A SAMPLE AUGMENTATION: PROOFS

In this section we give the proofs omitted in Section 2.

Proof of Lemma 7 First, suppose that $k = d$, in which case $\det(\mathbf{A}^\top\mathbf{B}) = \det(\mathbf{A})\det(\mathbf{B})$. Recall that by definition the determinant can be written as:

$$\det(\mathbf{C}) = \sum_{\sigma\in\mathscr{S}_d} \mathrm{sgn}(\sigma)\prod_{i=1}^d c_{i,\sigma_i},$$

where $\mathscr{S}_d$ is the set of all permutations of $(1..d)$, and $\mathrm{sgn}(\sigma) = \mathrm{sgn}\big((1..d),\sigma\big) \in \{-1,1\}$ is the parity of the number of swaps from $(1..d)$ to σ. Using this formula and denoting $c_{ij} = \big(\mathbb{E}[\mathbf{a}\mathbf{b}^\top]\big)_{ij}$, we can rewrite the expectation as:

$$\begin{aligned}\mathbb{E}\big[\det(\mathbf{A})\det(\mathbf{B})\big] &= \sum_{\sigma,\sigma'\in\mathscr{S}_d} \mathrm{sgn}(\sigma)\,\mathrm{sgn}(\sigma')\prod_{i=1}^d \mathbb{E}\big[a_{i\sigma_i}b_{i\sigma'_i}\big] \\ &= \sum_{\sigma\in\mathscr{S}_d}\sum_{\sigma'\in\mathscr{S}_d} \mathrm{sgn}(\sigma,\sigma')\prod_{i=1}^d c_{\sigma_i\sigma'_i} \\ &= d!\sum_{\sigma'\in\mathscr{S}_d} \mathrm{sgn}(\sigma')\prod_{i=1}^d c_{i\sigma'_i} \\ &= d!\det\big(\mathbb{E}[\mathbf{a}\mathbf{b}^\top]\big),\end{aligned}$$

which completes the proof for $k = d$. The case of $k > d$ follows by induction via a standard determinantal formula:

$$\begin{aligned}\mathbb{E}\big[\det(\mathbf{A}^\top\mathbf{B})\big] &\overset{(*)}{=} \mathbb{E}\bigg[\frac{1}{k-d}\sum_{i=1}^k \det\big(\mathbf{A}_{-i}^\top\mathbf{B}_{-i}\big)\bigg] \\ &= \frac{k}{k-d}\,\mathbb{E}\big[\det\big(\mathbf{A}_{-k}^\top\mathbf{B}_{-k}\big)\big],\end{aligned}$$

where $(*)$ follows from the Cauchy-Binet formula and $\mathbf{A}_{-i}$ denotes matrix $\mathbf{A}$ with the ith row removed. ∎

Next, we state a formula which we used in the proof of Theorem 2. This lemma is an immediate implication of a result shown by [8].

Lemma 15 *Given full rank* $\mathbf{X} \in \mathbb{R}^{k\times d}$ *and* $\mathbf{y} \in \mathbb{R}^k$, *we have:*

$$\mathbf{w}^*(\mathbf{X},\mathbf{y}) = \sum_{i=1}^k \frac{\det(\mathbf{X}_{-i}^\top\mathbf{X}_{-i})}{(k-d)\det(\mathbf{X}^\top\mathbf{X})}\mathbf{w}^*(\mathbf{X}_{-i},\mathbf{y}_{-i}),$$

where $\mathbf{w}^*(\mathbf{X},\mathbf{y}) = \mathbf{X}^+\mathbf{y}$ *is the least squares solution for* $(\mathbf{X},\mathbf{y})$, *and* $\mathbf{X}^+$ *is the pseudoinverse of* $\mathbf{X}$.

Proof Let $\mathbf{I}_{-i}$ denote the identity matrix with ith diagonal entry set to zero. Note that we can write $\mathbf{w}^*(\mathbf{X}_{-i},\mathbf{y}_{-i}) = (\mathbf{I}_{-i}\mathbf{X})^+\mathbf{y}$. Moreover, by Sylvester's theorem we have

$$\frac{\det(\mathbf{X}_{-i}^\top\mathbf{X}_{-i})}{\det(\mathbf{X}^\top\mathbf{X})} = 1 - \mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i.$$

Thus, it suffices to show that

$$\mathbf{X}^+ = \sum_{i=1}^k \frac{1-\mathbf{x}_i^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{x}_i}{k-d}(\mathbf{I}_{-i}\mathbf{X})^+,$$

which is in fact precisely the formula shown in [8] (see proof of Theorem 5). ∎

B VOLUME-RESCALED GAUSSIAN: PROOFS

In this section we give the proofs omitted in Section 3.

Proof of Lemma 9 Since we are conditioning on an event which may have probability 0, this requires a careful limiting argument. Let A be any measurable event over the random matrix $\widetilde{\mathbf{X}}$ and let

$$C_{\mathbf{\Sigma}}^\epsilon \overset{def}{=} \big\{\mathbf{B}\in\mathbb{R}^{d\times d} \,:\, \|\mathbf{B}-\mathbf{\Sigma}\| \le \epsilon\big\}$$

be an ϵ-neighborhood of $\mathbf{\Sigma}$ w.r.t. the matrix 2-norm. We write the conditional probability of $\widetilde{\mathbf{X}} \in A$ given that $\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}} \in C_{\mathbf{\Sigma}}^\epsilon$ as:

$$\begin{aligned}\Pr\big(\widetilde{\mathbf{X}}\in A \,|\, \widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}}\in C_{\mathbf{\Sigma}}^\epsilon\big) &= \frac{\Pr\big(\widetilde{\mathbf{X}}\in A \,\wedge\, \widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}}\in C_{\mathbf{\Sigma}}^\epsilon\big)}{\Pr\big(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}}\in C_{\mathbf{\Sigma}}^\epsilon\big)} \\ &= \frac{\mathbb{E}\big[\mathbf{1}_{[\mathbf{X}\in A]}\mathbf{1}_{[\mathbf{X}^\top\mathbf{X}\in C_{\mathbf{\Sigma}}^\epsilon]}\det(\mathbf{X}^\top\mathbf{X})\big]}{\mathbb{E}\big[\mathbf{1}_{[\mathbf{X}^\top\mathbf{X}\in C_{\mathbf{\Sigma}}^\epsilon]}\det(\mathbf{X}^\top\mathbf{X})\big]} \\ &\le \frac{\mathbb{E}\big[\mathbf{1}_{[\mathbf{X}\in E]}\mathbf{1}_{[\mathbf{X}^\top\mathbf{X}\in C_{\mathbf{\Sigma}}^\epsilon]}\det(\mathbf{\Sigma})(1+\epsilon)^d\big]}{\mathbb{E}\big[\mathbf{1}_{[\mathbf{X}^\top\mathbf{X}\in C_{\mathbf{\Sigma}}^\epsilon]}\det(\mathbf{\Sigma})(1-\epsilon)^d\big]} \\ &= \frac{\mathbb{E}\big[\mathbf{1}_{[\mathbf{X}\in A]}\mathbf{1}_{[\mathbf{X}^\top\mathbf{X}\in C_{\mathbf{\Sigma}}^\epsilon]}\big]}{\mathbb{E}\big[\mathbf{1}_{[\mathbf{X}^\top\mathbf{X}\in C_{\mathbf{\Sigma}}^\epsilon]}\big]}\bigg(\frac{1+\epsilon}{1-\epsilon}\bigg)^d \\ &= \Pr\big(\mathbf{X}\in A \,|\, \mathbf{X}^\top\mathbf{X}\in C_{\mathbf{\Sigma}}^\epsilon\big)\bigg(\frac{1+\epsilon}{1-\epsilon}\bigg)^d \\ &\overset{\epsilon\to 0}{\longrightarrow} \Pr\big(\mathbf{X}\in A \,|\, \mathbf{X}^\top\mathbf{X}=\mathbf{\Sigma}\big).\end{aligned}$$

We can obtain a lower-bound analogous to the above upper-bound, namely $\Pr\big(\mathbf{X}\in A \,|\, \mathbf{X}^\top\mathbf{X}\in C_{\mathbf{\Sigma}}^\epsilon\big)\big(\frac{1-\epsilon}{1+\epsilon}\big)^d$, which also converges to $\Pr\big(\mathbf{X}\in A \,|\, \mathbf{X}^\top\mathbf{X}=\mathbf{\Sigma}\big)$. Thus, we conclude that:

$$\begin{aligned}\Pr\big(\widetilde{\mathbf{X}}\in A \,|\, \widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}}=\mathbf{\Sigma}\big) &= \lim_{\epsilon\to 0}\Pr\big(\widetilde{\mathbf{X}}\in A \,|\, \widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}}\in C_{\mathbf{\Sigma}}^\epsilon\big) \\ &= \Pr\big(\mathbf{X}\in A \,|\, \mathbf{X}^\top\mathbf{X}=\mathbf{\Sigma}\big),\end{aligned}$$

completing the proof. ∎

C GENERAL ALGORITHM: PROOFS

In this section we give proofs omitted in Section 4.

Proof of Lemma 12 The distribution $\mathrm{Lev}_{\widehat{\mathbf{\Sigma}},\mathcal{X}}$ integrates to one because for $\mathbf{x} \sim \mathrm{D}_{\mathcal{X}}$:

$$\mathbb{E}\big[\mathbf{x}^\top \widehat{\mathbf{\Sigma}}^{-1}\mathbf{x}\big] = \mathbb{E}\Big[\mathrm{tr}\big(\mathbf{x}\mathbf{x}^\top \widehat{\mathbf{\Sigma}}^{-1}\big)\Big] = \mathrm{tr}\big(\mathbf{\Sigma}_{\mathrm{D}_\mathcal{X}} \widehat{\mathbf{\Sigma}}^{-1}\big).$$

Next, we use the geometric-arithmetic mean inequality for the eigenvalues of matrix $\widetilde{\mathbf{\Sigma}}$ to show that:

$$\begin{aligned}\det\big(\widetilde{\mathbf{\Sigma}}\widehat{\mathbf{\Sigma}}^{-1}\big) &\leq \Big(\frac{1}{d}\mathrm{tr}\big(\widetilde{\mathbf{\Sigma}}\widehat{\mathbf{\Sigma}}^{-1}\big)\Big)^d \\ &= \Big(\frac{1}{d\,t}\sum_{i=1}^t \frac{d}{l_{\widehat{\mathbf{\Sigma}}}(\mathbf{x}_i)}\mathbf{x}_i^\top \widehat{\mathbf{\Sigma}}^{-1}\mathbf{x}_i\Big)^d = 1.\end{aligned}$$

Next, we use the formula for the normalization constant in Theorem 1 but with a modified random vector. Specifically, let $\widetilde{\mathbf{x}}_i = \sqrt{\frac{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{D}_\mathcal{X}}\widehat{\mathbf{\Sigma}}^{-1})}{l_{\widehat{\mathbf{\Sigma}}}(\mathbf{x}_i)}}\mathbf{x}_i$. Then $\mathbb{E}[\widetilde{\mathbf{x}}_i\widetilde{\mathbf{x}}_i^\top] = \mathbf{\Sigma}_{\mathrm{D}_\mathcal{X}}$ and

$$\widetilde{\mathbf{\Sigma}} = \frac{1}{t}\sum_{i=1}^t \frac{d}{l_{\widehat{\mathbf{\Sigma}}}(\mathbf{x}_i)}\mathbf{x}_i\mathbf{x}_i^\top = \frac{d}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{D}_\mathcal{X}}\widehat{\mathbf{\Sigma}}^{-1})}\frac{1}{t}\sum_{i=1}^t \widetilde{\mathbf{x}}_i\widetilde{\mathbf{x}}_i^\top.$$

So, using Lemma 7 on the vectors $\widetilde{\mathbf{x}}_i$, we have:

$$\begin{aligned}\mathbb{E}\big[\det(\widetilde{\mathbf{\Sigma}}\widehat{\mathbf{\Sigma}}^{-1})\big] &= \bigg(\frac{d}{\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{D}_\mathcal{X}}\widehat{\mathbf{\Sigma}}^{-1})}\bigg)^d \frac{\mathbb{E}[\det(\sum_i \widetilde{\mathbf{x}}_i\widetilde{\mathbf{x}}_i^\top)]}{t^d\det(\widehat{\mathbf{\Sigma}})} \\ &= \frac{d!\binom{t}{d}\det\big(\mathbb{E}[\widetilde{\mathbf{x}}_1\widetilde{\mathbf{x}}_1^\top]\big)}{t^d(\frac{1}{d}\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{D}_\mathcal{X}}\widehat{\mathbf{\Sigma}}^{-1}))^d\det(\widehat{\mathbf{\Sigma}})} \\ &= \bigg(\prod_{i=0}^{d-1}\frac{t-i}{t}\bigg)\frac{\det(\mathbf{\Sigma}_{\mathrm{D}_\mathcal{X}}\widehat{\mathbf{\Sigma}}^{-1})}{(\frac{1}{d}\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{D}_\mathcal{X}}\widehat{\mathbf{\Sigma}}^{-1}))^d} \\ &\geq \bigg(1-\frac{d}{t}\bigg)^d\frac{\det(\mathbf{\Sigma}_{\mathrm{D}_\mathcal{X}}\widehat{\mathbf{\Sigma}}^{-1})}{(\frac{1}{d}\mathrm{tr}(\mathbf{\Sigma}_{\mathrm{D}_\mathcal{X}}\widehat{\mathbf{\Sigma}}^{-1}))^d}.\end{aligned}$$

Applying Bernoulli's inequality concludes the proof. ■

Proof of Lemma 13 Let $\mathbf{X} \in \mathbb{R}^{k\times d}$ be the matrix with rows $\mathbf{x}_i^\top$ and let $q_i(\mathbf{X})$ denote the sampling probability in line 4 of Algorithm 2, given the set of row vectors. We will show that if $\mathbf{x}_1,\dots,\mathbf{x}_k \sim \mathrm{VS}_{\mathrm{D}_\mathcal{X}}^k$, then after one step of the algorithm, the remaining vectors are distributed according to $\mathrm{VS}_{\mathrm{D}_\mathcal{X}}^{k-1}$. Let A denote a measurable event over the space $(\mathbb{R}^d)^{k-1}$, and let $A' = A\times\mathbb{R}^d$ be that event marginalized over the space $(\mathbb{R}^d)^k$. We wish to compute the probability $\Pr(A)$ over the sample returned by the algorithm given input set $\{\mathbf{x}_1,\dots,\mathbf{x}_k\}$ and sampling size $k-1$. Note that since the sample $\mathbf{x}_1,\dots,\mathbf{x}_k$ is symmetric under permutations, the probability of A should not depend on which index i is selected in line 5 of Algorithm 2, so we have

$$\begin{aligned}\Pr(A) &= k\ \Pr(A \mid \text{Alg. 2 selected } i\!=\!k) \\ &\propto \mathbb{E}_{\mathrm{D}_\mathcal{X}^k}\Big[\mathbf{1}_{A'}\,q_k(\mathbf{X})\det\big(\mathbf{X}^\top\mathbf{X}\big)\Big] \\ &\propto \mathbb{E}_{\mathrm{D}_\mathcal{X}^k}\bigg[\mathbf{1}_{A'}\,\frac{\det(\mathbf{X}_{-k}^\top\mathbf{X}_{-k})}{\det(\mathbf{X}^\top\mathbf{X})}\det\big(\mathbf{X}^\top\mathbf{X}\big)\bigg] \\ &= \mathbb{E}_{\mathrm{D}_\mathcal{X}^k}\Big[\mathbf{1}_{A'}\,\det(\mathbf{X}_{-k}^\top\mathbf{X}_{-k})\Big] \\ &\propto \mathrm{VS}_{\mathrm{D}_\mathcal{X}}^{k-1}(A),\end{aligned}$$

where in the above we skipped constant factors, since they fall into the normalization constant. The lemma now follows by induction over increasing k. ■