MULTIPLIER BOOTSTRAP FOR BURES–WASSERSTEIN BARYCENTERS

BY ALEXEY KROSHNIN[1,3], VLADIMIR SPOKOINY[2] AND ALEXANDRA SUVORIKOVA[2,1]

[1]*National Research University Higher School of Economics, akroshnin@hse.ru*

[2]*Weierstrass Institute for Applied Analysis and Stochastics, spokoiny@wias-berlin.de; suvorikova@wias-berlin.de*

[3]*Institute for Information Transmission Problems RAS*

Bures–Wasserstein barycenter is a popular and promising tool in analysis of complex data like graphs, images etc. In many applications the input data are random with an unknown distribution, and uncertainty quantification becomes a crucial issue. This paper offers an approach based on multiplier bootstrap to quantify the error of approximating the true Bures–Wasserstein barycenter Q_* by its empirical counterpart Q_n. The main results state the bootstrap validity under general assumptions on the data generating distribution P and specifies the approximation rates for the case of sub-exponential P. The performance of the method is illustrated on synthetic data generated from the weighted stochastic block model.

1. Introduction. Let $\mathbb{H}_+(d)$ be the space of positive semi-definite d-dimensional Hermitian operators, while $\mathbb{H}_{++}(d)$ is the space of positive definite Hermitian operators. Bhatia et al. [2018] introduced the Bures–Wasserstein distance $d_{\mathrm{B}}(\cdot,\cdot)$ on $\mathbb{H}_+(d)$: for any $S, Q \in \mathbb{H}_+(d)$

$$d_{\mathrm{B}}^2(Q,S) \stackrel{\mathrm{def}}{=} \operatorname{tr} Q + \operatorname{tr} S - 2\operatorname{tr}\left(S^{1/2} Q S^{1/2}\right)^{1/2}. \tag{1.1}$$

This tool appeared to be very useful in many real-world problems with the natural formulation in terms of graphs, covariance matrices, elliptical distribution, etc. One may also refer to Muzellec and Cuturi [2018] or Kroshnin et al. [2021], which discuss the relation of the Bures–Wasserstein distance to optimal transportation metrics.

The current study considers the following statistical setting. Let $S_1, \ldots, S_n$ be an i.i.d. sample with values in $\mathbb{H}_+(d)$, $S_i \stackrel{iid}{\sim} P$. We assume $P\left(\mathbb{H}_{++}(d)\right) > 0$. The population barycenter Q_*, and its empirical counterpart be Q_n are defined as

$$Q_* \stackrel{\mathrm{def}}{=} \operatorname*{argmin}_{Q \in \mathbb{H}_+(d)} \mathbb{E}\, d_{\mathrm{B}}^2(Q,S), \quad Q_n \stackrel{\mathrm{def}}{=} \operatorname*{argmin}_{Q \in \mathbb{H}_+(d)} \frac{1}{n}\sum_{i=1}^n d_{\mathrm{B}}^2(Q,S_i).$$

Theorem 2.1 by Kroshnin et al. [2021] ensures the existence and uniqueness of Q_* and Q_n. We note that unlike the Frobenius mean, the Bures–Wasserstein barycenter is not a linear function of a sample, instead it is the solution of a fixed-point equation:

$$Q_* = \mathbb{E}\left(Q_*^{1/2} S Q_*^{1/2}\right)^{1/2}, \quad Q_n = \frac{1}{n}\sum_{i=1}^n \left(Q_n^{1/2} S_i Q_n^{1/2}\right)^{1/2}. \tag{1.2}$$

For a discussion of computational aspects one may refer to Chewi et al. [2020].

MSC2020 subject classifications: Primary 62F40.

Keywords and phrases: Bures-Wasserstein barycenter, Wasserstein barycenter, multiplier bootstrap, distance between graphs.

1.1. *Motivation and goals.* As a possible practical application, we consider analyzing the brain connectomes. Namely, the connectomes can be described using graphs, where nodes correspond to particular anatomical regions, and edges denote structural or functional connections Bullmore and Bassett [2011]. Given an i.i.d. set, we are interested in performing statistical analysis, such as computing the barycenters, testing the hypothesis of homogeneity, e.t.c. Section 6 provides more details on using Bures-Wasserstein for the solution of practical problems.

The present study aims to answer the question concerning the approximation of the distribution of $\sqrt{n} d_{\mathrm{B}}(Q_n, Q_*)$ by some non-asymptotic data-driven counterpart. In essence, the idea is to use multiplier bootstrap consisting in the random reweighting of the summands in (1.2). Namely, the approximation relies on using a new distribution P_u generating the weights.

In general, since its introduction in the seminal work Efron [1979], the bootstrapping techniques have attracted much attention due to algorithmic simplicity, computational tractability and nice theoretical properties. For instance, Chazal et al. [2014] suggest using the approach for statistical analysis of distributions of persistence diagrams. For more examples, we recommend an excellent survey by Mammen and Nandi [2012], and a brief analysis from the practical point of view by Efron [2000]. Finally, among the recent works on multiplier bootstrap, one may mention Spokoiny and Zhilova [2015], which considers likelihood-based confidence sets. The work by Chen and Zhou [2020] investigates the case of heavy-tailed data. Naumov et al. [2019] validates bootstrap approximation spectral projectors under the assumption of the Gaussianity of the data. We note that the Bures-Wasserstein barycenter is an M-estimators. In this regard, it is worth noting the work of Cheng and Huang [2010], which provides approximation rates for multiplier bootstrap for M-estimators in semi-parametric models. Furthermore, Lee and Yang [2020] propose resampling procedure for M-estimators for nonstandard cases. Finally, it is worth noting that the current study covers the setting of Ebert et al. [2019].

1.2. *Theoretical contribution of the study.* The main result is the validation of construction of a non-asymptotic approximation of the distribution $\sqrt{n}\rho(Q_n, Q_*)$, where ρ denotes either the Bures-Wasserstein distance d_{B}, or the Frobenius norm $\|\cdot\|_F$. We will consider a set of non-negative i.i.d. weights $u_1, \ldots, u_n$, $u_i \overset{\text{iid}}{\sim} P_u$ s.t.

$$\mathbb{E}_u\, u_i = 1, \quad \mathrm{Var}_u\, u_i = 1.$$

Some specific examples are the exponential distribution $u_i \overset{\text{iid}}{\sim} \mathrm{Exp}(1)$, or the Poisson distribution, $u_i \overset{\text{iid}}{\sim} \mathrm{Po}(1)$, which roughly corresponds to sampling with replacement from $\{S_1, \ldots, S_n\}$. Following Agueh and Carlier [2011], we introduce a *reweighted* barycenter Q_u:

$$Q_u \overset{\text{def}}{=} \operatorname*{argmin}_{Q \in \mathbb{H}_+(d)} \sum_i d^2_{BW}(Q, S_i) u_i. \tag{1.3}$$

We refer a reader to the work Kroshnin et al. [2021], which ensures the existence of Q_u, its uniqueness, and measurability. The main theoretical results of the current study are Theorems 4.1 and 4.2. Theorem 4.1 ensures that $\mathcal{L}\left(\sqrt{n}\|Q_n - Q_*\|_F\right)$ is close to $\mathcal{L}\left(\sqrt{n}\|Q_u - Q_n\|_F\right)$ in the Kolmogorov distance. Here and in what follows $\mathcal{L}(X)$ denotes the distribution law of the random variable X. Theorem 4.2 deals with the Bures–Wasserstein case proving that $\mathcal{L}\left(\sqrt{n} d_{\mathrm{B}}(Q_n, Q_*)\right)$ is close to $\mathcal{L}\left(\sqrt{n} d_{\mathrm{B}}(Q_u, Q_n)\right)$.

Along with the main results, Section 2 establishes a connection between the Bures–Wasserstein distance and the Frobenius norm. In particular, we show that for any $Q_1, Q_2 \in$

$\mathbb{H}_{++}(d)$ with $\mathbb{H}_{++}(d)$ being the set of positive-definite Hermitian operators,

$$\left|\frac{d_{\mathrm{B}}(Q_1, Q_2)}{\|\boldsymbol{A}(Q_1)(Q_1 - Q_2)\|_F} - 1\right| \le 2\left\|Q_1^{-1/2} Q_2 Q_1^{-1/2} - I\right\|, \quad \boldsymbol{A}(Q_1) : \mathbb{H}(d) \to \mathbb{H}(d),$$

where $\boldsymbol{A}(Q_1)$ is a linear operator, and $\|\cdot\|$ and $\|\cdot\|_{\mathrm{F}}$ stand for the operator norm and the Frobenius norm, respectively.

1.3. *Proofs in a nutshell.* Let us briefly recall the concept of optimal transportation maps (sometimes referred to as optimal push-forwards). Given two points $Q, S \in \mathbb{H}_+(d)$ the optimal push-forward from Q to S is written as

$$T_Q^S = S^{1/2}\big(S^{1/2} Q S^{1/2}\big)^{-1/2} S^{1/2},$$

by $\big(S^{1/2} Q S^{1/2}\big)^{-1/2}$ we denote the pseudo-inverse matrix $\Big(\big(S^{1/2} Q S^{1/2}\big)^{1/2}\Big)^{+}$. To know more on T_Q^S we recommend Kroshnin et al. [2021], Agueh and Carlier [2011], Brenier [1991]. In particular, Lemma A.2 by Kroshnin et al. [2021] asserts that optimal push-forwards are differentiable in the Fréchet sense, i.e.

$$T_{Q+X}^S = T_Q^S + \boldsymbol{dT}_Q^S(X) + o(\|X\|), \quad X \in \mathbb{H}(d), \tag{1.4}$$

where X is an infinitesimal element of $\mathbb{H}(d)$ with $\mathbb{H}(d)$ being the space of all d-dimensional Hermitian operators. We also note that the linear operator $\boldsymbol{dT}_Q^S(X)$ is self-adjoint and negative definite. Lemma A.3 by Kroshnin et al. [2021] proves the above facts and validates some other properties of $\boldsymbol{dT}_Q^S$ that we will widely use in what follows.

Given a random sample, $S_1, \ldots, S_n$, we look more closely at a random set of optimal push-forwards $T_{Q_*}^{S_i}$ and $T_{Q_n}^{S_i}$. Theorem 2.1 by Kroshnin et al. [2021] claims that $\mathbb{E}\, T_{Q_*}^{S_i} = I$ and $\frac{1}{n}\sum_i T_{Q_n}^{S_i} = I$, where I denotes the identity matrix. From now on, we primarily consider the centred counterparts of $T_{Q_*}^{S_i}$ and $T_{Q_n}^{S_i}$,

$$T_i \overset{\text{def}}{=} T_{Q_*}^{S_i} - I, \quad \hat{T}_i \overset{\text{def}}{=} T_{Q_n}^{S_i} - I. \tag{1.5}$$

The covariance of T_i and its empirical counterpart are written as

$$\boldsymbol{\Sigma} \overset{\text{def}}{=} \mathbb{E}\, T_i \otimes T_i, \quad \boldsymbol{\Sigma_n} \overset{\text{def}}{=} \frac{1}{n}\sum_i T_i \otimes T_i, \tag{1.6}$$

where $\otimes$ stands for the tensor product in $\mathbb{H}(d)$. We also introduce the empirical covariance of $\hat{T}_i$,

$$\hat{\boldsymbol{\Sigma}} \overset{\text{def}}{=} \frac{1}{n}\sum_i \hat{T}_i \otimes \hat{T}_i, \tag{1.7}$$

For T_i and $\hat{T}_i$ (1.4) yields

$$\hat{T}_i \approx T_i + \boldsymbol{dT}_{Q_*}^{S_i}(Q_n - Q_*).$$

For the sake of simplicity, we denote

$$\boldsymbol{dT}_{Q_*}^{S_i} \overset{\text{def}}{=} \boldsymbol{dT}_i, \quad \boldsymbol{dT}_{Q_n}^{S_i} \overset{\text{def}}{=} \boldsymbol{d\hat{T}}_i.$$

A connection between the real and the bootstrap worlds is established by means of "gluing" operators:

$$\boldsymbol{F} \overset{\text{def}}{=} -\mathbb{E}\,\boldsymbol{dT}_i, \quad \boldsymbol{F}_n \overset{\text{def}}{=} -\frac{1}{n}\sum_i \boldsymbol{dT}_i, \tag{1.8}$$

$$\hat{\boldsymbol{F}} \stackrel{\text{def}}{=} -\frac{1}{n}\sum_i \boldsymbol{d}\hat{\boldsymbol{T}}_i, \quad \hat{\boldsymbol{F}}_{\boldsymbol{u}} \stackrel{\text{def}}{=} -\frac{1}{n}\sum_i u_i \boldsymbol{d}\hat{\boldsymbol{T}}_i. \tag{1.9}$$

REMARK 1.1 (Accepted notations). *In the rest of the text, we will often mention the objects coming from "the real world" and the ones coming from "the bootstrap world". The latter is always conditioned on the observed sample and are ticked with a hat sign, e.g. $\hat{\boldsymbol{F}}$, or $\hat{\boldsymbol{F}}_u$. In addition, the subscript "u" of a "hat" object emphasizes its dependency on the random weights $u_1, \dots, u_n$, e.g. $\hat{\boldsymbol{F}}_u$. Consequently, this object is random in both the bootstrap and the real worlds.*

Finally, we define two centred Gaussian vectors Z and $\hat{Z}$,

$$Z \sim \mathcal{N}(0, \Xi), \quad \Xi \stackrel{\text{def}}{=} \boldsymbol{F}^{-1}\boldsymbol{\Sigma}\boldsymbol{F}^{-1}, \tag{1.10}$$

$$\hat{Z} \sim \mathcal{N}(0, \hat{\Xi}), \quad \hat{\Xi} \stackrel{\text{def}}{=} \hat{\boldsymbol{F}}^{-1}\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{F}}^{-1}. \tag{1.11}$$

The proof of the bootstrap validity is based on a chain of approximations presented in Table 1. Each approximation holds under some assumptions on the data generating distribution P and the weight generating distribution P_u. First, we introduce a very general setting in Sections 3 and 4. Then, we discuss how they can be derived from a direct specifications of the data-generating model P, and the weight-generating distribution P_u. Namely, Section 5 illustrates this with an example of P coming from some sub-exponential family, and P_u being sub-Gaussian. For this case we also obtain explicit approximation rates.

$$\sqrt{n}\rho(Q_*, Q_n) \underset{by A.(T), A.(F)}{\approx} \left\| \boldsymbol{L}_\rho \boldsymbol{F}^{-1} \tfrac{1}{\sqrt{n}} \textstyle\sum_i T_i \right\|_F \underset{by A.(Z)}{\overset{\mathrm{d}}{\approx}} \|\boldsymbol{L}_\rho Z\|_F$$

$$\|\boldsymbol{L}_\rho Z\|_F \underset{by A.(\Sigma)}{\overset{\mathrm{d}}{\approx}} \left\|\hat{\boldsymbol{L}}_\rho \hat{Z}\right\|_F$$

$$\sqrt{n}\rho(Q_n, Q_u) \underset{by A.(\hat{T}), A.(\hat{F})}{\approx} \left\| \hat{\boldsymbol{L}}_\rho \hat{\boldsymbol{F}}^{-1} \tfrac{1}{\sqrt{n}} \textstyle\sum_i (u_i - 1)\hat{T}_i \right\|_F \underset{by A.(\hat{Z})}{\overset{\mathrm{d}}{\approx}} \left\|\hat{\boldsymbol{L}}_\rho \hat{Z}\right\|_F$$

TABLE 1

Schematic diagram of the proof of the bootstrap validity. The terms $\boldsymbol{L}_\rho$, $\hat{\boldsymbol{L}}_\rho$ are some linear operators depending on choice of the distance ρ.

2. On geometry of Bures–Wasserstein barycenters.

This section establishes a connection between the Bures–Wasserstein distance and the Frobenius norm, and sets deterministic bounds on the deviation of an empirical weighted barycenter from its population counterpart.

Connection between $\|\cdot\|_{\mathrm{F}}$ and d_{B}. The first lemma quantifies the relation between the Bures–Wasserstein distance and the Frobenius norm. The result is "gluing", as it plays a crucial part in the transition from the case $\rho = \|\cdot\|_{\mathrm{F}}$ to the case $\rho = d_{\mathrm{B}}$. The proof is technical and is postponed to Appendix.

LEMMA 2.1. *Let $Q_0, Q_1, Q_2 \in \mathbb{H}_{++}(d)$. We set*

$$Q_1' = Q_0^{-1/2} Q_1 Q_0^{-1/2}, \quad Q_2' = Q_0^{-1/2} Q_2 Q_0^{-1/2},$$

and assume $\|Q_1' - I\| \le 1/2$ and $\|Q_2' - I\| \le 1/2$. Then

$$\left| \frac{d_{\mathrm{B}}(Q_1, Q_2)}{\|\boldsymbol{A}_0(Q_1 - Q_2)\|_F} - 1 \right| \le 4\left\|Q_1' - I\right\| + 2\left\|Q_2' - I\right\|,$$

with $\boldsymbol{A}_0 = \left(-\frac{1}{2}\boldsymbol{d}\boldsymbol{T}_{Q_0}^{Q_0}\right)^{1/2}$ and $\boldsymbol{d}\boldsymbol{T}_{Q_0}^{Q_0}$ coming from (1.4).

Concentration of barycenters. Let $S_1, \ldots, S_n \in \mathbb{H}_+(d)$ be a fixed set of observations and $w_1, \ldots, w_n$ be a deterministic non-degenerate set of non-negative weights: $w_i \geq 0$ and $\sum_i w_i > 0$. Let

$$Q_w \stackrel{\text{def}}{=} \operatorname*{argmin}_{Q \in \mathbb{H}_+(d)} \sum_i w_i d_{\mathrm{B}}^2(Q, S_i).$$

We fix some $Q_* \in \mathbb{H}_{++}(d)$, which does not necessarily coincide with the population barycenter Q_* (1.2). This trick will allow us to use all the notations from Section 1.3 throughout this section without redefining them. Let

$$\overline{T}_w \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n w_i T_i, \tag{2.1}$$

with T_i coming from (1.5). Generalizing the result by Kroshnin et al. [2021] (see the proof of Theorem 2.2), we write the Taylor expansion in integral form for Q_w in the neighbourhood of Q_*. Let $Q_t \stackrel{\text{def}}{=} tQ_w + (1-t)Q_*$, $t \in [0, 1]$, then

$$Q_w - Q_* = \boldsymbol{D}_w^{-1} \overline{T}_w, \quad \boldsymbol{D}_w \stackrel{\text{def}}{=} -\frac{1}{n} \sum_{i=1}^n w_i \int_0^1 \boldsymbol{dT}_{Q_t}^{S_i} \, dt, \tag{2.2}$$

From now on, we will denote

$$\boldsymbol{F}_w \stackrel{\text{def}}{=} -\frac{1}{n} \sum_i w_i \boldsymbol{dT}_{Q_*}^{S_i}, \quad \hat{\boldsymbol{F}}_w \stackrel{\text{def}}{=} -\frac{1}{n} \sum_i w_i \boldsymbol{dT}_{Q_w}^{S_i}, \tag{2.3}$$

where $\boldsymbol{dT}_{Q_*}^S$ and $\boldsymbol{dT}_{Q_w}^S$ come from (1.4). Lemma A.3 presents technical bounds on $\boldsymbol{D}_w$ and $\boldsymbol{F}_w$. The first result in this section ensures the concentration of Q_w in the vicinity of Q_*. Note that from now on we will denote as $\lambda_{\min}(X)$ and $\lambda_{\max}(X)$ the smallest and the largest eigenvalues of X, respectively.

LEMMA 2.2 (Concentration of Q_w). *Denote*

$$\boldsymbol{\xi}(X) \stackrel{\text{def}}{=} Q_*^{1/2} \boldsymbol{F} \left(Q_*^{1/2} X Q_*^{1/2} \right) Q_*^{1/2}, \quad X \in \mathbb{H}(d).$$

Let $\left\| \boldsymbol{F}^{-1/2} \boldsymbol{F}_w \boldsymbol{F}^{-1/2} - \boldsymbol{I} \right\| \leq \frac{1}{2}$ *and* $\left\| \overline{T}_w \right\|_F \leq \frac{\lambda_{\min}(\boldsymbol{\xi})}{3\|Q_*\|}$. *Then*

$$\left\| Q_*^{-1/2} Q_w Q_*^{-1/2} - I \right\|_F \leq \frac{4\|Q_*\|}{\lambda_{\min}(\boldsymbol{\xi})} \left\| \overline{T}_w \right\|_F.$$

PROOF. Lemma B.1 by Kroshnin et al. [2021] ensures

$$\left\| Q_*^{-1/2} Q_w Q_*^{-1/2} - I \right\|_F \leq \frac{\zeta_w}{1 - \frac{3}{4}\zeta_w},$$

with

$$\zeta_w \stackrel{\text{def}}{=} \frac{1}{\lambda_{\min}(\boldsymbol{\xi}_w)} \left\| Q_*^{1/2} \overline{T}_w Q_*^{1/2} \right\|_F, \quad \boldsymbol{\xi}_w(X) \stackrel{\text{def}}{=} Q_*^{1/2} \boldsymbol{F}_w \left(Q_*^{1/2} X Q_*^{1/2} \right) Q_*^{1/2},$$

provided that $\zeta_w < \frac{4}{3}$. It is easy to see that $\left\| \boldsymbol{F}^{-1/2} \boldsymbol{F}_w \boldsymbol{F}^{-1/2} - \boldsymbol{I} \right\| \leq \frac{1}{2}$ implies the same bound for rescaled operators: $\left\| \boldsymbol{\xi}^{-1/2} \boldsymbol{\xi}_w \boldsymbol{\xi}^{-1/2} - \boldsymbol{I} \right\| \leq \frac{1}{2}$, thus $\lambda_{\min}(\boldsymbol{\xi}_w) \geq \frac{\lambda_{\min}(\boldsymbol{\xi})}{2}$. Therefore, the assumptions of the lemma ensure

$$\zeta_w \leq \frac{2\|Q_*\|}{\lambda_{\min}(\boldsymbol{\xi})} \left\| \overline{T}_w \right\|_F \leq \frac{2}{3},$$

and hence

$$\left\| Q_*^{-1/2} Q_w Q_*^{-1/2} - I \right\|_F \leq 2\zeta_w \leq \frac{4\|Q_*\|}{\lambda_{\min}(\boldsymbol{\xi})} \left\| \overline{T}_w \right\|_F. \qquad \square$$

Relation between OT maps and barycenters. Now we are ready to draw a connection between barycenters and OT maps. Let

$$q_w \stackrel{\text{def}}{=} \left\| Q_*^{-1/2} Q_w Q_*^{-1/2} - I \right\|, \quad f_w \stackrel{\text{def}}{=} \left\| \boldsymbol{F}^{-1/2} \boldsymbol{F}_w \boldsymbol{F}^{-1/2} - I \right\|. \tag{2.4}$$

The quantities q_w and f_w are auxiliary and will appear in the rest of the text as bounding terms. Given an operator or a matrix X, its condition number is defined as

$$\kappa(X) \stackrel{\text{def}}{=} \|X\| \cdot \|X^{-1}\|.$$

We combine these notations and introduce another bounding term

$$\eta_w \stackrel{\text{def}}{=} 2\sqrt{\kappa(\boldsymbol{F})}(q_w + f_w). \tag{2.5}$$

LEMMA 2.3 (Barycenters vs. OT maps: Frobenius norm). *Let $q_w + f_w \le \frac{1}{2}$. Then*

$$\frac{\left\| Q_w - Q_* - \boldsymbol{F}^{-1}\overline{T}_w \right\|_F}{\left\| \boldsymbol{F}^{-1}\overline{T}_w \right\|_F} \le \eta_w.$$

PROOF. We note that the expansion (2.4) yields the following equality

$$Q_w - Q_* - \boldsymbol{F}^{-1}\overline{T}_w = \left(\boldsymbol{D}_w^{-1}\boldsymbol{F} - \boldsymbol{I} \right) \boldsymbol{F}^{-1}\overline{T}_w,$$

therefore, we obtain

$$\frac{\left\| Q_w - Q_* - \boldsymbol{F}^{-1}\overline{T}_w \right\|_F}{\left\| \boldsymbol{F}^{-1}\overline{T}_w \right\|_F} \le \left\| \boldsymbol{D}_w^{-1}\boldsymbol{F} - \boldsymbol{I} \right\|.$$

We use the bound on $\boldsymbol{D}_w$ from Lemma A.3 and get

$$(1 - q_w - f_w)\, \boldsymbol{F}^{-1} \preccurlyeq \boldsymbol{D}_w^{-1} \preccurlyeq (1 + 2q_w + 2f_w)\, \boldsymbol{F}^{-1}.$$

Therefore,

$$\left\| \boldsymbol{D}_w^{-1}\boldsymbol{F} - \boldsymbol{I} \right\| \le \sqrt{\kappa(\boldsymbol{F})} \left\| \boldsymbol{F}^{1/2}\boldsymbol{D}_w^{-1}\boldsymbol{F}^{1/2} - \boldsymbol{I} \right\| \le \eta_w,$$

thus the claim follows. □

A similar result holds for the Bures–Wasserstein distance.

LEMMA 2.4 (Barycenters vs. OT maps: BW-distance). *Let $q_w + f_w \le \frac{1}{2}$. Then*

$$\left| \frac{d_{\mathrm{B}}(Q_w, Q_*)}{\left\| \boldsymbol{A}\boldsymbol{F}^{-1}\overline{T}_w \right\|_F} - 1 \right| \le 3\sqrt{\kappa(Q_*)}\eta_w,$$

with

$$\boldsymbol{A} \stackrel{\text{def}}{=} \left(-\frac{1}{2} \boldsymbol{dT}_{Q_*}^{Q_*} \right)^{1/2}. \tag{2.6}$$

PROOF. We use the result (2.1) from Lemma 2.1 and set $Q_0 = Q_1 = Q_*$, $Q_2 = Q_w$. This ensures

$$\left| \frac{d_{\mathrm{B}}(Q_w, Q_*)}{\|\boldsymbol{A}(Q_w - Q_*)\|_F} - 1 \right| \le 2q_w.$$

It is easy to see that using the triangle inequality and the same line of reasoning as in Lemma 2.3, we get

$$\left|\frac{\|\boldsymbol{A}\left(Q_w - Q_*\right)\|_F}{\left\|\boldsymbol{A}\boldsymbol{F}^{-1}\overline{T}_w\right\|_F} - 1\right| \le \frac{\left\|\boldsymbol{A}\left(Q_w - Q_* - \boldsymbol{F}^{-1}\overline{T}_w\right)\right\|_F}{\left\|\boldsymbol{A}\boldsymbol{F}^{-1}\overline{T}_w\right\|_F}$$
$$\le \kappa(\boldsymbol{A})\frac{\left\|Q_w - Q_* - \boldsymbol{F}^{-1}\overline{T}_w\right\|_F}{\left\|\boldsymbol{F}^{-1}\overline{T}_w\right\|_F} \overset{\text{by } L.2.3}{\le} \sqrt{\kappa(Q_*)}\eta_w.$$

Here we use the fact that $\kappa(\boldsymbol{A}) = \sqrt{\kappa(Q_*)}$ by Lemma A.1. Combining the above inequalities, one obtains the final bound:

$$\left|\frac{d_{\mathrm{B}}(Q_w, Q_*)}{\left\|\boldsymbol{A}\boldsymbol{F}^{-1}\overline{T}_w\right\|_F} - 1\right| \le 2q_w + (1 + 2q_w)\left|\frac{\|\boldsymbol{A}\left(Q_w - Q_*\right)\|_F}{\left\|\boldsymbol{A}\boldsymbol{F}^{-1}\overline{T}_w\right\|_F} - 1\right|$$
$$\le 2q_w + 2\left|\frac{\|\boldsymbol{A}\left(Q_w - Q_*\right)\|_F}{\left\|\boldsymbol{A}\boldsymbol{F}^{-1}\overline{T}_w\right\|_F} - 1\right| \le 3\sqrt{\kappa(Q_*)}\eta_w,$$

where the second and third inequality rely on the assumption $q_w + f_w \le \frac{1}{2}$ and the bound $2q_w \le \eta_w$. □

3. Gaussian approximation.

This section aims to show that the Gaussian approximations presented in Table 1 holds. Namely,

$$\sqrt{n}\rho(Q_*, Q_n) \overset{\mathrm{d}}{\approx} \|\boldsymbol{L}_\rho Z\|_F, \quad \sqrt{n}\rho(Q_n, Q_u) \overset{\mathrm{d}}{\approx} \|\hat{\boldsymbol{L}}_\rho \hat{Z}\|_F. \tag{3.1}$$

We begin with the key ingredient used in the proofs.

3.1. *Key lemma for Gaussian approximation.*

Following Götze et al. [2019], we define $\varkappa(\boldsymbol{\Psi})$. Let $\{\lambda_k\}_k$ be the eigenvalues of an Hermitian operator $\boldsymbol{\Psi}$ arranged in non-increasing order. We set

$$\varkappa(\boldsymbol{\Psi}) \overset{\text{def}}{=} (\varLambda_1\varLambda_2)^{-1/2} \quad \text{with} \quad \varLambda_r^2 \overset{\text{def}}{=} \sum_{k\ge r}\lambda_k^2, \text{ where } r = 1, 2. \tag{3.2}$$

Its properties are investigated in Lemma B.1. We also need

$$\gamma(\boldsymbol{\Psi}) \overset{\text{def}}{=} \varkappa(\boldsymbol{\Psi})\,\mathrm{tr}(\boldsymbol{\Psi}) \ge 1, \quad \gamma_\kappa(\boldsymbol{\Psi}) \overset{\text{def}}{=} \sqrt{\kappa(Q_*)}\varkappa(\boldsymbol{\Psi})\,\mathrm{tr}(\boldsymbol{\Psi}) \ge 1. \tag{3.3}$$

The lower bound on $\gamma(\boldsymbol{\Psi})$ is trivial. It follows from the fact that for any $r \ge 1$

$$\varLambda_r^2 \le \left(\sum_{k\ge r}\lambda_k\right)^2 \le \left(\mathrm{tr}(\boldsymbol{\Psi})\right)^2.$$

LEMMA 3.1 (GAR). *Let $X, Y \in \mathbb{R}_+$ be random variables s.t. there exist constants $m, \delta > 0$, and $\eta \in \left[0, \frac{1}{2}\right]$, s.t.*

$$\mathbb{P}\left(|X - Y| \le \eta Y + m\right) \ge 1 - \delta. \tag{GAR-I}$$

Assume that for Y the Gaussian approximation holds, i.e. there exists a centred Gaussian vector $G \sim \mathcal{N}(0, \boldsymbol{K})$ taking values in a Hilbert space H, and a constant $\varDelta \in (0, 1)$, s.t.

$$\sup_{z>0}\left|\mathbb{P}\left\{Y \le z\right\} - \mathbb{P}\left\{\|G\|_H \le z\right\}\right| \le \varDelta, \tag{GAR-II}$$

with $\|\cdot\|_H$ denoting the norm induced by the scalar product in H. Then

$$\sup_{z>0}|\mathbb{P}\{X \le z\} - \mathbb{P}\{\|G\|_H \le z\}| \le \varDelta + \delta + \mathtt{C}\gamma(\boldsymbol{K})\left(\frac{m}{\sqrt{\operatorname{tr}(\boldsymbol{K})}} + \eta\right),$$

where $\gamma(\boldsymbol{K})$ comes from (3.3).

PROOF. Union bound ensures that

$$\mathbb{P}\{X \le z\} \le \mathbb{P}\left\{Y \le \tfrac{z+m}{1-\eta}\right\} + \mathbb{P}\{|X-Y| > \eta Y + m\} \le \mathbb{P}\left\{Y \le \tfrac{z+m}{1-\eta}\right\} + \delta,$$

$$\mathbb{P}\left\{Y \le \tfrac{z-m}{1+\eta}\right\} \le \mathbb{P}\{X \le z\} + \mathbb{P}\{|X-Y| > \eta Y + m\} \le \mathbb{P}\{X \le z\} + \delta.$$

Thus

$$\mathbb{P}\left\{Y \le \tfrac{z-m}{1+\eta}\right\} - \delta \le \mathbb{P}\{X \le z\} \le \mathbb{P}\left\{Y \le \tfrac{z+m}{1-\eta}\right\} + \delta.$$

Assumption (GAR-II) yields

$$\mathbb{P}\left\{\|G\|_H \le \tfrac{z-m}{1+\eta}\right\} - \delta - \varDelta \le \mathbb{P}\{X \le z\} \le \mathbb{P}\left\{\|G\|_H \le \tfrac{z+m}{1-\eta}\right\} + \delta + \varDelta.$$

Now one has to bound $\mathbb{P}\left\{\|G\|_H \le \tfrac{z-m}{1+\eta}\right\}$ and $\mathbb{P}\left\{\|G\|_H \le \tfrac{z+m}{1-\eta}\right\}$. Since for any $x, h > 0$ it holds $(x+\varepsilon)^2 \le x^2 + \frac{x\varepsilon}{h} + h\varepsilon + \varepsilon^2$, Theorem 2.7 by Götze et al. [2019] yields

$$\begin{aligned}\mathbb{P}\{x \le \|G\|_H \le x+\varepsilon\} \le& \mathbb{P}\left\{x^2 \le \|G\|_H^2 \le x^2 + 2h\varepsilon + 2\varepsilon^2\right\} \\ &+ \mathbb{P}\left\{x \le \|G\|_H \le x\left(1 + \frac{\varepsilon}{h}\right)\right\} \\ \le& \mathtt{C}\varkappa(\boldsymbol{K})\left(h\varepsilon + \varepsilon^2 + \frac{\varepsilon}{h}\operatorname{tr}(\boldsymbol{K})\right) \le \mathtt{C}\varkappa(\boldsymbol{K})\left(\varepsilon\sqrt{\operatorname{tr}(\boldsymbol{K})} + \varepsilon^2\right),\end{aligned}$$

where the last inequality is ensured by $h = \sqrt{\operatorname{tr}\boldsymbol{K}}$. The above inequality can be rewritten as

$$\mathbb{P}\{x \le \|G\|_H \le x+\varepsilon\} \le \mathtt{C}\gamma(\boldsymbol{K})\left(\frac{\varepsilon}{\sqrt{\operatorname{tr}(\boldsymbol{K})}} + \frac{\varepsilon^2}{\operatorname{tr}(\boldsymbol{K})}\right),$$

and since $\gamma(\boldsymbol{K}) \ge 1$ and the probability on the l.h.s. is bounded by 1, it is enough to consider the case $\varepsilon \le \sqrt{\operatorname{tr}(\boldsymbol{K})}$. Thus, we obtain

$$\mathbb{P}\{x \le \|G\|_H \le x+\varepsilon\} \le \mathtt{C}\gamma(\boldsymbol{K})\frac{\varepsilon}{\sqrt{\operatorname{tr}(\boldsymbol{K})}}.$$

Taking into account that $\eta \in \left[0, \frac{1}{2}\right]$, we get

$$\mathbb{P}\left\{\|G\|_H \le \tfrac{z-m}{1+\eta}\right\} \ge \mathbb{P}\left\{\|G\|_H \le \tfrac{z}{1+\eta}\right\} - \mathtt{C}\gamma(\boldsymbol{K})\frac{m}{\sqrt{\operatorname{tr}(\boldsymbol{K})}},$$

$$\mathbb{P}\left\{\|G\|_H \le \tfrac{z+m}{1-\eta}\right\} \le \mathbb{P}\left\{\|G\|_H \le \tfrac{z}{1-\eta}\right\} + \mathtt{C}\gamma(\boldsymbol{K})\frac{m}{\sqrt{\operatorname{tr}(\boldsymbol{K})}}.$$

Now we consider a Gaussian r.v. αG with some $\alpha > 0$. Note that by definition $\varkappa(\alpha^2\boldsymbol{K}) = \frac{1}{\alpha^2}\varkappa(\boldsymbol{K})$. To compare G and αG we use Corollary 2.3 by Götze et al. [2019]. This ensures for any $z > 0$

$$\begin{aligned}\left|\mathbb{P}\left\{\|G\|_H \le \tfrac{z}{\alpha}\right\} - \mathbb{P}\{\|G\|_H \le z\}\right| &\le \mathtt{C}\left(\varkappa(\boldsymbol{K}) + \varkappa(\alpha^2\boldsymbol{K})\right)\left\|\boldsymbol{K} - \alpha^2\boldsymbol{K}\right\|_1 \\ &= \mathtt{C}\left(1 + \tfrac{1}{\alpha^2}\right)|1-\alpha^2|\varkappa(\boldsymbol{K})\operatorname{tr}(\boldsymbol{K}).\end{aligned}$$

Setting $\alpha = 1 + \eta$ and taking into account that $\eta \in [0, \frac{1}{2}]$, we obtain

$$\mathbb{P}\left\{\|G\|_H \le \tfrac{z}{1+\eta}\right\} \ge \mathbb{P}\left\{\|G\|_H \le z\right\} - \mathtt{C}\gamma(\boldsymbol{K})\eta.$$

In a similar way

$$\mathbb{P}\left\{\|G\|_H \le \tfrac{z}{1-\eta}\right\} \le \mathbb{P}\left\{\|G\|_H \le z\right\} + \mathtt{C}\gamma(\boldsymbol{K})\eta.$$

Collecting all the bounds, we get the result. □

To get the Gaussian approximations (3.1), we combine this lemma with the results of Section 2.

3.2. *Gaussian approximation in the real world.* From now on we assume that Q_* comes from (1.2), and adapt the notations from Section 2. The counterpart of $\overline{T}_w$ (2.1) in the real world is

$$\overline{T}_n \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} T_i, \tag{3.4}$$

with T_i coming from (1.5), and $w_i = 1/n$ for all i. The counterparts of q_w and f_w (2.4) are

$$q_n \stackrel{\text{def}}{=} \left\| Q_*^{-1/2} Q_n Q_*^{-1/2} - I \right\|, \quad f_n \stackrel{\text{def}}{=} \left\| \boldsymbol{F}^{-1/2} \boldsymbol{F}_n \boldsymbol{F}^{-1/2} - I \right\|. \tag{3.5}$$

Finally, the counterpart of $\boldsymbol{A}$ (2.6) has the same form:

$$\boldsymbol{A} \stackrel{\text{def}}{=} \left(-\frac{1}{2}\boldsymbol{dT}_{Q_*}^{Q_*}\right)^{1/2}. \tag{3.6}$$

To validate the Gaussian approximation, we assume the data distribution P to be s.t. for any $\mathrm{x} \ge 0$

$$\mathbb{P}\left\{\left\|\overline{T}_n\right\|_F > \varepsilon_T(\mathrm{x})\right\} \le \delta_T(\mathrm{x}). \tag{T}$$

Furthermore, for any $\mathrm{x} \ge 0$

$$\mathbb{P}\left\{\left\|\boldsymbol{F}^{-1/2}\boldsymbol{F_n}\boldsymbol{F}^{-1/2} - \boldsymbol{I}\right\| > \varepsilon_F(\mathrm{x})\right\} \le \delta_F(\mathrm{x}), \tag{F}$$

where $\boldsymbol{I}$ denotes the identity operator acting on the same space as $\boldsymbol{F}$. These assumptions combined together ensure the concentration of Q_n in the vicinity of Q_*. This fact follows from Lemma 2.2 and is validated by Corollary B.2.

Our third assumption ensures that the Frobenius norm of the sum of rescaled optimal push-forwards is close in distribution to the Frobenius norm of the Gaussian vector Z. For all $z \ge 0$

$$\left|\mathbb{P}\left\{\left\|\sqrt{n}\boldsymbol{F}^{-1}\overline{T}_n\right\|_F \le z\right\} - \mathbb{P}\left\{\|Z\|_F \le z\right\}\right| \le \varDelta_Z. \tag{Z}$$

Combined together, the assumptions ensure GAR in the real world. The proof is technical and is postponed to Appendix.

LEMMA 3.2 (GAR in the real world). *Let Assumptions (T), (F), (Z) be fulfilled. We choose* $\mathrm{x} \in \mathbb{R}_+^2$ *and denote*

$$\delta_\eta(\mathrm{x}) \stackrel{\text{def}}{=} \delta_\eta(\mathrm{x}_1, \mathrm{x}_2) = \delta_F(\mathrm{x}_1) + \delta_T(\mathrm{x}_2), \tag{3.7}$$

$$\varepsilon_\eta(\mathrm{x}) \stackrel{\text{def}}{=} \varepsilon_\eta(\mathrm{x}_1, \mathrm{x}_2) = 2\sqrt{\kappa(\boldsymbol{F})}\left(\varepsilon_F(\mathrm{x}_1) + \varepsilon_Q(\mathrm{x}_2)\right), \tag{3.8}$$

with $\varepsilon_Q(\mathrm{x}_2) \propto \varepsilon_T(\mathrm{x}_2)$ *coming from Corollary B.2. Take* $Z \sim \mathcal{N}(0, \Xi)$*. Then the following Gaussian approximations hold.*

Frobenius norm. *For all* $z \ge 0$

$$\left|\mathbb{P}\left\{\sqrt{n}\|Q_n - Q_*\|_F \le z\right\} - \mathbb{P}\left\{\|Z\|_F \le z\right\}\right| \le \Omega_{\mathrm{F}}(n),$$

$$\Omega_{\mathrm{F}}(n) \stackrel{\text{def}}{=} \Delta_Z + \inf_{\mathrm{x}} \left\{\delta_\eta(\mathrm{x}) + \mathtt{C}\gamma(\boldsymbol{\Xi})\varepsilon_\eta(\mathrm{x})\right\}, \tag{3.9}$$

where the infimum is taken over all x*, s.t.* $\varepsilon_\eta(\mathrm{x}) \le \frac{1}{2}$.

Bures–Wasserstein distance. *For all* $z \ge 0$

$$\left|\mathbb{P}\left\{\sqrt{n}d_{\mathrm{B}}(Q_n, Q_*) \le z\right\} - \mathbb{P}\left\{\|\boldsymbol{A}Z\|_F \le z\right\}\right| \le \Omega_{\mathrm{B}}(n),$$

$$\Omega_{\mathrm{B}}(n) \stackrel{\text{def}}{=} \Delta_Z + \inf_{\mathrm{x}} \left\{\delta_\eta(\mathrm{x}) + \mathtt{C}\gamma_\kappa(\boldsymbol{\Xi})\varepsilon_\eta(\mathrm{x})\right\}, \tag{3.10}$$

where the infimum is taken over all x*, s.t.* $\varepsilon_\eta(\mathrm{x}) \le \frac{1}{6\sqrt{\kappa(Q_*)}}$.

3.3. *Gaussian approximation in the bootstrap world.* The similar results hold in the bootstrap world. The counterpart of $\overline{T}_w$ (2.1) is

$$\overline{T}_u \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n u_i \hat{T}_i, \tag{3.11}$$

with $\hat{T}_i$ coming from (1.5), and $w_i = u_i$ for all i. The counterparts of q_w and f_w (2.4) are

$$q_u \stackrel{\text{def}}{=} \left\|Q_n^{-1/2} Q_u Q_n^{-1/2} - I\right\|, \quad f_u \stackrel{\text{def}}{=} \left\|\hat{\boldsymbol{F}}^{-1/2}\hat{\boldsymbol{F}}_u\hat{\boldsymbol{F}}^{-1/2} - I\right\|. \tag{3.12}$$

The counterpart of $\boldsymbol{A}$ is

$$\hat{\boldsymbol{A}} \stackrel{\text{def}}{=} \left(-\frac{1}{2}\boldsymbol{dT}_{Q_n}^{Q_n}\right)^{1/2}. \tag{3.13}$$

We are now ready to introduce the assumptions about the concentration of quantities conditioned on the observed sample. Thus, all of them hold with some P-probability. First, we assume that for any $\mathrm{x}, \mathrm{s} \ge 0$, with P-probability at least $1 - \alpha_{\hat{T}}(\mathrm{s})$,

$$\mathbb{P}_u\left\{\left\|\overline{T}_u\right\|_F > \varepsilon_{\hat{T}}(\mathrm{x};\mathrm{s})\right\} \le \delta_{\hat{T}}(\mathrm{x};\mathrm{s}). \tag{$\hat{T}$}$$

For any $\mathrm{x}, \mathrm{s} \ge 0$, with P-probability at least $1 - \alpha_{\hat{F}}(\mathrm{s})$,

$$\mathbb{P}_u\left\{\left\|\hat{\boldsymbol{F}}^{-1/2}\hat{\boldsymbol{F}}_u\hat{\boldsymbol{F}}^{-1/2} - \boldsymbol{I}\right\| > \varepsilon_{\hat{F}}(\mathrm{x};\mathrm{s})\right\} \le \delta_{\hat{F}}(\mathrm{x};\mathrm{s}). \tag{$\hat{F}$}$$

Assumptions ($\hat{T}$) and ($\hat{F}$) ensure the concentration of Q_u in the vicinity of Q_n, see Corollary B.2. Finally, we require that with P-probability at least $1 - \alpha_{\hat{Z}}(\mathrm{s})$, $\mathrm{s} \ge 0$, for all $z \ge 0$

$$\left|\mathbb{P}_u\left\{\left\|\sqrt{n}\hat{\boldsymbol{F}}^{-1}\overline{T}_u\right\|_F \le z\right\} - \mathbb{P}\left\{\left\|\hat{Z}\right\|_F \le z\right\}\right| \le \Delta_{\hat{Z}}(\mathrm{s}), \tag{$\hat{Z}$}$$

LEMMA 3.3. *Let Assumptions* ($\hat{T}$), ($\hat{F}$) *and* ($\hat{Z}$) *be fulfilled. Let* $\mathrm{s} \in \mathbb{R}_+^3$*. We denote*

$$\hat{\alpha}(\mathrm{s}) = \alpha_{\hat{T}}(\mathrm{s}_1) + \alpha_{\hat{F}}(\mathrm{s}_2) + \alpha_{\hat{Z}}(\mathrm{s}_3). \tag{3.14}$$

Take $\mathrm{x} \in \mathbb{R}_+^2$ *and define*

$$\delta_{\hat{\eta}}(\mathrm{x};\mathrm{s}) \stackrel{\text{def}}{=} \delta_{\hat{F}}(\mathrm{x}_1;\mathrm{s}_1) + \delta_{\hat{T}}(\mathrm{x}_2;\mathrm{s}_2), \tag{3.15}$$

$$\varepsilon_{\hat{\eta}}(\mathrm{x};\mathrm{s}) \stackrel{\text{def}}{=} 6\sqrt{\kappa(\boldsymbol{F})}\left(\varepsilon_{\hat{F}}(\mathrm{x}_1;\mathrm{s}_1) + \varepsilon_{\hat{Q}}(\mathrm{x}_2;\mathrm{s}_2)\right), \tag{3.16}$$

with $\varepsilon_{\hat{Q}}(\mathrm{x}_2;\mathrm{s}_2) \propto \varepsilon_{\hat{T}}(\mathrm{x}_2;\mathrm{s}_2)$ *coming from Corollary B.2. Let* $\hat{Z} \sim \mathcal{N}\left(0, \hat{\boldsymbol{\Xi}}\right)$*. Then the following approximations hold with* P*-probability at least* $1 - \hat{\alpha}(\mathrm{s})$*, provided that* $f_n, q_n \le \frac{1}{2}$, $\gamma(\hat{\boldsymbol{\Xi}}) \le 2\gamma(\boldsymbol{\Xi})$.

Frobenius norm. *For all* $z > 0$

$$\left|\mathbb{P}_u\left\{\sqrt{n}\|Q_u - Q_n\|_F \le z\right\} - \mathbb{P}\left\{\|\hat{Z}\|_F \le z\right\}\right| \le \hat{\Omega}_{\mathrm{F}}(n; \mathrm{s}),$$

$$\hat{\Omega}_{\mathrm{F}}(n; \mathrm{s}) \stackrel{\mathrm{def}}{=} \Delta_{\hat{Z}}(\mathrm{s}) + \inf_{\mathrm{x}}\left\{\delta_{\hat{\eta}}(\mathrm{x}; \mathrm{s}) + \mathtt{C}\gamma(\boldsymbol{\Xi})\varepsilon_{\hat{\eta}}(\mathrm{x}; \mathrm{s})\right\}, \tag{3.17}$$

where the infimum is taken over all x *s.t.* $\varepsilon_{\hat{\eta}}(\mathrm{x}; \mathrm{s}) \le \frac{1}{2}$*;*

Bures–Wasserstein distance. *For all* $z \ge 0$

$$\left|\mathbb{P}_u\left\{\sqrt{n} d_{\mathrm{B}}(Q_u, Q_n) \le z\right\} - \mathbb{P}\left\{\|\hat{\boldsymbol{A}}\hat{Z}\|_F \le z\right\}\right| \le \hat{\Omega}_{\mathrm{B}}(n; \mathrm{s}),$$

$$\hat{\Omega}_{\mathrm{B}}(n; \mathrm{s}) \stackrel{\mathrm{def}}{=} \Delta_{\hat{Z}}(\mathrm{s}) + \inf_{\mathrm{x}}\left\{\delta_{\hat{\eta}}(\mathrm{x}; \mathrm{s}) + \mathtt{C}\gamma_{\kappa}(\boldsymbol{\Xi})\varepsilon_{\hat{\eta}}(\mathrm{x}; \mathrm{s})\right\}, \tag{3.18}$$

where the infimum is taken over all x *s.t.* $\varepsilon_{\hat{\eta}}(\mathrm{x}; \mathrm{s}) \le \frac{1}{6\sqrt{3\kappa(Q_*)}}$.

The proof is technical and is postponed to the Appendix.

REMARK 3.4. *We note that the assumptions* $f_n, q_n \le \frac{1}{2}$ *and* $\gamma(\hat{\boldsymbol{\Xi}}) \le 2\gamma(\boldsymbol{\Xi})$ *can be replaced with some suitable concentration bounds. This will yield an additional term to* P*-probability, which controls the approximation. The replacement is done in Section 4 and is discussed in more detail in Remark C.3.*

4. Non-asymptotic approximation based on multiplier bootstrap. This section introduces the last assumption, which ensures the approximation

$$\|\boldsymbol{L}_{\rho} Z\|_F \stackrel{\mathrm{d}}{\approx} \left\|\hat{\boldsymbol{L}}_{\rho}\hat{Z}\right\|_F. \tag{4.1}$$

Let for any $\mathrm{x} \ge 0$

$$\mathbb{P}\left\{\|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}\| > \varepsilon_{\Sigma}(\mathrm{x})\right\} \le \delta_{\Sigma}(\mathrm{x}). \tag{Σ}$$

This guarantees that with high P-probability for any $\mathrm{x} \in \mathbb{R}^3_+$

$$\|\hat{\boldsymbol{\Xi}} - \boldsymbol{\Xi}\|_1 \lesssim \varepsilon_{\Xi}(\mathrm{x}),$$

$$\varepsilon_{\Xi}(\mathrm{x}) \stackrel{\mathrm{def}}{=} d^2\varepsilon_{\Sigma}(\mathrm{x}) + \varepsilon_{\eta}(\mathrm{x})\left(\mathrm{tr}\,\boldsymbol{\Sigma} + d\right), \tag{4.2}$$

with $\|\cdot\|_1$ being the 1-Schatten (nuclear) norm. This fact is crucial for the Gaussian comparison (4.1) and is validated by Lemma C.2. The proof of (4.1) is contained within the proofs of the main results. We postpone it to the Appendix.

Now we are ready to present the main results in a rigorous way. We begin with the Frobenius norm. Let $\mathrm{y} \in \mathbb{R}^3_+$, and denote

$$\delta(\mathrm{y}) \stackrel{\mathrm{def}}{=} \delta_T(\mathrm{y}) + \delta_F(\mathrm{y}) + \delta_{\Sigma}(\mathrm{y}).$$

From now on we will denote generic absolute constants as $\mathtt{C}$.

THEOREM 4.1 (Bootstrap validity for Frobenius norm). *Let all Assumptions* (T) – (Σ) *be fulfilled. Take* $\mathrm{s} \in \mathbb{R}^3_+$ *and* $\mathrm{y} \in \mathbb{R}^3_+$ *s.t.*

$$\varepsilon_{\eta}(\mathrm{y}) \le 1, \quad \varepsilon_{\Xi}(\mathrm{y}) \le \mathtt{C}\lambda^2_{\min}(\boldsymbol{F})\frac{\Lambda_2^2(\boldsymbol{\Xi})}{\|\boldsymbol{\Xi}\|}.$$

Then with P-probability at least $1-\delta(\mathrm{y})-\hat{\alpha}(\mathrm{s})$, with $\hat{\alpha}(\mathrm{s})$ coming from (3.14), for all $z \ge 0$

$$|\mathbb{P}_u\{\|Q_u - Q_n\|_F \le z\} - \mathbb{P}\{\|Q_n - Q_*\|_F \le z\}| \le \mathit{\Gamma}_F(n;\mathrm{y},\mathrm{s}),$$

$$\mathit{\Gamma}_F(n;\mathrm{y},\mathrm{s}) \stackrel{\text{def}}{=} \mathtt{C}\frac{\varkappa(\Xi)}{\lambda^2_{\min}(\boldsymbol{F})}\varepsilon_{\Xi}(\mathrm{y}) + \mathit{\Omega}_{\mathrm{F}}(n) + \hat{\mathit{\Omega}}_{\mathrm{F}}(n;\mathrm{s}),$$

where $\mathit{\Omega}_{\mathrm{F}}(n)$ comes from (3.9) and $\hat{\mathit{\Omega}}_{\mathrm{F}}(n;\mathrm{s})$ comes from (3.17).

THEOREM 4.2 (Bootstrap validity for BW distance). *Let all Assumptions (T) – (Σ) be fulfilled. Take $\mathrm{s} \in \mathbb{R}^3_+$ and $\mathrm{y} \in \mathbb{R}^3_+$ s.t.*

$$\varepsilon_\eta(\mathrm{y}) \le 1, \quad \varepsilon_\Xi(\mathrm{y}) \le \mathtt{C}\lambda_{\min}(Q_*)\lambda^2_{\min}(\boldsymbol{F})\frac{\Lambda_2^2(\boldsymbol{A}\Xi\boldsymbol{A})}{\|\boldsymbol{A}\Xi\boldsymbol{A}\|}.$$

Then with P-probability at least $1-\delta(\mathrm{y})-\hat{\alpha}(\mathrm{s})$, with $\hat{\alpha}(\mathrm{s})$ coming from (3.14), for all $z \ge 0$

$$|\mathbb{P}_u\{d_{\mathrm{B}}(Q_u, Q_n) \le z\} - \mathbb{P}\{d_{\mathrm{B}}(Q_n, Q_*) \le z\}| \le \mathit{\Gamma}_{\mathrm{B}}(n;\mathrm{y},\mathrm{s}),$$

$$\mathit{\Gamma}_{\mathrm{B}}(n;\mathrm{y},\mathrm{s}) \stackrel{\text{def}}{=} \mathtt{C}\frac{\varkappa(\boldsymbol{A}\Xi\boldsymbol{A})}{\lambda^2_{\min}(\boldsymbol{F})\lambda_{\min}(Q_*)}\varepsilon_{\Xi}(\mathrm{y}) + \mathit{\Omega}_{\mathrm{B}}(n) + \hat{\mathit{\Omega}}_{\mathrm{B}}(n;\mathrm{s}),$$

where $\mathit{\Omega}_{\mathrm{B}}(n)$ comes from (3.10) and $\hat{\mathit{\Omega}}_{\mathrm{B}}(n;\mathrm{s})$ comes from (3.18).

5. Sub-Gaussian case. To illustrate the approximation rates we consider a particular choice of the model: the distribution P is sub-exponential and P_u is sub-Gaussian.

ASSUMPTION 1. *Let $\|\cdot\|_\psi$ be the Orlicz ψ-norm. We assume that:*

- *P is s.t. $\|\operatorname{tr} S\|_{\psi_1} = \nu_S < \infty$,*
- *P_u is s.t. $\|u-1\|_{\psi_2} = \nu_u < \infty$.*

This assumption immediately ensures the sub-Gaussianity of the following terms.

LEMMA 5.1 (Sub-Gaussianity of other quantities). *Assumption 1 ensures that*

1. *$\|S_i\|^{1/2}$ is sub-Gaussian with some parameter $\nu_{\|S\|}$,*
2. *$\|T_i\|$ is a sub-Gaussian r.v. with some parameter $\nu_{\|T\|}$,*
3. *T_i is sub-Gaussian with some parameter ν_T,*
4. *$\|\boldsymbol{d}\boldsymbol{T}_i\|$ sub-Gaussian with $\big\|\|\boldsymbol{d}\boldsymbol{T}_i\|\big\|_{\psi_2} = \nu_{\|\boldsymbol{d}\boldsymbol{T}\|}$.*

In order to specify the results of Theorem 4.1 and Theorem 4.2 we have to show that all the Assumptions (T)–(Σ) hold. Validation of Assumptions (T) and ($\hat{T}$) is technical. The key ingredient for verification of Assumptions (F), ($\hat{F}$), and (Σ) is Lemma D.6. It slightly extends Proposition 2 by Koltchinskii et al. [2011]: we obtain a similar concentration result for the case of independent but not identically distributed observations. Finally, to check Assumptions (Z) and ($\hat{Z}$), we use the result by Bentkus [2003], and Theorem 3.5 by Chen and Fang [2011], respectively. We specify the rates coming from the assumptions and the auxiliary bounds in the tables below. Namely, Table 2 deals with Gaussian approximation, while Table 3 presents the results on Gaussian comparison.

Of note, the largest input in terms of dimension d comes from Assumptions ($\hat{Z}$): it yields an approximation error of order at least $\frac{d^2}{\sqrt{n}}$. An additional multiplier $\sqrt{\mathrm{s}+\ln n}$ comes from controlling the concentration of maxima of i.i.d. sub-Gaussian random variables $\max_i\|T_i\|_F$ while estimating $\Delta_{\hat{Z}}(s)$.

	Real world	**Bootstrap world**
Assumptions validating GAR	Assumption (T) holds due to Lemma D.1 $$\delta_T(\mathrm{x}) = e^{-\mathrm{x}},$$ $$\varepsilon_T(\mathrm{x}) \lesssim \sqrt{\frac{\mathrm{x}+d^2}{n}}.$$	Assumption ($\hat{T}$) holds due to Lemma D.3 $$\delta_{\hat{T}}(\mathrm{x};\mathrm{s}) = e^{-\mathrm{x}}, \quad \alpha_{\hat{T}}(\mathrm{s}) = e^{-\mathrm{s}}$$ $$\varepsilon_{\hat{T}}(\mathrm{x};\mathrm{s}) \lesssim \sqrt{\frac{\mathrm{x}+d^2}{n}}\left(1+\sqrt{\frac{\mathrm{s}}{n}}\right).$$
	Assumption (F) holds due to Lemma D.7 $$\delta_F(\mathrm{x}) = e^{-\mathrm{x}},$$ $$\varepsilon_F(\mathrm{x}) \lesssim \sqrt{\frac{\mathrm{x}+\ln d}{n}}.$$	Assumption ($\hat{F}$) holds due to Lemma D.9 $$\delta_{\hat{F}}(\mathrm{x};\mathrm{s}) = e^{-\mathrm{x}}, \quad \alpha_{\hat{F}}(\mathrm{s}) = 3e^{-\mathrm{s}},$$ $$\varepsilon_{\hat{F}}(\mathrm{x};\mathrm{s}) \lesssim \sqrt{\frac{\mathrm{x}+\ln d}{n}}\left(1+\sqrt{\frac{\mathrm{s}}{n}}\right).$$
	Assumption (Z) holds due Lemma D.4 $$\Delta_Z \lesssim \frac{d^2}{\sqrt{n}}.$$	Assumption ($\hat{Z}$) holds due to Lemma D.5 $$\alpha_{\hat{Z}}(\mathrm{s}) = 2e^{-\mathrm{s}},$$ $$\Delta_{\hat{Z}}(\mathrm{s}) \lesssim d^2\sqrt{\frac{\mathrm{s}+\ln n}{n}}.$$
GAR bounds	Lemma 3.2 holds due to Lemma D.9: for sufficiently large n $$\Omega_\rho(n) \lesssim \sqrt{\frac{\ln n}{n}}.$$	Lemma 3.3 holds due to Lemma D.11 and Theorem D.12: for sufficiently large n $$\hat{\Omega}_\rho(n;\mathrm{s}) \lesssim \sqrt{\frac{\mathrm{s}+\ln n}{n}}.$$

TABLE 2
GAR bounds; ρ stands either for $\|\cdot\|_F$ or d_{B}.

Cov. comp.	Assumption (Σ) holds due Lemma D.8 $$\delta_\Sigma(\mathrm{x}) = e^{-\mathrm{x}}, \quad \varepsilon_\Sigma(\mathrm{x}) \lesssim \sqrt{\frac{\mathrm{x}+\ln d}{n}}.$$
Gauss. comp.	For all $z>0$ with P-probability at least $1-3e^{-\mathrm{x}}$ $$\left\|\mathbb{P}\left\{\|\boldsymbol{L}_\rho Z\|_F \le z\right\} - \mathbb{P}\left\{\left\|\hat{\boldsymbol{L}}_\rho \hat{Z}\right\|_F \le z\right\}\right\| \lesssim d^2\sqrt{\frac{\mathrm{x}+\ln d}{n}} + (\operatorname{tr}\boldsymbol{\Sigma}+d)\sqrt{\frac{\mathrm{x}+d^2}{n}}$$

TABLE 3
Comparison of covariances and Gaussian comparison; ρ stands either for $\|\cdot\|_F$ or d_{B}.

Finally, the bounds $\Gamma_{\mathrm{F}}(n;\mathrm{s})$ and $\Gamma_{\mathrm{B}}(n;\mathrm{s})$ are given by Theorem D.12. For sufficiently large sample size n they are written as

$$\Gamma_{\mathrm{F}}(n;\mathrm{s}) \lesssim \sqrt{\frac{\mathrm{s}+\ln n}{n}} \quad \text{and} \quad \Gamma_{\mathrm{B}}(n;\mathrm{s}) \lesssim \sqrt{\frac{\mathrm{s}+\ln n}{n}}.$$

Note that the notation $\lesssim$ suppresses multiplicative constants depending on the distribution.

6. Experiments. This section presents application of bootstrap approach for study of the data represented by graphs. For instance, such models are used for description of 3D genome folding via Hi-C matrices [Van Berkum et al., 2010, Yaffe and Tanay, 2011], or for the analysis of the structure and functions of the brain networks represented by connectomes [Bullmore and Sporns, 2009, Fornito et al., 2016].

It is well-known that the specific structures of a graph can be revealed through analysis of its Laplacian matrix. Let G be an undirected graph without self loops. We denote as A_G its adjacency matrix and as D_G its degree matrix:

$$A_G = \begin{pmatrix} 0 & a_{12} \dots a_{1d} \\ a_{d1} & a_{d2} \dots \ 0 \end{pmatrix}, \quad D_G = \begin{pmatrix} \sum_i a_{1i} & 0 \dots & 0 \\ 0 & 0 \dots & \sum_i a_{di} \end{pmatrix},$$

where $a_{ij} \in \{0,1\}$ if G is an unweighted graph, and $a_{ij} \geq 0$ otherwise. The graph Laplacian is defined as

$$L_G \stackrel{\text{def}}{=} D_G - A_G,$$

by construction $L_G \in \mathbb{H}_+(d)$. Many clustering methods are based on the analysis of spectral properties of L_G. For instance, its second smallest eigenvector which is also referred to as Fiedler's vector [Fiedler, 1989] provides information about the connectivity of the graph [Van De Ville et al., 2017, Chen et al., 2017]. In general, the spectral clustering methods are often used in practice, as they works well for the case of non-convex supports [Ng et al., 2002] because of their close relation to kernel PCA methods, see Bengio et al. [2003]. For further details on the topic we recommend and excellent tutorial by Von Luxburg [2007].

We use stochastic block model to generate the synthetic data, as it is widely used for description of natural phenomenon. For instance, Faskowitz et al. [2018] fit it to human connectomes for different age groups and then compare the obtained structures. This approach opens the door for further statistical analysis using the Bures-Wasserstein barycenters.

As a toy example we consider a set of weighted graphs $G_1, \dots, G_n$ with $d = 20$ nodes split into two communities. We write the corresponding adjacency matrix $A_{G_i} \in \mathbb{R}^{20\times 20}$ as

$$A_G = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix},$$

where C_{11} and C_{22} are the blocks representing intra-community interaction. The corresponding edge probabilities for each block are $p_{11} = 0.8$, $p_{22} = 0.5$, $p_{12} = p_{21} = 0.2$. The distributions of the weights are Poisson, with $a \sim \mathrm{Po}(12)$ for $a \in C_{11}$, $a \sim \mathrm{Po}(7)$ for $a \in C_{22}$, and $a \sim \mathrm{Po}(2)$ for $a \in C_{12}, C_{21}$. The number of nodes in each community is randomly generated as $\mathrm{size}(C_{11}) = 10 + \mathrm{Unif}([-2,2])$, where $\mathrm{Unif}([-2,2])$ is the uniform distribution on the discrete set $\{-2,-1,0,1,2\}$.

The recent work by Petric Maretic et al. [2019] applied the 2-Wasserstein distance to graph alignment and indicated, that it accounts well for the global structure of graphs. The authors suggest using the pseudo-inverse of L_G. We follow this framework and consider the inverted Laplacians

$$S_i = (L_{G_i} + rI)^{-1},$$

with I being the $d \times d$ identity matrix and $r = 1$ being a regularization parameter. This model fits to the setting of Section 5.

We used 80000 matrices to estimate the true barycenter Q_* and $n = 800$ to compute Q_n. For bootstrapping we use uniform and Poisson weights. The result is illustrated by Fig. 1. Fig. 2 illustrates the convergence result presented in Section 5 for different sample sizes $n \in \{20, 200, 800, 1000\}$ for $u \sim 2\mathrm{Be}(1/2)$. Confidence bands are estimated using 100 bootstrap CDFs.

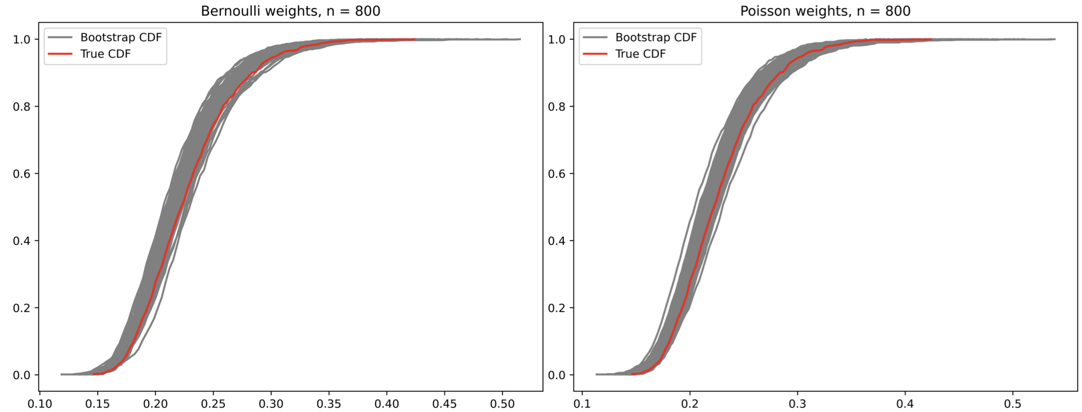

Fig 1: Empirical distribution functions. On both images red is empirical distribution function of $\sqrt{800}d_{\mathrm{B}}(Q_*, Q_{800})$ estimated from 1000 samples. Gray lines are 100 empirical distribution functions of $\sqrt{800}d_{\mathrm{B}}(Q_{800}, Q_u)$. Each is estimated from 1000 samples. At the left image we observe $u \sim 2\mathrm{Be}(1/2)$. The right image depicts the case of $u \sim \mathrm{Po}(1)$.

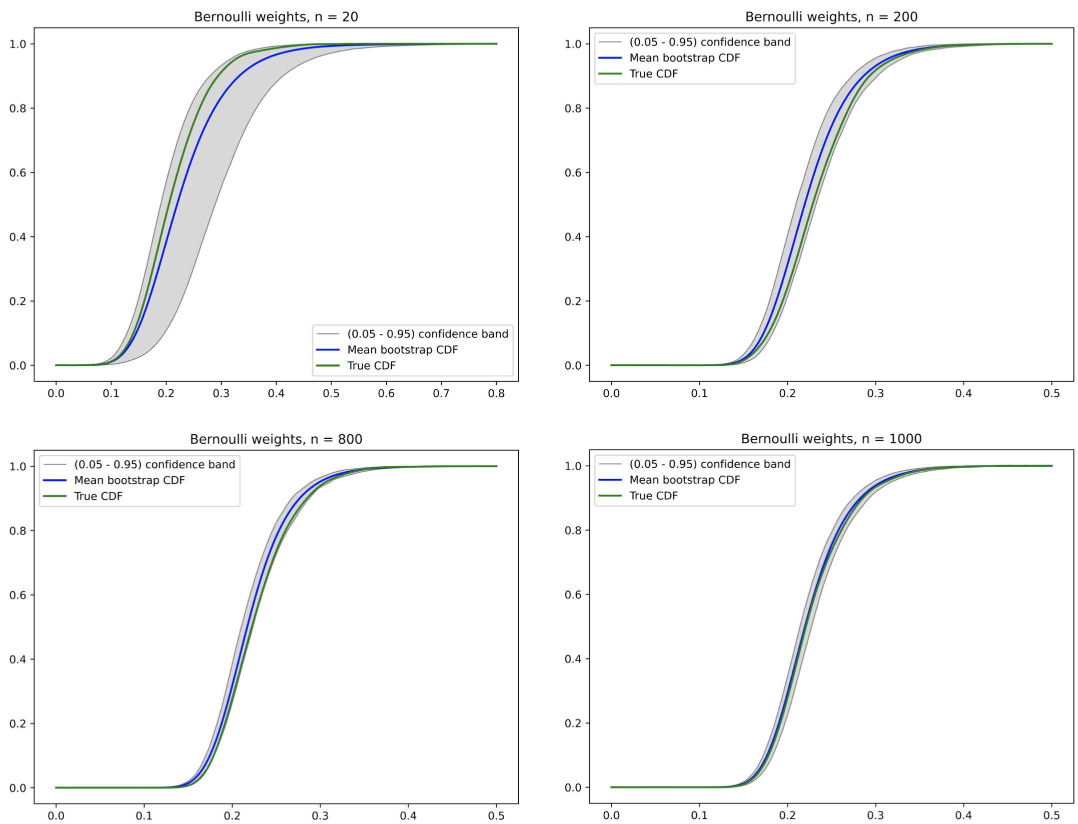

Fig 2: Confidence bands for the distribution of bootstrap CDFs generated using the Bernoulli weights, $u \sim 2\mathrm{Be}(1/2)$. The blue line is the mean bootstrap CDF. The green line is the true CDF.

REFERENCES

Martial Agueh and Guillaume Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

Yoshua Bengio, Pascal Vincent, Jean-François Paiement, Olivier Delalleau, Marie Ouimet, and Nicolas Le Roux. *Spectral clustering and kernel PCA are learning eigenfunctions*, volume 1239. Citeseer, 2003.

Vidmantas Bentkus. On the dependence of the berry–esseen bound on dimension. *Journal of Statistical Planning and Inference*, 113(2):385–402, 2003.

Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 2018.

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198, 2009.

Edward T Bullmore and Danielle S Bassett. Brain graphs: graphical models of the human brain connectome. *Annual review of clinical psychology*, 7:113–140, 2011.

Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. Stochastic convergence of persistence landscapes and silhouettes. In *Proceedings of the thirtieth annual symposium on Computational geometry*, pages 474–483, 2014.

Louis HY Chen and Xiao Fang. Multivariate normal approximation by Stein's method: The concentration inequality approach. *arXiv preprint arXiv:1111.4073*, 2011.

Xi Chen and Wen-Xin Zhou. Robust inference via multiplier bootstrap. *The Annals of Statistics*, 48(3):1665–1691, 2020.

Yannan Chen, Liqun Qi, and Xiaoyan Zhang. The fiedler vector of a laplacian tensor for hypergraph partitioning. *SIAM Journal on Scientific Computing*, 39(6):A2508–A2537, 2017.

Guang Cheng and Jianhua Z Huang. Bootstrap consistency for general semiparametric m-estimation. *The Annals of Statistics*, 38(5):2884–2915, 2010.

Sinho Chewi, Tyler Maunu, Philippe Rigollet, and Austin J Stromme. Gradient descent algorithms for bures-wasserstein barycenters. In *Conference on Learning Theory*, pages 1276–1304. PMLR, 2020.

Johannes Ebert, Vladimir Spokoiny, and Alexandra Suvorikova. Elements of statistical inference in 2-wasserstein space. In *Topics in Applied Analysis and Optimisation*, pages 139–158, Cham, 2019. Springer International Publishing. ISBN 978-3-030-33116-0.

Bradley Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1 – 26, 1979.

Bradley Efron. The bootstrap and modern statistics. *Journal of the American Statistical Association*, 95(452): 1293–1296, 2000.

Joshua Faskowitz, Xiaoran Yan, Xi-Nian Zuo, and Olaf Sporns. Weighted stochastic block models of the human connectome across the life span. *Scientific reports*, 8(1):1–16, 2018.

Miroslav Fiedler. Laplacian of graphs and algebraic connectivity. *Banach Center Publications*, 25(1):57–70, 1989.

Alex Fornito, Andrew Zalesky, and Edward Bullmore. *Fundamentals of brain network analysis*. Academic Press, 2016.

Friedrich Götze, Alexey Naumov, Vladimir Spokoiny, and Vladimir Ulyanov. Large ball probabilities, gaussian comparison and anti-concentration. *Bernoulli*, 25(4A):2538–2563, 2019.

Daniel Hsu, Sham Kakade, Tong Zhang, et al. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17, 2012.

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

Alexey Kroshnin, Vladimir Spokoiny, and Alexandra Suvorikova. Statistical inference for Bures–Wasserstein barycenters. *The Annals of Applied Probability*, 31(3):1264 – 1298, 2021.

Stephen MS Lee and Puyudi Yang. Bootstrap confidence regions based on m-estimators under nonstandard conditions. *The Annals of Statistics*, 48(1):274–299, 2020.

Enno Mammen and Swagata Nandi. Bootstrap and resampling. In *Handbook of computational statistics*, pages 499–527. Springer, 2012.

Boris Muzellec and Marco Cuturi. Generalizing point embeddings using the Wasserstein space of elliptical distributions. In *Advances in Neural Information Processing Systems*, pages 10237–10248, 2018.

Alexey Naumov, Vladimir Spokoiny, and Vladimir Ulyanov. Bootstrap confidence sets for spectral projectors of sample covariance. *Probability Theory and Related Fields*, 174(3):1091–1132, 2019.

Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856, 2002.

Hermina Petric Maretic, Mireille El Gheche, Giovanni Chierchia, and Pascal Frossard. Got: An optimal transport framework for graph comparison. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Vladimir Spokoiny and Mayya Zhilova. Bootstrap confidence sets under model misspecification. *The Annals of Statistics*, 43(6):2653–2675, 2015.

Nynke L Van Berkum, Erez Lieberman-Aiden, Louise Williams, Maxim Imakaev, Andreas Gnirke, Leonid A Mirny, Job Dekker, and Eric S Lander. Hi-c: a method to study the three-dimensional architecture of genomes. *JoVE (Journal of Visualized Experiments)*, (39):e1869, 2010.

Dimitri Van De Ville, Robin Demesmaeker, and Maria Giulia Preti. When slepian meets fiedler: Putting a focus on the graph spectrum. *IEEE Signal Processing Letters*, 24(7):1001–1004, 2017.

Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

Eitan Yaffe and Amos Tanay. Probabilistic modeling of hi-c contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nature genetics*, 43(11):1059–1065, 2011.

APPENDIX A: GEOMETRY OF THE BURES-WASSERSTEIN SPACE

This section contains a list of notations and auxiliary bounds used further in the proofs.

A, B	Matrices or vectors
$\boldsymbol{A}$, $\boldsymbol{B}$	Operators
$\lambda_{\max}(X)$, $\lambda_{\min}(X)$	Largest and smallest eigenvalue
$\\|X\\|$	Operator norm
$\\|X\\|_F$	Frobenius norm
$\\|X\\|_1$	1-Schatten (nuclear) norm
$\\|X\\|_{\psi_1}$	ψ_1 Orlicz norm
$\\|X\\|_{\psi_2}$	ψ_2 Orlicz norm
$\langle X, Y\rangle$	Inner product associated to the Frobenius norm
$\kappa(X) = \\|X\\| \cdot \\|X^{-1}\\|$	Condition number of an operator or a matrix
$\otimes$	Tensor product
$\log(x)$	$\log(x) \stackrel{\text{def}}{=} \max\{1, \ln(x)\}$

Let $Q \in \mathbb{H}_{++}(d)$, then we define an Hermitian operator on $\mathbb{H}(d)$

$$\boldsymbol{A}_Q \stackrel{\text{def}}{=} \left(-\frac{1}{2}\boldsymbol{dT}_Q^Q\right)^{1/2} \tag{A.1}$$

existing due to Lemma A.3 by Kroshnin et al. [2021].

LEMMA A.1 (Properties of of $\boldsymbol{A}$). *Consider $\boldsymbol{A}_Q$ from (A.1). It holds*

$$\|\boldsymbol{A}_Q\| = \frac{1}{2\sqrt{\lambda_{\min}(Q)}}, \quad \left\|\boldsymbol{A}_Q^{-1}\right\| = 2\sqrt{\lambda_{\max}(Q)}. \tag{A.2}$$

Moreover, there is a unitary operator $\boldsymbol{U}_Q \in U(\mathbb{H}(d))$, where $U(\mathbb{H}(d))$ is the set of unitary operators on $\mathbb{H}(d)$, s.t. for any $X \in \mathbb{H}(d)$ the following equation holds:

$$(\boldsymbol{U}_Q\boldsymbol{A}_Q)X = Q^{1/2}\boldsymbol{dT}_Q^Q(X). \tag{A.3}$$

PROOF. First we prove (A.3). Without loss of generality we set the matrix Q to be diagonal, $Q = \operatorname{diag}(q_1, \dots, q_d)$, and write down the explicit form of the operator $\boldsymbol{dT}_Q^Q(X)$ coming from formula (A.2) by Kroshnin et al. [2021]. In what follows we denote the scalar product as $\langle\cdot,\cdot\rangle$. Taking into account that $X = (X_{ij})$ with $i, j \in \{1, \dots, d\}$, we get

$$-\left\langle \boldsymbol{dT}_Q^Q(X), X\right\rangle = \sum_{i,j=1}^d \frac{X_{ij}}{q_i + q_j} X_{ij} = \sum_{i,j=1}^d (q_i + q_j)\left(\frac{X_{ij}}{q_i + q_j}\right)^2$$

$$= 2\sum_{i,j=1}^d \left(\sqrt{q_i}\frac{X_{ij}}{q_i + q_j}\right)^2 = 2\left\|Q^{1/2}\boldsymbol{dT}_Q^Q(X)\right\|_F^2,$$

thus $\|\boldsymbol{A}_Q(X)\|_F = \left\|Q^{1/2}\boldsymbol{dT}_Q^Q(X)\right\|_F$, and these operators are unitary equivalent.

Now we prove (A.2). The above chain of equations ensures

$$\|\boldsymbol{A}_Q(X)\|_F^2 = \frac{1}{2}\sum_{i,j=1}^{d}\frac{X_{ij}^2}{q_i+q_j}.$$

This yields

$$\frac{1}{4\lambda_{\max}(Q)}\|X\|_F^2 \le \|\boldsymbol{A}_Q(X)\|_F^2 \le \frac{1}{4\lambda_{\min}(Q)}\|X\|_F^2.$$

One can show in the same way as in the proof of Corollary A.2 by Kroshnin et al. [2021] that these inequalities are sharp. The result follows immediately. □

LEMMA A.2 (Local Lipschitz continuity of $\boldsymbol{A}_Q$). *Let $Q_0, Q_1 \in \mathbb{H}_{++}(d)$ and $\boldsymbol{A}_i = \boldsymbol{A}_{Q_i}$, $i = 1, 2$. We set $Q_1' = Q_0^{-1/2}Q_1Q_0^{-1/2}$. If $\|Q_1' - I\| \le 1/2$, then*

$$\|\boldsymbol{A}_1 - \boldsymbol{A}_0\| \le \|Q_1' - I\| \cdot \|\boldsymbol{A}_0\|.$$

PROOF. We recall that by Lemma A.3 by Kroshnin et al. [2021] $Q \mapsto \boldsymbol{dT}_Q^Q$ is monotone and (-1)-homogeneous. Then $Q \mapsto \boldsymbol{A}_Q$ is antimonotone and $(-\frac{1}{2})$-homogeneous, thus

$$\left(1 - \frac{1}{2}\|Q_1' - I\|\right)\boldsymbol{A}_0 \preccurlyeq \frac{1}{\sqrt{\lambda_{\max}(Q_1')}}\boldsymbol{A}_0 \preccurlyeq \boldsymbol{A}_1,$$

$$\boldsymbol{A}_1 \preccurlyeq \frac{1}{\sqrt{\lambda_{\min}(Q_1')}}\boldsymbol{A}_0 \preccurlyeq \left(1 + \|Q_1' - I\|\right)\boldsymbol{A}_0.$$

This yields $\|\boldsymbol{A}_1 - \boldsymbol{A}_0\| \le \|Q_1' - I\| \cdot \|\boldsymbol{A}_0\|$. □

Now we are ready to prove Lemma 2.1.

PROOF OF LEMMA 2.1. To prove (2.1) we use the quadratic approximation from Lemma A.6 by Kroshnin et al. [2021]. We set $Q_0 = Q_0$, $Q_1 = Q_1$, and $S = Q_1$. This yields

$$-\frac{2}{\left(1+\lambda_{\max}^{1/2}(Q_{21}')\right)^2}\left\langle \boldsymbol{dT}_{Q_1}^{Q_1}(Q_2 - Q_1), Q_2 - Q_1\right\rangle \le d_{\mathrm{B}}^2(Q_2, Q_1)$$

$$\le -\frac{2}{\left(1+\lambda_{\min}^{1/2}(Q_{21}')\right)^2}\left\langle \boldsymbol{dT}_{Q_1}^{Q_1}(Q_2 - Q_1), Q_2 - Q_1\right\rangle,$$

where $Q_{21}' = Q_1^{-1/2}Q_2Q_1^{-1/2}$. Due to monotonicity and homogeneity of the operator $\boldsymbol{dT}_Q^S$ (see (IV) and (V) in Lemma A.3 by Kroshnin et al. [2021]), it holds that

$$\boldsymbol{dT}_{Q_1}^{Q_1} \preccurlyeq \boldsymbol{dT}_{\lambda_{\max}(Q_1')Q_0}^{\lambda_{\max}(Q_1')Q_0} = \frac{1}{\lambda_{\max}(Q_1')}\boldsymbol{dT}_{Q_0}^{Q_0},$$

$$\boldsymbol{dT}_{Q_1}^{Q_1} \succcurlyeq \boldsymbol{dT}_{\lambda_{\min}(Q_1')Q_0}^{\lambda_{\min}(Q_1')Q_0} = \frac{1}{\lambda_{\min}(Q_1')}\boldsymbol{dT}_{Q_0}^{Q_0}.$$

Combining these inequalities with (A.3) we get

$$\text{(A.4)} \qquad \frac{4\|\boldsymbol{A}_0(Q_2 - Q_1)\|_F^2}{\lambda_{\max}(Q_1')\left(1 + \lambda_{\max}^{1/2}(Q_{21}')\right)^2} \le d_{\mathrm{B}}^2(Q_2, Q_1) \le \frac{4\|\boldsymbol{A}_0(Q_2 - Q_1)\|_F^2}{\lambda_{\min}(Q_1')\left(1 + \lambda_{\min}^{1/2}(Q_{21}')\right)^2}.$$

The following bounds on $\lambda_{\min}(Q_1')$ and $\lambda_{\max}(Q_1')$ hold

$$1 - \|Q_1' - I\| \le \lambda_{\min}(Q_1') \le \lambda_{\max}(Q_1') \le 1 + \|Q_1' - I\|.$$

Assumption $\|Q_1' - I\| \le \frac{1}{2}$ yields

$$\lambda_{\max}^{-1/2}(Q_1') \ge 1 - \frac{1}{2}\|Q_1' - I\|, \quad \lambda_{\min}^{-1/2}(Q_1') \le 1 + 2\|Q_1' - I\|.$$

Further, for Q_{21}' assumptions $\|Q_1' - I\| \le \frac{1}{2}$ and $\|Q_2' - I\| \le \frac{1}{2}$ yield

$$\lambda_{\min}(Q_{21}') \ge \frac{\lambda_{\min}(Q_2')}{\lambda_{\max}(Q_1')} \ge 1 - \|Q_1' - I\| - \|Q_2' - I\|,$$

$$\lambda_{\max}(Q_{21}') \le \frac{\lambda_{\max}(Q_2')}{\lambda_{\min}(Q_1')} \le 1 + 2\|Q_1' - I\| + 2\|Q_2' - I\|.$$

Then

$$2\left(1 + \lambda_{\max}^{1/2}(Q_{21}')\right)^{-1} \ge 1 - \frac{1}{2}\|Q_1' - I\| - \frac{1}{2}\|Q_2' - I\|,$$

$$2\left(1 + \lambda_{\min}^{1/2}(Q_{21}')\right)^{-1} \le 1 + \|Q_1' - I\| + \|Q_2' - I\|.$$

Thus we obtain

$$2\lambda_{\max}^{-1/2}(Q_1')\left(1 + \lambda_{\max}^{1/2}(Q_{21}')\right)^{-1} \ge 1 - \|Q_1' - I\| - \frac{1}{2}\|Q_2' - I\|,$$

$$2\lambda_{\min}^{-1/2}(Q_1')\left(1 + \lambda_{\min}^{1/2}(Q_{21}')\right)^{-1} \le 1 + 4\|Q_1' - I\| + 2\|Q_2' - I\|.$$

Combining these inequalities with (A.4) we get the result. □

The next lemma is crucial for the proof of Lemma 2.2. We recall $\boldsymbol{D}_w$ and $\hat{\boldsymbol{F}}_w$ defined in (2.2), and (2.3), respectively.

LEMMA A.3 (Bounds on $\boldsymbol{D}_w$ and $\hat{\boldsymbol{F}}_w$). *Let $q_w + f_w \le \frac{1}{2}$, then*

$$\frac{1}{1 + 2(q_w + f_w)}\boldsymbol{F} \preccurlyeq \boldsymbol{D}_w \preccurlyeq \frac{1}{1 - (q_w + f_w)}\boldsymbol{F},$$

$$\frac{1}{1 + \frac{5}{2}(q_w + f_w)}\boldsymbol{F} \preccurlyeq \hat{\boldsymbol{F}}_w \preccurlyeq \frac{1}{1 - \frac{3}{2}(q_w + f_w)}\boldsymbol{F}.$$

PROOF. Lemma A.4 by Kroshnin et al. [2021] ensures that for $Q_t = (1-t)Q_* + tQ_w$, and $Q' = Q_*^{-1/2} Q_w Q_*^{-1/2}$

$$\frac{1}{1 - q_w}\boldsymbol{dT}_{Q_*}^S \preccurlyeq \int_0^1 \boldsymbol{dT}_{Q_t}^S \, dt \preccurlyeq \frac{1}{1 + \frac{3}{4}q_w}\boldsymbol{dT}_{Q_*}^S.$$

This yields

$$\frac{1}{1+\frac{3}{4}q_w}\boldsymbol{F}_w \preccurlyeq \boldsymbol{D}_w \preccurlyeq \frac{1}{1-q_w}\boldsymbol{F}_w.$$

Further,

$$(1-f_w)\,\boldsymbol{F} \preccurlyeq \boldsymbol{F}_w \preccurlyeq (1+f_w)\,\boldsymbol{F}. \tag{A.5}$$

Combining these bounds, we obtain:

$$\frac{1}{1+2(q_w+f_w)}\boldsymbol{F} \preccurlyeq \frac{1-f_w}{1+\frac{3}{4}q_w}\boldsymbol{F} \preccurlyeq \boldsymbol{D}_w \preccurlyeq \frac{1+f_w}{1-q_w}\boldsymbol{F} \preccurlyeq \frac{1}{1-(q_w+f_w)}\boldsymbol{F}.$$

In a similar way,

$$\frac{1}{1+2q_w}\boldsymbol{F}_w \preccurlyeq (1+q_w)^{-3/2}\boldsymbol{F}_w \preccurlyeq \hat{\boldsymbol{F}}_w \preccurlyeq (1-q_w)^{-3/2}\boldsymbol{F}_w \preccurlyeq \frac{1}{1-\frac{3}{2}q_w}\boldsymbol{F}_w,$$

$$\frac{1}{1+\frac{5}{2}(q_w+f_w)}\boldsymbol{F} \preccurlyeq \frac{1-f_w}{1+2q_w}\boldsymbol{F} \preccurlyeq \hat{\boldsymbol{F}}_w \preccurlyeq \frac{1+f_w}{1-\frac{3}{2}q_w}\boldsymbol{F} \preccurlyeq \frac{1}{1-\frac{3}{2}(q_w+f_w)}\boldsymbol{F}.$$

□

APPENDIX B: GAUSSIAN APPROXIMATION

The first result in this section investigates the properties of $\varkappa(\cdot)$ introduced by (3.2).

LEMMA B.1 (Bounds on $\varkappa(\cdot)$). *Let $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ be symmetric operators, s.t.* $\|\boldsymbol{\Phi}-\boldsymbol{\Psi}\|_1 \le \frac{\Lambda_2^2(\boldsymbol{\Psi})}{4\|\boldsymbol{\Psi}\|}$. *Then the following bounds hold:*

$$\varkappa(\boldsymbol{\Phi}) \le 2\varkappa(\boldsymbol{\Psi}), \quad \operatorname{tr}\boldsymbol{\Phi} \le \tfrac{5}{4}\operatorname{tr}\boldsymbol{\Psi}.$$

PROOF. Note, that $\Lambda_2^2(\boldsymbol{\Psi}) \le \Lambda_1^2(\boldsymbol{\Psi}) \le \|\boldsymbol{\Psi}\|\operatorname{tr}(\boldsymbol{\Psi})$ and therefore,

$$\operatorname{tr}(\boldsymbol{\Phi}) \le \operatorname{tr}(\boldsymbol{\Psi}) + \|\boldsymbol{\Phi}-\boldsymbol{\Psi}\|_1 \le \frac{5}{4}\operatorname{tr}(\boldsymbol{\Psi}).$$

By the definition of $\Lambda_r^2(\cdot)$, $\Lambda_r^2(\boldsymbol{\Phi}) \ge \Lambda_r^2(\boldsymbol{\Psi}) - \|\boldsymbol{\Psi}\|\|\boldsymbol{\Phi}-\boldsymbol{\Psi}\|_1$, then

$$\Lambda_1^2(\boldsymbol{\Phi})\,\Lambda_2^2(\boldsymbol{\Phi}) \ge \Lambda_1^2(\boldsymbol{\Psi})\,\Lambda_2^2(\boldsymbol{\Psi}) - \left(\Lambda_1^2(\boldsymbol{\Psi}) + \Lambda_2^2(\boldsymbol{\Psi})\right)\|\boldsymbol{\Psi}\|\|\boldsymbol{\Phi}-\boldsymbol{\Psi}\|_1.$$

Then it follows that

$$\begin{aligned}\varkappa(\boldsymbol{\Phi}) &\le \varkappa(\boldsymbol{\Psi})\left(1-\tfrac{\Lambda_1^2(\boldsymbol{\Psi})+\Lambda_2^2(\boldsymbol{\Psi})}{\Lambda_1^2(\boldsymbol{\Psi})\Lambda_2^2(\boldsymbol{\Psi})}\|\boldsymbol{\Psi}\|\|\boldsymbol{\Phi}-\boldsymbol{\Psi}\|_1\right)^{-1} \\ &\le \varkappa(\boldsymbol{\Psi})\left(1-2\tfrac{\|\boldsymbol{\Psi}\|}{\Lambda_2^2(\boldsymbol{\Psi})}\|\boldsymbol{\Phi}-\boldsymbol{\Psi}\|_1\right)^{-1} \le 2\varkappa(\boldsymbol{\Psi}).\square\end{aligned}$$

Now we are ready to prove the concentration results for Q_n and Q_u. Both facts follow immediately from Lemma 2.2.

COROLLARY B.2 (Concentration of Q_n and Q_u). *Let $f_n \le \frac{1}{2}$. Under Assumption (T) for all* $\mathrm{x} > 0$ *s.t.* $\varepsilon_T(\mathrm{x}) \le \frac{\lambda_{\min}(\boldsymbol{\xi})}{3\|Q_*\|}$, *it holds with P-probability at least* $1-\delta_T(\mathrm{x})$ *that*

$$\left\|Q_*^{-1/2}Q_nQ_*^{-1/2} - I\right\|_F \le \varepsilon_Q(\mathrm{x}) \stackrel{\text{def}}{=} \frac{4\|Q_*\|}{\lambda_{\min}(\boldsymbol{\xi})}\varepsilon_T(\mathrm{x}). \tag{B.1}$$

Let $q_n + f_n \le \frac{1}{2}$, *and* $f_u \le \frac{1}{2}$. *Under Assumption* ($\hat{T}$) *for all* $\mathrm{x}, \mathrm{s} > 0$ *s.t.* $\varepsilon_{\hat{T}}(\mathrm{x};\mathrm{s}) \le \frac{\lambda_{\min}(\boldsymbol{\xi})}{81\|Q_*\|}$, *it holds with* P*-probability at least* $1 - \alpha_{\hat{T}}(\mathrm{s})$ *and with* P_u*-probability at least* $1 - \delta_{\hat{T}}(\mathrm{x};\mathrm{s})$ *that*

$$\left\| Q_n^{-1/2} Q_u Q_n^{-1/2} - I \right\|_F \le \varepsilon_{\hat{Q}}(\mathrm{x};\mathrm{s}) \stackrel{\text{def}}{=} \frac{108\|Q_*\|}{\lambda_{\min}(\boldsymbol{\xi})} \varepsilon_{\hat{T}}(\mathrm{x};\mathrm{s}). \tag{B.2}$$

PROOF. Bound (B.1) follows immediately from Lemma 2.2 by setting $w_i = 1$ for all i. Now we consider (B.2). Set $w_i = u_i$ for all i. Let for any $X \in \mathbb{H}(d)$

$$\hat{\boldsymbol{\xi}}(X) \stackrel{\text{def}}{=} Q_n^{1/2} \hat{\boldsymbol{F}} \left(Q_n^{1/2} X Q_n^{1/2} \right) Q_n^{1/2}, \quad \boldsymbol{\xi}_u(X) \stackrel{\text{def}}{=} Q_n^{1/2} \boldsymbol{F}_u \left(Q_n^{1/2} X Q_n^{1/2} \right) Q_n^{1/2}.$$

We note that conditions $q_n + f_n \le \frac{1}{2}$, and $f_u \le \frac{1}{2}$ yield $\|Q_n\| \le \frac{3}{2}\|Q_*\|$ and

$$\begin{aligned}
\langle X, \boldsymbol{\xi}_u(X) \rangle &\ge \frac{1}{2} \langle X, \hat{\boldsymbol{\xi}}(X) \rangle \\
&= \frac{1}{2} \langle Q_n^{1/2} X Q_n^{1/2}, \hat{\boldsymbol{F}}(Q_n^{1/2} X Q_n^{1/2}) \rangle \\
&\ge \frac{1}{2} \frac{1}{1 + \frac{5}{4}} \langle Q_n^{1/2} X Q_n^{1/2}, \boldsymbol{F}(Q_n^{1/2} X Q_n^{1/2}) \rangle \\
&\ge \frac{2}{9} \lambda_{\min}(\boldsymbol{\xi}) \left\| Q_*^{-1/2} Q_n^{1/2} X Q_n^{1/2} Q_*^{-1/2} \right\|_F^2 \\
&\ge \frac{1}{18} \lambda_{\min}(\boldsymbol{\xi}) \|X\|_F^2.
\end{aligned}$$

Thus $\lambda_{\min}(\boldsymbol{\xi}_u) \ge \frac{1}{18}\lambda_{\min}(\boldsymbol{\xi})$ and $\frac{\|Q_n\|}{\lambda_{\min}(\boldsymbol{\xi}_u)} \le 27 \frac{\|Q_*\|}{\lambda_{\min}(\boldsymbol{\xi})}$. Then (B.2) follows from Lemma 2.2. □

Next, we define the counterparts of η_w coming from (2.5) in the real world and the bootstrap one:

$$\eta_n \stackrel{\text{def}}{=} 2\sqrt{\kappa(\boldsymbol{F})}(q_n + f_n), \quad \eta_u \stackrel{\text{def}}{=} 2\sqrt{\kappa(\hat{\boldsymbol{F}})}(q_u + f_u), \tag{B.3}$$

with q_n, f_n coming from (3.5) and q_u, f_u coming from (3.12). The next lemma is trivial. It validates the concentrations of η_n and η_u. We provide it only for the completeness of presentation.

LEMMA B.3 (Concentrations of η_n and η_u). *Let Assumptions* (T) *and* (F) *be fulfilled. We recall* $\delta_\eta(\mathrm{x})$ *defined in* (3.7) *and* $\varepsilon_\eta(\mathrm{x})$ *defined in* (3.8). *For any* $\mathrm{x} \in \mathbb{R}_+^2$ *s.t.* $\varepsilon_\eta(\mathrm{x}) \le 1$*:*

$$\mathbb{P}\left\{ \eta_n > \varepsilon_\eta(\mathrm{x}) \right\} \le \delta_\eta(\mathrm{x}).$$

Let Assumptions ($\hat{T}$) *and* ($\hat{F}$) *be fulfilled, and let* $f_n + q_n \le \frac{1}{2}$. *We recall* $\delta_{\hat{\eta}}(\mathrm{x};\mathrm{s})$ *defined in* (3.15) *and* $\varepsilon_{\hat{\eta}}(\mathrm{x};\mathrm{s})$ *defined in* (3.16). *Then with* P*-probability at least* $1 - \alpha_{\hat{T}}(\mathrm{s}) - \alpha_{\hat{F}}(\mathrm{s})$ *for all* $\mathrm{x}, \mathrm{s} \in \mathbb{R}_+^2$ *s.t.* $\varepsilon_{\hat{\eta}}(\mathrm{x};\mathrm{s}) \le 1$ *it holds that*

$$\mathbb{P}_u\left\{ \eta_u > \varepsilon_{\hat{\eta}}(\mathrm{x};\mathrm{s}) \right\} \le \delta_{\hat{\eta}}(\mathrm{x};\mathrm{s}).$$

PROOF. The proof of the first result follows immediately from the definitions and Corollary B.2. The proof of the second one relies on the fact that

$$\kappa(\hat{\boldsymbol{F}}) \le \frac{1 + \frac{5}{4}}{1 - \frac{3}{4}} \kappa(\boldsymbol{F}) \le 9\kappa(\boldsymbol{F}),$$

whenever $q_n + f_n \le \frac{1}{2}$. □

Now we are ready to prove Lemma 3.2 and Lemma 3.3.

PROOF OF LEMMA 3.2. The proof is split into two parts.

Frobenius case. We have to check the conditions of Theorem 3.1. Denote

$$X = \sqrt{n}\|Q_n - Q_*\|_F, \quad Y = \sqrt{n}\left\|\boldsymbol{F}^{-1}\overline{T}_n\right\|_F, \quad G = Z.$$

Condition (GAR-I) follows from

$$|X - Y| \overset{\text{by } L.2.3}{\leq} \eta_n Y \overset{\text{by } L.B.3}{\leq} \varepsilon_\eta(\mathrm{x}) Y.$$

The last inequality holds with P-probability at least $1-\delta_\eta(\mathrm{x})$ for all x s.t. $\varepsilon_\eta(\mathrm{x}) \leq \frac{1}{2}$. Condition (GAR-II) is fulfilled due to Assumption (Z). Using notations accepted in Theorem 3.1 we get

$$\eta = \varepsilon_\eta(\mathrm{x}), \quad m = 0, \quad \delta = \delta_\eta(\mathrm{x}), \quad \varDelta = \varDelta_Z.$$

The result of the lemma follows immediately.

Bures–Wasserstein case. We set

$$X = \sqrt{n} d_{\mathrm{B}}(Q_n, Q_*), \quad Y = \sqrt{n}\left\|\boldsymbol{A}\boldsymbol{F}^{-1}\overline{T}_n\right\|_F, \quad G = \boldsymbol{A}Z.$$

Condition (GAR-I) holds due to

$$|X - Y| \overset{\text{by } L.2.4}{\leq} 3\sqrt{\kappa(Q_*)}\eta_n Y \overset{\text{by } L.B.3}{\leq} 3\sqrt{\kappa(Q_*)}\varepsilon_\eta(\mathrm{x}) Y,$$

where the last inequality holds with P-probability at least $1-\delta_\eta(\mathrm{x})$ for all x, s.t. $\varepsilon_\eta(\mathrm{x}) \leq \frac{1}{6\sqrt{\kappa(Q_*)}}$. Condition (GAR-II) is fulfilled due to Assumption (Z). In terms of notations accepted in Theorem 3.1 we get

$$\eta = 3\sqrt{\kappa(Q_*)}\varepsilon_\eta(\mathrm{x}), \quad m = 0, \quad \delta = \delta_\eta(\mathrm{x}), \quad \varDelta = \varDelta_Z.$$

Then the result follows immediately. □

PROOF OF LEMMA 3.3. We have to check the conditions of Theorem 3.1. The proof is split into three parts. First we investigate Frobenius case, than we consider Bures–Wasserstein case, and, finally, we obtain the probabilistic bounds.

Frobenius case. We set

$$X = \sqrt{n}\|Q_n - Q_u\|_F, \quad Y = \sqrt{n}\left\|\hat{\boldsymbol{F}}^{-1}\overline{T}_u\right\|_F, \quad G = \hat{Z}.$$

The proof is exactly the same as in Frobenius case in Lemma 3.2. In terms of notations accepted in Theorem 3.1 we get

$$\eta = \varepsilon_{\hat{\eta}}(\mathrm{x};\mathrm{s}), \quad m = 0, \quad \delta = \delta_{\hat{\eta}}(\mathrm{x};\mathrm{s}), \quad \varDelta = \delta_{\hat{T}}(\mathrm{s}).$$

The result follows.

Bures–Wasserstein case. Denote

$$X = \sqrt{n} d_{\mathrm{B}}(Q_u, Q_n), \quad Y = \sqrt{n}\left\|\hat{\boldsymbol{A}}\hat{\boldsymbol{F}}^{-1}\overline{T}_u\right\|_F, \quad G = \hat{\boldsymbol{A}}\hat{Z}.$$

First we note that $\kappa(Q_n) \leq 3\kappa(Q_*)$ since $q_n \leq \frac{1}{2}$. The rest of proof is similar to the proof of Bures–Wasserstein case in Lemma 3.2. This yields

$$\eta = 3\sqrt{3\kappa(Q_*)}\varepsilon_{\hat{\eta}}(\mathrm{x};\mathrm{s}), \quad m = 0, \quad \delta = \delta_{\hat{\eta}}(\mathrm{x};\mathrm{s}), \quad \varDelta = \delta_{\hat{T}}(\mathrm{s}).$$

The result follows from Theorem 3.1. □

APPENDIX C: BOOTSTRAP VALIDITY

In this section we present the proofs and auxiliary results concerning the bootstrap validity. The first lemma provides a deviation bound on $\hat{\boldsymbol{\Sigma}}$.

LEMMA C.1 (Deviations of $\hat{\boldsymbol{\Sigma}}$). *Let* $q_n \le \frac{1}{2}$. *Then*

$$\left\| \hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_n \right\|_1 \le 2q_n \left(2 \operatorname{tr} \boldsymbol{\Sigma} + 2d^2 \|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}\| + d \right).$$

PROOF. Recall that $T_Q^S = S^{1/2}\left(S^{1/2} Q S^{1/2}\right)^{-1/2} S^{1/2}$ is antimonotone and $(-\frac{1}{2})$-homogeneous w.r.t. Q. Thus

$$\left(1 - \frac{1}{2} q_n\right) T_{Q_*}^S \preccurlyeq \frac{1}{\sqrt{1+q_n}} T_{Q_*}^S \preccurlyeq T_{Q_n}^S \preccurlyeq \frac{1}{\sqrt{1-q_n}} T_{Q_*}^S \preccurlyeq (1+q_n) T_{Q_*}^S$$

and

$$\left\| T_{Q_*}^S - T_{Q_n}^S \right\|_F \le q_n \left\| T_{Q_*}^S \right\|_F. \tag{C.1}$$

Then we obtain

$$\begin{aligned}
\left\| \hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_n \right\|_1 &\le \frac{1}{n} \sum_{i=1}^n \left\| \hat{T}_i \otimes \hat{T}_i - T_i \otimes T_i \right\|_1 \\
&\le \frac{1}{n} \sum_{i=1}^n \left(2 \left\| \hat{T}_i - T_i \right\|_F \|T_i\|_F + \left\| \hat{T}_i - T_i \right\|_F^2 \right) \\
&\le \frac{1}{n} \sum_{i=1}^n \left(2q_n \|T_i + I\|_F \|T_i\|_F + q_n^2 \|T_i + I\|_F^2 \right) \\
&\le \frac{q_n}{n} \sum_{i=1}^n \left(4 \|T_i\|_F^2 + 2d \right) \\
&= 2q_n \left(2 \operatorname{tr} \hat{\boldsymbol{\Sigma}}_n + d \right) \\
&\le 2q_n \left(2 \operatorname{tr} \boldsymbol{\Sigma} + 2d^2 \|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}\| + d \right). \qquad \text{(C.2)}
\end{aligned}$$

□

Next, we ensure the concentration of the covariance operator $\hat{\boldsymbol{\Xi}}$.

LEMMA C.2 (Concentration of $\hat{\boldsymbol{\Xi}}$). *Let* $q_n + f_n \le \frac{1}{2}$. *Then*

$$\left\| \hat{\boldsymbol{\Xi}} - \boldsymbol{\Xi} \right\|_1 \le \left\| \boldsymbol{F}^{-1} \right\|^2 \left(42 d^2 \|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}\| + 71 (q_n + f_n) (\operatorname{tr} \boldsymbol{\Sigma} + d) \right).$$

PROOF. First we note that the proof of Theorem 2.2 (B) by Kroshnin et al. [2021] (see paragraph *Convergence of* $\hat{\boldsymbol{F}}$ *to* $\boldsymbol{F}$ therein) being combined with (A.5) ensures

$$\hat{\boldsymbol{F}} \overset{\text{proof of T.2.2 (B)}}{\preccurlyeq} (1 - q_n)^{-3/2} \boldsymbol{F}_n \overset{\text{(A.5)}}{\preccurlyeq} \frac{1 + f_n}{(1 - q_n)^{3/2}} \boldsymbol{F} \preccurlyeq \frac{1}{1 - 2q_n - f_n} \boldsymbol{F},$$

$$\hat{\boldsymbol{F}} \overset{\text{proof of T.2.2 (B)}}{\succcurlyeq} (1 + q_n)^{-3/2} \boldsymbol{F}_n \overset{\text{(A.5)}}{\succcurlyeq} \frac{1 - f_n}{(1 + q_n)^{3/2}} \boldsymbol{F} \succcurlyeq \frac{1}{1 + 3q_n + 3f_n} \boldsymbol{F}.$$

Now we consider

$$\left\|\hat{\Xi}-\Xi\right\|_1 = \left\|\hat{\boldsymbol{F}}^{-1}\left(\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\right)\hat{\boldsymbol{F}}^{-1}\right\|_1 + \left\|\hat{\boldsymbol{F}}^{-1}\boldsymbol{\Sigma}\hat{\boldsymbol{F}}^{-1}-\boldsymbol{F}^{-1}\boldsymbol{\Sigma}\boldsymbol{F}^{-1}\right\|$$
$$\leq \left\|\hat{\boldsymbol{F}}^{-1}\right\|^2\left\|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\right\|_1 + \left\|\hat{\boldsymbol{F}}^{-1}-\boldsymbol{F}^{-1}\right\|\left(\left\|\hat{\boldsymbol{F}}^{-1}\right\|+\left\|\boldsymbol{F}^{-1}\right\|\right)\operatorname{tr}\boldsymbol{\Sigma}$$
$$\leq 14\left\|\boldsymbol{F}^{-1}\right\|^2\left\|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\right\|_1 + 15(q_n+f_n)\left\|\boldsymbol{F}^{-1}\right\|^2\operatorname{tr}\boldsymbol{\Sigma}.$$

To continue this inequality we first note that due to Lemma C.1

$$\left\|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\right\|_1 \leq \left\|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}_n\right\|_1 + d^2\|\boldsymbol{\Sigma}_n-\boldsymbol{\Sigma}\|$$
$$\leq 2q_n\left(2\operatorname{tr}\boldsymbol{\Sigma}+d\right)+3d^2\|\boldsymbol{\Sigma}_n-\boldsymbol{\Sigma}\|.$$

Finally,

$$\left\|\hat{\Xi}-\Xi\right\|_1 \leq \left\|\boldsymbol{F}^{-1}\right\|^2\left(28q_n\left(2\operatorname{tr}\boldsymbol{\Sigma}+d\right)+42d^2\|\boldsymbol{\Sigma}_n-\boldsymbol{\Sigma}\|+15(q_n+f_n)\operatorname{tr}\boldsymbol{\Sigma}\right)$$
$$\leq \left\|\boldsymbol{F}^{-1}\right\|^2\left(42d^2\|\boldsymbol{\Sigma}_n-\boldsymbol{\Sigma}\|+71(q_n+f_n)\left(\operatorname{tr}\boldsymbol{\Sigma}+d\right)\right).$$

□

Now we are ready to prove the main results of the paper. Before doing that we note the following fact.

REMARK C.3. *Under the assumptions of Theorem 4.1 with P-probability at least $1-\delta(\mathrm{y})$ it holds that $\eta_n \leq \varepsilon_\eta(\mathrm{y})$ and $\|\boldsymbol{\Sigma}_n-\boldsymbol{\Sigma}\| \leq \varepsilon_\Sigma(\mathrm{y})$ simultaneously. In particular, these inequalities imply $f_n+q_n \leq \frac{1}{2}$ and $\gamma(\hat{\Xi}) \leq 2\gamma(\Xi)$ (due to Lemmas B.1 and C.2), cf. Remark 3.4.*

PROOF OF THEOREM 4.1. To prove the result we use the Gaussian approximations from Lemma 3.2 and Lemma 3.3. For all $z \geq 0$ with P-probability at least $1-\delta(\mathrm{y})-\hat{\alpha}(\mathrm{s})$ it holds

$$\left|\mathbb{P}\left\{\sqrt{n}\|Q_n-Q_*\|_F \leq z\right\}-\mathbb{P}\left\{\|Z\|_F \leq z\right\}\right| \overset{\text{by L.3.2 (F)}}{\leq} \Omega_{\mathrm{F}}(n),$$
$$\left|\mathbb{P}_u\left\{\sqrt{n}\|Q_u-Q_n\|_F \leq z\right\}-\mathbb{P}\left\{\|\hat{Z}\|_F \leq z\right\}\right| \overset{\text{by L.3.3 (F)}}{\leq} \hat{\Omega}_{\mathrm{F}}(n;\mathrm{s}).$$

This yields for all $z \geq 0$

$$\left|\mathbb{P}\left\{\sqrt{n}\|Q_n-Q_*\|_F \leq z\right\}-\mathbb{P}_u\left\{\sqrt{n}\|Q_u-Q_n\|_F \leq z\right\}\right|$$
$$\leq \left|\mathbb{P}\left\{\|Z\|_F \leq z\right\}-\mathbb{P}\left\{\|\hat{Z}\|_F \leq z\right\}\right|+\Omega_{\mathrm{F}}(n)+\hat{\Omega}_{\mathrm{F}}(n;\mathrm{s}) \tag{C.3}$$

The final step is a Gaussian comparison. We use Corollary 2.3 by Götze et al. [2019]. It claims that for all $z \geq 0$

$$\left|\mathbb{P}\left\{\|Z\|_F \leq z\right\}-\mathbb{P}\left\{\|\hat{Z}\|_F \leq z\right\}\right| \leq \mathtt{C}\left(\varkappa(\Xi)+\varkappa(\hat{\Xi})\right)\|\Xi-\hat{\Xi}\|_1. \tag{C.4}$$

Lemma C.2 and Remark C.3 yield with P-probability at least $1-\delta(\mathrm{y})$

$$\left\|\Xi-\hat{\Xi}\right\|_1 \leq \mathtt{C}\left\|\boldsymbol{F}^{-1}\right\|^2\varepsilon_\Xi(\mathrm{y}). \tag{C.5}$$

Furthermore, $\varkappa(\hat{\Xi}) \leq 2\varkappa(\Xi)$ by Lemma B.1. Combining this fact with (C.3) and (C.4) we get the result. □

PROOF OF THEOREM 4.2. The proof is similar to the Frobenius-case. For all $z \ge 0$ with P-probability at least $1 - \delta(\mathrm{y}) - \hat{\alpha}(\mathrm{s})$ it holds

$$\left|\mathbb{P}\left\{d_{\mathrm{B}}(Q_n, Q_*) \le z\right\} - \mathbb{P}\left\{\|\boldsymbol{A}Z\|_F \le z\right\}\right| \overset{\text{by L.3.2(BW)}}{\le} \Omega_{\mathrm{B}}(n),$$

$$\left|\mathbb{P}_u\left\{\sqrt{n}d_{\mathrm{B}}(Q_u, Q_n) \le z\right\} - \mathbb{P}\left\{\left\|\hat{\boldsymbol{A}}\hat{Z}\right\|_F \le z\right\}\right| \overset{\text{by L.3.3(BW)}}{\le} \hat{\Omega}_{\mathrm{B}}(n;\mathrm{s}).$$

This yields

$$\begin{aligned}\text{(C.6)}\qquad &\left|\mathbb{P}\left\{d_{\mathrm{B}}(Q_n, Q_*) \le z\right\} - \mathbb{P}_u\left\{\sqrt{n}d_{\mathrm{B}}(Q_u, Q_n) \le z\right\}\right| \\ &\le \left|\mathbb{P}\left\{\|\boldsymbol{A}Z\|_F \le z\right\} - \mathbb{P}\left\{\left\|\hat{\boldsymbol{A}}\hat{Z}\right\|_F \le z\right\}\right| + \Omega_{\mathrm{B}}(n) + \hat{\Omega}_{\mathrm{B}}(n;\mathrm{s}).\end{aligned}$$

Gaussian comparison. To compare $\boldsymbol{A}\Xi\boldsymbol{A}^*$ and $\hat{\boldsymbol{A}}\Xi_n\hat{\boldsymbol{A}}^*$ we first recall that $\boldsymbol{A}$ and $\hat{\boldsymbol{A}}$ are self-adjoint and then use the same idea as in the proof of Theorem 4.1. First we apply Corollary 2.3 by Götze et al. [2019]: for all $z > 0$

$$\begin{aligned}\text{(C.7)}\qquad &\left|\mathbb{P}\left\{\|\boldsymbol{A}Z\|_F \le z\right\} - \mathbb{P}\left\{\left\|\hat{\boldsymbol{A}}\hat{Z}\right\|_F \le z\right\}\right| \\ &\le \mathtt{C}\left(\varkappa(\boldsymbol{A}\Xi\boldsymbol{A}) + \varkappa\left(\hat{\boldsymbol{A}}\hat{\Xi}\hat{\boldsymbol{A}}\right)\right)\left\|\boldsymbol{A}\Xi\boldsymbol{A} - \hat{\boldsymbol{A}}\hat{\Xi}\hat{\boldsymbol{A}}\right\|_1.\end{aligned}$$

Consider

$$\left\|\boldsymbol{A}\Xi\boldsymbol{A} - \hat{\boldsymbol{A}}\hat{\Xi}\hat{\boldsymbol{A}}\right\|_1 \le \|\boldsymbol{A}\|^2\left\|\Xi - \hat{\Xi}\right\|_1 + \left\|\boldsymbol{A} - \hat{\boldsymbol{A}}\right\|\left(\|\boldsymbol{A}\| + \left\|\hat{\boldsymbol{A}}\right\|\right)\operatorname{tr}\hat{\Xi}.$$

We note that with P-probability at least $1 - \delta_T(\mathrm{y})$

$$\|\boldsymbol{A} - \hat{\boldsymbol{A}}\| \overset{\text{by L.A.2}}{\le} \varepsilon_Q(\mathrm{y})\|\boldsymbol{A}\|.$$

The conditions of the theorem ensure

$$\|\hat{\boldsymbol{A}}\| \le \frac{3}{2}\|\boldsymbol{A}\|, \quad \operatorname{tr}\hat{\Xi} \overset{\text{by L.B.1}}{\le} \frac{5}{4}\operatorname{tr}\Xi.$$

Recall that by Lemma A.1 $\|\boldsymbol{A}\| = \frac{1}{2\lambda_{\min}^{1/2}(Q_*)}$. We combine these results with (C.5) and get

$$\begin{aligned}\left\|\boldsymbol{A}\Xi\boldsymbol{A} - \hat{\boldsymbol{A}}\hat{\Xi}\hat{\boldsymbol{A}}\right\|_1 &\le \frac{1}{4\lambda_{\min}(Q_*)}\left\|\Xi - \hat{\Xi}\right\|_1 + \frac{25}{8}\varepsilon_Q(\mathrm{y})\|\boldsymbol{A}\|^2\operatorname{tr}\Xi \\ &\le \frac{\mathtt{C}}{\lambda_{\min}(Q_*)}\left\|\boldsymbol{F}^{-1}\right\|^2\left(\varepsilon_Q(\mathrm{y})\operatorname{tr}\boldsymbol{\Sigma} + \varepsilon_\Xi(\mathrm{y})\right) \\ &\overset{\text{by (4.2)}}{\le} \frac{\mathtt{C}}{\lambda_{\min}^2(\boldsymbol{F})\lambda_{\min}(Q_*)}\varepsilon_\Xi(\mathrm{y}).\end{aligned}$$

The last step is to recall that $\varkappa\left(\hat{\boldsymbol{A}}\hat{\Xi}\hat{\boldsymbol{A}}\right) \le 2\varkappa(\boldsymbol{A}\Xi\boldsymbol{A})$ by Lemma B.1. Combining these bounds with (C.6) and (C.7) we get the result. □

APPENDIX D: SUB-EXPONENTIAL CASE

We begin with the proof of Lemma 5.1.

PROOF OF LEMMA 5.1. Note that 1) follows from

$$\left\| \|S\|^{1/2} \right\|_{\psi_2} \le \left\| \sqrt{\operatorname{tr} S} \right\|_{\psi_2} < +\infty. \tag{D.1}$$

The explicit formula $T^S_{Q_*} = Q_*^{-1/2} \left(Q_*^{1/2} S Q_*^{1/2} \right)^{1/2} Q_*^{-1/2}$ yields

$$\|T^S_{Q_*}\| \le \frac{\lambda^{1/2}_{\max}(Q_*)}{\lambda_{\min}(Q_*)} \|S\|^{1/2}.$$

Combining these facts with (D.1) we get 2) and 3). Result (III) in Lemma A.3 by Kroshnin et al. [2021] ensures

$$\|\boldsymbol{dT}^S_{Q_*}\| \le \frac{\lambda^{1/2}_{\max}\left(S^{1/2} Q_* S^{1/2}\right)}{2\lambda^2_{\min}(Q_*)} \le \frac{\lambda^{1/2}_{\max}(Q_*)}{2\lambda^2_{\min}(Q_*)} \|S\|^{1/2}. \tag{D.2}$$

This fact together with (D.1) ensures 4). □

Assumption (T) follows from Lemma 5.1 directly.

LEMMA D.1 (Assumption (T)). *Assumption 1 ensures that for all* $\mathrm{x} > 0$

$$\delta_T(\mathrm{x}) = e^{-\mathrm{x}}, \quad \varepsilon_T(\mathrm{x}) = \nu_T \frac{d + \sqrt{\mathrm{x}}}{\sqrt{n}},$$

with ν_T *coming from Lemma 5.1, 3).*

PROOF. Lemma 5.1 ensures T_i to be sub-Gaussian with parameter ν_T. Applying Theorem 2.1 by Hsu et al. [2012] with $\Sigma = I$, $\dim(I) = d^2 \times d^2$ we get the result. □

D.1. Assumptions ($\hat{T}$), (Z) and ($\hat{Z}$). The next lemma deals with concentrations of sub-exponential r.v. The first two results are well-known and we provide them for completeness of the text.

LEMMA D.2. *Let* $X_1, \dots, X_n$, $X_i \in \mathbb{R}_+$ *be i.i.d. sub-exponential r.v. with* $\|X\|_{\psi_1} = \nu$. *Then with probability at least* $1 - e^{-\mathrm{x}}$ *the following bound holds:*

$$\frac{1}{n} \sum_{i=1}^n X_i \le \mathbb{E}\, X + \mathtt{C} \nu \max\left\{ \frac{\mathrm{x}}{n}, \sqrt{\frac{\mathrm{x}}{n}} \right\}. \tag{I}$$

Further, with probability at least $1 - e^{-\mathrm{x}}$

$$\max_i X_i \le \nu\big(\mathrm{x} + \ln(2n)\big). \tag{II}$$

And finally, for any $p \ge 1$ *there exists a constant* $C = C(p) > 0$ *such that*

$$\mathbb{E}\, X^p \le C\, \mathbb{E}\, X \left(\nu \log\left(\frac{\nu}{\mathbb{E}\, X} \right) \right)^{p-1}. \tag{III}$$

PROOF OF LEMMA D.2. To obtain the first bound we use a Bernstein-type inequality on the sub-exponential random variables (see e.g. the proof of Proposition 5.16 by Vershynin [2010]),

$$\mathbb{P}\left\{ \frac{1}{n} \sum_{i=1}^n X_i \ge \mathbb{E}\, X + t \right\} \le \exp\left(-C \min\left\{ \frac{nt^2}{K^2}, \frac{nt}{K} \right\} \right)$$

where

$$K = \|X - \mathbb{E}\, X\|_{\psi_1} \le 2\|X\|_{\psi_1} = 2\nu.$$

To obtain the second bound we use a well-known line of reasoning:

$$\mathbb{P}\left\{\max_i X_i \ge t\right\} = \mathbb{P}\left\{\bigcup_i \{X_i \ge t\}\right\} \le 2ne^{-t/\nu} = e^{\ln(2n) - t/\nu}.$$

Thus we obtain the result.

□

For the sake of transparency we further use

$$E(X, \nu; \mathrm{x}) \stackrel{\text{def}}{=} \mathbb{E}\, X + \nu \max\left\{\sqrt{\frac{\mathrm{x}}{n}}, \frac{\mathrm{x}}{n}\right\}. \tag{D.3}$$

LEMMA D.3 (Assumption ($\hat{T}$)). *Let $q_n \le 1/2$. Assumption 1 ensures that for all $\mathrm{x}, \mathrm{s} > 0$*

$$\delta_{\hat{T}}(\mathrm{x}; \mathrm{s}) = \delta_{\hat{T}}(\mathrm{x}) = e^{-\mathrm{x}}, \quad \alpha_{\hat{T}}(\mathrm{s}) = e^{-\mathrm{s}},$$

$$\varepsilon_{\hat{T}}(\mathrm{x}; \mathrm{s}) = \mathtt{C}\nu_u \frac{d + \sqrt{\mathrm{x}}}{\sqrt{n}} \cdot \sqrt{\frac{E\left(S_i, \nu_S^2; \mathrm{s}\right)}{\lambda_{\min}(Q_*)}},$$

with ν_u and ν_S coming from Assumption 1.

PROOF. First we note that $\overline{T}_u$ is centred in the bootstrap world. We note that by definition of the barycenter $\frac{1}{n}\sum_{i=1}^n \hat{T}_i = 0$. Thus

$$\overline{T}_u = \frac{1}{n}\sum_{i=1}^n u_i \hat{T}_i - \frac{1}{n}\sum_{i=1}^n \hat{T}_i = \frac{1}{n}\sum_{i=1}^n (u_i - 1)\hat{T}_i.$$

The sum is sub-Gaussian due to Assumption 1. We apply Theorem 2.2 by Hsu et al. [2012] and get with P-probability at least $1 - e^{-\mathrm{x}}$

$$\frac{1}{n}\sum_{i=1}^n \left\|\overline{T}_u\right\| \le \frac{d + \sqrt{\mathrm{x}}}{\sqrt{n}} \cdot \sqrt{\frac{\nu_u^2}{n}\sum_{i=1}^n \left\|\hat{T}_i\right\|_F^2}.$$

Now we consider

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \left\|\hat{T}_i\right\|_F^2 &\le \frac{1}{n\lambda_{\min}(Q_n)}\sum_{i=1}^n \operatorname{tr}\left(\hat{T}_i Q_n \hat{T}_i\right) = \frac{1}{n\lambda_{\min}(Q_n)}\sum_{i=1}^n d_{\mathrm{B}}^2(Q_n, S_i) \\
&\le \frac{1}{n\lambda_{\min}(Q_n)}\sum_{i=1}^n d_{\mathrm{B}}^2(0, S_i) = \frac{1}{n\lambda_{\min}(Q_n)}\sum_{i=1}^n \operatorname{tr} S_i \\
&\stackrel{\text{by } L.D.2(\mathrm{I})}{\le} \frac{1}{n\lambda_{\min}(Q_n)} E(S_1, \nu_S, \mathrm{s}).
\end{aligned}$$

The last inequality holds with P-probability at least $1 - e^{-\mathrm{s}}$. Finally, note that

$$\lambda_{\min}(Q_n) \ge (1 - q_n)\lambda_{\min}(Q_*) \ge \frac{1}{2}\lambda_{\min}(Q_*).$$

Combining all the bounds we get the result. □

LEMMA D.4 (Assumption (Z)). *Under Assumption 1 it holds, that*

$$\Delta_Z \le \mathtt{C} d^2 \sqrt{\frac{\|\boldsymbol{\Sigma}^{-1}\|\nu_{\|T\|}^2}{n} \log\left(\frac{\|\boldsymbol{\Sigma}^{-1}\|\nu_{\|T\|}^2}{d^2}\right)},$$

with $\nu_{\|T\|}$ coming from Lemma 5.1, 2).

PROOF. The result follows from Theorem 1.1 by Bentkus [2003]. Using the notations accepted there we set $X_i = \boldsymbol{\Sigma}^{-1/2} T_i$ for all $i = 1, \ldots, n$. Thus

$$\Delta_Z = \frac{\mathtt{C}}{\sqrt{n}} \mathbb{E}\left\|\boldsymbol{\Sigma}^{-1/2} T_i\right\|_F^3.$$

Since $\|T_i\|_F$ is sub-Gaussian r.v. with parameter $\nu_{\|T\|}$ by Lemma 5.1, one obtains $\left\|\left\|\boldsymbol{\Sigma}^{-1/2} T_i\right\|_F^2\right\|_{\psi_1} \le \mathtt{C}\left\|\boldsymbol{\Sigma}^{-1}\right\|\nu_{\|T\|}^2$. Since $\mathbb{E}\left\|\boldsymbol{\Sigma}^{-1/2} T_i\right\|_F^2 = \dim T_i \le d^2$,

$$\mathbb{E}\left\|\boldsymbol{\Sigma}^{-1/2} T_i\right\|_F^3 \overset{\text{by } L.D.2(\text{III})}{\le} \mathtt{C} d^2 \sqrt{\|\boldsymbol{\Sigma}^{-1}\|\nu_{\|T\|}^2 \log\left(\frac{\|\boldsymbol{\Sigma}^{-1}\|\nu_{\|T\|}^2}{d^2}\right)}.$$

The claim follows. □

LEMMA D.5 (Assumption ($\hat{Z}$)). *Let $q_n + f_n \le 1/2$ and*

$$q_n\left(\operatorname{tr}\boldsymbol{\Sigma} + d\right) + 3\|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}\| \le \frac{1}{2}\lambda_{\min}(\boldsymbol{\Sigma}). \tag{D.4}$$

Then for any $\mathrm{s} > 0$ it holds under Assumption 1 that

$$\alpha_{\hat{Z}}(\mathrm{s}) = e^{-\mathrm{s}}, \quad \Delta_{\hat{Z}}(\mathrm{s}) \le \mathtt{C} d^2 m_3(u) \sqrt{\frac{\|\boldsymbol{\Sigma}^{-1}\|\nu_{\|T\|}^2}{n}(\mathrm{s} + \ln n)},$$

with $m_3(u) = \mathbb{E}_u|u-1|^3$, and $\nu_{\|T\|}$ coming from Lemma 5.1, 2).

PROOF. We use Theorem 3.5 by Chen and Fang [2011] follow their notations:

$$X_i = \frac{u_i - 1}{\sqrt{n}} \hat{\boldsymbol{\Sigma}}^{-1/2} \hat{T}_i, \quad \Phi \sim \mathcal{N}(0, \boldsymbol{I}).$$

Now we show concentration of $\mathbb{E}_u \sum_{i=1}^n \|X_i\|_F^3$. Let $m_3(u) \overset{\text{def}}{=} \mathbb{E}_u|u-1|^3$, thus

$$\mathbb{E}_u \sum_{i=1}^n \|X_i\|_F^3 = \frac{m_3(u)}{\sqrt{n}} \frac{1}{n} \sum_{i=1}^n \left\|\hat{\boldsymbol{\Sigma}}^{-1/2} \hat{T}_i\right\|_F^3 \le d^2 m_3(u) \sqrt{\frac{\|\hat{\boldsymbol{\Sigma}}^{-1}\|}{n}} \max_i \|\hat{T}_i\|_F.$$

We recall bound (C.1) and combine it with the condition $q_n \le 1/2$. This yields $\|\hat{T}_i\|_F \le \|T_i\|_F + q_n\|T_i - I\|_F \le \frac{3}{2}\|T_i\|_F + q_n\|I\|_F$. Since $\|T_i\|_F$ are sub-Gaussian with parameter $\nu_{\|T\|}$, we get

$$\max_i \|\hat{T}_i\|_F \le \frac{3}{2} \max_i \|T_i\|_F + q_n\sqrt{d} \overset{\text{by } L.D.2(\text{II})}{\le} \mathtt{C}\nu_{\|T\|}\sqrt{\mathrm{s} + \ln n} + q_n\sqrt{d},$$

where the last inequality holds with P-probability at least $1 - e^{-\mathrm{s}}$. Note that by assumption (D.4),

$$q_n^2 d \le \frac{1}{4\operatorname{tr}\boldsymbol{\Sigma}} \lambda_{\min}^2(\boldsymbol{\Sigma}) \le \operatorname{tr}\boldsymbol{\Sigma} = \mathbb{E}\|T_i\|^2 \le \nu_{\|T\|}^2.$$

Moreover,

$$q_n \le \frac{\lambda_{\min}(\boldsymbol{\Sigma})}{2\operatorname{tr}\boldsymbol{\Sigma}} \le \frac{1}{2\dim\boldsymbol{\Sigma}} \le \frac{1}{d^2}.$$

Then condition $q_n + f_n \le 1/2$ and bound (C.2) yield

$$\left\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\| \le q_n\left(\operatorname{tr}\boldsymbol{\Sigma} + d\right) + \|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_n\|\left(1 + 2d^2 q_n\right) \le \frac{1}{2}\lambda_{\min}(\boldsymbol{\Sigma}).$$

This fact ensures $\left\|\boldsymbol{\Sigma}^{-1/2}\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2} - \boldsymbol{I}\right\| \le \frac{1}{2}$. Thus $\left\|\hat{\boldsymbol{\Sigma}}^{-1}\right\| \le 2\left\|\boldsymbol{\Sigma}^{-1}\right\|$. Now we apply Theorem 3.5 by Chen and Fang [2011] and get the result. □

D.2. Assumptions ($\boldsymbol{F}$), ($\hat{\boldsymbol{F}}$), and ($\boldsymbol{\Sigma}$). The key ingredient of this section is the following lemma.

LEMMA D.6. *Let $X_1, \dots, X_n$ be independent Hermitian matrices, $X_i \in \mathbb{H}(d)$, s.t. for all i, $\mathbb{E}\, X_i = 0$, and $\sigma_i^2 \stackrel{\text{def}}{=} \left\|\mathbb{E}\, X_i^2\right\|$. Denote for some fixed $\alpha \ge 1$*

$$U_i^{(\alpha)} \ge \max\left\{\left\|\|X_i\|\right\|_{\psi_\alpha}, \sigma_i\right\}, \quad U \stackrel{\text{def}}{=} \max_i \left\{U_i^{(\alpha)} \log^{\frac{1}{\alpha}}\left(\frac{U_i^{(\alpha)}}{\sigma_i}\right)\right\},$$

(recall that $\log(x) \stackrel{\text{def}}{=} \max\{1, \ln(x)\}$).

In particular, for $\alpha = 1, 2$ one can take $U_i^{(\alpha)} = \left\|\|X_i\|\right\|_{\psi_\alpha}$. Then for all $\mathrm{x} > 0$ with probability at least $1 - e^{-\mathrm{x}}$

$$\left\|\frac{1}{n}\sum_{i=1}^n X_i\right\| \le \mathsf{C}\sigma\sqrt{\frac{\mathrm{x} + \ln d}{n}} \max\left\{1, \frac{U}{\sigma}\sqrt{\frac{\mathrm{x} + \ln d}{n}}\right\},$$

where $\sigma^2 \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n \sigma_i^2$.

PROOF. We set $Y_n = \sum_{i=1}^n X_i$ and note that $\|Y_n\| < t$ is if and only if $-tI \preccurlyeq Y_n \preccurlyeq tI$. This implies

$$\mathbb{P}\left\{\|Y_n\| \ge t\right\} \le \mathbb{P}\left\{Y_n \not\preccurlyeq tI\right\} + \mathbb{P}\left\{Y_n \not\succcurlyeq -tI\right\}. \tag{D.5}$$

Taking the matrix exponent we obtain

$$\mathbb{P}\left\{Y_n \not\preccurlyeq tI\right\} = \mathbb{P}\left\{e^{\lambda Y_n} \not\preccurlyeq e^{\lambda t I}\right\} \le \mathbb{P}\left\{\operatorname{tr}\left(e^{\lambda Y_n}\right) \ge e^{\lambda t}\right\} \le e^{-\lambda t}\,\mathbb{E}\operatorname{tr}\left(e^{\lambda Y_n}\right). \tag{D.6}$$

Independence of $X_1, \dots, X_n$ and the Golden–Thompson inequality ensure

$$\mathbb{E}\operatorname{tr}\left(e^{\lambda Y_n}\right) \le \mathbb{E}\operatorname{tr}\left(e^{\lambda Y_{n-1}}\right)\left\|\mathbb{E}\, e^{\lambda X_n}\right\|.$$

By induction we obtain that

$$\mathbb{E}\operatorname{tr}\left(e^{\lambda Y_n}\right) \le \mathbb{E}\operatorname{tr}\left(e^{\lambda X_1}\right)\prod_{i=2}^n\left\|\mathbb{E}\, e^{\lambda X_i}\right\| \le d\prod_{i=1}^n\left\|\mathbb{E}\, e^{\lambda X_i}\right\|, \tag{D.7}$$

where the last inequality follows from $\mathbb{E}\operatorname{tr}\left(e^{\lambda X_1}\right) = \operatorname{tr}\left(\mathbb{E}\, e^{\lambda X_1}\right) \le d\left\|\mathbb{E}\, e^{\lambda X_1}\right\|$. So as to bound $\left\|\mathbb{E}\, e^{\lambda X_i}\right\|$ we use the following fact obtained in the proof of Proposition 2 by Koltchinskii et al. [2011] (see p. 2949). There exist constants C, C_2 s.t. for all λ satisfying

$$\lambda U_i^{(\alpha)}\left(\ln\left(\frac{U_i^{(\alpha)}}{\sigma_i}\right)\right)^{1/\alpha} \le C_2,$$

it holds that

$$\left\|\mathbb{E}\, e^{\lambda X_i}\right\| \le \exp\left\{C\lambda^2\sigma_i^2\right\}.$$

Combining this bound with (D.5), (D.6), and (D.7), we obtain

$$\mathbb{P}\left\{\|Y_n\| \ge t\right\} \le 2d\exp\left\{-\lambda t + Cn\lambda^2\sigma^2\right\}, \quad \text{whenever} \quad \lambda U \le C_2.$$

Minimization over all admissible λ yields the result.

Furthermore, for $\alpha = 2$ by Jensens's inequality we have

$$\left\|\mathbb{E}\, X^2\right\| \le \mathbb{E}\|X\|^2 \le \left\|\|X\|\right\|_{\psi_2}^2 \ln 2 \le \left\|\|X\|\right\|_{\psi_2}^2.$$

Respectively, for $\alpha = 1$ we have

$$\left\|\mathbb{E}\, X^2\right\| \le \mathbb{E}\|X\|^2 \le \left\|\|X\|\right\|_{\psi_1}^2 \left[\mathbb{E}\exp\left(\frac{\|X\|}{\left\|\|X\|\right\|_{\psi_1}}\right) - 1\right] \le \left\|\|X\|\right\|_{\psi_1}^2$$

since $x^2 \le e^x - 1$ for all $x \ge 0$. □

Since the above bound appears often, we introduce the following terms. For $\sigma, U > 0$, s.t. $\frac{U}{\sigma} \ge 1$, we define $\vartheta \overset{\text{def}}{=} \frac{U}{\sigma}$ and

$$M^{(\alpha)}(\vartheta; \mathrm{x}) \overset{\text{def}}{=} \max\left\{1,\ \vartheta \log^{\frac{1}{\alpha}}(\vartheta)\sqrt{\frac{\mathrm{x} + \ln d}{n}}\right\}, \tag{D.8}$$

$$M(\vartheta; \mathrm{x}) \overset{\text{def}}{=} \max\left\{1,\ \vartheta\sqrt{\frac{\mathrm{x} + \ln d}{n}}\right\}. \tag{D.9}$$

LEMMA D.7 (Assumption ($\boldsymbol{F}$)). *Assumption 1 ensures that for all* $\mathrm{x} > 0$

$$\delta_F(\mathrm{x}) = e^{-\mathrm{x}}, \quad \varepsilon_F(\mathrm{x}) = \mathtt{C}\left\|\boldsymbol{F}^{-1}\right\|\sigma_F\sqrt{\frac{\mathrm{x} + \ln d}{n}} M^{(2)}(\vartheta_F; \mathrm{x}),$$

where $\vartheta_F \overset{\text{def}}{=} \frac{\nu_{\|\boldsymbol{dT}\|}}{\sigma_F}$, $\sigma_F^2 \overset{\text{def}}{=} \left\|\mathbb{E}\left(\boldsymbol{dT}_i - \mathbb{E}\,\boldsymbol{dT}_i\right)^2\right\|$, *and* $\nu_{\|\boldsymbol{dT}\|}$ *comes from Lemma 5.1, 4).*

PROOF. The concentration result follows directly from Lemma D.6. We set $X_i = \boldsymbol{dT}_i - \mathbb{E}\,\boldsymbol{dT}_i$, and choose $\alpha = 2$. Since all X_i are i.i.d. we get

$$\sigma_i^2 = \sigma_F^2 \overset{\text{def}}{=} \left\|\mathbb{E}\left(\boldsymbol{dT}_i - \mathbb{E}\,\boldsymbol{dT}_i\right)^2\right\|.$$

Moreover, for all $i = 1, \ldots, n$

$$U_i^{(2)} = \left\|\|\boldsymbol{dT}_i - \mathbb{E}\,\boldsymbol{dT}_i\|\right\|_{\psi_2} \overset{\text{by } L.5.1}{\le} 2\nu_{\|\boldsymbol{dT}\|}.$$

Thus with P-probability at least $1 - e^{-\mathrm{x}}$ we get

$$\|\boldsymbol{F}_n - \boldsymbol{F}\| \le \mathtt{C}\sigma_F\sqrt{\frac{\mathrm{x} + \ln d}{n}} M^{(2)}(\vartheta_F; \mathrm{x}), \quad \vartheta_F \overset{\text{def}}{=} \frac{\nu_{\|\boldsymbol{dT}\|}}{\sigma_F}.$$

The claim follows. □

LEMMA D.8 (Assumption ($\boldsymbol{\Sigma}$)). *Assumption 1 ensures that for all* $\mathrm{x} > 0$

$$\delta_\Sigma(\mathrm{x}) = e^{-\mathrm{x}}, \quad \varepsilon_\Sigma(\mathrm{x}) = \mathtt{C}\sigma_\Sigma\sqrt{\frac{\mathrm{x} + \ln d}{n}} M^{(1)}(\vartheta_\Sigma; \mathrm{x}),$$

where $\vartheta_\Sigma \overset{\text{def}}{=} \frac{\nu_{\|T\|}}{\sigma_\Sigma}$, $\sigma_\Sigma^2 \overset{\text{def}}{=} \left\|\mathbb{E}\left(T_i \otimes T_i - \boldsymbol{\Sigma}\right)^2\right\|$, *and* $\nu_{\|T\|}$ *comes from Lemma 5.1, 2).*

PROOF. The result follows directly from Lemma D.6. We set $X_i = T_i \otimes T_i - \boldsymbol{\Sigma}$ and select $\alpha = 1$. Since all X_i are i.i.d. we get

$$\sigma_i^2 = \sigma_\Sigma^2 \stackrel{\text{def}}{=} \left\| \mathbb{E}\left(T_i \otimes T_i - \boldsymbol{\Sigma}\right)^2 \right\|.$$

Moreover, for all $i = 1, \ldots, n$

$$U_i^{(1)} = \big\| \|T_i \otimes T_i - \boldsymbol{\Sigma}\| \big\|_{\psi_1} \le 2 \big\| \|T_i \otimes T_i\| \big\|_{\psi_1} = 2 \big\| \|T_i\|_F^2 \big\|_{\psi_1} \stackrel{\text{by } L.5.1}{=} 2\nu_{\|T\|}^2.$$

Then with P-probability at least $1 - e^{-\mathrm{x}}$ we get

$$\|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}\| \le \mathtt{C}\sigma_\Sigma \sqrt{\frac{\mathrm{x} + \ln d}{n}} M^{(1)}(\vartheta_\Sigma; \mathrm{x}), \quad \vartheta_\Sigma \stackrel{\text{def}}{=} \frac{\nu_{\|T\|}}{\sigma_\Sigma}.$$

The claim follows. □

LEMMA D.9 (Assumption ($\hat{F}$)). *Let $q_n + f_n \le 1/2$. Under Assumption 1 for all* $\mathrm{x}, \mathrm{s} > 0$

$$\delta_{\hat{F}}(\mathrm{x}; \mathrm{s}) = \delta_{\hat{F}}(\mathrm{x}) = e^{-\mathrm{x}}, \quad \alpha_{\hat{F}}(\mathrm{s}) = 3e^{-\mathrm{s}},$$

$$\varepsilon_{\hat{F}}(\mathrm{x}; \mathrm{s}) = \mathtt{C}\big\|\boldsymbol{F}^{-1}\big\| \sigma_{\hat{F}}(\mathrm{s}) \sqrt{\frac{\ln d + \mathrm{x}}{n}} M(\vartheta_{\hat{F}}(\mathrm{s}); \mathrm{x}),$$

where $\vartheta_{\hat{F}}(\mathrm{s}) \stackrel{\text{def}}{=} \frac{U_{\hat{F}}(\mathrm{s})}{\sigma_{\hat{F}}(\mathrm{s})}$, $\sigma_{\hat{F}}^2(\mathrm{s}) \stackrel{\text{def}}{=} E\left(\|\boldsymbol{dT}_i\|^2, \nu_{\|\boldsymbol{dT}\|}^2; \mathrm{s}\right)$, *with* $\nu_{\|\boldsymbol{dT}\|}$ *coming from Lemma 5.1, 4) and* $U_{\hat{F}}(\mathrm{s}) \stackrel{\text{def}}{=} \mathtt{C}\nu_u \log^{\frac{1}{2}}(\nu_u) \nu_{\|\boldsymbol{dT}\|} \sqrt{\mathrm{s} + \ln n}$.

PROOF. First we note that

$$\left\| \hat{\boldsymbol{F}}^{-1/2} \hat{\boldsymbol{F}}_u \hat{\boldsymbol{F}}^{-1/2} - \boldsymbol{I} \right\| \le \left\| \hat{\boldsymbol{F}}^{-1} \right\| \cdot \left\| \hat{\boldsymbol{F}}_u - \hat{\boldsymbol{F}} \right\|, \ \text{ and } \ \hat{\boldsymbol{F}}_u - \hat{\boldsymbol{F}} = \frac{1}{n} \sum_{i=1}^n (u_i - 1) \boldsymbol{d}\hat{\boldsymbol{T}}_i.$$

Now we apply Lemma D.6 with $\alpha = 2$ and $X_i = (u_i - 1)\boldsymbol{d}\hat{\boldsymbol{T}}_i$. This yields

$$\sigma_i^2 = \left\| \mathbb{E}_u \left[(u_i - 1)\boldsymbol{d}\hat{\boldsymbol{T}}_i\right]^2 \right\| = \left\| \left[\boldsymbol{d}\hat{\boldsymbol{T}}_i\right]^2 \right\| = \left\| \boldsymbol{d}\hat{\boldsymbol{T}}_i \right\|^2.$$

The last equality follows from the fact that $\boldsymbol{d}\hat{\boldsymbol{T}}_i$ is a Hermitian operator. Thus

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n \left\| \boldsymbol{d}\hat{\boldsymbol{T}}_i \right\|^2 \le 8 \frac{1}{n} \sum_{i=1}^n \|\boldsymbol{dT}_i\|^2,$$

since $q_n \le \frac{1}{2}$ and due to monotonicity and homogeneity of $\boldsymbol{dT}_Q^S$. Note that

$$\big\| \|X_i\| \big\|_{\psi_2} = \left\| |u_i - 1| \cdot \left\| \boldsymbol{d}\hat{\boldsymbol{T}}_i \right\| \right\|_{\psi_2} = \left\| \boldsymbol{d}\hat{\boldsymbol{T}}_i \right\| \cdot \|u_i - 1\|_{\psi_2} \stackrel{\text{by } A.1}{=} \nu_u \left\| \boldsymbol{d}\hat{\boldsymbol{T}}_i \right\|,$$

thus we set $U_i = \nu_u \left\| \boldsymbol{d}\hat{\boldsymbol{T}}_i \right\|$. This yields

$$\begin{aligned} U = \max_i \left\{ U_i \log^{\frac{1}{2}} \left(\frac{U_i}{\sigma_i} \right) \right\} &= \nu_u \log^{\frac{1}{2}}(\nu_u) \max_i \left\| \boldsymbol{d}\hat{\boldsymbol{T}}_i \right\| \\ &\le 2^{3/2} \nu_u \log^{\frac{1}{2}}(\nu_u) \max_i \|\boldsymbol{dT}_i\|. \end{aligned}$$

Then we get with P_u-probability at least $1 - e^{-\mathrm{x}}$,

$$\left\| \hat{\boldsymbol{F}} - \hat{\boldsymbol{F}}_u \right\| \le \mathtt{C}\sigma \sqrt{\frac{\ln d + \mathrm{x}}{n}} M(\vartheta; \mathrm{x}), \quad \vartheta \stackrel{\text{def}}{=} \frac{U}{\sigma}.$$

Now we have to bound the terms σ^2 and U. First we consider σ^2 and recall that due to Lemma 5.1 $\|\boldsymbol{dT}_i\|$ are sub-Gaussian r.v. We get with P-probability at least $1 - e^{-\mathrm{s}}$

$$\sigma^2 \overset{\text{by } L.D.2(I)}{\le} 8\sigma^2_{\hat{F}}(\mathrm{s}), \quad \sigma^2_{\hat{F}}(\mathrm{s}) \overset{\text{def}}{=} E\left(\|\boldsymbol{dT}_i\|^2, \nu^2_{\|\boldsymbol{dT}\|}; \mathrm{s}\right).$$

Now we consider term U. With P-probability at least $1 - e^{-\mathrm{s}}$

$$U \overset{\text{by } L.D.2(II)}{\le} 2^{3/2} U_{\hat{F}}(\mathrm{s}), \quad U_{\hat{F}}(\mathrm{s}) \overset{\text{def}}{=} \mathtt{C}\nu_u \log^{\frac{1}{2}}(\nu_u)\nu_{\|\boldsymbol{dT}\|}\sqrt{\mathrm{s} + \ln n}.$$

Finally, Lemma A.3 together with the assumption $q_n + f_n \le 1/2$ ensures $\left\|\hat{\boldsymbol{F}}^{-1}\right\| \le \frac{9}{4}\left\|\boldsymbol{F}^{-1}\right\|$. Collecting the above bounds we get the result. □

D.3. GAR and bootstrap validity. Now we collect the bounds obtained in the previous section to get GAR results. First, we introduce the scaling multipliers

$$C_T \overset{\text{def}}{=} \frac{\nu_T^2 \kappa(\boldsymbol{F})\|Q_*\|^2}{\lambda^2_{\min}(\boldsymbol{\xi})}, \quad C_F \overset{\text{def}}{=} \sigma_F^2 \kappa(\boldsymbol{F})\left\|\boldsymbol{F}^{-1}\right\|^2, \quad C_Z \overset{\text{def}}{=} \left\|\boldsymbol{\Sigma}^{-1}\right\|\nu^2_{\|T\|}.$$

Let $N(n,d)$ be the "threshold" sample-size

$$N(n,d) \overset{\text{def}}{=} \max\left\{\vartheta_F^2 \log(\vartheta_F)\log\left(\frac{nd}{C_F}\right), \, C_F \ln d, \, C_T d^2\right\},$$

These bound on the sample size are crucial for asymptotic results.

LEMMA D.10 (Lemma 3.2). *Let* $\mathrm{x} \in \mathbb{R}^2_+$. *Under Assumption 1:*

Asymptotic for Frobenius case: bound (3.9). *Let the sample size* $n \gtrsim N(n,d)$. *Then the following bound holds*

$$\Omega_{\mathrm{F}}(n) \lesssim d^2\sqrt{\frac{C_Z}{n}\log\frac{C_Z}{d}} + \gamma(\boldsymbol{\Xi})\sqrt{\frac{C_T}{n}\left(d^2 + \log\frac{n}{C_T}\right)} + \gamma(\boldsymbol{\Xi})\sqrt{\frac{C_F}{n}\log\frac{nd}{C_F}} \lesssim \sqrt{\frac{\ln n}{n}}. \tag{D.10}$$

Asymptotic for Bures–Wasserstein case: bound (3.10). *Let the sample size* $n \gtrsim \kappa(Q_*)N(n,d)$. *Then the following bound holds*

$$\Omega_{\mathrm{B}}(n) \lesssim d^2\sqrt{\frac{C_Z}{n}\log\frac{C_Z}{d}} + \gamma_\kappa(\boldsymbol{\Xi})\sqrt{\frac{C_T}{n}\left(d^2 + \log\frac{n}{C_T}\right)} + \gamma_\kappa(\boldsymbol{\Xi})\sqrt{\frac{C_F}{n}\log\frac{nd}{C_F}} \lesssim \sqrt{\frac{\ln n}{n}}. \tag{D.11}$$

PROOF. First we obtain bounds on $\delta_\eta(\mathrm{x})$ from (3.7), and $\varepsilon_\eta(\mathrm{x})$ from (3.8).

- *Term* (3.7). To get it we use (3.7), and combine it with Lemma D.1, and Lemma D.7. This yields $\delta_\eta(\mathrm{x}) = e^{-\mathrm{x}_1} + e^{-\mathrm{x}_2}$.
- *Term* (3.8). To get it we note that the term $\varepsilon_Q(\mathrm{x}_1)$ coming from (B.1) is written as

$$\varepsilon_Q(\mathrm{x}_1) \overset{\text{by L.}D.1}{=} \frac{4\|Q_*\|\nu_T}{\lambda_{\min}(\boldsymbol{\xi})}\frac{d + \sqrt{\mathrm{x}_1}}{\sqrt{n}} \le \frac{\mathtt{C}}{\sqrt{\kappa(\boldsymbol{F})}}\sqrt{\frac{C_T}{n}}(d^2 + \mathrm{x}_1).$$

We combine this fact with the definition (3.8), Lemma D.1, and Lemma D.7:

$$\varepsilon_\eta(\mathrm{x}) = \mathtt{C}\sqrt{\frac{C_T}{n}}(d^2 + \mathrm{x}_1) + \mathtt{C}\sqrt{\frac{C_F}{n}}(\ln d + \mathrm{x}_2)M^{(2)}(\vartheta_F; \mathrm{x}_2)$$

Asymptotic bound: Frobenius case. We set

$$\mathrm{x}_1^* = \frac{1}{2}\ln\left(\frac{n}{C_T}\right), \quad \mathrm{x}_2^* = \frac{1}{2}\ln\left(\frac{n}{C_F}\right), \mathrm{x}^* = (\mathrm{x}_1^*, \mathrm{x}_2^*).$$

We note that $\mathrm{x}_1^*, \mathrm{x}_2^* > 0$ due to $n \gtrsim N(d,s)$.

- ∘ Term $\delta_\eta(\mathrm{x}^*)$ turns into

$$\delta_\eta(\mathrm{x}^*) = e^{-\mathrm{x}_1^*} + e^{-\mathrm{x}_2^*} = \sqrt{\frac{C_F}{n}} + \sqrt{\frac{C_T}{n}}.$$

- ∘ Now we show that $\varepsilon_\eta(\mathrm{x}^*) \lesssim 1$. Note that condition $n \gtrsim N(n,s)$ ensures three following facts:
  - · $n \gtrsim C_F$, this ensures $\frac{n}{C_F} \gtrsim \log\left(\frac{n}{C_F}\right)$,
  - · $n \gtrsim C_T$, this ensures $\frac{n}{C_T} \gtrsim \log\left(\frac{n}{C_T}\right)$,
  - · $M^{(2)}(\vartheta_F; \mathrm{x}_2^*) = 1$.

  For simplicity we denote for any $t > 0$

$$f_1(t) \stackrel{\text{def}}{=} \sqrt{\frac{t}{n}\left(\ln d + \log\frac{n}{t}\right)}, \quad f_2(t) \stackrel{\text{def}}{=} \sqrt{\frac{t}{n}\left(d^2 + \log\frac{n}{t}\right)}. \tag{D.12}$$

  Thus we get

$$\varepsilon_\eta(\mathrm{x}^*) \lesssim f_1(C_F) + f_2(C_T) \lesssim 1.$$

  Since $\gamma(\boldsymbol{\Xi}) \geq 1$, we derive (D.10):

$$\begin{aligned}\varOmega_{\mathrm{F}}(n) &\lesssim \varDelta_Z + \gamma(\boldsymbol{\Xi})\left(\sqrt{\frac{C_F}{n}} + \sqrt{\frac{C_T}{n}} + f_1(C_F) + f_2(C_T)\right)\\ &\lesssim \varDelta_Z + \gamma(\boldsymbol{\Xi})\big(f_1(C_F) + f_2(C_T)\big).\end{aligned}$$

Asymptotic bound: Bures–Wasserstein case. We set

$$\mathrm{x}_1^* = \frac{1}{2}\ln\left(\frac{n}{\kappa(Q_*)C_T}\right), \quad \mathrm{x}_2^* = \frac{1}{2}\ln\left(\frac{n}{\kappa(Q_*)C_F}\right), \quad \mathrm{x}^* = (\mathrm{x}_1^*, \mathrm{x}_2^*).$$

We note that $\mathrm{x}_1^*, \mathrm{x}_2^* > 0$ due to $n \gtrsim \kappa(Q_*)N(n,d)$.

- ∘ Term $\delta_\eta(\mathrm{x}^*)$ is written as

$$\delta_\eta(\mathrm{x}^*) = e^{-\mathrm{x}_1^*} + e^{-\mathrm{x}_2^*} = \sqrt{\frac{\kappa(Q_*)C_T}{n}} + \sqrt{\frac{\kappa(Q_*)C_F}{n}}.$$

- ∘ Now we have to show that $\sqrt{\kappa(Q_*)}\varepsilon_\eta(\mathrm{x}^*) \lesssim 1$. Condition $n \gtrsim \kappa(Q_*)N(n,d)$ yields:
  - · $n \gtrsim \kappa(Q_*)C_F$, this ensures $\frac{n}{\kappa(Q_*)C_F} \gtrsim \log\left(\frac{n}{\kappa(Q_*)C_F}\right)$,
  - · $n \gtrsim \kappa(Q_*)C_T$, this ensures $\frac{n}{\kappa(Q_*)C_T} \gtrsim \log\left(\frac{n}{\kappa(Q_*)C_T}\right)$,
  - · $M^{(2)}(\vartheta_F; \mathrm{x}_2^*) = 1$.

  Therefore,

$$\sqrt{\kappa(Q_*)}\varepsilon_\eta(\mathrm{x}^*) \lesssim f_1(\kappa(Q_*)C_F) + f_2(\kappa(Q_*)C_T) \lesssim 1.$$

Now by Lemma 3.3 we obtain (D.11):

$$\varOmega_{\mathrm{B}}(n) \lesssim \varDelta_Z + \gamma(\boldsymbol{\Xi})\left(\sqrt{\tfrac{\kappa(Q_*)C_F}{n}} + \sqrt{\tfrac{\kappa(Q_*)C_T}{n}} + f_1(\kappa(Q_*)C_F) + f_2(\kappa(Q_*)C_T)\right)$$
$$\lesssim \varDelta_Z + \gamma(\boldsymbol{\Xi})\big(f_1(\kappa(Q_*)C_F) + f_2(\kappa(Q_*)C_T)\big)$$
$$\lesssim \varDelta_Z + \gamma_\kappa(\boldsymbol{\Xi})\big(f_1(C_F) + f_2(C_T)\big).\square$$

To validate GAR in the bootstrap world we proceed in the same way as before and define some auxiliary multipliers. Let

$$C_{\hat{T}}(n;\mathrm{s}) \overset{\text{def}}{=} \frac{\nu_u^2 E\left(S_1, \nu_S^2; \mathrm{s}\right)}{\nu_T^2 \lambda_{\min}(Q_*)} C_T, \quad C_{\hat{F}}(n;\mathrm{s}) \overset{\text{def}}{=} \frac{E\left(\|\boldsymbol{dT}_i\|^2, \nu_{\|\boldsymbol{dT}\|}^2; \mathrm{s}\right)}{\sigma_F^2} C_F,$$
$$C_{\hat{Z}}(n;\mathrm{s}) \overset{\text{def}}{=} m_3^2(u)(\mathrm{s} + \ln n) C_Z.$$

For the sake of transparency we further refer to $C_{\hat{T}}(n;\mathrm{s})$, $C_{\hat{F}}(n;\mathrm{s})$, and $C_{\hat{Z}}(n;\mathrm{s})$ as $C_{\hat{T}}$, $C_{\hat{F}}$, and $C_{\hat{Z}}$. Let us also define the effective sample size

$$\hat{N}(n,d;\mathrm{s}) \overset{\text{def}}{=} \max\left\{\vartheta_{\hat{F}}^2(\mathrm{s})\log\left(\tfrac{nd}{C_{\hat{F}}}\right),\ C_{\hat{T}}d^2,\ C_{\hat{F}}\ln d, \mathrm{s}\right\}.$$

LEMMA D.11 (Lemma 3.3). *Let $q_n + f_n \le 1/2$ and let condition (D.4) be fulfilled. Choose $\mathrm{x} \in \mathbb{R}_+^2$). Under Assumption 1 it holds with P-probability at least $1 - 7e^{-\mathrm{s}}$:*

Asymptotic for Frobenius case. *Let the sample size $n \gtrsim \hat{N}(n,d;\mathrm{s})$. Then with P-probability at least $1 - 7e^{-\mathrm{s}}$*

$$\hat{\varOmega}_{\mathrm{F}}(n;\mathrm{s}) \lesssim d^2\sqrt{\frac{C_{\hat{Z}}}{n}} + \gamma(\boldsymbol{\Xi})\sqrt{\frac{C_{\hat{F}}}{n}\log\frac{nd}{C_{\hat{F}}}} + \gamma(\boldsymbol{\Xi})\sqrt{\frac{C_{\hat{T}}}{n}\left(d^2 + \log\frac{n}{C_{\hat{T}}}\right)} \tag{D.13}$$
$$\lesssim \sqrt{\frac{\mathrm{s}+\ln n}{n}}.$$

Asymptotic for Bures–Wasserstein case. *Let $n \gtrsim \kappa(Q_*)\hat{N}(n,d;\mathrm{s})$. Then with P-probability at least $1 - 7e^{-\mathrm{s}}$ it holds*

$$\hat{\varOmega}_{\mathrm{B}}(n;\mathrm{s}) \lesssim d^2\sqrt{\frac{C_{\hat{Z}}}{n}} + \gamma_\kappa(\boldsymbol{\Xi})\sqrt{\frac{C_{\hat{F}}}{n}\log\frac{nd}{C_{\hat{F}}}} + \gamma_\kappa(\boldsymbol{\Xi})\sqrt{\frac{C_{\hat{T}}}{n}\left(d^2 + \log\frac{n}{C_{\hat{T}}}\right)} \tag{D.14}$$
$$\lesssim \sqrt{\frac{\mathrm{s}+\ln n}{n}}.$$

PROOF. As before, we split proof into several parts.

- *Term* (3.15). It comes from Lemma D.3 and Lemma D.9: $\delta_{\hat{\eta}}(\mathrm{x};\mathrm{s}) = \delta_{\hat{\eta}}(\mathrm{x}) = e^{-\mathrm{x}_1} + e^{-\mathrm{x}_2}$.
- *Term* (3.16) First we note that the term $\varepsilon_{\hat{Q}}(\mathrm{x};\mathrm{s})$ introduced by (B.2) is written as

$$\varepsilon_{\hat{Q}}(\mathrm{x};\mathrm{s}) \overset{\text{by L.}D.3}{=} \mathtt{C}\frac{\|Q_*\|\nu_u E^{1/2}\left(S_1, \nu_S^2; \mathrm{s}\right)}{\lambda_{\min}(\boldsymbol{\xi})\lambda_{\min}^{1/2}(Q_*)} \cdot \frac{d + \sqrt{\mathrm{x}_1}}{\sqrt{n}} \le \frac{\mathtt{C}}{\kappa(\boldsymbol{F})}\sqrt{\frac{C_{\hat{T}}}{n}}(d^2 + \mathrm{x}_1).$$

To get the result we combine this fact with the results of Lemma D.3 and D.9. The equality holds with $\alpha_{\hat{T}}(\mathrm{s}) + \alpha_{\hat{F}}(\mathrm{s}) = 4e^{-\mathrm{s}}$:

$$\varepsilon_{\hat{\eta}}(\mathrm{x};\mathrm{s}) = \mathtt{C}\sqrt{\frac{C_{\hat{T}}}{n}}(d^2 + \mathrm{x}_1) + \mathtt{C}\sqrt{\frac{C_{\hat{F}}}{n}}(\ln d + \mathrm{x}_2) M(\vartheta_{\hat{F}}(\mathrm{s}); \mathrm{x}_2),$$

Asymptotic for Frobenius case. We choose

$$\mathrm{x}_1^* = \frac{1}{2}\ln\left(\frac{n}{C_{\hat{T}}}\right), \quad \mathrm{x}_2^* = \frac{1}{2}\ln\left(\frac{n}{C_{\hat{F}}}\right), \quad \mathrm{x}^* = (\mathrm{x}_1^*, \mathrm{x}_2^*).$$

Condition $n \gtrsim \hat{N}(n,d;\mathrm{s})$ ensures $\mathrm{x}^* \in \mathbb{R}_+^2$.

- Term $\delta_{\hat{\eta}}(\mathrm{x}^*)$ turns into

$$\delta_{\hat{\eta}}(\mathrm{x}^*) = e^{-\mathrm{x}_1^*} + e^{-\mathrm{x}_2^*} = \sqrt{\tfrac{C_{\hat{F}}}{n}} + \sqrt{\tfrac{C_{\hat{T}}}{n}}.$$

- Now we have to show that $\varepsilon_{\hat{\eta}} \lesssim 1$. Condition $n \gtrsim \hat{N}(n,d,\mathrm{s})$ ensures
  - $n \gtrsim C_{\hat{F}}$, this yields $\frac{n}{C_{\hat{F}}} \gtrsim \log\left(\frac{n}{C_{\hat{F}}}\right)$,
  - $n \gtrsim C_{\hat{T}}$, this yields $\frac{n}{C_{\hat{T}}} \gtrsim \log\left(\frac{n}{C_{\hat{T}}}\right)$,
  - $E^{1/2}\left(S_1, \nu_S^2; \mathrm{s}_1\right) \lesssim (\mathbb{E}\,\mathrm{tr}\, S_1 + \nu_S)^{1/2} \lesssim 1$,
  - $M(\vartheta_{\hat{F}}(\mathrm{s}); \mathrm{x}_2^*) = 1$.

  Now use the notations accepted in (D.12) and get

$$\varepsilon_{\hat{\eta}}(\mathrm{x}^*; \mathrm{s}) \lesssim f_1\left(C_{\hat{F}}\right) + f_2\left(C_{\hat{T}}\right) \lesssim 1.$$

  Thus (D.13) follows from

$$\begin{aligned}\hat{\Omega}_{\mathrm{F}}(n;\mathrm{s}) &\lesssim \Delta_{\hat{Z}}(\mathrm{s}) + \sqrt{\frac{C_{\hat{F}}}{n}} + \sqrt{\frac{C_{\hat{T}}}{n}} + \gamma(\boldsymbol{\Xi})\left(f_1\left(C_{\hat{F}}\right) + f_2\left(C_{\hat{T}}\right)\right) \\ &\lesssim \Delta_{\hat{Z}}(\mathrm{s}) + \gamma(\boldsymbol{\Xi})\left(f_1\left(C_{\hat{F}}\right) + f_2\left(C_{\hat{T}}\right)\right).\end{aligned}$$

Asymptotic result for Bures–Wasserstein case. We set

$$\mathrm{x}_1^* = \frac{1}{2}\ln\left(\frac{n}{\kappa(Q_*)C_{\hat{T}}}\right), \quad \mathrm{x}_2^* = \frac{1}{2}\ln\left(\frac{n}{\kappa(Q_*)C_{\hat{F}}}\right), \quad \mathrm{x}^* = (\mathrm{x}_1^*, \mathrm{x}_2^*).$$

Condition $n \gtrsim \kappa(Q_*)\hat{N}(n,d;\mathrm{s})$ ensures $\mathrm{x}^* \in \mathbb{R}_+^2$.

- Term $\delta_{\hat{\eta}}(\mathrm{x}^*)$ turns into

$$\delta_{\hat{\eta}}(\mathrm{x}^*) = e^{-\mathrm{x}_1^*} + e^{-\mathrm{x}_2^*} = \sqrt{\tfrac{\kappa(Q_*)C_{\hat{F}}}{n}} + \sqrt{\tfrac{\kappa(Q_*)C_{\hat{T}}}{n}}.$$

- Now we have to show that $\sqrt{\kappa(Q_*)}\varepsilon_{\hat{\eta}}(\mathrm{x}^*;\mathrm{s}) \lesssim 1$. Condition $n \gtrsim \kappa(Q_*)\hat{N}(n,d;\mathrm{s})$ ensures
  - $n \gtrsim \kappa(Q_*)C_{\hat{F}}$, this yields $\frac{n}{\kappa(Q_*)C_{\hat{F}}} \gtrsim \log\left(\frac{n}{\kappa(Q_*)C_{\hat{F}}}\right)$,
  - $n \gtrsim \kappa(Q_*)C_{\hat{T}}$, this yields $\frac{n}{\kappa(Q_*)C_{\hat{T}}} \gtrsim \log\left(\frac{n}{\kappa(Q_*)C_{\hat{T}}}\right)$,
  - $E^{1/2}\left(S_1, \nu_S^2; \mathrm{s}\right) \lesssim (\mathbb{E}\,\mathrm{tr}\, S + \nu_S)^{1/2} \lesssim 1$,
  - $M(\vartheta_{\hat{F}}(\mathrm{s}); \mathrm{x}_2^*) = 1$.

  Taking into account the above facts and using notations accepted in (D.12), we get

$$\sqrt{\kappa(Q_*)}\varepsilon_{\hat{\eta}}(\mathrm{x}^*;\mathrm{s}) \lesssim f_1(\kappa(Q_*)C_{\hat{F}}) + f_2(\kappa(Q_*)C_{\hat{T}}) \lesssim 1.$$

  Combining the above facts we get (D.14)

$$\begin{aligned}\hat{\Omega}_{\mathrm{B}}(n;\mathrm{s}) &\lesssim \Delta_{\hat{Z}}(\mathrm{s}) + \sqrt{\tfrac{\kappa(Q_*)C_{\hat{F}}}{n}} + \sqrt{\tfrac{\kappa(Q_*)C_{\hat{T}}}{n}} + \gamma(\boldsymbol{\Xi})\left(f_1(\kappa(Q_*)C_{\hat{F}}) + f_2(\kappa(Q_*)C_{\hat{T}})\right) \\ &\lesssim \Delta_{\hat{Z}}(\mathrm{s}) + \gamma(\boldsymbol{\Xi})\sqrt{\kappa(Q_*)}\left(f_1(C_{\hat{F}}) + f_2(C_{\hat{T}})\right). \quad \square\end{aligned}$$

Now we are ready to write down the main results. In what follows we use notations accepted in Lemma D.10 and D.11. Further we assume $\mathrm{y} = \mathrm{s} \in \mathbb{R}^3_+$, s.t. $\mathrm{s}_1 = \mathrm{s}_2 = \mathrm{s}_3 \stackrel{\mathrm{def}}{=} \mathrm{s}$.

THEOREM D.12. *Let* $\mathrm{s} \in \mathbb{R}_+$ *be s.t.*

$$\varepsilon_\eta(\mathrm{s}) \leq 1, \quad \varepsilon_\Xi(\mathrm{s}) \leq \mathtt{C} \min\left\{\frac{1}{2}\lambda_{\min}(\boldsymbol{\Sigma}), \lambda^2_{\min}(\boldsymbol{F})\frac{\Lambda^2_2(\Xi)}{\|\Xi\|}\right\}.$$

Then the following bounds hold with P-probability at least $1 - 7e^{-\mathrm{s}}$.

Asymptotic for Frobenius case. *Let*

$$n \gtrsim \max\left\{\vartheta^2_\Sigma \ln^2(\vartheta_\Sigma)(\mathrm{s} + \ln d),\, N(n,d),\, \hat{N}(n,d;\mathrm{s})\right\},$$

then

$$\Gamma_{\mathrm{F}}(n) \lesssim \sqrt{\frac{\mathrm{s} + \ln n}{n}}.$$

Asymptotic for Bures–Wasserstein case. *Let*

$$n \gtrsim \max\left\{\vartheta^2_\Sigma \ln^2(\vartheta_\Sigma)(\mathrm{s} + \ln d),\, \kappa(Q_*)N(n,d),\, \kappa(Q_*)\hat{N}(n,d;\mathrm{s})\right\},$$

then

$$\Gamma_{\mathrm{B}}(n) \lesssim \sqrt{\frac{\mathrm{s} + \ln n}{n}}.$$

PROOF. First we obtain an asymptotic bound on term $\varepsilon_\Xi(\mathrm{s})$ defined in (4.2). Let the sample size be $n \geq \vartheta^2_\Sigma \ln^2(\vartheta_\Sigma)(\mathrm{s} + \ln d)$, then

$$\varepsilon_\Sigma(\mathrm{s}) \lesssim \sqrt{\frac{\mathrm{s} + \ln d}{n}} \lesssim \sqrt{\frac{\mathrm{s} + \ln n}{n}}.$$

If at the same time $n \gtrsim \max\left\{N(n,d), \hat{N}(n,d;\mathrm{s}),\right\}$, we recall (D.10), (D.13) and combine it with Theorem 4.1. This ensures

$$\Gamma_{\mathrm{F}}(n;\mathrm{s}) \lesssim \sqrt{\frac{\mathrm{s} + \ln n}{n}}.$$

In a similar way, if $n \gtrsim \kappa(Q_*) \max\left\{N(n,d), \hat{N}(n,d;\mathrm{s}),\right\}$, we use (D.11), (D.14). Together with Theorem (4.2) it yields

$$\Gamma_{\mathrm{B}}(n;\mathrm{s}) \lesssim \sqrt{\frac{\mathrm{s} + \ln n}{n}}.$$

□

Acknowledgments. The authors would like to thank Alexandra Galitsyna, Dmitrii Kriukov and Alexander Favorov for fruitful discussions concerning applying the developed methodology to the analysis of 3D chromosome folding. It is an ongoing collaboration.