X_j\rangle\Big]. \tag{5}$$

Replacing the empirical $\|\cdot\|_{L_2(\Pi_n)}$-norm with the "true" $\|\cdot\|_{L_2(\Pi)}$-norm (which could make sense in the case when the design distribution Π is known) yields the following *modified least squares* estimator studied in Koltchinskii et al. (2011), Koltchinskii (2013a):

$$\check{\rho} := \operatorname*{arg\,min}_{S\in\mathcal{S}_m}\Big[\|S\|^2_{L_2(\Pi)} - \frac{2}{n}\sum_{j=1}^{n} Y_j\langle S, X_j\rangle\Big]. \tag{6}$$

Another estimator was proposed in Koltchinskii (2011a) and it is based on an idea of using so called *von Neumann entropy* as a penalizer in least squares method. Von Neumann entropy is a canonical extension of Shannon's entropy to the quantum setting. For a density matrix $S \in \mathcal{S}_m$, it is defined as $\mathcal{E}(S) := -\mathrm{tr}(S\log S)$. The estimator proposed in Koltchinskii (2011a) is defined as follows

$$\tilde{\rho}^{\varepsilon} := \operatorname*{arg\,min}_{S\in\mathcal{S}_m}\Big[\frac{1}{n}\sum_{j=1}^{n}(Y_j - \langle S, X_j\rangle)^2 + \varepsilon\mathrm{tr}(S\log S)\Big]. \tag{7}$$

Essentially, it is based on a trade-off between fitting the model via the least squares method in the class of all density matrices and maximizing the entropy of the quantum state. Note that (7) is also a convex optimization problem (due to concavity of von Neumann entropy, see Nielsen and Chuang 2000) and its solution $\tilde{\rho}^{\varepsilon}$ is a full rank matrix (see Koltchinskii 2011a, the proof of Proposition 3). It should be also mentioned that the idea of estimation of a density matrix of a quantum state by maximizing the von Neumann entropy subject to constraints based on the data has been used in quantum state tomography earlier (see Bužek 2004 and references therein).

1.4 Distances between Density Matrices

The main purpose of this paper is to study the optimality properties of estimator $\tilde{\rho}^{\epsilon}$ with respect to a variety of statistically meaningful distances, in the case when the underlying density matrix ρ is low rank. These distances include Schatten p-norm distances for $p \in$

$[1, 2]$,[1] but also quantum versions of Hellinger distance and Kullback-Leibler divergence that are of importance in quantum statistics and quantum information. A version of the (squared) Hellinger distance that will be studied is defined as

$$H^2(S_1, S_2) := 2 - 2\mathrm{tr}\sqrt{S_1^{\frac{1}{2}} S_2 S_1^{\frac{1}{2}}}$$

for $S_1, S_2 \in \mathcal{S}_m$ (see also Nielsen and Chuang 2000). Clearly, $0 \leq H^2(S_1, S_2) \leq 2$. In quantum information literature, it is usually called Bures distance and it does not coincide with $\mathrm{tr}(\sqrt{S_1} - \sqrt{S_2})^2$ (which is another possible non-commutative extension of the classical Hellinger distance). In fact, $H^2(S_1, S_2) \leq \mathrm{tr}(\sqrt{S_1} - \sqrt{S_2})^2, S_1, S_2 \in \mathcal{S}_m$, but the opposite inequality does not necessarily hold. The quantity $\mathrm{tr}\sqrt{S_1^{\frac{1}{2}} S_2 S_1^{\frac{1}{2}}}$ in the right hand side of the definition of H^2 is a quantum version of Hellinger affinity.

The noncommutative Kullback-Leibler divergence (or relative entropy distance) $K(\cdot\|\cdot)$ is defined as (see also Nielsen and Chuang 2000):

$$K(S_1\|S_2) := \langle S_1, \log S_1 - \log S_2 \rangle.$$

If $\log S_2$ is not well-defined (for instance, some of the eigenvalues of S_2 are equal to 0) we set $K(S_1\|S_2) = +\infty$. The symmetrized version of Kullback-Leibler divergence is defined as

$$K(S_1; S_2) := K(S_1\|S_2) + K(S_2\|S_1) = \langle S_1 - S_2, \log S_1 - \log S_2 \rangle.$$

The following very useful inequality is a noncommutative extension of similar classical inequalities for total variation, Hellinger and Kullback-Leibler distances. It follows from representing the "noncommutative distances" involved in the inequality as suprema of the corresponding classical distances between the distributions of outcomes of measurements for two states S_1, S_2 over all possible measurements represented by positive operator valued measures (see, Nielsen and Chuang 2000, Klauck et al. 2007, Koltchinskii 2011a, Section 3 and references therein).

Lemma 2 *For all $S_1, S_2 \in \mathcal{S}_m$, the following inequalities hold:*

$$\frac{1}{4}\|S_1 - S_2\|_1^2 \leq H^2(S_1, S_2) \leq (K(S_1\|S_2) \wedge \|S_1 - S_2\|_1). \tag{8}$$

1.5 Matrix Bernstein Inequalities

Non-commutative (matrix) versions of Bernstein inequality will be used in what follows. The most common version is stated (in a convenient form for our applications) in the following lemma.

Lemma 3 *Let $X, X_1, \ldots, X_n \in \mathbb{H}_m$ be i.i.d. random matrices with $\mathbb{E}X = 0$, $\sigma_X^2 := \|\mathbb{E}X^2\|_\infty$ and $\|X\|_\infty \leq U$ a.s. for some $U > 0$. Then, for all $t \geq 0$ with probability at least $1 - e^{-t}$,*

$$\Big\|\frac{1}{n}\sum_{j=1}^n X_j\Big\|_\infty \leq 2\Big[\sigma_X\sqrt{\frac{t + \log(2m)}{n}} \bigvee U\frac{t + \log(2m)}{n}\Big].$$

[1] Similar problems for estimators $\hat{\rho}, \check{\rho}$ and for Schatten p-norm distances with $p \in (2, +\infty]$ are studied in a related paper by Koltchinskii and Xia (2015+)

The proof of such bounds could be found, e.g., in Tropp (2012). Other versions on matrix Bernstein type inequalities for not necessarily bounded random matrices will be also used in what follows and they could be found in Koltchinskii (2011b), Koltchinskii (2013a). A simple consequence of the inequality of Lemma 3 is the following expectation bound:

$$\mathbb{E}\Big\|\frac{1}{n}\sum_{j=1}^{n} X_j\Big\|_{\infty} \lesssim \Big[\sigma_X\sqrt{\frac{\log(2m)}{n}}\bigvee U\frac{\log(2m)}{n}\Big].$$

It follows from the exponential bound by integrating the tail probabilities.

The paper is organized as follows. In Section 2, minimax lower bounds on estimation error of low rank density matrices are provided in Schatten p-norm, Hellinger (Bures) and Kullback-Leibler distances. In Section 3.1, sharp low rank oracle inequalities for von Neumann entropy penalized least squares estimator are derived in the case of trace regression model with bounded response. In Section 3.2, low rank oracle inequalities are established in the case of trace regression with Gaussian noise. In addition to this, in these two sections, upper bounds on estimation error with respect to Kullback-Leibler distance are obtained. In Section 3.3, they are further developed and extended to other distances (Hellinger distance, Schatten p-norm distances for $p \in [1,2]$) showing the minimax optimality (up to logarithmic factors) of the error rates of the least squares estimator with von Neumann entropy penalization.

2. Minimax Lower Bounds

In this section, we provide main results on the minimax lower bounds on the risk of estimation of density matrices with respect to Schatten p-norm (or, rather q-norm in the notations used below) distances as well as Hellinger-Bures distance and Kullback-Leibler divergence.

Minimax lower bounds will be derived for the class $\mathcal{S}_{r,m} := \{S \in \mathcal{S}_m : \mathrm{rank}(S) \le r\}$ consisting of all density matrices of rank at most r (the low rank case). We will start with the case of trace regression with Gaussian noise. Given that the sample $(X_1, Y_1), \dots, (X_n, Y_n)$ satisfies Assumption 4 with the target density matrix $\rho \in \mathcal{S}_m$ and noise variance σ_ξ^2, let $\mathbb{P}_\rho$ denote the corresponding probability distribution.

Note that Ma and Wu (2013) developed a method of deriving minimax lower bounds for distances based on unitary invariant norms, including Schatten p-norms in matrix problems, and obtained such lower bounds, in particular, in matrix completion problem. The approach used in our paper is somewhat different and the aim is to develop such bounds under an additional constraint that the target matrix is a density matrix. The resulting bounds are also somewhat different, they involve an additional term that does not depend on the rank, but does depend on q. Essentially, it means that the "complexity" of the problem is controlled by a "truncated rank" $r \wedge \frac{1}{\tau}$, where $\tau = \frac{\sigma_\xi m^{3/2}}{\sqrt{n}}$ rather than by the actual rank r. The upper bounds of Section 3.3 show that such a structure of the bound is, indeed, necessary. It should be also mentioned that minimax lower bounds on the nuclear norm error of estimation of density matrices have been obtained earlier in Flammia et al. (2012) (see Remark 11 below).

Theorem 4 *For all $q \in [1, +\infty]$, there exist constants $c, c' > 0$ such that, the following bounds hold:*

$$\inf_{\hat{\rho}} \sup_{\rho \in \mathcal{S}_{r,m}} \mathbb{P}_\rho \Big\{ \|\hat{\rho} - \rho\|_q \geq c \Big(\frac{\sigma_\xi m^{\frac{3}{2}} r^{1/q}}{\sqrt{n}} \bigwedge \Big(\frac{\sigma_\xi m^{3/2}}{\sqrt{n}} \Big)^{1-\frac{1}{q}} \bigwedge 1 \Big) \Big\} \geq c', \tag{9}$$

$$\inf_{\hat{\rho}} \sup_{\rho \in \mathcal{S}_{r,m}} \mathbb{P}_\rho \Big\{ H^2(\hat{\rho}, \rho) \geq c \Big(\frac{\sigma_\xi m^{\frac{3}{2}} r}{\sqrt{n}} \bigwedge 1 \Big) \Big\} \geq c', \tag{10}$$

and

$$\inf_{\hat{\rho}} \sup_{\rho \in \mathcal{S}_{r,m}} \mathbb{P}_\rho \Big\{ K(\rho \| \hat{\rho}) \geq c \Big(\frac{\sigma_\xi m^{\frac{3}{2}} r}{\sqrt{n}} \bigwedge 1 \Big) \Big\} \geq c', \tag{11}$$

where $\inf_{\hat{\rho}}$ denotes the infimum over all estimators $\hat{\rho}$ in $\mathcal{S}_m$ based on the data $(X_1, Y_1), \ldots, (X_n, Y_n)$ satisfying the Gaussian trace regression model with noise variance σ_ξ^2.

Proof A couple of preliminary facts will be needed in the proof. We start with bounds on the packing numbers of Grassmann manifold $\mathcal{G}_{k,l}$, which is the set of all k-dimensional subspaces L of the l-dimensional space $\mathbb{R}^l$. Given such a subspace $L \subset \mathbb{R}^l$ with $\dim(L) = k$, let P_L be the orthogonal projection onto L and let $\mathfrak{P}_{k,l} := \{P_L : L \in \mathcal{G}_{k,l}\}$. The set of all k-dimensional projectors $\mathfrak{P}_{k,l}$ will be equipped with Schatten q-norm distances for all $q \in [1, +\infty]$ (which also could be viewed as distances on the Grassmannian itself): $d_q(Q_1, Q_2) := \|Q_1 - Q_2\|_q, Q_1, Q_2 \in \mathfrak{P}_{k,l}$. Recall that the *$\varepsilon$-packing number* of a metric space (T, d) is defined as

$$D(T, d, \varepsilon) = \max \Big\{ n : \text{there are } t_1, \ldots, t_n \in T, \text{such that } \min_{i \neq j} d(t_i, t_j) > \varepsilon \Big\}.$$

The following lemma (see Pajor 1998, Proposition 8) will be used to control the packing numbers of $\mathfrak{P}_{k,l}$ with respect to Schatten distances d_q.

Lemma 5 *For all integer $1 \leq k \leq l$ such that $k \leq l - k$, and all $1 \leq q \leq \infty$, the following bounds hold*

$$\Big(\frac{c}{\varepsilon} \Big)^d \leq D(\mathfrak{P}_{k,l}, d_q, \varepsilon k^{1/q}) \leq \Big(\frac{C}{\varepsilon} \Big)^d, \ \varepsilon > 0 \tag{12}$$

with $d = k(l - k)$ and universal positive constants c, C.

In addition to this, we need the following well known information-theoretic bound frequently used in derivation of minimax lower bounds (see Tsybakov 2008, Theorem 2.5). Let $\Theta = \{\theta_0, \theta_1, \ldots, \theta_M\}$ be a finite parameter space equipped with a metric d and let $\mathcal{P} := \{\mathbb{P}_\theta : \theta \in \Theta\}$ be a family of probability distributions in some sample space. Given $\mathbb{P}, \mathbb{Q} \in \mathcal{P}$, let $K(\mathbb{P} \| \mathbb{Q}) := \mathbb{E}_{\mathbb{P}} \log \frac{d\mathbb{P}}{d\mathbb{Q}}$ be the Kullback-Leibler divergence between $\mathbb{P}$ and $\mathbb{Q}$.

Proposition 6 *Suppose that the following conditions hold:*

(i) for some $s > 0$, $d(\theta_j, \theta_k) \geq 2s > 0, 0 \leq j < k \leq M$;

(ii) for some $0 < \alpha < 1/8$, $\frac{1}{M} \sum_{j=1}^{M} K(\mathbb{P}_{\theta_j} \| \mathbb{P}_{\theta_0}) \leq \alpha \log M$

Then, for a positive constant c_α,

$$\inf_{\hat\theta}\sup_{\theta\in\Theta}\mathbb{P}_\theta\{d(\hat\theta,\theta)\geq s\}\geq c_\alpha,$$

where the infimum is taken over all estimators $\hat\theta\in\Theta$ *based on an observation sampled from* $\mathbb{P}_\theta$.

We now turn to the actual proof of Theorem 4. Under Assumption 4, the following computation is well known: for $\rho_1,\rho_2\in\mathcal{S}_{r,m}$,

$$\begin{aligned}K(\mathbb{P}_{\rho_1}\|\mathbb{P}_{\rho_2})&=\mathbb{E}_{\mathbb{P}_{\rho_1}}\log\frac{\mathbb{P}_{\rho_1}}{\mathbb{P}_{\rho_2}}\Big(X_1,Y_1,\dots,X_n,Y_n\Big)\\&=\mathbb{E}_{\mathbb{P}_{\rho_1}}\sum_{j=1}^n\Big[-\frac{(Y_j-\langle\rho_1,X_j\rangle)^2}{2\sigma_\xi^2}+\frac{(Y_j-\langle\rho_2,X_j\rangle)^2}{2\sigma_\xi^2}\Big]\\&=\mathbb{E}\sum_{j=1}^n\frac{\langle\rho_1-\rho_2,X_j\rangle^2}{2\sigma_\xi^2}=\frac{n}{2\sigma_\xi^2}\|\rho_1-\rho_2\|_{L_2(\Pi)}^2.\end{aligned}\tag{13}$$

It is enough to prove the bounds for $2\leq r\leq m/2$. The proof in the case $r=1$ is simpler and the case $r>m/2$ easily reduces to the case $r\leq m/2$. We will use Lemma 5 to construct a well separated (with respect to d_q) subset of density matrices in $\mathcal{S}_{r,m}$. To this end, first choose a subset $\mathcal{D}_q\subset\mathfrak{P}_{r-1,m-1}$ such that $\mathrm{card}(\mathcal{D}_q)\geq 2^{(r-1)(m-r)}$ and, for some constant c', $\|Q_1-Q_2\|_q\geq c'(r-1)^{1/q}$, $Q_1,Q_2\in\mathfrak{P}_{r-1,m-1},Q_1\neq Q_2$. Such a choice is possible due to the lower bound on the packing numbers of Lemma 5. For $Q\in\mathcal{D}_q$ (note that Q can be viewed as an $(m-1)\times(m-1)$ matrix with real entries) and $\kappa\in(0,1)$, consider the following $m\times m$ matrix

$$S=S_Q=\begin{pmatrix}1-\kappa & \mathbf{0}'\\ \mathbf{0} & \kappa\frac{Q}{r-1}\end{pmatrix}.\tag{14}$$

Note that S is symmetric positively-semidefinite real matrix of unit trace. It is straightforward to check that it defines a Hermitian positively-semidefinite operator in $\mathbb{C}^m$ of unit trace, and it can be identified with a density matrix $S\in\mathcal{S}_m$. Clearly, S is of rank r, so, $S\in\mathcal{S}_{r,m}$.

We will take $\kappa:=c_1\frac{\sigma_\xi m^{3/2}(r-1)}{\sqrt{n}}$ with a small enough absolute constant $c_1>0$ and first assume that $\kappa<1$ (as it is needed in definition Equation 14).

Let $\mathcal{S}_q':=\{S_Q:Q\in\mathcal{D}_q\}$ and consider a family of $M+1=\mathrm{card}(\mathcal{D}_q)\geq 2^{(r-1)(m-r)}$ distributions $\{\mathbb{P}_S:S\in\mathcal{S}_q'\}$. It is immediate that for $S_1=S_{Q_1},S_2=S_{Q_2}$, $Q_1,Q_2\in\mathcal{D}_q,Q_1\neq Q_2$, we have

$$\begin{aligned}\|S_1-S_2\|_q&=\frac{\kappa}{r-1}\|Q_1-Q_2\|_q\geq c'\kappa(r-1)^{1/q-1}\\&\geq c'c_1\frac{\sigma_\xi m^{3/2}(r-1)^{1/q}}{\sqrt{n}}\geq c\frac{\sigma_\xi m^{3/2}r^{1/q}}{\sqrt{n}}\end{aligned}\tag{15}$$

with some constant $c>0$, implying condition (i) of Proposition 6 with $s=\frac{c}{2}\frac{\sigma_\xi m^{3/2}r^{1/q}}{\sqrt{n}}$.

We will now check its condition (ii) . In view of (13), we have, for all $S_1 = S_{Q_1}, S_2 = S_{Q_2} \in \mathcal{S}'_q$,

$$\begin{aligned} K(\mathbb{P}_{S_1} \| \mathbb{P}_{S_2}) &= \frac{n}{2\sigma_\xi^2} \|S_1 - S_2\|^2_{L_2(\Pi)} = \frac{n}{2\sigma_\xi^2 m^2} \|S_1 - S_2\|_2^2 \\ &= \frac{n\kappa^2}{2\sigma_\xi^2 m^2 (r-1)^2} \|Q_1 - Q_2\|_2^2 \le \frac{4n(r-1)\kappa^2}{2\sigma_\xi^2 m^2 (r-1)^2} = 2c_1^2 m(r-1) \\ &\le \alpha m(r-1)/\log(2)/4 \le \frac{\alpha}{2}(r-1)(m-r)\log(2) \le \alpha \log M, \end{aligned} \tag{16}$$

provided that constant c_1 is small enough, so, condition (ii) of Proposition 6 is also satisfied. Proposition 6 implies that, under the assumption $\kappa = c_1 \frac{\sigma_\xi m^{3/2}(r-1)}{\sqrt{n}} < 1$, the following minimax lower bound holds for some $c, c' > 0$:

$$\inf_{\hat\rho} \sup_{\rho \in \mathcal{S}_{r,m}} \mathbb{P}_\rho \left\{ \|\hat\rho - \rho\|_q \ge c \frac{\sigma_\xi m^{\frac{3}{2}} r^{1/q}}{\sqrt{n}} \right\} \ge c'. \tag{17}$$

In the case when

$$c_1 \frac{\sigma_\xi m^{3/2}}{\sqrt{n}} < 1 \le c_1 \frac{\sigma_\xi m^{3/2}(r-1)}{\sqrt{n}},$$

one can choose $2 \le r' < r-1$ such that, for some constant $c_2 > 0$,

$$c_2 < c_1 \frac{\sigma_\xi m^{3/2}(r'-1)}{\sqrt{n}} < 1.$$

For such a choice of r', it follows from (17) that

$$\inf_{\hat\rho} \sup_{\rho \in \mathcal{S}_{r',m}} \mathbb{P}_\rho \left\{ \|\hat\rho - \rho\|_q \ge c \frac{\sigma_\xi m^{\frac{3}{2}} (r')^{1/q}}{\sqrt{n}} \right\} \ge c'. \tag{18}$$

The definition of r' implies that

$$r' \asymp r' - 1 \asymp \left(\frac{\sigma_\xi m^{3/2}}{\sqrt{n}} \right)^{-1}.$$

Therefore,

$$\frac{\sigma_\xi m^{\frac{3}{2}} (r')^{1/q}}{\sqrt{n}} \asymp \left(\frac{\sigma_\xi m^{3/2}}{\sqrt{n}} \right)^{1-1/q},$$

and, since $\mathcal{S}_{r',m} \subset \mathcal{S}_{r,m}$, bound (18) yields

$$\inf_{\hat\rho} \sup_{\rho \in \mathcal{S}_{r,m}} \mathbb{P}_\rho \left\{ \|\hat\rho - \rho\|_q \ge c \left(\frac{\sigma_\xi m^{3/2}}{\sqrt{n}} \right)^{1-1/q} \right\} \ge \inf_{\hat\rho} \sup_{\rho \in \mathcal{S}_{r',m}} \mathbb{P}_\rho \left\{ \|\hat\rho - \rho\|_q \ge c \left(\frac{\sigma_\xi m^{3/2}}{\sqrt{n}} \right)^{1-1/q} \right\} \ge c' \tag{19}$$

for some constants $c, c' > 0$. This allows us to recover the second term in the minimum in bound (9). Finally, in the case when $c_1 \frac{\sigma_\xi m^{3/2}}{\sqrt{n}} > 1$, the minimax lower bound becomes a constant (and the proof is based on a simplified version of the above argument that could be done for $r = 1$). This completes the proof of bound (9) for Schatten q-norms.

The proof of bound (10) for the Hellinger distance is similar. In the case $r \geq 2$, we will use a "well separated" set of density matrices $\mathcal{S}'_q \subset \mathcal{S}_{r,m}$ for $q = 1$ constructed above. We still use $\kappa := c_1 \frac{\sigma_\xi m^{3/2}(r-1)}{\sqrt{n}}$ assuming first that $\kappa \in (0, 1)$. For $S_{Q_1}, S_{Q_2} \in \mathcal{S}'_q$ with $Q_1 \neq Q_2$, it follows by a simple computation and using bound (8) that, for some $c'' > 0$,

$$\begin{aligned} H^2(S_{Q_1}, S_{Q_2}) &= \kappa H^2\Big(\frac{Q_1}{r-1}, \frac{Q_2}{r-1}\Big) \\ &\geq \frac{1}{4}\frac{\kappa}{(r-1)^2}\|Q_1 - Q_2\|_1^2 \geq \frac{(c')^2}{4}\kappa \geq c'' \frac{\sigma_\xi m^{3/2}(r-1)}{\sqrt{n}}. \end{aligned}$$

Repeating the argument based on Proposition 6 yields bound (10) in the case when $\kappa = c_1 \frac{\sigma_\xi m^{3/2}(r-1)}{\sqrt{n}} < 1$, and in the opposite case it is easy to see that the lower bound is a constant.

Finally, bound (11) for the Kullback–Leibler divergence follows from (10) and the inequality $K(\rho \| \hat{\rho}) \geq H^2(\hat{\rho}, \rho)$ (see inequality 8).

■

Next we state similar results in the case of trace regression model with bounded response (see Assumption 3). Denote by $\mathcal{P}_{r,m}(\bar{U})$ the class of all distributions P of (X, Y) such that Assumption 3 holds for some $\bar{U}$ and $\mathbb{E}(Y|X) = \langle \rho_P, X\rangle$ for some $\rho_P \in \mathcal{S}_{r,m}$. Given P, $\mathbb{P}_P$ denotes the corresponding probability measure (such that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. copies of (X, Y) sampled from P).

Theorem 7 *Suppose $\bar{U} \geq 2U$. For all $q \in [1, +\infty]$, there exist absolute constants $c, c' > 0$ such that the following bounds hold:*

$$\inf_{\hat{\rho}} \sup_{P \in \mathcal{P}_{r,m}(\bar{U})} \mathbb{P}_P\Big\{\|\hat{\rho} - \rho_P\|_q \geq c\Big(\frac{\bar{U} m^{\frac{3}{2}} r^{1/q}}{\sqrt{n}} \bigwedge \Big(\frac{\bar{U} m^{3/2}}{\sqrt{n}}\Big)^{1-\frac{1}{q}} \bigwedge 1\Big)\Big\} \geq c', \tag{20}$$

$$\inf_{\hat{\rho}} \sup_{P \in \mathcal{P}_{r,m}(\bar{U})} \mathbb{P}_P\Big\{H^2(\hat{\rho}, \rho_P) \geq c\Big(\frac{\bar{U} m^{\frac{3}{2}} r}{\sqrt{n}} \bigwedge 1\Big)\Big\} \geq c', \tag{21}$$

and

$$\inf_{\hat{\rho}} \sup_{P \in \mathcal{P}_{r,m}(\bar{U})} \mathbb{P}_P\Big\{K(\rho_P \| \hat{\rho}) \geq c\Big(\frac{\bar{U} m^{\frac{3}{2}} r}{\sqrt{n}} \bigwedge 1\Big)\Big\} \geq c', \tag{22}$$

where $\inf_{\hat{\rho}}$ denotes the infimum over all estimators $\hat{\rho}$ in $\mathcal{S}_m$ based on the data $(X_1, Y_1), \ldots, (X_n, Y_n)$.

Proof The proof relies on an idea already used in a context of matrix completion by Koltchinskii et al. (2011) (see their Theorem 7). We need the same family $\mathcal{S}'_q \subset \mathcal{S}_{r,m}$ of "well separated" density matrices of rank r as in the proof of Theorem 4. For a density matrix ρ, let (X, Y) be a random couple such that X is sampled from the uniform distribution Π in $\mathcal{E}$ and, conditionally on X, Y takes value $+\bar{U}$ with probability $p_\rho(X) := \frac{1}{2} + \frac{\langle \rho, X\rangle}{2\bar{U}}$ and value

$-\bar{U}$ with probability $q_\rho(X) := \frac{1}{2} - \frac{\langle\rho, X\rangle}{2\bar{U}}$. Since $\bar{U} \geq 2U$ and $|\langle\rho, X\rangle| \leq \|\rho\|_1 \|X\|_\infty \leq U$, we have $p_\rho(X), q_\rho(X) \in [1/4, 3/4]$ (so, they are bounded away from 0 and from 1). Clearly, $\mathbb{E}_\rho(Y|X) = \langle\rho, X\rangle$. Let P_ρ denote the distribution of such a couple and $\mathbb{P}_\rho$ denote the corresponding distribution of the data $(X_1, Y_1), \ldots, (X_n, Y_n)$. Then, for all $\rho \in \mathcal{S}_{r,m}$, $P_\rho \in \mathcal{P}_{r,m}(\bar{U})$. The only difference with the proof of Theorem 4 is in the bound on Kullback-Leibler divergence $K(\mathbb{P}_{\rho_1} \| \mathbb{P}_{\rho_2})$ (see Equation 13). It is easy to see that

$$K(\mathbb{P}_{\rho_1} \| \mathbb{P}_{\rho_2}) = n\mathbb{E}\left(p_{\rho_1}(X) \log \frac{p_{\rho_1}(X)}{p_{\rho_2}(X)} + q_{\rho_1}(X) \log \frac{q_{\rho_1}(X)}{q_{\rho_2}(X)}\right). \tag{23}$$

The following simple inequality will be used: for all $a, b \in [1/4, 3/4]$,

$$a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b} \leq 12(a-b)^2.$$

It implies that

$$K(\mathbb{P}_{\rho_1} \| \mathbb{P}_{\rho_2}) \leq 3n\mathbb{E}\frac{\langle\rho_1 - \rho_2, X\rangle^2}{\bar{U}^2} \leq \frac{3n}{\bar{U}^2}\|\rho_1 - \rho_2\|^2_{L_2(\Pi)}.$$

This bound is used instead of identity (13) from the proof of Theorem 4. The rest of the proof is the same. ■

Note that the proof requires the possible range $[-\bar{U}, \bar{U}]$ of response variable Y to be larger than the possible range $[-U, U]$ of Fourier coefficients $\langle\rho, E_j\rangle, j = 1, \ldots, m^2$. This is not the case for standard QST model described in the introduction (see also the example of Pauli measurements) and it is of interest to prove a version of minimax lower bounds without this constraint, including the case when $\bar{U} = U$. The following theorem is a result in this direction.

Theorem 8 *Suppose Assumption 1 is satisfied and, moreover, for some constant $\gamma \in (0,1)$,*

$$\Big|\mathrm{tr}(E_k)\Big| \leq (1-\gamma)Um, \ k = 1, \ldots, m^2. \tag{24}$$

Then, for all $q \in [1, +\infty]$, there exist constants $c_\gamma, c'_\gamma > 0$ such that the following bounds hold:

$$\inf_{\hat{\rho}} \sup_{P \in \mathcal{P}_{r,m}(U)} \mathbb{P}_P\Big\{\|\hat{\rho} - \rho_P\|_q \geq c_\gamma\Big(\frac{Um^{\frac{3}{2}} r^{1/q}}{\sqrt{n}} \bigwedge \Big(\frac{Um^{3/2}}{\sqrt{n}}\Big)^{1-\frac{1}{q}} \bigwedge 1\Big)\Big\} \geq c'_\gamma, \tag{25}$$

$$\inf_{\hat{\rho}} \sup_{P \in \mathcal{P}_{r,m}(U)} \mathbb{P}_P\Big\{H^2(\hat{\rho}, \rho_P) \geq c_\gamma\Big(\frac{Um^{\frac{3}{2}} r}{\sqrt{n}} \bigwedge 1\Big)\Big\} \geq c'_\gamma, \tag{26}$$

and

$$\inf_{\hat{\rho}} \sup_{P \in \mathcal{P}_{r,m}(U)} \mathbb{P}_P\Big\{K(\rho_P \| \hat{\rho}) \geq c_\gamma\Big(\frac{Um^{\frac{3}{2}} r}{\sqrt{n}} \bigwedge 1\Big)\Big\} \geq c'_\gamma, \tag{27}$$

where $\inf_{\hat{\rho}}$ *denotes the infimum over all estimators $\hat{\rho}$ in $\mathcal{S}_m$ based on the data $(X_1, Y_1), \ldots, (X_n, Y_n)$.*

Proof The proof is based on the following lemma:

Lemma 9 *Suppose assumption (24) holds. Let K be a sufficiently large absolute constant (to be chosen later) and let m satisfy the condition $K\frac{\log m}{\sqrt{m}} \le \frac{\gamma}{2}$ (which means that $m \ge A_\gamma$ for some constant A_γ). Then there exists $v \in \mathbb{C}^m$ with $\|v\| = 1$ such that*

$$\Big|\langle E_k v, v\rangle\Big| \le (1-\gamma/2)U, k = 1, \ldots, m^2. \tag{28}$$

Proof We will prove this fact by a probabilistic argument. Namely, set $v := m^{-1/2}(\varepsilon_1, \ldots, \varepsilon_m)$, where $\varepsilon_j = \pm 1$. We will show that there is a random choice of "signs" ε_j such that (28) holds. Assume that $\varepsilon_j, j = 1, \ldots, m$ are i.i.d. and take values ± 1 with probability $1/2$ each. Let $E_k := (a_{ij}^{(k)})_{i,j=1,\ldots,m}$. For simplicity, assume that $(a_{ij}^{(k)})_{i,j=1,\ldots,m}$ is a symmetric real matrix (in the complex case, the proof can be easily modified). We have

$$\langle E_k v, v\rangle = \frac{1}{m}\sum_{i=1}^m a_{ii}^{(k)}\varepsilon_i^2 + \frac{1}{m}\sum_{i\neq j} a_{ij}^{(k)}\varepsilon_i\varepsilon_j = \frac{\mathrm{tr}(E_k)}{m} + \frac{1}{m}\sum_{i\neq j} a_{ij}^{(k)}\varepsilon_i\varepsilon_j.$$

It is well known that

$$\mathrm{Var}\Big(\sum_{i\neq j} a_{ij}^{(k)}\varepsilon_i\varepsilon_j\Big) = \mathbb{E}\Big(\sum_{i\neq j} a_{ij}^{(k)}\varepsilon_i\varepsilon_j\Big)^2 = 2\sum_{i\neq j}\Big(a_{ij}^{(k)}\Big)^2 \le 2\sum_{i,j}\Big(a_{ij}^{(k)}\Big)^2 = 2\|E_k\|_2^2 = 2.$$

Moreover, it follows from exponential inequalities for Rademacher chaos (see, e.g., Corollary 3.2.6 in de la Peña and Giné 1999) that for some absolute constant $K > 0$ and for all $t > 0$, with probability at least $1 - e^{-t}$

$$\left|\langle E_k v, v\rangle - \frac{\mathrm{tr}(E_k)}{m}\right| = \left|\frac{1}{m}\sum_{i\neq j} a_{ij}^{(k)}\varepsilon_i\varepsilon_j\right| \le \frac{Kt}{m}.$$

Taking $t = 2\log m$ and using the union bound, we conclude that with probability at least $1 - me^{-2\log m} = 1 - \frac{1}{m} > 0$,

$$\max_{1\le k\le m^2}\left|\langle E_k v, v\rangle - \frac{\mathrm{tr}(E_k)}{m}\right| \le \frac{K\log m}{m} \le \frac{K\log m}{\sqrt{m}}U \le \frac{\gamma}{2}U,$$

where we also used the fact that $U \ge m^{-1/2}$. Thus, there exists a choice of signs ε_j such that

$$\max_{1\le k\le m^2}\Big|\langle E_k v, v\rangle\Big| \le \max_{1\le k\le m}\left|\frac{\mathrm{tr}(E_k)}{m}\right| + \frac{\gamma}{2}U,$$

which, under condition (24), implies (28). ■

We set $e_1 := v$ (where v is the unit vector introduced in Lemma 9) and construct an orthonormal basis $e_1, \ldots, e_m$. Assume that matrices S_Q defined by (14) represent linear transformations in basis $e_1, \ldots, e_m$. Then we have

$$\langle S_Q, E_k\rangle = (1-\kappa)\langle E_k v, v\rangle + \frac{\kappa}{r-1}\langle Q, E_k\rangle.$$

Therefore,

$$\Big|\langle S_Q, E_k\rangle\Big| \le (1-\kappa)\Big|\langle E_k v, v\rangle\Big| + \frac{\kappa}{r-1}\|E_k\|_\infty\|Q\|_1 \le (1-\kappa)(1-\gamma/2)U + \kappa U = (1-(1-\kappa)(\gamma/2))U.$$

Assuming that $\kappa \le 1/2$, we get

$$\Big|\langle S_Q, E_k\rangle\Big| \le (1-\gamma/4)U,\ k = 1,\dots,m^2. \tag{29}$$

The rest of the proof becomes similar to the proof of Theorem 7 (with $\bar{U} = U$). Namely, bound (29) implies that, for $\rho = S_Q$ and X being sampled from the orthonormal basis $\{E_1,\dots,E_{m^2}\}$, probabilities $p_\rho(X)$ and $q_\rho(X)$ are bounded away from 0 and from 1 : $p_\rho(X), q_\rho(X) \in [\gamma/8, 1-\gamma/8]$. This allows us to complete the argument of the proof of Theorem 7. ■

Theorem 8 does not apply directly to the Pauli basis since condition (24) fails in this case. Indeed, by the definition of Pauli basis, $U = m^{-1/2}$ and $\mathrm{tr}(E_1) = \sqrt{m} = Um > (1-\gamma)Um$. Note also that $\mathrm{tr}(E_j) = 0, j = 2,\dots,m^2$. Thus, for Pauli basis, E_1 is the only matrix for which condition (24) fails. However, for this matrix $\langle \rho, E_1\rangle = m^{-1/2}\mathrm{tr}(\rho) = m^{-1/2} = U$ for all density matrices $\rho \in \mathcal{S}_m$. This immediately implies that $p_\rho(E_1) = 1$ and $q_\rho(E_1) = 0$ for all $\rho \in \mathcal{S}_m$ and, as a result, the value $X = E_1$ does not have an impact on the computation of Kullback-Leibler divergence in (23). For the rest of the matrices in the Pauli basis, condition (24) holds implying also bound (28). Therefore, if $X \ne E_1$, we still have that, for $\rho = S_Q$, $p_\rho(X), q_\rho(X) \in [\gamma/8, 1-\gamma/8]$, and the proof of Theorem 7 can be completed in this case, too. Note also that, given X sampled from the Pauli basis, the binary random variable Y taking values $\pm U = \pm\frac{1}{\sqrt{m}}$ with probabilities $p_\rho(X)$ and $q_\rho(X)$, respectively (this is exactly the random variable used in the construction of the proof of Theorem 7) coincides with an outcome of a Pauli measurement for the system prepared in state ρ. These considerations yield the following minimax lower bounds for Pauli measurements.

Theorem 10 *Let $\{E_1,\dots,E_{m^2}\}$ be the Pauli basis in the space $\mathbb{H}_m$ of $m\times m$ Hermitian matrices and let $X_1,\dots,X_n$ be i.i.d. random variables sampled from the uniform distribution in $\{E_1,\dots,E_{m^2}\}$. Let $Y_1,\dots,Y_n$ be outcomes of measurements of observables $X_1,\dots,X_n$ for the system being identically prepared n times in state ρ. The corresponding distribution of the data $(X_1,Y_1),\dots,(X_n,Y_n)$ will be denoted by $\mathbb{P}_\rho$. Then, for all $q \in [1,+\infty]$, there exist constants $c, c' > 0$ such that the following bounds hold:*

$$\inf_{\hat\rho}\sup_{\rho\in\mathcal{S}_{r,m}} \mathbb{P}_\rho\Big\{\|\hat\rho-\rho\|_q \ge c\Big(\frac{mr^{1/q}}{\sqrt{n}} \bigwedge \Big(\frac{m}{\sqrt{n}}\Big)^{1-\frac{1}{q}} \bigwedge 1\Big)\Big\} \ge c', \tag{30}$$

$$\inf_{\hat\rho}\sup_{\rho\in\mathcal{S}_{r,m}} \mathbb{P}_\rho\Big\{H^2(\hat\rho,\rho) \ge c\Big(\frac{mr}{\sqrt{n}} \bigwedge 1\Big)\Big\} \ge c', \tag{31}$$

and

$$\inf_{\hat\rho}\sup_{\rho\in\mathcal{S}_{r,m}} \mathbb{P}_\rho\Big\{K(\rho\|\hat\rho) \ge c\Big(\frac{mr}{\sqrt{n}} \bigwedge 1\Big)\Big\} \ge c', \tag{32}$$

where $\inf_{\hat\rho}$ denotes the infimum over all estimators $\hat\rho$ in $\mathcal{S}_m$ based on the data $(X_1,Y_1),\dots,(X_n,Y_n)$.

Remark 11 *Minimax lower bounds on nuclear norm error of density matrix estimation close to bound (30) for $q = 1$ (but for a somewhat different "estimation protocol" and stated in a different form) were obtained earlier in Flammia et al. (2012). This paper also contains upper bounds on the errors of matrix LASSO and Dantzig selector estimators in the nuclear norm matching the lower bounds up to log-factors.*

Remark 12 *It is easy to see that, if constant $\gamma \in (0,1)$ is small enough (namely, $\gamma < 1-\frac{1}{\sqrt{2}}$), then, in an arbitrary orthonormal basis $\{E_1, \dots, E_{m^2}\}$, there is at most one matrix E_j such that $|\mathrm{tr}(E_j)| > (1-\gamma)Um$. Indeed, note that $\mathrm{tr}(E_j) = \langle E_j, I_m\rangle$. Since*

$$\sum_{j=1}^{m^2} \langle E_j, I_m\rangle^2 = \|I_m\|_2^2 = m$$

and $U^2 m \geq 1$, we have

$$\mathrm{card}\Bigl(\Bigl\{j : |\langle E_j, I_m\rangle| > (1-\gamma)Um\Bigr\}\Bigr) \leq \frac{1}{(1-\gamma)^2 U^2 m^2} \sum_{j=1}^{m^2} \langle E_j, I_m\rangle^2$$

$$\leq \frac{m}{(1-\gamma)^2 U^2 m^2} = \frac{1}{(1-\gamma)^2 U^2 m} \leq \frac{1}{(1-\gamma)^2} < 2,$$

provided that $\gamma < 1 - \frac{1}{\sqrt{2}}$.

Remark 13 *It will be shown in Section 3.3 that the minimax rates of theorems 4, 7, 8 and 10 are attained up to logarithmic factors for the von Neumann entropy penalized least squares estimator.*

Remark 14 *Similar minimax lower bounds could be proved in certain classes of "nearly low rank" density matrices. Consider, for instance, the following class*

$$B_p(d;m) := \Bigl\{S \in \mathcal{S}_m : \sum_{j=1}^{m} |\lambda_j(S)|^p \leq d\Bigr\} \tag{33}$$

for some $d > 0$ and $p \in [0,1]$, where $\lambda_1(S) \geq \cdots \geq \lambda_m(S)$ denote the eigenvalues of S. This set consists of density matrices with the eigenvalues decaying at a certain rate (nearly low rank case) and, for $p = 0$, $d = r$ it coincides with $\mathcal{S}_{r,m}$. It turns out that minimax lower bounds of theorems 4 and 7 hold for the class $B_p(d;m)$ (instead of $\mathcal{S}_{r,m}$) with r replaced by

$$\bar{r} := \bar{r}(\tau, d, m, p) = d\tau^{-p} \wedge m,$$

where $\tau := \frac{\sigma_\xi m^{3/2}}{\sqrt{n}}$ in the case of trace regression with Gaussian noise and $\tau := \frac{\bar{U} m^{3/2}}{\sqrt{n}}$ in the case of trace regression with bounded response. These minimax bounds are attained up to logarithmic factors for a slightly modified von Neumann entropy penalized least squares estimator.

Note that, for $\rho \in B_p(d,m)$ with eigenvalues $\lambda_1(\rho) \geq \cdots \geq \lambda_m(\rho)$, we have $\lambda_j(\rho) \leq \frac{d^{1/p}}{j^{1/p}}, j = 1, \dots, m$. Therefore, for $j \geq \bar{r}$, $\lambda_j(\rho) \leq \tau$. Note also that τ characterizes the *minimax rate of estimation of $\rho \in \mathcal{S}_{r,m}$ in the operator norm for any value of the rank r (see bound (9) for $q = +\infty$; the corresponding upper bound also holds for the least squares estimator up to a logarithmic factor, see Koltchinskii and Xia 2015+). Roughly speaking, τ is a threshold below which the estimation of eigenvalues $\lambda_j(\rho)$ becomes impossible and $\bar{r}$ can be viewed as an "effective rank" of nearly low rank density matrices in the class $B_p(d,m)$.*

3. Von Neumann Entropy Penalization: Optimality and Oracle Inequalities

The goal of this section is to study optimality properties of von Neumann entropy penalized least squares estimator $\tilde{\rho}^{\varepsilon}$ defined by (7). In particular, we establish oracle inequalities for such estimators in the cases of trace regression with bounded response (Subsection 3.1) and trace regression with Gaussian noise (Subsection 3.2), and prove upper bounds on their estimation errors measured by Schatten q-norm distances for $q \in [1,2]$ and also by Hellinger and Kullback-Leibler distances (Subsection 3.3).

3.1 Oracle Inequalities for Trace Regression with Bounded Response

In this subsection, we prove a *sharp low rank oracle inequality* for estimator $\tilde{\rho}^{\varepsilon}$ defined by (7). It is done in the case of trace regression model with bounded response (that is, under Assumption 3). The results of this type show some form of optimality of the estimation method, namely, that the estimator provides an optimal trade-off between the "approximation error" of the target density matrix by a low rank "oracle" and the "estimation error" of the "oracle" that is proportional to its rank. Sharp oracle inequalities (in which the leading constant in front of the "approximation error" is equal to 1, so that the bound mimics precisely the approximation by the oracle) are usually harder to prove. In the case of low rank matrix completion, the first result of this type was proved by Koltchinskii et al. (2011) for a modified least squares estimator with nuclear norm penalty. A version of such inequality for empirical risk minimization with nuclear norm penalty (that includes matrix LASSO) was first proved by Koltchinskii (2013b). Low rank oracle inequalities for von Neumann entropy penalized least squares method with the leading constant larger than 1 were proved by Koltchinskii (2011a). The main result of this section refines these previous bounds by proving a sharp oracle inequality, improving the logarithmic factors and removing superfluous assumptions, but also by establishing the inequality in the whole range of values of regularization parameter $\varepsilon \geq 0$ (including the value $\varepsilon = 0$, for which $\tilde{\rho}^{\varepsilon}$ coincides with the least squares estimator $\hat{\rho}$). In addition to this, for a special choice of regularization parameter ε, the theorem below also provides an upper bound on the Kullback-Leibler error $K(\rho \| \tilde{\rho}^{\varepsilon})$ of $\tilde{\rho}^{\varepsilon}$ that matches the minimax lower bound (22) up to log-factors (and "second order terms"). It turns out that, for this choice of ε, the estimator satisfies exactly the same low rank oracle inequality as the best inequalities known for LASSO estimator and minimax optimal error rates are attained for $\tilde{\rho}^{\varepsilon}$ also with respect to Hellinger distance and Schatten q-norm distances for all $q \in [1,2]$ (see Section 3.3). For simplicity, it will be assumed that constants U in Assumption 1 and $\bar{U}$ in Assumption 3 coincide (in the upper bounds, one can always replace U and $\bar{U}$ by $U \vee \bar{U}$).

Theorem 15 *Suppose Assumption 3 holds with constant $\bar{U} = U$ and let $\varepsilon \in [0,1]$. Then, there exists a constant $C > 0$ such that for all $t \geq 1$ with probability at least $1 - e^{-t}$*

$$\|f_{\tilde{\rho}^{\varepsilon}} - f_{\rho}\|^2_{L_2(\Pi)} \leq \inf_{S \in \mathcal{S}_m}\biggl[\|f_S - f_{\rho}\|^2_{L_2(\Pi)} + C\biggl(\mathrm{rank}(S) m^2 \varepsilon^2 \log^2(mn) + U^2 \frac{\mathrm{rank}(S) m \log(2m)}{n} + U^2 \frac{t + \log\log_2(2n)}{n}\biggr)\biggr]. \tag{34}$$

In particular, this implies that

$$\|f_{\tilde{\rho}^{\varepsilon}} - f_{\rho}\|^2_{L_2(\Pi)} \leq C\biggl[\mathrm{rank}(\rho) m^2 \varepsilon^2 \log^2(mn) + U^2 \frac{\mathrm{rank}(\rho) m \log(2m)}{n} + U^2 \frac{t + \log\log_2(2n)}{n}\biggr]. \tag{35}$$

Moreover, if

$$\varepsilon := \frac{1}{\log(mn)}\biggl[U\sqrt{\frac{\log(2m)}{nm}} \bigvee U^2 \frac{\log(2m)}{n}\biggr],$$

then, with some constant C and with probability at least $1 - e^{-t}$

$$\|f_{\tilde{\rho}^{\varepsilon}} - f_{\rho}\|^2_{L_2(\Pi)} \leq C\biggl[U^2 \frac{\mathrm{rank}(\rho) m \log(2m)}{n}\biggl(1 \bigvee U^2 \frac{m \log(2m)}{n}\biggr) + U^2 \frac{t + \log\log_2(2n)}{n}\biggr] \tag{36}$$

and

$$K(\rho \| \tilde{\rho}^{\varepsilon}) \leq CU\biggl[\frac{\mathrm{rank}(\rho) m^{3/2} \sqrt{\log(2m)} \log(mn)}{\sqrt{n}}\biggl(1 \bigvee U\sqrt{\frac{m \log(2m)}{n}}\biggr) + \sqrt{\frac{m}{n}} \frac{(t + \log\log_2(2n)) \log(mn)}{\sqrt{\log(2m)}}\biggr]. \tag{37}$$

Proof The following notations will be used in the proof. Let $\ell(y,u) := (u-y)^2, y, u \in \mathbb{R}$ be the quadratic loss function. For $f : \mathbb{H}_m \mapsto \mathbb{R}$, denote

$$(\ell \bullet f)(x,y) = (f(x) - y)^2, \quad (\ell' \bullet f)(x,y) = 2(f(x) - y)$$

and

$$P(\ell \bullet f) = \mathbb{E}(Y - f(X))^2, \quad P_n(\ell \bullet f) = n^{-1}\sum_{j=1}^{n}(Y_j - f(X_j))^2.$$

For $A \in \mathbb{H}_m$, let $f_A(x) = \langle A, x\rangle, x \in \mathbb{H}_m$. Since for density matrices $S \in \mathcal{S}_m$, $\|S\|_1 = \mathrm{tr}(S) = 1$, the estimator $\tilde{\rho} = \tilde{\rho}^{\varepsilon}$ can be equivalently defined by the following convex optimization problem:

$$\tilde{\rho} = \mathrm{argmin}_{S \in \mathcal{S}_m} L_n(S), \quad L_n(S) := \Bigl[P_n(\ell \bullet f_S) + \varepsilon \mathrm{tr}(S \log S) + \bar{\varepsilon}\|S\|_1\Bigr]$$

for an arbitrary $\bar{\varepsilon} > 0$.

The following lemma will be crucial in the proofs of Theorem 15 as well Theorem 19 in the following subsection. Note that it does not rely on Assumption 3, only Assumptions 1 and 2 are needed.

Lemma 16 *Suppose Assumptions 1 and 2 hold. Let $\delta \in (0,1)$ and $S := (1-\delta)S' + \delta\frac{I_m}{m}$, where $S' \in \mathcal{S}_m$, $\mathrm{rank}(S') = r$ and I_m is the $m \times m$ identity matrix. Then the following bound holds:*

$$\begin{aligned}\|f_{\tilde\rho} - f_\rho\|^2_{L_2(\Pi)} + \tfrac{1}{2}\|f_{\tilde\rho} - f_S\|^2_{L_2(\Pi)} + \varepsilon K(\tilde\rho; S) + \bar\varepsilon\left\|\mathcal{P}_L^\perp(\tilde\rho)\right\|_1 \\ \le \|f_S - f_\rho\|^2_{L_2(\Pi)} + rm^2\varepsilon^2\log^2(m/\delta) + rm^2\bar\varepsilon^2 \\ +4\bar\varepsilon\delta + (P-P_n)(\ell' \bullet f_{\tilde\rho})(f_{\tilde\rho} - f_S).\end{aligned} \tag{38}$$

Lemma 16 will be often used together with the following simple bound:

$$\begin{aligned}\|f_S - f_\rho\|^2_{L_2(\Pi)} = \tfrac{1}{m^2}\|S-\rho\|_2^2 \le \\ \tfrac{1}{m^2}\|S'-\rho\|_2^2 + \tfrac{2}{m^2}\|S'-\rho\|_2\|S'-S\|_2 + \tfrac{1}{m^2}\|S'-S\|_2^2 \\ \le \|f_{S'} - f_\rho\|^2_{L_2(\Pi)} + \tfrac{8\delta}{m^2} + \tfrac{4\delta^2}{m^2} \le \|f_{S'} - f_\rho\|^2_{L_2(\Pi)} + \tfrac{12\delta}{m^2}.\end{aligned} \tag{39}$$

Together, they imply that

$$\begin{aligned}\|f_{\tilde\rho} - f_\rho\|^2_{L_2(\Pi)} + \tfrac{1}{2}\|f_{\tilde\rho} - f_S\|^2_{L_2(\Pi)} + \varepsilon K(\tilde\rho; S) + \bar\varepsilon\left\|\mathcal{P}_L^\perp(\tilde\rho)\right\|_1 \\ \le \|f_{S'} - f_\rho\|^2_{L_2(\Pi)} + rm^2\varepsilon^2\log^2(m/\delta) + rm^2\bar\varepsilon^2 \\ +4\bar\varepsilon\delta + \tfrac{12\delta}{m^2} + (P-P_n)(\ell' \bullet f_{\tilde\rho})(f_{\tilde\rho} - f_S).\end{aligned} \tag{40}$$

We will now give the proof of Lemma 16.

Proof By standard necessary conditions of extremum in convex problems, we get that, for all $S \in \mathcal{S}_m$ and for some $\tilde V \in \partial\|\tilde\rho\|_1$,

$$P_n(\ell' \bullet f_{\tilde\rho})(f_{\tilde\rho} - f_S) + \varepsilon\langle \log\tilde\rho, \tilde\rho - S\rangle + \bar\varepsilon\langle \tilde V, \tilde\rho - S\rangle \le 0$$

(see, e.g., Aubin and Ekeland 2006, Chapter 2, Corollary 6; see also Koltchinskii 2011b, pp. 198–199; for the computation of derivative of the function $\mathrm{tr}(S\log S)$, see Lemma 1 in Koltchinskii 2011a). Replacing in the left hand side P by P_n, we get

$$P(\ell' \bullet f_{\tilde\rho})(f_{\tilde\rho} - f_S) + \varepsilon\langle \log\tilde\rho, \tilde\rho - S\rangle + \bar\varepsilon\langle \tilde V, \tilde\rho - S\rangle \le (P-P_n)(\ell' \bullet f_{\tilde\rho})(f_{\tilde\rho} - f_S).$$

It is easy to check that for the quadratic loss

$$P(\ell' \bullet f_{\tilde\rho})(f_{\tilde\rho} - f_S) = P(\ell \bullet f_{\tilde\rho}) - P(\ell \bullet f_S) + \|f_{\tilde\rho} - f_S\|^2_{L_2(\Pi)},$$

implying that

$$P(\ell \bullet f_{\tilde\rho}) - P(\ell \bullet f_S) + \|f_{\tilde\rho} - f_S\|^2_{L_2(\Pi)} + \varepsilon\langle \log\tilde\rho, \tilde\rho - S\rangle + \bar\varepsilon\langle \tilde V, \tilde\rho - S\rangle$$

$$\le (P-P_n)(\ell' \bullet f_{\tilde\rho})(f_{\tilde\rho} - f_S).$$

Also, for the quadratic loss,

$$P(\ell \bullet f) - P(\ell \bullet f_\rho) = \|f - f_\rho\|^2_{L_2(\Pi)}.$$

Therefore,

$$\|f_{\tilde\rho}-f_\rho\|_{L_2(\Pi)}^2+\|f_{\tilde\rho}-f_S\|_{L_2(\Pi)}^2+\varepsilon\langle\log\tilde\rho,\tilde\rho-S\rangle+\bar\varepsilon\langle\tilde V,\tilde\rho-S\rangle$$
$$\le\|f_S-f_\rho\|_{L_2(\Pi)}^2+(P-P_n)(\ell'\bullet f_{\tilde\rho})(f_{\tilde\rho}-f_S).$$

Recall that we have set $S=(1-\delta)S'+\delta\frac{I_m}{m}$, where $S'\in\mathcal{S}_m$, $\mathrm{rank}(S')=r$, $\delta\in(0,1)$. Clearly,

$$\Bigl|\langle\tilde V,S-S'\rangle\Bigr|\le\|\tilde V\|_\infty\|S-S'\|_1\le\|S-S'\|_1=\delta\Bigl\|S'-\frac{I_m}{m}\Bigr\|_1\le 2\delta,$$

where we used the fact that $\|\tilde V\|_\infty\le 1$ for $\tilde V\in\partial\|\tilde\rho\|_1$. This implies

$$\begin{aligned}&\|f_{\tilde\rho}-f_\rho\|_{L_2(\Pi)}^2+\|f_{\tilde\rho}-f_S\|_{L_2(\Pi)}^2+\varepsilon\langle\log\tilde\rho,\tilde\rho-S\rangle+\bar\varepsilon\langle\tilde V,\tilde\rho-S'\rangle\\&\quad\le\|f_S-f_\rho\|_{L_2(\Pi)}^2+2\bar\varepsilon\delta+(P-P_n)(\ell'\bullet f_{\tilde\rho})(f_{\tilde\rho}-f_S).\end{aligned}\tag{41}$$

Recall formula (2) for the subdifferential of nuclear norm. Let $L=\mathrm{supp}(S')$. By the duality between the operator and nuclear norms, there exists $M\in\mathbb{H}_m$ with $\|M\|_\infty\le 1$ such that

$$\langle\mathcal{P}_L^\perp(M),\tilde\rho-S'\rangle=\langle M,\mathcal{P}_L^\perp(\tilde\rho-S')\rangle=\Bigl\|\mathcal{P}_L^\perp(\tilde\rho-S')\Bigr\|_1=\Bigl\|\mathcal{P}_L^\perp(\tilde\rho)\Bigr\|_1.$$

With $V=\mathrm{sign}(S')+\mathcal{P}_L^\perp(M)\in\partial\|S'\|_1$, by monotonicity of subdifferential, we get that

$$\langle\mathrm{sign}(S'),\tilde\rho-S'\rangle+\Bigl\|\mathcal{P}_L^\perp(\tilde\rho)\Bigr\|_1=\langle V,\tilde\rho-S'\rangle\le\langle\tilde V,\tilde\rho-S'\rangle.\tag{42}$$

In addition to this, we have

$$\langle\log\tilde\rho,\tilde\rho-S\rangle=\langle\log\tilde\rho-\log S,\tilde\rho-S\rangle+\langle\log S,\tilde\rho-S\rangle=K(\tilde\rho;S)+\langle\log S,\tilde\rho-S\rangle.\tag{43}$$

Substituting (42) and (43) into (41), we get

$$\begin{aligned}&\|f_{\tilde\rho}-f_\rho\|_{L_2(\Pi)}^2+\|f_{\tilde\rho}-f_S\|_{L_2(\Pi)}^2+\varepsilon K(\tilde\rho;S)+\bar\varepsilon\Bigl\|\mathcal{P}_L^\perp(\tilde\rho)\Bigr\|_1\\&\quad\le\|f_S-f_\rho\|_{L_2(\Pi)}^2+\varepsilon\langle\log S,S-\tilde\rho\rangle+\bar\varepsilon\langle\mathrm{sign}(S'),S'-\tilde\rho\rangle\\&\qquad+2\bar\varepsilon\delta+(P-P_n)(\ell'\bullet f_{\tilde\rho})(f_{\tilde\rho}-f_S).\end{aligned}\tag{44}$$

The following bound on $\bar\varepsilon\langle\mathrm{sign}(S'),S'-\tilde\rho\rangle$ is straightforward:

$$\begin{aligned}&\bar\varepsilon\langle\mathrm{sign}(S'),S'-\tilde\rho\rangle\le\bar\varepsilon\langle\mathrm{sign}(S'),S-\tilde\rho\rangle+\bar\varepsilon\|\mathrm{sign}(S')\|_\infty\|S-S'\|_1\\&\quad\le\bar\varepsilon\|\mathrm{sign}(S')\|_2\|S-\tilde\rho\|_2+2\bar\varepsilon\delta\le\bar\varepsilon\sqrt{r}m\|f_S-f_{\tilde\rho}\|_{L_2(\Pi)}+2\bar\varepsilon\delta\\&\quad\le rm^2\bar\varepsilon^2+\tfrac14\|f_S-f_{\tilde\rho}\|_{L_2(\Pi)}^2+2\bar\varepsilon\delta.\end{aligned}\tag{45}$$

A similar bound on $\varepsilon\langle\log S,S-\tilde\rho\rangle$ is only slightly more complicated. Suppose S' has the following spectral representation: $S'=\sum_{k=1}^r\lambda_kP_k$ with eigenvalues $\lambda_k\in(0,1]$ (repeated with their multiplicities) and one-dimensional orthogonal eigenprojectors P_k. We will extend $P_j,j=1,\dots,r$ to the complete orthogonal resolution of the identity $P_j,j=1,\dots,m$. Then

$$\log S=\log\Bigl((1-\delta)S'+\delta\frac{I_m}{m}\Bigr)=\sum_{j=1}^r\log\Bigl((1-\delta)\lambda_j+\delta/m\Bigr)P_j+\sum_{j=r+1}^m\log(\delta/m)P_j$$

$$= \sum_{j=1}^{r} \log\Big(1 + (1-\delta)m\lambda_j/\delta\Big)P_j + \log(\delta/m)I_m$$

and

$$\langle \log S, S - \tilde\rho\rangle = \Big\langle \sum_{j=1}^{r} \log\Big(1 + (1-\delta)m\lambda_j/\delta\Big)P_j, S - \tilde\rho\Big\rangle + \log(\delta/m)\langle I_m, S-\tilde\rho\rangle$$
$$= \Big\langle \sum_{j=1}^{r} \log\Big(1 + (1-\delta)m\lambda_j/\delta\Big)P_j, S - \tilde\rho\Big\rangle$$

where we used the fact that $\langle I_m, S-\tilde\rho\rangle = \mathrm{tr}(S) - \mathrm{tr}(\tilde\rho) = 0$. Therefore,

$$\varepsilon\langle \log S, S-\tilde\rho\rangle \le \varepsilon \Big\| \textstyle\sum_{j=1}^{r} \log\Big(1 + (1-\delta)m\lambda_j/\delta\Big)P_j\Big\|_2 \|S-\tilde\rho\|_2 \tag{46}$$
$$= \varepsilon m \Big(\textstyle\sum_{j=1}^{r} \log^2\Big(1 + (1-\delta)m\lambda_j/\delta\Big)\Big)^{1/2} \|f_S - f_{\tilde\rho}\|_{L_2(\Pi)}$$
$$\le \varepsilon\sqrt{r}m\log(m/\delta)\|f_S - f_{\tilde\rho}\|_{L_2(\Pi)} \le rm^2\varepsilon^2\log^2(m/\delta) + \tfrac{1}{4}\|f_S - f_{\tilde\rho}\|^2_{L_2(\Pi)},$$

where it was used that for $\lambda_j \in [0,1]$

$$\log\Big(1 + (1-\delta)m\lambda_j/\delta\Big) \le \log\Big(\frac{\delta + (1-\delta)m}{\delta}\Big) \le \log(m/\delta).$$

Substituting bounds (45) and (46) in (44) we easily get bound (38), as claimed in the lemma. ■

We will also need the following simple lemma that provides a bound on $K(S'\|\tilde\rho)$ in terms of $K(S\|\tilde\rho)$.

Let

$$h(\delta) := \delta\log\frac{1}{\delta} + (1-\delta)\log\frac{1}{1-\delta}.$$

Observe that

$$h(\delta) = \delta\log\frac{1}{\delta} + (1-\delta)\log\Big(1 + \frac{\delta}{1-\delta}\Big) \le \delta\log\frac{1}{\delta} + (1-\delta)\frac{\delta}{1-\delta} \le \delta\log\frac{e}{\delta}$$

(this bound will be used in what follows).

Lemma 17 *Let $\delta \in (0,1)$, $S' \in \mathcal{S}_m$ with $\mathrm{rank}(S') = r$ and $S = (1-\delta)S' + \delta\frac{I_m}{m}$. Then, for any $U \in \mathcal{S}_m$,*

$$K(S'\|U) \le \frac{K(S\|U) + h(\delta)}{1-\delta}.$$

Proof The following identities are straightforward:

$$K(S\|U) = \mathrm{tr}(S(\log S - \log U))$$
$$= (1-\delta)\mathrm{tr}(S'(\log S - \log U)) + \delta\mathrm{tr}((I_m/m)(\log S - \log U))$$
$$= (1-\delta)\mathrm{tr}(S'(\log S' - \log U)) + (1-\delta)\mathrm{tr}(S'(\log S - \log S'))$$
$$+\delta\mathrm{tr}((I_m/m)(\log S - \log(I_m/m))) + \delta\mathrm{tr}((I_m/m)(\log(I_m/m) - \log U))$$
$$= (1-\delta)K(S'\|U) - (1-\delta)K(S'\|S) + \delta K(I_m/m\|U) - \delta K(I_m/m\|S).$$

Since $K(I_m/m\|U) \geq 0$, it follows that

$$K(S'\|U) \leq \frac{K(S\|U)}{1-\delta} + K(S'\|S) + \frac{\delta}{1-\delta} K(I_m/m\|S). \tag{47}$$

Assuming that S' has spectral representation $S' = \sum_{j=1}^r \lambda_j P_j$ with eigenvalues $\lambda_j > 0$ and one-dimensional projectors P_j, we get

$$-K(S'\|S) = \sum_{j=1}^r \lambda_j \log \frac{(1-\delta)\lambda_j + \delta/m}{\lambda_j}$$

$$= \sum_{j=1}^r \lambda_j \log\Bigl(1-\delta+\frac{\delta}{m\lambda_j}\Bigr) \geq \log(1-\delta)\sum_{j=1}^r \lambda_j = \log(1-\delta),$$

implying that $K(S'\|S) \leq \log\frac{1}{1-\delta}$. On the other hand,

$$K(I_m/m\|S) = \frac{1}{m}\sum_{j=1}^m \log\frac{1/m}{(1-\delta)\lambda_j + \delta/m} \leq \frac{1}{m}\sum_{j=1}^m \log\frac{1}{\delta} = \log\frac{1}{\delta}.$$

Substituting these bounds in (47) yields the result. ■

To complete the proof of Theorem 15, we need to control the empirical process $(P - P_n)(\ell' \bullet f_{\tilde\rho})(f_{\tilde\rho} - f_S)$ in the right hand side of bound (38). Our approach is based on the following empirical processes bound that is a slight modification of Lemma 1 in Koltchinskii (2013b). As before, we assume that $S = (1-\delta)S' + \delta\frac{I_m}{m}$ with $S' \in \mathcal{S}_m$, $\mathrm{rank}(S') = r$. We will set $\delta := \frac{1}{m^2n^2}$.

Let $\Xi_\varepsilon := n^{-1}\sum_{j=1}^n \varepsilon_j X_j$, where ε_j are i.i.d. Rademacher random variables (that is, ε_j takes values $+1$ and -1 with probability $1/2$ each) and $\{\varepsilon_j\}, \{X_j\}$ are independent.

Lemma 18 *Given $\delta_1, \delta_2 > 0$, denote*

$$\alpha_n(\delta_1,\delta_2) := \sup\Bigl\{\Bigl|(P_n - P)(\ell' \bullet f_A)(f_A - f_S)\Bigr| : A \in \mathcal{S}_m, \|f_A - f_S\|_{L_2(\Pi)} \leq \delta_1, \|\mathcal{P}_L^\perp A\|_1 \leq \delta_2\Bigr\}.$$

Let $0 < \delta_1^- < \delta_1^+, 0 < \delta_2^- < \delta_2^+$. For $t \geq 1$, denote

$$\bar t := t + \log\Bigl([\log_2(\delta_1^+/\delta_1^-)] + 2\Bigr) + \log\Bigl([\log_2(\delta_2^+/\delta_2^-)] + 2\Bigr) + \log 3.$$

Then, with probability at least $1 - e^{-t}$, for all $\delta_1 \in [\delta_1^-, \delta_1^+], \delta_2 \in [\delta_2^-, \delta_2^+]$,

$$\alpha_n(\delta_1,\delta_2) \leq C_1 U \mathbb{E}\|\Xi_\varepsilon\|_\infty\Bigl(\sqrt{r}m\delta_1 + \delta_2 + \delta\Bigr) + C_2 U\delta_1\sqrt{\frac{\bar t}{n}} + C_3 U^2\frac{\bar t}{n},$$

where $C_1, C_2, C_3 > 0$ are constants.

We will use this lemma to control the term $(P-P_n)(\ell'\bullet f_{\tilde\rho})(f_{\tilde\rho}-f_S)$ in bound (38). Let $\delta_1 := \|f_{\tilde\rho}-f_S\|_{L_2(\Pi)}$ and $\delta_2 := \|\mathcal{P}_L^{\perp}\tilde\rho\|_1$. Define also

$$\delta_1^+ := \frac{2}{m},\ \delta_2^+ := 1,\ \delta_1^- = \delta_2^- := \frac{1}{mn},$$

so that $\bar t \le t + 2\log(\log_2(mn)+3)+\log 3$. It is easy to see that $\delta_1 \le \delta_1^+$ and $\delta_2 \le \delta_2^+$. If, in addition, $\delta_1 \ge \delta_1^-$, $\delta_2 \ge \delta_2^-$, the bound of Lemma 18 implies that with probability at least $1-e^{-t}$

$$(P-P_n)(\ell'\bullet f_{\tilde\rho})(f_{\tilde\rho}-f_S) \le \alpha_n(\delta_1,\delta_2)$$
$$\le C_1 U\mathbb{E}\|\Xi_\varepsilon\|_\infty\Big(\sqrt{r}m\delta_1+\delta_2+\delta\Big)+C_2U\delta_1\sqrt{\frac{\bar t}{n}}+C_3U^2\frac{\bar t}{n}$$

If $\bar\varepsilon \ge C_1U\mathbb{E}\|\Xi_\varepsilon\|_\infty$, the last bound implies that

$$\begin{aligned}&(P-P_n)(\ell'\bullet f_{\tilde\rho})(f_{\tilde\rho}-f_S)\\ &\le \tfrac14\|f_{\tilde\rho}-f_S\|_{L_2(\Pi)}^2+rm^2\bar\varepsilon^2+\bar\varepsilon\|\mathcal{P}_L^{\perp}\tilde\rho\|_1+\bar\varepsilon\delta\\ &+\tfrac14\|f_{\tilde\rho}-f_S\|_{L_2(\Pi)}^2+(C_2^2+C_3)U^2\tfrac{\bar t}{n}.\end{aligned}\tag{48}$$

Substituting this bound in the right hand side of (40), we get

$$\begin{aligned}&\|f_{\tilde\rho}-f_\rho\|_{L_2(\Pi)}^2+\varepsilon K(\tilde\rho;S)\\ &\le \|f_{S'}-f_\rho\|_{L_2(\Pi)}^2+rm^2\varepsilon^2\log^2(m/\delta)+2rm^2\bar\varepsilon^2\\ &+5\bar\varepsilon\delta+CU^2\tfrac{\bar t}{n}+\tfrac{12\delta}{m^2},\end{aligned}\tag{49}$$

where $C := C_2^2+C_3$.

In the case when $\delta_1 = \|f_{\tilde\rho}-f_S\|_{L_2(\Pi)} \le \delta_1^- = \frac{1}{mn}$ or $\delta_2 = \|\mathcal{P}_L^{\perp}\tilde\rho\|_1 \le \delta_2^- = \frac{1}{mn}$, we can replace the terms $\frac14\|f_{\tilde\rho}-f_S\|_{L_2(\Pi)}^2$ or $\|\mathcal{P}_L^{\perp}\tilde\rho\|_1$ in bound (48) by their respective upper bounds ($\frac14(\delta_1^-)^2 = \frac{1}{4m^2n^2}$, or $\delta_2^- = \frac{1}{mn}$), which would be smaller than $CU^2\frac{\bar t}{n}$ for large enough $C>0$, so bound (49) still holds (recall that $U \ge m^{-1/2}$). Note also that $\frac{12\delta}{m^2} = 12\frac{1}{m^4n^2} \le 12U^2\frac{\bar t}{n}$. Thus, increasing the value of constant C, one can rewrite (49) in a simpler form as

$$\begin{aligned}&\|f_{\tilde\rho}-f_\rho\|_{L_2(\Pi)}^2+\varepsilon K(\tilde\rho;S)\\ &\le \|f_{S'}-f_\rho\|_{L_2(\Pi)}^2+rm^2\varepsilon^2\log^2(m/\delta)+2rm^2\bar\varepsilon^2\\ &+5\bar\varepsilon\delta+CU^2\tfrac{\bar t}{n}.\end{aligned}\tag{50}$$

The following expectation bound is a consequence of a matrix version of Bernstein inequality for $\|\Xi_\varepsilon\|_\infty$ (it follows by integrating out its exponential tails):

$$\mathbb{E}\|\Xi_\varepsilon\|_\infty \le 4\Big[\sqrt{\frac{\log(2m)}{nm}}\bigvee U\frac{\log(2m)}{n}\Big]$$

(it is also used in this computation that, in the case of uniform sampling from an orthonormal basis, $\sigma_{\varepsilon X}^2 = \|\mathbb{E}X^2\|_\infty = \frac{1}{m}$, a simple fact often used in the literature; see, e.g., Koltchinskii 2011a, Section 5). Let

$$\bar\varepsilon := D'U\sqrt{\frac{\log(2m)}{nm}}$$

for some constant D'. If D' is sufficiently large and

$$U\frac{\log(2m)}{n} \leq \sqrt{\frac{\log(2m)}{nm}}, \tag{51}$$

then the condition $\bar{\varepsilon} \geq C_1 U \mathbb{E}\|\Xi_\varepsilon\|_\infty$ is satisfied and bound (50) holds with probability at least $1-e^{-t}$. Moreover, $\bar{\varepsilon}\delta \lesssim_{D'} \delta \lesssim_{D'} U^2\frac{\bar{t}}{n}$, implying that the term $5\bar{\varepsilon}\delta$ in (50) can be dropped at a price of further increasing the value of constant C.

If (51) does not hold, we still have that

$$\|f_{\tilde{\rho}} - f_\rho\|^2_{L_2(\Pi)} = \frac{\|\tilde{\rho}-\rho\|_2^2}{m^2} \leq \frac{2}{m^2} \leq CU^2\frac{\bar{t}}{n}.$$

Recalling that $\bar{t} \leq t + 2\log(\log_2(mn)+3)$ and $\log(m/\delta) \lesssim \log(mn)$, we deduce from (50) that with some constant C and with probability at least $1-e^{-t}$

$$\begin{aligned}\|f_{\tilde{\rho}} - f_\rho\|^2_{L_2(\Pi)} \leq \|f_{S'} - f_\rho\|^2_{L_2(\Pi)} + C\Big[& rm^2\varepsilon^2\log^2(mn) \\ &+U^2\tfrac{rm\log(2m)}{n} + U^2\tfrac{t+\log(\log_2(mn)+3)}{n}\Big].\end{aligned} \tag{52}$$

Note that, for $n \geq 2$,

$$\log(\log_2(mn)+3) = \log\Big(\log_2(4m)+\log_2(2n)\Big) \leq \log\log_2(4m) + \log\log_2(2n), \tag{53}$$

since $\log_2(4m)+\log_2(2n) \leq \log_2(4m)\log_2(2n)$. Since also, for $r \geq 1$,

$$U^2\frac{t+\log\log_2(4m)}{n} \lesssim U^2\frac{rm\log(2m)}{n}, \tag{54}$$

we can replace in bound (52) the term $U^2\frac{t+\log(\log_2(mn)+3)}{n}$ with the term $U^2\frac{t+\log\log_2(2n)}{n}$ (increasing the value of the constant C accordingly). This yields bound (34) of the theorem. For $S' = \rho$, it yields bound (35), and, moreover, for $S' = \rho$ and $S = (1-\delta)\rho + \delta\frac{I_m}{m}$ with $\delta = \frac{1}{m^2n^2}$, bound (50) also implies that

$$\begin{aligned}\varepsilon K(\tilde{\rho}; S) \leq \mathrm{rank}(\rho)m^2\varepsilon^2\log^2(m/\delta) + 2\mathrm{rank}(\rho)m^2\bar{\varepsilon}^2 \\ +5\bar{\varepsilon}\delta + CU^2\tfrac{\bar{t}}{n}.\end{aligned} \tag{55}$$

We will now take

$$\bar{\varepsilon} := D'\Big[U\sqrt{\frac{\log(2m)}{nm}}\bigvee U^2\frac{\log(2m)}{n}\Big]$$

for a large enough constant D' so that $\bar{\varepsilon} \geq C_1 U\mathbb{E}\|\Xi_\varepsilon\|_\infty$. Assume that

$$\varepsilon := \frac{1}{\log(mn)}\Big[U\sqrt{\frac{\log(2m)}{nm}}\bigvee U^2\frac{\log(2m)}{n}\Big].$$

As before, the term $\bar{\varepsilon}\delta$ in bound (55) will be absorbed by the term $CU^2\frac{\bar{t}}{n}$ with a larger value of C and also

$$\operatorname{rank}(\rho)m^2\varepsilon^2\log^2(m/\delta) \asymp_{D'} \operatorname{rank}(\rho)m^2\bar{\varepsilon}^2 \asymp_{D'} U^2\frac{\operatorname{rank}(\rho)m\log(2m)}{n}\left(1\bigvee U^2\frac{m\log(2m)}{n}\right).$$

As a result, taking into account (53), (54), bound (55) can be rewritten as follows:

$$\varepsilon K(\tilde{\rho};S) \le CU^2\left[\tfrac{\operatorname{rank}(\rho)m\log(2m)}{n}\left(1\bigvee U^2\tfrac{m\log(2m)}{n}\right)\right. \\ \left. +\tfrac{t+\log\log_2(2n)}{n}\right]. \tag{56}$$

Using the bound of Lemma 17 along with the bound

$$h(\delta) \le \delta\log(e/\delta) = \frac{1}{m^2n^2}\log(em^2n^2) \lesssim U\sqrt{\frac{m}{n}}\frac{(t+\log\log_2(2n))\log(mn)}{\sqrt{\log(2m)}},$$

we easily get that (37) holds. ■

3.2 Oracle Inequalities for Trace Regression with Gaussian Noise

In this subsection, we establish oracle inequalities for the von Neumann entropy penalized least squares estimator $\tilde{\rho}^\varepsilon$ in the case of trace regression model with Gaussian noise (Assumption 4). Unlike in the case of Theorem 15 of the previous section, our aim is not to obtain sharp oracle inequality, but rather to get a clean main term of the random error bound part of the inequality, namely, the term $\sigma_\xi^2\frac{\operatorname{rank}(S)m(t+\log(2m))}{n}$ in inequality (58) below. Note that this term depends only on the variance of the noise σ_ξ^2, but not on the constant U from Assumption 1 (the constant U is involved only in the higher order $O(n^{-2})$ terms of the bound). Note also that there are no constraints on the variance σ_ξ^2 that could be arbitrarily small, or even equal to 0 (in which case only higher order terms are present in the bound). This improvement comes at a price of having the leading constant 2 in the oracle inequality and also of imposing assumption (57) that requires the regularization parameter ε to be bounded away from 0 (again, unlike Theorem 15, where it could be arbitrarily small). As in the previous section, we also obtain a bound on Kullback–Leibler divergence $K(\rho\|\tilde{\rho}^\varepsilon)$.

Theorem 19 *Let $t \ge 1$. Suppose*

$$\varepsilon \in \left[DU^2\frac{t+\log^3 m\log^2 n}{n}, \frac{D_1\sigma_\xi}{\log(mn)}\sqrt{\frac{t+\log(2m)}{nm}}\bigvee DU^2\frac{t+\log^3 m\log^2 n}{n}\right] \tag{57}$$

with large enough constants $D, D_1 > 0$. There exists a constant $C > 0$ such that with probability at least $1-e^{-t}$

$$\|f_{\tilde{\rho}^\varepsilon} - f_\rho\|_{L_2(\Pi)}^2 \le \inf_{S\in\mathcal{S}_m}\left[2\|f_S - f_\rho\|_{L_2(\Pi)}^2 + C\left(\sigma_\xi^2\frac{\operatorname{rank}(S)m(t+\log(2m))}{n}\right.\right. \\ \left.\left. +\sigma_\xi^2U^2\frac{\operatorname{rank}(S)m^2(t+\log(2m))^2\log(2m)}{n^2} + U^4\frac{\operatorname{rank}(S)m^2(t+\log^3 m\log^2 n)^2\log^2(mn)}{n^2}\right)\right]. \tag{58}$$

In particular,

$$\|f_{\tilde\rho^\varepsilon} - f_\rho\|_{L_2(\Pi)}^2 \le C\Big[\sigma_\xi^2 \tfrac{\mathrm{rank}(\rho)m(t+\log(2m))}{n} \tag{59}$$

$$+\sigma_\xi^2 U^2 \tfrac{\mathrm{rank}(\rho)m^2(t+\log(2m))^2\log(2m)}{n^2} + U^4 \tfrac{\mathrm{rank}(\rho)m^2(t+\log^3 m\log^2 n)^2\log^2(mn)}{n^2}\Big].$$

Moreover, if

$$\varepsilon := \frac{D_1\sigma_\xi}{\log(mn)}\sqrt{\frac{t+\log(2m)}{nm}}\bigvee DU^2\frac{t+\log^3 m\log^2 n}{n}$$

for large enough constants D, D_1, then with some constant C and with the same probability both (59) and the following bound hold:

$$K(\rho\|\tilde\rho^\varepsilon) \le C\Big[\sigma_\xi \tfrac{\mathrm{rank}(\rho)m^{3/2}(t+\log(2m))^{1/2}\log(mn)}{\sqrt{n}} \tag{60}$$

$$+\sigma_\xi^2 \tfrac{\mathrm{rank}(\rho)m^2(t+\log(2m))\log(2m)}{n} + U^2\tfrac{\mathrm{rank}(\rho)m^2(t+\log^3 m\log^2 n)\log^2(mn)}{n}\Big].$$

Proof As in in the proof of Theorem 15, we rely on Lemma 16, but we use a different approach to bounding the empirical process $(P-P_n)(\ell'\bullet f_{\tilde\rho})(f_{\tilde\rho}-f_S)$. The following identity follows from the definition of quadratic loss ℓ

$$(\ell'\bullet f)(x,y)(f(x)-f_S(x)) = 2(f(x)-f_S(x))^2 + 2(f_S(x)-y)(f(x)-f_S(x))$$

and it implies that

$$(P-P_n)(\ell'\bullet f_{\tilde\rho})(f_{\tilde\rho}-f_S) = -2(P_n-P)(f_{\tilde\rho}-f_S)^2 - 2\langle\Xi,\tilde\rho - S\rangle \tag{61}$$

where

$$\Xi := n^{-1}\sum_{j=1}^n (f_S(X_j)-Y_j)X_j - \mathbb{E}(f_S(X)-Y)X.$$

We will bound $(P_n-P)(f_{\tilde\rho}-f_S)^2$ in representation (61) as follows:

$$\Big|(P_n-P)(f_{\tilde\rho}-f_S)^2\Big| \le \|\tilde\rho - S\|_1^2\beta_n\bigg(\frac{\|f_{\tilde\rho}-f_S\|_{L_2(\Pi)}}{\|\tilde\rho-S\|_1}\bigg), \tag{62}$$

where

$$\beta_n(\Delta) := \sup\Big\{\Big|(P_n-P)(f_A^2)\Big| : A\in\mathbb{H}_m, \|A\|_1\le 1, \|f_A\|_{L_2(\Pi)}\le\Delta\Big\}.$$

The next lemma provides a bound on $\beta_n(\Delta)$. Its proof is somewhat involved and it will not be given here. It is based on Rudelson's $L_\infty(P_n)$ generic chaining bound for empirical processes indexed by squares of functions and on the ideas of the paper by Guédon et al. (2008) combined with Talagrand's concentration inequality (see also Aubrun 2009, Liu 2011 and Theorem 3.16, Lemma 9.8 and Proposition 9.2 in Koltchinskii 2011b for similar arguments).

Lemma 20 *Given $0 < \delta^- < \delta^+$ and $t \geq 1$, let*

$$\bar{t} := t + \log\Bigl(\log_2(\delta^+/\delta^-) + 3\Bigr).$$

Then, with some constant C and with probability at least $1 - e^{-t}$, the following bound holds for all $\Delta \in [\delta^-, \delta^+]$:

$$\beta_n(\Delta) \leq C\Bigl[\Delta U \frac{\log^{3/2} m \log n}{\sqrt{n}} + U^2 \frac{\log^3 m \log^2 n}{n} + \Delta U \sqrt{\frac{\bar{t}}{n}} + U^2 \frac{\bar{t}}{n}\Bigr]. \tag{63}$$

We will use Lemma 20 to control $\beta_n(\Delta)$ for $\Delta := \frac{\|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)}}{\|\tilde{\rho} - S\|_1}$. Let $\delta^+ := \frac{1}{m}$ and $\delta^- := \frac{1}{mn}$. With this choice, $\bar{t} \leq t + \log(\log_2 n + 3)$. Note that for $A = \frac{\tilde{\rho} - S}{\|\tilde{\rho} - S\|_1}$, $\|f_A\|_{L_2(\Pi)} = \frac{\|A\|_2}{m} \leq \frac{\|A\|_1}{m} = m^{-1} = \delta^+$. If also $\|f_A\|_{L_2(\Pi)} \geq \delta^-$, then we can substitute bound (63) on $\beta_n(\Delta)$ into (62) that yields:

$$\begin{aligned}
\Bigl|(P_n - P)(f_{\tilde{\rho}} - f_S)^2\Bigr| &\leq C\Bigl[\|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)} \|\tilde{\rho} - S\|_1 U \tfrac{\log^{3/2} m \log n}{\sqrt{n}} \\
&\quad + \|\tilde{\rho} - S\|_1^2 U^2 \tfrac{\log^3 m \log^2 n}{n} + \|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)} \|\tilde{\rho} - S\|_1 U \sqrt{\tfrac{\bar{t}}{n}} \\
&\quad + \|\tilde{\rho} - S\|_1^2 U^2 \tfrac{\bar{t}}{n}\Bigr] \\
&\leq \tfrac{1}{32}\|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)}^2 + 8(C^2 + C/8) U^2 \tfrac{\log^3 m \log^2 n}{n} \|\tilde{\rho} - S\|_1^2 \\
&\quad + \tfrac{1}{32}\|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)}^2 + 8(C^2 + C/8) U^2 \tfrac{\bar{t}}{n} \|\tilde{\rho} - S\|_1^2 \\
&\leq \tfrac{1}{16}\|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)}^2 + C' U^2 \tfrac{\log^3 m \log^2 n + \bar{t}}{n} \|\tilde{\rho} - S\|_1^2,
\end{aligned} \tag{64}$$

where $C' := 8(C^2 + C/8)$. If, on the other hand, $\|f_A\|_{L_2(\Pi)} \leq \delta^- = \frac{1}{mn}$, then $\|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)}$ in the above bound can be replaced by $\frac{1}{mn}\|\tilde{\rho} - S\|_1$ and the proof that follows only simplifies since

$$\frac{1}{16}\|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)}^2 \leq \frac{1}{16}\frac{1}{m^2 n^2}\|\tilde{\rho} - S\|_1^2 \leq \frac{1}{16} U^2 \frac{\log^3 m \log^2 n + \bar{t}}{n} \|\tilde{\rho} - S\|_1^2.$$

Another term in the right hand side of representation (61) to be controlled is $\langle \Xi, \tilde{\rho} - S\rangle$. Note that $\Xi = \Xi_1 + \Xi_2$, where

$$\Xi_1 := -n^{-1} \sum_{j=1}^n \xi_j X_j$$

and

$$\Xi_2 := n^{-1} \sum_{j=1}^n (f_S(X_j) - f_\rho(X_j)) X_j - \mathbb{E}(f_S(X) - f_\rho(X))X.$$

Recall that $S = (1-\delta)S' + \delta \frac{I_m}{m}$ with $S' \in \mathcal{S}_m$, $\operatorname{rank}(S') = r$, $\operatorname{supp}(S') = L$ and $\delta = \frac{1}{m^2 n^2}$.

The term with Ξ_1 is controlled as follows:

$$\begin{aligned}
&\Big|\langle \Xi_1, \tilde{\rho} - S\rangle\Big| \\
&\le \Big|\langle \mathcal{P}_L(\Xi_1), \tilde{\rho} - S'\rangle\Big| + \Big|\langle \Xi_1, \mathcal{P}_L^{\perp}(\tilde{\rho} - S')\rangle\Big| + \Big|\langle \mathcal{P}_L^{\perp}(\Xi_1), S' - S\rangle\Big| \\
&\le \|\mathcal{P}_L(\Xi_1)\|_2\|\tilde{\rho} - S'\|_2 + \|\Xi_1\|_\infty\|\mathcal{P}_L^{\perp}(\tilde{\rho})\|_1 + \Big\|\mathcal{P}_L^{\perp}(\Xi_1)\Big\|_\infty\|S' - S\|_1 \\
&\le 2\sqrt{2r}m\|\Xi_1\|_\infty\|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)} + \|\Xi_1\|_\infty\|\mathcal{P}_L^{\perp}(\tilde{\rho})\|_1 + 4\delta\|\Xi_1\|_\infty \\
&\qquad \le 32rm^2\|\Xi_1\|_\infty^2 + \tfrac{1}{16}\|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)}^2 \\
&\qquad +\|\Xi_1\|_\infty\|\mathcal{P}_L^{\perp}(\tilde{\rho})\|_1 + 4\delta\|\Xi_1\|_\infty.
\end{aligned} \tag{65}$$

We also have

$$\begin{aligned}
\Big|\langle \Xi_2, \tilde{\rho} - S\rangle\Big| &\le \|\Xi_2\|_\infty\|\tilde{\rho} - S\|_1 \le \|\Xi_2\|_\infty\|\tilde{\rho} - S'\|_1 + \|\Xi_2\|_\infty\|S' - S\|_1 \\
&\le \|\Xi_2\|_\infty\|\tilde{\rho} - S'\|_1 + 2\delta\|\Xi_2\|_\infty.
\end{aligned} \tag{66}$$

Thus,

$$\begin{aligned}
&\Big|\langle \Xi, \tilde{\rho} - S\rangle\Big| \le 32rm^2\|\Xi_1\|_\infty^2 + \tfrac{1}{16}\|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)}^2 \\
&+\|\Xi_1\|_\infty\|\mathcal{P}_L^{\perp}(\tilde{\rho})\|_1 + 4\delta\|\Xi_1\|_\infty + \|\Xi_2\|_\infty\|\tilde{\rho} - S'\|_1 + 2\delta\|\Xi_2\|_\infty.
\end{aligned} \tag{67}$$

It follows from (61), (64) and (67) that with some constant C'

$$\begin{aligned}
&(P - P_n)(\ell' \bullet f_{\tilde{\rho}})(f_{\tilde{\rho}} - f_S) \le \\
&\tfrac{1}{4}\|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)}^2 + C'U^2\tfrac{\log^3 m \log^2 n + \bar{t}}{n}\|\tilde{\rho} - S\|_1^2 \\
&+64rm^2\|\Xi_1\|_\infty^2 + 2\|\Xi_1\|_\infty\|\mathcal{P}_L^{\perp}(\tilde{\rho})\|_1 + 8\delta\|\Xi_1\|_\infty \\
&+2\|\Xi_2\|_\infty\|\tilde{\rho} - S'\|_1 + 4\delta\|\Xi_2\|_\infty.
\end{aligned} \tag{68}$$

This bound will be substituted in (38). Note that, if assumption (57) on ε holds with a sufficiently large constant D, then we have

$$\varepsilon \ge 8C'U^2\frac{\log^3 m \log^2 n + \bar{t}}{n}$$

(this follows from the fact that $\bar{t} \le t + \log(\log_2 n + 3) \le t + c\log^3 m \log^2 n$ for some constant $c > 0$). Assume also that $\bar{\varepsilon} \ge 4\|\Xi_1\|_\infty$ and recall that $K(\tilde{\rho}; S) \ge \frac{1}{4}\|\tilde{\rho} - S\|_1^2$ (see inequality 8). Taking all this into account, (38) implies that

$$\begin{aligned}
&\|f_{\tilde{\rho}} - f_\rho\|_{L_2(\Pi)}^2 + \tfrac{1}{4}\|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)}^2 + \tfrac{\varepsilon}{2}K(\tilde{\rho}; S) + \tfrac{\bar{\varepsilon}}{2}\|\mathcal{P}_L^{\perp}\tilde{\rho}\|_1 \\
&\le \|f_S - f_\rho\|_{L_2(\Pi)}^2 + rm^2\varepsilon^2\log^2(m/\delta) + 5rm^2\bar{\varepsilon}^2 + 6\bar{\varepsilon}\delta \\
&+2\|\Xi_2\|_\infty\|\tilde{\rho} - S'\|_1 + 4\|\Xi_2\|_\infty\delta.
\end{aligned} \tag{69}$$

It remains to control $\|\Xi_1\|_\infty$ and $\|\Xi_2\|_\infty$. To this end, we use matrix versions of Bernstein inequality. To bound $\|\Xi_2\|_\infty$, we use its standard version which yields that with probability at least $1 - e^{-t}$

$$\|\Xi_2\|_\infty \leq 2\biggl[\Bigl\|\mathbb{E}(f_S(X) - f_\rho(X))^2 X^2\Bigr\|_\infty^{1/2} \sqrt{\tfrac{t+\log(2m)}{n}}$$
$$\bigvee \Bigl\|(f_S(X) - f_\rho(X))\|X\|_\infty\Bigr\|_{L_\infty} \tfrac{t+\log(2m)}{n}\biggr],$$

where $\|\cdot\|_{L_\infty}$ denotes the essential supremum norm in the space of random variables. Since

$$\Bigl\|\mathbb{E}(f_S(X) - f_\rho(X))^2 X^2\Bigr\|_\infty \leq U^2\|f_S - f_\rho\|^2_{L_2(\Pi)}$$

and

$$\Bigl\|(f_S(X) - f_\rho(X))\|X\|_\infty\Bigr\|_{L_\infty} \leq 2U^2,$$

we get

$$\|\Xi_2\|_\infty \leq 4\biggl[\|f_S - f_\rho\|_{L_2(\Pi)} U\sqrt{\tfrac{t+\log(2m)}{n}} + U^2\tfrac{t+\log(2m)}{n}\biggr]. \tag{70}$$

This implies that

$$\begin{aligned}2\|\Xi_2\|_\infty\|\tilde{\rho} - S'\|_1 \leq \|f_S - f_\rho\|^2_{L_2(\Pi)} + 16U^2\tfrac{t+\log(2m)}{n}\|\tilde{\rho} - S'\|_1^2 \\ +8U^2\tfrac{t+\log(2m)}{n}\|\tilde{\rho} - S'\|_1.\end{aligned} \tag{71}$$

Note that

$$\begin{aligned}&16U^2\tfrac{t+\log(2m)}{n}\|\tilde{\rho} - S'\|_1^2 \\ &\leq 16U^2\tfrac{t+\log(2m)}{n}\|\tilde{\rho} - S\|_1^2 + 16U^2\tfrac{t+\log(2m)}{n}(4\delta + \delta^2)\end{aligned} \tag{72}$$

and

$$\begin{aligned}&8U^2\tfrac{t+\log(2m)}{n}\|\tilde{\rho} - S'\|_1 \\ &\leq 8U^2\tfrac{t+\log(2m)}{n}\|\mathcal{P}_L^\perp\tilde{\rho}\|_1 + 8U^2\tfrac{t+\log(2m)}{n}\|\mathcal{P}_L(\tilde{\rho} - S')\|_1 \\ &\leq 8U^2\tfrac{t+\log(2m)}{n}\|\mathcal{P}_L^\perp\tilde{\rho}\|_1 + 8U^2\tfrac{t+\log(2m)}{n}\|\mathcal{P}_L(\tilde{\rho} - S)\|_1 + 16U^2\tfrac{t+\log(2m)}{n}\delta.\end{aligned} \tag{73}$$

Since, for some constant $C'' > 0$,

$$\begin{aligned}&8U^2\tfrac{t+\log(2m)}{n}\|\mathcal{P}_L(\tilde{\rho} - S)\|_1 \leq 8\sqrt{2}U^2\tfrac{t+\log(2m)}{n}\sqrt{r}\|\mathcal{P}_L(\tilde{\rho} - S)\|_2 \\ &\leq 8\sqrt{2}U^2\tfrac{t+\log(2m)}{n}\sqrt{r}m\|f_{\tilde{\rho}} - f_S\|_{L_2(\Pi)} \leq \tfrac{1}{4}\|f_{\tilde{\rho}} - f_S\|^2_{L_2(\Pi)} + C''U^4\tfrac{rm^2(t+\log(2m))^2}{n^2},\end{aligned}$$

it follows from (71), (72) and (73) that

$$\begin{aligned}2\|\Xi_2\|_\infty\|\tilde{\rho} - S'\|_1 \leq \|f_S - f_\rho\|^2_{L_2(\Pi)} + \\ +16U^2\tfrac{t+\log(2m)}{n}\|\tilde{\rho} - S\|_1^2 + 16U^2\tfrac{t+\log(2m)}{n}(4\delta + \delta^2) \\ +8U^2\tfrac{t+\log(2m)}{n}\|\mathcal{P}_L^\perp\tilde{\rho}\|_1 + 16U^2\tfrac{t+\log(2m)}{n}\delta \\ +\tfrac{1}{4}\|f_{\tilde{\rho}} - f_S\|^2_{L_2(\Pi)} + C''U^4\tfrac{rm^2(t+\log(2m))^2}{n^2}.\end{aligned} \tag{74}$$

Note that (70) also implies that

$$\|\Xi_2\|_\infty \le 4\Bigl[\tfrac{2U}{m}\sqrt{\tfrac{t+\log(2m)}{n}} + U^2\tfrac{t+\log(2m)}{n}\Bigr] \tag{75}$$

(since $\|f_S - f_\rho\|_{L_2(\Pi)} \le m^{-1}\|S-\rho\|_2 \le 2m^{-1}$). Let us substitute (74) and (75) in the last line of (69). Assume that

$$\bar\varepsilon \ge 16U^2\frac{t+\log(2m)}{n}$$

and that constant D in assumption (57) is large enough so that

$$16U^2\frac{t+\log(2m)}{n}\|\tilde\rho - S\|_1^2 \le \frac{\varepsilon}{4}K(\tilde\rho, S)$$

(recall inequality 8). It easily follows that with some constants C_1, C_2,

$$\begin{aligned} &\|f_{\tilde\rho} - f_\rho\|_{L_2(\Pi)}^2 + \tfrac{\varepsilon}{4}K(\tilde\rho; S) \\ &\le 2\|f_S - f_\rho\|_{L_2(\Pi)}^2 + C_1 rm^2\varepsilon^2\log^2(m/\delta) + 5rm^2\bar\varepsilon^2 \\ &+C_2\bar\varepsilon\delta + 32\tfrac{U}{m}\sqrt{\tfrac{t+\log(2m)}{n}}\delta \end{aligned} \tag{76}$$

(note that the term $C''U^4\frac{rm^2(t+\log(2m))^2}{n^2}$ of bound (74) is "absorbed" by the term $C_1 rm^2\varepsilon^2\log^2(m/\delta)$ of bound (76) provided that constant C_1 is large enough). Since

$$\delta = \frac{1}{m^2n^2} \le U^2\frac{t+\log(2m)}{n} \le \bar\varepsilon$$

(recall that $U^2 \ge m^{-1}$), we have $\bar\varepsilon\delta \le \bar\varepsilon^2$. Also, since $U \ge m^{-1/2}$,

$$\frac{U}{m}\sqrt{\frac{t+\log(2m)}{n}}\delta = U\sqrt{\frac{t+\log(2m)}{n}}\frac{1}{m^3n^2} \le U^4\Bigl(\frac{t+\log(2m)}{n}\Bigr)^2 \le \bar\varepsilon^2.$$

Therefore, (76) implies that with some constant C

$$\begin{aligned} &\|f_{\tilde\rho} - f_\rho\|_{L_2(\Pi)}^2 + \tfrac{\varepsilon}{4}K(\tilde\rho; S) \\ &\le 2\|f_S - f_\rho\|_{L_2(\Pi)}^2 + C\Bigl(rm^2\varepsilon^2\log^2(m/\delta) + rm^2\bar\varepsilon^2\Bigr). \end{aligned} \tag{77}$$

To bound $\|\Xi_1\|_\infty$, we use a version of matrix Bernstein type inequality due to Koltchinskii (2011b) (see bound (2.7) of Theorem 2.7). Its version for $\alpha = 2$ (with $U^{(\alpha)} \asymp U\sigma_\xi$) implies that for some constant $K > 0$ with probability at least $1 - e^{-t}$

$$\|\Xi_1\|_\infty \le K\Bigl[\sigma_\xi\sqrt{\frac{t+\log(2m)}{nm}} \bigvee \sigma_\xi U\frac{(t+\log(2m))\log^{1/2}(2Um^{1/2})}{n}\Bigr]. \tag{78}$$

We choose

$$\bar\varepsilon := D_2\Bigl[\sigma_\xi\sqrt{\frac{t+\log(2m)}{nm}} \bigvee (\sigma_\xi \vee U)U\frac{(t+\log(2m))\log^{1/2}(2m)}{n}\Bigr]$$

with a sufficiently large constant D_2 to satisfy the condition $\|\Xi_1\|_\infty \le 4\bar\varepsilon$ with probability at least $1-e^{-t}$ (the rest of the assumptions we made on $\bar\varepsilon$ are also satisfied with this choice).

Bound (77) then implies that with some constant C and with probability at least $1-3e^{-t}$ the following inequality holds:

$$\begin{aligned}&\|f_{\tilde\rho^\varepsilon}-f_\rho\|^2_{L_2(\Pi)} \le 2\|f_S-f_\rho\|^2_{L_2(\Pi)}\\ &+C\Big[\sigma_\xi^2\frac{rm(t+\log(2m))}{n}+\sigma_\xi^2U^2\frac{rm^2(t+\log(2m))^2\log(2m)}{n^2}\\ &+U^4\frac{rm^2(t+\log^3 m\log^2 n)^2\log^2(mn)}{n^2}\Big].\end{aligned}\tag{79}$$

Using bound (39) to replace S in $\|f_S-f_\rho\|^2_{L_2(\Pi)}$ with S' and adjusting the value of constant C to rewrite the probability bound as $1-e^{-t}$, it is easy to complete the proof of (58). If $S'=\rho$, this also yields bound (59). Moreover, with a larger value of regularization parameter

$$\varepsilon := \frac{D_1\sigma_\xi}{\log(mn)}\sqrt{\frac{t+\log(2m)}{nm}}\bigvee DU^2\frac{t+\log^3 m\log^2 n}{n},$$

bound (77) and Lemma 17 easily imply bound (60). ■

3.3 Optimality Properties of von Neumann Entropy Penalized Estimator $\tilde\rho^\epsilon$

We start with upper bounds on the error of estimator $\tilde\rho^\epsilon$ (von Neumann entropy penalized least squares estimator defined by (7)) in Hellinger, Kullback-Leibler and Schatten q-norm distances for $q\in[1,2]$ for the trace regression model with Gaussian noise (Assumption 4). To avoid the impact of "second order terms" on the upper bounds, we will make the following simplifying assumptions:

$$U\sqrt{\frac{m}{n}}\log m \lesssim 1 \ \text{ and } \ U^2\sqrt{\frac{m}{n}}\log^{5/2} m\log^2 n\log(mn)\lesssim \sigma_\xi.\tag{80}$$

Recall that, for the Pauli basis, $U=m^{-1/2}$, so, the above assumptions hold if $n\gtrsim \log^2 m$ and σ_ξ is larger than $\frac{1}{\sqrt{mn}}$ (times a logarithmic factor). We will choose regularization parameter ε as follows:

$$\varepsilon := \frac{D_1\sigma_\xi}{\log(mn)}\sqrt{\frac{\log(2m)}{nm}}\tag{81}$$

with a sufficiently large constant $D_1>0$. The next result shows that minimax rates of Theorem 4 are attained up to logarithmic factors for the estimator $\tilde\rho^\varepsilon$.

Theorem 21 *There exists a constant $C>0$ such that the following bounds hold for all $r=1,\dots,m$, for all $\rho\in\mathcal{S}_{r,m}$ and for all $q\in[1,2]$ with probability at least $1-m^{-2}$:*

$$\|\tilde\rho^\varepsilon-\rho\|_q\le C\Big(\frac{\sigma_\xi m^{\frac32}r^{1/q}}{\sqrt n}\sqrt{\log m}\log^{(2-q)/q}(mn)\bigwedge\Big(\frac{\sigma_\xi m^{3/2}}{\sqrt n}\Big)^{1-\frac1q}(\log m)^{\frac12-\frac1{2q}}\Big)\bigwedge 2,\tag{82}$$

$$H^2(\tilde{\rho}^{\varepsilon},\rho)\le C\frac{\sigma_{\xi}m^{\frac{3}{2}}r}{\sqrt{n}}\sqrt{\log m}\log(mn)\bigwedge 2 \tag{83}$$

and

$$K(\rho\|\tilde{\rho}^{\varepsilon})\le C\frac{\sigma_{\xi}m^{\frac{3}{2}}r}{\sqrt{n}}\sqrt{\log m}\log(mn). \tag{84}$$

Proof We will need the following simple lemma.

Lemma 22 *For all $\rho\in\mathcal{S}_m$ and all $l=1,\dots,m$, there exists $\rho'\in\mathcal{S}_{l,m}$ such that*

$$\|\rho-\rho'\|_2^2\le\frac{1}{l}.$$

Proof Suppose that $\rho=\sum_{j=1}^m\lambda_jP_j$, where λ_j are the eigenvalues of ρ repeated with their multiplicities and P_j are orthogonal one-dimensional projectors. Note that $\{\lambda_j:j=1,\dots,m\}$ is a probability distribution on the set $\{1,\dots,m\}$. Let ν be a random variable sampled from this distribution and $\nu_1,\dots,\nu_l$ be its i.i.d. copies. Then $\mathbb{E}P_\nu=\rho$ and

$$\mathbb{E}\Bigl\|l^{-1}\sum_{j=1}^l P_{\nu_j}-\rho\Bigr\|_2^2=\frac{\mathbb{E}\|P_\nu-\rho\|_2^2}{l}=\frac{\mathbb{E}\|P_\nu\|_2^2-\|\rho\|_2^2}{l}=\frac{1-\|\rho\|_2^2}{l}\le\frac{1}{l}.$$

Therefore, there exists a realization $\nu_1=k_1,\dots,\nu_l=k_l$ of r.v. $\nu_1,\dots,\nu_l$ such that

$$\Bigl\|l^{-1}\sum_{j=1}^l P_{k_j}-\rho\Bigr\|_2^2\le\frac{1}{l}.$$

Denote $\rho':=l^{-1}\sum_{j=1}^l P_{k_j}$. Then, $\rho'\in\mathcal{S}_{l,m}$ and $\|\rho-\rho'\|_2^2\le\frac{1}{l}$. ■

First, we will prove bound (82) for $q=2$. To this end, we use oracle inequality (58) with $t=2\log m+\log 2$ and with oracle $S=\rho'\in\mathcal{S}_{l,m}$ such that $\|\rho-\rho'\|_2^2\le\frac{1}{l}$. Under simplifying assumptions (80) it yields that with probability at least $1-\frac{1}{2}m^{-2}$

$$\|\tilde{\rho}^{\varepsilon}-\rho\|_2^2=m^2\|f_{\tilde{\rho}^{\varepsilon}}-f_\rho\|_{L_2(\Pi)}^2\lesssim\Bigl[\frac{1}{l}+\tau^2l\log m\Bigr],$$

where $\tau:=\frac{\sigma_\xi m^{3/2}}{\sqrt{n}}$. On the other hand, using the same inequality with $S=\rho\in\mathcal{S}_{r,m}$ yields the bound

$$\|\tilde{\rho}^{\varepsilon}-\rho\|_2^2\lesssim\tau^2r\log m$$

that also holds with probability at least $1-\frac{1}{2}m^{-2}$. Therefore, with probability at least $1-m^{-2}$

$$\|\tilde{\rho}^{\varepsilon}-\rho\|_2^2\lesssim\Bigl(\frac{1}{l}+\tau^2l\log m\Bigr)\bigwedge\tau^2r\log m. \tag{85}$$

Let $\bar{l}=\frac{1}{\tau\sqrt{\log m}}$. If $\bar{l}\in[1,m]$, set $l:=[\bar{l}]$. Otherwise, if $\bar{l}>m$, set $l:=m$ and, if $\bar{l}<1$, set $l:=1$. An easy computation shows that with such a choice of l bound (85) implies (82) for $q=2$.

Next we use bound (60) that, for $t = 2\log m$, implies under assumptions (80) that with some constant C and with probability at least $1 - m^{-2}$

$$K(\rho\|\tilde{\rho}^{\varepsilon}) \leq C\sigma_{\xi}\frac{rm^{3/2}\sqrt{\log m}\log(mn)}{\sqrt{n}}, \tag{86}$$

which is bound (84). Bound (83) also holds in view of inequality (8).

Now, we prove bound (82) for $q = 1$ (the bound for $q \in [1,2]$ will then follow by interpolation). To this end, we will use the following lemma (see Proposition 1 in Koltchinskii 2011a) that shows that if two density matrices are close in Hellinger distance and one of them is "concentrated around a subspace" L, then another one is also "concentrated around" L.

Lemma 23 *For any $L \subset \mathbb{C}^m$ and all $S_1, S_2 \in \mathcal{S}_m$,*

$$\|\mathcal{P}_L^{\perp}S_1\|_1 \leq 2\|\mathcal{P}_L^{\perp}S_2\|_1 + 2H^2(S_1, S_2).$$

We apply this lemma to $S_1 = \tilde{\rho}^{\varepsilon}$, $S_2 = \rho$ and $L = \mathrm{supp}(\rho)$ so that $\mathcal{P}_L^{\perp}\rho = 0$. It yields that

$$\|\mathcal{P}_L^{\perp}\tilde{\rho}^{\varepsilon}\|_1 \leq 2H^2(\tilde{\rho}^{\varepsilon}, \rho).$$

Therefore,

$$\|\tilde{\rho}^{\varepsilon}-\rho\|_1 \leq \|\mathcal{P}_L(\tilde{\rho}^{\varepsilon}-\rho)\|_1+\|\mathcal{P}_L^{\perp}(\tilde{\rho}^{\varepsilon}-\rho)\|_1 \leq \sqrt{2r}\|\tilde{\rho}^{\varepsilon}-\rho\|_2+\|\mathcal{P}_L^{\perp}\tilde{\rho}^{\varepsilon}\|_1 \leq \sqrt{2r}\|\tilde{\rho}^{\varepsilon}-\rho\|_2+2H^2(\tilde{\rho}^{\varepsilon},\rho). \tag{87}$$

Using bounds (82) for $q = 2$ and (83), we get from (87) that

$$\|\tilde{\rho}^{\varepsilon} - \rho\|_1 \leq C\frac{\sigma_{\xi}m^{\frac{3}{2}}r}{\sqrt{n}}\sqrt{\log m}\log(mn) \bigwedge 2, \tag{88}$$

which is equivalent to (82) for $q = 1$. Note that by choosing $t = 2\log m + \log 2 + 2$ (which might have an impact only on the constant), we could make probability bounds in (82) for $q = 2$ and (83) to be at least $1 - \frac{1}{2}m^{-2}$ implying that (88) holds with probability at least $1 - m^{-2}$, as it is claimed in the theorem.

To complete the proof, it is enough to use the interpolation inequality of Lemma 1. It follows that, for $q \in (1,2)$,

$$\|\tilde{\rho}^{\varepsilon} - \rho\|_q \leq \|\tilde{\rho}^{\varepsilon} - \rho\|_1^{\frac{2}{q}-1}\|\tilde{\rho}^{\varepsilon} - \rho\|_2^{2-\frac{2}{q}}.$$

Substituting bound (82) for $q = 1$ and $q = 2$ into the last inequality yields the result for an arbitrary $q \in (1,2)$. ■

Similarly, in the case of trace regression with bounded response (see Assumption 3), minimax rates of Theorem 7 are also attained for the estimator $\tilde{\rho}^{\varepsilon}$ (up to log factors). In this case, assume that Assumption 3 holds with $\bar{U} = U$ and, in addition, let us make the following simplifying assumptions:

$$U\sqrt{\frac{m\log m}{n}} \lesssim 1 \quad \text{and} \quad \log\log_2 n \lesssim m\log m. \tag{89}$$

For the Pauli basis ($U = m^{-1/2}$), the first assumption holds if $n \gtrsim \log m$. The second assumption does hold unless n is extremely large ($n \sim 2^{\exp\{m \log m\}}$). Under these assumptions, we will use the following value of regularization parameter ε :

$$\varepsilon := \frac{U}{\log(mn)}\sqrt{\frac{\log(2m)}{nm}}.$$

The following version of Theorem 21 holds in the bounded regression case (with a similar proof).

Theorem 24 *There exists a constant $C > 0$ such that the following bounds hold for all $r = 1, \dots, m$, for all $\rho \in \mathcal{S}_{r,m}$ and for all $q \in [1, 2]$ with probability at least $1 - m^{-2}$:*

$$\|\tilde{\rho}^{\varepsilon} - \rho\|_q \leq C\Big(\frac{Um^{\frac{3}{2}}r^{1/q}}{\sqrt{n}}\sqrt{\log m}\log^{(2-q)/q}(mn)\bigwedge\Big(\frac{Um^{3/2}}{\sqrt{n}}\Big)^{1-\frac{1}{q}}(\log m)^{\frac{1}{2}-\frac{1}{2q}}\Big)\bigwedge 2, \tag{90}$$

$$H^2(\tilde{\rho}^{\varepsilon}, \rho) \leq C\frac{Um^{\frac{3}{2}}r}{\sqrt{n}}\sqrt{\log m}\log(mn)\bigwedge 2 \tag{91}$$

and

$$K(\rho\|\tilde{\rho}^{\varepsilon}) \leq C\frac{Um^{\frac{3}{2}}r}{\sqrt{n}}\sqrt{\log m}\log(mn). \tag{92}$$

Remark 25 *In the case of Pauli basis, the minimax optimal rates (up to constants and logarithmic factors) are: $\frac{mr^{1/q}}{\sqrt{n}} \wedge (\frac{m}{\sqrt{n}})^{1-\frac{1}{q}} \wedge 2$ for Schatten q-norm distances for $q \in [1, 2]$; $\frac{mr}{\sqrt{n}}$ for nuclear norm, squared Hellinger and Kullback-Leibler distances (provided the $mr \lesssim \sqrt{n}$).*

References

Jean-Pierre Aubin and Ivar Ekeland. *Applied Nonlinear Analysis*. Courier Corporation, 2006.

Guillaume Aubrun. On almost randomizing channels with a short Kraus decomposition. *Communications in Mathematical Physics*, 288(3):1103–1116, 2009.

Vladimír Bužek. Quantum tomography from incomplete data via maxent principle. In *Quantum State Estimation*, pages 189–234. Springer, 2004.

Tony Cai, Donggyu Kim, Yazhen Wang, Ming Yuan, and Harrison H Zhou. Optimal large-scale quantum state tomography with Pauli measurements. `http://www-stat.wharton.upenn.edu/~tcai/paper/Estimating-Density-Matrix-Pauli.pdf`, 2015.

Emmanuel J Candés and Yaniv Plan. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory*, 57(4):2342–2359, 2011.

Victor H. de la Peña and Evarist Giné. *Decoupling. From Dependence to Independence.* Springer, 1999.

Steven T Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, 2012.

David Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, 2011.

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.

Olivier Guédon, Shahar Mendelson, Alain Pajor, and Nicole Tomczak-Jaegermann. Majorizing measures and proportional subsets of bounded orthonormal systems. *Revista Matemática Iberoamericana*, 24(3):1075–1095, 2008.

Amir Kalev, Robert L Kosut, and Ivan H Deutsch. Informationally complete measurements from compressed sensing methodology. *arXiv preprint arXiv:1502.00536*, 2015.

Hartmut Klauck, Ashwin Nayak, Amnon Ta-Shma, and David Zuckerman. Interaction in quantum communication. *IEEE Transactions on Information Theory*, 53(6):1970–1982, 2007.

Vladimir Koltchinskii. von Neumann entropy penalization and low-rank matrix estimation. *The Annals of Statistics*, 39(6):2936–2973, 2011a.

Vladimir Koltchinskii. *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: École d'Été de Probabilités de Saint-Flour XXXVIII-2008*. Springer, 2011b.

Vladimir Koltchinskii. A remark on low rank matrix recovery and noncommutative Bernstein type inequalities. In *From Probability to Statistics and Back: High-Dimensional Models and Processes–A Festschrift in Honor of Jon A. Wellner*, pages 213–226. Institute of Mathematical Statistics, 2013a.

Vladimir Koltchinskii. Sharp oracle inequalities in low rank estimation. In *Empirical Inference*, pages 217–230. Springer, 2013b.

Vladimir Koltchinskii and Dong Xia. Schatten p-norm distances in low rank density matrix estimation. 2015+.

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.

Yi-Kai Liu. Universal low-rank matrix recovery from Pauli measurements. In *Advances in Neural Information Processing Systems*, pages 1638–1646, 2011.

Zongming Ma and Yihong Wu. Volume ratio, sparsity, and minimaxity under unitarily invariant norms. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium*, pages 1027–1031. IEEE, 2013.

Sahand Negahban and Martin J. Wainwright. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics*, 2010.

M.A. Nielsen and I.L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

Alain Pajor. Metric entropy of the Grassmann manifold. *Convex Geometric Analysis*, 34: 181–188, 1998.

Angelika Rohde and Alexandre B Tsybakov. Estimation of high-dimensional low-rank matrices. *The Annals of Statistics*, 39(2):887–930, 2011.

Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation.* Springer, 2008.