A Notations.

In this section, we recall and introduce some notation which will be used throughout the appendix.

Block norms. By default, $\|\cdot\|$ is the Euclidean norm for vector and spectral norm for matrices. For a vector $x = [x_1, \dots, x_s] \in \mathbb{C}^{sd}$ formed of s blocks $x_i \in \mathbb{C}^d$, $1 \leqslant i \leqslant s$, we define the block norm

$$\|x\|_{\text{block}} \stackrel{\text{def.}}{=} \sup_{1 \leqslant i \leqslant s} \|x_i\|_2$$

For a vector $q = [q_1, \dots, q_s, Q_1, \dots, Q_s] \in \mathbb{C}^{s(d+1)}$ decomposed such that $q_i \in \mathbb{C}$ and $Q_i \in \mathbb{C}^d$, we define

$$\|q\|_{\text{Block}} \stackrel{\text{def.}}{=} \max_{i=1}^{s} \{|q_i|\,,\ \|Q_i\|\}.$$

Kernel The empirical kernel is defined as

$$\hat{K}(x, x') = \frac{1}{m} \sum_{k=1}^{m} \overline{\varphi_{\omega_k}(x)} \varphi_{\omega_k}(x')$$

and the limit kernel is $K(x, x) \stackrel{\text{def.}}{=} \mathbb{E}_\omega[\overline{\varphi_\omega(x)} \varphi_\omega(x')]$. The metric tensor associated to this kernel is

$$\mathbf{H}_x \stackrel{\text{def.}}{=} \mathbb{E}_\omega[\overline{\nabla \varphi_\omega(x)} \nabla \varphi_\omega(x)^\top]$$

Given an event E, we write $K_E(x, x') \stackrel{\text{def.}}{=} \mathbb{E}_\omega[\hat{K}(x, x') | E]$ to denote the conditional expectation on E.

Derivatives Given $f \in \mathscr{C}^\infty(\mathcal{X})$, by interpreting the r^{th} derivative as a multilinear map: $\nabla^r f : (\mathbb{C}^d)^r \to \mathbb{C}$, so given $Q \stackrel{\text{def.}}{=} \{q_\ell\}_{\ell=1}^r \in (\mathbb{C}^d)^r$,

$$\nabla^r f[Q] = \sum_{i_1, \cdots, i_r} \partial_{i_1} \cdots \partial_{i_r} f(x) q_{1,i_1} \cdots q_{r,i_r}.$$

and we define the r^{th} normalized derivative of f as

$$\mathrm{D}_r\left[f\right](x)[Q] \stackrel{\text{def.}}{=} \nabla^r f(x)[\{\mathbf{H}_x^{-\frac{1}{2}} q_i\}_{i=1}^r]$$

with norm $\|\mathrm{D}_r\left[f\right](x)\| \stackrel{\text{def.}}{=} \sup_{\forall \ell, \|q_\ell\| \leqslant 1} |\mathrm{D}_r\left[f\right](x)[Q]|$. We will sometimes make use the the multiarray interpretation: $\mathrm{D}_0\left[f\right] = f$, $\mathrm{D}_1\left[f\right](x) = \mathbf{H}_x^{-\frac{1}{2}} \nabla f(x) \in \mathbb{C}^d$, $\mathrm{D}_2\left[f\right](x) = \mathbf{H}_x^{-\frac{1}{2}} \nabla^2 f(x) \mathbf{H}_x^{-\frac{1}{2}} \in \mathbb{C}^{d \times d}$.

For a bivariate function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$, $\partial_{1,i}$ (resp. $\partial_{2,i}$) designates the derivative with respect to the i^{th} coordinate of the first variable (resp. second variable), and similarly ∇_i and ∇_i^2 denote the gradient and Hessian on the i^{th} coordinate respectively.

For $i, j \in \{0, 1, 2\}$, let $K^{(ij)}(x, x')$ be a "bi"-multilinear map, defined for $Q \in (\mathbb{C}^d)^i$ and $V \in (\mathbb{C}^d)^j$ as

$$[Q] K^{(ij)}(x, x')[V] \stackrel{\text{def.}}{=} \mathbb{E}[\overline{\mathrm{D}_i\left[\varphi_\omega\right](x)[Q]} \mathrm{D}_j\left[\varphi_\omega\right](x')[V]]$$

and $\left\|K^{(ij)}(x, x')\right\| \stackrel{\text{def.}}{=} \sup_{Q,V} \left\|[Q] K^{(ij)}(x, x')[V]\right\|$ where the supremum is defined over all $Q \stackrel{\text{def.}}{=} \{q_\ell\}_{\ell=1}^i$, $V \stackrel{\text{def.}}{=} \{v_\ell\}_{\ell=1}^j$ with $\|q_\ell\| \leqslant 1$, $\|v_\ell\| \leqslant 1$.

When $i + j \leqslant 2$, an equivalent definition is $K^{(ij)}(x, x') = \mathbb{E}[\overline{\mathrm{D}_i\left[\varphi_\omega\right](x)} \mathrm{D}_j\left[\varphi_\omega\right](x')^\top]$, and we note that $K^{(00)} = K$, and we have normalized so that $\mathrm{Re}\left(K^{(11)}(x, x)\right) = -\mathrm{Re}\left(K^{(02)}(x, x)\right)$. Finally, we will make use of the still equivalent definition: $[q] K^{(12)}(x, x') = \mathbb{E}[\overline{q^\top \mathrm{D}_1\left[\varphi_\omega\right](x)} \mathrm{D}_2\left[\varphi_\omega\right](x')^\top] \in \mathbb{C}^{d \times d}$.

Kernel constants For for $i,j \in \{(0,0),(0,1)\}$, define $B_{ij} \stackrel{\text{def.}}{=} \sup_{x,x' \in \mathcal{X}} \left|K^{(ij)}(x,x')\right|$, for $(i,j) \in \{(0,2),(1,2)\}$,

$$B_{ij} \stackrel{\text{def.}}{=} \sup\left\{ \left\|K^{(ij)}(x,x')\right\| \ ; \ d_{\mathbf{H}}(x,x') \leqslant r_{\text{near}} \text{ or } d_{\mathbf{H}}(x,x') > \Delta/2 \right\}.$$

and define for $i = 1,2$

$$B_{ii} \stackrel{\text{def.}}{=} \sup_{x \in \mathcal{X}} \left\|K^{(ii)}(x,x)\right\|.$$

For convenience, we define

$$B_i \stackrel{\text{def.}}{=} B_{0i} + B_{1i} + 1, \quad B \stackrel{\text{def.}}{=} \sum_{\substack{i,j \in \{0,1,2\} \\ i+j \leqslant 3}} B_{ij} + 1. \tag{A.1}$$

Matrices and vectors We will make use of the following vectors and matrices throughout: Given $X \stackrel{\text{def.}}{=} \{x_j\}_{j=1}^s \in \mathcal{X}^s$ and $a \in \mathbb{C}^s$ which are always clear from context, define the vector $\gamma_X(\omega) \in \mathbb{C}^{s(d+1)}$ as

$$\gamma_X(\omega) \stackrel{\text{def.}}{=} \left(\left(\overline{\varphi_\omega(x_i)}\right)_{i=1}^s, \left(\overline{\mathrm{D}_1\left[\varphi_\omega\right](x_i)}^\top\right)_{i=1}^s \right)^\top, \tag{A.2}$$

and

$$\begin{aligned} \Upsilon_X &\stackrel{\text{def.}}{=} \mathbb{E}_\omega[\gamma(\omega)\gamma(\omega)^*] \in \mathbb{C}^{s(d+1)\times s(d+1)} \\ \mathbf{f}_X(x) &\stackrel{\text{def.}}{=} \mathbb{E}_\omega[\gamma(\omega)\varphi_\omega(x)] \in \mathbb{C}^{s(d+1)} \\ \alpha &\stackrel{\text{def.}}{=} \Upsilon_X^{-1}\mathbf{u}_s, \qquad \mathbf{u}_s = \begin{pmatrix} \operatorname{sign}(a) \\ 0_{sd} \end{pmatrix}. \end{aligned}$$

Note that the diagonal of Υ has only 1's. For $\omega_1, \ldots, \omega_m$, we denote their empirical versions as:

$$\begin{aligned} \hat{\Upsilon}_X &\stackrel{\text{def.}}{=} \frac{1}{m}\sum_{k=1}^m \gamma(\omega_k)\gamma(\omega_k)^*, \\ \hat{\mathbf{f}}_X(x) &\stackrel{\text{def.}}{=} \frac{1}{m}\sum_{k=1}^m \gamma(\omega_k)\varphi_{\omega_k}(x), \quad \hat{\alpha} \stackrel{\text{def.}}{=} \hat{\Upsilon}_X^{-1}\mathbf{u}_s. \end{aligned}$$

which will serve us to construct our certificate, using the properties of their respective limit version.

We remark that $\mathbf{G}_X^{-1/2}\Gamma_X^*\Gamma_X\mathbf{G}_X^{-1/2} = \hat{\Upsilon}_X$, where Γ_X is defined in the main paper and

$$\mathbf{G}_X = \begin{pmatrix} \mathrm{Id}_s & & & 0 \\ & \mathbf{H}_{x_1} & & \\ & & \ddots & \\ 0 & & & \mathbf{H}_{x_s} \end{pmatrix} \tag{A.3}$$

The vanishing derivative pre-certificate $\hat{\eta}_{X,a}$ is $\hat{\alpha}^\top\hat{\mathbf{f}}_X(\cdot)$ and the limit pre-certificate is $\eta_{X,a} \stackrel{\text{def.}}{=} \alpha^\top\mathbf{f}_X(\cdot)$. When the set of points X and amplitudes a are clear from context, we will drop the subscripts and write instead γ, Υ, $\mathbf{f}$, η, and so on.

Metric induced distances Given $X = (x_j)_{j=1}^s \in \mathcal{X}^s$ and $X' = (x'_j)_{j=1}^s \in \mathcal{X}^s$, denote $d_{\mathbf{H}}(X,X') \stackrel{\text{def.}}{=} \sqrt{\sum_j d_{\mathbf{H}}(x_j,x'_j)^2}$. Observe also that $\mathbf{G}_X$ is positive definite for all X and induces a metric on $\mathbb{R}^s \times \mathcal{X}^s$ so that given $a, a' \in \mathbb{R}^s$ and $X, X' \in \mathcal{X}^s$,

$$d_G((a,X),(a',X')) = \sqrt{\|a-a'\|_2^2 + d_{\mathbf{H}}(X,X')^2}.$$

Stochastic gradient bounds For $r \in \mathbb{N}$, define the following random variable

$$L_r(\omega) = \sup_{x \in \mathcal{X}} \|\mathrm{D}_r [\varphi_\omega](x)\|,$$

and for $i, j \in \mathbb{N}$, define $L_{ij}(\omega) \stackrel{\text{def.}}{=} \sqrt{L_i(\omega)^2 + L_j(\omega)^2}$. For $i = 0, 1, 2, 3$, let F_i be such that

$$\mathbb{P}_\omega (L_j(\omega) > t) \leqslant F_i(t),$$

Throughout, for $(\bar{L}_j)_{j=0}^3 \in \mathbb{R}_+^4$, the event $\bar{E}$ is defined as

$$\bar{E} \stackrel{\text{def.}}{=} \bigcap_{k=1}^m E_{\omega_k} \quad \text{where} \quad E_\omega \stackrel{\text{def.}}{=} \{L_j(\omega) \leqslant \bar{L}_j,\ \forall j = 0, 1, 2, 3\}. \tag{A.4}$$

B Proof of Theorem 2

In this section, we consider the (limit) vanishing derivative pre-certificate

$$\eta(x) = \mathbf{u}^\top \Upsilon_X^{-1} \mathbf{f}_X(x).$$

Note that

$$\mathrm{D}_2 [\eta](x) = \sum_{i=1}^s \alpha_{1,i} K^{(02)}(x_i, x) + [\alpha_{2,i}] K^{(12)}(x_i, x)$$

where we have decomposed $\alpha = [\alpha_{1,1}, \ldots, \alpha_{1,s}, \alpha_{2,1}, \ldots, \alpha_{2,s}] \in \mathbb{C}^{s(d+1)}$ where $\alpha_{2,i} \in \mathbb{C}^d$.

We aim to prove that η is nondegenerate if K is an admissible kernel. Our first lemma shows that nondegeneracy of η within each small neighbourhood of x_i can be established by controlling the real and imaginary parts of $\mathrm{D}_2 [\eta]$ in each small region:

Lemma B.1. *Let $\varepsilon > 0$. Let $x_0 \in \mathcal{X}$ and let $\sigma \in \mathbb{C}$ be such that $|\sigma| = 1$. Suppose that $\eta \in \mathscr{C}^2(\mathcal{X}; \mathbb{C})$ is such that $\eta(x_0) = \sigma$, $\nabla \eta(x_0) = 0$ and $\mathrm{Re}(\overline{\sigma} \mathrm{D}_2 [\eta](x_0)) \prec -\varepsilon \mathrm{Id}$. Then, $\nabla^2 |\eta|^2 (x_0) \prec -2\varepsilon \mathrm{Id}$. If in addition, we have $c, r > 0$ with $\varepsilon r < 1$ and $c^2 \leqslant (1 - \varepsilon r^2)/(\varepsilon r^2)$ such that for all x such that $d_{\mathbf{H}}(x, x_0) \leqslant r$,*

$$\mathrm{Re}(\overline{\sigma} \mathrm{D}_2 [\eta](x)) \prec -\varepsilon \mathrm{Id} \quad \text{and} \quad \|\mathrm{Im}(\overline{\sigma} \mathrm{D}_2 [\eta](x))\| \leqslant c\varepsilon,$$

then, $|\eta(x)|^2 \leqslant 1 - \varepsilon^2 d_{\mathbf{H}}(x, x_0)^2$ for all x such that $d_{\mathbf{H}}(x, x_0) \leqslant r$.

Proof. The first claim follows immediately from the computation: by writing $\eta = \eta_r(x) + \mathrm{i}\eta_i(x)$ where η_i and η_r are real valued functions,

$$\frac{1}{2} \mathrm{D}_2 \left[|\eta|^2\right] = \mathrm{Re}\left(\overline{\mathrm{D}_1 [\eta]} \mathrm{D}_1 [\eta]^\top + \mathrm{D}_2 [\eta] \overline{\eta}\right),$$

and evaluation at x_0 gives the required result.

Let $\gamma : [0, 1] \to \mathcal{X}$ be a piecewise smooth path such that $\gamma(0) = x_0$, $\gamma(1) = x$.

$$\begin{aligned}
\eta(x) &= \eta(x_0) + \int_0^1 (1 - t) \langle \nabla^2 \eta(\gamma(t)) \gamma'(t),\ \gamma'(t) \rangle \mathrm{d}t \\
&= \eta(x_0) + \int_0^1 (1 - t) \langle \mathrm{D}_2 [\eta](\gamma(t)) \mathbf{H}_{\gamma(t)}^{\frac{1}{2}} \gamma'(t),\ \mathbf{H}_{\gamma(t)}^{\frac{1}{2}} \gamma'(t) \rangle \mathrm{d}t.
\end{aligned}$$

So,

$$\mathrm{Re}(\overline{\sigma} \eta(x)) = 1 + \inf_\gamma \mathrm{Re}\left(\overline{\sigma} \int_0^1 (1 - t) \langle \mathrm{D}_2 [\eta](\gamma(t)) \mathbf{H}_{\gamma(t)}^{\frac{1}{2}} \gamma'(t),\ \mathbf{H}_{\gamma(t)}^{\frac{1}{2}} \gamma'(t) \rangle \mathrm{d}t\right) \leqslant 1 - \varepsilon d_{\mathbf{H}}(x, x')^2$$

if we minimise over all paths from x to x_0. Similarly,

$$\|\mathrm{Im}\left(\overline{\sigma}\eta(x)\right)\| \leqslant c\varepsilon d_{\mathbf{H}}(x,x_0)^2$$

Therefore,

$$\begin{aligned}|\eta(x)|^2 &\leqslant \left|1-\varepsilon d_{\mathbf{H}}(x,x_0)^2\right|^2 + \left|c\varepsilon d_{\mathbf{H}}(x,x_0)^2\right|^2 \\ &\leqslant 1-2\varepsilon d_{\mathbf{H}}(x,x_0)^2 + \varepsilon^2 d_{\mathbf{H}}(x,x_0)^4 + c^2\varepsilon^2 d_{\mathbf{H}}(x,x_0)^4 \\ &= 1-\varepsilon d_{\mathbf{H}}(x,x_0)^2 - \varepsilon d_{\mathbf{H}}(x,x_0)^2\left(1-\varepsilon d_{\mathbf{H}}(x,x_0)^2\left(1+c^2\right)\right) \leqslant 1-\varepsilon d_{\mathbf{H}}(x,x_0)^2.\end{aligned}$$

□

Proof of Theorem 2. In order to show that η is $(\varepsilon_0/2, \varepsilon_2/2)$-nondegenerate, it is enough to show that

$$\forall x \in \mathcal{X}^{\text{far}}, \quad |\eta(x)| \leqslant 1-\varepsilon_0/2 \tag{B.1}$$

$$\forall x \in \mathcal{X}^{\text{near}}, \quad \mathrm{Re}\left(\overline{\mathrm{sign}(a_j)}\mathrm{D}_2\left[\eta\right](x)\right) \prec -\frac{\varepsilon_2}{2}\mathrm{Id} \quad \text{and} \quad \left\|\mathrm{Im}\left(\overline{\mathrm{sign}(a_j)}\mathrm{D}_2\left[\eta\right](x)\right)\right\| \leqslant \frac{p}{4}\varepsilon_2 \tag{B.2}$$

where $p = \sqrt{\frac{1-\varepsilon_2 r_{\text{near}}^2/2}{\varepsilon_2 r_{\text{near}}^2/2}}$.

We first prove that the matrix Υ is invertible. To this end, we write

$$\Upsilon = \begin{pmatrix}\Upsilon_0 & \Upsilon_1^\top \\ \Upsilon_1 & \Upsilon_2\end{pmatrix} \tag{B.3}$$

where $\Upsilon_0 \overset{\text{def.}}{=} (K(x_i,x_j))_{i,j=1}^s \in \mathbb{C}^{s\times s}$, $\Upsilon_1 \overset{\text{def.}}{=} (K^{(10)}(x_i,x_j))_{i,j=1}^s \in \mathbb{C}^{sd\times s}$, and $\Upsilon_2 \overset{\text{def.}}{=} (K^{(11)}(x_i,x_j))_{i,j=1}^s \in \mathbb{C}^{sd\times sd}$. By definition of $K^{(ij)}$, Υ (and also Υ_0 and Υ_2) has only 1's on its diagonal.

To prove the invertibility of Υ, we use the Schur complement of Υ, and in particular it suffices to prove that Υ_2 and the Schur complement $\Upsilon_S \overset{\text{def.}}{=} \Upsilon_0 - \Upsilon_1\Upsilon_2^{-1}\Upsilon_1^\top$ are both invertible. To show that Υ_2 is invertible, we define $A_{ij} = K^{(11)}(x_i,x_j)$. So Υ_2 has the form:

$$\Upsilon_2 = \begin{pmatrix}\mathrm{Id} & A_{12} & \dots & A_{1s} \\ A_{21} & \mathrm{Id} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ A_{s1} & \dots & \dots & \mathrm{Id}\end{pmatrix}$$

and by Lemma G.6, we have

$$\|\mathrm{Id}-\Upsilon_2\|_{\text{block}} \leqslant \max_i \sum_j \|A_{ij}\| \leqslant 1/4.$$

Since $\|\mathrm{Id}-\Upsilon_2\|_{\text{block}} < 1$, Υ_2 is invertible, and we have $\left\|\Upsilon_2^{-1}\right\|_{\text{block}} \leqslant \frac{1}{1-\|I-\Upsilon_2\|_{\text{block}}} \leqslant \frac{4}{3}$. Next, again with Lemma G.6, we can bound

$$\begin{aligned}\|I-\Upsilon_0\|_\infty &= \max_i \sum_{j\neq i} |K(x_i,x_j)| \leqslant \frac{\varepsilon_0}{16} \\ \|\Upsilon_1\|_{\infty\to\text{block}} &\leqslant \max_i \sum_j \left\|K^{(10)}(x_i,x_j)\right\| \leqslant h \quad \text{since } K^{(10)}(x,x)=0 \\ \left\|\Upsilon_1^\top\right\|_{\text{block}\to\infty} &\leqslant \max_i \sum_j \left\|K^{(10)}(x_j,x_i)\right\| \leqslant h\end{aligned}$$

Hence, we have

$$\|I-\Upsilon_S\|_\infty \leqslant \|I-\Upsilon_0\|_\infty + \left\|\Upsilon_1^\top\right\|_{\text{block}\to\infty}\left\|\Upsilon_2^{-1}\right\|_{\text{block}}\|\Upsilon_1\|_{\infty\to\text{block}} \leqslant \frac{\varepsilon_0}{16} + \frac{4}{3}h^2 \leqslant \frac{\varepsilon_0}{8} \tag{B.4}$$

since $h \leqslant \frac{\varepsilon_0}{32}$. Therefore the Schur complement of Υ is invertible and so is Υ.

Expression of η. By definition, $\eta =$ satisfies $\eta(x_i) = \mathrm{sign}(a_i)$ and $\nabla\eta(x_i) = 0$.

We divide:

$$\alpha = \Upsilon^{-1}\mathbf{u}_s = \begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}$$

where $\alpha_1 \in \mathbb{C}^s$ and $\alpha_2 \in \mathbb{C}^{sd}$, and we denote $\alpha_{2,i} \in \mathbb{C}^d$ blocks such that $\alpha_2 = [\alpha_{2,1}, \dots, \alpha_{2,s}]$.

The Schur's complement of Υ allows us to express α_1 and α_2 as

$$\alpha_1 = \Upsilon_S^{-1}\,\mathrm{sign}(a), \qquad \alpha_2 = -\Upsilon_2^{-1}\Upsilon_1\Upsilon_S^{-1}\,\mathrm{sign}(a) \tag{B.5}$$

and therefore we can bound

$$\|\alpha_1\|_\infty \leqslant \frac{1}{1-\varepsilon_0/8} \tag{B.6}$$

$$\|\alpha_2\|_{\text{block}} \leqslant \frac{8}{3}h \leqslant 4h \tag{B.7}$$

Moreover, we have

$$\|\alpha_1 - \mathrm{sign}(a)\|_\infty \leqslant \left\|I - \Upsilon_S^{-1}\right\|_\infty \leqslant \left\|\Upsilon_S^{-1}\right\|_\infty \|I - \Upsilon_S\|_\infty \leqslant \frac{1}{4} \tag{B.8}$$

Non-degeneracy. We can now prove that η is non-degenerate.

Let x be such that $d_{\mathbf{H}}(x_i, x) \leqslant r_{\text{near}}$. We need to prove that for all x such that $d_{\mathbf{H}}(x, x_i) \leqslant r$,

$$\mathrm{Re}\left(\overline{\mathrm{sign}(a_i)}\mathrm{D}_2\left[\eta\right](x)\right) \prec -\frac{\varepsilon_2}{2}\mathrm{Id} \quad \text{and} \quad \left\|\mathrm{Im}\left(\overline{\mathrm{sign}(a_i)}\mathrm{D}_2\left[\eta\right](x)\right)\right\| \leqslant \frac{\varepsilon_2}{2}\sqrt{\frac{2-\varepsilon r_{\text{near}}^2}{\varepsilon_2 r_{\text{near}}^2}}.$$

Then, since $r_{\text{near}} \leqslant \Delta/2$ and the x_i's are Δ-separated, for all $j \neq i$ we have $d_{\mathbf{H}}(x, x_j) \geqslant \Delta/2$. Then, we have

$$\begin{aligned}\overline{\mathrm{sign}(a_i)}\mathrm{D}_2\left[\eta\right](x) = \overline{\mathrm{sign}(a_i)}\Bigg[&\alpha_{1,i}K^{(02)}(x_i,x) + \sum_{j\neq i}\alpha_{1,j}K^{(02)}(x_j,x)\\ &+ [\alpha_{2,i}]K^{(12)}(x_i,x) + \sum_{j\neq i}[\alpha_{2,j}]K^{(12)}(x_j,x)\Bigg]\end{aligned}$$

$$\begin{aligned}\mathrm{Re}\left(\overline{\mathrm{sign}(a_i)}\mathrm{D}_2\left[\eta\right](x)\right) &\preccurlyeq (1 - \|\alpha_1 - \mathrm{sign}(a)\|_\infty)\mathrm{Re}\left(K^{(02)}(x_i,x)\right) + \|\alpha_1\|_\infty \sum_{j\neq i}\left\|K^{(02)}(x_j,x)\right\|\mathrm{Id}\\ &\quad + \left(\left\|K^{(12)}(x_i,x)\right\| + \sum_{j\neq i}\left\|K^{(12)}(x_j,x)\right\|\right)\|\alpha_2\|_{\text{block}}\,\mathrm{Id}\\ &\preccurlyeq \left(-\frac{3}{4}\varepsilon_2 + \frac{1}{1-\varepsilon_0/8}\frac{\varepsilon_2}{16} + 4h(B_{12}+1)\right)\mathrm{Id} \preccurlyeq \varepsilon_2\left(-\frac{3}{4}+\frac{1}{4}\right)\mathrm{Id} \preccurlyeq -\frac{\varepsilon_2}{2}\mathrm{Id}\,.\end{aligned}$$

Taking the imaginary part, we have

$$\begin{aligned}\left\|\mathrm{Im}\left(\overline{\mathrm{sign}(a_i)}\mathrm{D}_2\left[\eta\right](x)\right)\right\| &\leqslant\ (1 + \|\alpha_1 - \mathrm{sign}(a)\|)\left\|\mathrm{Im}\left(K^{(02)}(x_i,x)\right)\right\| + \|\alpha_1\|_\infty\sum_{j\neq i}\left\|K^{(02)}(x_j,x)\right\|\\ &\quad + \left(\left\|K^{(12)}(x_i,x)\right\| + \sum_{j\neq i}\left\|K^{(12)}(x_j,x)\right\|\right)\|\alpha_2\|_{\text{block}}\\ &\leqslant \left(\frac{5c\varepsilon_2}{4} + \frac{1}{(1-\varepsilon_0/8)}h + 4h(B_{12}+1)\right) \leqslant \frac{5c\varepsilon_2}{4} + h\left(4B_{12}+6\right) \leqslant \frac{\varepsilon_2}{2}\sqrt{\frac{2-\varepsilon r_{\text{near}}^2}{\varepsilon_2 r_{\text{near}}^2}}.\end{aligned}$$

So, by Lemma B.1, for each $i = 1, \ldots, s$, $|\eta(x)| \leqslant 1 - \varepsilon_2/2 d_{\mathbf{H}}(x, x_i)$ for all $x \in \mathcal{X}$ such that $d_{\mathbf{H}}(x, x_i) \leqslant r_{\text{near}}$.

Next, for any x such that $d_{\mathbf{H}}(x, x_i) \geqslant r_{\text{near}}$ for all x_i's, we can say that there exists (at most) one index i such that $d_{\mathbf{H}}(x, x_i) \geqslant r_{\text{near}}$ and for all $j \neq i$ we have $d_{\mathbf{H}}(x, x_j) \geqslant \Delta/2$. We have

$$\begin{aligned}
|\eta(x)| &= \Bigg| \alpha_{1,i} K(x_i, x) + \sum_{j \neq i} \alpha_{1,j} K(x_j, x) \\
&\quad + K^{(10)}(x_i, x)^\top \alpha_{2,i} + \sum_{j \neq i} K^{(10)}(x_j, x)^\top \alpha_{2,j} \Bigg| \\
&\leqslant \|\alpha_1\|_\infty \left(|K(x_i, x)| + \sum_{j \neq i} |K(x_j, x)| \right) \\
&\quad + \|\alpha_2\|_{\text{block}} \left(\left\| K^{(10)}(x_i, x) \right\| + \sum_{j \neq i} \left\| K^{(10)}(x_j, x) \right\| \right) \\
&\leqslant \frac{1 - \varepsilon_0 + \varepsilon_0/16}{1 - \varepsilon_0/8} + 4h(B_{10} + 1) \leqslant 1 - \frac{\varepsilon_0}{2}\,.
\end{aligned}$$

□

Remark B.1. *Assuming that the derivatives of the kernel decay like a function $f(\|x - x'\|)$ when, there is always a separation $\Delta \propto f^{-1}(1/(C s_{\max}))$ such that the kernel is admissible. Ex: when $f = x^{-p}$, we have $\Delta \propto s_{\max}^{1/p}$ (eg Cauchy). When $f = e^{-x^p}$, we have $\Delta \propto \log^{1/p}(s_{\max})$ (eg Gaussian).*

C Preliminaries

In this section, we present some preliminary results which will be used for proving our main results. We assume that K is admissible, and given a set of points $X \in \mathcal{X}^s$, let $\mathcal{X}_j^{\text{near}} \overset{\text{def.}}{=} \{x \in \mathcal{X} \; ; \; d_{\mathbf{H}}(x, x_j) \leqslant r_{\text{near}}\}$, $\mathcal{X}^{\text{near}} \overset{\text{def.}}{=} \bigcup_{j=1}^s \mathcal{X}_j^{\text{near}}$ and $\mathcal{X}^{\text{far}} \overset{\text{def.}}{=} \mathcal{X} \setminus \mathcal{X}^{\text{near}}$.

C.1 On the determistic kernel

For an admissible kernel, we have the following additional bounds that will be handy.

Lemma C.1. *Assume K is an admissible kernel, let $X \in \mathcal{X}^s$ be Δ-separated points. Then we have the following:*

(i) Υ *is invertible and satisfies*

$$\|\text{Id} - \Upsilon\| \leqslant \frac{1}{2} \quad \textit{and} \quad \|\text{Id} - \Upsilon\|_{\text{Block}} \leqslant \frac{1}{2}. \tag{C.1}$$

(ii) For any vector $q \in \mathbb{C}^{s(d+1)}$ and any $x \in \mathcal{X}^{\text{far}}$, we have

$$\|\mathbf{f}(x)\| \leqslant B_0 \quad \textit{and} \quad \left| q^\top \mathbf{f}(x) \right| \leqslant B_0 \|q\|_{\text{Block}} \tag{C.2}$$

(iii) For any vector $q \in \mathbb{C}^{s(d+1)}$ and any $x \in \mathcal{X}^{\text{near}}$ we have the bound:

$$\left\| \mathrm{D}_2 \left[q^\top \mathbf{f}(.) \right](x) \right\| \leqslant \|q\| B_2 \quad \textit{and} \quad \left\| \mathrm{D}_2 \left[q^\top \mathbf{f}(.) \right](x) \right\| \leqslant \|q\|_{\text{Block}} B_2 \tag{C.3}$$

Proof. We bound the spectral norm of $\mathrm{Id} - \Upsilon$. Define $y \in \mathbb{C}^{s(d+1)}$ decomposed as $y = [y_1, \ldots, y_s, Y_1, \ldots, Y_s]$ where $Y_i \in \mathbb{R}^d$, such that $\|y\| \leqslant 1$. We have

$$\begin{aligned}
\|(\mathrm{Id} - \Upsilon)y\|^2 = &\sum_{i=1}^{s} \left| \sum_{j \neq i} K(x_i, x_j) y_j + \sum_{j=1}^{s} K^{(10)}(x_i, x_j)^\top Y_j \right|^2 \\
&+ \left\| \sum_{j} y_j K^{(10)}(x_i, x_j) + \sum_{j \neq i} K^{(11)}(x_i, x_j) Y_j \right\|^2 \\
\leqslant &\sum_{i=1}^{s} \left(\sum_{j \neq i} |K(x_i, x_j)| \, |y_j| + \sum_{j=1}^{s} \left\| K^{(10)}(x_i, x_j) \right\| \|Y_j\| \right)^2 \\
&+ \left(\sum_{j} |y_j| \left\| K^{(10)}(x_i, x_j) \right\| + \sum_{j \neq i} \left\| K^{(11)}(x_i, x_j) \right\| \|Y_j\| \right)^2 \\
\leqslant &\max_{d_{\mathbf{H}}(x, x') \geqslant \Delta} \left(|K(x, x')| , \left\| K^{(10)}(x, x') \right\| , \left\| K^{(11)}(x, x') \right\| \right)^2 \sum_i 2 \left(\sum_j |y_j| + \|Y_j\| \right)^2 \\
\leqslant &\, 4s^2 \max_{d_{\mathbf{H}}(x, x') \geqslant \Delta} \left(|K(x, x')| , \left\| K^{(10)}(x, x') \right\| , \left\| K^{(11)}(x, x') \right\| \right)^2
\end{aligned}$$

by Cauchy-Schwartz inequality and since $K^{(10)}(x, x) = 0$ for all $x \in \mathcal{X}$. Since by hypothesis we have

$$\max_{d_{\mathbf{H}}(x, x') \geqslant \Delta} \left(|K(x, x')| , \left\| K^{(10)}(x, x') \right\| , \left\| K^{(11)}(x, x') \right\| \right) \leqslant \frac{1}{4 s_{\max}} ,$$

we obtain

$$\|\mathrm{Id} - \Upsilon\| \leqslant \frac{1}{2} \tag{C.4}$$

and we deduce (i). A near identical argument also yields $\|\Upsilon - \mathrm{Id}\|_{\text{Block}} \leqslant \frac{1}{4}$.

For (ii), let $x \in \mathcal{X}^{\text{far}}$, then we have

$$\begin{aligned}
\|\mathbf{f}(x)\| &\leqslant \left(\sum_{i=1}^{s} |K(x_i, x)|^2 + \left\| K^{(10)}(x_i, x) \right\|^2 \right)^{\frac{1}{2}} \\
&\leqslant \left(B_{00}^2 + \frac{(s-1)\varepsilon_0^2}{(16 s_{\max})^2} + B_{10}^2 + \frac{(s-1)}{s_{\max}^2} \right)^{\frac{1}{2}} \leqslant B_0
\end{aligned}$$

for which, similar to the proof above, we have used the fact that x is $\Delta/2$-separated from at least $s-1$ points x_i. Similarly, for any vector $q = [q_1, \ldots, q_s, Q_1, \ldots, Q_s] \in \mathbb{C}^{s(d+1)}$ and any $x \in \mathcal{X}^{\text{far}}$, we have

$$\begin{aligned}
\left\| q^\top \mathbf{f}(x) \right\| &\leqslant \sum_{i=1}^{s} |q_i| \, |K(x_i, x)| + \|Q_i\| \left\| K^{(10)}(x_i, x) \right\| \\
&\leqslant \|q\|_{\text{Block}} \left(B_{00} + \frac{(s-1)\varepsilon_0}{32 s_{\max})} + B_{10} + \frac{(s-1)\varepsilon_0}{32 s_{\max}} \right) \leqslant B_0 \|q\|_{\text{Block}} .
\end{aligned}$$

For any $x \in \mathcal{X}^{\text{near}}$ we have the bound:

$$\begin{aligned}\left\|\mathrm{D}_2\left[q^\top \mathbf{f}\right](x)\right\| &= \left\|\sum_{i=1}^s q_i K^{(02)}(x_i,x) + [Q_i]K^{(12)}(x_i,x)\right\| \\ &\leqslant \|q\| \left(\sum_{i=1}^s \left\|K^{(02)}(x_i,x)\right\|^2 + \left\|K^{(12)}(x_i,x)\right\|^2\right)^{\frac{1}{2}} \\ &\leqslant \|q\| B_2\end{aligned}$$

and

$$\begin{aligned}\left\|\mathrm{D}_2\left[q^\top \mathbf{f}\right](x)\right\| &= \left\|\sum_{i=1}^s q_i K^{(02)}(x_i,x) + [Q_i]K^{(12)}(x_i,x)\right\| \\ &\leqslant \|q\|_{\text{Block}} \left(\sum_{i=1}^s \left\|K^{(02)}(x_i,x)\right\| + \left\|K^{(12)}(x_i,x)\right\|\right) \\ &\leqslant \|q\|_{\text{Block}} B_2\end{aligned}$$

□

C.2 Lipschitz bounds

Lemma C.2 (Local Lipschitz constant of φ_ω and higher order derivatives)**.** *Suppose that* $\|\mathrm{D}_j[\varphi_\omega](x)\| \leqslant \bar{L}_j$ *for all* $x \in \mathcal{X}$*. For all* x, x' *with* $d_{\mathbf{H}}(x,x') \leqslant r_{\text{near}}$*, we have*

(i) $|\varphi_\omega(x) - \varphi_\omega(x')| \leqslant \mathcal{L}_0 d_{\mathbf{H}}(x,x')$,

(ii) $\|\mathrm{D}_1[\varphi_\omega](x) - \mathrm{D}_1[\varphi_\omega](x')\| \leqslant \mathcal{L}_1 d_{\mathbf{H}}(x,x')$,

(iii) $\|\mathrm{D}_2[\varphi_\omega](x) - \mathrm{D}_2[\varphi_\omega](x')\| \leqslant \mathcal{L}_2 d_{\mathbf{H}}(x,x')$,

where $\mathcal{L}_0 \overset{\text{def.}}{=} \bar{L}_1$*,* $\mathcal{L}_1 \overset{\text{def.}}{=} \bar{L}_1 C_{\mathbf{H}} + \bar{L}_2(1 + C_{\mathbf{H}} r_{\text{near}})$ *and* $\mathcal{L}_2 \overset{\text{def.}}{=} \bar{L}_2\left(C_{\mathbf{H}} + C_{\mathbf{H}}^2 r_{\text{near}} + 1\right) + \bar{L}_3(1 + C_{\mathbf{H}} r_{\text{near}})^2$*. As a consequence, for all* $X = (x_j)$ *and* $X' = (x'_j)$ *such that* $d_{\mathbf{H}}(x_j, x'_j) \leqslant r_{\text{near}}$*, we have*

$$\sup_{\|q\|=1} \left\|\mathrm{D}_r\left[q^\top(\hat{\mathbf{f}}_X - \hat{\mathbf{f}}_{X'})\right](y)\right\| \leqslant \bar{L}_r \sqrt{\mathcal{L}_0^2 + \mathcal{L}_1^2}\, d_{\mathbf{H}}(X,X').$$

Proof. Let $x, x' \in \mathcal{X}$ with $d_{\mathbf{H}}(x,x') \leqslant r_{\text{near}}$. Recall that $\left\|\mathbf{H}_{x'}^{\frac{1}{2}}\mathbf{H}_x^{-\frac{1}{2}} - \mathrm{Id}\right\| \leqslant C_{\mathbf{H}} d_{\mathbf{H}}(x,x')$, and so, $\left\|\mathbf{H}_{x'}^{\frac{1}{2}}\mathbf{H}_x^{-\frac{1}{2}}\right\| \leqslant 1 + C_{\mathbf{H}} r_{\text{near}}$.

Let $p : [0,1] \to \mathcal{X}$ be a piecewise smooth path such that $p(0) = x'$, $p(1) = x$. Then, by Taylor's theorem,

$$\varphi_\omega(x) - \varphi_\omega(x') = \int_{t=0}^1 \langle \mathbf{H}_{p(t)}^{-\frac{1}{2}} \nabla\varphi_\omega(p(t)),\, \mathbf{H}_{p(t)}^{\frac{1}{2}} p'(t)\rangle \mathrm{d}t \leqslant \bar{L}_1 \int_0^1 \left\|\mathbf{H}_{p(t)}^{\frac{1}{2}} p'(t)\right\| \mathrm{d}t \tag{C.5}$$

so taking the minimum over all paths p yields $|\varphi_\omega(x) - \varphi_\omega(x')| \leqslant \bar{L}_1 d_{\mathbf{H}}(x,x')$.

Given $q \in \mathbb{R}^d$, by Taylor's theorem,

$$\begin{aligned}&\mathrm{D}_1[\varphi_\omega](x)[q] = \nabla\varphi(x)[\mathbf{H}_x^{-\frac{1}{2}} q] = \nabla\varphi(x')[\mathbf{H}_x^{-\frac{1}{2}} q] + \int \nabla^2\varphi_\omega(p(t))[\mathbf{H}_x^{-\frac{1}{2}} q, p'(t)]\mathrm{d}t \\ &= \mathrm{D}_1[\varphi_\omega](x')[q] + \mathrm{D}_1[\varphi_\omega](x')[(\mathbf{H}_{x'}^{\frac{1}{2}}\mathbf{H}_x^{-\frac{1}{2}} - \mathrm{Id})q] + \int \mathrm{D}_2[\varphi_\omega](p(t))[\mathbf{H}_{p(t)}^{\frac{1}{2}}\mathbf{H}_x^{-\frac{1}{2}} q, \mathbf{H}_{p(t)}^{\frac{1}{2}} p'(t)]\mathrm{d}t\end{aligned} \tag{C.6}$$

Therefore,

$$\|\mathrm{D}_1[\varphi_\omega](x) - \mathrm{D}_1[\varphi_\omega](x')\| \leqslant \bar{L}_1 C_{\mathbf{H}} d_{\mathbf{H}}(x,x') + \bar{L}_2(1 + C_{\mathbf{H}} r_{\text{near}}) d_{\mathbf{H}}(x,x').$$

Finally, for all $q_1, q_2 \in \mathbb{R}^d$, by Taylor's theorem

$$\begin{aligned}
&\mathrm{D}_2\left[\varphi_\omega\right](x)[q_1, q_2] - \mathrm{D}_2\left[\varphi_\omega\right](x')[q_1, q_2] \\
&= \nabla^2\varphi_\omega(x)[\mathbf{H}_x^{-\frac{1}{2}} q_1, \mathbf{H}_x^{-\frac{1}{2}} q_2] - \nabla^2\varphi_\omega(x')[\mathbf{H}_{x'}^{-\frac{1}{2}} q_1, \mathbf{H}_{x'}^{-\frac{1}{2}} q_2] \\
&= \mathrm{D}_2\left[\varphi_\omega\right](x')[\mathbf{H}_{x'}^{\frac{1}{2}}\mathbf{H}_x^{-\frac{1}{2}} q_1, (\mathbf{H}_{x'}^{\frac{1}{2}}\mathbf{H}_x^{-\frac{1}{2}} - \mathrm{Id}) q_2] + \mathrm{D}_2\left[\varphi_\omega\right](x')[(\mathbf{H}_{x'}^{\frac{1}{2}}\mathbf{H}_x^{-\frac{1}{2}} - \mathrm{Id}) q_1, q_2] \\
&\qquad + \int \mathrm{D}_3\left[\varphi_\omega\right](p(t))[\mathbf{H}_{p(t)}^{\frac{1}{2}}\mathbf{H}_x^{-\frac{1}{2}} q_1, \mathbf{H}_{p(t)}^{\frac{1}{2}}\mathbf{H}_x^{-\frac{1}{2}} q_2, \mathbf{H}_{p(t)}^{\frac{1}{2}} p'(t)]\mathrm{d}t.
\end{aligned} \tag{C.7}$$

Therefore,

$$\|\mathrm{D}_2\left[\varphi_\omega\right](x) - \mathrm{D}_2\left[\varphi_\omega\right](x')\| \leqslant \left(\bar{L}_2\left((1 + C_{\mathbf{H}} r_{\text{near}})C_{\mathbf{H}} + 1\right) + \bar{L}_3(1 + C_{\mathbf{H}} r_{\text{near}})^2\right) d_{\mathbf{H}}(x, x').$$

By applying these Lipschitz bounds, we obtain

$$\begin{aligned}
&\sup_{\|q\|=1} \left\|\mathrm{D}_r\left[q^\top(\hat{\mathbf{f}}_X - \hat{\mathbf{f}}_{X'})\right](y)\right\|^2 \\
&\quad\leqslant \sum_{j=1}^s \left\|\hat{K}^{(0r)}(x_j, y) - \hat{K}^{(0r)}(x'_j, y)\right\|^2 + \sum_{j=1}^s \left\|\hat{K}^{(1r)}(x_j, y) - \hat{K}^{(1r)}(x'_j, y)\right\|^2 \\
&\quad\leqslant \sum_{j=1}^s \mathcal{L}_0^2 \bar{L}_r^2 d_{\mathbf{H}}(x_j, x'_j)^2 + \sum_{j=1}^s \mathcal{L}_1^2 \bar{L}_r^2 d_{\mathbf{H}}(x_j, x'_j)^2 \\
&\quad= \left(\mathcal{L}_0^2 + \mathcal{L}_1^2\right) \bar{L}_r^2 d_{\mathbf{H}}(X, X')^2
\end{aligned}$$

□

Lemma C.3 (Local Lipschitz constant of $\hat{K}^{(ij)}$)**.** *Let $x_1, x_0 \in \mathcal{X}$. Let $i, j \in \{0, 1, 2\}$ with $i + j \leqslant 3$. Define*

$$A_{ij} = \sup_x \left\|\hat{K}^{(ij)}(x, x_0)\right\|$$

where x ranges over $d_{\mathbf{H}}(x, x_1) \leqslant r_{\text{near}}$. Then, for all x such that $d_{\mathbf{H}}(x, x_1) \leqslant r_{\text{near}}$,

$$\begin{aligned}
\left\|\hat{K}^{(0j)}(x, x_0) - \hat{K}^{(0j)}(x_1, x_0)\right\| &\leqslant A_{1j} d_{\mathbf{H}}(x, x_1) \\
\left\|\hat{K}^{(1j)}(x, x_0) - \hat{K}^{(1j)}(x_1, x_0)\right\| &\leqslant \left(C_{\mathbf{H}} A_{1j} + (1 + C_{\mathbf{H}} r_{\text{near}}) A_{2j}\right) d_{\mathbf{H}}(x, x_1)
\end{aligned}$$

The same results hold if we replace $\hat{K}$ by K.

Proof. The Lipschitz bounds on $\hat{K}^{ij}$ follow by combining

$$\begin{aligned}
&[q_1, \ldots, q_i](\hat{K}^{(ij)}(x, x_0) - \hat{K}^{(ij)}(x_1, x_0))[v_1, \ldots, v_j] \\
&= \hat{\mathbb{E}}\mathrm{Re}\left(\overline{(\mathrm{D}_i\left[\varphi_\omega\right](x) - \mathrm{D}_i\left[\varphi_\omega\right](x_1))[q_1, \ldots, q_i]}\mathrm{D}_j\left[\varphi_j\right](x_0)[v_1, \ldots, v_j]\right)
\end{aligned}$$

where $\hat{\mathbb{E}}$ indicates either empirical expectation or true expectation with (C.5), (C.6) and (C.7).

□

C.3 Probability bounds

In the proof of our main results, we will often assume that event $\bar{E}$ (see (A.4)) holds since our assumptions in Section 2.3 of the main paper imply that $\mathbb{P}(\bar{E}^c) \leqslant \rho/m$. The following lemma shows that our assumptions also imply that $\mathbb{E}_\omega[L_i(\omega)^2 1_{E_\omega^c}] \leqslant \frac{\varepsilon}{m}$. and this is a condition which our proofs will often rely upon.

Lemma C.4. *The following holds.* $\mathbb{P}(E^c_\omega) \leqslant \sum_i F_i(\bar{L}_i)$ *and*

$$\mathbb{E}_\omega[L_j(\omega)^2 1_{E^c_\omega}] \leqslant 2\int_{\bar{L}_j}^{\infty} tF_j(t)\mathrm{d}t + \bar{L}_j^2 \sum_i F_i(\bar{L}_i)$$

Proof. Let $E_{\omega,j}$ be the event that $L_r(\omega) \leqslant \bar{L}_j$, so $E_\omega = \cap_{j=0}^3 E_{\omega,j}$. By the union bound, $\mathbb{P}(E^c_\omega) \leqslant \sum_j \mathbb{P}(E^c_{\omega,j}) \leqslant \sum_i F_i(\bar{L}_i)$.

For the second claim, observe that $E^c_\omega = \cup_i E^c_{\omega,i}$ so that $\mathbb{E}[L_j(\omega)^2 1_{E^c_\omega}] \leqslant \sum_i \mathbb{E}[L_j(\omega)^2 1_{E^c_{\omega,i}}]$ and we have

$$\begin{aligned}
\mathbb{E}[L_j(\omega)^2 1_{E^c_{\omega,i}}] &= \int_0^\infty \mathbb{P}(L_j(\omega)^2 1_{E^c_{\omega,i}} \geqslant t)\mathrm{d}t \\
&= \int_0^\infty \mathbb{P}\left((L_j(\omega)^2 \geqslant t) \cap (L_i(\omega) \geqslant \bar{L}_i)\right)\mathrm{d}t \\
&\leqslant \bar{L}_j^2 F_i(\bar{L}_i) + \int_{\bar{L}_j^2}^\infty F_j(\sqrt{t})\mathrm{d}t = \bar{L}_j^2 F_i(\bar{L}_i) + 2\int_{\bar{L}_j}^\infty tF_j(t)\mathrm{d}t
\end{aligned}$$

where we have bounded $\mathbb{P}\left((L_j(\omega)^2 \geqslant t) \cap (L_i(\omega) \geqslant \bar{L}_i)\right)$ by respectively $\mathbb{P}(L_i(\omega) \geqslant \bar{L}_i) \leqslant F_i(\bar{L}_i)$ in the first term and by $\mathbb{P}(L_j(\omega)^2 \geqslant t) \leqslant F_j(\sqrt{t})$ in the second term. $\square$

C.3.1 Concentration inequalities

The following result is an adaption of the Matrix Bernstein inequality for dealing with conditional probabilities.

Lemma C.5 (Adapted unbounded Matrix Bernstein)**.** *Let* $A_j \in \mathbb{R}^{d_1\times d_2}$ *be a family of iid matrices for* $j = 1,\ldots,m$*. Let* $Z = \frac{1}{m}\sum_{j=1}^m A_j$ *and let* $\bar{Z} = \mathbb{E}[Z]$*. Let* $t \in (0, 4\|\mathbb{E}[A_1]\|]$*. Let events* E_j *be independent events such that* $E_j \subseteq \{\|A_j\| \leqslant L\}$ *and let* $E = \cap_j E_j$*. Suppose that we have*

$$\mathbb{P}(E_j^c) \leqslant \frac{t}{t + 4\|\mathbb{E}[A_1]\|} \quad \textit{and} \quad \mathbb{E}[\|A_j\| \mathbb{1}_{E_j^c}] \leqslant \frac{t}{4}$$

Then a first consequence is that we have $\mathbb{E}_E[Z] = \mathbb{E}_{E_j}[A_j]$ *for all* j *and* $\|\mathbb{E}[Z] - \mathbb{E}_E[Z]\| \leqslant \frac{t}{2}$*.*

Finally, assuming that

$$\sigma^2 \stackrel{\text{def.}}{=} \max_j\{\left\|\mathbb{E}_{E_j}[A_jA_j^*]\right\|, \left\|\mathbb{E}_{E_j}[A_j^*A_j]\right\|\} < \infty$$

we have

$$\mathbb{P}_E\left(\|Z - \mathbb{E}[Z]\| \geqslant t\right) \leqslant (d_1 + d_2)\exp\left(-\frac{mt^2/4}{\sigma^2 + Lt/3}\right).$$

Proof. We first bound $\|\mathbb{E}[Z] - \mathbb{E}_E[Z]\|$. First observe that $\mathbb{E}[Z] = \mathbb{E}_{E_1}[A_1]$ and $\mathbb{E}_E Z = \mathbb{E}_{E_1}[A_1]$ since A_j are iid. Moreover,

$$\mathbb{E}[A_1] = \mathbb{E}[A_1\mathbb{1}_{E_1}] + \mathbb{E}[A_1\mathbb{1}_{E_1^c}] = \mathbb{E}[A_1|E_1]\mathbb{P}(E_1) + \mathbb{E}[A_1\mathbb{1}_{E_1^c}].$$

Hence,

$$\begin{aligned}
&\|\mathbb{E}[A_1] - \mathbb{E}_{E_1}[A_1]\| = \left\|(P(E_1) - 1)\mathbb{E}_{E_1}[A_1] + \mathbb{E}[A_1\mathbb{1}_{E_1^c}]\right\| \\
&\leqslant \mathbb{P}(E_1^c)\|\mathbb{E}[A_1]\| + P(E_1^c)\|\mathbb{E}[A_1] - \mathbb{E}_{E_1}[A_1]\| + \mathbb{E}[\|A_1\|\mathbb{1}_{E_1^c}].
\end{aligned}$$

Therefore,

$$\|\mathbb{E}[A_1] - \mathbb{E}_{E_1}[A_1]\| \leqslant \frac{P(E_1^c)\|\mathbb{E}[A_1]\| + \mathbb{E}[\|A_1\|\mathbb{1}_{E_1^c}]}{1 - \mathbb{P}(E_1^c)} \leqslant \frac{t}{2}$$

For the second statement,

$$\mathbb{P}_E(\|Z-\mathbb{E}[Z]\| \geqslant t) \leqslant \mathbb{P}_E(\|Z-\mathbb{E}_E[Z]\| \geqslant t - \|\mathbb{E}[Z]-\mathbb{E}_E[Z]\|)$$
$$\leqslant \mathbb{P}_E(\|Z-\mathbb{E}_E[Z]\| \geqslant t/2).$$

To conclude, we apply Bernstein's inequality (Lemma G.2) to $Y_j = A_j - \mathbb{E}[A_j|E] = Y_j = A_j - \mathbb{E}[A_j|E_j]$ conditional to E. Observe that

$$0 \preceq \mathbb{E}_E[Y_jY_j^\top] \preceq \mathbb{E}_E[A_jA_j^\top] - \mathbb{E}_E[A_j]\mathbb{E}_E[A_j]^\top] \preceq \mathbb{E}_E[A_jA_j^\top],$$

which yields $\left\|\mathbb{E}_E[Y_jY_j^\top]\right\| \leqslant \left\|\mathbb{E}[A_jA_j^\top]\right\|$ and similarly, $\left\|\mathbb{E}_E[Y_j^\top Y_j]\right\| \leqslant \left\|\mathbb{E}_E[A_j^\top A_j]\right\|$. So by Bernstein's inequality

$$\mathbb{P}_E(\|Z-\mathbb{E}_E[Z]\| \geqslant t/2) \leqslant 2(d_1+d_2)\exp\left(-\frac{mt^2/4}{\sigma^2+Lt/3}\right).$$

□

Corollary C.1. *Let $x, x' \in \mathcal{X}$. If*

$$\mathbb{P}(E_\omega^c) \leqslant \frac{t}{t+4\left\|K^{(ij)}(x,x')\right\|} \quad \textit{and} \quad \mathbb{E}[L_{ij}(\omega)\mathbb{1}_{E_\omega^c}] \leqslant \frac{t}{4}$$

then $\left\|K_{\bar{E}}^{(ij)}(x,x') - K^{(ij)}(x,x')\right\| \leqslant t/2$.

Proposition C.1. *Let $t > 0$ and assume that*

$$\mathbb{P}(E_\omega^c) \leqslant \frac{t}{t+6} \quad \textit{and} \quad \mathbb{E}[L_{01}(\omega)^2\mathbb{1}_{E_\omega^c}] \leqslant \frac{t}{4s}$$

then $\|\Upsilon - \Upsilon_{\bar{E}}\| \leqslant t/2$ *and*

$$\mathbb{P}_{\bar{E}}(\left\|\Upsilon - \hat{\Upsilon}\right\| \geqslant t) \leqslant 4(d+1)s\exp\left(-\frac{mt^2/4}{s\bar{L}_{01}^2(3+t/3)}\right)$$

Consequently,

$$\mathbb{P}_{\bar{E}}(\left\|\Upsilon^{-1} - \hat{\Upsilon}^{-1}\right\| \geqslant t) \leqslant 4(d+1)s\exp\left(-\frac{mt^2}{16s\bar{L}_{01}^2(3+2\tilde{t})}\right).$$

Proof. We apply Lemma C.5 to $A_j = \gamma(\omega_j)\gamma(\omega_j)^*$ with the following observations:

- for each ω,
$$\|\gamma(\omega)\gamma(\omega)^*\| \leqslant \|\gamma(\omega)\|^2 \leqslant s\max_{x\in\mathcal{X}}\{\|\mathrm{D}_1[\varphi_\omega](x)\|^2 + |\varphi_\omega(x)|^2\},$$
so under event $\bar{E}$, $\|A_j\| \leqslant s\bar{L}_{01}^2$.
- By Lemma C.1, $\|\mathbb{E}[A_j]\| = \|\Upsilon\| \leqslant 3/2$,
- We may set $\sigma^2 = \bar{L}_{01}(3/2+t/2)$ since
$$0 \preceq \mathbb{E}_{\bar{E}}[A_1A_1^*] = \mathbb{E}_{\bar{E}}[A_1^*A_1] = \mathbb{E}_{\bar{E}}[\|\gamma(\omega_j)\|^2\gamma(\omega_j)\gamma(\omega_j)^*] \preceq \bar{L}_{01}(\|\mathbb{E}[A_j]\|+t/2)\mathrm{Id}.$$

The last claim is because $\left\|\Upsilon-\hat{\Upsilon}\right\| \leqslant t$ implies that $\|\Upsilon\| \leqslant 3/2+t$, $\left\|\Upsilon^{-1}\right\| \leqslant \frac{\|\Upsilon\|}{1-\left\|\Upsilon-\hat{\Upsilon}\right\|\|\Upsilon^{-1}\|} \leqslant \frac{3}{2-4t}$ and $\left\|\Upsilon^{-1}-\hat{\Upsilon}^{-1}\right\| \leqslant \left\|\Upsilon^{-1}\right\|\left\|\Upsilon-\hat{\Upsilon}\right\|\left\|\hat{\Upsilon}^{-1}\right\| \leqslant \frac{3t}{1-2t}$ and writing $\tilde{t} = \frac{3t}{1-2t}$ is equivalent to $t = \tilde{t}/(3+2\tilde{t})$. □

Bounds on $\hat{\mathbf{f}}_X$ applied to a fixed vector

Proposition C.2. *Let $t \in (0,1)$, $r \in \{0,2\}$, $q \in \mathbb{C}^{s(d+1)}$ and $y \in \mathcal{X}_r$, where $\mathcal{X}_0 \overset{\text{def.}}{=} \mathcal{X}$ and $\mathcal{X}_2 \overset{\text{def.}}{=} \mathcal{X}^{\text{near}}$. If*

$$\mathbb{P}(E_\omega^c) \leqslant \frac{t}{t+4B_r} \quad \textit{and} \quad \mathbb{E}[L_{01}(\omega)L_r(\omega)\mathbb{1}_{E_\omega^c}] \leqslant \frac{t}{4\sqrt{s}}$$

then

$$\mathbb{P}_{\bar{E}}\left(\left\|\mathrm{D}_r\left[(\hat{\mathbf{f}}_{X_0} - \mathbf{f}_{X_0})^\top q\right](y)\right\| \geqslant t\left\|q\right\|\right) \leqslant 2\tilde{d}\exp\left(\frac{-mt^2/4}{2\bar{L}_r^2 + \bar{L}_r\bar{L}_{01}t/(3\sqrt{s})}\right)$$

where $\tilde{d} = 1$ if $r = 0$ and $\tilde{d} = d$ if $r = 2$.

As a consequence, since $\sqrt{2s}\left\|q\right\|_{\text{Block}} \geqslant \left\|q\right\|_2$, we have

$$\mathbb{P}_E\left(\left\|\mathrm{D}_r\left[(\mathbf{f}_{X_0} - \hat{\mathbf{f}}_{X_0})^\top q\right](y)\right\| \geqslant t\left\|q\right\|_{\text{Block}}\right) \leqslant 2\tilde{d}\exp\left(\frac{-mt^2}{16s(\bar{L}_r^2 + 8\bar{L}_r\bar{L}_{01}t/(3\sqrt{2}))}\right)$$

provided that

$$\mathbb{P}(E_\omega^c) \leqslant \frac{t}{t+4\sqrt{2s}B_r} \quad \textit{and} \quad \mathbb{E}[L_{01}(\omega)L_r(\omega)\mathbb{1}_{E_\omega^c}] \leqslant \frac{t}{4\sqrt{2}s}.$$

Proof. Without loss of generality, assume that $\left\|q\right\| = 1$. First note that

$$\mathrm{D}_r\left[(\hat{\mathbf{f}}_{X_0} - \mathbf{f}_{X_0})^\top q\right](y) = \frac{1}{m}\sum_{k=1}^m q^\top\gamma(\omega_k)\mathrm{D}_r\left[\varphi_{\omega_k}\right](y) - \mathbb{E}[q^\top\gamma(\omega_k)\mathrm{D}_r\left[\varphi_{\omega_k}\right](y)].$$

We first consider the case of $r = 0$. We apply Lemma C.5 to $A_k \overset{\text{def.}}{=} q^\top\gamma(\omega_k)\varphi_{\omega_k}(y) \in \mathbb{C}$: Note that $|A_k| \leqslant \sqrt{s}L_{01}(\omega_k)L_0(\omega_k)$ and $|\mathbb{E}[A_k]| \leqslant B_0$.

- Under event E_{ω_k}, $|A_k| \leqslant \bar{L}_2\bar{L}_{01}\sqrt{s} \overset{\text{def.}}{=} L$.
- $\mathbb{E}_{\bar{E}}\left|A_k\right|^2 = \mathbb{E}_{\bar{E}}[\langle\gamma(\omega_k)\gamma(\omega_k)^*q,\, q\rangle\left|\varphi_{\omega_k}(y)\right|^2] \leqslant \bar{L}_0^2\left\|\Upsilon_{\bar{E}}\right\| \leqslant (3/2 + t/2)\,\bar{L}_0^2 \leqslant 2\bar{L}_0^2 \overset{\text{def.}}{=} \sigma^2$.

For the case $r = 2$, we apply Lemma C.5 with $A_k \overset{\text{def.}}{=} q^\top\gamma(\omega_k)\mathrm{D}_2\left[\varphi_{\omega_k}\right](y) \in \mathbb{C}^{d\times d}$. Then, $\left\|A_k\right\| \leqslant \sqrt{s}L_{01}(\omega_k)L_2(\omega_k)$, $\left\|\mathbb{E}[A_k]\right\| \leqslant B_2$, under event E_{ω_k}, $\left\|A_k\right\| \leqslant \bar{L}_2\bar{L}_{01}\sqrt{s} \overset{\text{def.}}{=} L$ and

$$\left\|\mathbb{E}_{\bar{E}}[A_kA_k^*]\right\| = \left\|\mathbb{E}_{\bar{E}}[A_k^*A_k]\right\| = \left\|\mathbb{E}_{\bar{E}}[\mathrm{D}_2\left[\varphi_{\omega_k}\right](y)^*\mathrm{D}_2\left[\varphi_{\omega_k}\right](y)\left|q^\top\gamma(\omega_k)\right|^2]\right\| \leqslant \bar{L}_2^2\mathbb{E}_{\bar{E}}[\left|q^\top\gamma(\omega_k)\right|^2] \leqslant 2\bar{L}_2^2 \overset{\text{def.}}{=} \sigma^2.$$

□

Lemma C.6. *Assume that*

$$\mathbb{P}(E_\omega^c) \leqslant \frac{t}{t+6\sqrt{2s}} \quad \textit{and} \quad \mathbb{E}[L_{01}(\omega)^2\mathbb{1}_{\bar{E}^c}] \leqslant \frac{t}{4\sqrt{2}s^{3/2}}$$

Let $q \in \mathbb{C}^{s(d+1)}$. Then, for all $t \geqslant \frac{2\sqrt{2}s\bar{L}_{01}\bar{L}_1}{m} + \sqrt{\frac{8s^2\bar{L}_{01}^2\bar{L}_1^2}{m^2} + \frac{144s\bar{L}_1^2}{m}}$, we have for each $x_i \in X$,

$$\mathbb{P}_E\left(\left\|\mathrm{D}_1\left[q^\top(\mathbf{f}_X - \hat{\mathbf{f}}_X)\right](x_i)\right\|_2 > 2t\left\|q\right\|_{\text{Block}}\right) \leqslant 28\exp\left(-\frac{mt^2/(4s)}{2\bar{L}_1^2 + \sqrt{2}t\bar{L}_1\bar{L}_{01}/3}\right).$$

Proof. For each $x_i \in X$,

$$\left\|\mathrm{D}_1\left[(\mathbb{E}_{\bar{E}}[q^\top\hat{\mathbf{f}}_X] - q^\top\mathbf{f}_X)\right](x_i)\right\| \leqslant \left\|\Upsilon - \Upsilon_{\bar{E}}\right\|\left\|q\right\| \leqslant \frac{t}{\sqrt{2s}}\left\|q\right\|,$$

by Proposition C.1. For convenience, we drop the subscript X from $\mathbf{f}_X$. Fix $i \in \{1, \ldots, s\}$. Observe that

$$\begin{aligned}
\mathbb{P}_E\left(\left\|\mathrm{D}_1\left[q^\top(\mathbf{f}-\hat{\mathbf{f}})\right](x_i)\right\|_2 > 2t\,\|q\|_{\text{Block}}\right) &\leqslant \mathbb{P}_E\left(\left\|\mathrm{D}_1\left[q^\top(\mathbf{f}-\hat{\mathbf{f}})\right](x_i)\right\|_2 > \frac{2t}{\sqrt{2s}}\,\|q\|_2\right)\\
&\leqslant \mathbb{P}_E\left(\left\|\mathrm{D}_1\left[q^\top(\mathbb{E}_{\bar{E}}[\hat{\mathbf{f}}]-\hat{\mathbf{f}})\right](x_i)\right\|_2 > \frac{t}{\sqrt{2s}}\,\|q\|_2\right)
\end{aligned}$$

The claim of this lemma follows by applying Lemma G.3: Let

$$Y_k = \mathrm{D}_1\left[\varphi_{\omega_k}\right](x_i)\gamma(\omega_k)^\top q - \mathbb{E}_{\bar{E}}\mathrm{D}_1\left[\varphi_{\omega_k}\right](x_i)\gamma(\omega)^\top q \in \mathbb{C}^d,$$

and observe that $\mathrm{D}_1\left[q^\top(\hat{\mathbf{f}}-\mathbb{E}_{\bar{E}}[\hat{\mathbf{f}}])\right](x_i) = \frac{1}{m}\sum_k Y_k$. Without loss of generality, assume that $\|q\|_2 = 1$. We apply Lemma G.3. Observe that conditional on event E,

- $\|Y_k\|_2 \leqslant 2\,\|q\|_2\,\|\gamma(\omega_k)\|_2\,\|\mathrm{D}_1\left[\varphi_{\omega_k}\right](x_i)\|_2 \leqslant 2\sqrt{s}\bar{L}_{01}\bar{L}_1$.
- $\mathbb{E}_E\,\|Y_k\|^2 \leqslant \mathbb{E}_E[\left|\gamma(\omega_k)^\top q\right|^2 \mathrm{D}_1\left[\varphi_{\omega_k}\right](x_i)\mathrm{D}_1\left[\varphi_{\omega_k}\right](x_i)^\top] \leqslant \bar{L}_1^2\,\|\Upsilon_E\|$. So, $\sigma^2 \leqslant m\bar{L}_1^2\,\|\Upsilon_E\| \leqslant m\bar{L}_1^2(t+\|\Upsilon\|) \leqslant m\bar{L}_1^2(t/2+3/2) \leqslant 2m\bar{L}_1^2$ (here we are talking about the σ^2 in Lemma G.3).

Therefore, for all

$$t \geqslant \frac{2\sqrt{2}s\bar{L}_{01}\bar{L}_1}{m} + \sqrt{\frac{8s^2\bar{L}_{01}^2\bar{L}_1^2}{m^2} + \frac{144s\bar{L}_1^2}{m}}$$

$$\mathbb{P}\left(\left\|\frac{1}{m}\sum_{k=1}^m Y_k\right\|_2 \geqslant \frac{t}{\sqrt{2s}}\right) \leqslant 28\exp\left(-\frac{mt^2/(4s)}{2\bar{L}_1^2 + \sqrt{2}t\bar{L}_1\bar{L}_{01}/3}\right)$$

□

Proposition C.3 (Block norm bound on $\hat{\Upsilon}$ applied to a fixed vector)**.** *Suppose that*

$$\mathbb{P}(E_\omega^c) \leqslant \frac{t}{t+6\sqrt{s}(B_0+1)} \quad \textit{and} \quad \mathbb{E}[L_{01}(\omega)^2\mathbb{1}_{\bar{E}^c}] \leqslant \frac{t}{4s^{3/2}(1+4B_0)}$$

Then, for all

$$t \geqslant \left(\frac{4\sqrt{2}s\bar{L}_{01}\bar{L}_1}{m} + \sqrt{\frac{32s^2\bar{L}_{01}^2\bar{L}_1^2}{m^2} + \frac{576s\bar{L}_1^2}{m}}\right)$$

we have

$$\mathbb{P}_E\left(\left\|(\Upsilon-\hat{\Upsilon})q\right\|_{\text{Block}} \geqslant t\,\|q\|_{\text{Block}}\right) \leqslant 32s\exp\left(-\frac{mt^2}{s\left(32\bar{L}_1^2 + 34t\bar{L}_1\bar{L}_{01}\right)}\right). \tag{C.8}$$

Proof. Let $S_0 \overset{\text{def.}}{=} \{1, \ldots, s\}$ and $S_j \overset{\text{def.}}{=} \{s+(j-1)d+1, \ldots, s+jd\}$ for $j = 1, \ldots, s$. Observe that by the union bound

$$\begin{aligned}
&\mathbb{P}_E\left(\left\|(\Upsilon-\hat{\Upsilon})q\right\|_{\text{Block}} \geqslant t\,\|q\|_{\text{Block}}\right)\\
&\leqslant \mathbb{P}_E\left(\left\|((\Upsilon-\hat{\Upsilon})q)_{S_0}\right\|_\infty \geqslant t\,\|q\|_{\text{Block}}\right) + \sum_{j=1}^s \mathbb{P}_E\left(\left\|((\Upsilon-\hat{\Upsilon})q)_{S_j}\right\|_2 \geqslant t\,\|q\|_{\text{Block}}\right)\\
&\leqslant \sum_{j=1}^s \mathbb{P}_E\left(\left|((\Upsilon-\hat{\Upsilon})q)_j\right| \geqslant t\,\|q\|_{\text{Block}}\right) + \sum_{j=1}^s \mathbb{P}_E\left(\left\|((\Upsilon-\hat{\Upsilon})q)_{S_j}\right\|_2 \geqslant t\,\|q\|_{\text{Block}}\right).
\end{aligned} \tag{C.9}$$

To bound the first sum, observe that $((\Upsilon-\hat{\Upsilon})q)_j = (\mathbf{f}(x_j)-\hat{\mathbf{f}}(x_j))^\top q$ and $((\Upsilon-\hat{\Upsilon})q)_{S_j} = \mathrm{D}_1\left[q^\top(\mathbf{f}-\hat{\mathbf{f}})\right](x_j)$. So, the first sum can be bounded by applying Proposition C.2. The second sum can be bounded by applying Lemma C.6.

□

Norm bounds for $\hat{\mathbf{f}}$ We will repeatedly make use of the following result on $\hat{\mathbf{f}}_X$. This result is due to concentration bounds on the kernel $\hat{K}$ which are derived subsequently.

Proposition C.4 (Bound on $\hat{\mathbf{f}}_X$)**.** *Let $X \in \mathcal{X}^s$. Let $\rho > 0$. Assume that for all $(i,j) \in \{(0,0),(1,0),(0,2),(1,2)\}$,*

$$\mathbb{P}(E_\omega^c) \leqslant \frac{t}{t + 4\sqrt{s}\max\{B_0, B_2\}}, \quad \mathbb{E}[L_i(\omega)L_j(\omega)\mathbb{1}_{E_\omega^c}] \leqslant \frac{t}{4\sqrt{s}}$$

Then, given any $y \in \mathcal{X}$,

$$\mathbb{P}_{\bar{E}}\left(\left\|\hat{\mathbf{f}}_X(y) - \mathbf{f}_X(y)\right\| \geqslant t\right) \leqslant 4sd\exp\left(-\frac{mt^2/8}{3s\bar{L}_{01}^2}\right). \tag{C.10}$$

and given any $y \in \mathcal{X}^{\text{near}}$, writing $\hat{\mathbf{f}}_X = (\hat{f}_j)_{j=1}^p$ and $\mathbf{f}_X = (f_j)_{j=1}^p$ with $p = s(d+1)$, we have

$$\mathbb{P}_{\bar{E}}\left(\sup_{\|q\|=1}\sqrt{\sum_{j=1}^p \left\|\mathrm{D}_2\left[\hat{f}_j - f_j\right](y)q\right\|^2} > t\right) \leqslant s(3d+d^2)\exp\left(-\frac{mt^2/8}{s(\bar{L}_2^2 B_{11} + \bar{L}_1^2 B_{22} + \bar{L}_{01}\bar{L}_2)}\right). \tag{C.11}$$

Proof. Let $i, j \in \mathbb{N}_0$ with $i + j \leqslant 2$. Let $[s] \stackrel{\text{def.}}{=} \{1,\ldots,s\}$ and $I \stackrel{\text{def.}}{=} \{(0,0),(1,0)\}$, By Lemma C.7 and the union bound,

$$\mathbb{P}_{\bar{E}}\left(\exists (i,j) \in I, \exists \ell \in [s], \left\|\hat{K}^{(ij)}(x_\ell, y) - K^{(ij)}(x_\ell, y)\right\| \geqslant \frac{t}{\sqrt{s}}\right) \leqslant 4sd\exp\left(-\frac{mt^2/4}{3s\bar{L}_{01}^2}\right). \tag{C.12}$$

So, (C.10) follows because

$$\left\|\hat{\mathbf{f}}_X(y) - \mathbf{f}_X(y)\right\| \leqslant \sqrt{\sum_{i=1}^s \left|\hat{K}(x_i,y) - K(x_i,y)\right|^2 + \left\|\hat{K}^{(10)}(x_i,y) - K^{(10)}(x_i,y)\right\|^2} \leqslant \sqrt{2}t.$$

By Lemma C.7, Lemma C.9 and the union bound, letting $I_2 \stackrel{\text{def.}}{=} \{(0,2),(1,2)\}$, we have

$$\begin{aligned}\mathbb{P}_{\bar{E}}\left(\exists (i,j) \in I_2, \exists \ell \in [s], \left\|\hat{K}^{(ij)}(x_\ell, y) - K^{(ij)}(x_\ell, y)\right\| \geqslant \frac{t}{\sqrt{s}}\right) &\leqslant 2sd\exp\left(-\frac{mt^2/4}{2s(\bar{L}_2^2 + \bar{L}_0\bar{L}_2)}\right) \\ &+ s(d+d^2)\exp\left(-\frac{mt^2/4}{s(\bar{L}_2^2 B_{11} + \bar{L}_1^2 B_{22} + \bar{L}_1\bar{L}_2)}\right).\end{aligned} \tag{C.13}$$

and (C.11) follows since given $q \in \mathbb{C}^d$, $\|q\| = 1$, we have

$$\sum_{j=1}^p \left\|\mathrm{D}_2\left[\hat{f}_j - f_j\right](y)q\right\|^2 \leqslant \sum_{j=1}^s \left(\left\|\hat{K}^{(02)}(x_j,y) - K^{(02)}(x_j,y)\right\|^2 + \left\|\hat{K}^{(12)}(x_j,y) - K^{(12)}(x_j,y)\right\|^2\right) \leqslant 2t^2$$

□

Lemma C.7 (Concentration on kernel)**.** *Let $t > 0$, $x, x' \in \mathcal{X}$. Let $i, j \in \mathbb{N}_0$ with $i + j \leqslant 2$. Assume*

$$\mathbb{P}(E_\omega^c) \leqslant \frac{t}{t + 4\left\|K^{(ij)}(x,x')\right\|}, \quad \mathbb{E}[L_i(\omega)L_j(\omega)\mathbb{1}_{E_\omega^c}] \leqslant \frac{t}{4}$$

then

$$\mathbb{P}_{\bar{E}}\left(\left\|\hat{K}^{(ij)}(x,x') - K^{(ij)}(x,x')\right\| \geqslant t\right) \leqslant 2d\exp\left(-\frac{mt^2}{\bar{L}_p^2(b_{ij}+1) + \bar{L}_i\bar{L}_j t/3}\right)$$

where $p = \max(i,j)$ and $b_{ij} = 1$ if $\min(i,j) = 0$ and $b_{ij} \stackrel{\text{def.}}{=} \left\|K^{(11)}(x,x')\right\|$ otherwise.

Proof. It is an immediate application of Lemma C.5 with $A_k = \mathrm{Re}\left(\overline{\mathrm{D}_i\left[\varphi_{\omega_k}\right](x)}\mathrm{D}_j\left[\varphi_{\omega_k}\right](x')^\top\right)$ for $k = 1, \ldots, m$. Note that $A_k \in (\mathbb{R}^d)^{i+j}$ if $(i,j) \in \{(0,0),(0,1),(1,0)\}$ and $A_k \in \mathbb{R}^{d\times d}$ if $\max(i,j) = 2$. noting that under $\bar{E}$, $\|A_k\| \leqslant \bar{L}_i \bar{L}_j$. Next, we need to bound $\|\mathbb{E}_{\bar{E}}[A_k A_k^*]\|$ and $\|\mathbb{E}_{\bar{E}}[A_k^* A_k]\|$. We present only the argument for $(i,j) = (0,2)$, since all the other cases are similar:

$$
\begin{aligned}
0 \preceq \mathbb{E}_{\bar{E}} A_k A_k^* &\preceq \mathbb{E}_{\bar{E}}[\|\varphi_{\omega_k}(x')\|^2 \mathrm{D}_2\left[\varphi_{\omega_k}\right](x)\mathrm{D}_2\left[\varphi_{\omega}\right](x)^*] \\
&\preceq \bar{L}_2^2 \mathbb{E}_{\bar{E}} \|\varphi_{\omega_k}(x')\|^2 \,\mathrm{Id} = \bar{L}_2^2 \left|K_{\bar{E}}(x',x')\right| \mathrm{Id} \preceq (1+t/2)\bar{L}_2^2 \mathrm{Id}
\end{aligned}
$$

so $\|\mathbb{E}_{\bar{E}} A_k A_k^*\| \leqslant (1+t/2)\bar{L}_2^2$. Similarly, $\|\mathbb{E}_{\bar{E}} A_k^* A_k\| \leqslant (1+t/2)\bar{L}_2^2$ and

$$
\|\mathbb{E}_{\bar{E}} A_k^* A_k\|, \|\mathbb{E}_{\bar{E}} A_k A_k^*\| \leqslant L_p^2 (B_{qq} + t/2)
$$

where $p = \max(i,j)$ and $q = \min(i,j)$.

□

Applying a grid on $\mathcal{X}^{\text{near}}$, we get a uniform version.

Lemma C.8. *Let $i, j \in \mathbb{N}_0$ with $i + j \leqslant 2$, and assume that*

$$
\mathbb{P}(E_\omega^c) \leqslant \frac{t}{t + 16 B_{ij}}, \quad \mathbb{E}[L_i(\omega) L_j(\omega) \mathbb{1}_{E_\omega^c}] \leqslant \frac{t}{16}.
$$

Then

$$
\begin{aligned}
&\mathbb{P}_{\bar{E}}\left(\exists\, x, x' \in \mathcal{X}^{\text{near}},\ \left\|\hat{K}^{(ij)}(x,x') - K^{(ij)}(x,x')\right\| \geqslant t\right) \\
&\qquad \leqslant 2ds^2 \exp\left(-\frac{mt^2/16}{L_p^2(B_{qq}+1) + \bar{L}_i \bar{L}_j t/12} + 2d \log\left(\frac{4(\mathcal{L}_i \bar{L}_j + \bar{L}_i \mathcal{L}_j)}{t}\right)\right).
\end{aligned}
$$

where $p = \max(i,j)$ and $q = \min(i,j)$ and $\mathcal{L}_i, \mathcal{L}_j$ are as in Lemma C.2

Proof. We define a δ-covering of $\mathcal{X}^{\text{near}}$ for the metric $d_{\mathbf{H}}$ with $\delta = \min\left(r_{\text{near}}, \frac{t}{4(\mathcal{L}_i \bar{L}_j + \bar{L}_i \mathcal{L}_j)}\right)$ of size $s\left(\frac{r_{\text{near}}}{\delta}\right)^d$. Let this covering be denoted by $\mathcal{X}^{\text{grid}}$.

By the union bound and Lemma C.7,

$$
\mathbb{P}_{\bar{E}}\left(\exists x, x' \in \mathcal{X}^{\text{grid}} \text{ s.t. } \left\|\hat{K}^{(ij)}(x,x') - K^{(ij)}(x,x')\right\| \geqslant t/4\right) \leqslant 2ds^2 \left(\frac{r_{\text{near}}}{\delta}\right)^{2d} \exp\left(-\frac{mt^2/16}{L_p^2(B_{qq}+1) + \bar{L}_i \bar{L}_j t/12}\right)
$$

where $p = \max(i,j)$ and $q = \min(i,j)$. This gives the required upper bound: Given any $x, x' \in \mathcal{X}$, let $x_{\text{grid}}, x'_{\text{grid}} \in \mathcal{X}^{\text{grid}}$ be such that $d_{\mathbf{H}}(x, x_{\text{grid}}), d_{\mathbf{H}}(x', x'_{\text{grid}}) \leqslant \delta$. Then, under event $\bar{E}$, by Lemma C.2,

$$
\left\|\hat{K}^{(ij)}(x,x') - \hat{K}^{(ij)}(x_{\text{grid}}, x'_{\text{grid}})\right\| \leqslant (\mathcal{L}_i \bar{L}_j + \bar{L}_i \mathcal{L}_j)\delta \leqslant t/4.
$$

By Jensen's inequality and since $\left\|K_{\bar{E}}^{(ij)}(x,x') - K^{(ij)}(x,x')\right\| \leqslant t/4$ for all x, x', we have

$$
\left\|K^{(ij)}(x,x') - K^{(ij)}(x_{\text{grid}}, x'_{\text{grid}})\right\| \leqslant t/2.
$$

□

We now derive analogous results for the kernel differentiated 3 times.

Lemma C.9 (Concentration on order 3 kernel)**.** *Let $x, x' \in \mathcal{X}^{\mathrm{near}}$. Assume that*

$$\mathbb{P}(E_\omega^c) \leqslant \frac{t}{t + 4\max\{B_{12}, B_{22}\}}, \quad \mathbb{E}[(L_1(\omega)L_2(\omega) + L_2^2(\omega))\mathbb{1}_{E_\omega^c}] \leqslant \frac{t}{4}$$

For $j = 1, \ldots, m$, let $a_i = (\mathrm{D}_1\left[\overline{\varphi_{\omega_j}}\right](x))_i \in \mathbb{C}$, $D \stackrel{\text{def.}}{=} \mathrm{D}_2\left[\varphi_\omega\right](x') \in \mathbb{C}^{d\times d}$ and

$$A_j \stackrel{\text{def.}}{=} \begin{pmatrix} a_1 D & a_2 D & \cdots & a_d D \end{pmatrix}^\top \in \mathbb{C}^{d^2\times d} \tag{C.14}$$

Let $Z \stackrel{\text{def.}}{=} \frac{1}{m}\sum_{j=1}^m (A_j - \mathbb{E}[A_j])$. Then, given

$$\begin{aligned} g(x') \stackrel{\text{def.}}{=} (g_i(x'))_{i=1}^d &\stackrel{\text{def.}}{=} \sum_{k=1}^m \left(\overline{\mathrm{D}_1\left[\varphi_{\omega_k}\right](x)}\varphi_\omega(x') - \mathbb{E}[\overline{\mathrm{D}_1\left[\varphi_{\omega_k}\right](x)}\varphi_\omega(x')]\right) \\ &= \hat{K}^{(10)}(x, x') - K^{(10)}(x, x'), \end{aligned}$$

(i) $\sup_{q\in\mathbb{C}^d, \|q\|\leqslant 1} \sum_{i=1}^d \left\|\mathrm{D}_2\left[g_i\right](x')q\right\|^2 = \|Z\|^2$,

(ii) $\sup_{q\in\mathbb{C}^d, \|q\|\leqslant 1} \left\|\mathrm{D}_2\left[q^\top g\right](x')\right\| = \left\|\hat{K}^{(12)}(x, x') - K^{(12)}(x, x')\right\| \leqslant \|Z\|$.

and

$$\mathbb{P}_{\bar{E}}\left(\|Z\| \geqslant t\right) \leqslant (d + d^2)\exp\left(-\frac{mt^2/4}{\tilde{B} + \bar{L}_1\bar{L}_2 t/3}\right)$$

where $\tilde{B} \stackrel{\text{def.}}{=} \max\{\bar{L}_2^2(B_{11} + t/2), \bar{L}_1^2(B_{22} + t/2)\}$.

Proof. The claim (i) is simply by definition, since $Zq = (\mathrm{D}_2\left[g_i\right](x')q)_{i=1}^d \in \mathbb{C}^{d^2}$. For (ii), the first equality is simply be definition, and for the inequality, observe that

$$\begin{aligned} \sup_{q\in\mathbb{C}^d, \|q\|\leqslant 1} \left\|\mathrm{D}_2\left[q^\top g\right](x')\right\| &= \sup_{q\in\mathbb{C}^d, \|q\|\leqslant 1}\ \sup_{p\in\mathbb{C}^d, \|p\|\leqslant 1} \left\|\sum_{i=1}^d q_i \mathrm{D}_2\left[g_i\right](x')p\right\| \\ &\leqslant \sup_{q\in\mathbb{C}^d, \|q\|\leqslant 1}\ \sup_{p\in\mathbb{C}^d, \|p\|\leqslant 1} \|q\| \sqrt{\sum_{i=1}^d \left\|\mathrm{D}_2\left[g_i\right](x')p\right\|^2} \leqslant \|Z\|. \end{aligned}$$

Finally, the probability bound follows by applying Lemma C.5: First note that under $\bar{E}$, $\|A_j\| \leqslant \bar{L}_1\bar{L}_2$. It remains to bound $\left\|\mathbb{E}_{\bar{E}}[A_j^* A_j]\right\|$ and $\left\|\mathbb{E}_{\bar{E}}[A_j A_j^*]\right\|$:

$$\begin{aligned} \sup_{\|q\|\leqslant 1} \mathbb{E}_{\bar{E}}\langle A_j^* A_j q,\, q\rangle &= \sup_{\|q\|\leqslant 1} \mathbb{E}_E \sum_{i=1}^d \left|(\mathrm{D}_1\left[\varphi_{\omega_j}\right](x))_i\right|^2 \left\|\mathrm{D}_2\left[\varphi_\omega\right](x')q\right\|^2 \\ &\leqslant \sup_{\|q_k\|\leqslant 1} \bar{L}_1^2 \mathbb{E}_{\bar{E}} \overline{\mathrm{D}_2\left[\varphi_\omega\right](x')[q_1, q_2]}\mathrm{D}_2\left[\varphi_\omega\right](x')[q_3, q_4] \\ &\leqslant \bar{L}_1^2 \left\|K_{\bar{E}}^{(22)}(x, x)\right\| \leqslant \bar{L}_1^2(B_{22} + t/2). \end{aligned}$$

Given $p_i \in \mathbb{C}^d$ for $i = 1, \ldots, d$ such that $\sum_i \|p_i\|^2 \leqslant 1$, write $P = \begin{pmatrix} p_1 & p_2 & \cdots p_d \end{pmatrix} \in \mathbb{C}^{d\times d}$ and $\bar{p} =$

$\begin{pmatrix} p_1^\top & p_2^\top & \cdots p_d^\top \end{pmatrix}^\top \in \mathbb{C}^{d^2}$. Then,

$$\begin{aligned}
\mathbb{E}_E \langle A_j A_j^* \bar{p},\, \bar{p} \rangle &= \mathbb{E}_E \left\| \sum_{i=1}^d (\mathrm{D}_1 \left[\varphi_{\omega_j}\right](x))_i \mathrm{D}_2 \left[\varphi_{\omega_j}\right](x') p_i \right\|^2 \\
&= \mathbb{E}_E \left\| \mathrm{D}_2 \left[\varphi_{\omega_j}\right](x') P \mathrm{D}_1 \left[\varphi_{\omega_j}\right](x) \right\|^2 \\
&\leqslant \bar{L}_2^2 \mathbb{E}_E \sum_i \left| \sum_k p_{i,k} (\mathrm{D}_1 \left[\varphi_{\omega_j}\right](x))_k \right|^2 \\
&= \bar{L}_2^2 \sum_i \langle \hat{K}_{\bar{E}}^{(11)}(x,x) p_i,\, p_i \rangle \leqslant \bar{L}_2^2 \left\| \hat{K}_{\bar{E}}^{(11)}(x,x) \right\|^2 \sum_i \|p_i\|^2 \leqslant \bar{L}_2^2 (B_{11} + t/2).
\end{aligned}$$

□

Lemma C.10 (Uniform concentration on order 3 kernel)**.** *Assume*

$$\mathbb{P}(E_\omega^c) \leqslant \frac{t}{t + 16 \max\{B_{12}, B_{22}\}}, \quad \mathbb{E}[L_1(\omega) L_2(\omega) \mathbb{1}_{E_\omega^c}] \leqslant \frac{t}{16}$$

then

$$\begin{aligned}
&\mathbb{P}_{\bar{E}} \left(\exists x, x' \in \mathcal{X}^{\text{near}},\ \left\| \hat{K}^{(12)}(x,x') - K^{(12)}(x,x') \right\| \geqslant t \right) \\
&\qquad \leqslant s^2 (d + d^2) \exp \left(- \frac{m t^2 / 16}{\tilde{B} + \bar{L}_1 \bar{L}_2 t / 6} + 2d \log \left(\frac{8(\mathcal{L}_1 \bar{L}_2 + \bar{L}_2 \mathcal{L}_2)}{t} \right) \right)
\end{aligned}$$

where $\tilde{B} \stackrel{\text{def.}}{=} \max\{\bar{L}_2^2 (B_{11} + t/2), \bar{L}_1^2 (B_{22} + t/2)\}$*,* $\mathcal{L}_1$*,* $\mathcal{L}_2$ *are as in Lemma C.2.*

Proof. Let $\mathcal{X}^{\text{grid}}$ be a δ-covering of $\mathcal{X}^{\text{near}}$ for the metric $d_{\mathbf{H}}$ with $\delta = \min \left(r_{\text{near}}, \frac{t}{8(\mathcal{L}_1 \bar{L}_2 + \mathcal{L}_2 \bar{L}_2)} \right)$ of size at most $s \left(\frac{8(\mathcal{L}_1 \bar{L}_2 + \mathcal{L}_2 \bar{L}_2)}{t} \right)^d$. By Lemma C.9 and the union bound,

$$\begin{aligned}
&\mathbb{P}_{\bar{E}} \left(\exists x, x' \in \mathcal{X}^{\text{grid}},\ \left\| \hat{K}^{(ij)}(x,x') - K^{(ij)}(x,x') \right\| \geqslant t/2 \right) \\
&\qquad \leqslant s^2 (d + d^2) \left(\frac{8(\bar{L}_1 \bar{L}_2 + \bar{L}_2^2)}{t} \right)^{2d} \exp \left(- \frac{m t^2 / 16}{\bar{L}_2^2 (B_{11} + t/4) + \bar{L}_1 \bar{L}_2 t / 6} \right) \stackrel{\text{def.}}{=} \rho.
\end{aligned}$$

Moreover, under event $\bar{E}$, given any $x, x' \in \mathcal{X}^{\text{near}}$, there exists grid points x_{grid}, x'_{grid} such that

$$d_{\mathbf{H}}(x, x_{\text{grid}}), d_{\mathbf{H}}(x', x'_{\text{grid}}) \leqslant \delta$$

and

$$\begin{aligned}
\left\| \left(\hat{K}^{(12)}(x,x') - K^{(12)}(x,x') \right) \right\| \leqslant &\left\| \left(\hat{K}^{(12)}(x_{\text{grid}}, x'_{\text{grid}}) - K^{(12)}(x_{\text{grid}}, x'_{\text{grid}}) \right) \right\| \\
&+ \left\| \left(\hat{K}^{(12)}(x,x') - \hat{K}^{(12)}(x_{\text{grid}}, x'_{\text{grid}}) \right) \right\| \\
&+ \left\| \left(K^{(12)}(x,x') - K^{(12)}(x_{\text{grid}}, x'_{\text{grid}}) \right) \right\|,
\end{aligned}$$

and by Lemma C.2, under event $\bar{E}$,

$$\left\| \left(\hat{K}^{(12)}(x,x') - \hat{K}^{(12)}(x_{\text{grid}}, x'_{\text{grid}}) \right) \right\| \leqslant (\mathcal{L}_1 \bar{L}_2 + \mathcal{L}_2 \bar{L}_2) \delta \leqslant t/8.$$

and by Jensen's inequality and since $\left\|K^{(12)}(x,y) - K^{(12)}_{\bar{E}}(x,y)\right\| \leqslant t/8$,

$$\left\|\left(K^{(12)}(x,y) - K^{(12)}(x_{\text{grid}},y)\right)\right\| \leqslant 3t/8.$$

Therefore, conditional on $\bar{E}$, $\left\|\left(\hat{K}^{(12)}(x,y) - K^{(12)}(x,y)\right)\right\| < t$ with probability at least $1-\rho$.

□

D Proof of Theorem 3

In all the rest of the proofs we fix $X_0 \in \mathcal{X}^s$ to be Δ-separated points, $a_0 \in \mathbb{C}^s$, and let $\mathbf{u} = (\text{sign}(a_0), 0_{sd})$. We denote $\mathcal{X}_i^{\text{near}} = \{x \in \mathcal{X}\ ;\ d_{\mathbf{H}}(x, x_{0,i}) \leqslant r_{\text{near}}\}$ and $\mathcal{X}^{\text{near}} = \cup_i \mathcal{X}_i^{\text{near}}$ and $\mathcal{X}^{\text{far}} = \mathcal{X} \backslash \mathcal{X}^{\text{near}}$.

Since K is an admissible kernel, from (B.2) and (B.1) in the proof of Theorem 2, η_{X_0,a_0} satisfies

(i) for all $y \in \mathcal{X}^{\text{far}}$, $|\eta_{X_0,a_0}(y)| \leqslant 1 - \frac{1}{2}\varepsilon_0$,

(ii) for all $y \in \mathcal{X}^{\text{near}}(i)$, $-\text{Re}\left(\text{sign}(a_i)\text{D}_2\left[\eta_{X_0,a_0}\right](y)\right) \succcurlyeq \frac{1}{2}\varepsilon_2\text{Id}$ and $\left\|\text{Im}\left(\text{sign}(a_i)\text{D}_2\left[\eta_{X_0,a_0}\right](y)\right)\right\| \leqslant (\frac{p}{2})\frac{1}{2}\varepsilon_2$.

$$p \stackrel{\text{def.}}{=} \sqrt{(1-\varepsilon_2 r_{\text{near}}^2/2)/(\varepsilon_2 r_{\text{near}}^2/2)} \geqslant 1,$$

since $\varepsilon_2 r_{\text{near}}^2 \leqslant 1$ by assumption of K being admissible. We aim to show that, for X close to X_0, $\hat{\eta}_X$ is nondegenerate by showing that $\left\|\text{D}_r\left[\hat{\eta}_X\right] - \text{D}_r\left[\eta_{X_0,a_0}\right]\right\| \leqslant c\varepsilon_r$ for some positive constant c sufficiently small.

D.1 Nondegeneracy of $\hat{\eta}_{X_0,a_0}$

We first establish the nondegeneracy of $\hat{\eta}_{X_0,a_0}$, our proof can be seen as a generalisation of the techniques in [9] to the multidimensional setting with general sampling operators:

Theorem D.1. *Let $\rho > 0$ and assume that the assumptions in Section 2.3 hold. Assume also that either (a) or (b) holds:*

(a) $\text{sign}(a_0)$ *is a Steinhaus sequence and*

$$m \gtrsim C \cdot s \cdot \log\left(\frac{N^d}{\rho}\right)\log\left(\frac{s}{\rho}\right)$$

(b) $\text{sign}(a_0)$ *is an arbitrary sequence from the complex unit circle, and*

$$m \gtrsim C \cdot s^{3/2} \cdot \log\left(\frac{N^d}{\rho}\right)$$

where C, N are defined in the main paper. Then with probability at least $1-\rho$, the following hold: For all $y \in \mathcal{X}^{far}$, $|\hat{\eta}_{X_0,a_0}(y)| \leqslant 1 - \frac{7}{16}\varepsilon_0$, and for all $y \in \mathcal{X}^{near}(i)$, $-\text{Re}\left(\text{sign}(a_i)\text{D}_2\left[\hat{\eta}_{X_0,a_0}\right](y)\right) \succcurlyeq \frac{7}{16}\varepsilon_2\text{Id}$ and $\left\|\text{Im}\left(\text{sign}(a_i)\text{D}_2\left[\hat{\eta}_{X_0,a_0}\right](y)\right)\right\| \leqslant (\frac{p}{2}+\frac{p}{8})\frac{1}{2}\varepsilon_2$ and hence, $\hat{\eta}_{X_0,a_0}$ is $(\frac{7}{16}\varepsilon_0, \frac{7}{16}\varepsilon_2)$-nondegenerate.

Proof. Note that

$$\frac{8}{7}\left(\frac{p}{2}+\frac{p}{8}\right) = \frac{5}{7}p < \sqrt{\frac{1-7\varepsilon_2 r_{\text{near}}^2/16}{7\varepsilon_2 r_{\text{near}}^2/16}}$$

so $\hat{\eta}_{X_0,a_0}$ is $(\frac{7}{16}\varepsilon_0, \frac{7}{16}\varepsilon_2)$-nondegenerate by Lemma B.1

Let $c \stackrel{\text{def.}}{=} 1/32$. Observe that by assumption and Lemma C.4, $\mathbb{P}(\bar{E}) \leqslant \rho/2$. Therefore, it is sufficient to prove that conditional on $\bar{E}$, with probability at least $1-\delta$ with $\delta \stackrel{\text{def.}}{=} \rho/2$, $\hat{\eta}_{X_0,a_0}$ is nondegenerate.

We will repeatedly use the fact that our assumptions (by Lemma C.4) also imply that

$$\mathbb{P}(E_\omega^c) \leqslant \frac{\varepsilon}{m}, \quad \mathbb{E}[L_i(\omega)L_j(\omega)\mathbb{1}_{E_\omega^c}] \leqslant \frac{\varepsilon}{m}$$

for all $(i,j) \in \{(0,0),(1,0),(0,2),(1,2)\}$,

Step I: Proving nondegeneracy on a finite grid.

Let $\mathcal{X}_{\text{grid}}^{\text{far}} \subset \mathcal{X}^{\text{far}}$ and $\mathcal{X}_{\text{grid}}^{\text{far}} \subset \mathcal{X}^{\text{near}}$, be finite point sets. Let

$$Q_r(y) \stackrel{\text{def.}}{=} \left\| \mathrm{D}_r\left[\hat{\eta}_{X_0,a_0}\right](y) - \mathrm{D}_r\left[\eta_{X_0,a_0}\right](y) \right\|, \qquad r = 0, 2.$$

We first prove that conditional on $\bar{E}$, with probability at least $1-\delta$ where $\delta \stackrel{\text{def.}}{=} \rho/2$, that $Q_0(y) \leqslant c\varepsilon_0$ for all $y \in \mathcal{X}_{\text{grid}}^{\text{far}}$ and $Q_2(y) \leqslant c\varepsilon_2$ for all $y \in \mathcal{X}_{\text{grid}}^{\text{far}}$.

Let us first recall some facts which were proven in the previous section: Let $a, t \in (0,1)$ and write $\mathbf{f} = (\bar{f}_j)_{j=1}^{s(d+1)}$ and $\hat{\mathbf{f}} = (f_j)_{j=1}^{s(d+1)}$. Let $q_0 \stackrel{\text{def.}}{=} \Upsilon^{-1}\mathbf{u}$, so $\|q_0\| \leqslant 2\sqrt{s}$. Let F be the event that

(a) $\left\| \Upsilon^{-1} - \hat{\Upsilon}^{-1} \right\| \leqslant t$,

(b) $\forall y \in \mathcal{X}_{\text{grid}}^{\text{far}}, \left\| \hat{\mathbf{f}}_{X_0}(y) - \mathbf{f}_{X_0}(y) \right\| \leqslant a\varepsilon_0$,

(c) $\forall y \in \mathcal{X}_{\text{grid}}^{\text{near}}, \sup_{q \in \mathbb{C}^d,\ \|q\|=1} \sqrt{\sum_{j=1}^p \left\| \mathrm{D}_2\left[f_j - \bar{f}_j\right](y) q \right\|^2} \leqslant a\varepsilon_2$,

Let G be the event that

(d) $\forall y \in \mathcal{X}_{\text{grid}}^{\text{far}}, \left| (\hat{\mathbf{f}}_{X_0}(y) - \mathbf{f}_{X_0}(y))^\top q_0 \right| \leqslant 2a\varepsilon_0$

(e) $\forall y \in \mathcal{X}_{\text{grid}}^{\text{near}}, \left\| \mathrm{D}_2\left[(\hat{\mathbf{f}}_{X_0} - \mathbf{f}_{X_0})^\top q_0 \right](y) \right\| \leqslant 2a\varepsilon_2$

then provided that

$$\mathbb{P}(E_\omega^c) \leqslant \frac{u}{u + \max\{4\sqrt{s}B_{ij}, 6\}}, \quad \mathbb{E}[L_i(\omega)L_j(\omega)\mathbb{1}_{E_\omega^c}] \leqslant \frac{u}{4s} \tag{D.1}$$

where $u = \min\{a\varepsilon_i, t\}$, we have

$$\begin{aligned}
\mathbb{P}_{\bar{E}}(F^c) \leqslant & 4(d+1)s \exp\left(-\frac{mt^2}{16s\bar{L}_{01}^2(3+2t)} \right) \\
& + 4sd \left| \mathcal{X}_{\text{grid}}^{\text{far}} \right| \exp\left(-\frac{m(a\varepsilon_0)^2/8}{s(\bar{L}_{01}^2(B_{11}+1) + \bar{L}_{01}^2)} \right) \\
& + s(3d+d^2) \left| \mathcal{X}_{\text{grid}}^{\text{near}} \right| \exp\left(-\frac{m(a\varepsilon_2)^2/8}{s(\bar{L}_2^2 B_{11} + \bar{L}_1^2 B_{22}) + \bar{L}_{01}\bar{L}_2)} \right) \\
\mathbb{P}_{\bar{E}}(G^c) \leqslant & 2 \left| \mathcal{X}_{\text{grid}}^{\text{far}} \right| \exp\left(-\frac{ma^2\varepsilon_0^2}{s(8\bar{L}_0^2 + \frac{4}{3}\bar{L}_0\bar{L}_{01}a\varepsilon_0)} \right) \\
& + 2d \left| \mathcal{X}_{\text{grid}}^{\text{near}} \right| \exp\left(-\frac{ma^2\varepsilon_2^2}{s(8\bar{L}_2^2 + \frac{4}{3}\bar{L}_2\bar{L}_{01}a\varepsilon_2)} \right),
\end{aligned} \tag{D.2}$$

where for $\mathbb{P}_{\bar{E}}(F^c)$, the first term on the right is due to Proposition C.1, the second and third are due to Proposition C.4 while the bound on $\mathbb{P}_{\bar{E}}(G^c)$ is due to Proposition C.2 (noting that, since this probability bound over the ω_j is valid for all fixed $\mathbf{u}$, and the ω_j and the signs are independent, it is valid with the same probability over both ω_j and $\mathbf{u}$).

Observe that

$$\begin{aligned}\left\|\mathrm{D}_j\left[\hat{\eta}_{X_0,a_0}\right](y)-\mathrm{D}_j\left[\eta_{X_0,a_0}\right](y)\right\| &= \left\|\mathrm{D}_j\left[(\hat{\alpha}_{X_0}-\alpha_{X_0})^\top\hat{\mathbf{f}}_{X_0}\right](y)+\mathrm{D}_j\left[\alpha_{X_0}^\top(\hat{\mathbf{f}}_{X_0}-\mathbf{f}_{X_0})\right](y)\right\| \\ &\leqslant \left\|\mathrm{D}_j\left[\mathbf{u}^\top\left((\hat{\Upsilon}^{-1}-\Upsilon^{-1})\hat{\mathbf{f}}_{X_0}+\Upsilon^{-1}(\hat{\mathbf{f}}_{X_0}-\mathbf{f}_{X_0})\right)\right](y)\right\|\end{aligned} \tag{D.3}$$

Step I (a): Random signs

We first bound (D.3) in the case where $\mathbf{u}$ is a Steinhaus sequence.

Let $\beta_1(y) \overset{\text{def.}}{=} (\hat{\Upsilon}^{-1}-\Upsilon^{-1})\hat{\mathbf{f}}_{X_0}(y)$ and $\beta_2(y) \overset{\text{def.}}{=} \Upsilon^{-1}(\hat{\mathbf{f}}_{X_0}(y)-\mathbf{f}_{X_0}(y))$. Then, event F implies that $\|\beta_1(y)\| \leqslant t(B_0+a\varepsilon_0)$ for all $y\in\mathcal{X}_{\text{grid}}^{\text{far}}$, and event G implies that $\left|\mathbf{u}^\top\beta_2(y)\right|\leqslant 2a\varepsilon_0$. So,

$$\begin{aligned}&\mathbb{P}_{\bar{E}}\left(\left|\exists y\in\mathcal{X}_{\text{grid}}^{\text{far}},\ \mathbf{u}^\top(\beta_1+\beta_2)(y)\right|>c\varepsilon_0\right)\\ &\leqslant \mathbb{P}_{F\cap\bar{E}}\left(\exists y\in\mathcal{X}_{\text{grid}}^{\text{far}},\ \left|\mathbf{u}^\top\beta_1(y)\right|>\frac{c}{2}\varepsilon_0\right)\mathbb{P}_{\bar{E}}(F)+\mathbb{P}_{\bar{E}}\left(F^c\right)\\ &\quad+\mathbb{P}_{G\cap\bar{E}}\left(\exists y\in\mathcal{X}_{\text{grid}}^{\text{far}},\ \left|\mathbf{u}^\top\beta_2(y)\right|>\frac{c}{2}\varepsilon_0\right)\mathbb{P}_{\bar{E}}(G)+\mathbb{P}_{\bar{E}}\left(G^c\right)\\ &\leqslant \mathbb{P}_{F\cap\bar{E}}\left(\exists y\in\mathcal{X}_{\text{grid}}^{\text{far}},\ \left|\mathbf{u}^\top\beta_1\right|>\frac{c}{2}\varepsilon_0\right)+\mathbb{P}_{\bar{E}}\left(F^c\right)+\mathbb{P}_{\bar{E}}\left(G^c\right)\\ &\leqslant 4\left|\mathcal{X}_{\text{grid}}^{\text{far}}\right|e^{-\frac{(c/4)^2\varepsilon_0^2}{8t^2(B_0+a\varepsilon_0)^2}}+\mathbb{P}_{\bar{E}}(F^c)+\mathbb{P}_{\bar{E}}\left(G^c\right).\end{aligned}$$

where we set $a=c/4$ for the second inequality and the last inequality follows from Lemma G.4 and because $\mathbf{u}$ consists if random signs.

Now consider $Q_2(y)=\mathrm{D}_2\left[\mathbf{u}^\top\beta\right](y)$. Under event G, $\left\|\mathrm{D}_2\left[\mathbf{u}^\top\beta_2\right](y)\right\|\leqslant\frac{c}{2}\varepsilon_2$. Writing $M=(\hat{\Upsilon}^{-1}-\Upsilon^{-1})$, we have

$$\mathrm{D}_2\left[\mathbf{u}^\top\beta_1\right](y)=\mathrm{D}_2\left[\mathbf{u}^\top\left(M\hat{\mathbf{f}}_{X_0}\right)\right](y)=\sum_{\ell=1}^p\mathbf{u}_\ell\left(\sum_{j=1}^p M_{\ell j}\mathrm{D}_2\left[f_j\right](y)\right). \tag{D.4}$$

We aim to bound (D.4) by applying the Matrix Hoeffding's inequality (Corollary G.1): let

$$Y_\ell \overset{\text{def.}}{=} \mathrm{Re}\left(\sum_{j=1}^p M_{\ell j}\mathrm{D}_2\left[f_j\right](y)\right)\in\mathbb{R}^{d\times d}$$

which is a symmetric matrix. Note that

$$\left\|\sum_{\ell=1}^p Y_\ell^2\right\|=\sup_{q\in\mathbb{R}^d,\|q\|=1}\sum_{\ell=1}^p\langle Y_\ell^2 q,\,q\rangle=\sup_{q\in\mathbb{R}^d,\|q\|=1}\sum_{\ell=1}^d\|Y_\ell q\|^2\leqslant\sup_{q\in\mathbb{R}^d,\|q\|=1}\left\|\sum_{j=1}^p M_{\ell,j}(\mathrm{D}_2\left[f_j\right](y)q)\right\|^2.$$

Then, for a vector q of unit norm, let $V_{j,n}\overset{\text{def.}}{=}(\mathrm{D}_2\left[f_j\right](y)q)_n$ for $j=1,\dots,p$ and $n=1,\dots,d$, then

$$\begin{aligned}\sum_{\ell=1}^p\left\|\sum_{j=1}^p M_{\ell,j}(\mathrm{D}_2\left[f_j\right](y)q)\right\|^2 &=\sum_{\ell=1}^p\sum_{n=1}^d\left|\sum_{j=1}^p M_{\ell,j}V_{j,n}\right|^2=\sum_{n=1}^d\|MV_{\cdot,n}\|^2\leqslant\|M\|^2\sum_{n=1}^d\|V_{\cdot,n}\|^2\\ &=\|M\|^2\sum_{n=1}^d\sum_{j=1}^p|V_{j,n}|^2=\|M\|^2\sum_{j=1}^p\|\mathrm{D}_2\left[f_j\right](y)q\|^2.\end{aligned}$$

Under event F, we have $\|M\|^2\sum_{j=1}^p\|\mathrm{D}_2\left[f_j\right](y)q\|^2\leqslant t^2(B_2+a\varepsilon_2)^2$. Then,

$$\mathbb{P}_{F\cap\bar{E}}\left(\left\|\mathrm{D}_2\left[\mathbf{u}^\top\mathrm{Re}\left(M\hat{\mathbf{f}}_{X_0}\right)\right](y)\right\|\geqslant\frac{c\varepsilon_2}{\sqrt{2}}\right)\leqslant 2d\exp\left(-\frac{(c/2)^2\varepsilon_2^2}{4t^2(B_2+a\varepsilon_2)^2}\right).$$

By repeating this argument for the imaginary part, we obtain

$$\mathbb{P}_{F\cap\bar{E}}\left(\left\|\mathrm{D}_2\left[\mathbf{u}^\top \mathrm{Im}\left(M\hat{\mathbf{f}}_{X_0}\right)\right](y)\right\| \geqslant \frac{c\varepsilon_2}{\sqrt{2}}\right) \leqslant 2d\exp\left(-\frac{(c/2)^2\varepsilon_2^2}{4t^2(B_2+a\varepsilon_2)^2}\right).$$

So,

$$\begin{aligned}
&\mathbb{P}_{\bar{E}}\left(\exists y \in \mathcal{X}_{\text{grid}}^{\text{near}},\ \left\|\mathrm{D}_2\left[\mathbf{u}^\top\beta(y)\right]\right\| > c\varepsilon_2\right)\\
&\leqslant \mathbb{P}_{F\cap\bar{E}}\left(\exists y \in \mathcal{X}_{\text{grid}}^{\text{near}},\ \left\|\mathrm{D}_2\left[\mathbf{u}^\top \mathrm{Re}\left(M\hat{\mathbf{f}}_{X_0}\right)\right](y)\right\| \geqslant \frac{c}{2}\varepsilon_2\right) + \mathbb{P}_{\bar{E}}(F^c) + \mathbb{P}_{\bar{E}}(G^c)\\
&\leqslant 4d\left|\mathcal{X}_{\text{grid}}^{\text{near}}\right|\exp\left(-\frac{(c/2)^2\varepsilon_2^2}{4t^2(B_2+a\varepsilon_2)^2}\right) + \mathbb{P}_{\bar{E}}(F^c) + \mathbb{P}_{\bar{E}}(G^c).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&1-\mathbb{P}\left(Q_0(y_0)\leqslant c\varepsilon_0 \text{ and } Q_2(y_2)\leqslant c\varepsilon_2, \forall y_0\in\mathcal{X}_{\text{grid}}^{\text{far}}, \forall y_2 \in \mathcal{X}_{\text{grid}}^{\text{near}}\right)\\
&\leqslant 4\left|\mathcal{X}_{\text{grid}}^{\text{far}}\right|\exp\left(-\frac{(c/2)^2\varepsilon_0^2}{32t^2(B_0+a\varepsilon_0)^2}\right) + 4d\left|\mathcal{X}_{\text{grid}}^{\text{near}}\right|\exp\left(-\frac{(c/2)^2\varepsilon_2^2}{16t^2(B_2+a\varepsilon_2)^2}\right) + 2\mathbb{P}_{\bar{E}}(F^c) + 2\mathbb{P}_{\bar{E}}(G^c).
\end{aligned}$$

The first 2 terms are each bounded by $\delta/7$ by setting t such that

$$\frac{1}{t^2} = 2^{13}\log\left(\frac{112\bar{N}d}{\delta}\right)\frac{\left(\bar{B}+1\right)}{c^2\varepsilon^2}$$

where $\bar{B} \stackrel{\text{def.}}{=} \max\{B_0, B_2\}$, $\varepsilon \stackrel{\text{def.}}{=} \min\{\varepsilon_0,\varepsilon_2\}$ and $\bar{N} = \max\left(\left|\mathcal{X}_{\text{grid}}^{\text{near}}\right|, \left|\mathcal{X}_{\text{grid}}^{\text{far}}\right|\right)$. The first term of (D.2) is bounded by $\delta/7$ if

$$m \geqslant \frac{1}{t^2}\log\left(\frac{28(d+1)s}{\delta}\right)64s\bar{L}_{01}^2 = s\bar{L}_{01}^2\frac{2^{19}\left(\bar{B}+1\right)}{c^2\varepsilon^2}\log\left(\frac{112\bar{N}d}{\delta}\right)\log\left(\frac{28(d+1)s}{\delta}\right)$$

and the last 4 terms of (D.2) are each bounded by $\delta/7$ provided that

$$m \gtrsim \log\left(\frac{28(s+d)d\bar{N}}{\delta}\right)\frac{16s(\bar{L}_2^2B_{11}+\bar{L}_1^2B_{22}+\bar{L}_{01}\bar{L}_2)}{c^2\varepsilon^2}$$

So, to summarise, recalling that $\delta=\rho/2$, $\hat{\eta}_{X_0,a_0}$ is nondegenerate on $\mathcal{X}_{\text{grid}}^{\text{near}}$ and $\mathcal{X}_{\text{grid}}^{\text{far}}$ with probability at least $1-\delta$ (conditional on $\bar{E}$) provided that

$$m \gtrsim \log\left(\frac{sdN}{\rho}\right)\log\left(\frac{sd}{\rho}\right)\frac{s(\bar{L}_2^2B_{11}+\bar{L}_1^2B_{22}+\bar{B}\bar{L}_{01}^2+\bar{L}_{01}\bar{L}_2)}{\varepsilon^2}$$

and

$$\mathbb{P}(E_\omega^c) \lesssim \frac{\varepsilon}{\bar{B}^{3/2}\sqrt{s}\sqrt{\log(\bar{N}d/\rho)}} \quad \text{and} \quad , \quad \mathbb{E}[L_i(\omega)L_j(\omega)\mathbb{1}_{E_\omega^c}] \lesssim \frac{\varepsilon}{4s\sqrt{B}\sqrt{\log(\bar{N}d/\rho)}}$$

Step I (b): Deterministic signs Assume now that $\mathbf{u}$ consists of arbitrary signs. We will show that (D.3) can be bounded by $c\varepsilon$ when m is chosen as in condition (b) of this theorem. Let F' be the event that

(a') $\left\|\Upsilon-\hat{\Upsilon}\right\| \leqslant \frac{t}{s^{1/4}}$ and $\left\|\Upsilon^{-1}-\hat{\Upsilon}^{-1}\right\| \leqslant \frac{t}{s^{1/4}}$

(b') $\forall y \in \mathcal{X}_{\text{grid}}^{\text{far}}$, $\left\|(\hat{\mathbf{f}}_{X_0}(y)-\mathbf{f}_{X_0}(y))\right\| \leqslant \frac{a\varepsilon_0}{s^{1/4}}$

(c') $\forall y \in \mathcal{X}_{\text{grid}}^{\text{near}}$, $\sup_{\|q\|=1}\left\|\mathrm{D}_2\left[(\hat{\mathbf{f}}_{X_0}-\mathbf{f}_{X_0})^\top q\right](y)\right\| \leqslant \frac{a\varepsilon_2}{s^{1/4}}$

(f) $\left\|(\Upsilon - \hat{\Upsilon})\Upsilon^{-1}\mathbf{u}\right\|_{\text{Block}} \leqslant a\varepsilon \left\|\Upsilon^{-1}\mathbf{u}\right\|_{\text{Block}} \leqslant 2a\varepsilon.$

Then, provided that

$$\mathbb{P}(E_\omega^c) \leqslant \frac{u}{u + 6s(B_0 + B_2)} \quad \text{and} \quad \mathbb{E}[L_{01}(\omega)^2 \mathbb{1}_{\bar{E}^c}] \leqslant \frac{u}{4\bar{B}s^{3/2}},$$

with $u = \min\{a\varepsilon_i, t\}$ as before, we have

$$\begin{aligned}\mathbb{P}_{\bar{E}}((F')^c) \leqslant & 4(d+1)s \exp\left(-\frac{mt^2}{16s^{3/2}\bar{L}_{01}^2(3+2t)}\right) \\ & + 4sd\left|\mathcal{X}_{\text{grid}}^{\text{far}}\right| \exp\left(-\frac{m(a\varepsilon_0)^2/8}{s^{3/2}(\bar{L}_{01}^2(B_{11}+1)+\bar{L}_{01}^2)}\right) \\ & + s(3d+d^2)\left|\mathcal{X}_{\text{grid}}^{\text{near}}\right| \exp\left(-\frac{m(a\varepsilon_2)^2/8}{s^{3/2}(\bar{L}_2^2 B_{11} + \bar{L}_1^2 B_{22} + \bar{L}_{01}\bar{L}_2)}\right) \\ & + 32s\exp\left(-\frac{m4a^2\varepsilon^2}{s\left(32L_1^2 + 68a\varepsilon L_1 \bar{L}_{01}\right)}\right).\end{aligned}$$

where the first bound is from Proposition C.1, the second and third are from Proposition C.4 and the final bound is due to Proposition C.3.

To bound (D.3), we first observe that if event G holds, then just as observed previously, $\left|\mathrm{D}_r\left[\mathbf{u}^\top \beta_2\right](y)\right| \leqslant 2a\varepsilon_r$. To bound $\left|\mathbf{u}^\top \beta_1(y)\right|$, observe that

$$\begin{aligned}\mathbf{u}^\top \beta_1(y) &= \mathbf{u}^\top(\Upsilon^{-1} - \hat{\Upsilon}^{-1})(\hat{\mathbf{f}}_{X_0} - \mathbf{f}_{X_0}) + \mathbf{u}^\top(\Upsilon^{-1} - \hat{\Upsilon}^{-1})\mathbf{f}_{X_0} \\ &= \mathbf{u}^\top(\Upsilon^{-1} - \hat{\Upsilon}^{-1})(\hat{\mathbf{f}}_{X_0} - \mathbf{f}_{X_0}) + \mathbf{u}^\top\Upsilon^{-1}(\hat{\Upsilon} - \Upsilon)\hat{\Upsilon}^{-1}\mathbf{f}_{X_0} \\ &= \mathbf{u}^\top(\Upsilon^{-1} - \hat{\Upsilon}^{-1})(\hat{\mathbf{f}}_{X_0} - \mathbf{f}_{X_0}) + \mathbf{u}^\top\Upsilon^{-1}(\hat{\Upsilon} - \Upsilon)(\hat{\Upsilon}^{-1} - \Upsilon^{-1})\mathbf{f}_{X_0} + \mathbf{u}^\top\Upsilon^{-1}(\hat{\Upsilon} - \Upsilon)\Upsilon^{-1}\mathbf{f}_{X_0}\end{aligned}$$

Under event F',

- $\left|\mathbf{u}^\top(\Upsilon^{-1} - \hat{\Upsilon}^{-1})(\hat{\mathbf{f}}_{X_0} - \mathbf{f}_{X_0})\right| \leqslant \sqrt{s}\left\|\Upsilon^{-1} - \hat{\Upsilon}^{-1}\right\|\left\|\hat{\mathbf{f}}_{X_0} - \mathbf{f}_{X_0}\right\| \leqslant ta\varepsilon$
- $\left|\mathbf{u}^\top\Upsilon^{-1}(\hat{\Upsilon} - \Upsilon)(\hat{\Upsilon}^{-1} - \Upsilon^{-1})\mathbf{f}_{X_0}\right| \leqslant \sqrt{s}\cdot 2 \cdot \left\|\hat{\Upsilon} - \Upsilon\right\|\left\|\hat{\Upsilon}^{-1} - \Upsilon^{-1}\right\| B_0 \leqslant 2t^2 B_0$
- $\left\|\Upsilon^{-1}(\hat{\Upsilon} - \Upsilon)\Upsilon^{-1}\mathbf{u}\right\|_{\text{Block}} \leqslant \left\|\Upsilon^{-1}\right\|_{\text{Block}}\left\|(\hat{\Upsilon} - \Upsilon)\Upsilon^{-1}\mathbf{u}\right\|_{\text{Block}} \leqslant 4a\varepsilon.$

Finally, given any vector q such that $\|q\|_{\text{Block}} \leqslant 4a\varepsilon$, we have $\left|q^\top \mathbf{f}_{X_0}\right| \leqslant 4a\varepsilon B_0$. Therefore,

$$\left|\mathbf{u}^\top \beta_1(y)\right| \leqslant ta + 2t^2 + 4a\varepsilon B_0,$$

and in a similar manner, we can show that the same upper bound holds for $\left\|\mathrm{D}_2\left[\mathbf{u}^\top \beta_1\right](y)\right\|$.

Therefore,

$$\left\|\mathrm{D}_r\left[\mathbf{u}^\top \beta\right](y)\right\| \leqslant c\varepsilon_r \tag{D.5}$$

if both F' and G hold, so conditional on $\bar{E}$, (D.5) holds with probability at least $1-\delta$ provided that

$$m \gtrsim s^{3/2} \cdot \frac{(\bar{L}_2^2 B_{11} + \bar{L}_1^2 B_{22} + \bar{B}\bar{L}_{01}^2 + \bar{L}_{01}\bar{L}_2)}{\varepsilon^2} \cdot \log\left(\frac{\bar{N}ds}{\rho}\right)$$

and

$$\mathbb{P}(E_\omega^c) \lesssim \frac{\varepsilon}{\bar{B}^{3/2} s\sqrt{\log(\bar{N}d/\rho)}} \quad \text{and} \quad , \quad \mathbb{E}[L_i(\omega)L_j(\omega)\mathbb{1}_{E_\omega^c}] \lesssim \frac{\varepsilon}{s^{3/2}\sqrt{B}\sqrt{\log(\bar{N}d/\rho)}}$$

Step II: Extending to the entire space To prove that $\hat{\eta}_{X_0,a_0}$ is nondegenerate on the entire space $\mathcal{X}$, we first show that $\hat{\eta}_{X_0,a_0}$ is locally Lipschitz (and hence determine how fine our grids $\mathcal{X}^{\text{near}}_{\text{grid}}$, $\mathcal{X}^{\text{far}}_{\text{grid}}$ need to be): for $x, x' \in \mathcal{X}$ with $d_{\mathbf{H}}(x,x') \leqslant r_{\text{near}}$,

$$\begin{aligned}
\left\|\mathrm{D}_r\left[\hat{\eta}_{X_0,a_0}\right](x) - \mathrm{D}_r\left[\hat{\eta}_{X_0,a_0}\right](x')\right\| &= \Big\| \frac{1}{m}\sum_{k=1}^{m} \mathrm{D}_r\left[\mathrm{Re}\left((\hat{\Upsilon}_X^{-1}\mathbf{u})^\top \gamma(\omega_k)\varphi_{\omega_k}\right)\right](x) \qquad \text{(D.6)}\\
&\qquad - \mathrm{D}_r\left[\mathrm{Re}\left((\hat{\Upsilon}_X^{-1}\mathbf{u})^\top \gamma(\omega_k)\varphi_{\omega_k}\right)\right](x')\Big\| \\
&= \left\| \frac{1}{m}\sum_{j=1}^{m} \mathrm{Re}\left(\left((\hat{\Upsilon}_X^{-1}\mathbf{u})^\top \gamma(\omega_k)\right)\cdot\left(\mathrm{D}_r\left[\varphi_{\omega_k}\right](x) - \mathrm{D}_r\left[\varphi_{\omega_k}\right](x')\right)\right)\right\| \\
&\leqslant \left\|\hat{\Upsilon}_X^{-1}\right\| \|\mathbf{u}\| \sqrt{s}\bar{L}_{01}\left\|\mathrm{D}_r\left[\varphi_{\omega_k}\right](x) - \mathrm{D}_r\left[\varphi_{\omega_k}\right](x')\right\| \qquad \text{(D.7)}\\
&\leqslant 4s\bar{L}_{01} d_{\mathbf{H}}(x,x')\mathcal{L}_r \leqslant c\varepsilon_r. \qquad \text{(D.8)}
\end{aligned}$$

where we have applied Lemma C.2 to obtain the last line.

Choosing $\mathcal{X}^{\text{far}}_{\text{grid}}$ to be a $\delta_0 \stackrel{\text{def.}}{=} \frac{c\varepsilon_0}{4\mathcal{L}_0\bar{L}_{01}s}$-covering of $\mathcal{X}^{\text{near}}$ (of size at most $\mathcal{O}(R_{\mathcal{X}}/\delta_0)$), $\mathcal{X}^{\text{far}}_{\text{grid}}$ to be a $\delta_2 \stackrel{\text{def.}}{=} \frac{c\varepsilon_2}{4\mathcal{L}_2\bar{L}_{01}s}$-covering of $\mathcal{X}^{\text{far}}$ (of size at most $\mathcal{O}(R_{\mathcal{X}}/\delta_2)$). Then for any $x \in \mathcal{X}^{\text{near}}$ and $x' \in \mathcal{X}^{\text{near}}_{\text{grid}}$ such that $d_{\mathbf{H}}(x,x') \leqslant \delta_0$,

$$|\hat{\eta}_{X_0,a_0}(x)| \leqslant |\hat{\eta}_{X_0,a_0}(x')| + |\hat{\eta}_{X_0,a_0}(x) - \hat{\eta}_{X_0,a_0}(x')| \leqslant 1 - \varepsilon_0 + 2c\varepsilon_0.$$

and given any $x \in \mathcal{X}^{\text{far}}$, let $x' \in \mathcal{X}^{\text{far}}_{\text{grid}}$ be such that $d_{\mathbf{H}}(x,x') \leqslant \delta_2$, so

$$\begin{aligned}
\mathrm{Re}\left(\overline{\mathrm{sign}(a_{0,i})}\mathrm{D}_2\left[\hat{\eta}_{X_0,a_0}\right](x)\right) &\preceq \mathrm{Re}\left(\overline{\mathrm{sign}(a_{0,i})}\mathrm{D}_2\left[\hat{\eta}_{X_0,a_0}\right](x')\right) + \left\|\mathrm{D}_2\left[\hat{\eta}_{X_0,a_0}\right](x) - \mathrm{D}_2\left[\hat{\eta}_{X_0,a_0}\right](x')\right\|\mathrm{Id}\\
&\preceq (-\varepsilon_2 + 2c\varepsilon_2)\mathrm{Id},
\end{aligned}$$

and

$$\left\|\mathrm{Im}\left(\overline{\mathrm{sign}(a_{0,i})}\mathrm{D}_2\left[\hat{\eta}_{X_0,a_0}\right](x)\right)\right\| \leqslant \left\|\mathrm{Im}\left(\overline{\mathrm{sign}(a_{0,i})}\mathrm{D}_2\left[\hat{\eta}_{X_0,a_0}\right](x')\right)\right\| + c\varepsilon_2 \leqslant (c_2 + c)\varepsilon_2.$$

□

D.2 Nondegeneracy transfer to $\hat{\eta}_{X,a}$.

We are now ready to prove Theorem 3, which we restate below for clarity.

Theorem D.2. *Suppose that the assumptions of Theorem D.1 hold, and the following holds with probability at least $1-\rho$: for all X such that*

$$d_{\mathbf{H}}(X, X_0) \lesssim \min\left(r_{\text{near}}, \varepsilon_r(C_{\mathbf{H}}B\sqrt{s})^{-1}, \varepsilon_r(C_{\mathbf{H}}\bar{L}_{12}\bar{L}_r\sqrt{s})^{-1}\right), \qquad \text{(D.9)}$$

and $\|a - a_0\| \lesssim \frac{\varepsilon_r}{\max(B_r)}\min_i |a_{0,i}|$. Then, $\hat{\eta}_{X,a} \stackrel{\text{def.}}{=} \Phi^*\Gamma_X^{*,\dagger}\binom{\mathrm{sign}(a)}{0}$ satisfies

(i) for all $y \in \mathcal{X}^{far}$, $|\hat{\eta}_{X,a}(y)| \leqslant 1 - \frac{13}{32}\varepsilon_0$

(ii) for all $y \in \mathcal{X}^{near}(i)$, $-\mathrm{Re}\left(\overline{\mathrm{sign}(a_i)}\mathrm{D}_2\left[\hat{\eta}_{X,a}\right](y)\right) \succcurlyeq \frac{13\varepsilon_2}{32}\mathrm{Id}$ and $\left\|\mathrm{Im}\left(\overline{\mathrm{sign}(a_i)}\mathrm{D}_2\left[\hat{\eta}_{X,a}\right](y)\right)\right\| \leqslant (\frac{p}{2} + \frac{3p}{16})\frac{1}{2}\varepsilon_2$.

Hence, $\hat{\eta}_{X,a}$ is $(\frac{13}{32}\varepsilon_0, \frac{13}{32}\varepsilon_2)$-nondegenerate.

The proof essentially exploits the fact that $\hat{\Upsilon}_X$, $\hat{\mathbf{f}}_X$ are locally Lipschitz in X with respect to the metric $d_{\mathbf{H}}$, and consequently nondegeneracy of $\hat{\eta}_{X_0,a_0}$ implies nondegeneracy of $\hat{\eta}_{X,a}$ whenever $d_{\mathbf{H}}(X, X_0)$ and $\|a - a_0\|_2$ are sufficiently small.

D.2.1 Proof of Theorem D.2

We begin with a lemma which shows that $\hat{\Upsilon}_X$ is locally Lipschitz in X.

Lemma D.1 (Lipschitz bound of $\hat{\Upsilon}_X$)**.** *Let $X_0 \in \mathcal{X}^s$ be Δ-separated points. Assume that for all $i + j \leqslant 3$*

$$\mathbb{P}(E_\omega^c) \leqslant \frac{1}{1 + 16\sqrt{s}B_{ij}}, \quad \mathbb{E}[L_i(\omega)L_j(\omega)\mathbb{1}_{E_\omega^c}] \leqslant \frac{1}{16\sqrt{s}}$$

for all $i, j = 0, ..., 2$. Let $\rho > 0$ and

$$m \gtrsim s(\bar{L}_2^2 B_{11} + \bar{L}_1^2 B_{22} + \bar{L}_{01}\bar{L}_2)\left(\log\left(\frac{sd}{\rho}\right) + d\log\left(sC_{\mathbf{H}} \max_{i=0}^{3} \bar{L}_i\right)\right)$$

Then, conditional on event $\bar{E}$, with probability at least $1 - \rho$, the following hold:

- *(i) for all X such that $d_{\mathbf{H}}(x_i, x_{0,i}) \leqslant r_{\text{near}}$, we have*

$$\left\|\hat{\Upsilon}_X - \hat{\Upsilon}_{X_0}\right\| \lesssim C_{\mathbf{H}} B d_{\mathbf{H}}(X, X_0)\,.$$

- *(ii) for all X such that $d_{\mathbf{H}}(X, X_0) \lesssim \min\left(r_{\text{near}}, \frac{1}{C_{\mathbf{H}}B}\right)$, we have $\left\|\mathrm{Id} - \hat{\Upsilon}_X\right\| \leqslant \frac{3}{4}$ and $\left\|\mathbf{G}_X^{-\frac{1}{2}}\Gamma_X^*\right\| \lesssim 1$.*

Proof. By Lemma C.8 and Lemma C.10, with probability at least $1 - \rho$ conditonal on $\bar{E}$, for all $(i, j) \in \{(0, 0), (0, 1), (1, 1), (1, 2)\}$ and all $x, y \in \mathcal{X}^{\text{near}}$,

$$\left\|\hat{K}^{(ij)}(x, y)\right\| \leqslant \left\|K^{(ij)}(x, y)\right\| + \frac{1}{\sqrt{s}},$$

note that this also holds for $\hat{K}^{(ji)}(x, y)$ since $\hat{K}^{(ij)}(x, y) = \overline{\hat{K}^{(ij)}(y, x)}$.

In particular, for all x, x' such that $d_{\mathbf{H}}(x, x') \geqslant \Delta/4$, we have $\left\|\hat{K}^{(ij)}(x, x')\right\| \leqslant \frac{2}{\sqrt{s}}$. Take any X such that $d_{\mathbf{H}}(x_i, x_{0,i}) \leqslant r_{\text{near}}$, we have that both $x_i, x_{0,i}$ are at least $\Delta/4$-separated from x_j and $x_{0,j}$. Therefore, for $k, \ell \in \{0, 1\}$, using Lemma C.3:

$$\begin{aligned} \left\|\hat{K}^{(k\ell)}(x_i, x_j) - \hat{K}^{(k\ell)}(x_{i,0}, x_{j,0})\right\| &\lesssim \frac{C_{\mathbf{H}}}{\sqrt{s}}\sqrt{d_{\mathbf{H}}(x_i, x_{0,i})^2 + d_{\mathbf{H}}(x_j, x_{0,j})^2} \\ \left\|\hat{K}^{(k\ell)}(x_i, x_i) - \hat{K}^{(k\ell)}(x_{i,0}, x_{i,0})\right\| &\lesssim C_{\mathbf{H}}\left(B_{k+1,\ell} + B_{k,\ell+1}\right) d_{\mathbf{H}}(x_i, x_{0,i}) \end{aligned} \tag{D.10}$$

and therefore by Lemma G.6:

$$\begin{aligned} \left\|\hat{\Upsilon}_X - \hat{\Upsilon}_{X_0}\right\|^2 &\leqslant \sum_{i,j=1}^{s}\sum_{k,\ell=0}^{1}\left\|\hat{K}^{(k\ell)}(x_i, x_j) - \hat{K}^{(k\ell)}(x_{0,i}, x_{0,j})\right\|^2 \\ &\leqslant 2\sum_{i,j=1}^{s}\sum_{k,\ell=0}^{1}\left\|\hat{K}^{(k\ell)}(x_i, x_j) - \hat{K}^{(k\ell)}(x_{0,i}, x_j)\right\|^2 + \left\|\hat{K}^{(\ell k)}(x_j, x_{0,i}) - \hat{K}^{(\ell k)}(x_{0,j}, x_{0,i})\right\|^2 \\ &\lesssim C_{\mathbf{H}}^2\left(\sum_{\substack{k,l\in\{0,1,2\}\\ k+\ell\leqslant 3}} B_{k\ell}\right)^2 \sum_i d_{\mathbf{H}}(x_i, x_{0,i})^2 + \frac{1}{s}\sum_{j\neq i} d_{\mathbf{H}}(x_j, x_{0,j})^2 \end{aligned}$$

which yields the desired result.

For the second statement, using Proposition C.1, $\mathbb{P}_{\bar{E}}(\left\|\hat{\Upsilon}_{X_0} - \Upsilon_{X_0}\right\| > \frac{1}{8}) \leqslant \rho$, so conditional on $\bar{E}$, we have with probability $1-\rho$, $\left\|\hat{\Upsilon}_X - \hat{\Upsilon}_{X_0}\right\| \leqslant \frac{1}{8}$ and the claim follows since $\|\mathrm{Id} - \Upsilon_{X_0}\| \leqslant \frac{1}{2}$ (due to Lemma C.1) implies that $\left\|\mathrm{Id} - \hat{\Upsilon}_X\right\| \leqslant \frac{3}{4}$ and

$$\left\|\hat{\Upsilon}_X\right\| \leqslant 7/4 \quad \text{and} \quad \left\|\mathbf{G}_X^{-\frac{1}{2}}\Gamma_X^*\right\| = \sqrt{\left\|\hat{\Upsilon}_X\right\|} \lesssim \sqrt{7}/2.$$

□

Proof of Theorem D.2. Since $\hat{\eta}_{X_0,a_0}$ is nondegenerate with probability at least $1-\rho$, the conclusion follows if we prove that for all $x \in \mathcal{X}^{\mathrm{far}}$ and all $y \in \mathcal{X}^{\mathrm{near}}$,

$$\|\mathrm{D}_0\left[\hat{\eta}_{X,a} - \hat{\eta}_{X_0,a_0}\right](x)\| \leqslant \varepsilon_0/32 \quad \text{and} \quad \|\mathrm{D}_2\left[\hat{\eta}_{X,a} - \hat{\eta}_{X_0,a_0}\right](y)\| \leqslant p\varepsilon_2/32 \tag{D.11}$$

with probability at least $1-\rho$. We first write

$$\hat{\eta}_{X,a}(y) - \hat{\eta}_{X_0,a_0}(y) = \left(\hat{\Upsilon}_X^{-1}\begin{pmatrix}\mathrm{sign}(a)\\ 0_{sd}\end{pmatrix}\right)^\top (\hat{\mathbf{f}}_X - \hat{\mathbf{f}}_{X_0}) + \left(\hat{\Upsilon}_X^{-1}\begin{pmatrix}\mathrm{sign}(a)\\ 0_{sd}\end{pmatrix} - \hat{\Upsilon}_{X_0}^{-1}\begin{pmatrix}\mathrm{sign}(a_0)\\ 0_{sd}\end{pmatrix}\right)^\top \hat{\mathbf{f}}_{X_0} \tag{D.12}$$

Conditional on $\bar{E}$, with probability at least $1-\rho/2$, we have by Lemma D.1 (note that our assumptions imply the assumptions of Lemma D.1), $\|\Upsilon_X - \Upsilon_{X_0}\| \lesssim C_{\mathbf{H}} B d_{\mathbf{H}}(X, X_0)$ and $\left\|\Upsilon_X^{-1}\right\| \leqslant 4$. Combining this with Lemma C.2, we obtain $\left\|\mathrm{D}_r\left[\left(\hat{\Upsilon}_X^{-1}\binom{\mathrm{sign}(a)}{0_{sd}}\right)(\hat{\mathbf{f}}_{X_0} - \hat{\mathbf{f}}_X)\right](y)\right\| \leqslant 4\sqrt{s}\bar{L}_r\sqrt{\mathcal{L}_0^2 + \mathcal{L}_1^2}d_{\mathbf{H}}(X, X_0)$. For the second term of (D.12),

$$\begin{aligned}
&\left\|\hat{\Upsilon}_X^{-1}\begin{pmatrix}\mathrm{sign}(a)\\ 0_{sd}\end{pmatrix} - \hat{\Upsilon}_{X_0}^{-1}\begin{pmatrix}\mathrm{sign}(a_0)\\ 0_{sd}\end{pmatrix}\right\| \\
&= \left\|\hat{\Upsilon}_X^{-1}\left(\begin{pmatrix}\mathrm{sign}(a)\\ 0_{sd}\end{pmatrix} - \begin{pmatrix}\mathrm{sign}(a_0)\\ 0_{sd}\end{pmatrix}\right) + \left(\hat{\Upsilon}_X^{-1} - \hat{\Upsilon}_{X_0}^{-1}\right)\begin{pmatrix}\mathrm{sign}(a_0)\\ 0_{sd}\end{pmatrix}\right\| \\
&\leqslant 4\,\|\mathrm{sign}(a) - \mathrm{sign}(a_0)\| + 8\sqrt{s}\left\|\hat{\Upsilon}_X - \hat{\Upsilon}_{X_0}\right\| \leqslant 8\frac{\|a-a_0\|}{\min_i |a_{0,i}|} + 8\sqrt{s}C_{\mathbf{H}} B d_{\mathbf{H}}(X, X_0).
\end{aligned}$$

So, $\left\|\mathrm{D}_r\left[\left(\hat{\Upsilon}_X^{-1}\binom{\mathrm{sign}(a)}{0_{sd}} - \hat{\Upsilon}_{X_0}^{-1}\binom{\mathrm{sign}(a_0)}{0_{sd}}\right)^\top \hat{\mathbf{f}}_{X_0}\right](y)\right\| \leqslant 16 B_r\left(\frac{\|a-a_0\|}{\min_i|a_{0,i}|} + \sqrt{s}C_{\mathbf{H}} B d_{\mathbf{H}}(X, X_0)\right)$.

Finally, since $\mathbb{P}(\bar{E}^c) \leqslant \rho/2$, we have with probability at least $1-\rho$, for all $y \in \mathcal{X}$, (D.11) holds provided that (D.9) holds. Combining with the nondegeneracy of $\hat{\eta}_{X_0,a_0}$, the conclusion follows with probability $1-2\rho$.

□

E Supplementary results to the proof Theorem 1

Let $\Phi_X : \mathbb{C}^s \to \mathbb{C}^m$ and its adjoint $\Phi_X^* : \mathbb{C}^m \to \mathbb{C}^s$ be defined by

$$\forall a \in \mathbb{C}^s,\ \Phi_X a = \sum_{j=1}^s a_j \varphi(x_j) \in \mathbb{C}^m, \quad \text{and} \quad \forall q \in \mathbb{C}^m,\ \Phi_X^* q = \left(\langle \varphi(x_j),\, q\rangle\right)_{j=1}^s,$$

and $(\Phi_X^{(1)}) : \mathbb{C}^{sd} \to \mathbb{C}^m$,

$$\forall P_i \in \mathbb{C}^d,\ (\Phi_X^{(1)})[P_1, \ldots, P_s] = \left(\sum_{i=1}^s \langle \nabla\varphi_{\omega_k}(x_i),\, P_i\rangle\right)_{k=1}^m$$

with adjoint

$$\forall q \in \mathbb{C}^m,\ (\Phi_X^{(1)})^*\ q = (\nabla_x \langle \varphi(x_i),\, q\rangle)_{i=1}^s\,.$$

In the following, we interpret $\Phi_X \in \mathbb{C}^{s\times m}$ and $\Phi_X^{(1)} \in \mathbb{C}^{sd\times m}$ as matrices.

Recall that in the proof of Theorem 1, we defined the function $f : \mathbb{R}^{2s} \times \mathcal{X}^s \times \mathbb{R}_+ \times \mathbb{R}^{2m} \to \mathbb{R}^{2s} \times \mathbb{C}^{sd}$ by

$$f(u,v) \overset{\text{def.}}{=} \begin{pmatrix} \mathrm{Re}\,(\Phi_X^*(\Phi_X(a_\mathrm{r}+\mathrm{i}a_\mathrm{i}) - \Phi_{X_0}a_0 - (w_\mathrm{r}+\mathrm{i}w_\mathrm{i}))) \\ \mathrm{Im}\,(\Phi_X^*(\Phi_X(a_\mathrm{r}+\mathrm{i}a_\mathrm{i}) - \Phi_{X_0}a_0 - (w_\mathrm{r}+\mathrm{i}w_\mathrm{i}))) \\ (\Phi_X^{(1)})^*(\Phi_X(a_\mathrm{r}+\mathrm{i}a_\mathrm{i}) - \Phi_{X_0}a_0 - (w_\mathrm{r}+\mathrm{i}w_\mathrm{i})) \end{pmatrix} + \lambda \begin{pmatrix} \left(\frac{a_{\mathrm{r}\,i}}{|a_i|}\right)_{i=1}^s \\ \left(\frac{a_{\mathrm{i}\,i}}{|a_i|}\right)_{i=1}^s \\ 0_{sd} \end{pmatrix}$$

where $u = (a_\mathrm{r}, a_\mathrm{i}, X)$, $v = (\lambda, w_\mathrm{r}, w_\mathrm{i})$ for $a_\mathrm{r}, a_\mathrm{i} \in \mathbb{R}^s$, $X \in \mathcal{X}^s$, $\lambda > 0$, $w_\mathrm{r}, w_\mathrm{i} \in \mathbb{R}^m$, and $a \overset{\text{def.}}{=} a_\mathrm{r} + \imath a_\mathrm{i} \in \mathbb{C}^s$, $w \overset{\text{def.}}{=} w_\mathrm{r} + \mathrm{i}w_\mathrm{i}$.

Differentiability of f The function f is differentiable at all (u,v) such that $i = 1,\ldots,s$, $a_\mathrm{r} + \mathrm{i}a_\mathrm{i} \neq 0$. Its differential can be written as

$$\partial_{w_\mathrm{r}} f = -\begin{pmatrix} \mathrm{Re}\,(\Phi_X^*) \\ \mathrm{Im}\,(\Phi_X^*) \\ (\Phi_X^{(1)})^* \end{pmatrix}, \quad \partial_{w_\mathrm{i}} f = -\mathrm{i}\begin{pmatrix} \mathrm{Re}\,(\Phi_X^*) \\ \mathrm{Im}\,(\Phi_X^*) \\ (\Phi_X^{(1)})^* \end{pmatrix}, \quad \partial_\lambda f = \begin{pmatrix} \left(\frac{a_{\mathrm{r}\,i}}{|a_i|}\right)_i \\ \left(\frac{a_{\mathrm{i}\,i}}{|a_i|}\right)_i \\ 0_{sd} \end{pmatrix} \tag{E.1}$$

so

$$\partial_v f(u,v) = \left(\begin{pmatrix} \left(\frac{a_{\mathrm{r}\,i}}{|a_i|}\right)_i \\ \left(\frac{a_{\mathrm{i}\,i}}{|a_i|}\right)_i \\ 0_{sd} \end{pmatrix},\ -\begin{pmatrix} \mathrm{Re}\,(\Phi_X^*) \\ \mathrm{Im}\,(\Phi_X^*) \\ (\Phi_X^{(1)})^* \end{pmatrix},\ -\mathrm{i}\begin{pmatrix} \mathrm{Re}\,(\Phi_X^*) \\ \mathrm{Im}\,(\Phi_X^*) \\ (\Phi_X^{(1)})^* \end{pmatrix} \right) \in \mathbb{C}^{(2s+sd)\times(1+2m)}, \tag{E.2}$$

and $\partial_u f(u,v) = (M_1(u,v) + M_2(u,v)) \begin{pmatrix} \mathrm{Id}_{s\times s} & 0 & 0 \\ 0 & \mathrm{Id}_{s\times s} & \\ 0 & 0 & J_a \end{pmatrix}$ with $M_1(u,v) \overset{\text{def.}}{=} \begin{pmatrix} D_{0,X} & \tilde{D}_{1,X} \\ D_{1,X} & D_{2,X} \end{pmatrix}$ and

$$M_2(u,v) \overset{\text{def.}}{=} \begin{pmatrix} C_1 & C_2 & \begin{pmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & A_{12} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{1s} \\ A_{21} & 0 & \cdots & 0 \\ 0 & A_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{2s} \end{pmatrix} \end{pmatrix} \begin{pmatrix} \mathrm{Id} & 0 & 0 \\ 0 & \mathrm{Id} & \\ 0 & 0 & J_a^{-1} \end{pmatrix} \tag{E.3}$$

where

$$D_{0,X} \overset{\text{def.}}{=} \begin{pmatrix} \mathrm{Re}\,(\Phi_X^*\Phi_X) & -\mathrm{Im}\,(\Phi_X^*\Phi_X) \\ \mathrm{Im}\,(\Phi_X^*\Phi_X) & \mathrm{Re}\,(\Phi_X^*\Phi_X) \end{pmatrix}, \quad \tilde{D}_{1,X} \overset{\text{def.}}{=} \begin{pmatrix} \mathrm{Re}\left(\Phi_X^*\Phi_X^{(1)}J_a\right)J_a^{-1} \\ \mathrm{Im}\left(\Phi_X^*\Phi_X^{(1)}J_a\right)J_a^{-1} \end{pmatrix},$$

$$D_{1,X} \overset{\text{def.}}{=} \left((\Phi_X^{(1)})^*\Phi_X \quad \mathrm{i}(\Phi_X^{(1)})^*\Phi_X\right), \quad \text{and} \quad D_{2,X} \overset{\text{def.}}{=} (\Phi_X^{(1)})^*\Phi_X^{(1)}$$

and $C_1, C_2 \in \mathbb{R}^{(ds+2s)\times s}$ are defined as

$$C_1 \overset{\text{def.}}{=} \lambda \begin{pmatrix} \mathrm{diag}\left(\left(\frac{1}{|a_i|} - \frac{a_{\mathrm{r}\,i}^2}{|a_i|^3}\right)_i\right) \\ \mathrm{diag}\left(\left(-\frac{a_{\mathrm{i}\,i}a_{\mathrm{r}\,i}}{|a_i|^3}\right)_i\right) \\ 0_{sd\times s} \end{pmatrix}, \qquad C_2 \overset{\text{def.}}{=} \lambda \begin{pmatrix} \mathrm{diag}\left(\left(-\frac{a_{\mathrm{i}\,i}a_{\mathrm{r}\,i}}{|a_i|^3}\right)_i\right) \\ \mathrm{diag}\left(\left(\frac{1}{|a_i|} - \frac{a_{\mathrm{i}\,i}^2}{|a_i|^3}\right)_i\right) \\ 0_{sd\times s} \end{pmatrix},$$

$$A_{1j} \stackrel{\text{def.}}{=} \begin{pmatrix} \mathrm{Re}\left(\nabla_x \langle \varphi(x_j), z\rangle\right)^\top \\ \mathrm{Im}\left(\nabla_x \langle \varphi(x_j), z\rangle\right)^\top \end{pmatrix}, \quad A_{2j} \stackrel{\text{def.}}{=} \nabla_x^2 \langle \varphi(x_j), z\rangle, \quad z \stackrel{\text{def.}}{=} (\Phi_X a - \Phi_{X_0} a_0 - w) \tag{E.4}$$

and $J_a \in \mathbb{R}^{sd\times sd}$ is the diagonal matrix:

$$J_a = \begin{pmatrix} a_1 \mathrm{Id}_{d\times d} & & 0 \\ & \ddots & \\ 0 & & a_s \mathrm{Id}_{d\times d} \end{pmatrix}.$$

Letting $u_0 = (\mathrm{Re}\,(a_0), \mathrm{Im}\,(a_0), X_0)$ and $v_0 = (0,0,0)$, we have that $M_2(u_0,v_0) = 0$ and $\partial_u f(u_0,v_0)$ is invertible since $\left\|\hat{\Upsilon}_{X_0} - \mathrm{Id}\right\| \leqslant 1/8$. Moreover, $f(u_0,v_0) = 0$. Hence, by the Implicit Function Theorem, there exists a neighbourhood V of v_0, a neighbourhood U of u_0 and a differentiable function $g : V \to U$ such that for all $(u,v) \in U \times V$, $f(u,v) = 0$ if and only if $u = g(v)$. To conclude, we simply need to bound the size of the region on which g is well defined, and to bound the error between $g(v)$ and $g(0)$. This is done with the following theorem. Let us first remark that our assumptions imply that $\mathbb{P}(\bar{E}^c) \leqslant \rho/2$ and

$$\mathbb{P}(E_\omega^c) \leqslant \frac{1}{1+16\sqrt{s}B_{ij}}, \quad \mathbb{E}[L_i(\omega)L_j(\omega)\mathbb{1}_{E_\omega^c}] \leqslant \frac{1}{16\sqrt{s}}, \tag{E.5}$$

for all $i,j = 0,...,2$. Therefore, it is sufficient to prove the existence of g conditional on event $\bar{E}$:

Theorem E.1. *Assume that for all $i+j \leqslant 3$*

$$\mathbb{P}(E_\omega^c) \leqslant \frac{1}{1+16\sqrt{s}B_{ij}}, \quad \mathbb{E}[L_i(\omega)L_j(\omega)\mathbb{1}_{E_\omega^c}] \leqslant \frac{1}{16\sqrt{s}}$$

for all $i,j = 0,...,2$. Let $\rho > 0$ and suppose that

$$m \gtrsim s(\bar{L}_2^2 B_{11} + \bar{L}_1^2 B_{22} + \bar{L}_{01}\bar{L}_2)\left(\log\left(\frac{sd}{\rho}\right) + d\log\left(sC_{\mathbf{H}}\mathbb{L}_3\right)\right)$$

where $\mathbb{L}_r \stackrel{\text{def.}}{=} \max_{i\leqslant r} L_r$. Then, conditional on event $\bar{E}$, with probability at least $1-\rho$: there exists a $\mathscr{C}^1$ function g such that, for all $v = (\lambda, w)$ such that $\|v\| \leqslant r$ with r satisfying

$$r = \mathcal{O}\left(\frac{1}{\sqrt{s}}\min\left(\frac{\min\{r_{\text{near}},(C_{\mathbf{H}}B)^{-1}\}}{\min_i |a_{0,i}|},\ \frac{1}{\bar{L}_{01}\bar{L}_{12}(1+\|a_0\|)},\right)\right) \tag{E.6}$$

we have $f(g(v),v) = 0$ and $g(0) = u_0$. Furthermore, given (λ,w) in this ball, $(a,X) \stackrel{\text{def.}}{=} g((\lambda,w))$ satisfies

$$\|a - a_0\| + d_{\mathbf{H}}(X,X_0) \leqslant \frac{\sqrt{s}(\lambda + \|w\|)}{\min_i |a_{0,i}|}. \tag{E.7}$$

Before proceeding to prove this result, we first remark that as discussed in the main paper, Lemma E.1 and Lemma E.2 below imply that given $v = (\lambda, w_{\mathrm{r}}, w_{\mathrm{i}}) \in V$, $u = g(v)$ indeed correspond to the unique solution of the BLASSO with regularisation parameter λ and noise $w = w_{\mathrm{r}} + \mathrm{i}w_{\mathrm{i}}$. In particular, the combination of these two lemmas imply that the certificate $\eta_{\lambda,w} \stackrel{\text{def.}}{=} \Phi^* p_{\lambda,w}$ associated to a and X is close to the nondegenerate certificate $\eta_{X,a} \stackrel{\text{def.}}{=} \Phi^* p_{X,a}$ when $\|w\|/\lambda$ and λ are sufficiently small. In the following, $\Pi_X \stackrel{\text{def.}}{=} (\mathrm{Id} - \Gamma_X\Gamma_X^\dagger)$ is the orthogonal projection onto $\mathrm{Im}(\Gamma_X)^\perp$.

Lemma E.1. *Given $u = (a_{\mathrm{r}}, a_{\mathrm{i}}, X)$ and $v = (\lambda, w_{\mathrm{r}}, w_{\mathrm{i}})$ such that $f(u,v) = 0$, write $a = a_{\mathrm{r}} + \mathrm{i}a_{\mathrm{i}}$ and $w = w_{\mathrm{r}} + \mathrm{i}w_{\mathrm{i}}$. Let $p_{\lambda,w} \stackrel{\text{def.}}{=} \frac{1}{\lambda}\left(\Phi_{X_0}a_0 - \Phi_X a + w\right)$. Then,*

$$p_{\lambda,w} = p_{X,a} + \frac{1}{\lambda}\Pi_X w + \frac{1}{\lambda}\Pi_X \Phi_{X_0} a_0.$$

Proof. The equation $f(u,v)=0$ can be written as

$$\Gamma_X^* \left(\Gamma_X \begin{pmatrix} a \\ 0_{sd} \end{pmatrix} - \Gamma_{X_0} \begin{pmatrix} a_0 \\ 0_{sd} \end{pmatrix} - w \right) + \lambda \begin{pmatrix} \mathrm{sign}(a) \\ 0_{sd} \end{pmatrix} = 0$$

By applying $\tilde{\Gamma}_X(\tilde{\Gamma}_X^* \tilde{\Gamma}_X)^\dagger$ to the above equation, we obtain

$$\Gamma_X \begin{pmatrix} a \\ 0_{sd} \end{pmatrix} - \Gamma_X \Gamma_X^\dagger \Gamma_{X_0} \begin{pmatrix} a_0 \\ 0_{sd} \end{pmatrix} - \Gamma_X \Gamma_X^\dagger w + \lambda \Gamma_X^{*,\dagger} \begin{pmatrix} \mathrm{sign}(a) \\ 0_{sd} \end{pmatrix} = 0 \tag{E.8}$$

Therefore, since $\Pi_X = (\mathrm{Id} - \Gamma_X \Gamma_X^\dagger)$, we have

$$-\Phi_X a + \Phi_{X_0} a_0 + w = \Pi_X \Phi_{X_0} a_0 + \Pi_X w + \lambda \Gamma_X^{*,\dagger} \begin{pmatrix} \mathrm{sign}(a) \\ 0_{sd} \end{pmatrix} \tag{E.9}$$

and by dividing by λ, we obtain the desired equation.

□

Lemma E.2. *Asssume that event $\bar{E}$ occurs. Then, for all $X \in \mathcal{X}^s$ such that $d_{\mathbf{H}}(x_i, x_{0,i}) \leqslant r_{\text{near}}$ and $a \in \mathbb{C}^s$,*

$$\|\Pi_X \Gamma_{X_0} a\| \lesssim \begin{cases} \bar{L}_2 \|a\|_1 \max_i d_{\mathbf{H}}(x_i, x_{0,i})^2 \\ \bar{L}_2 \|a\|_\infty d_{\mathbf{H}}(X, X_0)^2 \end{cases}$$

Proof. Let $\gamma_i : [0,1] \to \mathcal{X}$ be any piecewise smooth curve such that $\gamma_i(1) = x_{0,i}$ and $\gamma_i(0) = x_i$. Then, by Taylor expanding $\varphi_{\omega_k}(\gamma_i(t))$ about $t=0$, we obtain

$$\varphi_{\omega_k}(x_{0,i}) = \varphi(x_i) + \langle \nabla \varphi_{\omega_k}(x_i),\, \gamma_i'(0) \rangle + \int_0^1 \frac{1}{2} \langle \nabla^2 \varphi_{\omega_k}(\gamma_i(t)) \gamma_i'(t),\, \gamma_i'(t) \rangle \mathrm{d}t.$$

Therefore, since $\mathrm{Im}(\Gamma_X) = \{\varphi(x_i), J_\varphi(x_i)\}_i$ where J_φ denotes the Jacobian of φ, and Π_X is a projector on $\mathrm{Im}(\Gamma_X)^\perp$,

$$\Pi_X \Gamma_{X_0} a = \Pi_X \left(\sum_{i=1}^s a_i \varphi(x_{0,i}) \right) = \Pi_X \left(\sum_{i=1}^s \frac{a_i}{2} \int_0^1 \langle \nabla^2 \varphi_{\omega_k}(\gamma_i(t)) \gamma_i'(t),\, \gamma_i'(t) \rangle \mathrm{d}t \right)_{k=1}^m$$

Taking the norm implies

$$\|\Pi_X \Gamma_{X_0} a\| \leqslant \sum_{i=1}^s \frac{|a_i|}{2} \int_0^1 \bar{L}_2 \left\| \mathbf{H}_{\gamma_i(t)} \gamma_i'(t) \right\|^2 \mathrm{d}t \tag{E.10}$$

since for $d_{\mathbf{H}}(x_i, x_{0,i}) \leqslant r_{\text{near}}$, we have $\left\| \mathbf{H}_{x_{0,i}}^{-\frac{1}{2}} \mathbf{H}_{x_i}^{\frac{1}{2}} \right\| \lesssim 1$, and hence, under $\bar{E}$:

$$\left\| \mathbf{H}_{x_{0,i}}^{-\frac{1}{2}} \nabla^2 \varphi_{\omega_j}(x_i) \mathbf{H}_{x_{0,i}}^{-\frac{1}{2}} \right\| \lesssim \left\| \mathrm{D}_2 \left[\varphi_{\omega_j} \right](x_i) \right\| \leqslant \bar{L}_2.$$

Taking the infimum over all paths γ_i in (E.10) yields

$$\|\Pi_X \Gamma_{X_0} a\| \leqslant \bar{L}_2 \sum_i |a_i| \, d_{\mathbf{H}}(x_i, x_{0,i})^2.$$

□

E.0.1 Proof of Theorem E.1

E.0.2 Preliminary results

We first discuss the invertibility of $\partial_u f$. To this end, we make the following definitions.

Let $u = (a_{\rm r}, a_{\rm i}, X)$, $u_0 = (\mathrm{Re}\,(a_0), \mathrm{Im}\,(a_0), X_0)$, $v = (\lambda, w_{\rm r}, w_{\rm i})$ and $v_0 = (0,0,0)$. We define the block diagonal matrices

$$\mathbf{F}_X \stackrel{\text{def.}}{=} \begin{pmatrix} \mathrm{Id}_{s\times s} & 0 & 0 \\ 0 & \mathrm{Id}_{s\times s} & 0 \\ 0 & 0 & \mathbf{G}_X \end{pmatrix} \quad \text{where} \quad \mathbf{G}_X \stackrel{\text{def.}}{=} \begin{pmatrix} \mathbf{H}_{x_1} & & 0 \\ & \ddots & \\ 0 & & \mathbf{H}_{x_s} \end{pmatrix}.$$

For (u,v) sufficiently close to (u_0, v_0), we aim to show that $\partial_u f(u,v)$ is invertible and to control $\left\|\mathbf{F}_X^{-\frac{1}{2}} \partial_v f(u,v)\right\|$ and $\left\|\mathbf{F}_X^{\frac{1}{2}} \partial_u f(u,v)^{-1} \mathbf{F}_X^{\frac{1}{2}}\right\|$. Using Lemma D.1, conditional on event $\bar{E}$, with probability $1-\rho$ we have

$$\left\|\mathbf{F}_X^{-\frac{1}{2}} \partial_v f(u,v)\right\| \leqslant \|\mathbf{u}\| + \left\| \begin{pmatrix} \mathrm{Re}\,(\Phi_X^*) \\ \mathrm{Im}\,(\Phi_X^*) \\ \mathbf{G}_X^{-\frac{1}{2}} (\Phi_X^{(1)})^* \end{pmatrix} \right\| \lesssim \sqrt{s} \tag{E.11}$$

To deduce invertibility of $\partial_u f(u,v)$ and to bound $\left\|\mathbf{F}_X^{\frac{1}{2}} \partial_u f(u,v)^{-1} \mathbf{F}_X^{\frac{1}{2}}\right\|$, first observe that

$$\mathbf{F}_X^{-1/2} \partial_u f(u,v) \mathbf{F}_X^{-1/2} = \left(\mathbf{F}_X^{-1/2} M_1(u,v) \mathbf{F}_X^{-1/2} + \mathbf{F}_X^{-1/2} M_2(u,v) \mathbf{F}_X^{-1/2}\right) \begin{pmatrix} \mathrm{Id} & 0 & 0 \\ 0 & \mathrm{Id} & \\ 0 & 0 & J_a \end{pmatrix}$$

where $\mathbf{F}_X^{-1/2} M_2(u,v) \mathbf{F}_X^{-1/2}$ is

$$\left(C_1 \quad C_2 \quad \begin{pmatrix} \frac{1}{a_1}\left(\mathbf{H}_{x_1}^{-\frac{1}{2}} \mathrm{Re}\,(\nabla[\langle \varphi, z\rangle](x_1))\right)^\top & 0 & \cdots & 0 \\ \frac{1}{a_1}\left(\mathbf{H}_{x_1}^{-\frac{1}{2}} \mathrm{Im}\,(\nabla[\langle \varphi, z\rangle](x_1))\right)^\top & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{1}{a_s}\left(\mathbf{H}_{x_s}^{-\frac{1}{2}} \mathrm{Re}\,(\nabla[\langle \varphi, z\rangle](x_s))\right)^\top \\ 0 & \cdots & 0 & \frac{1}{a_s}\left(\mathbf{H}_{x_s}^{-\frac{1}{2}} \mathrm{Im}\,(\nabla[\langle \varphi, z\rangle](x_s))\right)^\top \\ \frac{1}{a_1} \mathbf{H}_{x_1}^{-\frac{1}{2}} \nabla^2[\langle \varphi, z\rangle](x_1) \mathbf{H}_{x_1}^{-\frac{1}{2}} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{1}{a_s} \mathbf{H}_{x_s}^{-\frac{1}{2}} \nabla^2[\langle \varphi, z\rangle](x_s) \mathbf{H}_{x_s}^{-\frac{1}{2}} \end{pmatrix} \right), \tag{E.12}$$

where $z = (\Phi_X a - \Phi_{X_0} a_0 - w)$. Now, let us study the invertibility of $\mathbf{F}_X^{-1/2} M_1(u,v) \mathbf{F}_X^{-1/2} + \mathbf{F}_X^{-1/2} M_2(u,v) \mathbf{F}_X^{-1/2}$ and bound the norm of its inverse.

Lemma E.3 (Bound on $M_2(u,v)$)**.** *Assume that $\bar{E}$ occurs and given $\varepsilon > 0$, let $c_\varepsilon \stackrel{\text{def.}}{=} \frac{\varepsilon \, \min_i |a_{0,i}|}{2\bar{L}_{12}}$. Then, for all $X \in \mathcal{X}^s$, $a \in \mathbb{C}^s$ and $w \in \mathbb{C}^m$ such that*

$$\lambda \leqslant \frac{\min_i |a_{0,i}|}{4}, \quad \|a - a_0\| \leqslant \frac{c_\varepsilon}{4\bar{L}_0}, \quad \|w\| \leqslant \frac{c_\varepsilon}{4} \quad \textit{and} \quad d_{\mathbf{H}}(X, X_0) \leqslant \min\left(r_{\rm near}, \frac{c_\varepsilon}{4\bar{L}_1 \|a_0\|}\right),$$

we have for $u \stackrel{\text{def.}}{=} (\text{Re}(a), \text{Im}(a), X)$ *and* $v \stackrel{\text{def.}}{=} (\text{Re}(w), \text{Im}(w), X)$*,*

$$\left\|\mathbf{F}_X^{-1/2} M_2(u,v) \mathbf{F}_X^{-1/2}\right\| \leqslant \varepsilon \quad \textit{and} \quad \left\|\mathbf{F}_X^{-1/2} M_2(u,v) \mathbf{F}_X^{-1/2}\right\|_{\text{Block}} \leqslant \varepsilon$$

Proof. First note that for $r \in \mathbb{N}_0$,

$$\left\|\mathrm{D}_r \left[\varphi^\top z\right](x_i)\right\| \leqslant \frac{1}{\sqrt{m}} \sum_{j=1}^m \left\|z_j \mathrm{D}_r \left[\varphi_{\omega_j}\right](x_i)\right\| \leqslant \bar{L}_r \|z\|$$

Now, for $\bar{q} = [P_1, P_2, Q_1, \dots, Q_s] \in \mathbb{C}^{s(d+2)}$, where $P_i \in \mathbb{C}^s$ and $Q_i \in \mathbb{C}^d$, and $\|\bar{q}\| = 1$, we have

$$\begin{aligned}
&\left\|\mathbf{F}_X^{-1/2} M_2(u,v) \mathbf{F}_X^{-1/2} \bar{q}\right\|^2 \\
&\lesssim \|C_1 P_1\|^2 + \|C_2 P_2\|^2 + \sum_{i=1}^s \left|\frac{1}{a_i}\left(\mathbf{H}_{x_i}^{-\frac{1}{2}} \nabla[\varphi^\top z](x_i)\right)^\top Q_i\right|^2 + \left\|\frac{1}{a_i} \mathbf{H}_{x_i}^{-\frac{1}{2}} \nabla^2[\varphi^\top z](x_i) \mathbf{H}_{x_i}^{-\frac{1}{2}} Q_i\right\|^2 \\
&\lesssim \frac{\lambda^2}{\min_i |a_{0,i}|^2} + \frac{4}{\min_i |a_{0,i}|^2} \max_i \left(\left\|\mathbf{H}_{x_i}^{-\frac{1}{2}} \nabla[\varphi^\top z](x_i)\right\|^2 + \left\|\mathbf{H}_{x_i}^{-\frac{1}{2}} \nabla^2[\varphi^\top z](x_i) \mathbf{H}_{x_i}^{-\frac{1}{2}}\right\|^2\right) \\
&= \frac{\lambda^2}{\min_i |a_{0,i}|^2} + \frac{4}{\min_i |a_{0,i}|^2} \max_i \left(\left\|\mathrm{D}_1\left[\varphi^\top z\right](x_i)\right\|^2 + \left\|\mathrm{D}_2\left[\varphi^\top z\right](x_i)\right\|^2\right) \\
&\leqslant \frac{\lambda^2}{\min_i |a_{0,i}|^2} + \frac{4}{\min_i |a_{0,i}|^2} (\bar{L}_1^2 + \bar{L}_2^2) \|z\|^2
\end{aligned}$$

where we have used the fact that $\min_i |a_i| \geqslant \min_i |a_{0,i}|/2$. If $\|\bar{q}\|_{\text{Block}} = 1$, then

$$\begin{aligned}
\left\|\mathbf{F}_X^{-1/2} M_2(u,v) \mathbf{F}_X^{-1/2} \bar{q}\right\|_{\text{Block}} &\lesssim \frac{\lambda}{\min_i |a_{0,i}|} + \max_i \{\left|\left(\mathbf{H}_{x_i}^{-\frac{1}{2}} \nabla[\varphi^\top z](x_i)\right)^\top Q_i\right|, \left\|\mathbf{H}_{x_i}^{-\frac{1}{2}} \nabla[\varphi^\top z](x_i) \mathbf{H}_{x_i}^{-\frac{1}{2}} Q_i\right\|^2\} \\
&\leqslant \frac{\lambda}{\min_i |a_{0,i}|} + \max_i \{\left\|\mathbf{H}_{x_i}^{-\frac{1}{2}} \nabla[\varphi^\top z](x_i)\right\|, \left\|\mathbf{H}_{x_i}^{-\frac{1}{2}} \nabla[\varphi^\top z](x_i) \mathbf{H}_{x_i}^{-\frac{1}{2}}\right\|^2\}
\end{aligned}$$

and the same bound holds.

Now it remains to bound $\|z\|$ (recall the definition of z from (E.4)). Writing $\varphi(x) \stackrel{\text{def.}}{=} (\varphi_{\omega_k}(x))_{k=1}^m$, we have

$$\begin{aligned}
\|z\| &= \left\|\sum_i (a_i \varphi(x_i) - a_{0,i} \varphi(x_{0,i})) - w\right\| \\
&\leqslant \bar{L}_0 \|a - a_0\| + \|a_0\| \max_k \sqrt{\sum_i |\varphi_{\omega_k}(x_i) - \varphi_{\omega_k}(x_{0,i})|^2} + \|w\| \\
&\leqslant \bar{L}_0 \|a - a_0\| + \|a_0\| \bar{L}_1 d_{\mathbf{H}}(X, X_0) + \|w\|
\end{aligned}$$

where the last inequality follows from Lemma C.2.

□

Lemma E.4. *If* $\|a - a_0\| \leqslant \frac{1}{2} \min |a_{0,i}|$ *and* $\left\|\hat{\Upsilon}_X - \text{Id}\right\| \leqslant \varepsilon < \frac{1}{3}$*, then* $M_1(u,v)$ *is invertible and*

$$\left\|\mathbf{F}_X^{\frac{1}{2}} M_1(u,v)^{-1} \mathbf{F}_X^{\frac{1}{2}}\right\| \leqslant \frac{4}{1 - \varepsilon - 4\varepsilon^2}.$$

Proof. By considering the Schur complement, we have that $M_1(u,v)$ is invertible provided that

(i) $D_{2,X}$ is invertible

(ii) $S \stackrel{\text{def.}}{=} D_{0,X} - \tilde{D}_{1,X} D_{2,X}^{-1} D_{1,X}$ is invertible.

In this case,

$$M_1(u,v)^{-1} = \begin{pmatrix} S^{-1} & -S^{-1}\tilde{D}_{1,X} D_{2,X}^{-1} \\ -D_{2,X}^{-1} D_{1,X} S^{-1} & D_{2,X}^{-1} + D_{2,X}^{-1} S^{-1} \tilde{D}_{1,X} D_{2,X}^{-1} \end{pmatrix}.$$

To establish (i) and (ii): Note that $\left\|\mathbf{G}_X^{-\frac{1}{2}}(\Phi_X^{(1)})^*\Phi_X\right\|, \left\|\mathbf{G}_X^{-\frac{1}{2}}(\Phi_X^{(1)})^*\Phi_X^{(1)}\mathbf{G}_X^{-\frac{1}{2}} - \mathrm{Id}\right\|, \|\Phi_X^*\Phi_X - \mathrm{Id}\| \leqslant \left\|\hat{\Upsilon}_X - \mathrm{Id}\right\|$. So, provided that $\left\|\hat{\Upsilon}_X - \mathrm{Id}\right\| \leqslant \varepsilon < 1$, (i) is satisfied, and note that $D_{0,X}$ on $\mathbb{R}^{2s}$ is invertible if and only if $\Phi_X^*\Phi_X$ is invertible on $\mathbb{C}^s$ and $\left\|(\Phi_X^*\Phi_X)^{-1}\right\| = \left\|D_{0,X}^{-1}\right\| \leqslant \frac{1}{(1-\varepsilon)}$ and since

$$\left\|\tilde{D}_{1,X} D_{2,X}^{-1} D_{1,X}\right\| \leqslant 2\left\|\mathbf{G}_X^{\frac{1}{2}}\left((\Phi_X^{(1)})^*\Phi_X^{(1)}\right)^{-1}\mathbf{G}_X^{\frac{1}{2}}\right\| \left\|\mathbf{G}_X^{-\frac{1}{2}}(\Phi_X^{(1)})^*\Phi_X\right\|^2 \leqslant \frac{4\varepsilon^2}{1-\varepsilon},$$

S is invertible provided that $4\varepsilon^2 < (1-\varepsilon)^2$, which is true when $\varepsilon < 1/3$, and we have

$$\left\|S^{-1}\right\| \leqslant \frac{\left\|D_{0,X}^{-1}\right\|}{1 - \left\|\tilde{D}_{1,X} D_{2,X}^{-1} D_{1,X}\right\|} \leqslant \frac{1}{1-\varepsilon-4\varepsilon^2}.$$

Note that $\left\|\mathbf{G}_X^{-\frac{1}{2}}\tilde{D}_{1,X}\right\|, \left\|\mathbf{G}_X^{-\frac{1}{2}} D_{1,X}\right\| \leqslant \sqrt{2}\varepsilon$. Then, by combining the above bounds, we have

$$\left\|\mathbf{F}_X^{\frac{1}{2}} M_1(u,v)^{-1}\mathbf{F}_X^{\frac{1}{2}}\right\| \leqslant \frac{4}{1-\varepsilon-4\varepsilon^2}$$

□

In the following given a metric d on some space $\mathcal{Y}$, $x \in \mathcal{Y}$ and $r > 0$, the ball of radius r around x is denoted by $\mathcal{B}_d(x,r) \stackrel{\text{def.}}{=} \{x' \,;\, d(x',x) \leqslant r\}$.

Theorem E.2 (Quantitative implicit function theorem, adapted from [4]). *Let $F : \mathcal{H} \times \mathcal{Y} \to \mathbb{C}^n$ be a differentiable mapping where $\mathcal{H}$ is a Hilbert space, $\mathcal{Y} \subseteq \mathbb{C}^{2s} \times \mathbb{C}^{sd}$, $n = s(d+2)$, $\|\cdot\|$ be a norm on $\mathcal{H}$. For each $y \in \mathcal{Y}$, suppose that there exists a positive definite matrix $\mathbf{F}_y$, and let $d_{\mathbf{F}}$ be the associated metric. Let $x_0 \in \mathcal{H}$, $y_0 \in \mathcal{Y}$ and $r_1, r_2 > 0$ be such that $F(x_0,y_0) = 0$ and for $x \in \mathcal{B}_{\|\cdot\|}(x_0,r_1), y \in \mathcal{B}_{d_{\mathbf{F}}}(y_0,r_2)$, $\partial_y F(x,y)$ is invertible and*

$$\left\|\mathbf{F}_y^{-\frac{1}{2}}\partial_x F(x,y)\right\| \leqslant D_1 \quad \text{and} \quad \left\|\mathbf{F}_y^{\frac{1}{2}}\partial_y F(x,y)^{-1}\mathbf{F}_x^{\frac{1}{2}}\right\| \leqslant D_2\,.$$

Then, defining $R = \min\left(\frac{r_2}{D_1 D_2}, r_1\right)$, there exists a unique Fréchet- differentiable mapping $g : \mathcal{B}_{\|\cdot\|}(x_0,R) \to \mathcal{B}_{d_{\mathbf{F}}}(y_0,r_2)$ such that $g(x_0) = y_0$ and for all $x \in \mathcal{B}_{\|\cdot\|}(x_0,R)$, $F(x,g(x)) = 0$. Furthermore

$$\mathrm{d}g(x) = -(\partial_y F(x,g(x)))^{-1}\partial_x F(x,g(x))$$

and consequently $\left\|\mathbf{F}_{g(x)}^{\frac{1}{2}}\mathrm{d}g(x)\right\| \leqslant D_1 D_2$.

Proof. Let $V^* = \cup_{V\in\mathcal{V}} V$, where $\mathcal{V}$ is the collection of all open sets V of $\mathcal{H}$ such that

1. $x_0 \in V$,
2. V is star-shaped with respect to x_0,
3. $V \subset \mathcal{B}_{\|\cdot\|}(x_0,r_1)$,
4. there exists a $\mathscr{C}^1$ function $g : V \to \mathcal{B}_{d_{\mathbf{F}}}(y_0,r_2)$ such that $g(x_0) = y_0$ and $F(x,g(x)) = 0$ for all $x \in V$.

Observe that $\mathcal{V}$ is non-empty by the (classical) Implicit Function Theorem. Moreover, $\mathcal{V}$ is stable by union: indeed, all conditions expect the last one are easy to check. Now, let $V, \tilde{V} \in \mathcal{V}$ and $g, \tilde{g}$ be corresponding functions. The set $\overline{V} = \{x \in V \cap \tilde{V},\ g(x) = \tilde{g}(x)\}$ is non-empty (it contains x_0), and closed in $V \cap \tilde{V}$. Moreover, it is open: for any $x \in \overline{V}$, by our assumptions $\partial_y F(x, g(x))$ is invertible and the Implicit Function theorem applies at $(x, g(x))$, and by the uniqueness of the mapping resulting from it we obtain an open set around x in which g and $\tilde{g}$ coincide. Hence $\overline{V}$ is both closed and open in $V \cap \tilde{V}$, and by the connectedness of it $\overline{V} = V \cap \tilde{V}$. Therefore, there exists a function g' defined on $V \cup \tilde{V}$ that satisfies condition 4 above (it is defined as g on V and $\tilde{g}$ on $\tilde{V}$, which is well-posed for their intersection), and $\mathcal{V}$ is indeed stable by union.

Hence $V^* \in \mathcal{V}$, let g^* be its corresponding function. It is unique by the arguments above, satisfies $F(x, g^*(x)) = 0$ and

$$\begin{aligned}\mathbf{F}_{g^*(x)}^{\frac{1}{2}} \mathrm{d}g^*(x) &= -\mathbf{F}_{g^*(x)}^{\frac{1}{2}} (\partial_y F(x, g^*(x)))^{-1} \partial_x F(x, g^*(x)) \\ &= -(\mathbf{F}_{g^*(x)}^{-\frac{1}{2}} \partial_y F(x, g^*(x)) \mathbf{F}_{g^*(x)}^{-\frac{1}{2}})^{-1} \mathbf{F}_{g^*(x)}^{-\frac{1}{2}} \partial_x F(x, g^*(x))\end{aligned}$$

for all $x \in V^*$. Note that by our assumptions $\left\| \mathbf{F}_{g^*(x)}^{\frac{1}{2}} \mathrm{d}g^*(x) \right\| \leqslant D_1 D_2$.

We finish the proof by showing that V^* contains a ball of radius $r_2/(D_1 D_2)$. Let $x \in \mathcal{H}$ with $\|x\| = 1$, $R_x = \sup\{R\ ;\ x_0 + Rx \in V^*\}$, and $x^* = x_0 + R_x x \in \partial V^*$. Clearly $0 < R_x \leqslant r_1$ since V^* is open, assume $R_x < r_1$. Our goal is to show that in that case $R_x \geqslant \frac{r_1}{D_1 D_2}$. Since $\mathrm{d}g^*$ is bounded, g^* is uniformly continuous on V^* and it can be extended on ∂V^*, and by continuity $F(x^*, g^*(x^*)) = 0$. By contradiction, if $g^*(x^*) \in \mathcal{B}_{d_{\mathbf{F}}}(y_0, r_2)$, by our assumptions we can apply the Implicit Function Theorem at $(x^*, g^*(x^*))$, and therefore extend g^* on an open set V that is not included in V^* such that $V \cup V^* \in \mathcal{V}$, which contradicts the maximality of V^*. Hence $d_{\mathbf{F}}(g^*(x^*), y_0) = r_2$. Let $\gamma : [0, 1] \to \mathcal{Y}$ be defined by $\gamma(t) \stackrel{\text{def.}}{=} g^*(x^* + t(x_0 - x^*))$, so $\gamma'(t) = \mathrm{d}g^*(\gamma(t))(x_0 - x^*)$. Then,

$$\begin{aligned} r_2 = d_{\mathbf{F}}(g^*(x^*), g^*(x_0)) &\leqslant \sqrt{\int_0^1 \langle \mathbf{F}_{g^*(\gamma(t))} \gamma'(t),\ \gamma'(t) \rangle \mathrm{d}t} \\ &= \sqrt{\int_0^1 \left\| \mathbf{F}_{g^*(\gamma(t))}^{\frac{1}{2}} \mathrm{d}g^*(\gamma(t))(x_0 - x^*) \right\|^2 \mathrm{d}t} \leqslant D_1 D_2 R_x. \end{aligned}$$

□

E.0.3 Proof of Theorem E.1

Our goal is to apply Theorem E.2.

Since $\left\| \mathrm{Id} - \hat{\Upsilon}_X \right\| \leqslant \frac{1}{8}$ by Lemma D.1, and by applying Lemma E.4, we have that $\left\| \left(\mathbf{F}_X^{-1/2} M_1(u, v) \mathbf{F}_X^{-1/2} \right)^{-1} \right\| \leqslant 5$. From Lemma E.3, under event $\bar{E}$, by taking

$$c \stackrel{\text{def.}}{=} \frac{\min_i |a_{0,i}|}{16 \bar{L}_{12}} \tag{E.13}$$

for all $X \in \mathcal{X}^s$, $a \in \mathbb{C}^s$ and $w \in \mathbb{C}^m$ such that

$$\lambda \leqslant \frac{\min_i |a_{0,i}|}{4}, \quad \|a - a_0\| \leqslant \frac{c}{4\bar{L}_0}, \quad \|w\| \leqslant \frac{c}{4} \quad \text{and} \quad d_{\mathbf{H}}(X, X_0) \leqslant \min\left(r_{\text{near}}, \frac{c}{4\bar{L}_1 \|a_0\|} \right),$$

we have $\left\| \mathbf{F}_X^{-1/2} M_2(u, v) \mathbf{F}_X^{-1/2} \right\| \leqslant \frac{1}{8}$.

In this case, $\partial_u f(u, v)$ is invertible, and we have

$$\left\| (\mathbf{F}_X^{-\frac{1}{2}} \partial_u f(u, v) \mathbf{F}_X^{-\frac{1}{2}})^{-1} \right\| \lesssim \frac{1}{\min_i |a_{0,i}|}$$

since $\|a - a_0\| \lesssim \min_i |a_{0,i}|$ by assumption.

Therefore we can apply Theorem E.2 with (recalling the definition of c in (E.13) and the bound (E.11)) with $\mathcal{H} = \mathbb{R}_+ \times \mathbb{R}^{2m}$,

$$r_1 = c,\; D_1 = \mathcal{O}\left(\sqrt{s}\right),\; r_2 = \mathcal{O}\left(\min\left(r_{\text{near}},\, \frac{c}{\bar{L}_1\|a_0\|}, \frac{c}{\bar{L}_0}, \frac{1}{C_{\mathbf{H}}B}\right)\right),\; D_2 = \mathcal{O}\left(\frac{1}{\min_i |a_{0,i}|}\right)$$

with $B = \sum_{i+j\leqslant 3} B_{ij}$, we obtain that $g(v)$ is defined for $v \in V \overset{\text{def.}}{=} \mathcal{B}_{\|\cdot\|_2}(0, r)$ with

$$r \overset{\text{def.}}{=} \min\left(\frac{r_2}{D_1 D_2}, r_1\right) = \frac{r_2}{D_1 D_2} = \mathcal{O}\left(\min\left(\frac{r_{\text{near}}}{\sqrt{s}\min_i|a_{0,i}|},\; \frac{1}{\sqrt{s}\bar{L}_1\bar{L}_{12}\|a_0\|}, \frac{1}{\sqrt{s}\bar{L}_{12}\bar{L}_0}, \frac{1}{\sqrt{s}\min_i|a_{0,i}|C_{\mathbf{H}}B}\right)\right)$$

such that g is $\mathcal{C}^1$, $f(g(v), v) = 0$, $g(v_0) = u_0$, where we recall that $u_0 = (a_0, X_0)$ and $v_0 = (0, 0)$.

Finally, from Theorem E.2 we also have that

$$\left\|\mathbf{F}_X^{\frac{1}{2}} \mathrm{d}g(v)\right\| \leqslant D_1 D_2 \lesssim \frac{\sqrt{s}}{\min_i |a_{0,i}|}$$

and by defining $\gamma(t) = g(v_0 + t(v - v_0))$ for $t \in [0, 1]$, we have the following error bound between $u = g(v)$ and $u_0 = g(v_0)$:

$$\begin{aligned}
d_{\mathbf{F}}(u, u_0) = \sqrt{\|a - a_0\|_2^2 + d_{\mathbf{H}}(X, X_0)^2} &\leqslant \sqrt{\int_0^1 \langle \mathbf{F}_{\gamma(t)}\gamma'(t),\, \gamma'(t)\rangle \mathrm{d}t} \\
&= \sqrt{\int_0^1 \langle \mathbf{F}_{\gamma(t)}\mathrm{d}g(tv)v,\, \mathrm{d}g(tv)v\rangle \mathrm{d}t} \\
&\leqslant \frac{\sqrt{s}}{\min_i |a_{0,i}|}\, \|v\|\,.
\end{aligned}$$

F Examples

F.1 Jackson kernel

Let $f \in \mathbb{N}$ and $\mathcal{X} \in \mathbb{T}^d$ the d-dimensional torus. We consider the Jackson kernel

$$K(x, x') = \prod_{i=1}^{d} \kappa(x_i - x'_i),$$

where $\kappa(x) \overset{\text{def.}}{=} \left(\frac{\sin\left(\left(\frac{f}{2}+1\right)\pi x\right)}{\left(\frac{f}{2}+1\right)\sin(\pi x)}\right)^4$, with constant metric tensor

$$\mathbf{H}_x = C_f \mathrm{Id} \quad \text{and} \quad d_{\mathbf{H}}(x, x') = C_f^{\frac{1}{2}} \|x - x'\|_2\,.$$

where $C_f \overset{\text{def.}}{=} -\kappa''(0) = \frac{\pi^2}{3} f(f+4) \sim f^2$. Note that $K^{(ij)} = C_f^{-(i+j)/2} \nabla_1^i \nabla_2^j K$ and since the metric is constant, we can set $C_{\mathbf{H}} \overset{\text{def.}}{=} 0$.

F.1.1 Discrete Fourier sampling

A random feature expansion associated with the Jackson kernel is obtained by choosing $\Omega = \left\{\omega \in \mathbb{Z}^d\,;\; \|\omega\|_\infty \leqslant f\right\}$, $\varphi_\omega(x) \overset{\text{def.}}{=} e^{\mathrm{i}2\pi\omega^\top x}$, and $\Lambda(\omega) = \prod_{j=1}^d g(\omega_j)$ where $g(j) = \frac{1}{f}\sum_{k=\max(j-f,-f)}^{\min(j+f,f)} (1 - |k/f|)(1 - |(j-k)/f|)$. Note that this corresponds to sampling *discrete* Fourier frequencies. In this case, the derivatives of the random features are uniformly bounded with $\left\|\nabla^j \varphi_\omega(x)\right\| = \|\omega\|^j = \mathcal{O}(C_f^{j/2} d^{j/2})$. So, we can set $\bar{L}_i = \mathcal{O}(d^{i/2})$.

F.1.2 Admissibility of the kernel

Theorem F.1. *Suppose that $f \geqslant 128$. Then, K is an admissible kernel with $r_{\text{near}} = 1/(8\sqrt{2})$, $\varepsilon_2 = 0.941$, $\varepsilon_0 = 0.00097$, $h = \mathcal{O}(d^{-1/2})$ and $\Delta = \mathcal{O}(d^{1/2} s_{\max}^{1/4})$, $B_{00} = B_{11} = B_{20} = \mathcal{O}(1)$, $B_{01} = \mathcal{O}(d^{1/2})$ and $B_{22} = \mathcal{O}(d)$.*

The remainder of this section is dedicated to proving this theorem. The uniform bounds on B_{ij} are due to Lemma F.4 (uniform bounds), and the bound on Δ and h are due to Lemma F.3. From Lemma F.1, we see that by setting $r_{\text{near}} \stackrel{\text{def.}}{=} \frac{1}{8\sqrt{2}}$, for all $d_{\mathbf{H}}(x, x') \leqslant r_{\text{near}}$, $K^{(20)}(x, x') \prec -\varepsilon_2 \text{Id}$ with $\varepsilon_2 = (1 - 6r_{\text{near}}^2)(1 - r_{\text{near}}^2/(2 - r_{\text{near}}^2) - r_{\text{near}}^2) \geqslant 0.941$. Finally, from Lemma F.2, we have that for for all $d_{\mathbf{H}}(x, x') \geqslant r_{\text{near}}$, $|K| \leqslant 1 - 1/(8^3 \cdot 2)$, so we can set $\varepsilon_0 \stackrel{\text{def.}}{=} 0.00097$.

Before proving these lemmas, we first summarise in Section F.1.3 some key properties of the univariate Jackson kernel κ when $f \geqslant 128$ which were derived in [2].

For notational convenience, write $t_i \stackrel{\text{def.}}{=} x_i - x_i'$, $\kappa_i \stackrel{\text{def.}}{=} \kappa(t_i)$, $\kappa_i' \stackrel{\text{def.}}{=} \kappa'(t_i)$, and so on. Let

$$K_i \stackrel{\text{def.}}{=} \prod_{\substack{k=1 \\ k \neq i}}^{d} \kappa_k, \quad K_{ij} \stackrel{\text{def.}}{=} \prod_{\substack{k=1 \\ k \neq i,j}}^{d} \kappa_k \quad \text{and} \quad K_{ij\ell} \stackrel{\text{def.}}{=} \prod_{\substack{k=1 \\ k \neq i,j,\ell}}^{d} \kappa_k.$$

With this, we have:

$$\begin{aligned} \partial_{1,i} K(x, x') &= \kappa_i' K_i \\ \partial_{1,i}\partial_{2,i} K(x, x') &= -\kappa_i'' K_i, \quad \text{and} \quad \forall i \neq j,\ \partial_{1,i}\partial_{2,j} K(x, x') = -\kappa_i' \kappa_j' K_{ij}. \end{aligned}$$

Where convenient, we sometimes write $K(t) = K(x - x') \stackrel{\text{def.}}{=} K(x, x')$.

F.1.3 Properties of κ

From [2, Equations (2.20)-(2.24) and (2.29)], for all $t \in [-1/2, 1/2]$ and $\ell = 0, 1, 2, 3$:

$$\begin{aligned} 1 - \frac{C_f}{2} t^2 \leqslant \kappa(t) \leqslant 1 - \frac{C_f}{2} t^2 + 8 \left(\frac{1 + 2/f}{1 + 2/(2+f)} \right)^2 C_f^2 t^4 \leqslant 1 - \frac{C_f}{2} t^2 + 8 C_f^2 t^4 \\ |\kappa'(t)| \leqslant C_f t, \quad |\kappa''(t)| \leqslant C_f, \quad |\kappa'''(t)| \leqslant 3 \left(\frac{1 + 2/f}{1 + 2/(2+f)} \right)^2 C_f^2 t \leqslant 12 C_f^2 t \\ \kappa'' \leqslant -C_f + \frac{3}{2} \left(\frac{1 + 2/f}{1 + 2/(2+f)} \right)^2 C_f^2 t^2 \leqslant -C_f + 6 C_f^2 t^2. \end{aligned} \tag{F.1}$$

By [2, Lemma 2.6],

$$\left| \kappa^{(\ell)}(t) \right| \leqslant \begin{cases} \frac{\pi^\ell H_\ell(t)}{(f+2)^{4-\ell} t^4}, & t \in [\frac{1}{2f}, \frac{\sqrt{2}}{\pi}] \\ \frac{\pi^\ell H_\ell^\infty}{(f+2)^{4-\ell} t^4}, & t \in [\frac{\sqrt{2}}{\pi}, \frac{1}{2}), \end{cases}$$

where $H_0^\infty \stackrel{\text{def.}}{=} 1$, $H_1^\infty \stackrel{\text{def.}}{=} 4$, $H_2^\infty \stackrel{\text{def.}}{=} 18$ and $H_3^\infty \stackrel{\text{def.}}{=} 77$, and $H_\ell(t) \stackrel{\text{def.}}{=} \alpha^4(t) \beta_\ell(t)$, with

$$\alpha(t) \stackrel{\text{def.}}{=} \frac{2}{\pi(1 - \frac{\pi^2 t^2}{6})}, \quad \bar{\beta}(t) \stackrel{\text{def.}}{=} \frac{\alpha(t)}{ft} = \frac{2}{ft\pi(1 - \pi^2 t^2/6)}$$

and $\beta_0(t) \stackrel{\text{def.}}{=} 1$, $\beta_1(t) \stackrel{\text{def.}}{=} 2 + 2\bar{\beta}(t)$, $\beta_2 \stackrel{\text{def.}}{=} 4 + 7\bar{\beta}(t) + 6\bar{\beta}(t)^2$ and $\beta_3(t) \stackrel{\text{def.}}{=} 8 + 24\bar{\beta} + 30\bar{\beta}(t)^2 + 15\bar{\beta}(t)^3$. Let us first remark that $\bar{\beta}$ is decreasing on $I \stackrel{\text{def.}}{=} [\frac{1}{2f}, \frac{\sqrt{2}}{\pi}]$, so $\left|\bar{\beta}(t)\right| \leqslant \left|\bar{\beta}(1/(2f))\right| \approx 1.2733$, and $a(t) \leqslant a(\sqrt{2}/\pi) = \frac{3}{\pi}$

on I. Therefore, on I, $H_0(t) \leqslant \frac{3}{\pi}$, $H_1(t) \leqslant 3.79$, $H_2(t) \leqslant 18.83$ and $H_3(t) \leqslant 98.26$, and we can conclude that on $[\frac{1}{2f}, \frac{1}{2})$, we have

$$\left|\kappa^{(\ell)}(t)\right| \leqslant \frac{\pi^\ell \bar{H}_\ell^\infty}{(f+2)^{4-\ell}t^4}$$

where $\bar{H}_0^\infty = 1$, $\bar{H}_1^\infty \stackrel{\text{def.}}{=} 4$, $\bar{H}_2^\infty \stackrel{\text{def.}}{=} 19$, $\bar{H}_3^\infty \stackrel{\text{def.}}{=} 99$. Combining with (F.1), we have $\left\|\kappa^{(\ell)}\right\|_\infty \leqslant \kappa_\ell^\infty$ where $\kappa_0^\infty \stackrel{\text{def.}}{=} 1$, $\kappa_2^\infty \stackrel{\text{def.}}{=} C_f$,

$$\kappa_1^\infty \stackrel{\text{def.}}{=} \sqrt{C_f} \max\left(\frac{2\pi^4}{(\frac{1}{2}+\frac{1}{f})^3}\frac{f}{\sqrt{C_f}}, \frac{\sqrt{C_f}}{2f}\right) = \mathcal{O}(\sqrt{C_f})$$

$$\kappa_3^\infty \stackrel{\text{def.}}{=} (C_f)^{3/2} \max\left(\frac{99\pi^3}{(\frac{1}{2}+\frac{1}{f})}\left(\frac{2f}{\sqrt{C_f}}\right)^4, \frac{6\sqrt{C_f}}{f}\right) = \mathcal{O}((C_f)^{3/2}).$$

Finally, given $p \in (0,1)$,

$$(f+2)^4 t^4 \geqslant (1 + p(f+2)^2 t^2)^2, \qquad \forall\, t \geqslant \frac{1}{\sqrt{(1-p)}(f+2)}.$$

Choosing $p = \frac{1}{2}$ and using $(f+2)^2 = (\frac{3}{\pi^2}C_f + 4) \geqslant \frac{3}{\pi^2}C_f$, we have

$$\left|\kappa^{(\ell)}(t)\right| \leqslant \frac{\kappa_\ell^\infty}{(1+\frac{3}{2\pi^2}C_f t^2)^2}, \qquad \forall\, t^2 \geqslant \frac{2\pi^2}{3C_f}, \tag{F.2}$$

F.1.4 Bounds in neighbourhood of $x' = x$

Lemma F.1. *Suppose that $C_f \|t\|_2^2 \leqslant c$ with $c > 0$ such that*

$$\varepsilon \stackrel{\text{def.}}{=} (1-6c)\left(1 - \frac{c}{2-c}\right) - c > 0$$

Then, $\hat{K}^{02}(t) \preceq -\varepsilon \mathrm{Id}$.

Proof. We need to show that $\lambda_{\min}(-K^{(02)}(t)) \geqslant b$. Let $q \in \mathbb{R}^d$, and note that

$$\begin{aligned} -\langle \nabla_2^2 K q,\, q\rangle &= -\sum_i \left(q_i \kappa_i'' K_i - \kappa_i' \sum_{j\neq i} q_j \kappa_j' K_{ij}\right) q_i \\ &= -\left(\sum_i q_i^2 \kappa_i'' K_i - \sum_i q_i \kappa_i \sum_{j\neq i} q_j \kappa_j K_{ij}\right) \qquad \text{(F.3)} \\ &\geqslant \|q\|^2 \left(-\max_i\{\kappa_i'' K_i\} - \sum_j \left|\kappa_j'\right|^2\right). \end{aligned}$$

We first consider $\kappa_i'' K_i$:

$$\begin{aligned} &\kappa_i'' \leqslant -C_f + 6C_f^2 t_i^2, \\ &K_i \geqslant \prod_{j\neq i}\left(1 - \frac{C_f}{2}t_i^2\right) \geqslant 1 - \frac{C_f}{2}\|t\|_2^2 - \left(\frac{C_f}{2}\|t\|_2^2\right)^3 - \left(\frac{C_f}{2}\|t\|_2^2\right)^5 - \cdots \\ &\qquad \geqslant 1 - \frac{C_f \|t\|_2^2}{2(1-\frac{C_f}{2}\|t\|_2^2)}. \end{aligned}$$

and hence,

$$\kappa_i'' K_i \leqslant \left(-C_f + 6C_f^2 \left\|t\right\|_2^2\right) \left(1 - \frac{C_f \left\|t\right\|_2^2}{2(1 - \frac{C_f}{2} \left\|t\right\|_2^2)}\right)$$

For the second term,

$$\sum_j \left|\kappa_j'\right|^2 \leqslant C_f^2 \left\|t\right\|_2^2.$$

Therefore,

$$\lambda_{\min}(-K^{(02)}(t)) \geqslant \left(1 - 6C_f \left\|t\right\|_2^2\right) \left(1 - \frac{C_f \left\|t\right\|_2^2}{2(1 - \frac{C_f}{2} \left\|t\right\|_2^2)}\right) - C_f \left\|t\right\|_2^2$$

□

Lemma F.2. *Assume that* $\frac{1}{8\sqrt{C_f}} \geqslant \left\|t\right\|_2$ *Then,*

$$K(t) \leqslant 1 - \frac{C_f}{4} \left\|t\right\|_2^2 + 16C_f^2 \left\|t\right\|_2^4.$$

Consequently, for all

$$0 < c \leqslant \frac{1}{8\sqrt{2C_f}},$$

and all t *such that* $\left\|t\right\|_2 \geqslant c$*,*

$$|K(t)| \leqslant 1 - \frac{C_f}{8} c^2.$$

Proof. First note that

$$|\kappa(u)| \leqslant 1 - \frac{C_f}{2} u^2 + 32C_f^2 u^4 = 1 - u^2 g(u)$$

where

$$g(u) \stackrel{\text{def.}}{=} C_f \left(\frac{1}{2} - 32C_f u^2\right),$$

and note that $g(u) \in (0, \frac{C_f}{2})$ for $u \in (0, 1/(8\sqrt{C_f})$. So, writing $t = (t_i)_{i=1}^d$ and $g_j \stackrel{\text{def.}}{=} g(t_j)$, we have

$$\begin{aligned} K(t) &= \prod_{j=1}^d \kappa(t_i) \leqslant \prod_{j=1}^d \left(1 - t_j^2 \cdot g(t_j)\right) \\ &= 1 - \sum_{j=1}^d t_j^2 g_j + \sum_{j \neq k} t_j^2 t_k^2 g_j g_k - \sum_{j \neq k \neq \ell} t_j^2 t_k^2 t_\ell^2 g_j g_k g_\ell + \cdots \end{aligned}$$

Note that

$$\begin{aligned} &- \sum_{j \neq k \neq \ell} t_j^2 t_k^2 t_\ell^2 \cdot g_j g_k g_\ell + \sum_{j \neq k \neq \ell \neq n} t_j^2 t_k^2 t_\ell^2 t_n^2 \cdot g_j g_k g_\ell g_n \\ &\leqslant - \sum_{j \neq k \neq \ell} t_j^2 t_k^2 t_\ell^2 \cdot g_j g_k g_\ell + \left(\sum_{j \neq k \neq \ell} t_j^2 t_k^2 t_\ell^2 \cdot g_j g_k g_\ell\right) \left(\sum_n t_n^2 g_n\right) \\ &\leqslant - \sum_{j \neq k \neq \ell} t_j^2 t_k^2 t_\ell^2 \cdot g_j g_k g_\ell \left(1 - \frac{C_f}{2} \left\|t\right\|_2^2\right) < 0 \end{aligned}$$

since $\left(1-\frac{C_f}{2}\|t\|_2^2\right)>0$. Also,

$$\sum_{j=1}^d t_j^2 g_j \leqslant \frac{C_f}{2}\sum_{j=1}^d t_j^2 < 1,$$

by assumption. So,

$$\begin{aligned} K(t) &\leqslant 1-\sum_{j=1}^d t_j^2 g_j + \sum_{j\neq k} t_j^2 t_k^2 g_j g_k \\ &\leqslant 1-\sum_{j=1}^d t_j^2 g_j + \frac{1}{2}\left(\sum_j t_j^2 g_j\right)^2 \leqslant 1-\frac{1}{2}\sum_{j=1}^d t_j^2 g_j \\ &\leqslant 1-\frac{C_f}{2}\left(\frac{1}{2}\sum_{j=1}^d t_j^2 - 32C_f\sum_{j=1}^d t_j^4\right) \leqslant 1-\frac{C_f}{4}\|t\|_2^2 + 16C_f^2\|t\|_2^4. \end{aligned}$$

Finally, observe that the function

$$q(z) \stackrel{\text{def.}}{=} \frac{C_f}{4}z^2 - 16C_f^2 z^4$$

is positive and increasing on the interval $[0,\frac{1}{8\sqrt{2C_f}}]$. So, for t satisfing

$$c \leqslant \|t\|_2 \leqslant \frac{1}{8\sqrt{2C_f}}, \tag{F.4}$$

we have $|K(t)| \leqslant 1-q(c) \leqslant 1-\frac{C_f}{8}c^2$. Finally, since $|K(t)|$ is decreasing as t increases, we trivially have that $|K(t)|\leqslant 1-q(c)$ for all t with $\|t\|_2 \geqslant c$.

□

F.1.5 Bounds under separation

Lemma F.3. *Let $i,j\in\{0,1,2\}$ with $i+j\leqslant 3$. Let $\bar{A}\geqslant\sqrt{\frac{4\pi^2}{3}}$ and $\|t\|_2\geqslant \bar{A}\sqrt{d}s_{\max}^{1/4}/\sqrt{C_f}$. Then, we have $\left\|K^{(ij)}(t)\right\|\leqslant d^{\frac{i+j-4}{2}}(\bar{A}^4 s_{\max})^{-1}$.*

Proof. Write $t=(t_j)_{j=1}^d$. To bound $K(t)=\prod_{j=1}^d \kappa(a_j)$, we want to make use of the form (F.2). We can do this for each t_j such that $|t_j|\geqslant\sqrt{\frac{2\pi^2}{3C_f}}$. Note that there exists at least one such t_j since $\|t\|_\infty \geqslant \|t\|_2/\sqrt{d} \geqslant \bar{A}s_{\max}^{1/4}/\sqrt{C_f} \geqslant \sqrt{\frac{2\pi^2}{3C_f}}$. If $\{|t_j|\}_{j=1}^k \subset [0,\sqrt{\frac{2\pi^2}{3C_f}})$ for $k\leqslant d-1$, then

$$k\frac{2\pi^2}{3C_f}+\sum_{j=k+1}^d t_j^2 \geqslant \|t\|_2^2 \geqslant \frac{\bar{A}^2 d s_{\max}^{1/2}}{C_f},$$

which implies that $\sum_{j=k+1}^d t_j^2 \geqslant \frac{1}{C_f}\left(\bar{A}^2 d s_{\max}^{1/2} - \frac{2\pi^2(d-1)}{3}\right) \geqslant \frac{\bar{A}^2 d s_{\max}^{1/2}}{2C_f}$, by our assumptions on $\bar{A}$. Therefore, we may assume that we have some $d\geqslant p\geqslant 1$ such that $\{b_j\}_{j=1}^p \subseteq \{t_j\}$ with $|b_j|\geqslant\sqrt{\frac{2\pi^2}{3C_f}}$ and $\|b\|_2 \geqslant \frac{\bar{A}\sqrt{d}\sqrt[4]{s_{\max}}}{\sqrt{2C_f}}$. Observe that

$$\prod_{j=1}^p\left(1+\frac{3C_f}{2\pi^2}b_j^2\right) \geqslant 1+\frac{3C_f}{2\pi^2}\sum_{j=1}^p b_j^2 = 1+\frac{3C_f}{2\pi^2}\|b\|_2^2 \geqslant 1+\frac{3}{4\pi^2}\bar{A}^2 d\sqrt{s_{\max}}.$$

So, by applying the fact that $|\kappa| \leqslant 1$, $\kappa_0^\infty = 1$ and (F.2), we have

$$|K(t)| \leqslant \prod_{j=1}^{p} |\kappa(b_j)| \leqslant \prod_{j=1}^{p} \frac{1}{\left(1 + \frac{3C_f}{2\pi^2} b_j^2\right)^2} \leqslant \frac{1}{\left(1 + \frac{3}{4\pi^2} \bar{A}^2 d \sqrt{s_{\max}}\right)^2}.$$

For $|\kappa_i' K_i|$, if $i \notin \left\{ j \; ; \; |t_j| > \sqrt{\frac{2\pi^2}{3C_f}} \right\}$, then

$$|\kappa_i' K_i| \leqslant \|\kappa_i'\|_\infty \prod_{j=1}^{p} |\kappa(b_j)| \leqslant \frac{\|\kappa_i'\|_\infty}{\left(1 + \frac{3}{4\pi^2} \bar{A}^2 d \sqrt{s_{\max}}\right)^2},$$

and otherwise, we have $|\kappa_i' K_i| \leqslant |\kappa'(t_i)| \prod_{j \neq i} |\kappa(b_j)| \leqslant \frac{\kappa_1^\infty}{\left(1 + \frac{3}{4\pi^2} \bar{A}^2 d \sqrt{s_{\max}}\right)^2}$, In a similar manner, writing $V \stackrel{\text{def.}}{=} \left(1 + \frac{3}{4\pi^2} \bar{A}^2 d \sqrt{s_{\max}}\right)^{-2}$, we can deduce that

$$\begin{aligned}
&|\kappa_i' K_i| \leqslant \kappa_1^{\max} V, \qquad |\kappa_i'' K_i| \leqslant \kappa_2^{\max} V, \qquad \left|\kappa_i' \kappa_j' K_{ij}\right|^2 \leqslant (\kappa_1^{\max})^2 V \\
&\left|\kappa_i''' K_i\right|^3 \leqslant \kappa_3^{\max} V, \qquad \left|\kappa_i'' \kappa_j' K_{ij}\right|^3 \leqslant \kappa_2^{\max} \kappa_1^{\max} V, \qquad \left|\kappa_i' \kappa_j' \kappa_\ell' K_{ij\ell}\right| \leqslant (\kappa_1^{\max})^3 V.
\end{aligned}$$

Therefore,

$$\left\|K^{(10)}\right\| = \frac{1}{\sqrt{C_f}} \|\nabla_1 K\| \leqslant \frac{1}{\sqrt{C_f}} \sqrt{\sum_{j=1}^{d} \left|\kappa_j' K_j\right|^2} \leqslant \frac{\kappa_1^\infty}{\sqrt{C_f}} V \sqrt{d} \lesssim \frac{1}{\bar{A}^4 d^{3/2} s_{\max}}.$$

Using Gershgorin theorem, we have

$$\left\|\nabla_2^2 K(x, x')\right\| \leqslant \max_{1 \leqslant i \leqslant d} \{|\kappa_i'' K_i| + |\kappa_i'| \sum_{j \neq i} \left|\kappa_j'\right| |K_{ij}|\}$$

and hence,

$$\begin{aligned}
\left\|K^{(02)}\right\| &= \frac{1}{C_f} \left\|\nabla_2^2 K\right\| \leqslant \frac{1}{C_f} \max_{i=1}^{d} \{|\kappa_i'' K_i| + |\kappa_i'| \sum_{j \neq i} \left|\kappa_j' K_{ij}\right|\} \\
&\leqslant \frac{1}{C_f} V \left(\kappa_2^{\max} + (\kappa_1^{\max})^2 (d-1)\right) \leqslant \frac{\max\{\kappa_2^\infty, (\kappa_1^\infty)^2\}}{C_f} V d \lesssim \frac{1}{\bar{A}^4 d s_{\max}}.
\end{aligned}$$

Note also that $\left\|K^{(11)}\right\| = \left\|K^{(02)}\right\|$. Finally, since

$$\begin{aligned}
\left\|\partial_{1,i} \nabla_2^2 K(x, x')\right\| \leqslant \max \Bigg\{ & |\kappa_i''' K_i| + |\kappa_i''| \sum_{j \neq i} \left|\kappa_j'\right| |K_{ij}|, \\
& \max_{j \neq i} \{\left|\kappa_j'' \kappa_i' K_{ij}\right| + \left|\kappa_j' \kappa_i'' K_{ij}\right| + |\kappa_i'| \left|\kappa_j'\right| \sum_{l \neq i,j} |\kappa_l'| |K_{ij\ell}|\} \Bigg\},
\end{aligned}$$

we have

$$\begin{aligned}
\left\|K^{(12)}\right\| &= \frac{1}{C_f^{3/2}} \left\|\nabla_1 \nabla_2^2 K\right\| \\
&\leqslant \frac{1}{C_f^{3/2}} \sqrt{d} V \max\left(\kappa_3^{\max} + \kappa_2^{\max} \kappa_1^{\max} (d-1), 2\kappa_2^{\max} \kappa_1^\infty + (d-1)(\kappa_1^\infty)^3\right) \\
&\leqslant d^{3/2} \max\{\kappa_3^\infty, \kappa_1^\infty \kappa_2^\infty, (\kappa_1^\infty)^3\} \frac{1}{C_f^{3/2}} V \lesssim \frac{1}{\bar{A}^4 d^{1/2} s_{\max}}
\end{aligned}$$

□

F.1.6 Uniform bounds

Lemma F.4. *If* $r_{\text{near}} \sim 1/\sqrt{C_f}$*, then* $B_0 = \mathcal{O}(1)$*,* $B_{01} = \mathcal{O}(\sqrt{d})$*,* $B_{02} = B_{12} = B_{11} = \mathcal{O}(1)$ *and* $B_{22} = \mathcal{O}(d)$*.*

Proof. We have $|K| \leqslant 1$, and

$$\|\nabla K\|^2 \leqslant \sum_i |\kappa_i|^2 |K_i|^2 \leqslant d(\kappa_1^\infty)^2 \lesssim C_f d,$$

so $B_{01} = \mathcal{O}(\sqrt{d})$.

From (F.3), for all $\|q\| = 1$,

$$\langle \nabla_2^2 K(t) q,\, q\rangle \leqslant \max_i |\kappa_i''| \, \|q\|_2^2 + \|q\|_2^2 \sum_i |\kappa_i|^2 \leqslant C_f + C_f^2 \|t\|^2 = \mathcal{O}(C_f),$$

for $\|t\| \lesssim 1/\sqrt{C_f}$. So, since $r_{\text{near}} \leqslant 2/\sqrt{C_f}$, $\left\|K^{02}(t)\right\| \leqslant 2 \overset{\text{def.}}{=} B_{02}$. The norm bound for K^{11} is the same.

$$\begin{aligned}
\left\|K^{(12)}\right\| &= \sup_{\|q\|=\|p\|=1} \frac{1}{C_f^{3/2}} \Bigg(\sum_k \sum_{k\neq i} \partial_{1,i} \left(\partial_{2,k}^2 K p_i q_k^2 + \partial_{1,i}\partial_{2,i}\partial_{2,k} K p_i q_i q_k\right) \\
&\quad + \sum_i \sum_k \sum_j \partial_{1,i}\partial_{2,j}\partial_{2,k} p_i p_j p_k + \sum_i \sum_{j\neq i} \partial_{1,i}\partial_{2,i}\partial_{2,j} K p_i q_i q_j + \sum_i \partial_{1,i}\partial_{2,j}^2 K p_i q_i^2 \Bigg) \\
&= \sup_{\|q\|=\|p\|=1} \frac{1}{C_f^{3/2}} \Bigg(\sum_k \sum_{k\neq i} \kappa_i' \kappa_k'' K_{ik} p_i q_k^2 + \kappa_i'' \kappa_k' K_{ik} p_i q_i q_k \\
&\quad + \sum_i \sum_k \sum_j \kappa_i' \kappa_k' \kappa_j' K_{ijk} p_i p_j p_k + \sum_i \sum_{j\neq i} \kappa_i'' \kappa_j' K_{ij} p_i q_i q_j + \sum_i \kappa_i' \kappa_j'' K_{ij} p_i q_i^2 \Bigg) \\
&\leqslant \frac{1}{C_f^{3/2}} \left(3 \|\kappa''\|_\infty \sqrt{\sum_i |\kappa_k'|^2} + \left(\sum_i |\kappa_k'|^2\right)^{3/2} + \|\kappa'\|_\infty \|\kappa''\|_\infty \right) \\
&\leqslant \frac{1}{C_f^{3/2}} \left(3 C_f^2 \|t\| + C_f^3 \|t\|^3 + \mathcal{O}(C_f^{3/2}) \right) = \mathcal{O}(1)
\end{aligned}$$

for $\|t\| \leqslant 1/C_f^{1/2}$.

We finally consider $K^{(22)}(x,x)$: for $\|p\| = 1$,

$$\begin{aligned}
\sum_i \sum_k \sum_j \partial_{1,k}\partial_{1,i}\partial_{2,j}\partial_{2,i} K p_j p_k &= \sum_i \sum_{k\neq i} \kappa_i'' \kappa_k'' p_j^2 K_{ik} + \sum_i \sum_{k\neq i} \kappa_i''' \kappa_k' p_i p_k K_{ik} \\
&\quad + \sum_i \sum_k \sum_j \kappa_i'' \kappa_j' \kappa_k' K_{ijk} p_j p_k + \sum_i \sum_j \kappa_i''' \kappa_j' p_j p_i K_{ij} + \sum_i \kappa_i'''' p_i^2 K_i \\
&= \sum_i \sum_{k\neq i} \kappa_i'' \kappa_k'' p_j^2 K_{ik} + \sum_i \kappa_i'''' p_i^2 \\
&= d\mathcal{O}(C_f^2)
\end{aligned}$$

since $\kappa'(0) = \kappa'''(0) = 0$ and $|\kappa''(0)| = \mathcal{O}(C_f)$, $|\kappa''''(0)| = \mathcal{O}(C_f^2)$. So, $B_{22} = \mathcal{O}(d)$. $\square$

F.2 The Gaussian kernel

We consider the Gaussian kernel $K(x,x') = \exp\left(-\frac{1}{2}\left\|x-x'\right\|^2_{\Sigma^{-1}}\right)$ in $\mathbb{R}^d$. Note that K is translation invariant, so that $\mathbf{H}_x$ will be constant and equal to $-\nabla^2 K(x,x)$. For simplicity define $t = x - x'$, $\hat{K}_\Sigma(t) = \exp\left(-\frac{1}{2}\left\|t\right\|^2_{\Sigma^{-1}}\right)$ and for $u \in \mathbb{R}$, $\kappa(u) = \exp\left(-\frac{1}{2}u^2\right)$. Denote by $\{e_i\}$ the canonical basis of $\mathbb{R}^d$, and by $f_i = \Sigma^{-1}e_i$ the i^{th} row of Σ^{-1}. We have the following:

$$\begin{aligned}
\nabla \hat{K}_\Sigma(t) &= -\Sigma^{-1}t\hat{K}_\Sigma(t) \\
\nabla^2 \hat{K}_\Sigma(t) &= \left(-\Sigma^{-1} + \Sigma^{-1}tt^\top\Sigma^{-1}\right)\hat{K}_\Sigma(t) \\
\partial_{1,i}\nabla^2 \hat{K}_\Sigma(t) &= \left(\Sigma^{-1}tf_i^\top + f_i t^\top\Sigma^{-1} - (-\Sigma^{-1} + \Sigma^{-1}tt^\top\Sigma^{-1})(t^\top f_i)\right)\hat{K}_\Sigma(t)
\end{aligned}$$

Hence we have $\mathbf{H}_x = -\nabla^2\hat{K}_\Sigma(0) = \Sigma^{-1}$, and, defining $d_{\mathbf{H}}(x,x') = \left\|x-x'\right\|_{\Sigma^{-1}} = \left\|\Sigma^{-\frac{1}{2}}(x-x')\right\|$, we have $C_{\hat{K}} = 1, C_{\mathbf{H}} = 0$ (that is, the metric tensor of the kernel is constant, and $d_{\mathbf{H}}$ is defined as the corresponding normalized norm).

Then, we have

$$\begin{aligned}
\left\|K^{(10)}(x,x')\right\| = \left\|K^{(01)}(x,x')\right\| &= d_{\mathbf{H}}(x,x')\kappa(d_{\mathbf{H}}(x,x')) \\
\left\|K^{(02)}(x,x')\right\| = \left\|K^{(11)}(x,x')\right\| &\leqslant (d_{\mathbf{H}}(x,x')^2 + 1)\kappa(d_{\mathbf{H}}(x,x')) \\
K^{(02)}(x,x') &\preccurlyeq (d_{\mathbf{H}}(x,x')^2 - 1)\kappa(d_{\mathbf{H}}(x,x'))\mathrm{Id}
\end{aligned}$$

and for $q \in \mathbb{R}^d$ with $\|q\| = 1$, since

$$\sum_i (\Sigma^{\frac{1}{2}}\nabla\varphi_\omega)_i q_i = \nabla\varphi_\omega^\top(\Sigma^{\frac{1}{2}}q) = \sum_i \partial_i\varphi_\omega(q^\top\Sigma^{\frac{1}{2}}e_i)$$

we can write

$$K^{(12)}(x,x')q = \sum_{i=1}^d (q^\top\Sigma^{\frac{1}{2}}e_i)\Sigma^{\frac{1}{2}}\partial_{1,i}\nabla^2\hat{K}_\Sigma(t)\Sigma^{\frac{1}{2}}$$

Thus we examine each term in $\partial_{1,i}\nabla^2\hat{K}_\Sigma$. We have

$$\sum_i (q^\top\Sigma^{\frac{1}{2}}e_i)\Sigma^{\frac{1}{2}}\Sigma^{-1}tf_i^\top\Sigma^{\frac{1}{2}} = \Sigma^{-\frac{1}{2}}t\left(\sum_i q^\top\Sigma^{\frac{1}{2}}e_ie_i^\top\Sigma^{-\frac{1}{2}}\right) = \Sigma^{-\frac{1}{2}}tq^\top$$

and similarly $\sum_i (q^\top\Sigma^{\frac{1}{2}}e_i)\Sigma^{\frac{1}{2}}f_it^\top\Sigma^{-1}\Sigma^{\frac{1}{2}} = qt^\top\Sigma^{\frac{1}{2}}$. Then

$$\sum_i (q^\top\Sigma^{\frac{1}{2}}e_i)(t^\top\Sigma^{-1}e_i)\Sigma^{\frac{1}{2}}\Sigma^{-1}\Sigma^{\frac{1}{2}} = t^\top\Sigma^{-1}(\sum_i e_ie_i^\top)\Sigma^{\frac{1}{2}}q = (t^\top\Sigma^{\frac{1}{2}}q)\mathrm{Id}$$

and similarly $\sum_i\sum_i (q^\top\Sigma^{\frac{1}{2}}e_i)(t^\top\Sigma^{-1}e_i)\Sigma^{\frac{1}{2}}\Sigma^{-1}tt^\top\Sigma^{-1}\Sigma^{\frac{1}{2}} = (t^\top\Sigma^{\frac{1}{2}}q)\Sigma^{-\frac{1}{2}}tt^\top\Sigma^{-\frac{1}{2}}$.

Hence at the end of the day

$$\left\|K^{(12)}(x,x')\right\| \leqslant (3d_{\mathbf{H}}(x,x') + d_{\mathbf{H}}(x,x')^3)\kappa(d_{\mathbf{H}}(x,x'))$$

and this bound is automatically valid for $K^{(21)}$ as well.

Finally, note that

$$\left\|K^{(22)}(x,x)\right\| = \sup_{\|p\|\leqslant 1}\langle\Sigma^{1/2}\nabla_2\nabla_2\cdot\left(\Sigma^{1/2}K^{(2,0)}(x,x)p\right), p\rangle$$

where $\nabla_2\cdot$ is the divergence operator on the 2nd variable, and one can show that $\left\|K^{(22)}(x,x)\right\| = (d+1)$.

We are then going to use the fact that for any $q \geqslant 1$ the function $f(r) = r^q e^{-\frac{1}{2}r^2}$ defined on $\mathbb{R}_+$ is increasing on $[0, \sqrt{q}]$ and decreasing after, and its maximum value is $f(\sqrt{q}) = \left(\frac{q}{e}\right)^{q/2}$. Furthermore, it is easy to see that we have $f(r) = r^q e^{-r^2/2} \leqslant \left(\frac{2q}{2}\right)^{\frac{q}{2}} e^{-r^2/4}$ and therefore $f(r) \leqslant \varepsilon$ if $r \geqslant 2\left(\log\left(\frac{1}{\varepsilon}\right) + \frac{q}{2}\log\left(\frac{2q}{e}\right)\right)$.

We define $r_{\text{near}} = 1/\sqrt{2}$ and $\Delta = C_1\sqrt{\log(s_{\max})} + C_2$ for some C_1 and C_2.

1. *Global Bounds.* From what preceeds, we have

$$\left\|K^{(10)}\right\| \leqslant \frac{1}{\sqrt{e}}, \quad \left\|K^{(02)}\right\| \leqslant \frac{2}{e} + 1, \quad \left\|K^{(12)}\right\| \leqslant \frac{3}{\sqrt{e}} + \left(\frac{3}{e}\right)^{\frac{3}{2}}$$

   and note that $\left\|K^{(11)}\right\| = \left\|K^{(02)}\right\|$, so for all $i+j \leqslant 3$ $B_{ij} = \mathcal{O}(1)$.

2. *Near* 0 For $d_{\mathbf{H}}(x,x') \leqslant r_{\text{near}}$, we have

$$K^{(02)} \preccurlyeq -\frac{e^{-\frac{1}{4}}}{2}\mathrm{Id}$$

   and for $d_{\mathbf{H}}(x,x') \geqslant \frac{1}{2}$,

$$|K| \leqslant e^{-\frac{1}{4}} = 1 - (1 - e^{-\frac{1}{4}})$$

   and $\left\|K^{(22)}(x,x)\right\| = d+1$, so we have also $\varepsilon_i = \mathcal{O}(1)$, so $B_i = B_{0i} + B_{1i} + 1 = \mathcal{O}(1)$ and $B_{22} = d+1$.

3. *Separation.* Since $\varepsilon_i = \mathcal{O}(1)$ and $B_{ij} = \mathcal{O}(1)$, every condition $\left\|K^{(ij)}\right\| \lesssim \frac{1}{s_{\max}}$ is satisfied if $\Delta \geqslant C_1\sqrt{\log(s_{\max})} + C_2$ for some constant C_1 and C_2.

F.2.1 Fourier measurements with Gaussian frequencies

The random feature expansion for K is $\varphi_\omega(x) = e^{i\omega^\top x}$ and $\Lambda = \mathcal{N}(0, \Sigma^{-1})$. We have immediately $L_0 = 1$. For $j \geqslant 1$, we have $\mathrm{D}_j\left[\varphi_\omega\right](x)[q_1, \ldots, q_j] = \left(\prod_i \omega^\top(\Sigma^{\frac{1}{2}} q_i)\right)\varphi_\omega(x)$ and therefore

$$\|\mathrm{D}_j\left[\varphi_\omega\right]\| \leqslant \|\omega\|_\Sigma^j$$

Now, we use $\|\omega\|_\Sigma^j = (\left\|\Sigma^{\frac{1}{2}}\omega\right\|^2)^{\frac{j}{2}} = W^{\frac{j}{2}}$ where W is a χ^2 variable with d degrees of freedom. Then, we use the following Chernoff bound [3]: for $x \geqslant d$, we have

$$\mathbb{P}(W \geqslant x) \leqslant \left(\frac{ex}{d}e^{-\frac{x}{d}}\right)^{\frac{d}{2}} \leqslant \left(e\left(\sqrt{\frac{x}{d}}\right)^2 e^{-\frac{1}{2}\cdot\left(\sqrt{\frac{x}{d}}\right)^2} e^{-\frac{x}{2d}}\right)^{\frac{d}{2}} \leqslant 2^{\frac{d}{2}} e^{-\frac{x}{4}}$$

by using $x^2 e^{-\frac{x^2}{2}} \leqslant \frac{2}{e}$.

Hence we can define the F_j such that, for all $t \geqslant d^{j/2}$, $\mathbb{P}(L_j(\omega) \geqslant t) \leqslant F_j(t) = 2^{\frac{d}{2}}\exp\left(-\frac{t^{\frac{2}{j}}}{4}\right)$, and $F_j(\bar{L}_j)$ is smaller than some δ if $\bar{L}_j \propto \left(d + \log\frac{1}{\delta}\right)^{\frac{j}{2}}$. Then we must choose the L_j such that $\int_{\bar{L}_j} tF_j(t)\mathrm{d}t$ is bounded by some δ. Taking $L_j \geqslant d^{j/2}$ in any case, we have

$$\begin{aligned}\int_{\bar{L}_j} tF_j(t)\mathrm{d}t &= 2^{\frac{d}{2}}\int_{\bar{L}_j} t\exp\left(-\frac{t^{\frac{2}{j}}}{4}\right)\mathrm{d}t = 2^{\frac{d}{2}}\int_{\bar{L}_j^{\frac{2}{j}}}(j/2)t^{j-1}\exp\left(-\frac{t}{4}\right)\mathrm{d}t \\ &= 2^{\frac{d}{2}}(j/2)\int_{\bar{L}_j^{\frac{2}{j}}}\left(t^{j-1}\exp\left(-\frac{t}{8}\right)\right)\exp\left(-\frac{t}{8}\right)\mathrm{d}t \leqslant 2^{\frac{d}{2}}(j/2)\left(\frac{8(j-1)}{e}\right)^{j-1}\int_{\bar{L}_j^{\frac{2}{j}}}\exp\left(-\frac{t}{8}\right)\mathrm{d}t \\ &= 2^{\frac{d}{2}} j\left(\frac{8(j-1)}{e}\right)^{j-1} 8\exp\left(-\bar{L}_j^{\frac{2}{j}}/8\right)\end{aligned}$$

Hence this quantity is bounded by δ if $\bar{L}_j \propto \left(d + \log\left(\frac{1}{\delta}\right)\right)^{\frac{j}{2}}$. Then we have $\bar{L}_j^2 F_i(\bar{L}_i) = \bar{L}_j^2 2^{\frac{d}{2}} \exp\left(-\frac{\bar{L}_i^{\frac{2}{i}}}{4}\right)$ which is also bounded by δ if $\bar{L}_j \propto \left(d + \left(\log\frac{d}{\delta}\right)^2\right)^{\frac{j}{2}}$. At the end of the day, our assumptions are satisfied for

$$\bar{L}_j \propto \left(d + \left(\log\frac{dm}{\rho}\right)^2\right)^{\frac{j}{2}}$$

F.2.2 Gaussian mixture model learning

We apply the mixture model framework with the base distribution:

$$P_\theta = \mathcal{N}(\theta, \Sigma)$$

The random features on the data space are $\varphi'_\omega(x) = Ce^{i\omega^\top x}$ with Gaussian distribution $\omega \sim \Lambda = \mathcal{N}(0, A)$ for some constant C and matrix A. Then, the features on the parameter space are $\varphi_\omega(\theta) = \mathbb{E}_{x\sim P_\theta}\varphi'_\omega(x) = Ce^{i\omega^\top\theta}e^{-\frac{1}{2}\|\omega\|_\Sigma^2}$ (that is, the characteristic function of Gaussians). Then, it is possible to show [5] that the kernel is

$$K(\theta, \theta') = C^2\frac{\left|A^{-1}\right|^{\frac{1}{2}}}{\left|2\Sigma + A^{-1}\right|^{\frac{1}{2}}}e^{-\frac{1}{2}\left\|\theta-\theta'\right\|^2_{(2\Sigma+A^{-1})^{-1}}}$$

Hence we choose $A = c\Sigma^{-1}$, $C = (1+2c)^{\frac{d}{4}}$, and we come back to the previous case $K(\theta, \theta') = e^{-\frac{1}{2}\left\|\theta-\theta'\right\|^2_{\tilde{\Sigma}^{-1}}}$ with covariance $\tilde{\Sigma} = (2+1/c)\Sigma$. Hence $\varepsilon_i = \mathcal{O}(1)$, $B_{ij} = \mathcal{O}(1)$, $d_{\mathbf{H}}(\theta, \theta') = \|\theta - \theta'\|_{\tilde{\Sigma}^{-1}} = \frac{1}{\sqrt{2+1/c}}\|\theta - \theta'\|_{\Sigma^{-1}}$.

Admissible features. Unlike the previous case, the features are directly bounded and Lipschitz. We have

$$|\varphi_\omega(\theta)| \leqslant C \stackrel{\text{def.}}{=} L_0,$$

$$\|\mathrm{D}_j\left[\varphi_\omega(\theta)\right]\| = C\left\|\tilde{\Sigma}^{\frac{1}{2}}\omega\right\|^j e^{-\frac{\|\omega\|_{\tilde{\Sigma}}^2}{2}} = C\left(2+1/c\right)^{\frac{j}{2}}\left\|\Sigma^{\frac{1}{2}}\omega\right\|^j e^{-\frac{\|\omega\|_{\tilde{\Sigma}}^2}{2}} \leqslant C\left(2+1/c\right)^{\frac{j}{2}}\left(\frac{j}{e}\right)^{\frac{j}{2}} \stackrel{\text{def.}}{=} L_j$$

Hence all constants L_j are in $\mathcal{O}\left(C(2+1/c)^{\frac{j}{2}}\right)$ by choosing $c = \frac{1}{d}$ they are in $\mathcal{O}\left(d^{\frac{j}{2}}\right)$.

F.3 The Laplace transform kernel

Let $\alpha \in \mathbb{R}_+^d$ and let $\mathcal{X} \subset \mathbb{R}_+^d$ be a compact domain. Define for $x \in \mathcal{X}$ and $\omega \in \mathbb{R}_+^d$,

$$\varphi_\omega(x) \stackrel{\text{def.}}{=} \exp(-\langle x,\, \omega\rangle)\prod_{i=1}^d\sqrt{\frac{(x_i+\alpha_i)}{\alpha_i}} \quad \text{and} \quad \Lambda(\omega) \stackrel{\text{def.}}{=} \exp(-\langle 2\alpha,\, \omega\rangle)\prod_{i=1}^d(2\alpha_i),$$

The associated kernel is $K(x, x') = \prod_{i=1}^d \kappa(x_i + \alpha_i, x'_i + \alpha_i)$ where κ is the 1D Laplace kernel

$$\kappa(u, v) \stackrel{\text{def.}}{=} 2\frac{\sqrt{uv}}{(u+v)}.$$

A direct computation shows that $\mathbf{H}_x \in \mathbb{R}^{d\times d}$ is the diagonal matrix with $(h_{x_i+\alpha_i})_{i=1}^d$ where $h_x \stackrel{\text{def.}}{=} \partial_x\partial_{x'}\kappa(x, x) = (2x)^{-2}$. Note that

$$d_\kappa(s, t) = \int_{\min\{s,t\}}^{\max\{s,t\}}(2x + 2\alpha)^{-1}\mathrm{d}x = \left|\log\left(\frac{t+\alpha}{s+\alpha}\right)\right| \tag{F.5}$$

and so, $d_{\mathbf{H}}(x, x') = \sqrt{\sum_{i=1}^d\left|\log\left(\frac{x_i+\alpha_i}{x'_i+\alpha_i}\right)\right|^2}$.

We have the following results concerning the boundedness of $\|\mathrm{D}_j\left[\varphi_\omega\right]\|$ and the admissiblity of K:

Theorem F.2 (Stochastic gradient bounds)**.** *Assume that the α_i's are all distinct. Then,* $\bar{L}_0(\omega) \leqslant \bar{L}_0 \stackrel{\text{def.}}{=} \left(1 + \frac{R_{\mathcal{X}}}{\min_i \alpha_i}\right)^d$ *and for* $j = 1, 2, 3$*,*

$$\mathbb{P}(L_j(\omega) \geqslant t) \leqslant F_j(t) \stackrel{\text{def.}}{=} \sum_{i=1}^{d} \beta_i \exp\left(-\alpha_i \left(\frac{1}{2(R_{\mathcal{X}} + \|\alpha\|_\infty)} \left(\frac{t}{\bar{L}_0}\right)^{1/j} - \sqrt{d}\right)\right)$$

and we have that $\sum_i F_j(\bar{L}_j) \leqslant \delta$ *and* $\bar{L}_j^2 \sum_i F_i(\bar{L}_i) + 2\int_{\bar{L}_j}^{\infty} tF_j(t)\mathrm{d}t \leqslant \delta$ *provided that*

$$\bar{L}_j \propto \bar{L}_0(R_{\mathcal{X}} + \|\alpha\|_\infty)^j \left(\sqrt{d} + \max_i \frac{1}{\alpha_i} \log\left(\frac{d\beta_i \bar{L}_0(R_{\mathcal{X}} + \|\alpha\|_\infty)}{\delta\alpha_i}\right)\right)^j.$$

where $\beta_i = \prod_{j \neq i} \frac{\alpha_j}{\alpha_j - \alpha_i}$*. Note that* $\alpha_i \sim d$ *implies that* $\bar{L}_0 \sim (1 + R_{\mathcal{X}}/d)^d \sim e^{R_{\mathcal{X}}}$*.*

Theorem F.3 (Admissiblity of K)**.** *The Laplace transform kernel K is admissible with* $r_{\text{near}} = 0.2$*,* $C_{\mathbf{H}} = 1.25$*,* $\varepsilon_0 = 0.005$*,* $\varepsilon_2 = 1.52$*. For all* $i + j \leqslant 3$*,* $B_{ij} = \mathcal{O}(1)$*,* $B_{22} = \mathcal{O}(d)$*,* $\Delta = \mathcal{O}(d + \log(d^{3/2} s_{\max}))$ *and* $h = \mathcal{O}(1)$*.*

The first result Theorem F.2 is proved in Section F.3.1 and the second result, Theorem F.4 is a direct consequence of Theorem F.4 and Lemma F.5 in Section F.3.2.

F.3.1 Stochastic gradient bounds

Proof of Theorem F.2. Let $V \stackrel{\text{def.}}{=} (1 - 2(x_i + \alpha_i)\omega_i)_{i=1}^d \in \mathbb{R}^d$. Then,

$$\begin{aligned}
\|V\| &= \sqrt{\sum_i (1 - 2(x_i + \alpha_i)\omega_i)^2} \\
&\leqslant \sqrt{\sum_i 1 + 4(x_i + \alpha_i)^2 \omega_i^2} \leqslant \sqrt{d + 4(R_{\mathcal{X}} + \|\alpha\|_\infty)^2 \|w\|^2} \\
&\leqslant \sqrt{d} + 2(R_{\mathcal{X}} + \|\alpha\|_\infty) \|w\|
\end{aligned}$$

We have the following bounds:

$$\begin{aligned}
|\varphi_\omega(x)| &\leqslant \prod_{i=1}^{d} \sqrt{1 + \frac{x_i}{\alpha_i}} \leqslant \left(1 + \frac{R_{\mathcal{X}}}{\min_i \alpha_i}\right)^d \stackrel{\text{def.}}{=} \bar{L}_0, \\
\mathrm{D}_1[\varphi_\omega](x) = \varphi_\omega(x)V &\implies \|\mathrm{D}_1[\varphi_\omega](x)\| \leqslant \bar{L}_0 \|V\| \\
\mathrm{D}_2[\varphi_\omega](x) = \varphi_\omega(x)(VV^\top - 2\mathrm{Id}) &\implies \|\mathrm{D}_2[\varphi_\omega](x)\| \leqslant \bar{L}_0 \min\{\|V\|^2, 2\}.
\end{aligned}$$

and given $u, q \in \mathbb{R}^d$,

$$\mathrm{D}_3[\varphi_\omega](x)[q, q, u] = \varphi_\omega(x)\left(\langle u, V\rangle\langle q, V\rangle^2 - 2\|q\|^2 - 4\langle u, q\rangle\langle q, V\rangle + 8\sum_i q_i^2 u_i\right),$$

so

$$\|\mathrm{D}_3[\varphi_\omega](x)\| \leqslant |\varphi_\omega(x)| \left(\|V\|^3 + 10 + 4\|V\|\right) \leqslant \bar{L}_0 5(\|V\|^3 + 3),$$

And therefore, in general,

$$\|\mathrm{D}_j[\varphi_\omega](x)\| \leqslant L_j(\omega) \stackrel{\text{def.}}{=} \bar{R}_{\mathcal{X}}^{j+1}\left(\sqrt{d} + \|\omega\|\right)^j$$

$$\|\mathrm{D}_j\left[\varphi_\omega\right](x)\| \lesssim L_j(\omega) \overset{\text{def.}}{=} \bar{L}_0\left(\sqrt{d}+2(R_{\mathcal{X}}+\|\alpha\|_\infty)\|w\|\right)^j$$

Assuming for simplicity that all α_j are distinct, we have [1]:

$$\mathbb{P}(\|w\| \geqslant t) \leqslant \mathbb{P}(\|\omega\|_1 \geqslant t) = \sum_{i=1}^d \beta_i e^{-\alpha_i t}$$

where $\beta_i = \prod_{j\neq i} \frac{\alpha_j}{\alpha_j-\alpha_i}$, using the fact that $\|\omega\|_1$ is a sum of independent exponential random variable.

Hence, for all $1 \leqslant j \leqslant 3$ and $t \geqslant d^{\frac{j}{2}}$ we have

$$\begin{aligned}
\mathbb{P}(L_j(\omega) \geqslant t) &\leqslant \mathbb{P}\left(\|w\| \geqslant \frac{1}{2(R_{\mathcal{X}}+\|\alpha\|_\infty)}\left(\frac{t}{\bar{L}_0}\right)^{1/j} - \sqrt{d}\right) \\
&\leqslant F_j(t) \overset{\text{def.}}{=} \sum_{i=1}^d \beta_i \exp\left(-\alpha_i\left(\frac{1}{2(R_{\mathcal{X}}+\|\alpha\|_\infty)}\left(\frac{t}{\bar{L}_0}\right)^{1/j} - \sqrt{d}\right)\right) \leqslant \delta
\end{aligned}$$

and $F_j(\bar{L}_j) \leqslant \delta$ if

$$\bar{L}_j \geqslant \bar{L}_0\left(2^j(R_{\mathcal{X}}+\|\alpha\|_\infty)^j\left(\sqrt{d}+\max_i \frac{1}{\alpha_i}\log\left(\frac{d\beta_i}{\delta}\right)\right)^j\right)$$

Next, in a similar manner to the Gaussian case, we compute

$$\begin{aligned}
\int_{\bar{L}_j} tF_j(t)\mathrm{d}t &= \sum_{i=1}^d \beta_i \int_{\bar{L}_j} t\exp\left(-\alpha_i\left(\frac{1}{2(R_{\mathcal{X}}+\|\alpha\|_\infty)}\left(\frac{t}{\bar{L}_0}\right)^{1/j} - \sqrt{d}\right)\right)\mathrm{d}t \\
&= \bar{L}_0^2 j\sum_{i=1}^d e^{\alpha_i\sqrt{d}}\beta_i \int_{(\bar{L}_j/\bar{L}_0)^{1/j}} \exp\left(\frac{-\alpha_i u}{2(R_{\mathcal{X}}+\|\alpha\|_\infty)}\right)u^{2j-1}\mathrm{d}u \\
&\leqslant \left(\frac{(2j-1)4(R_{\mathcal{X}}+\|\alpha\|_\infty)}{e\alpha_i}\right)^{2j-1}\bar{L}_0^2 j\sum_{i=1}^d e^{\alpha_i\sqrt{d}}\beta_i \int_{(\bar{L}_j/\bar{L}_0)^{1/j}} \exp\left(\frac{-\alpha_i u}{4(R_{\mathcal{X}}+\|\alpha\|_\infty)}\right)\mathrm{d}u \\
&\leqslant \left(\frac{4(R_{\mathcal{X}}+\|\alpha\|_\infty)}{\alpha_i}\right)^{2j}\left(\frac{2j-1}{e}\right)^{2j-1}\bar{L}_0^2 j\sum_{i=1}^d e^{\alpha_i\sqrt{d}}\beta_i \exp\left(\frac{-\alpha_i(\bar{L}_j/\bar{L}_0)^{1/j}}{4(R_{\mathcal{X}}+\|\alpha\|_\infty)}\right) \leqslant \delta
\end{aligned}$$

if for all $i = 1, \ldots, d$,

$$\frac{4(R_{\mathcal{X}}+\|\alpha\|_\infty)}{\alpha_i}\left(2j\log\left(\frac{4(2j-1)(R_{\mathcal{X}}+\|\alpha\|_\infty)}{e\alpha_i}\right)+\log(\bar{L}_0^2 j)+\alpha_i\sqrt{d}+\log\left(\frac{d\beta_i}{\delta}\right)\right) \leqslant \left(\frac{\bar{L}_j}{\bar{L}_0}\right)^{1/j}$$

that is,

$$\bar{L}_j \gtrsim \bar{L}_0\left(2^j(R_{\mathcal{X}}+\|\alpha\|_\infty)^j\left(\sqrt{d}+\max_i \frac{1}{\alpha_i}\log\left(\frac{d\beta_i}{\delta}\right)\right)^j\right).$$

It remains to bound $\bar{L}_j F_\ell(\bar{L}_\ell)$ with $\ell, j \in \{0,1,2,3\}$: Let $\bar{L}_\ell \geqslant \bar{L}_0 M^\ell$ for some M to be determined. Then,

$$\begin{aligned}
\bar{L}_j F_\ell(\bar{L}_\ell) &\leqslant \bar{L}_0 M^j \sum_{i=1}^d \beta_i \exp\left(\frac{-\alpha_i}{2(R_{\mathcal{X}} + \|\alpha\|_\infty)} M + \alpha_i \sqrt{d}\right) \\
&= \bar{L}_0 \sum_{i=1}^d \beta_i M^j \exp\left(\frac{-\alpha_i}{4(R_{\mathcal{X}} + \|\alpha\|_\infty)} M\right) \exp\left(\frac{-\alpha_i}{4(R_{\mathcal{X}} + \|\alpha\|_\infty)} M\right) e^{\alpha_i \sqrt{d}} \\
&\leqslant \bar{L}_0 e^{-j} \sum_{i=1}^d \left(\frac{4j(R_{\mathcal{X}} + \|\alpha\|_\infty)}{\alpha_i}\right)^j \beta_i \exp\left(\frac{-\alpha_i}{4(R_{\mathcal{X}} + \|\alpha\|_\infty)} M\right) e^{\alpha_i \sqrt{d}} \\
&\leqslant \bar{L}_0 e^{-3} \sum_{i=1}^d \left(\frac{12(R_{\mathcal{X}} + \|\alpha\|_\infty)}{\alpha_i}\right)^3 \beta_i \exp\left(\frac{-\alpha_i}{4(R_{\mathcal{X}} + \|\alpha\|_\infty)} M\right) e^{\alpha_i \sqrt{d}} \leqslant \delta
\end{aligned}$$

if for each $i = 1, \ldots, d$

$$M \geqslant 4(R_{\mathcal{X}} + \|\alpha\|_\infty) \left(\sqrt{d} + \max_i \frac{1}{\alpha_i} \log\left(\frac{\bar{L}_0 d\beta_i}{\delta e^3} \left(\frac{12(R_{\mathcal{X}} + \|\alpha\|_\infty)}{\alpha_i}\right)^3\right)\right).$$

Therefore, similar to the Gaussian case, the conclusion follows for $\bar{L}_0 = \left(1 + \frac{R_{\mathcal{X}}}{\min_i \alpha_i}\right)^d$, and for $j = 1, 2, 3$,

$$\bar{L}_j \propto \bar{L}_0 (R_{\mathcal{X}} + \|\alpha\|_\infty)^j \left(\sqrt{d} + \max_i \frac{1}{\alpha_i} \log\left(\frac{d\beta_i \bar{L}_0 (R_{\mathcal{X}} + \|\alpha\|_\infty)}{\delta \alpha_i}\right)\right)^j.$$

□

F.3.2 Admissiblity of the kernel

Metric variation We have the following lemma on the variation of the Fisher metric:

Lemma F.5. *Suppose that $d_{\mathbf{H}}(x, x') \leqslant c$, then* $\left\|\mathrm{Id} - \mathbf{H}_x^{1/2} \mathbf{H}_{x'}\right\| \leqslant (1 + ce^c) d_{\mathbf{H}}(x, x')$.

Proof. Note that $|1 - |(x_i + \alpha_i)/(x_i' + \alpha_i)|| \leqslant \max\{e^{d_\kappa(x_i, x_i')} - 1, 1 - e^{-d_\kappa(x_i, x_i')}\} \leqslant d_\kappa(x_i, x_i')(1 + ce^c)$ for all $d_\kappa(x_i, x_i') \leqslant c$. Therefore,

$$\|\mathrm{Id} - \mathbf{H}_x \mathbf{H}_{x'}\|^2 = \sum_i |1 - |(x_i + \alpha_i)/(x_i' + \alpha_i)||^2 \leqslant (1 + ce^c) d_{\mathbf{H}}(x, x')$$

provided that $d_{\mathbf{H}}(x, x') \leqslant c$.

□

Admissiblity of the kernel The following theorem provides bounds for K and its normalised derivatives.

Theorem F.4. *1.* $|K(x, x')| \leqslant \min\{2^d e^{-\frac{1}{2} d_{\mathbf{H}}(x, x')}, \frac{8}{8 + d_{\mathbf{H}}(x, x')^2}\}$.

2. $\left\|K^{(10)}(x, x')\right\| \leqslant \min\{2\sqrt{d} \, |K|, \sqrt{2}\}$.

3. $\left\|K^{(11)}\right\| \leqslant \min\{9d \, |K|, 8\}$

4. $\left\|K^{(20)}\right\| \leqslant \min\{10d \, |K|, 8\}$ *and* $\lambda_{\min}(-K^{(20)}) \geqslant \left(2 - 12 d_{\mathbf{H}}(x, x')^2\right) K$.

5. $\left\|K^{(12)}\right\| \leqslant \min\{66 \, |K| \, d^{3/2}, 16\sqrt{d} + 49\}$ *and* $\left\|K^{(12)}(x, x')\right\| \leqslant 34$ *if* $d_{\mathbf{H}}(x, x') \leqslant 1$.

6. $\left\|K^{(22)}\right\| \leqslant 16d + 9$.

In particular, for $d_{\mathbf{H}}(x, x') \geqslant 2d \log(2) + 2 \log\left(\frac{52 d^{3/2} s_{\max}}{h}\right)$, *we have* $\left\|K^{(ij)}(x, x')\right\| \leqslant \frac{h}{s_{\max}}$.

To prove this result, we first present some bounds for the univariate Laplace kernel in Section F.3.3 before applying these bounds in Section F.3.4.

F.3.3 1D Laplace kernel

In the following $\kappa^{(ij)}(x,x') \overset{\text{def.}}{=} h_x^{-i/2} h_{x'}^{-j/2} \partial_x^i \partial_{x'}^j \kappa(x,x')$.

Lemma F.6. *We have*

(i) $\kappa(x,x') = \operatorname{sech}\left(\frac{d_\kappa(x,x')}{2}\right) \leqslant 2e^{-\frac{1}{2}d_\kappa(x,x')}$,

(ii) $\left|\kappa^{(10)}(x,x')\right| = 2\left|\tanh\left(\frac{d_\kappa(x,x')}{2}\right)\kappa(x,x')\right|$, *and* $\left|\kappa^{(10)}\right| \leqslant 2\left|\kappa\right|$.

(iii) $\left|\kappa^{(11)}\right| \leqslant 4\left|\kappa\right|^3 + 4\left|\kappa\right|$

(iv) $\left|\kappa^{(20)}\right| \leqslant 6\left|\kappa\right|$ *and* $-\kappa^{(20)} \geqslant 2\kappa(x,x')\left(1 - 2\tanh\left(\frac{d_\kappa(x,x')}{2}\right)\right)$.

(v) $\left|\kappa^{(12)}\right| \leqslant 49\left|\kappa\right|$.

(vi) $\kappa^{(22)}(x,x) = 9$ *for all* x.

Proof. We first state the partial derivatives of κ:

$$
\begin{aligned}
\kappa(x,x') &= \frac{2\sqrt{xx'}}{x+x'}, \\
\partial_x\kappa(x,x') &= \frac{x'(x'-x)}{\sqrt{xx'}(x+x')^2} \\
\partial_x\partial_{x'}\kappa(x,x') &= \frac{-x^2+6xx'-(x')^2}{2\sqrt{xx'}(x+x')^3} \\
\partial_x^2\kappa(x,x') &= -\frac{(x')^2\left((x+x')^2+4x(x'-x)\right)}{2\left(xx'\right)^{3/2}(x+x')^3} \\
&= -\frac{(x')^2}{2\left(xx'\right)^{3/2}(x+x')} - \frac{2x'(x'-x)}{\left(xx'\right)^{1/2}(x+x')^3} \\
\partial_x\partial_{x'}^2\kappa(x,x') &= \frac{x^3+13x^2x'-33x(x')^2+3(x')^3)}{4x'(xx')^{1/2}(x+x')^4} \\
\partial_x^2\partial_{x'}^2\kappa(x,x') &= -\frac{3x^4+60x^3x'-270x^2(x')^2+60x(x')^3+3(x')^4}{8xx'(xx')^{1/2}(x+x')^5}
\end{aligned}
$$

(i)

$$
\kappa(x,x') = 2\left(\sqrt{\frac{x}{x'}} + \sqrt{\frac{x'}{x}}\right)^{-1} = \frac{2}{e^{-\frac{d_\kappa(x,x')}{2}} + e^{\frac{d_\kappa(x,x')}{2}}} = \frac{1}{\cosh(\frac{d_\kappa(x,x')}{2})} \leqslant 2e^{-\frac{1}{2}d_\kappa(x,x')},
$$

(ii) We have, assuming that $x > x'$,

$$\begin{aligned}
\kappa^{(10)}(x,x') &= 2x\partial_x \kappa(x,x') = 2\frac{x'-x}{x+x'}\kappa(x,x') \\
&= 2\left(\frac{1}{\frac{x}{x'}+1} - \frac{1}{1+\frac{x'}{x}}\right)\kappa(x,x') \\
&= 2\left(\frac{1}{1+\exp(d_\kappa(x,x'))} - \frac{1}{1+\exp(-d_\kappa(x,x'))}\right) \\
&= 2\left(\frac{\exp(-d_\kappa(x,x')) - \exp(d_\kappa(x,x'))}{2+\exp(d_\kappa(x,x'))+\exp(d_\kappa(x,x'))}\right) \\
&= \frac{-2\sinh(d_\kappa(x,x'))}{1+\cosh(d_\kappa(x,x'))}\kappa(x,x') \\
&= -2\tanh(d_\kappa(x,x')/2)\kappa(x,x'),
\end{aligned}$$

(iii)

$$\begin{aligned}
\kappa^{(11)} &= 4xx'\partial_{x'}\partial_x \kappa(x,x') = 4xx'\frac{4xx'-(x-x')^2}{2\sqrt{xx'}(x+x')^3} \\
&= 4\kappa(x,x')^3 - \frac{4(x-x')^2}{(x+x')^2}\kappa(x,x') \\
&= \kappa(x,x')\left(4\kappa(x,x')^2 - 4\tanh^2(d_\kappa(x,x')/2)\right)
\end{aligned}$$

so $\left|\kappa^{(11)}\right| \leqslant 4\left|\kappa\right|^3 + 4\left|\kappa\right|$.

(iv)

$$\begin{aligned}
\kappa^{(20)} &= 4x^2\partial_x^2\kappa(x,x') = -\frac{4\left(xx'\right)^{1/2}\left((x+x')^2+4x(x'-x)\right)}{2(x+x')^3} \\
&= -2\kappa(x,x')\left(1+\frac{2x(x'-x)}{(x+x')^2}\right)
\end{aligned}$$

so $\left|\kappa^{20}\right| \leqslant 6\left|\kappa\right|$. Also,

$$-\kappa^{(20)} \geqslant 2\kappa(x,x')\left(1-2\tanh(d_\kappa(x,x')/2)\right)$$

(v)

$$\begin{aligned}
\kappa^{(12)} &= 2x(2x')^2\partial_x\partial_{x'}^2\kappa(x,x') \\
&= \kappa(x,x')\left(1+\frac{2v(5u^2-18uv+v^2)}{(u+v)^3}\right)
\end{aligned}$$

so $\left|\kappa^{(12)}\right| \leqslant 49\left|\kappa\right|$.

(vi)

$$\begin{aligned}
\kappa^{(22)} &= 16(xx')^2\partial_x^2\partial_{x'}^2\kappa(x,x') \\
&= -3 - \frac{48xx'(x^2-6xx'+(x')^2)}{(x+x')^4}
\end{aligned}$$

and $\kappa^{(22)}(x,x) = 9$. $\square$

F.3.4 Proof of Theorem F.4

Let $\mathrm{d}_\ell \overset{\text{def.}}{=} d_\kappa(x_\ell + \alpha_\ell, x'_\ell + \alpha_\ell)$ and note that $d_{\mathbf{H}}(x, x') = \sqrt{\sum_\ell \mathrm{d}_\ell^2}$. Define $g = \left(2\tanh(\frac{\mathrm{d}_\ell}{2})\right)_{\ell=1}^d$. We first prove that

(i) $|K(x,x')| \leqslant \prod_{\ell=1}^d \operatorname{sech}(\mathrm{d}_\ell/2) \leqslant \prod_{\ell=1}^d \frac{1}{1+\mathrm{d}_\ell^2/8} \leqslant \frac{1}{1+\frac{1}{8}d_{\mathbf{H}}(x,x')^2}$.

(ii) $\left\|K^{(10)}(x,x')\right\| \leqslant \|g\|_2 |K|$.

(iii) $\left\|K^{(11)}\right\| \leqslant |K| \left(\|g\|_2^2 + 5\right)$

(iv) $\left\|K^{(20)}\right\| \leqslant |K| \left(\|g\|_2^2 + 6\right)$ and $\lambda_{\min}\left(K^{(20)}\right) \geqslant K \left(2 - 3\|g\|_2^2\right)$.

(v) $\left\|K^{(12)}\right\| \leqslant |K| \left(\|g\|_2^3 + 16\|g\|_2 + 49\right)$

(vi) $\left\|K^{(22)}\right\| \leqslant 16d + 9$.

The result would then follow because

- $\operatorname{sech}(x) \leqslant 2e^{-x}$ and $\operatorname{sech}(x) \leqslant (1 + x^2/2)^{-1}$.
- $|\tanh(x)| \leqslant \min\{x, 1\}$, so $\|g\| \leqslant \min\{d_{\mathbf{H}}(x,x'), 2\sqrt{d}\}$,

For example, $\left\|K^{(12)}\right\| \leqslant \frac{1}{1+\frac{1}{8}d_{\mathbf{H}}(x,x')^2}\left(d_{\mathbf{H}}(x,x')^3 + 16 d_{\mathbf{H}}(x,x') + 24\right) \leqslant 8 d_{\mathbf{H}}(x,x') + \frac{\sqrt{8}}{2} + 24 \leqslant 34$ when $d_{\mathbf{H}}(x,x') \leqslant 1$.

In the following, we write $\kappa_\ell^{(ij)} \overset{\text{def.}}{=} \kappa^{(ij)}(x_\ell + \alpha_\ell, x'_\ell + \alpha_\ell)$ and $\kappa_\ell \overset{\text{def.}}{=} \kappa_\ell^{(00)}$ and $K_i \overset{\text{def.}}{=} \prod_{j \neq i} \kappa_j$. Moreover, we will make use of the inequalities for $\kappa^{(ij)}$ derived in Lemma F.6.

(i) We have

$$|K(x,x')| \leqslant \prod_{\ell=1}^d \operatorname{sech}(\mathrm{d}_\ell) \leqslant \prod_{\ell=1}^d \left(1 + \frac{\mathrm{d}_\ell^2}{2}\right)^{-1} \leqslant \frac{1}{1 + d_{\mathbf{H}}(x,x')^2}.$$

(ii)

$$K^{(10)}(x,x') = \left(\kappa_\ell^{(10)} K_\ell\right)_{\ell=1}^d \implies \left\|K^{(10)}(x,x')\right\| \leqslant \|g\|_2 |K|.$$

(iii) For $i \neq j$

$$\left|K_{ij}^{(11)}\right| = \left|\kappa_i^{(10)} \kappa_j^{(01)} K_{ij}\right| \leqslant 4 \tanh\left(\frac{\mathrm{d}_i}{2}\right) \tanh\left(\frac{\mathrm{d}_j}{2}\right) |K|,$$

and $\left|K_{ii}^{(11)}\right| = \left|\kappa_i^{(11)} K_i\right| \leqslant 5|K|$. So, given $p \in \mathbb{R}^d$ of unit norm,

$$\begin{aligned}
\langle K^{(11)} p,\, p\rangle &= \sum_{i=1}^d \sum_{j \neq i} \kappa_i^{(10)} \kappa_j^{(01)} K_{ij} p_i p_j + \sum_{i=1}^d p_i^2 \kappa_i^{(11)} K_i \\
&\leqslant |K| \left(\sum_{i=1}^d \sum_{j \neq i} 4 \tanh(\mathrm{d}_i/2) \tanh(\mathrm{d}_j/2) p_i p_j + 5 \sum_{i=1}^d p_i^2\right) \\
&\leqslant |K| \left(\|g\|_2^2 + 5\right)
\end{aligned}$$

(iv) For $i \neq j$, $K^{(20)}_{ij} = \kappa^{(10)}_i \kappa^{(10)}_j K_{ij}$, and $\left|K^{(20)}_{ii}\right| = \left|\kappa^{(20)}_i K_i\right| \leqslant 6\,|K|$ and $-K^{(20)}_{ii} \geqslant 2K\left(1 - 2\tanh\left(\frac{\mathrm{d}_i}{2}\right)\right)$.

$$
\begin{aligned}
\langle K^{(20)}p,\, p\rangle &= \sum_{i=1}^{d}\sum_{j\neq i}\kappa^{(10)}_i\kappa^{(10)}_j K_{ij}p_ip_j + \sum_{i=1}^{d} p_i^2\kappa^{(20)}_i K_i \\
&\leqslant |K|\left(\sum_{i=1}^{d}\sum_{j\neq i} 4\tanh(\mathrm{d}_i/2)\tanh(\mathrm{d}_j/2)p_ip_j + 6\sum_{i=1}^{d}p_i^2\right) \\
&\leqslant |K|\left(\|g\|_2^2 + 6\right),
\end{aligned}
$$

and

$$
\langle -K^{(20)}p,\, p\rangle \geqslant K\left(2 - 2\,\|g\|_\infty - \|g\|_2^2\right)
$$

(v) For i, j, ℓ all distinct,

$$
K^{(12)}_{ij\ell} = \kappa^{(10)}_i\kappa^{(01)}_j\kappa^{(01)}_\ell K_{ij\ell} \leqslant 8\tanh\left(\frac{\mathrm{d}_i}{2}\right)\tanh\left(\frac{\mathrm{d}_j}{2}\right)\tanh\left(\frac{\mathrm{d}_\ell}{2}\right)K,
$$

for all i, ℓ,

$$
\begin{aligned}
K^{(12)}_{ii\ell} &= 8\kappa^{(11)}_i\kappa^{(01)}_\ell K_{i\ell} \leqslant 10\tanh\left(\frac{\mathrm{d}_\ell}{2}\right)K \\
K^{(12)}_{iji} &= \kappa^{(11)}_i\kappa^{(01)}_j K_{ij} \leqslant 10\tanh\left(\frac{\mathrm{d}_j}{2}\right)K,
\end{aligned}
$$

$K^{(12)}_{ijj} = \kappa^{(10)}_i\kappa^{(02)}_\ell K_{ij} \leqslant 12\tanh\left(\frac{\mathrm{d}_i}{2}\right)K$, and $K^{(12)}_{iii} = \kappa^{(12)}_i K_i \leqslant 26K$. So, for $p, q \in \mathbb{R}^d$ of unit norm,

$$
\begin{aligned}
\sum_i\sum_j\sum_\ell K^{(12)}_{ij\ell}p_jp_\ell q_i &= \sum_i\left(\sum_{j\neq i}\sum_\ell K^{(12)}_{ij\ell}p_jp_\ell q_i + \sum_\ell K^{(12)}_{ii\ell}p_ip_\ell q_i\right) \\
&= \sum_i\sum_{j\neq i}\left(\sum_{\ell\notin\{i,j\}}K^{(12)}_{ij\ell}p_jp_\ell q_i + K^{(12)}_{iji}p_jp_iq_i + K^{(12)}_{ijj}p_j^2q_i\right) \\
&+ \sum_i\sum_{\ell\neq i}K^{(12)}_{ii\ell}p_ip_\ell q_i + \sum_i K^{(12)}_{iii}p_i^2q_i \\
&\leqslant |K|\left(\|g\|_2^3 + 16\,\|g\|_2 + 49\right).
\end{aligned}
$$

(vi)

$$
\begin{aligned}
\left\|K^{(22)}(x,x)\right\| &= \sup_{\|p\|=1}\mathbb{E}[\langle \mathbf{H}_x^{-1/2}\nabla^2\varphi_\omega(x)\mathbf{H}_x^{-1/2}p,\, \mathbf{H}_x^{-1/2}\nabla^2\varphi_\omega(x)\mathbf{H}_x^{-1/2}p\rangle] \\
&\leqslant \sup_{\|p\|=1}\sum_i\sum_{k\neq i}\kappa^{(11)}_i\kappa^{(11)}_k p_i^2 + \sum_i\sum_{k\neq i}\kappa^{(12)}_i\kappa^{(10)}_k p_ip_k + \sum_i\sum_{k\neq i}\sum_{j\notin\{i,k\}}\kappa^{(11)}_i\kappa^{(10)}_k\kappa^{(01)}_j p_kp_j \\
&+ \sum_i\sum_{j\neq i}\kappa^{(21)}_i\kappa^{(01)}_j p_jp_i + \sum_i\kappa^{(22)}_i p_i^2 \\
&= \sup_{\|p\|=1}\sum_i\sum_{k\neq i}\kappa^{(11)}_i\kappa^{(11)}_k p_i^2 + \sum_i\kappa^{(22)}_i p_i^2 \\
&\leqslant d\left\|\kappa^{(11)}\right\|_\infty + \left\|\kappa^{(22)}\right\|_\infty \leqslant 16d + \left\|\kappa^{(22)}\right\|_\infty.
\end{aligned}
$$

since $\kappa^{(10)}(x,x) = \kappa^{(01)}(x,x) = 0$, and $\kappa^{(11)}(x,x) = 4$ from the proof of (iii) in Lemma F.6.

G Tools

G.1 Probability tools

Lemma G.1 (Bernstein's inequality ([8], Thm. 6)). *Let $x_1, \ldots, x_n \in \mathbb{C}$ be $i.i.d.$ bounded random variables such that $\mathbb{E}x_i = 0$, $|x_i| \leqslant M$ and $Var(x_i) \stackrel{\text{def.}}{=} \mathbb{E}[|x_i|^2] \leqslant \sigma^2$ for all i's.*

Then for all $t > 0$ we have

$$\mathcal{X}\left(\frac{1}{n}\sum_{i=1}^{n} x_i \geqslant t\right) \leqslant 4\exp\left(-\frac{nt^2/4}{\sigma^2 + Mt/(3\sqrt{2})}\right). \tag{G.1}$$

Lemma G.2 (Matrix Bernstein ([10], Theorem 6.1.1)). *Let $Y_1, ..., Y_m \in \mathbb{C}^{d_1,d_2}$ be complex random matrices with*

$$\mathbb{E}Y_j = 0, \quad \|Y_j\| \leqslant L, \quad v(Y_j) := \max(\|\mathbb{E}Y_jY_j^*\|, \|\mathbb{E}Y_j^*Y_j\|) \leqslant M$$

for each index $1 \leqslant j \leqslant m$. Introduce the random matrix

$$Z = \frac{1}{m}\sum_j Y_j.$$

Then

$$\mathbb{P}(\|Z\| \geqslant t) \leqslant 2(d_1 + d_2)e^{-\frac{mt^2/2}{M+Lt/3}} \tag{G.2}$$

Lemma G.3 (Vector Bernstein for complex vectors [7]). *Let $Y_1, \ldots, Y_M \in \mathbb{C}^d$ be a sequence of independent random vectors such that $\mathbb{E}[Y_i] = 0$, $\|Y_i\|_2 \leqslant K$ for $i = 1, \ldots, M$ and set*

$$\sigma^2 \stackrel{\text{def.}}{=} \sum_{i=1}^{M} \mathbb{E}\|Y_i\|_2^2.$$

Then, for all $t \geqslant (K + \sqrt{K^2 + 36\sigma^2})/M$,

$$\mathbb{P}\left(\left\|\frac{1}{M}\sum_{i=1}^{M} Y_i\right\|_2 \geqslant t\right) \leqslant 28\exp\left(-\frac{Mt^2/2}{\sigma^2/M + tK/3}\right)$$

Lemma G.4 (Hoeffding's inequality ([9], Lemma G.1)). *Let the components of $u \in C^k$ be drawn i.i.d. from a symmetric distribution on the complex unit circle or 0, consider a vector $w \in \mathbb{C}^k$. Then, with probability at least $1 - \rho$, we have*

$$\mathbb{P}(|\langle u, w\rangle| \geqslant t) \leqslant 4e^{-\frac{t^2}{4\|w\|^2}} \tag{G.3}$$

Lemma G.5. *[10, Theorem 4.1.1] Let the components of $u \in \mathbb{R}^k$ be a Rademacher sequence and let $Y_1, \ldots, Y_M \in \mathbb{C}^{d\times d}$ be self-adjoint matrices. Set $\sigma^2 \stackrel{\text{def.}}{=} \left\|\sum_{\ell=1}^{M} Y_\ell^2\right\|$. Then, for $t > 0$,*

$$\mathbb{P}\left(\left\|\sum_{\ell=1}^{M} u_\ell Y_\ell\right\| \geqslant t\right) \leqslant 2d\exp\left(-\frac{t^2}{2\sigma^2}\right). \tag{G.4}$$

We were only able to find a reference for this result in the case where u is a Rademacher sequence, however, by the contraction principle (see [6, Theorem 4.4]), a similar statement is true for Steinhaus sequences (we write only for the case of real symmetric matrices because this is all we require in this paper, but of course, the same argument extends to complex self-adjoint matrices):

Corollary G.1. *Let the components of $u \in \mathbb{C}^k$ i.i.d. from a symmetric distribution on the complex unit circle or 0 and let $B_1, \dots, B_M \in \mathbb{R}^{d\times d}$ be symmetric matrices. Set $\sigma^2 \stackrel{\text{def.}}{=} \left\| \sum_{\ell=1}^M B_\ell^2 \right\|$. Then, for $t > 0$,*

$$\mathbb{P}\left(\left\| \sum_{\ell=1}^M u_\ell B_\ell \right\| \geqslant t \right) \leqslant 4d \exp\left(-\frac{t^2}{4\sigma^2} \right). \tag{G.5}$$

Proof. By the union bound,

$$\mathbb{P}\left(\left\| \sum_{\ell=1}^M u_\ell B_\ell \right\| \geqslant t \right) \leqslant \mathbb{P}\left(\left\| \sum_{\ell=1}^M \mathrm{Re}\,(u_\ell) B_\ell \right\| \geqslant \frac{t}{\sqrt{2}} \right) + \mathbb{P}\left(\left\| \sum_{\ell=1}^M \mathrm{Im}\,(u_\ell) B_\ell \right\| \geqslant \frac{t}{\sqrt{2}} \right).$$

By the contraction principle [6, Theorem 4.4],

$$\mathbb{P}\left(\left\| \sum_{\ell=1}^M \mathrm{Re}\,(u_\ell) B_\ell \right\| \geqslant \frac{t}{\sqrt{2}} \right) \leqslant \mathbb{P}\left(\left\| \sum_{\ell=1}^M \xi_\ell B_\ell \right\| \geqslant \frac{t}{\sqrt{2}} \right)$$

where ξ is a Rademacher sequence, and the same argument applies to the case of $\mathrm{Im}\,(u_\ell)$. Therefore by Lemma G.5, we have $\mathbb{P}\left(\left\| \sum_{\ell=1}^M u_\ell B_\ell \right\| \geqslant t \right) \leqslant 4d \exp\left(-\frac{t^2}{4\sigma^2} \right)$. □

G.2 Linear algebra tools

The following simple lemma will be handy.

Lemma G.6. *For $1 \leqslant i, j \leqslant s$, take any scalars $a_{ij} \in \mathbb{R}$, vectors $Q_{ij}, R_{ij} \in \mathbb{R}^d$ and square matrices $A_{ij} \in \mathbb{R}^{d\times d}$.*

1. *Let $M \in \mathbb{R}^{sd\times sd}$ be a matrix formed by blocks :*

$$M = \begin{pmatrix} A_{11} & \dots & A_{1s} \\ \vdots & \ddots & \vdots \\ A_{s1} & \dots & A_{ss} \end{pmatrix}$$

Then we have

$$\|M\|_{\text{block}} = \sup_{\|x\|_{\text{block}}=1} \|Mx\|_{\text{block}} \leqslant \max_{1\leqslant i\leqslant s} \sum_{j=1}^s \|A_{ij}\| \tag{G.6}$$

Now, let $P \in \mathbb{R}^{sd\times s}$ be a rectangular matrix formed by stacking vectors $Q_{ij} \in \mathbb{R}^d$:

$$M = \begin{pmatrix} Q_{11} & \dots & Q_{1s} \\ \vdots & \ddots & \vdots \\ Q_{s1} & \dots & Q_{ss} \end{pmatrix}$$

Then,

$$\|M\|_{\infty\to\text{block}} \leqslant \max_{1\leqslant i\leqslant s} \sum_{j=1}^s \|Q_{ij}\|_2, \quad \left\|M^\top\right\|_{\text{block}\to\infty} \leqslant \max_{1\leqslant i\leqslant s} \sum_{j=1}^s \|Q_{ji}\|_2 \tag{G.7}$$

2. *Consider $A \in \mathbb{R}^{s(d+1)\times s(d+1)}$ decomposed as*

$$M = \begin{pmatrix} a_{11} & \dots & a_{1s} & Q_{11}^\top & \dots & Q_{1s}^\top \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{s1} & \dots & a_{ss} & Q_{s1}^\top & \dots & Q_{ss}^\top \\ R_{11} & \dots & R_{1s} & A_{11} & \dots & A_{1s} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ R_{s1} & \dots & R_{ss} & A_{s1} & \dots & A_{ss} \end{pmatrix}$$

Then,

$$\|M\| \leqslant \sqrt{\sum_{i,j} a_{ij}^2 + \|Q_{ij}\|^2 + \|R_{ij}\|^2 + \|A_{ij}\|^2},$$
$$\|M\|_{\text{Block}} \leqslant \max_i \{\sum_j |a_{ij}| + \|Q_{ij}\|,\ \sum_j \|R_{ij}\| + \|A_{ij}\|\}$$

Proof. The proof is simple linear algebra.

1. Let x be a vector with $\|x\|_{\text{block}} \leqslant 1$ decomposed into blocks $x = [x_1, \ldots, x_s]$ with $x_i \in \mathbb{R}^d$, we have

$$\|Mx\|_{\text{block}}^2 = \max_{1 \leqslant i \leqslant s} \left\| \sum_{j=1}^s A_{ij} x_j \right\| \leqslant \max_i \sum_j \|A_{ij}\| \|x_j\| \leqslant \max_i \sum_j \|A_{ij}\|$$

2. Similarly,

$$\left\|M^\top x\right\|_\infty = \max_{1 \leqslant i \leqslant s} \left\| \sum_{j=1}^s Q_{ji}^\top x_j \right\| \leqslant \max_i \sum_j \|Q_{ji}\| \|x_j\| \leqslant \max_i \sum_j \|Q_{ji}\|$$

Then, taking $x \in \mathbb{R}^s$ such that $\|x\|_\infty \leqslant 1$, we have

$$\|Mx\|_{\text{block}} = \max_{1 \leqslant i \leqslant s} \left\| \sum_{j=1}^s x_j Q_{ij} \right\| \leqslant \max_i \sum_j \|Q_{ij}\|$$

3. Taking $x = [x_1, \ldots, x_s, X_1, \ldots, X_s] \in \mathbb{R}^{s(d+1)}$ with $\|x\| = 1$, we have

$$\begin{aligned}
\|Mx\|^2 &= \sum_{i=1}^s \left(\sum_{j=1}^s a_{ij} x_j + Q_{ij}^\top X_j \right)^2 + \left\| \sum_{j=1}^s R_{ij} x_j + A_{ij} X_j \right\|^2 \\
&\leqslant \sum_{i=1}^s \left(\|x\| \sqrt{\sum_{j=1}^s a_{ij}^2 + \|Q_{ij}\|^2} \right)^2 + \left(\|x\| \sqrt{\sum_{j=1}^s \|R_{ij}\|^2 + \|A_{ij}\|^2} \right)^2 \\
&\leqslant \sum_{i,j} a_{ij}^2 + \|Q_{ij}\|^2 + \|R_{ij}\|^2 + \|A_{ij}\|^2
\end{aligned}$$

Now, if $\|x\|_{\text{Block}} = 1$, we have

$$\begin{aligned}
\|Mx\|_{\text{Block}} &= \max_i \left(\left| \sum_{j=1}^s a_{ij} x_j + Q_{ij}^\top X_j \right|, \left\| \sum_{j=1}^s R_{ij} x_j + A_{ij} X_j \right\| \right) \\
&\leqslant \max_i \left(\sum_{j=1}^s |a_{ij}| + \|Q_{ij}\|, \sum_{j=1}^s \|R_{ij} x_j + A_{ij} X_j\| \right)
\end{aligned}$$

□

References

[1] Mohamed Akkouchi. On the convolution of exponential distributions. *Journal of the Chungcheong Mathematical Society*, 21(4):501–510, 2008.

[2] Emmanuel J. Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[3] Sanjoy Dasgupta and Anupam Gupta. An Elementary Proof of a Theorem of Johnson and Lindenstrauss. *Random Structures and Algorithms*, 22(1):60–65, 2003.

[4] Quentin Denoyelle, Vincent Duval, and Gabriel Peyré. Support recovery for sparse super-resolution of positive measures. *Journal of Fourier Analysis and Applications*, 23(5):1153–1194, 2017.

[5] Rémi Gribonval, Gilles Blanchard, Nicolas Keriven, and Yann Traonmilin. Compressive statistical learning with random feature moments. *arXiv preprint arXiv:1706.07180*, 2017.

[6] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media, 2013.

[7] Stanislav Minsker. On some extensions of bernsteins inequality for self-adjoint operators. *Statistics & Probability Letters*, 127:111–119, 2017.

[8] Karthik Sridharan. A Gentle Introduction to Concentration Inequalities. Technical report, 2002.

[9] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed sensing off the grid. *IEEE transactions on information theory*, 59(11):7465–7490, 2013.

[10] Joel A Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1-2):1–230, 2015.