Supplementary Appendix for "The Voronoigram: Minimax Estimation of Bounded Variation Functions From Scattered Data"

Addison J. Hu   Alden Green   Ryan J. Tibshirani

A Added details and proofs for Sections 1 and 2

A.1 Discussion of sampling model for BV functions

We clarify what is meant by the sampling model in (1), since, strictly speaking, each element $f \in \mathrm{BV}(\Omega)$ is really an equivalence class of functions, defined only up to sets of Lebesgue measure zero. This issue is not simply a formality, and becomes a genuine problem for $d \geq 2$, as in this case the space $\mathrm{BV}(\Omega)$ does not compactly embed into $C^0(\Omega)$, the space of continuous functions on Ω (equipped with the L^∞ norm). A key implication of this is that the point evaluation operator is not continuous over $\mathrm{BV}(\Omega)$.

In order to make sense of the evaluation map, $x \mapsto f(x)$, we will pick a representative, denoted $f^\star \in f$, and speak of evaluations of this representative. Our approach here is the same as that taken in Green et al. (2021a,b), who study minimax estimation of Sobolev functions in the subcritical regime (and use an analogous random design model). We let $f^\star$ be the *precise representative*, defined (Evans and Gariepy, 2015) as:

$$f^\star(x) = \begin{cases} \lim\limits_{\epsilon \to 0} \dfrac{1}{\mu(B(x,\epsilon))} \displaystyle\int_{B(x,\epsilon)} f(z)\, dz & \text{if the limit exists} \\ 0 & \text{otherwise.} \end{cases}$$

Here μ denotes Lebesgue measure and $B(x,\epsilon)$ is the ball of radius ϵ centered at x.

Now we explain why the particular choice of representative is not crucial, and any choice of representative would have resulted in the same interpretation of function evaluations in (1), *almost surely, assuming that each x_i is drawn from a continuous distribution on Ω*. Recall that for a locally integrable function f on Ω, we say that a given point $x \in \Omega$ is a *Lebesgue point* of f provided that $\lim_{\epsilon \to 0}(\int_{B(x,\epsilon)} f(z)\, dz)/\mu(B(x,\epsilon))$ exists and equals $f(x)$. By the Lebesgue differentiation theorem (e.g., Theorem 1.32 of Evans and Gariepy, 2015), for any $f \in L^1(\Omega)$, almost every $x \in \Omega$ is a Lebesgue point of f. This means that each evaluation $f^\star(x_i)$ of the precise representative will equal the evaluation of any member of the equivalence class, almost surely (with respect to draws of x_i). This justifies the notation $f(x_i)$ used in the main text, for $f \in \mathrm{BV}(\Omega)$ and x_i drawn from a continuous probability distribution.

A.2 TV representation for piecewise constant functions

Here we will state and prove a more general result from which Proposition 1 will follow. First we give a more general definition of measure theoretic total variation, wherein the norm used to constrain the "test function" ϕ in the supremum is an arbitrary norm $\|\cdot\|$ on $\mathbb{R}^d$,

$$\mathrm{TV}(f; \Omega, \|\cdot\|) = \sup \left\{ \int_\Omega f(x) \,\mathrm{div}\, \phi(x)\, dx : \phi \in C_c^1(\Omega; \mathbb{R}^d),\ \|\phi(x)\| \leq 1 \text{ for all } x \in \Omega \right\}. \tag{S.1}$$

Note that our earlier definition in (2) corresponds to the special case $\mathrm{TV}(f; \Omega, \|\cdot\|_2)$, that is, corresponds to choosing $\|\cdot\| = \|\cdot\|_2$ in (S.1). In the more general TV context, this special case is often called *isotropic* TV.

Proposition S.1. *Let $V_1, \dots, V_n$ be an open partition of Ω such that each V_i is semialgebraic. Let f be of the form*

$$f = \sum_{i=1}^n \theta_i \cdot 1_{V_i},$$

for arbitrary $\theta_1, \dots, \theta_n \in \mathbb{R}$. Then, for any norm $\|\cdot\|$ and its dual norm $\|\cdot\|_*$ (induced by the Euclidean inner product), we have

$$\mathrm{TV}(f;\Omega,\|\cdot\|) = \sum_{i,j=1}^{n} \left(\int_{\partial V_i \cap \partial V_j} \|n_i(t)\|_* \, d\mathcal{H}^{d-1}(t) \right) \cdot |\theta_i - \theta_j|,$$

where $n_i(t)$ is the measure theoretic unit outer normal for V_i at a boundary point $t \in \partial V_i$. In particular, in the isotropic case $\|\cdot\| = \|\cdot\|_2$,

$$\mathrm{TV}(f;\Omega,\|\cdot\|_2) = \sum_{i,j=1}^{n} \mathcal{H}^{d-1}(\partial V_i \cap \partial V_j) \cdot |\theta_i - \theta_j|.$$

Remark 1. The condition that each V_i is semialgebraic may to weakened to what is called "polynomially bounded boundary measure." Namely, the proposition still holds if each map $r \mapsto \mathcal{H}^{d-1}(\partial V_i \cap B(0,r))$ is polynomially bounded (cf. Assumption 2.2 in Mikkelsen and Hansen, 2018). This is sufficient to guarantee a locally Lipschitz boundary (a prerequisite for the application of Gauss-Green) and to characterize the outer normals associated with the partition $V_1, \dots, V_n$.

Proof. We begin by deriving an equivalent expression of total variation of piecewise constant functions.

$$\begin{aligned}
&\mathrm{TV}(f;\Omega,\|\cdot\|) \\
&\quad = \sup\left\{ \int_\Omega f(x)\,\mathrm{div}\,\phi(x)dx : \phi \in C_c^1(\Omega;\mathbb{R}^d), \|\phi\|_* \le 1\ \forall x \right\} \\
&\quad = \sup\left\{ \sum_{i=1}^{n} \int_{V_i} \theta_i\,\mathrm{div}\,\phi(x)dx : \phi \in C_c^1(\Omega;\mathbb{R}^d), \|\phi\|_* \le 1\ \forall x \right\} \\
&\quad = \sup\left\{ \sum_{i=1}^{n} \theta_i \int_{\partial V_i} \langle \phi(t), n_i(t)\rangle d\mathcal{H}^{d-1}(t) : \phi \in C_c^1(\Omega;\mathbb{R}^d), \|\phi\|_* \le 1\ \forall x \right\} \qquad \text{(S.2)} \\
&\quad = \sup\Bigg\{ \sum_{i,j=1}^{n} \left(\theta_i \int_{\partial V_i \cap \partial V_j} \langle \phi(t), n_i(t)\rangle d\mathcal{H}^{d-1}(t) + \theta_j \int_{\partial V_i \cap \partial V_j} \langle \phi(t), n_j(t)\rangle d\mathcal{H}^{d-1}(t) \right) \\
&\qquad\qquad + \sum_{i:\bar{V}_i \cap \partial\Omega \neq \emptyset} \theta_i \underbrace{\int_{\partial V_i \cap \partial\Omega} \langle \phi(t), n_i(t)\rangle d\mathcal{H}^{d-1}(t)}_{=0;\ (\phi \text{ compactly supported})} : \phi \in C_c^1(\Omega;\mathbb{R}^d), \|\phi\|_* \le 1\ \forall t \Bigg\} \qquad \text{(S.3)} \\
&\quad = \sup\left\{ \sum_{i,j=1}^{n} \int_{\partial V_i \cap \partial V_j} (\theta_i - \theta_j)\langle \phi(t), n_i(t)\rangle d\mathcal{H}^{d-1}(t) : \phi \in C_c^1(\Omega;\mathbb{R}^d), \|\phi\|_* \le 1\ \forall t \right\} \qquad \text{(S.4)}
\end{aligned}$$

we obtain (S.2) by applying the Gauss-Green Theorem (Evans and Gariepy, 2015, Theorem 5.16); (S.3) by observing that when the boundaries of three or more $V_i \neq V_j \neq V_k \neq \cdots$ intersect, the outer normal vector is zero (Mikkelsen and Hansen, 2018, Lemma A.2(c)); and (S.4) because when the boundaries of exactly two $V_i \neq V_j$ intersect, they have opposing outer normals (Mikkelsen and Hansen, 2018, Lemma A.2(b)). Apply Hölder's inequality to obtain an upper bound,

$$\begin{aligned}
&\mathrm{TV}(f;\Omega,\|\cdot\|) \\
&\quad \le \sup\left\{ \sum_{i,j=1}^{n} |\theta_i - \theta_j| \int_{\partial V_i \cap \partial V_j} \|\phi(t)\|_* \|n_i(t)\| d\mathcal{H}^{d-1}(t) : \phi \in C_c^1(\Omega;\mathbb{R}^d), \|\phi\|_* \le 1\ \forall t \right\} \\
&\quad = \sum_{i,j=1}^{n} |\theta_i - \theta_j| \int_{\partial V_i \cap \partial V_j} \|n_i(t)\| d\mathcal{H}^{d-1}(t),
\end{aligned}$$

where recall $\|\cdot\|, \|\cdot\|_*$ are dual norms. Finally, we obtain a matching lower bound via a mollification argument. The target of our approximating sequence will be a pointwise duality map with respect to $\|\cdot\|$, but first we need to do a little bit of work. Define the function $\phi_0 : \cup_{i,j=1}^{n} \partial V_i \cap \partial V_j \to \mathbb{R}^d$ by

$$\phi_0(t) \in \{g/\|g\|_* : g \in F(n_i(t)), t \in \partial V_i \cap \partial V_j\},$$

and its piecewise constant extension to Ω, $\tilde{\phi}:\Omega\to\mathbb{R}^d$ by

$$\tilde{\phi}(x)=\phi_0\left(t\in\underset{t}{\operatorname{argmin}}\{\|x-t\|_2 : t\in\cup_{i,j=1}^n\partial V_i\cap\partial V_j\}\right),$$

where for a Banach space E and its continuous dual E^*, we write $F:E\to P(E^*)$ for the dual map defined by

$$F(x_0)=\left\{f_0\in E^* : \|f_0\|_{E^*}=\|x_0\|_E \text{ and } \langle f_0,x_0\rangle_{(E,E^*)}=\|x_0\|_E^2\right\},$$

and moreover when E^* is strictly convex, the duality map is singleton-valued (Brezis, 2011). Observe that $\tilde{\phi}\in L^p_{\text{loc}}(\Omega)$, $1\le p<\infty$, so there exists an approximating sequence $\tilde{\phi}_k\in C_c^\infty(\Omega,\mathbb{R}^d)$, $k=1,2,3,\ldots$, such that $\lim_{k\to\infty}\tilde{\phi}_k\to\tilde{\phi}$ μ-a.e. We invoke Fatou's Lemma and properties of the duality map to obtain a matching lower bound,

$$\begin{aligned}
&\mathrm{TV}(f;\Omega,\|\cdot\|)\\
&\quad=\sup\left\{\sum_{i,j=1}^n|\theta_i-\theta_j|\int_{\partial V_i\cap\partial V_j}\langle\phi(t),n_i(t)\rangle d\mathcal{H}^{d-1}(t) : \phi\in C_c^1(\Omega;\mathbb{R}^d),\|\phi\|_*\le 1\ \forall t\right\}\\
&\quad\ge\sum_{i,j=1}^n|\theta_i-\theta_j|\liminf_{k\to\infty}\int_{\partial V_i\cap\partial V_j}\langle\tilde{\phi}_k(t),n_i(t)\rangle d\mathcal{H}^{d-1}(t)\\
&\quad\ge\sum_{i,j=1}^n|\theta_i-\theta_j|\int_{\partial V_i\cap\partial V_j}\left\langle\liminf_{k\to\infty}\tilde{\phi}_k(t),n_i(t)\right\rangle d\mathcal{H}^{d-1}(t)\\
&\quad=\sum_{i,j=1}^n|\theta_i-\theta_j|\int_{\partial V_i\cap\partial V_j}\left\langle\lim_{k\to\infty}\tilde{\phi}_k(t),n_i(t)\right\rangle d\mathcal{H}^{d-1}(t)\\
&\quad=\sum_{i,j=1}^n|\theta_i-\theta_j|\int_{\partial V_i\cap\partial V_j}\langle\tilde{\phi}(t),n_i(t)\rangle d\mathcal{H}^{d-1}(t)\\
&\quad=\sum_{i,j=1}^n|\theta_i-\theta_j|\int_{\partial V_i\cap\partial V_j}\|n_i(t)\|d\mathcal{H}^{d-1}(t),
\end{aligned}$$

establishing equality. □

B Proofs for Section 3

B.1 Roadmap for the proof of Theorem 1

The proof of Theorem 1 consists of several parts, and we summarize them below. Some remarks on notation: throughout this section, we use σ_{Vor} for the constant c_d appearing in (24), and we abbreviate $\|\cdot\|=\|\cdot\|_2$. Also, we use $C^1(\Omega)$ and $C^2(\Omega)$ to denote the spaces of continuously differentiable and twice continuously differentiable functions, respectively, equipped with the L^∞ norm.

1. An edge $\{i,j\}$ in the Voronoi graph depends not only on x_i and x_j but also on all other design points $x_k, k\ne i,j$. In Lemma S.1, we start by showing that the randomness due this dependence on $x_k, k\ne i,j$ is negligible,

$$\mathbb{E}\Big[\big(\mathrm{DTV}(f;w^{\mathrm{V}})-U_{n,\mathrm{Vor}}(f)\big)^2\Big]\le C\frac{\|f\|^2_{C^1(\Omega)}(\log n)^{(d+2)/d}}{n^{1/d}}, \tag{S.5}$$

for a constant $C>0$. The functional $U_{n,\mathrm{Vor}}(f)$ is an order-2 U-statistic,

$$U_{n,\mathrm{Vor}}(f)=\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n\big|f(x_i)-f(x_j)\big|H_{\mathrm{Vor}}(x_i,x_j),$$

with kernel $H_{\mathrm{Vor}}(x,y)$ defined by

$$H_{\mathrm{Vor}}(x,y)=\mathbb{E}\big[\mathcal{H}^{d-1}(\partial V_i\cap\partial V_j)|x_i,x_j\big]=\int_{L\cap\Omega}\big(1-p_x(z)\big)^{(n-2)}\,dz.$$

Here $L = L_{xy}$ is the $(d-1)$-dimensional hyperplane $L = \{z : \|x - z\| = \|y - z\|\}$, and $p_x(z) = P(B(z, \|x - z\|))$. (Note that $p_x(z) = p_y(z)$ for all $z \in L$).

2. We proceed to separately analyze the variance and bias of $U_{n,\mathrm{Vor}}(f)$. In Lemma S.2, we establish that $U_{n,\mathrm{Vor}}(f)$ concentrates around its mean, giving the estimate, for a constant $C > 0$,

$$\mathrm{Var}\big[U_{n,\mathrm{Vor}}(f)\big] \leq C\frac{(\log n)^3}{n}\|f\|^2_{C^1(\Omega)}. \tag{S.6}$$

3. It remains to analyze the bias, the difference between the expectation of $U_{n,\mathrm{Vor}}(f)$ and continuum TV. Lemma S.3 leverages the fact that the kernel $H_{\mathrm{Vor}}(x, y)$ is close to a spherically symmetric kernel—at least at points x, y sufficiently far from the boundary of Ω—to show that the expectation of the U-statistic $U_{n,\mathrm{Vor}}(f)$ is close to (an appropriately rescaled version of) the nonlocal functional

$$\mathrm{TV}_{\varepsilon,K}\big(f;\Omega,h\big) := \int_\Omega\int_\Omega |f(x) - f(y)| K_{\mathrm{Vor}}\bigg(\frac{\|y - x\|}{\varepsilon(x)}\bigg) h(x)\, dy\, dx, \tag{S.7}$$

for bandwidth $\varepsilon(x) = (np(x))^{-1/d}$, weight $h(x) = (p(x))^{(d+1)/d}$, and kernel $K_{\mathrm{Vor}}(t)$ defined in (S.13). Lemma S.4 in turn shows that this nonlocal functional is close to (a scaling factor) times $\int_\Omega \|\nabla f\|$. Together, these lemmas imply that

$$\lim_{n\to\infty} \mathbb{E}[U_{n,\mathrm{Vor}}(f)] = \sigma_{\mathrm{Vor}} \int_\Omega \|\nabla f(x)\|\, dx. \tag{S.8}$$

Combining (S.5), (S.6), and (S.8) with Chebyshev's inequality implies the consistency result stated in (24). In the rest of this section, across Sections B.2–B.4, we state and prove the various lemmas referenced above.

B.2 Step 1: Voronoi TV approximates Voronoi U-statistic

Lemma S.1 upper bounds the expected squared difference between Voronoi TV and the U-statistic $U_{n,\mathrm{Vor}}(f)$.

Lemma S.1. *Suppose $x_{1:n}$ are sampled independently from a distribution P satisfying A1. There exists a constant $C > 0$ such that for all $n \in \mathbb{N}$ sufficiently large, and any $f \in C^1(\Omega)$,*

$$\mathbb{E}\Big[\big(\mathrm{DTV}(f; w^{\mathrm{V}}) - U_{n,\mathrm{Vor}}(f)\big)^2\Big] \leq C\frac{\|f\|^2_{C^1(\Omega)}(\log n)^{(d+2)/d}}{n^{1/d}}.$$

Proof of Lemma S.1. We begin by introducing some notation and basic inequalities used throughout this proof. Take $\varepsilon_0 = (\log n/n)^{1/d}$. Let $B_x(z) := B^o(z, \|x - z\|)$ denote the open ball centered at z of radius $\|x - z\|$, and note that by our assumptions on p, we have $p_x(z) := P(B_x(z))$. We will repeatedly use the estimates

$$p_x(z) \geq \frac{p_{\min}}{2d}\mu_d\|x - z\|^d,$$

and therefore for $c_1 = \frac{p_{\min}}{2d}\mu_d$,

$$(1 - p_x(z))^n \leq \exp(-c_1 n\|x - z\|^d).$$

It follows by Lemma S.18 that for any constants $a, c > 0$, there exists a constant $C > 0$ depending only on a, c and d such that

$$\int_{L\cap\Omega} (1 - cp_x(z))^n \leq C\bigg(\frac{1\{\|x - y\| \leq C\varepsilon_0\}}{n^{(d-1)/d}} + \frac{1}{n^5}\bigg).$$

We will assume $n \geq 8$, so that the same estimate holds with respect to $n - 4 \geq n/2$. Finally for simplicity write $\Delta(x_i, x_j) := |f(x_i) - f(x_j)|\big(\mathcal{H}^{d-1}(\partial V_i \cap \partial V_j) - H_{\mathrm{Vor}}(x_i, x_j)\big)$.

We note immediately that, because $x_{1:n}$ are identically distributed, it follows from linearity of expectation that

$$\begin{aligned}\mathbb{E}\Big[\big(\mathrm{DTV}_{n,\mathrm{Vor}}(f; w^{\mathrm{V}}) - U_{n,\mathrm{Vor}}(f)\big)^2\Big] &= \binom{n}{2}\mathbb{E}[\big(\Delta(x_1, x_2)\big)^2] \\ &\quad + \binom{n}{3}\mathbb{E}\big[\Delta(x_1, x_2)\Delta(x_1, x_3)\big]\end{aligned}$$

$$
+ \binom{n}{4}\mathbb{E}\big[\Delta(x_1,x_2)\Delta(x_3,x_4)\big]
$$

$$
=: \binom{n}{2}T_1 + \binom{n}{3}T_2 + \binom{n}{4}T_3.
$$

We separately upper bound $|T_1|$ (which will make the main contribution to the overall upper bound) and $|T_2|$ and $|T_3|$ (which will be comparably negligible). In each case, the general idea is to use the fact that the fluctuations of the Voronoi edge weights $\mathcal{H}^{d-1}(\partial V_1 \cap \partial V_2)$ around the conditional expectation $H_{\mathrm{Vor}}(x_1,x_2)$ are small unless x_1 and x_2 are close together.

Upper bound on T_1. We begin by conditioning on x_1, x_2, and considering the conditional expectation

$$
\mathbb{E}\big[(\Delta(x_1,x_2))^2|x_1,x_2\big] = |f(x_1)-f(x_2)|^2\mathrm{Var}(\mathcal{H}^{d-1}(\partial V_1\cap\partial V_2)|x_1,x_2).
$$

By Jensen's inequality,

$$
\begin{aligned}
\mathrm{Var}(\mathcal{H}^{d-1}(\partial V_1\cap\partial V_2)|x_1,x_2) &\le \mathcal{H}^{d-1}(L\cap\Omega)\int_{L\cap\Omega}\mathrm{Var}\big(1\{P_n(B_{x_1}(z))=0\}|x_1\big)\,dz \\
&= \mathcal{H}^{d-1}(L\cap\Omega)\int_{L\cap\Omega}\big(1-p_{x_1}(z)\big)^{(n-2)}\,dz \\
&\le C\Big(\frac{1}{n^{(d-1)/d}}1\{\|x_1-x_2\|\le C\varepsilon_0\}+\frac{1}{n^5}\Big).
\end{aligned}
$$

Taking expectation over x_1 and x_2 gives

$$
\begin{aligned}
T_1 &\le C\Big(\frac{\|f\|_{C^1(\Omega)}^2}{n^{(d-1)/d}}\int_\Omega\int_\Omega\|x-y\|^2 1\{\|x-y\|\le C\varepsilon_0\}\,dy\,dx + \frac{\|f\|_{L^\infty(\Omega)}^2}{n^5}\Big) \\
&\le C\Big(\frac{\|f\|_{C^1(\Omega)}^2\varepsilon_0^{(d+2)}}{n^{(d-1)/d}} + \frac{\|f\|_{L^\infty(\Omega)}^2}{n^5}\Big) \\
&= C\Big(\frac{\|f\|_{C^1(\Omega)}^2(\log n)^{(d+2)/d}}{n^{(2+1/d)}} + \frac{\|f\|_{L^\infty(\Omega)}^2}{n^5}\Big).
\end{aligned}
$$

Upper bound on T_2. Again we begin by conditioning, this time on $x_{1:3}$, meaning we consider

$$
\mathbb{E}\big[\Delta(x_1,x_2)\Delta(x_1,x_3)|x_{1:3}\big] = |f(x_1)-f(x_2)||f(x_1)-f(x_3)|\mathrm{Cov}\big[\mathcal{H}^{d-1}(\partial V_1\cap\partial V_2),\mathcal{H}^{d-1}(\partial V_1\cap\partial V_3)|x_{1:3}\big].
$$

We begin by focusing on this conditional covariance. Write $L = \{z\in\Omega : \|z-x_1\| = \|z-x_2\|\}$ and likewise $L' = \{z\in\Omega : \|z-x_1\| = \|z-x_3\|\}$. Exchanging covariance with integration gives

$$
\begin{aligned}
&\Big|\mathrm{Cov}\big[\mathcal{H}^{d-1}(\partial V_1\cap\partial V_2),\mathcal{H}^{d-1}(\partial V_1\cap\partial V_3)|x_{1:3}\big]\Big| \\
&\le \int_L\int_{L'}\big|\mathrm{Cov}\big[1\{P_n(B_{x_1}(z))=0\},1\{P_n(B_{x_1}(z'))=0\}|x_{1:3}\big]\big|\,dz'\,dz \\
&\overset{(i)}{\le} \int_L\int_{L'}\Big(1-\frac{p_{x_1}(z)+p_{x_1}(z')}{2}\Big)^{(n-3)}\,dz\,dz' \\
&+ \int_L\int_{L'}(1-p_{x_1}(z))^{(n-3)}(1-p_{x_1}(z'))^{(n-3)}\,dz'\,dz \\
&\le C\Big(\frac{1}{n^{(d-1)/d}}1\{\|x_1-x_2\|\le C\varepsilon_0\}+\frac{1}{n^5}\Big)\Big(\frac{1}{n^{(d-1)/d}}1\{\|x_1-x_3\|\le C\varepsilon_0\}+\frac{1}{n^5}\Big) \\
&\le C\Big(\frac{1}{n^{2(d-1)/d}}1\{\|x_1-x_2\|\le C\varepsilon_0\}1\{\|x_1-x_3\|\le C\varepsilon_0\}+\frac{1}{n^5}\Big).
\end{aligned}
\tag{S.9}
$$

The inequality (i) follows first from the standard fact that for positive random variables X and Y, $\big|\mathrm{Cov}[X,Y]\big| \le \mathbb{E}[XY]+\mathbb{E}[Y]\mathbb{E}[X]$, and second from the upper bound

$$
\mathbb{E}\Big[1\{P_n(B_{x_1}(z))=0\},1\{P_n(B_{x_1}(z'))=0\}\Big] \le \Big(1-P\big(B_{x_1}(z)\cup B_{x_1}(z')\big)\Big)^{(n-3)}
$$

$$\leq \Big(1 - \frac{P\big(B_{x_1}(z)\big) + P\big(B_{x_1}(z')\big)}{2}\Big)^{(n-3)}.$$

Taking expectation over $x_{1:3}$, we have

$$\begin{aligned}
T_2 &\leq C\Big(\frac{\|f\|_{C^1(\Omega)}^2}{n^{2(d-1)/d}}\int_\Omega\int_\Omega\int_\Omega \|x-y\|\|x-z\| 1\{\|x-y\| \leq C\varepsilon_0\} 1\{\|x-z\|\leq C\varepsilon_0\}\,dz\,dy\,dx + \frac{\|f\|_{L^\infty(\Omega)}^2}{n^5}\Big) \\
&\leq C\Big(\frac{\|f\|_{C^1(\Omega)}^2 \varepsilon_0^{2(d+1)}}{n^{2(d-1)/d}} + \frac{\|f\|_{L^\infty(\Omega)}^2}{n^5}\Big) \\
&= C\Big(\frac{\|f\|_{C^1(\Omega)}^2 (\log n)^{2(d+1)/d}}{n^4} + \frac{\|f\|_{L^\infty(\Omega)}^2}{n^5}\Big).
\end{aligned}$$

Upper bound on T_3. Again we begin by conditioning, this time on $x_{1:4}$, so that

$$\mathbb{E}\big[\Delta(x_1,x_2)\Delta(x_3,x_4)|x_{1:4}\big] = |f(x_1)-f(x_2)||f(x_3)-f(x_4)|\mathrm{Cov}\big[\mathcal{H}^{d-1}(\partial V_1 \cap \partial V_2), \mathcal{H}^{d-1}(\partial V_3\cap\partial V_4)|x_{1:4}\big],$$

Write $L = \{z \in \Omega : \|z-x_1\| = \|z-x_2\|\}$ and likewise $L' = \{z\in\Omega : \|z-x_3\| = \|z-x_4\|\}$, we focus on the conditional covariance

$$\mathrm{Cov}\big[\mathcal{H}^{d-1}(\partial V_1\cap\partial V_2), \mathcal{H}^{d-1}(\partial V_3\cap\partial V_4)|x_{1:4}\big] = \int_L\int_{L'} \mathrm{Cov}\big[1\{P_n(B_{x_1}(z))=0\}, 1\{P_n(B_{x_3}(z'))=0\}|x_{1:4}\big]\,dz'\,dz$$

We now show that this covariance is very small unless x_1 and x_3 are close. Specifically, suppose $\|x_1-x_3\| > \varepsilon_0$. Then either $\|z-x_1\|\geq\varepsilon_0/3$, or $\|z'-x_3\|\geq\varepsilon_0/3$, or $B_{x_1}(z)\cap B_{x_3}(z')=\emptyset$. In either of the first two cases, we have that

$$\begin{aligned}
&\Big|\mathrm{Cov}\big[1\{P_n(B_{x_1}(z))=0\}, 1\{P_n(B_{x_3}(z'))=0\}|x_{1:4}\big]\Big| \\
&\qquad\leq 2\exp(-\frac{p_{\min}}{4d}(n-4)\|x_1-z\|^d)\exp(-\frac{p_{\min}}{4d}(n-4)\|x_3-z'\|^d)\} \\
&\qquad\leq 2\exp(-\frac{p_{\min}}{4d}(n-4)\varepsilon_0^d) \leq \frac{C}{n^5}.
\end{aligned}$$

In the third case, it follows that $P(B_{x_1}(z)\cup B_{x_3}(z')) = p_{x_1}(z) + p_{x_3}(z)$. Assume $x_3, x_4 \notin B_{x_1}(z)$, and likewise $x_1, x_2 \notin B_{x_3}(z')$, otherwise there is nothing to prove. We use the definition of covariance $\mathrm{Cov}[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ to obtain the upper bound,

$$\begin{aligned}
&\Big|\mathrm{Cov}\big[1\{P_n(B_{x_1}(z))=0\}, 1\{P_n(B_{x_3}(z'))=0\}|x_{1:4}\big]\Big| \\
&= \big|(1-(p_{x_1}(z)+p_{x_3}(z)))^{(n-4)} - (1-p_{x_1}(z))^{(n-4)}(1-p_{x_3}(z))^{(n-4)}\big| \\
&= (1-p_{x_1}(z))^{(n-4)}(1-p_{x_3}(z))^{(n-4)}\Big|\Big(1 - \frac{p_{x_1}(z)p_{x_3}(z)}{(1-p_{x_1}(z))(1-p_{x_3}(z))}\Big)^{(n-4)} - 1\Big| \\
&\leq (1-p_{x_1}(z))^{(n-4)}(1-p_{x_3}(z))^{(n-4)}p_{x_1}(z)p_{x_3}(z)n \\
&\leq p_{\max}^2\mu_d^2\exp(-\frac{p_{\min}}{4d}(n-4)\|x_1-z\|^d)\exp(-\frac{p_{\min}}{4d}(n-4)\|x_2-z\|^d)\|x_1-z\|^d\|x_3-z'\|^d n \\
&\leq C\exp(-\frac{p_{\min}}{4d}(n-4)\|x_1-z\|^d)\exp(-\frac{p_{\min}}{4d}(n-4)\|x_2-z\|^d)\varepsilon_0^{2d}n
\end{aligned}$$

Integrating over z, z', it follows that if $\|x_1-x_3\| > \varepsilon_0$, then

$$\Big|\mathrm{Cov}\big[\mathcal{H}^{d-1}(\partial V_1\cap\partial V_2), \mathcal{H}^{d-1}(\partial V_3\cap\partial V_4)|x_{1:4}\big]\Big| \leq C\Big(\frac{\varepsilon_0^{2d}}{n^{(d-2)/d}}1\{\|x_1-x_2\|\leq C\varepsilon_0\}1\{\|x_3-x_4\|\leq C\varepsilon_0\} + \frac{1}{n^5}\Big).$$

Otherwise $\|x_1-x_3\|\leq\varepsilon_0$, and using the same inequalities as in (S.9), we find that

$$\Big|\mathrm{Cov}\big[\mathcal{H}^{d-1}(\partial V_1\cap\partial V_2), \mathcal{H}^{d-1}(\partial V_3\cap\partial V_4)|x_{1:4}\big]\Big|$$

$$\leq C\Big(\frac{1}{n^{2(d-1)/d}}1\{\|x_1 - x_2\| \leq C\varepsilon_0\}1\{\|x_3 - x_4\| \leq C\varepsilon_0\}\{\|x_1 - x_3\| \leq \varepsilon_0\} + \frac{1}{n^5}\Big).$$

Taking expectation over $x_{1:4}$, we conclude that

$$\begin{aligned}
T_3 \leq C\Big(&\frac{\varepsilon_0^{2d}\|f\|_{C^1(\Omega)}^2}{n^{(d-2)/d}} \int_\Omega\int_\Omega\int_\Omega\int_\Omega \|x-y\|\|h-z\|1\{\|x-y\| \leq C\varepsilon_0\}1\{\|h-z\| \leq C\varepsilon_0\}\, dh\, dz\, dy\, dx \\
&+ \frac{\|f\|_{C^1(\Omega)}^2}{n^{2(d-1)/d}} \int_\Omega\int_\Omega\int_\Omega\int_\Omega \|x-y\|\|h-z\|1\{\|x-y\|C \leq \varepsilon_0\}1\{\|h-z\| \leq C\varepsilon_0, \|x-h\| \leq \varepsilon_0\}\, dh\, dz\, dy\, dx \\
&+ \frac{\|f\|_{L^\infty(\Omega)}^2}{n^5}\Big) \\
\leq C\Big(&\frac{\|f\|_{C^1(\Omega)}^2\varepsilon_0^{4d+2}}{n^{(d-2)/d}} + \frac{\|f\|_{C^1(\Omega)}^2\varepsilon_0^{3d+2}}{n^{2(d-1)/d}} + \frac{\|f\|_{L^\infty(\Omega)}^2}{n^5}\Big) \\
= C\Big(&\frac{\|f\|_{C^1(\Omega)}^2(\log n)^{(4d+2)/d}}{n^5} + \frac{\|f\|_{L^\infty(\Omega)}^2}{n^5}\Big).
\end{aligned}$$

Combining our upper bounds on T_1-T_3 gives the claim of the lemma. □

B.3 Step 2: Variance of Voronoi U-statistic

Lemma S.2 leverages classical theory regarding order-2 U-statistics to show that the Voronoi U-statistic $U_{n,\mathrm{Vor}}(f)$ concentrates around its expectation. This is closely related to an estimate provided in García Trillos et al. (2017), but not strictly implied by that result: it handles a specific kernel H_{Vor} that is not compactly supported, and functions f besides $f(x) = 1\{x \in A\}$ for some $A \subseteq \Omega$.

Lemma S.2. *Suppose $x_{1:n}$ are sampled independently from a distribution P satisfying A1. There exists a constant $C > 0$ such that for any $f \in C^1(\Omega)$,*

$$\mathrm{Var}\big[U_{n,\mathrm{Vor}}(f)\big] \leq C\frac{(\log n)^3}{n}\|f\|_{C^1(\Omega)}^2. \tag{S.10}$$

Lemma S.2 can be strengthened in several respects. Under the assumptions of the lemma, better bounds are available than (S.10) which do not depend on factors of $\log n$. Additionally, under weaker assumptions than $f \in C^1(\Omega)$, it is possible to obtain bounds which are looser than (S.10) but which still imply that $\mathrm{Var}\big[U_{n,\mathrm{Vor}}(f)\big] \to 0$ as $n \to \infty$. Neither of these are necessary to prove Theorem 1, and so we do not pursue them further.

Proof of Lemma S.2. We will repeatedly use the following fact, which is a consequence of Lemma S.18: there exists a constant $C > 0$ not depending on n such that for any $x, y \in \Omega$,

$$H_{\mathrm{Vor}}(x,y) \leq \int_{L\cap\Omega} \exp\big(-(p_{\min}/2d)\|x-z\|^d\big)\, dz \leq C\Big(\frac{1}{n^{(d-1)/d}}1\{\|x-y\| \leq C\varepsilon_0\} + \frac{1}{n^2}\Big). \tag{S.11}$$

Now, we recall from Hoeffding's decomposition of U-statistics (Hoeffding, 1948) that the variance of $U_{n,\mathrm{Vor}}(f)$ can be written as

$$\mathrm{Var}[U_{n,\mathrm{Vor}}(f)] = \frac{1}{4}\Big(n(n-1)\mathrm{Var}[h(x_1,x_2)] + n(n-1)(n-2)\mathrm{Var}[h_1(x_1)]\Big) \tag{S.12}$$

where $h(x,y) = |f(x) - f(y)|H_{\mathrm{Vor}}(x,y)$ and $h_1(x) = \mathbb{E}[h(x_1,x_2)|x_1]$.

We now use (S.11) to upper bound the variance of h and h_1. For h, we have that

$$\begin{aligned}
\mathrm{Var}[h(x_1,x_2)] &\leq \mathbb{E}[h^2(x_1,x_2)] \\
&\leq p_{\max}^2\|f\|_{C^1(\Omega)}^2 \int_\Omega\int_\Omega \|y-x\|^2\big(H_{\mathrm{Vor}}(x,y)\big)^2\, dy\, dx \\
&\leq C\|f\|_{C^1(\Omega)}^2\Big(\frac{1}{n^{2(d-1)/d}}\int_\Omega\int_\Omega \|y-x\|^2 1\{\|x-y\| \leq C\varepsilon_0\}\, dy\, dx + \frac{1}{n^4}\Big)
\end{aligned}$$

$$\leq C\biggl(\varepsilon_0^{3d}\|f\|_{C^1(\Omega)}^2 + \frac{\|f\|_{C^1(\Omega)}^2}{n^4}\biggr).$$

For h_1, we have that for every $x \in \Omega$,

$$\begin{aligned}
|h_1(x)| &\leq \|f\|_{C^1(\Omega)} p_{\max} \int_\Omega |y-x| H_{\mathrm{Vor}}(y,x)\,dy \\
&\leq C\|f\|_{C^1(\Omega)}\biggl(\frac{1}{n^{(d-1)/d}}\int_\Omega |y-x| 1\{\|y-x\| \leq C\varepsilon_0\}\,dy + \frac{1}{n^2}\biggr) \\
&\leq C\|f\|_{C^1(\Omega)}\biggl(\varepsilon_0^{2d} + \frac{1}{n^2}\biggr).
\end{aligned}$$

Integrating over $x \in \Omega$, we conclude that

$$\mathrm{Var}[h_1(x_1)] \leq \mathbb{E}[(h_1(x_1))^2] \leq C\|f\|_{C^1(\Omega)}^2\biggl(\varepsilon_0^{4d} + \frac{1}{n^4}\biggr).$$

Plugging these estimates back into (S.12) gives the upper bound in (S.10). □

B.4 Step 3: Bias of Voronoi U-statistic

Under appropriate conditions, the expectation of $U_{n,\mathrm{Vor}}(f)$ is approximately equal to (an appropriately rescaled version of) the nonlocal functional (S.7) for bandwidth $\varepsilon_{(1)}(x) = (np(x))^{-1/d}$, weight $(p(x))^{(d+1)/d}$, and kernel

$$K_{\mathrm{Vor}}(t) = \int_0^\infty \exp\biggl(-\mu_d\Bigl\{\frac{t^2}{4} + s^2\Bigr\}^{d/2}\biggr) s^{d-2}\,ds. \tag{S.13}$$

Lemma S.3. *Suppose $x_{1:n}$ are sampled independently from a distribution P satisfying A1. For any $f \in C^1(\Omega)$,*

$$\mathbb{E}\bigl[U_{n,\mathrm{Vor}}(f)\bigr] = n^{(d+1)/d}\frac{\eta_{d-2}}{2}\cdot \mathrm{TV}_{\varepsilon_{(1)},K_{\mathrm{Vor}}}\bigl(f;\Omega,p^{(d+1)/d}\bigr) + O\biggl(\frac{(\log n)^{3+1/d}}{n^{1/d}}\|f\|_{C^1(\Omega)}\biggr).$$

Proof. We will use Lemma S.17, which shows that at points $x, y \in \Omega$ sufficiently far from the boundary of Ω, the kernel $H_{\mathrm{Vor}}(x,y)$ is approximately equal to a spherical kernel. To invoke this lemma, we need to restrict our attention to points sufficiently far from the boundary. In particular, letting $h = h_n$ be defined as in Lemma S.17, we conclude from (S.93) that

$$\begin{aligned}
&\int_\Omega\int_\Omega |f(y)-f(x)| H_{\mathrm{Vor}}(x,y)p(y)p(x)\,dy\,dx = \\
&\qquad\int_{\Omega_h}\int_\Omega |f(y)-f(x)| H_{\mathrm{Vor}}(x,y)p(y)p(x)\,dy\,dx + O\biggl(\frac{h}{n^2}\|f\|_{C^1(\Omega)}\biggr),
\end{aligned} \tag{S.14}$$

where we have used the assumption $f \in C^1(\Omega)$ and (S.93) to control the boundary term, since

$$\begin{aligned}
&\int_{\Omega\setminus\Omega_h}\int_\Omega |f(y)-f(x)| H_{\mathrm{Vor}}(x,y)p(y)p(x)\,dy\,dx \\
&\leq \frac{C_3 p_{\max}^2\eta_{d-2}\|f\|_{C^1(\Omega)}}{n^{(d-1)/d}}\int_{\Omega\setminus\Omega_h}\int_\Omega \|y-x\| K_{\mathrm{Vor}}\biggl(\frac{\|y-x\|}{C_4 n^{1/d}}\biggr)\,dy\,dx \\
&\overset{(i)}{\leq} \frac{C_3 C_4^{(d+1)/d} p_{\max}^2\eta_{d-2}\|f\|_{C^1(\Omega)}}{n^2}\int_{\Omega\setminus\Omega_h}\int_{\mathbb{R}^d}\|h\| K_{\mathrm{Vor}}(\|h\|)\,dh\,dx \\
&\overset{(ii)}{\leq} \frac{C_3 C_4^{(d+1)/d} p_{\max}^2\eta_{d-2}\eta_{d-1}\|f\|_{C^1(\Omega)}}{n^2}\int_{\Omega\setminus\Omega_h}\int_0^\infty t^d K_{\mathrm{Vor}}(t)\,dt\,dx \\
&\overset{(iii)}{\leq} \frac{C\|f\|_{C^1(\Omega)}}{n^2}\mu(\Omega\setminus\Omega_h) \\
&\leq \frac{Ch\|f\|_{C^1(\Omega)}}{n^2},
\end{aligned} \tag{S.15}$$

where (i) follows by changing variables $h = (y-x)/C_3 n^{1/d}$, (ii) by converting to polar coordinates, and (iii) upon noticing that $\int_0^\infty t^d K_{\mathrm{Vor}}(t) < \infty$.

Returning to the first-order term in (S.14), we can use (S.92) to replace the integral with H_{Vor} by an integral with the Voronoi kernel K_{Vor}. Precisely,

$$
\begin{aligned}
&\int_{\Omega_h}\int_\Omega |f(y)-f(x)| H_{\mathrm{Vor}}(x,y)p(y)p(x)\,dy\,dx \\
&\qquad = \frac{\eta_{d-2}}{n^{(d-1)/d}}\int_{\Omega_h}\int_\Omega |f(y)-f(x)| K_{\mathrm{Vor}}\biggl(\frac{\|x-y\|}{\varepsilon_{(1)}}\biggr)p(y)\bigl(p(x)\bigr)^{1/d}\,dy\,dx \\
&\qquad\quad + O\biggl(\frac{1}{n^3}\int_\Omega\int_\Omega |f(y)-f(x)|\,dy\,dx\biggr) \\
&\qquad\quad + O\biggl(\frac{(\log n)^2}{n}\int_\Omega\int_\Omega |f(y)-f(x)| 1\Bigl\{\|x-y\| \le C(\log n/n)^{1/d}\Bigr\}\,dy\,dx\biggr) \\
&\qquad = \frac{\eta_{d-2}}{n^{(d-1)/d}}\int_{\Omega_h}\int_\Omega |f(y)-f(x)| K_{\mathrm{Vor}}\biggl(\frac{\|x-y\|}{\varepsilon_{(1)}}\biggr)p(y)\bigl(p(x)\bigr)^{1/d}\,dy\,dx \\
&\qquad\quad + O\biggl(\frac{\|f\|_{C^1(\Omega)}}{n^3} + \frac{(\log n)^{3+1/d}}{n^{2+1/d}}\|f\|_{C^1(\Omega)}\biggr) \\
&\qquad = \frac{\eta_{d-2}}{n^{(d-1)/d}}\int_{\Omega}\int_\Omega |f(y)-f(x)| K_{\mathrm{Vor}}\biggl(\frac{\|x-y\|}{\varepsilon_{(1)}}\biggr)p(y)\bigl(p(x)\bigr)^{1/d}\,dy\,dx \\
&\qquad\quad + O\biggl(\frac{\|f\|_{C^1(\Omega)}}{n^3} + \frac{(\log n)^{3+1/d}}{n^{2+1/d}}\|f\|_{C^1(\Omega)} + \frac{h\|f\|_{C^1(\Omega)}}{n^2}\biggr),
\end{aligned}
\tag{S.16}
$$

with the second equality following from the upper bound (S.39), and the third equality from exactly the same argument as in (S.15). Finally, we use the Lipschitz property of p to conclude that

$$
\begin{aligned}
&\int_\Omega\int_\Omega |f(y)-f(x)| K_{\mathrm{Vor}}\biggl(\frac{\|x-y\|}{\varepsilon_{(1)}}\biggr)p(y)\bigl(p(x)\bigr)^{1/d}\,dy\,dx \\
&\qquad = \int_\Omega\int_\Omega |f(y)-f(x)| K_{\mathrm{Vor}}\biggl(\frac{\|x-y\|}{\varepsilon_{(1)}}\biggr)\bigl(p(x)\bigr)^{(d+1)/d}\,dy\,dx + O\biggl(\frac{\|f\|_{C^1(\Omega)}}{n^{(d+2)/2}}\biggr),
\end{aligned}
\tag{S.17}
$$

since

$$
\begin{aligned}
&\int_\Omega\int_\Omega |f(y)-f(x)| K_{\mathrm{Vor}}\biggl(\frac{\|x-y\|}{\varepsilon_{(1)}}\biggr)|p(y)-p(x)|\bigl(p(x)\bigr)^{1/d}\,dy\,dx \\
&\quad \le C\|f\|_{C^1(\Omega)} p_{\max}^{1/d}\int_\Omega\int_\Omega \|y-x\|^2 K_{\mathrm{Vor}}\biggl(\frac{\|x-y\|}{\varepsilon_{(1)}}\biggr)\,dy\,dx \\
&\quad \le C\frac{\|f\|_{C^1(\Omega)} p_{\max}^{1/d}}{p_{\min}^{1/d} n^{(2+d)/d}}\int_\Omega\int_{\mathbb{R}^d} \|h\|^2 K_{\mathrm{Vor}}(\|h\|)\,dh\,dx \\
&\quad = C\frac{\|f\|_{C^1(\Omega)} p_{\max}^{1/d}\eta_{d-1}}{p_{\min}^{1/d} n^{(2+d)/d}}\int_\Omega\int_0^\infty t^{d+1} K_{\mathrm{Vor}}(t)\,dt\,dx \\
&\quad \le C\frac{\|f\|_{C^1(\Omega)}}{n^{(2+d)/d}},
\end{aligned}
$$

with the last inequality following since $\int_0^\infty t^{d+1}K_{\mathrm{Vor}}(t)\,dt = C < \infty$. Combining (S.14), (S.16) and (S.17) yields the final claim. □

Finally, Lemma S.4 shows that the kernelized TV $\mathrm{TV}_{\varepsilon,K}(f;\Omega,h)$ converges to a continuum TV under appropriate conditions.

Assumption A2. The bandwidth $\varepsilon(x) = \bar{\varepsilon}_n g(x)$ for a sequence $\bar{\varepsilon}_n \to 0$ and a bounded function $g \in L^\infty(\Omega)$. The kernel function K satisfies $\int_0^\infty K(t)t^{d+1}\,dt < \infty$. The weight function $h \in L^\infty(\Omega)$.

Note that Assumption A1 implies that Assumption A2 is satisfied by bandwidth $\varepsilon_{(1)}$, kernel K_{Vor} and weight function $h = p^{(d+1)/d}$.

Lemma S.4. *Assuming A2, for any $f \in C^2(\Omega)$,*

$$\lim_{n\to\infty} (\bar{\varepsilon}_n)^{-(d+1)} \mathrm{TV}_{\varepsilon,K}(f;\Omega,h) = \sigma_K \int_\Omega \|\nabla f(x)\| h(x) (g(x))^{d+1}\, dx \tag{S.18}$$

where

$$\sigma_K := \frac{2\eta_{d-2}}{(d-1)} \int_0^\infty K(t) t^d\, dt. \tag{S.19}$$

Proof. The proof of Lemma S.4 follows closely the proof of some related results, e.g., Lemma 4.2 of García Trillos and Slepčev (2016). We begin by summarizing the major steps.

1. We use a 2nd-order Taylor expansion to replace differencing by derivative inside the nonlocal TV.
2. Naturally, the nonlocal TV behaves rather differently than a local functional near the boundary of Ω. We show that the contribution of the integral near the boundary is negligible.
3. Finally, we reduce from a double integral to a single integral involving the norm $\|\nabla f\|$.

Step 1: Taylor expansion. Since $f \in C^2(\Omega)$ we have that

$$f(y) - f(x) = \nabla f(x)^\top (y-x) + O(\|f\|_{C^2(\Omega)} \|y-x\|^2).$$

Consequently,

$$\mathrm{TV}_{\varepsilon,K}(f;\Omega,h) = \int_\Omega \int_\Omega \Big(|\nabla f(x)^\top (y-x)| + O(\|f\|_{C^2(\Omega)}) \Big) K\left(\frac{\|y-x\|}{\varepsilon(x)}\right) h(x)\, dy\, dx.$$

We now upper bound the contribution of the $O(\|y-x\|^2)$-term. For each $x \in \Omega$,

$$\int_\Omega \|y-x\| K\left(\frac{\|y-x\|^2}{\varepsilon(x)}\right) dy \le C|\varepsilon_n(x)|^{d+2} \int_{\mathbb{R}^d} \|z\|^2 K(\|z\|)\, dz \le C|\varepsilon_n(x)|^{d+2} \le C|\varepsilon_n(x)|^{d+2},$$

with the final inequality following from the assumption $\int_0^\infty t^{d+1} K(t)\, dt < \infty$. Integrating over Ω gives the upper bound

$$\int_\Omega \int_\Omega O(\|f\|_{C^2(\Omega)} \|y-x\|^2) K\left(\frac{\|y-x\|}{\varepsilon(x)}\right) h(x)\, dy\, dx = O(\|f\|_{C^2(\Omega)} \bar{\varepsilon}_n^{d+2}),$$

recalling that $h(x), g(x) \in L^\infty(\Omega)$.

Step 2: Contribution of boundary to nonlocal TV. Take $r = r_n$ to be any sequence such that $r_n/\bar{\varepsilon}_n \to \infty, r_n \to 0$. Breaking up the integrals in the definition of nonlocal TV gives

$$\begin{aligned}
\int_\Omega \int_\Omega |\nabla f(x)^\top (y-x)| K\left(\frac{\|y-x\|}{\varepsilon(x)}\right) h(x)\, dy\, dx &= \int_{\Omega_r} \int_{\mathbb{R}^d} |\nabla f(x)^\top (y-x)| K\left(\frac{\|y-x\|}{\varepsilon(x)}\right) h(x)\, dy\, dx \\
&\quad - \int_{\Omega_r} \int_{\mathbb{R}^d \setminus \Omega} |\nabla f(x)^\top (y-x)| K\left(\frac{\|y-x\|}{\varepsilon(x)}\right) h(x)\, dy\, dx \\
&\quad + \int_{\Omega \setminus \Omega_r} \int_{\Omega} |\nabla f(x)^\top (y-x)| K\left(\frac{\|y-x\|}{\varepsilon(x)}\right) h(x)\, dy\, dx \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

Now we are going to show that I_2 and I_3 are negligible. For I_2, noting that $r/\varepsilon(x) \to \infty$ for all x, we have that for any $x \in \Omega_r$,

$$\int_{\mathbb{R}^d \setminus \Omega} |\nabla f(x)^\top (y-x)| K\left(\frac{\|y-x\|}{\varepsilon(x)}\right) h(x)\, dy \le \|f\|_{C^1(\Omega)} \int_{\mathbb{R}^d \setminus \Omega} K\left(\frac{\|y-x\|}{\varepsilon(x)}\right) \|y-x\|\, dy$$

$$
\begin{aligned}
&\leq \|f\|_{C^1(\Omega)}(\varepsilon(x))^1 \int_{\mathbb{R}^d \setminus B(0,r/\varepsilon(x))} \|z\|K(\|z\|)\,dz \\
&\overset{(i)}{\leq} C\|f\|_{C^1(\Omega)}(\varepsilon(x))^{d+1} \int_{r/\varepsilon(x)}^{\infty} t^{d+1}K(t)\,dt \\
&\overset{(ii)}{=} o(\|f\|_{C^1(\Omega)}(\varepsilon(x))^{d+1}),
\end{aligned}
$$

where (i) follows from converting to polar coordinates and (ii) follows by the assumption $\int_0^\infty t^{d+1}K(t)\,dt < \infty$. Integrating over x yields $I_2 = o(\|f\|_{C^1(\Omega)}\bar{\varepsilon}_n^{d+1})$, since $h, g \in L^\infty(\Omega)$.

On the other hand for I_3, similar manipulations show that for every $x \in \Omega$,

$$
\int_\Omega |\nabla f(x)^\top (y-x)| K\left(\frac{\|y-x\|}{\varepsilon(x)}\right) dy \leq C\|f\|_{C^1(\Omega)}(\varepsilon(x))^{d+1}.
$$

Noting that the tube $\Omega \setminus \Omega_r$ has volume at most Cr, we conclude that

$$
I_3 \leq C\|f\|_{C^1(\Omega)}(\varepsilon(x))^{d+1}\mu(\Omega \setminus \Omega_r) \leq Cr\|f\|_{C^1(\Omega)}(\varepsilon(x))^{d+1} = o(\|f\|_{C^1(\Omega)}(\varepsilon(x))^{d+1}),
$$

with the last inequality following since $r = o(1)$.

Step 3: Double integral to single integral. Now we proceed to reduce the double integral in I_1 to a single integral. Changing variables to $z = (y-x)/\varepsilon(x)$, converting to polar coordinates, and letting $w(x) = \nabla f(x)/\|\nabla f(x)\|$, we have that

$$
\begin{aligned}
\int_{\mathbb{R}^d} \|\nabla f(x)^\top (y-x)| K\left(\frac{\|y-x\|}{\varepsilon(x)}\right) dy &= (\varepsilon(x))^{d+1} \int_{\mathbb{R}^d} |\nabla f(x)^\top z| K(\|z\|)\,dz \\
&= (\varepsilon(x))^{d+1} \left(\int_{\mathbb{S}^{d-1}} |\nabla f(x)^\top \phi|\,d\mathcal{H}^{d-1}\right)\left(\int_0^\infty t^d K(t)\,dt\right) \\
&= (\varepsilon(x))^{d+1}\|\nabla f(x)\| \left(\int_{\mathbb{S}^{d-1}} |w(x)^\top \phi|\,d\mathcal{H}^{d-1}\right)\left(\int_0^\infty t^d K(t)\,dt\right) \\
&= (\varepsilon(x))^{d+1}\|\nabla f(x)\| \left(\int_{\mathbb{S}^{d-1}} |\phi_1|\,d\mathcal{H}^{d-1}\right)\left(\int_0^\infty t^d K(t)\,dt\right) \\
&= \sigma_K(\varepsilon(x))^{d+1}\|\nabla f(x)\|,
\end{aligned}
$$

with the second to last equality following from the spherical symmetry of the integral, and the last equality by definition of σ_K. Integrating over $x \in \Omega_r$ gives

$$
\begin{aligned}
I_1 &= \sigma_K \bar{\varepsilon}_n^{d+1} \int_{\Omega_r} \|\nabla f(x)\| h(x)(g(x))^{d+1}\,dx \\
&= \sigma_K \bar{\varepsilon}_n^{d+1} \int_{\Omega} \|\nabla f(x)\| h(x)(g(x))^{d+1}\,dx + o(\bar{\varepsilon}_n^{d+1}\|f\|_{C^1(\Omega)}),
\end{aligned}
$$

with the second equality following from the same reasoning as was used in analyzing the integral I_3.

Putting the pieces together. We conclude that

$$
\begin{aligned}
&(\bar{\varepsilon}_n)^{-(d+1)}\mathrm{TV}_{\varepsilon,K}(f;\Omega,h) \\
&\quad= (\bar{\varepsilon}_n)^{-(d+1)} \int_\Omega \int_\Omega \Big(|\nabla f(x)^\top (y-x)|\Big) K\left(\frac{\|y-x\|}{\varepsilon(x)}\right) h(x)\,dy\,dx + O(\bar{\varepsilon}_n\|f\|_{C^2(\Omega)}) \\
&\quad= (\bar{\varepsilon}_n)^{-(d+1)} \int_{\Omega_r} \int_{\mathbb{R}^d} \Big(|\nabla f(x)^\top (y-x)|\Big) K\left(\frac{\|y-x\|}{\varepsilon(x)}\right) h(x)\,dy\,dx + O(\bar{\varepsilon}_n\|f\|_{C^2(\Omega)}) + o(\|f\|_{C^1(\Omega)}) \\
&\quad= \sigma_K \int_\Omega \int_\Omega \|\nabla f(x)\| h(x)(g(x))^{d+1}\,dx + O(\bar{\varepsilon}_n\|f\|_{C^2(\Omega)}) + o(\|f\|_{C^1(\Omega)}),
\end{aligned}
$$

completing the proof of Lemma S.4. □

C Sensitivity analysis for Section 4

In Section 4, we chose the scale k, ε in the k-nearest neighbor and ε-neighborhood graphs to be such that their average degree would roughly match that of the Voronoi adjacency graph, and we remarked that mildly better results are attainable if one increases the connectivity of the graphs. Here, we present an analogous set of results to those found in Section 4, where the average degree of the k-nearest neighbor and ε-neighborhood graphs are roughly twice that of the graphs in Section 4. All other details of the experimental setup remain the same.

- In Figure S.1, the estimates of TV by the k-nearest neighbor and ε-neighborhood graphs approach their density-weighted limits more quickly than in Section 4, with slightly narrower variability bands.
- In Figure S.2, we see that ε-neighborhood TV denoising is now competitive with k-nearest neighbor TV denoising and the unweighted Voronoigram for the "low inside tube" setting. In the "high inside tube" and uniform sampling settings, the performance of k-nearest neighbor TV denoising improves slightly.

As previously remarked, the Voronoigram has no such auxiliary tuning parameter, so the weighted and unweighted Voronoigram results here are the same as in Section 4. We also note that with greater connectivity in the k-nearest neighbor and ε-neighborhood graphs comes greater computational burden in storing the graphs, as well as performing calculations with them. Therefore, it is advantageous to the practitioner to use the sparsest graph capable of achieving favorable performance.

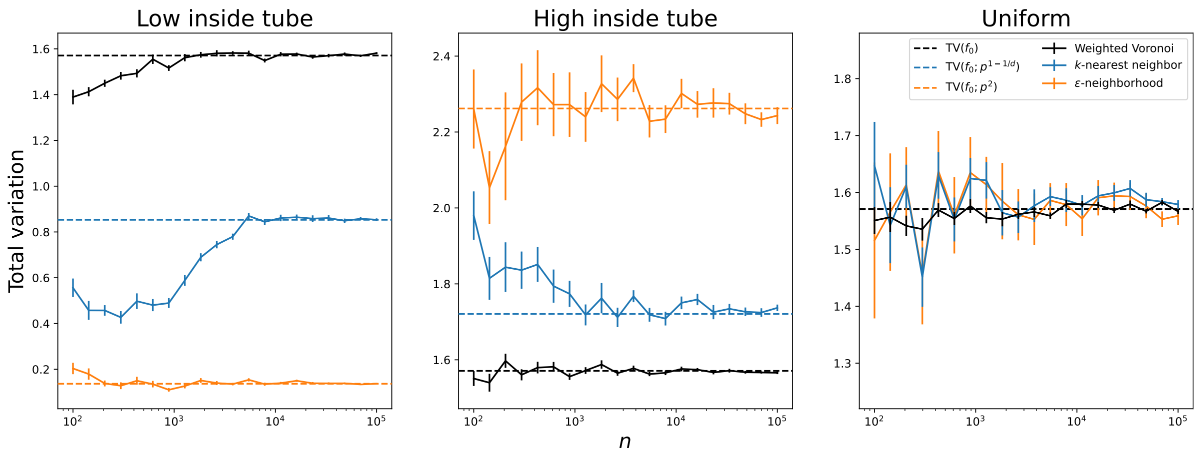

Figure S.1: *Results from the TV estimation experiment, with greater connectivity in the kNN and ε-neighborhood graphs. Compare these results to those in Figure 3.*

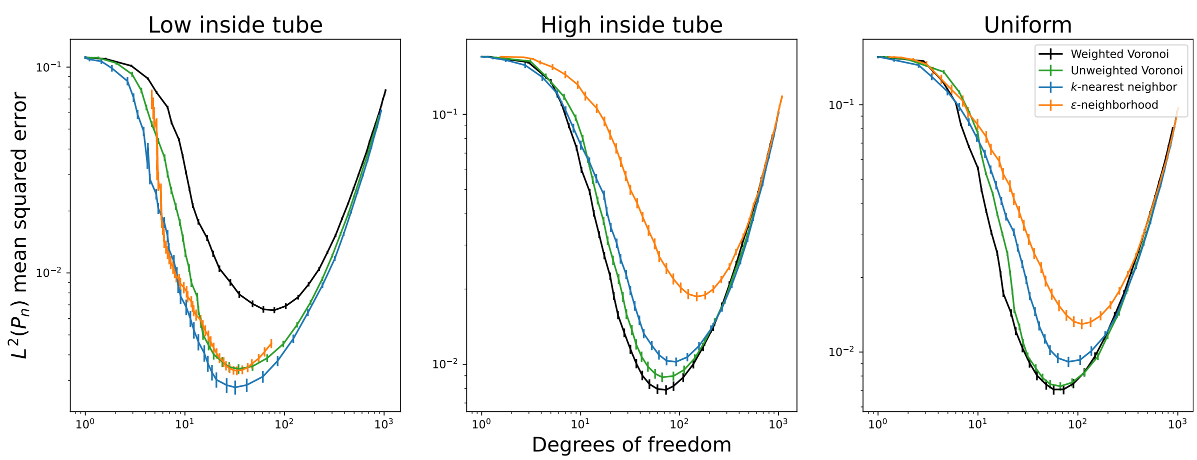

Figure S.2: *Results from the function estimation experiment, with greater connectivity in the kNN and ε-neighborhood graphs. Compare these results to those in Figure 5.*

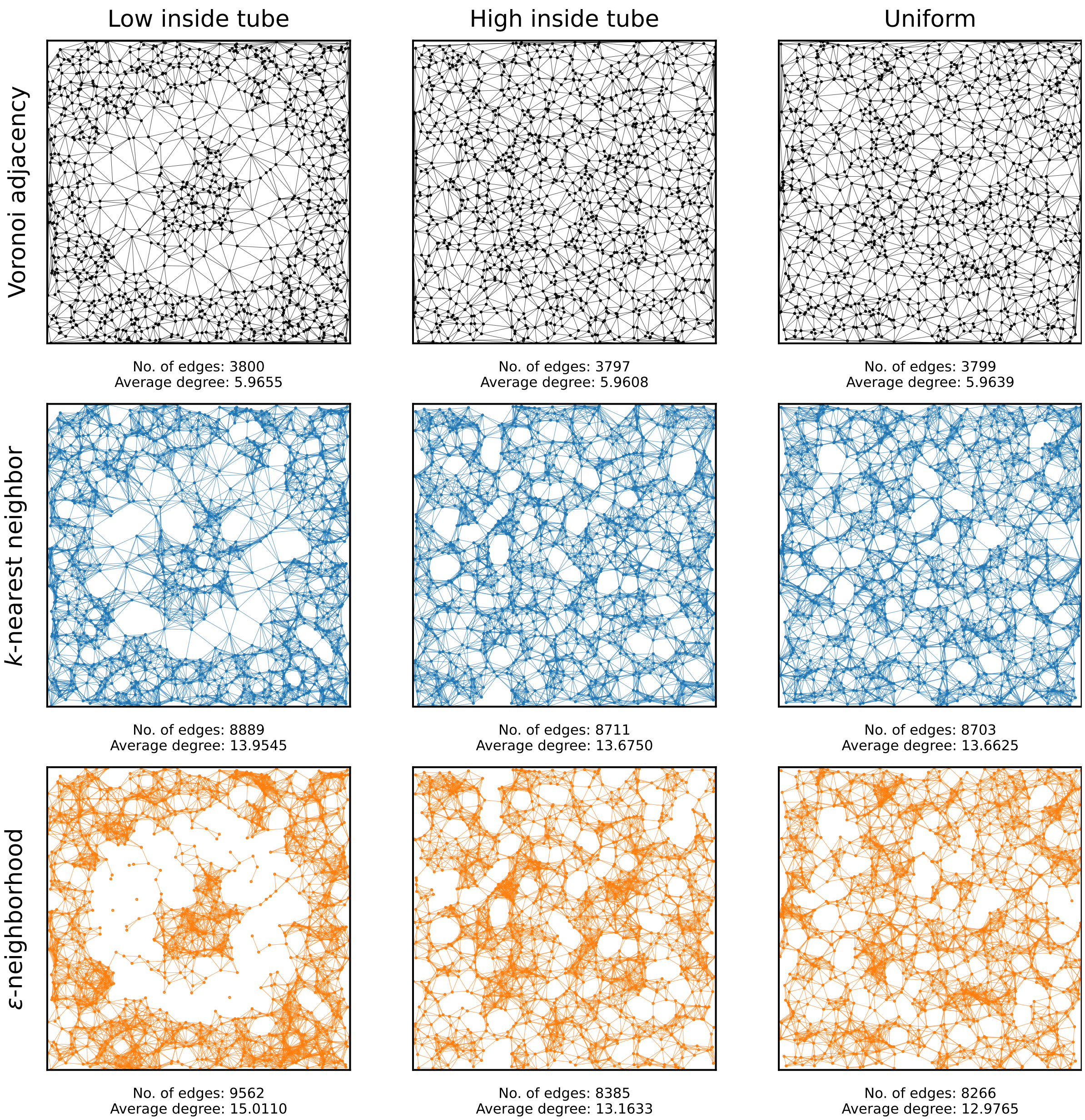

Figure S.3: *Visualization of the Voronoi, kNN, and ε-neighborhood graphs, with greater connectivity in the latter two graphs. (The Voronoi graph does not have such an auxiliary tuning parameter.) Compare these graphs to those in Figure 4.*

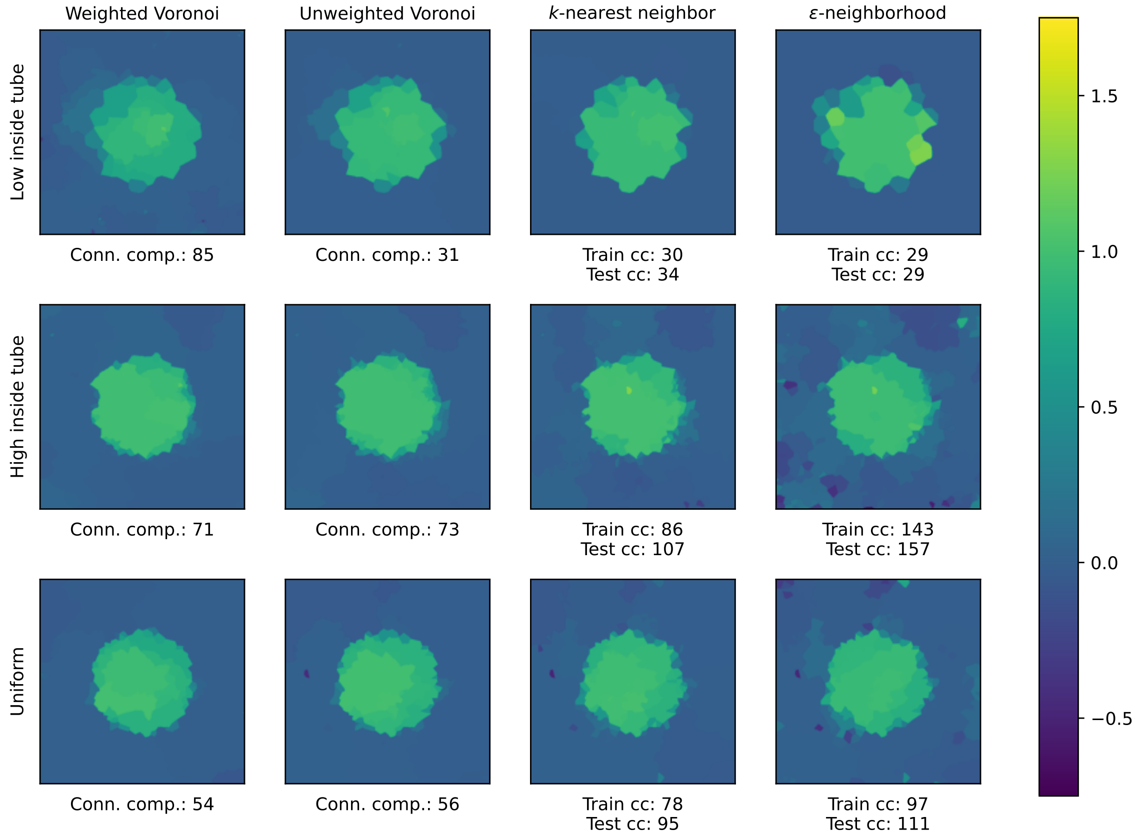

Figure S.4: *Extrapolants from graph TV denoising, with greater connectivity in the kNN and ε-neighborhood graphs. Compare these results to those in Figure 6.*

D Proofs for Section 5

D.1 Proof of Theorem 2

From (35), in the discussion preceding Lemma 3, we have

$$\mathbb{E}\|\hat{f} - f_0\|_{L^2(P)}^2 \le \mathbb{E}\Big[K_n\|\hat{f} - f_0\|_{L^2(P_n)}^2\Big] + 2\mathbb{E}\|\bar{f}_0 - f_0\|_{L^2(P)}^2, \tag{S.20}$$

where

$$K_n = 2p_{\max} n \cdot \Big(\max_{i=1,\dots,n} \mu(V_i)\Big).$$

The second term is bounded by Lemma 3. We now outline the analysis of the first term. As in the $L^2(P_n)$ case we will decompose the error into the case where the design points are well-spaced and the case where they are not. This is formalized by the set $\mathscr{X} = \mathscr{X}_1 \cap \mathscr{X}_2$, where $\mathscr{X}_1, \mathscr{X}_2$ are defined in Appendix F. $x_{1:n}$ falls within this set with probability at least $1 - 3/n^4$, and notably on this set,

$$\max_i \mu(V_i) \le C_1 \log n / n,$$

for some $C_1 > 0$, since $\mathscr{X}_2$ is the set upon which the conclusion of Lemma S.15 holds. We proceed by conditioning,

$$\begin{aligned}\mathbb{E}\left[K_n\|\hat{f} - f_0\|_{L^2(P_n)}^2\right] = 2p_{\max}\Bigg(&\mathbb{E}_x\left[\mathbb{E}_{z|x}\left[\max_i \mu(V_i)\|\hat{\theta} - \theta_0\|_2^2\right] 1\{x_{1:n} \in \mathscr{X}\}\right] \\ &+ \mathbb{E}_x\left[\mathbb{E}_{z|x}\left[\max_i \mu(V_i)\|\hat{\theta} - \theta_0\|_2^2\right] 1\{x_{1:n} \notin \mathscr{X}\}\right]\Bigg).\end{aligned} \tag{S.21}$$

Using the fact that $x_{1:n} \in \mathscr{X}$, the first term on the RHS of (S.21) may be bound,

$$\mathbb{E}_{z|x}\left[\max_i \mu(V_i)\|\hat{\theta}-\theta_0\|_2^2\right] 1\{x_{1:n}\in\mathscr{X}\} \le C_1(\log n)\,\mathbb{E}_{z|x}\left[\frac{1}{n}\|\hat{\theta}-\theta_0\|_2^2\right]\cdot 1\{x_{1:n}\in\mathscr{X}\}$$
$$\le C_2(\log n)\left(\frac{\lambda\|D\theta_0\|}{n}+\frac{\log^\alpha n}{n}\right), \tag{S.22}$$

where the latter inequality is obtained by following the analysis of Lemma 1. For the second term on the RHS of (S.21), we apply the crude upper bound that $\mu(V_i) \le \mu(\Omega) = 1$ for all $i = 1, \ldots, n$. Then apply (S.70) to obtain,

$$\begin{aligned}\mathbb{E}_{z|x}\left[\max_i \mu(V_i)\|\hat{\theta}-\theta_0\|_2^2\right] 1\{x_{1:n}\notin\mathscr{X}\} &\le \mathbb{E}_{z|x}\left[16\|z_{1:n}\|_2^2+2\lambda\|D\theta_0\|_1\right] 1\{x_{1:n}\notin\mathscr{X}\}\\ &= (16n+2\lambda\|D\theta_0\|_1)\,1\{x_{1:n}\notin\mathscr{X}\}.\\ &\le \left(16n+4n^2\lambda\|\theta_0\|_\infty\|w\|_\infty\right)1\{x_{1:n}\notin\mathscr{X}\}.\\ &\le \left(16n+4n^2\lambda\|\theta_0\|_\infty\right)1\{x_{1:n}\notin\mathscr{X}\},\end{aligned} \tag{S.23}$$

where we also use crude upper bounds on the discrete TV. Substitute (S.22) and (S.23) into (S.21) to obtain,

$$\begin{aligned}\mathbb{E}\left[K_n\|\hat{f}-f_0\|_{L^2(P_n)}^2\right] &\le C_3\left(\frac{(\log n)\lambda\mathbb{E}\|D\theta_0\|}{n}+\frac{(\log n)^{1+\alpha}}{n}+\lambda n^2\mathbb{P}\{x_{1:n}\notin\mathscr{X}\}\right)\\ &\le C_4\left(\frac{(\log n)\lambda\mathbb{E}\|D\theta_0\|}{n}+\frac{(\log n)^{1+\alpha}}{n}+\frac{\lambda}{n^2}\right)\\ &\le C_5\left(\frac{\sigma\tau_n(\log n)^{3/2+\alpha}\mathbb{E}\|D\theta_0\|}{n}+\frac{(\log n)^{1+\alpha}}{n}\right),\end{aligned} \tag{S.24}$$

where in the final line we have substituted in the value of $\lambda = c\sigma\tau_n(\log n)^{1/2+\alpha}$. Apply Lemma 2 to (S.24) and substitute back into (S.20) to obtain the claim. □

D.2 Proof of Theorem 3

To establish the lower bound in (30), we follow a classical approach, similar to that outlined in (del Álamo et al., 2021): first we reduce the problem to estimating binary sequences, then we apply Assouad's lemma (Lemma S.5). This results in a constrained maximization problem, which we analyze to establish the ultimate lower bound.

Step 1: Reduction to estimating binary sequences. We begin by associating functions f_θ with vertices of the hypercube $\Theta_S = \{0,1\}^S$, where $S \subseteq [m]^d$ for some $m \in \mathbb{N}$. To construct these functions f_θ, we partition Ω into cubes,

$$Q_i = \frac{1}{m}(i_1-1, i_1)\times\cdots\times\frac{1}{m}(i_d-1,i_d), \quad \text{for } i \in [m]^d,$$

and for each $\theta \in \Theta_S$ take f_θ to be the piecewise constant function

$$f_\theta(x) := a\cdot\sum_{i\in S}\theta_i 1_{Q_i}(x), \tag{S.25}$$

where $1_{Q_i}(x) = 1(x \in Q_i)$ is the characteristic function of Q_i. Observe that for all $\theta \in \Theta_S$, letting $\epsilon := 1/m$,

$$\mathrm{TV}(f_\theta) \le 2da|S|\epsilon^{d-1}, \quad \text{and} \quad \|f_\theta\|_{L^\infty(\Omega)} \le a. \tag{S.26}$$

So long as the constraints in (S.26) are satisfied $\{f_\theta : \theta \in \Theta_S\} \subseteq \mathrm{BV}_\infty(L, M)$, and consequently

$$\inf_{\widehat{f}}\sup_{f_0\in\mathrm{BV}_\infty(L,M)}\mathbb{E}_{f_0}\|\widehat{f}-f_0\|_{L^2(\Omega)}^2 \ge \inf_{\widehat{f}}\max_{\theta\in\Theta_S}\mathbb{E}_\theta\|\widehat{f}-f_\theta\|_{L^2(\Omega)}^2 \ge \frac{a^2\epsilon^d}{4}\inf_{\widehat{\theta}}\max_{\theta\in\Theta}\mathbb{E}_\theta\rho(\widehat{\theta},\theta), \tag{S.27}$$

where $\rho(\theta,\theta') = \sum_{i\in S} |\theta_i - \theta'_i|$ is the Hamming distance between vertices $\theta, \theta' \in \Theta_S$. The second inequality in (S.27) is verified as follows: for a given $\widehat{f}$, letting

$$\widehat{\theta}_i = \begin{cases} 1, & \text{if } \oint_{Q_i} \widehat{f}(x)\,dx \geq a/2, \\ 0, & \text{otherwise,} \end{cases}$$

it follows that

$$\begin{aligned} \|\widehat{f} - f_\theta\|_{L^2(P)}^2 &= \sum_{i\in[m]^d} \|\widehat{f} - f_\theta\|_{L^2(Q_i)}^2 \\ &\geq \sum_{i\in S} \|\widehat{f} - f_\theta\|_{L^2(Q_i)}^2 \\ &\geq \frac{a^2\epsilon^d}{4} \sum_{i\in S} 1\{\widehat{\theta}_i \neq \theta_i\}. \end{aligned}$$

Step 2: application of Assouad's lemma. Given a measurable space $(\mathcal{Z}, \mathcal{A})$, and a set of probability measures $\mathcal{M} = \{\mu_\theta : \theta \in \Theta\}$ on $(\mathcal{Z}, \mathcal{A})$, Assouad's lemma lower bounds the minimax risk over Θ_S, when loss is measured by the Hamming distance $\rho(\widehat{\theta}, \theta) := \sum_{i\in S} |\widehat{\theta}_i - \theta_i|$. We use a form of Assouad's lemma given in Tsybakov (2009).

Lemma S.5 (Lemma 2.12 of Tsybakov (2009))**.** *Suppose that for each $\theta, \theta' \in \Theta_S : \rho(\theta,\theta') = 1$, we have that $\mathrm{KL}(\mu_\theta, \mu_{\theta'}) \leq \alpha < \infty$. It follows that*

$$\inf_{\widehat{\theta}} \sup_{\theta\in\Theta_S} \mathbb{E}_\theta \rho(\widehat{\theta}, \theta) \geq \frac{|S|}{2} \max\left(\frac{1}{2}\exp(-\alpha), (1 - \sqrt{\alpha/2})\right).$$

To apply Assouad's lemma in our context, we take $\mathcal{Z} = (\Omega \times \mathbb{R})^{\otimes n}$, and associate each $\theta \in \Theta_S$ with the measure $\mu_\theta^{(n)}$, the n-times product of measure $\mu_\theta = \mathrm{Unif}(\Omega) \times N(f_\theta(x), 1)$. We now lower bound the KL divergence $\mathrm{KL}(\mu_\theta^{(n)}, \mu_{\theta'}^{(n)})$ when $\rho(\theta,\theta') = 1$; letting $i \in S$ be the single index at which $\theta_i \neq \theta'_i$,

$$\begin{aligned} \mathrm{KL}(\mu_\theta, \mu_{\theta'}) &= \int_{\Omega\times\mathbb{R}} \log\left(\frac{\phi(y - f_\theta(x))}{\phi(y - f_{\theta'}(x))}\right) \phi(y - f_\theta(x))\,dy\,dx \\ &= \int_{Q_i\times\mathbb{R}} \log\left(\frac{\phi(y - a\theta_i)}{\phi(y - a\theta'_i)}\right) \phi(y - a\theta_i)\,dy\,dx \\ &= \epsilon^d \int_{\mathbb{R}} \log\left(\frac{\phi(y - a\theta_i}{\phi(y - a\theta'_i)}\right) \phi(y - a\theta_i)\,dy \\ &= \frac{\epsilon^d a^2}{2}, \end{aligned}$$

and it follows that $\mathrm{KL}(\mu_\theta^{(n)}, \mu_{\theta'}^{(n)}) \leq n\epsilon^d a^2/2$. Consequently, so long as (S.26) is satisfied and

$$\frac{n\epsilon^d a^2}{2} \leq 1,$$

we may apply Lemma S.5, and deduce from (S.27) that

$$\inf_{\widehat{f}} \sup_{f_0\in \mathrm{BV}_\infty(L,M)} \mathbb{E}_{f_0} \|\widehat{f} - f_0\|_{L^2(\Omega)}^2 \geq \frac{a^2\epsilon^d}{4} \inf_{\widehat{\theta}} \max_{\theta\in\Theta} \mathbb{E}_\theta \rho(\widehat{\theta}, \theta) \geq \frac{a^2\epsilon^d |S|}{16\exp(1)}. \tag{S.28}$$

Step 3: Lower bound. The upshot of Steps 1 and 2 is that the solution to the following constrained maximization problem yields a lower bound on the minimax risk: letting $s = |S|$,

$$\begin{aligned} \text{maximize} \quad & \frac{a^2 \epsilon^d s}{16 \exp(1)}, \\ \text{subject to} \quad & 1 \le s \le \epsilon^{-d}, \\ & a s \epsilon^{d-1} \le \frac{L}{2d}, \\ & a \le M, \\ & \frac{n a^2 \epsilon^d}{2} \le 1. \end{aligned}$$

Setting $a = M, \epsilon = (\frac{2}{a^2 n})^{1/d}$, and $s = \frac{L}{2da}\epsilon^{-(d-1)}$ is feasible for this problem if $2dM(\frac{M^2 n}{2})^{-\frac{(d-1)}{d}} \le L \le 2dM(\frac{M^2 n}{2})^{1/d}$, and implies that the optimal value is at least $\frac{2^{1/d}}{32\exp(1)d} LM(M^2 n)^{-1/d}$. This implies the claim (30) upon suitable choices of constants. ☐

D.3 Proof of Lemma 1

In this proof, write $\theta_0 := (f_0(x_1), \dots, f_0(x_n))$ and $\mathbb{E}_{z|x}[\cdot] = \mathbb{E}[\cdot | x_{1:n}]$. We will use D to represent the modified edge incidence operator with either clipped edge weights or unit weights; the following analysis, which uses the scaling factor τ_n, applies to both. Let

$$\mathscr{X} = \mathscr{X}_1 \cap \mathscr{X}_2, \tag{S.29}$$

with $\mathscr{X}_1, \mathscr{X}_2$ as in Section F. By the law of iterated expectation,

$$\mathbb{E}\left[\frac{1}{n}\|\widehat{\theta} - \theta_0\|_2^2\right] = \mathbb{E}_x\left[\mathbb{E}_{z|x}\left[\frac{1}{n}\|\widehat{\theta} - \theta_0\|_2^2\right] \cdot 1\{x_{1:n} \in \mathscr{X}\}\right] + \mathbb{E}_x\left[\mathbb{E}_{z|x}\left[\frac{1}{n}\|\widehat{\theta} - \theta_0\|_2^2\right] \cdot 1\{x_{1:n} \notin \mathscr{X}\}\right]. \tag{S.30}$$

We now upper bound each term on the right hand side separately.

For the first term, we will proceed by comparing the penalty operator D to the averaging operator (S.74) and surrogate operator T corresponding to the graph (S.75). By construction $x_{1:n} \in \mathscr{X}$ implies, for (ξ_k, u_k) the kth singular value/left singular vector of T, that

$$\begin{aligned} \lambda &\ge C_1 \sigma \tau_n (\log n)^{1/2+\alpha} \\ &\ge \max\left\{ 8 \max_\ell |\mathcal{C}_\ell|^{1/2} \Phi_1(D,T,A) \cdot \sigma \sqrt{\log 2n^4 \cdot \sum_{k=2}^n \frac{\|u_k\|_\infty^2}{\xi_k^2}}, \Phi_2(D,T,A) \cdot \sigma \sqrt{2 \log n} \right\}, \end{aligned}$$

where the latter inequality follows from combining (S.71), (S.72) with (S.76), (S.77) in the clipped weights case, or (S.78), (S.79) in the unit weights case, for an appropriately chosen C_1. We may therefore apply Theorem S.1 with D, T, and A, which gives

$$\mathbb{E}_{z|x}\left[\frac{1}{n}\|\widehat{\theta} - \theta_0\|_2^2\right] \cdot 1\{x_{1:n} \in \mathscr{X}\} \le C\left(\frac{\lambda \|D\theta_0\|_1}{n} + \frac{\log^\alpha n}{n}\right), \tag{S.31}$$

On the other hand, to upper bound the second term in (S.30) we use (S.70),

$$\begin{aligned} \mathbb{E}_{z|x}\Big[\frac{1}{n}\|\widehat{\theta} - \theta_0\|_2^2\Big] \cdot 1\{x_{z|x} \notin \mathscr{X}\} &\le \mathbb{E}_{z|x}\Big[\frac{16\|z_{1:n}\|_2^2}{n} + \frac{2\lambda\|D\theta_0\|_1}{n}\Big] 1\{x_{1:n} \notin \mathscr{X}\} \\ &\le \Big(16 + \frac{2\lambda \|D\theta_0\|_1}{n}\Big) 1\{x_{1:n} \notin \mathscr{X}\}. \end{aligned} \tag{S.32}$$

Substituting (S.31) and (S.32) into (S.30), we conclude that

$$\mathbb{E}\left[\frac{1}{n}\|\widehat{\theta} - \theta_0\|_2^2\right] \le C\Big(\frac{\lambda \mathbb{E}\|D\theta_0\|_1}{n} + \frac{\log^\alpha n}{n} + \mathbb{P}(x_{1:n} \notin \mathscr{X})\Big)$$

$$\leq C\Big(\frac{\lambda \mathbb{E}\|D\theta_0\|_1}{n} + \frac{\log^\alpha n}{n}\Big)$$
$$= C\Big(\frac{\sigma\tau_n(\log n)^{1/2+\alpha}\mathbb{E}\|D\theta_0\|_1}{n} + \frac{\log^\alpha n}{n}\Big), \tag{S.33}$$

with the second inequality following from Lemma S.13, and the equality from the choice of $\lambda = C_1\sigma\tau_n(\log n)^{1/2+\alpha}$. □

D.4 Proof of Lemma 2

We prove the claim (33) separately for the unit weights and clipped weights case (recall that they differ by a scaling factor $\bar{\tau}_n$. We will subsequently abbreviate $f := f_0$ and use the notation $\mathrm{DTV}(\,\cdot\,; w^{\varepsilon\leftarrow r})$ to denote the ε-neighborhood graph TV, having set $\varepsilon = r$.

D.4.1 Unit weights

Our goal is to upper bound

$$\mathbb{E}\Big[\mathrm{DTV}\Big(f(x_{1:n});\, \check{w}^{\mathrm{V}}\Big)\Big] = n(n-1)\mathbb{E}\Big[|f(x_1) - f(x_2)| 1\{\mathcal{H}^{d-1}(\bar{V}_1 \cap \bar{V}_2) > 0\}\Big].$$

By conditioning, we can rewrite the expectation above as

$$p_{\max}^2 \int_\Omega \int_\Omega |f(y) - f(x)| \mathbb{P}_{x_{3:n}}\{\mathcal{H}^{d-1}(\bar{V}_x \cap \bar{V}_y) > 0\}\, dy\, dx, \tag{S.34}$$

where $V_x = \{z : \|z - x\|_2 < \|z - x_i\|\ \forall i = 2, 3, \ldots, n\}$, and likewise for V_y. Note that V_x and V_y are random subsets of $\mathbb{R}^d$.

We now give an upper bound on the probability that the random cells $\bar{V}_x$ and $\bar{V}_y$ intersect on a set of positive Hausdorff measure, by relating the problem to uniform concentration of the empirical mass of balls in $\mathbb{R}^d$. The upper bound will be crude, in that it may depend on suboptimal multiplicative constants, but sufficient for our purposes. Define $r(V_x) := \sup\{\|z - x\| : z \in V_x\}$. Observe that if $\|y - x\| > r(V_x) + r(V_y)$, then $\bar{V}_x \cap \bar{V}_y = \emptyset$, since for any $z \in V_x$, by the triangle inequality

$$\{\|z - y\| \geq \|y - x\| - \|z - x\| > r(V_y)\} \Longrightarrow \{z \notin V_y\};$$

therefore

$$\{\mathcal{H}^{d-1}(\bar{V}_x \cap \bar{V}_y) > 0\} \Longrightarrow \{\|y - x\| \leq r(V_x) + r(V_y)\}.$$

Now, choose $z \in V_x$ for which $\|z - x\| = r(V_x)$. Observe that the ball $B(z, r(V_x)/2)$ must have empirical mass 0, i.e., $B(z, r(V_x)/2) \cap \{x_3, \ldots, x_n\} = \emptyset$ (indeed, this same fact must hold for any $r < r(V_x)$). Therefore,

$$\mathbb{P}_{x_{3:n}}\{r(V_x) \geq t\} \leq \mathbb{P}_{x_{3:n}}\Big\{\exists z : B(z, t/2) \cap \{x_3, \ldots, x_n\} = \emptyset\Big\}.$$

It follows from Lemma S.14 that if $t_{n,\delta} = c\big(\frac{1}{n}(d\log n + \log(1/\delta)\big)^{1/d} < t_0$, where t_0 is a constant not depending on n, δ, then

$$\mathbb{P}_{x_{3:n}}\{\exists z : B(z, t_{n,\delta}/2) \cap \{x_3, \ldots, x_n\} = \emptyset\} \leq \delta.$$

Summarizing this reasoning, we have

$$\begin{aligned}
\mathbb{P}_{x_{3:n}}\{\mathcal{H}^{d-1}(\bar{V}_x \cap \bar{V}_y) > 0\} &\leq \mathbb{P}_{x_{3:n}}\Big\{\|y - x\| \leq r(V_x) + r(V_y)\Big\} \\
&\leq \mathbb{P}_{x_{3:n}}\Big\{\|y - x\| \leq 2r(V_x)\Big\} + \mathbb{P}_{x_{3:n}}\Big\{\|y - x\| \leq 2r(V_y)\Big\} \\
&\leq \mathbb{P}_{x_{3:n}}\Big\{\exists z : |B(z, \|x - y\|/4) \cap \{x_3, \ldots, x_n\}| = \emptyset\Big\} \\
&\quad + \mathbb{P}_{x_{3:n}}\Big\{\exists z : |B(z, \|x - y\|/4) \cap \{x_3, \ldots, x_n\}| = \emptyset\Big\}
\end{aligned}$$

$$\leq \begin{cases} 2, & \text{if } \|x-y\|_2 \leq 2t_{n,\delta}, \\ 2\delta, & \text{otherwise.} \end{cases}$$

Setting $\delta_n = n^{-(d+1)/d}$ and plugging this back into (S.34), we conclude that if $t_{n,\delta_n} < t_0$, then

$$\mathbb{E}\Big[\operatorname{DTV}\left(f(x_{1:n});\, \tilde{w}^{\mathrm{V}}\right)\Big] \leq 2n(n-1)\int_\Omega\int_\Omega |f(y)-f(x)|\Big(1\{\|x-y\| \leq 2t_{n,\delta_n}\} + 2\delta_n\Big)\,dy\,dx$$
$$\leq 2\mathbb{E}[\operatorname{DTV}(f; w^{\varepsilon\leftarrow t_{n,\delta}})] + 2n^{1-1/d}\int_\Omega\int_\Omega |f(y)-f(x)|\,dy\,dx. \tag{S.35}$$

Note that since $\lim_{n\to\infty} t_{n,\delta_n} = 0$, the condition $t_{n,\delta_n} < t_0$ will automatically be satisfied for all n sufficiently large.

We now conclude the proof by upper bounding each term in (S.35). The first term refers to the expected ε-neighborhood graph total variation of f when $\varepsilon = t_{n,\delta_n}$, and by (S.39) satisfies

$$\mathbb{E}[\operatorname{DTV}_{n,t_{n,\delta}}(f)] \leq Cn^2(t_{n,\delta_n})^{d+1}\operatorname{TV}(f;\Omega) \leq Cn^{1-1/d}(\log n)^{(d+1)/d}\operatorname{TV}(f;\Omega).$$

The second term above can be upper bounded using a Poincaré inequality for $\mathrm{BV}(\Omega)$ functions, i.e.,

$$\int_\Omega\int_\Omega |f(y)-f(x)|\,dy\,dx \leq 2\int_\Omega |f(x)-\bar{f}(x)|\,dx \leq C\operatorname{TV}(f;\Omega).$$

Plugging these upper bounds back into (S.35) yields the claimed result (33) in the unit weights case. □

D.4.2 Clipped weights

We now show (33) using clipped weights. Our goal is to upper bound

$$\mathbb{E}\Big[\operatorname{DTV}\left(f(x_{1:n});\, \tilde{w}^{\mathrm{V}}\right)\Big] = n(n-1)\mathbb{E}\Big[|f(x_1)-f(x_2)|\max\{c_0 n^{-(d-1)/d}1\{\mathcal{H}^{d-1}(\bar{V}_1\cap\bar{V}_2) > 0\}, \mathcal{H}^{d-1}(\bar{V}_1\cap\bar{V}_2)\}\Big].$$

By conditioning, we may rewrite the expectation above as

$$p_{\max}^2\int_\Omega\int_\Omega |f(y)-f(x)|\mathbb{E}_{x_{3:n}}\left[\max\{c_0 n^{-(d-1)/d}1\{\mathcal{H}^{d-1}(\bar{V}_x\cap\bar{V}_y) > 0\}, \mathcal{H}^{d-1}(\bar{V}_x\cap\bar{V}_y)\}\right]\,dy\,dx, \tag{S.36}$$

where $V_x = \{z : \|z-x\|_2 < \|z-x_i\|\ \forall i = 2,3,\ldots,n\}$, and likewise for V_y. Note that V_x and V_y are random subsets of $\mathbb{R}^d$. We now focus on controlling the inner expectation of (S.36). Upper bound the maximum of two positive functions with their sum to obtain,

$$\mathbb{E}_{3:n}\Big[\max\{c_0 n^{-(d-1)/d}1\{\mathcal{H}^{d-1}(\bar{V}_x\cap\bar{V}_y) > 0\}, \mathcal{H}^{d-1}(\bar{V}_x\cap\bar{V}_y)\}\Big]$$
$$\leq c_0 n^{-(d-1)/d}\mathbb{P}\{\mathcal{H}^{d-1}(\bar{V}_x\cap\bar{V}_y) > 0\} + \mathbb{E}\left[\mathcal{H}^{d-1}(\bar{V}_x\cap\bar{V}_y)\right]. \tag{S.37}$$

We recognize the first term on the RHS of (S.37) as having already been analyzed in the unit weights case; we now focus on the second term. The latter "Voronoi kernel" term may be rewritten,

$$\mathbb{E}_{x_{3:n}}\left[\mathcal{H}^{d-1}(\bar{V}_x\cap\bar{V}_y)\right] = \int_{L\cap\Omega}(1-p_x(z))^{n-2}dz,$$

where $L = \{z : \|x-z\| = \|y-z\|\}$ and $p_x(z) = P(B(z,\|x-z\|))$. Observe by Assumption A1 that $p_x(z) \geq p_{\min}\mu_d\|x-z\|^d$, and therefore

$$\int_{L\cap\Omega}(1-p_x(z))^{n-2} \leq \exp(-cn\|x-z\|^d),$$

for some $c > 0$. Apply Lemma S.19 with $a = 2$ to therefore bound,

$$\mathbb{E}_{x_{3:n}}\left[\mathcal{H}^{d-1}(\bar{V}_x\cap\bar{V}_y)\right] \leq C_1\left(\frac{1\{\|x-y\| \leq C_2(\log n/n)^{1/d}\}}{n^{(d-1)/d}} + \frac{1}{n^2}\right), \tag{S.38}$$

for constants $C_1, C_2 > 0$. Substitute (S.38) into (S.37) and (S.36) to obtain,

$$
\begin{aligned}
&\mathbb{E}\Big[\mathrm{DTV}\,(f(x_{1:n});\ \tilde{w})\Big] \\
&\leq p_{\max}^2 n^2 \int_\Omega \int_\Omega |f(y) - f(x)| \Bigg(c_0 n^{-(d-1)/d} \mathbb{P}_{3:n}\{\mathcal{H}(\bar{V}_x \cap \bar{V}_y) > 0\} \\
&\qquad\qquad + C_1 \frac{1\{\|x-y\| \leq C_2(\log n/n)^{1/d}\}}{n^{(d-1)/d}} + \frac{C_1}{n^2} \Bigg)\, dy\, dx \\
&\leq p_{\max}^2 c_0 n^{-(d-1)/d} \mathbb{E}\left[\mathrm{DTV}(f(x_{1:n}); \breve{w}^{\mathrm{V}})\right] + p_{\max}^2 C_1 n^{-(d-1)/d} \mathbb{E}\left[\mathrm{DTV}(f(x_{1:n}); w^{\varepsilon \leftarrow C_2(\log n/n)^{1/d}})\right] \\
&\qquad\qquad + p_{\max}^2 C_1 \int_\Omega \int_\Omega |f(y) - f(x)|\, dy\, dx \\
&= T_1 + T_2 + T_3.
\end{aligned}
$$

We bound each of the terms above in turn. The first term appeals to (33) in the unit weights case, which we have already proved.

$$
\begin{aligned}
T_1 &= p_{\max}^2 c_0 n^{-(d-1)/d} \mathbb{E}\left[\mathrm{DTV}(f(x_{1:n}); \breve{w}^{\mathrm{V}})\right] \\
&\leq C_3 n^{-(d-1)/d} n^{(d-1)/d} (\log n)^{1+1/d}\, \mathrm{TV}(f) \\
&= C_3 (\log n)^{1+1/d}\, \mathrm{TV}(f).
\end{aligned}
$$

The second term refers to the expected ε-neighborhood graph total variation of f when $\varepsilon = C_2(\log n/n)^{1/d}$, which by (S.39) satisfies,

$$
\begin{aligned}
T_2 &= p_{\max}^2 C_1 n^{-(d-1)/d} \mathbb{E}\left[\mathrm{DTV}\left(f(x_{1:n}); w^{\varepsilon \leftarrow C_2(\log n/n)^{1/d}}\right)\right] \\
&\leq C_4 n^{-(d-1)/d} n^2 (\log n/n)^{(d+1)/d}\, \mathrm{TV}(f) \\
&\leq C_4 (\log n)^{1+1/d}\, \mathrm{TV}(f).
\end{aligned}
$$

The third term can be controlled via the Poincaré inequality,

$$
\begin{aligned}
T_3 &= p_{\max}^2 C_1 \int_\Omega \int_\Omega |f(y) - \bar{f} + \bar{f} - f(x)|\, dy\, dx \\
&\leq C_5 \int_\Omega |f(x) - \bar{f}|\, dx \\
&\leq C_5\, \mathrm{TV}(f),
\end{aligned}
$$

where $\bar{f} := \fint_\Omega f$. ☐

D.4.3 ε-neighborhood and kNN expected discrete TV

Lemma S.6. *Under Assumption A1, there exist constants $c, C_1, C_2 > 0$ such that for all sufficiently large n and $f_0 \in \mathrm{BV}(\Omega)$,*

- *The ε-neighborhood graph total variation, for any $\varepsilon > 0$, satisfies*

$$
\mathbb{E}\Big[\mathrm{DTV}\left(f_0(x_{1:n};\ w^\varepsilon)\right)\Big] \leq C_1 n^2 \varepsilon^{d+1}\, \mathrm{TV}(f_0). \tag{S.39}
$$

- *The k-nearest neighbors graph total variation, for any $k \in \mathbb{N}$, satisfies*

$$
\mathbb{E}\Big[\mathrm{DTV}\left(f_0(x_{1:n};\ w^k)\right)\Big] \leq C_2 \left(n^{1-1/d} k^{(d+1)/d} + n^2 \exp(-ck)\right) \mathrm{TV}(f_0). \tag{S.40}
$$

Proof.

ε-neighborhood expected discrete TV. This follows the proof of Lemma 1 in Green et al. (2021a), with two adaptations to move from Sobolev $H^2(\Omega)$ to the $\mathrm{BV}(\Omega)$: we deal in absolute differences rather than squared differences, and an approximation argument is invoked at the end to account for the existence of non-weakly differentiable functions in $\mathrm{BV}(\Omega)$.

Begin by rewriting,

$$\mathbb{E}\left[\sum_{i,j=1}^{n}|f(x_i)-f(x_j)|\cdot 1\{\|x_i-x_j\|\leq\varepsilon\}\right]=\frac{n(n-1)}{2}\mathbb{E}\left[|f(X')-f(X)|K\left(\frac{\|X'-X\|}{\varepsilon}\right)\right], \tag{S.41}$$

where X and X' are random variables independently drawn from P following Assumption A1 and $K(t)=1\{t\leq 1\}$. Now, take Ω' to be an arbitrary bounded open set such that $B(x,c_0)\subseteq\Omega'$ for all $x\in\Omega$.

For the remainder of this proof, we assume that (i) $f\in BV(\Omega')$ and (ii) $\|f\|_{BV(\Omega')}\leq C'\|f\|_{BV(\Omega)}$ for some constant C' independent of f. These conditions are guaranteed by the Extension Theorem (Evans, 2010; Section 5.4 Theorem 1), which promises an extension operator $E: W^{1,p}(\Omega)\to W^{1,p}(\Omega)$ (take $p=1$ and the BV case is established through an approximation argument). We also assume that $f\in C^\infty(\Omega)$, which is addressed through via an approximation argument at the end. Since $f\in C^\infty(\Omega)$, we may rewrite a difference in terms of an integrated derivative:

$$f(x')-f(x)=\int_0^1\nabla f(x+t(x'-x))^\top(x'-x)dx. \tag{S.42}$$

It follows that

$$\mathbb{E}\left[|f(X')-f(X)|K\left(\frac{\|X'-X\|}{\varepsilon}\right)\leq p_{\max}^2\int_\Omega\int_\Omega|f(x')-f(x)|K\left(\frac{\|x'-x\|}{\varepsilon}\right)dx'dx\right], \tag{S.43}$$

and the final step is to bound the double integral. We have

$$\begin{aligned}
&\int_\Omega\int_\Omega|f(x')-f(x)|K\left(\frac{\|x'-x\|}{\varepsilon}\right)dx'dx\\
&=\int_\Omega\int_\Omega\left|\int_0^1\nabla f(x+t(x'-x))^\top(x'-x)dt\right|K\left(\frac{\|x'-x\|}{\varepsilon}\right)dx'dx && \text{(FTC)}\\
&\leq\int_\Omega\int_\Omega\int_0^1|\nabla f(x+t(x'-x))^\top(x'-x)|K\left(\frac{\|x'-x\|}{\varepsilon}\right)dtdx'dx && \text{(Jensen)}\\
&=\int_\Omega\int_{B(0,1)}\int_0^1|\nabla f(x+t\varepsilon z)^\top(\varepsilon z)|K(\|z\|)\varepsilon^d dtdzdx && (z=(x'-x)/\varepsilon)\\
&=\varepsilon^{d+1}\int_\Omega\int_{B(0,1)}\int_0^1|\nabla f(x+t\varepsilon z)^\top z|K(\|z\|)dtdzdx\\
&\leq\varepsilon^{d+1}\int_{\Omega'}\int_{B(0,1)}\int_0^1|\nabla f(\tilde{x})^\top z|K(\|z\|)dtdzd\tilde{x} && (\tilde{x}=x+t\varepsilon z).
\end{aligned}$$

Next, we apply the Cauchy-Schwarz to $|\nabla f(\tilde{x})^\top z|$ to obtain,

$$\begin{aligned}
\int_{B(0,1)}|\nabla f(\tilde{x})^\top z|K(\|z\|)dz&\leq\int_{B(0,1)}\|\nabla f(\tilde{x})\|\|z\|K(\|z\|)dz\\
&=\|\nabla f(\tilde{x})\|\int_{(B(0,1)}\|z\|K(\|z\|)dz\\
&=C_d\|\nabla f(\tilde{x})\|
\end{aligned}$$

Substituting back in to the previous derivation, we obtain

$$\int_\Omega\int_\Omega|f(x')-f(x)|K\left(\frac{\|x'-x\|}{\varepsilon}\right)dx'dx\leq C_d\varepsilon^{d+1}\int_{\Omega'}\int_0^1\|\nabla f(\tilde{x})\|_1dtd\tilde{x}$$

$$
\begin{aligned}
&= C_d \varepsilon^{d+1} \|Df\|(\Omega') \\
&\leq C_d C' \varepsilon^{d+1} \|Df\|(\Omega)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathbb{E}\left[\frac{1}{2}\sum_{i,j=1}^{n} |f(x_i) - f(x_j)| \cdot 1\{\|x_i - x_j\| \leq \varepsilon\}\right] &\leq \frac{n(n-1)}{2} p_{\max}^2 C_d C' \varepsilon^{d+1} \|Df\|(\Omega) \\
&\leq C_2 n^2 \varepsilon^{d+1} \|Df\|(\Omega)
\end{aligned}
$$

Finally, we provide an approximation argument to justify the assumption that $f \in C^1(\Omega')$. For a function $f \in \mathrm{BV}(\Omega')$, we may construct a sequence of functions $f_k \in C^\infty(\Omega')$ via mollification such that $f_k \to f$ μ-a.e. (specifically, at all Lebesgue points) and $\|Df_k\|(\Omega') \to \|Df\|(\Omega')$ as $k \to \infty$ (Evans and Gariepy, 2015; Theorems 4.1 & 5.3). Via an application of Fatou's lemma, we find that

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{2}\sum_{i,j=1}^{n} |f(x_i) - f(x_j)| \cdot 1\{\|x_i - x_j\| \leq \varepsilon\}\right] \\
&\qquad = \mathbb{E}\left[\frac{1}{2}\sum_{i,j=1}^{n} |\lim_{k\to\infty} f_k(x_i) - f_k(x_j)| \cdot 1\{\|x_i - x_j\| \leq \varepsilon\}\right] \\
&\qquad = \mathbb{E}\left[\liminf_{k\to\infty} \frac{1}{2}\sum_{i,j=1}^{n} |f_k(x_i) - f_k(x_j)| \cdot 1\{\|x_i - x_j\| \leq \varepsilon\}\right] && \text{(Continuity)} \\
&\qquad \leq \liminf_{k\to\infty} \mathbb{E}\left[\frac{1}{2}\sum_{i,j=1}^{n} |f_k(x_i) - f_k(x_j)| \cdot 1\{\|x_i - x_j\| \leq \varepsilon\}\right] && \text{(Fatou's lemma)} \\
&\qquad \leq \liminf_{k\to\infty} C n^2 \varepsilon^{d+1} \|Df_k\|(\Omega) \\
&\qquad = C n^2 \varepsilon^{d+1} \|Df\|(\Omega)
\end{aligned}
$$

k-nearest neighbors expected discrete TV. Let $\varepsilon_k(x) := \|x - x_{(k)}(x)\|_2$ and $\varepsilon_k(x,y) = \max\{\varepsilon_k(x), \varepsilon_k(y)\}$ be data-dependent radii. Notice that

$$
\mathrm{DTV}_{n,k}(f) = \frac{1}{2}\sum_{i,j=1}^{n} |f(x_i) - f(x_j)| \cdot 1\big\{\|x_i - x_j\| \leq \varepsilon_k(x_i, x_j)\big\}.
$$

By linearity of expectation and conditioning, the expected k-nearest neighbor TV can be written as a double integral,

$$
\begin{aligned}
\mathbb{E}[\mathrm{DTV}(f; w^k)] &= n(n-1)\mathbb{E}\Big[|f(x_i) - f(x_j)|\, 1\big\{\|x_i - x_j\| \leq \varepsilon_k(x_i, x_j)\big\}\Big] \\
&= n(n-1)\mathbb{E}\Big[\mathbb{E}\Big[|f(x_i) - f(x_j)|\, 1\big\{\|x_i - x_j\| \leq \varepsilon_k(x_i, x_j)\big\} | x_i, x_j\Big]\Big] \\
&\leq n(n-1)\int_\Omega \int_\Omega |f(y) - f(x)|\, \mathbb{P}\big\{\|x - y\| \leq \varepsilon_k(x,y)\big\}\, dx\, dy \\
&\leq n(n-1)\int_\Omega \int_\Omega |f(y) - f(x)|\, \Big(\mathbb{P}\big\{\|x - y\| \leq \varepsilon_k(x)\big\} + \mathbb{P}\big\{\|x - y\| \leq \varepsilon_k(x)\big\}\Big)\, dx\, dy
\end{aligned}
$$

(The first inequality above is nearly an equality for large n, and the second inequality follows by a union bound.)

We now derive an upper bound $\mathbb{P}\big\{\|x - y\| \leq \varepsilon_k(x)\big\}$. First, observe that the event $\|x - y\| \leq \varepsilon_k(x)$ is equivalent to $|B(x, \|y - x\|) \cap x_{1:n}| < k$. Suppose $\|y - x\| \geq C(k/n)^{1/d}$ for $C = (\frac{2d}{p_{\min}\mu_d})^{1/d}$. Then

$$
p_k(x,y) := P(B(x, \|y - x\|)) \geq \frac{p_{\min}}{2d}\mu_d \|y - x\|^d \geq \frac{2k}{n},
$$

and applying standard concentration bounds (Bernstein's inequality) to the tails of a binomial distribution, it follows that

$$\begin{aligned}\mathbb{P}\Big\{|B(x,\|y-x\|)\cap x_{1:n}|<k\Big\} &= \mathbb{P}\Big\{|B(x,\|y-x\|)\cap x_{1:n}|-np_k(x,y)<k-np_k(x,y)\Big\} \\ &\le \exp\Big(-\frac{c(np_k(x,y)-k)^2}{np_k(x,y)+|np_k(x,y)-k|}\Big) \\ &\le \exp(-ck).\end{aligned}$$

Otherwise if $\|y-x\|<C(k/n)^{1/d}$, we use the trivial upper bound 1 on the probability of an event. To summarize, we have shown

$$\mathbb{P}\big(\|x-y\|\le\varepsilon_k(x)\big)\le\begin{cases}1, & \text{if } \|x-y\|<C(k/n)^{1/d},\\ \exp(-ck), & \text{otherwise.}\end{cases}$$

It follows from (S.40) that

$$\begin{aligned}\mathbb{E}[\mathrm{DTV}(f;w^k)] &\le 2n^2\int_\Omega\int_\Omega |f(y)-f(x)|\,\Big(\big(1\{\|x-y\|<C(k/n)^{1/d}\}\big)+\exp(-ck)\Big)\,dx\,dy \\ &\le C\big(\mathbb{E}[\mathrm{DTV}(f;w^{\varepsilon\leftarrow C(k/n)^{1/d}})]+n^2\exp(-ck)\mathrm{TV}(f,\Omega)\big);\end{aligned} \tag{S.44}$$

the first term on the right hand side of the second inequality is the expected ε-neighborhood graph TV of f, with radius $C(k/n)^{1/d}$, while the second term is obtained from the Poincaré inequality

$$\int_\Omega\int_\Omega |f(y)-f(x)|\,dy\,dx=\int_\Omega\int_\Omega\Big|f(y)-\bar f+\bar f-f(x)\Big|\,dy\,dx\le C\big(\mathrm{TV}(f;\Omega)\big), \tag{S.45}$$

where $\bar f=\fint_\Omega f(x)\,dx$ is the average of f over Ω. The claimed upper bound (S.40) follows from applying inequality (S.39), with $\varepsilon=C(k/n)^{1/d}$, to (S.44). □

D.5 Proof of Lemma 3

Recall that $\|g\|_{L^2(P)}\le p_{\max}\|g\|_{L^2(\mu)}$ for any $g\in L^2(\mu)$ and note that $\|\bar f_0\|_{L^\infty(\mu)}\le M$ with probability one. By Hölder's inequality,

$$\begin{aligned}\mathbb{E}\|\bar f_0-f_0\|^2_{L^2(\mu)} &\le \mathbb{E}\Big[\|\bar f_0-f_0\|_{L^1(\mu)}\cdot\|\bar f_0-f_0\|_{L^\infty(\mu)}\Big] \\ &\le 2M\,\mathbb{E}\|\bar f_0-f_0\|_{L^1(\mu)},\end{aligned} \tag{S.46}$$

and the problem is reduced to upper bounding the expected $L^1(\mu)$ loss of $\bar f_0$. By Fubini's Theorem we may exchange expectation with integral, giving

$$\begin{aligned}\mathbb{E}\|\bar f_0-f_0\|_{L^1(\mu)} &= \int_\Omega \mathbb{E}|\bar f_0(x)-f_0(x)|\,dx \\ &= \int_\Omega\int_\Omega |f_0(y)-f_0(x)|p_x^{(1)}(y)\,dy\,dx,\end{aligned} \tag{S.47}$$

where $p_x^{(1)}(\cdot)$ is the density of $x_{(1)}(x)$. We now give a closed form expression for this density, before proceeding to lower bound (S.47).

Closed-form expression for $p_x^{(1)}$. Suppose P satisfies Assumption A1. For any $y\in\Omega$ and $0<r<\mathrm{dist}(y,\partial\Omega)$, we have

$$\begin{aligned}\mathbb{P}\big\{x_{(1)}(x)\in B(y,r)\big\} &\le n\,\mathbb{P}\big\{x_1\in B(y,r)\big\}\big(\mathbb{P}\{x_2\notin B(x,\|y-x\|\}\big)^{(n-1)} \\ &\le np_{\max}\mu\big(B(y,r)\big)\Big(1-P\big(B(x,\|y-x\|)\big)\Big)^{(n-1)}.\end{aligned}$$

Taking limits as $r\to 0$ gives

$$p_x^{(1)}(y)=\lim_{r\to 0}\frac{\mathbb{P}\big\{x_{(1)}(x)\in B(y,r)\big\}}{\mu(B(y,r))}=np_{\max}\Big(1-P\big(B(x,\|y-x\|)\big)\Big)^{(n-1)}.$$

Upper bound on (S.47). There exists a constant C_d such that for all $x, y \in \Omega$,

$$P\big(B(x, \|y-x\|)\big) \geq \frac{p_{\min}}{C_d}\mu(B(x,\|y-x\|)) = \frac{p_{\min}\mu_d}{C_d}\|y-x\|^d.$$

This implies an upper bound on the density of $x_{(1)}(x)$,

$$\begin{aligned} p_x^{(1)}(y) &\leq n\left(1 - \frac{p_{\min}\mu_d}{C_d}\|y-x\|^d\right)^{(n-1)} \\ &\leq n\exp\left(-\frac{p_{\min}\mu_d}{C_d}\left(\frac{\|y-x\|}{n^{-1/d}}\right)^d\right), \end{aligned}$$

where we have used the inequality $(1-x)^n \leq \exp(-nx)$ for $|x| \leq 1$. Using the inequality, valid for all monotone non-increasing functions $g : [0,\infty) \to [0,\infty)$, that $g(t) \leq 1\{t \leq t_0\}g(0) + g(t_0)$, we further conclude that

$$p_x^{(1)}(y) \leq n1\{\|y-x\| \leq \varepsilon_n^{(1)}\} + \frac{1}{n},$$

for $\varepsilon_n^{(1)} := (\frac{2C_d}{p_{\min}\mu_d}(\log n/n))^{1/d}$. Plugging back into (S.47), we see that the expected $L^1(\mu)$ error is upper bounded by the expected discrete TV of a neighborhood graph with particular kernel and radius, plus a remainder term. Specifically,

$$\begin{aligned} \mathbb{E}\|\bar{f}_0 - f_0\|_{L^1(\mu)} &\leq n\int_\Omega\int_\Omega |f_0(y) - f_0(x)|1\{\|y-x\| \leq \varepsilon_n^{(1)}\}\,dy\,dx + \frac{1}{n}\int_\Omega\int_\Omega |f_0(y)-f_0(x)|\,dy\,dx \\ &\leq n\int_\Omega\int_\Omega |f_0(y) - f_0(x)|1\{\|y-x\| \leq \varepsilon_n^{(1)}\}\,dy\,dx + \frac{C\,\mathrm{TV}(f_0;\Omega)}{n} \qquad \text{(S.48)} \\ &= \frac{1}{n}\mathrm{E}[\mathrm{DTV}(f_0; w^{\varepsilon\leftarrow\varepsilon_n^{(1)}}))] + \frac{C\;\mathrm{TV}(f_0;\Omega)}{n}, \end{aligned}$$

where (S.48) above follows from the Poincaré inequality (S.45). We can therefore apply (S.39), which upper bounds the expected ε-neighborhood graph TV, and conclude that

$$\mathbb{E}\|\bar{f}_0 - f_0\|_{L^1(\mu)} \leq C\left(\frac{(\log n)^{1+1/d}}{n^{1/d}} + \frac{1}{n}\right)\mathrm{TV}(f_0;\Omega) \leq C\left(\frac{L(\log n)^{1+1/d}}{n^{1/d}}\right).$$

Inserting this upper bound into (S.46) completes the proof of Lemma 3. □

D.6 Proof of Theorem 4

The analysis of the ε-neighborhood and kNN TV denoising estimators proceeds identically, so we consider them together. Henceforth let D denote the penalty operator for either estimator and $\hat{f}$ denote their 1NN extrapolants. Follow the proof of Theorem 2 (given in Appendix D.1) to decompose the $L^2(P)$ error for some $C > 0$,

$$\mathbb{E}\left[\|\hat{f} - f_0\|^2_{L^2(P)}\right] \leq C\left(\frac{\lambda\log n\;\mathbb{E}\|D\theta_0\|}{n} + \frac{(\log n)^{1+\alpha}}{n} + \frac{LM(\log n)^{1+1/d}}{n^{1/d}}\right), \qquad \text{(S.49)}$$

where we have applied Lemma 3 which controls the 1NN extrapolation error. Lemma S.6 provides that under the standard assumptions, there exist constants $C_1, C_1' > 0$ such that for all sufficiently large n and $\theta_0 = f_0(x_{1:n})$, $f_0 \in \mathrm{BV}(\Omega)$,

- setting $\varepsilon = c_1(\log^\alpha n/n)^{1/d}$,
$$\mathbb{E}\|D^\varepsilon\theta_0\|_1 \leq C_1 n^{(d-1)/d}(\log n)^{\alpha+\alpha/d}\,\mathrm{TV}(f_0); \qquad \text{(S.50)}$$
- setting $k = c_1'(\log n)^3$,
$$\mathbb{E}\|D^k\theta_0\|_1 \leq C_1' n^{(d-1)/d}(\log n)^{3+3/d}\,\mathrm{TV}(f_0). \qquad \text{(S.51)}$$

Take these values of ε, k and $\lambda = c\sigma(\log n)^{1/2-\alpha}$, $c = c_2, c_2'$, and substitute (S.50), (S.51) into (S.49) to obtain the claim. □

Note that the $L^2(P_n)$ in-sample error may be obtained similarly, beginning with an analysis identical to that of Lemma 1 to obtain the preliminary upper bound,

$$\mathbb{E}\left[\|\hat{f} - f_0\|^2_{L^2(P_n)}\right] \leq C\left(\frac{\lambda\;\mathbb{E}\|D\theta_0\|}{n} + \frac{(\log n)^{1+\alpha}}{n}\right).$$

D.7 Proof of Theorem 5

In this section we prove the upper bound (39). The proof is comprised of several steps and we start by giving a high-level summary.

- We begin in Section D.7.1 by formalizing the estimator $\widehat{f}_{\mathrm{wave}}$ alluded to in Theorem 5, based on hard thresholding of Haar wavelet empirical coefficients.
- Section D.7.2 reviews wavelet coefficient decay of $\mathrm{BV}(\Omega)$ and $L^\infty(\Omega)$ functions. These rates of decay imply that the wavelet coefficients of $f_0 \in \mathrm{BV}_\infty(L, M)$ must belong to the normed balls in a pair of Besov bodies, defined formally in (S.58). Besov bodies are sequence-based spaces that reflect the wavelet coefficient decay of functions in Besov spaces.
- Section D.7.3 gives a deterministic upper bound on the squared-ℓ^2 error of thresholding wavelet coefficients when the population-level coefficients belong to intersections of Besov bodies. This deterministic upper bound is based on analyzing two functionals—a modulus of continuity (S.101) and the tail width (S.102)— in the spirit of (Donoho et al., 1995); the difference is that we are considering *intersections* of Besov bodies.
- The aforementioned modulus of continuity measures the size of the ℓ^2-norm $\|\theta - \theta'\|_2$ relative to ℓ^∞-norm $\|\theta - \theta'\|_\infty$. In Section D.7.4, we give an upper bound on the ℓ^∞ norm of the difference between sample and population-level wavelet coefficients.
- Finally, in Section D.7.5 we combine the results of Sections D.7.3 and D.7.4 to establish upper bounds on the expected squared-ℓ^2 error of hard thresholding sample wavelet coefficients. The same upper bound will apply to the expected squared-$L^2(\Omega)$ error of $\widehat{f}_{\mathrm{wave}}$, by Parseval's theorem.

D.7.1 Step 1: Hard-thresholding of wavelet coefficients

To define the estimator $\widehat{f}_{\mathrm{wave}}$ that achieves the upper bound in (39), we first review the definition of tensor product Haar wavelets.

Definition 1 (Haar wavelet)**.** The *Haar wavelet* $\psi : (0,1) \to \mathbb{R}$ is defined by

$$\psi(x) := 1\{x \in (0, 1/2]\} - 1\{x \in (1/2, 1)\}. \tag{S.52}$$

For each $\mathbf{i} \in \{0,1\}^d \setminus \{(0,\ldots,0)\}$, the *tensor product Haar wavelet* $\Psi^{\mathbf{i}} : (0,1)^d \to \mathbb{R}$ is defined by

$$\Psi^i(x) := \psi^{i_1}(x_1)\ldots\psi^{i_d}(x_d), \tag{S.53}$$

where $\psi^1(x) = \psi(x)$ and $\psi^0(x) = 1$. To ease notation, let $\mathcal{I} = \{0,1\}^d \setminus \{(0,\ldots,0)\}$ and $\mathcal{K}(\ell) = [2^\ell - 1]^d$. For each $\ell \in \mathbb{N} \cup \{0\}$, $k \in \mathcal{K}(\ell)$ and $\mathbf{i} \in \mathcal{I}$, put $\Psi^{\mathbf{i}}_{\ell k}(x) := 2^{\ell d/2}\Psi^{\mathbf{i}}(2^\ell x - k)$. Finally, let $\Phi(x) = 1\{x \in (0,1)^d\}$. The Haar wavelet basis is the collection $\{\Psi^{\mathbf{i}}_{\ell k} : \ell \in \mathbb{N}, k \in \mathcal{K}(\ell), \mathbf{i} \in \mathcal{I}\} \cup \{\Phi\}$, and it forms an orthonormal basis of $L^2((0,1)^d)$.

We now describe the estimator $\widehat{f}_{\mathrm{wave}}$, which applies hard thresholding to sample wavelet coefficients. For each $\ell \in \mathbb{N} \cup \{0\}$, $k \in \mathcal{K}(\ell)$ and $\mathbf{i} \in \mathcal{I}$, write

$$\theta_{\ell k \mathbf{i}}(f) := \int_\Omega \Psi^{\mathbf{i}}_{\ell k}(x) f(x)\, dx, \quad \widetilde{\theta}_{\ell k \mathbf{i}}(f) := \frac{1}{n}\sum_{j=1}^n f(x_j)\Psi^{\mathbf{i}}_{\ell k}(x_j),$$

for the population-level and empirical wavelet coefficients of a given $f \in L^2(\Omega)$. The sample wavelet coefficient is $\widetilde{\theta}_{\ell k \mathbf{i}}(y_{1:n})$. The hard thresholding estimator we use is defined with respect to a threshold $\lambda > 0$ and a truncation level $\ell^* \in \mathbb{N} \cup \{0\}$ as

$$\widehat{\theta}^{(\lambda,\ell^*)}_{\ell k \mathbf{i}} := \begin{cases} \widetilde{\theta}_{\ell k \mathbf{i}}(y_{1:n}) \cdot 1\{\widetilde{\theta}_{\ell k \mathbf{i}}(y_{1:n}) \geq \lambda\}, & \ell = 0, \ldots, \ell^* \\ 0, & \ell \geq \log_2(n)/d + 1, \end{cases} \tag{S.54}$$

and we map the sequence estimate $\widehat{\theta}^{(\lambda,\ell^*)} = \big(\widehat{\theta}^{(\lambda,\ell^*)}_{\ell k \mathbf{i}} : \ell \in \mathbb{N}, k \in \mathcal{K}(\ell), \mathbf{i} \in \mathcal{I}\big)$ to the function

$$\widehat{f}^{(\lambda,\ell^*)}(x) = \overline{y} + \sum_{\ell \in \mathbb{N}} \sum_{k \in \mathcal{K}(\ell), \mathbf{i} \in \mathcal{I}} \widehat{\theta}^{(\lambda,\ell^*)}_{\ell k \mathbf{i}} \Psi^{\mathbf{i}}_{\ell k}(x), \tag{S.55}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$ is the sample average of the responses. (S.55) defines a family of estimators depending on the threshold λ, and the estimator $\widehat{f}_{\rm wave}$ is the hard thresholding estimate $\widehat{f}^{(\lambda,\ell^*)}$ with the specific choices $\lambda = 8n^{-1/2}\log^{3/2}(2n/\delta)$ and $\ell^* = \log_2(n)/d$.

D.7.2 Step 2: Wavelet decay

In this section we recall the wavelet coefficient decay of functions in $\mathrm{BV}(\Omega)$ and in $L^\infty(\Omega)$. For each $\ell \in \mathbb{N}\cup\{0\}$, define

$$\theta_{\ell\cdot}(f) = \big(\theta_{\ell k}^{\mathbf{i}}(f) : k \in \mathcal{K}(\ell), \mathbf{i} \in \mathcal{I}\big).$$

and write $\theta(f)$ for the vector with entries $\theta(f)_\ell := \theta_{\ell\cdot}(f)$.

Lemma S.7. *Let $f \in L^\infty(\Omega)$. Then for all $\ell \in \mathbb{N}\cup\{0\}$,*

$$\|\theta_{\ell\cdot}(f)\|_\infty \le 2^{-\ell d/2}\|f\|_{L^\infty(\Omega)}. \tag{S.56}$$

Lemma S.8. *There exists a constant C_1 such that for all $f \in \mathrm{BV}(\Omega)$ and $\ell \in \mathbb{N}\cup\{0\}$,*

$$\|\theta_{\ell\cdot}(f)\|_1 \le C_1 2^{-\ell(1-d/2)}\mathrm{TV}(f;\Omega). \tag{S.57}$$

The decay rates established by Lemmas S.7 and S.8 imply that if $f_0 \in \mathrm{BV}_\infty(L,M)$, then $\theta(f_0)$ belongs to $\Theta_\infty^{0,\infty}(M)$ and $\Theta_\infty^{1,1}(L)$, where $\Theta_\infty^{s,p}(C)$ consists of sequences θ for which

$$\|\theta\|_{\Theta_\infty^{s,p}} := \sup_{\ell\in\mathbb{N}\cup\{0\}} 2^{\ell(s+d/2-d/p)}\|\theta_{\ell\cdot}\|_p < C. \tag{S.58}$$

The sets $\Theta_\infty^{s,p}(C)$ can be interpreted as normed balls in Besov bodies, since a function f belongs to the Besov space $B_\infty^{s,p}$ if and only if its coefficients in a suitable wavelet basis satisfy $\|\theta(f)\|_{\Theta_\infty^{s,p}} < \infty$.

The conclusions of Lemmas S.7 and S.8 are generally well-understood (see for instance Giné and Nickl (2021) for the upper bound on wavelet decay of $L^\infty(\Omega)$ functions when $d=1$, and Cohen et al. (2003) for the wavelet decay of $\mathrm{BV}(\Omega)$ functions). For purposes of completeness only, we include proofs of these results in Appendix G.3.1.

D.7.3 Step 3: Deterministic upper bound on ℓ^2-error

In this section, we analyze the ℓ^2-error of the hard-thresholding estimator $\widehat{\theta}^{(\lambda,\ell^*)}$. Specifically, we upper bound the magnitude of $\|\widehat{\theta}^{(\lambda,\ell^*)} - \theta(f_0)\|_2$ as a function of the ℓ^∞ distance between the (truncated) sample and population-level wavelet coefficients, i.e the quantity

$$\epsilon_n := \|(\widetilde{\theta}(y_{1:n}) - \theta(f_0))_{\le \ell^*}\|_\infty,$$

where

$$(\theta_{\le\ell^*})_{\ell k}^{\mathbf{i}} := \begin{cases} \theta_{\ell k}^{\mathbf{i}}, & \text{if } \ell \le \ell^*, \\ 0, & \text{otherwise.}\end{cases}$$

Note that this upper bound is purely deterministic.

Proposition S.2. *Suppose $\theta(f_0) \in \Theta_\infty^{0,\infty}(M)\cap\Theta_\infty^{1,1}(L)$. Then there exists a constant C_3 that does not depend on n, M or L for which the following statement holds: if $\lambda \ge 2\epsilon_n$, then the estimator $\widehat{\theta}^{(\lambda,\ell^*)}$ of (S.55) satisfies the upper bound*

$$\|\widehat{\theta}^{(\lambda,\ell^*)} - \theta(f_0)\|_2^2 \le 4C_1 LM 2^{-\ell^*} + C_3\cdot\begin{cases} L\lambda\max\{1, 1/M, \log_2(M/\lambda)\}, & \text{if } d=2 \\ L^{2/d}\lambda^{4/(2+d)} + LM\Big(\dfrac{\lambda}{M}\Big)^{2/d}, & \text{if } d\ge 3.\end{cases} \tag{S.59}$$

The proof is deferred to Appendix G.3.2.

D.7.4 Step 4: Uniform convergence of wavelet coefficients

Lemma S.9 gives an upper bound on the maximum difference between sample and population-level wavelet coefficients that holds uniformly over all $\ell = 0, \ldots, \log_2(n)/d$, $k \in \mathcal{K}(\ell)$ and $\mathbf{i} \in \mathcal{I}$. Its proof is deferred to Appendix G.3.3.

Lemma S.9. *Suppose we observe data $(x_1, y_1), \ldots, (x_n, y_n)$ according to (1), where $f_0 \in L^\infty(\Omega; M)$. There exists a constant C_4 not depending on n such that the following statement holds for all $\delta > 0$: with probability at least $1 - C_4\delta$,*

$$\|(\widetilde{\theta}(y_{1:n}) - \theta(f_0))_{\leq \log_2(n)/d}\|_\infty \leq \underbrace{\frac{4\log^{3/2}(2n/\delta)}{\sqrt{n}} + \frac{\sqrt{12}M\sqrt{\log(2n/\delta)}}{\sqrt{n}}}_{:=\delta_n}. \tag{S.60}$$

D.7.5 Step 5: Upper bound on risk

We are now ready to prove the stated upper bound (39). In this section we take $\lambda = 2\delta_n$ and $\ell^* = \log_2(n)/d$. Combining Proposition S.2 and Lemma S.9, we have that with probability at least $1 - C_4\delta$,

$$\|\widehat{\theta}^{(\lambda,\ell^*)} - \theta_0(f)\|_2^2 \leq \frac{4C_1LM}{n^{1/d}} + C_3 \cdot \begin{cases} 2L\delta_n \max\{1, 1/M, \log_2(M/2\delta_n)\}, & \text{if } d = 2, \\ L^{2/d}(2\delta_n)^{4/(2+d)} + LM\Big(\dfrac{2\delta_n}{M}\Big)^{2/d}, & \text{if } d \geq 3. \end{cases} \tag{S.61}$$

The following lemma allows us to convert this upper bound, which holds with probability $1 - C_4\delta$, to an upper bound which holds in expectation. Its proof is deferred to Appendix G.3.4.

Lemma S.10. *Let $X > 0$ be a positive random variable. Suppose there exist positive numbers $A_1, \ldots, A_K$, $a_1, \ldots, a_K$, $b_1, \ldots, b_K > 1$ and B such that for all $\delta \in (0, 1)$,*

$$\mathbb{P}\Big(X > \sum_{k=1}^K A_k \log^{a_k}(b_k/\delta)\Big) \leq B\delta.$$

Then there exists a constant C_5 depending only on $a_1, \ldots, a_k$ and B such that

$$\mathbb{E}[X] \leq C_5 \sum_{k=1}^K A_k (\log b_k)^{a_k}.$$

Now we use Lemma S.10 to complete the proof of Theorem 5. Note that for any $a > 0$,

$$\delta_n^a \leq \frac{2^a}{\sqrt{n}}\Big((\log(2n/\delta))^{3a/2} + M\sqrt{\log(2n/\delta)}\Big).$$

Thus we may can apply Lemma S.10 to (S.61), which, setting $\delta_n^* = \frac{4}{\sqrt{n}}\Big((\log 2n)^{3/2} + M\log(n)^{1/2}\Big)$, gives

$$\mathbb{E}\|\widehat{\theta}^{(\lambda,\ell^*)} - \theta_0(f)\|_2^2 \leq \frac{4C_1LM}{n^{1/d}} + C_6 \cdot \begin{cases} 2L\delta_n^* \max\{1, 1/M, \log_2(M\sqrt{n})\}, & \text{if } d = 2, \\ L^{2/d}(2\delta_n^*)^{4/(2+d)} + LM\Big(\dfrac{2\delta_n^*}{M}\Big)^{2/d}, & \text{if } d \geq 3. \end{cases} \tag{S.62}$$

where $C_6 = 2C_3C_5$.

Finally, we translate this to an upper bound on the expected risk of $\widehat{f}_{\text{wave}} = \widehat{f}^{(\lambda,\ell^*)}$. Since $\{\Psi_{\ell k}^{\mathbf{i}}, \ell \in \mathbb{N} \cup \{0\}, k \in \mathcal{K}(\ell), \mathbf{i} \in \mathcal{I}\} \cup \{\Phi\}$ forms an orthonormal basis of $L^2(\Omega)$, by Parseval's theorem we have

$$\|\widehat{f}^{(\lambda,\ell^*)} - f_0\|_{L^2(\Omega)}^2 = (\overline{y} - \mathbb{E}[f_0])^2 + \|\widehat{\theta}^{(\lambda,\ell^*)} - \theta(f_0)\|_2^2.$$

Taking expectation on both sides, the claimed upper bound (39) follows by (S.62), upon proper choice of constant $C = \max\{4C_1, 16C_6\}$. □

E Analysis of graph TV denoising

In this section, we review tools for analyzing graph total variation denoising. Suppose an unknown $\theta_0 \in \mathbb{R}^n$ and observations $y_1, \dots, y_n$,

$$y_i = \theta_{0i} + z_i, \qquad i = 1, \dots, n, \tag{S.63}$$

where $z_i \sim \mathcal{N}(0, \sigma^2)$. The graph total variation denoising estimator $\hat{\theta}$ associated with a graph $G = (V, E)$, $|V| = n$, is given by

$$\hat{\theta} = \operatorname*{argmin}_{\theta \in \mathbb{R}^n} \frac{1}{2} \|y_{1:n} - \theta\|_2^2 + \lambda \|D\theta\|_1, \tag{S.64}$$

where $D \in \mathbb{R}^{m \times n}$ is the edge incidence matrix of G.

The initial analysis of graph total variation denoising was performed by Hutter and Rigollet (2016) for the two-dimensional grid. Sadhanala et al. (2016) subsequently generalized the analysis to d-dimensional lattices, and Wang et al. (2016) provided tools for the analysis of general graphs. These techniques rely on direct analysis of properties of graph G and the penalty D in induces, which is tractable when G has a known and regular properties (e.g., it is a lattice graph).

Unfortunately, direct analysis on D may not always be feasible. It may be possible, however, to compare the operator D to a *surrogate operator* whose properties we analyze instead. For our purposes, we compare D to a linear operator which first takes averages on a partition, and then computes differences across cells of the partition. Comparison to this type of surrogate operator was used by Padilla et al. (2020) to bound the risk of graph total variation denoising in probability; the following theorem provides an analogous risk bound in expectation. We note that elements of the "surrogate operator analysis" are also found in Padilla et al. (2018).

Theorem S.1. *Suppose we observe data according to model* (S.63) *and compute the graph TV denoising estimator* $\hat{\theta}$ *of* (S.64)*. Let* $A \in \mathbb{R}^{n \times n}$ *denote an averaging operator over* $\bar{N}$ *groups of the form,*

$$A = \begin{bmatrix} n_1^{-1} \mathbf{1}_{n_1} \mathbf{1}_{n_1}^\top & 0 & \dots & 0 \\ 0 & n_2^{-1} \mathbf{1}_{n_2} \mathbf{1}_{n_2}^\top & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n_{\bar{N}}^{-1} \mathbf{1}_{n_{\bar{N}}} \mathbf{1}_{n_{\bar{N}}}^\top \end{bmatrix},$$

with $M := \max_j n_j$*, and let* $\bar{A} \in \mathbb{R}^{\bar{N} \times n}$ *be the same matrix with redundant rows removed. Further let* $T \in \mathbb{R}^{\bar{m} \times \bar{N}}$ *be a surrogate penalty operator, with singular value decomposition* $T = U \Sigma V^\top$*, such that*

$$\|T\bar{A}\theta\|_1 \le \Phi_1(D, T, A) \|D\theta\|_1, \tag{S.65}$$

$$\|(I - A)\theta\|_1 \le \Phi_2(D, T, A) \|D\theta\|_1, \tag{S.66}$$

for quantities $\Phi_1(D, T, A), \Phi_2(D, T, A)$ *that may depend on* n*, for all* $\theta \in \mathbb{R}^n$*. If the penalty parameter*

$$\lambda > \max \left\{ 8M^{1/2} \Phi_1(D, T, A) \cdot \sigma \sqrt{\log(2n^4) \sum_{k=2}^{\bar{N}} \frac{\|u_k\|_\infty^2}{\xi_k^2}}, \Phi_2(D, T, A) \cdot \sigma \sqrt{2 \log(n)} \right\} \tag{S.67}$$

where u_k *is the* k*th column of* U *and* ξ_k *the* k*th diagonal entry of* Σ*, then there exists a constant* $C > 0$ *such that*

$$\mathbb{E}\left[\frac{1}{n} \|\hat{\theta} - \theta_0\|_2^2\right] \le C \left(\frac{\lambda \|D\theta_0\|_1}{n} + \frac{M \text{nullity}(T)}{n} \right). \tag{S.68}$$

Proof. We follow the approach of Padilla et al. (2020), with adaptations to provide a bound in expectation rather than in probability. From the basic inequality,

$$\|\hat{\theta} - \theta_0\|_2^2 \le 2\langle z_{1:n}, \hat{\theta} - \theta_0 \rangle + \lambda(\|D\theta_0\|_1 - \|D\hat{\theta}\|_1),$$

where $z_{1:n} \in \mathbb{R}^n$ is the vector of error terms z_i, $i = 1, \dots, n$. We provide two deterministic bounds under the "good case" that the error term falls into the set,

$$\mathcal{Z}_\lambda = \left\{ z_{1:n} : \max \left\{ M^{1/2} \Phi_1(D, T, A) \sup_{\bar{A}\theta \in \text{row}(T) : \|T\bar{A}\theta\|_1 \le 1} |\langle \overline{z_{1:n}}, \bar{A}\theta \rangle|, \Phi_2(D, T, A) \|z_{1:n}\|_\infty \right\} \le \frac{\lambda}{8} \right\}, \tag{S.69}$$

where $\overline{z_{1:n}} \in \mathbb{R}^{\bar{N}}$ has entries $\overline{z_{1:n}}_j = n_j^{1/2} (\bar{A} z_{1:n})_j$, and under the "bad case" that $z_{1:n} \notin \mathcal{Z}_\lambda$.

Upper bound in the "good case". Decompose the first term on the RHS,

$$\begin{aligned}\langle z_{1:n}, \hat\theta - \theta_0\rangle &= \langle z_{1:n}, \hat\theta - A\hat\theta\rangle + \langle z_{1:n}, A\theta_0 - \theta_0\rangle + \langle z_{1:n}, A(\theta_0 - \hat\theta)\rangle \\ &\le \langle z_{1:n}, A(\theta_0 - \hat\theta)\rangle + \|z_{1:n}\|_\infty(\|(I-A)\hat\theta\|_1 + \|(I-A)\theta_0\|_1) \\ &\le \langle z_{1:n}, A(\theta_0 - \hat\theta)\rangle + \|z_{1:n}\|_\infty \Phi(D,T,A)(\|D\theta_0\|_1 + \|D\hat\theta\|_1),\end{aligned}$$

where the final inequality follows from (S.66). Observe that we may rewrite, for any $\theta \in \mathbb{R}^n$,

$$\begin{aligned}\langle z_{1:n}, A\theta\rangle &= \sum_{j=1}^{\bar N}\sum_{i=1}^{n_j} z_{1:n(\sum_{k=1}^{j-1} n_k)+i}(\bar A\theta)_j \\ &\overset{d}{=} \sum_{j=1}^{\bar N} n_j^{1/2} z_{1:nj}(\bar A\theta)_j \\ \Rightarrow \langle z_{1:n}, A\theta\rangle &\le M^{1/2}|\langle \overline{z_{1:n}}, \bar A\theta\rangle| \\ &\le M^{1/2}\left|\langle \mathrm{proj}_V(\overline{z_{1:n}}), \bar A\theta\rangle + \langle \mathrm{proj}_{V^\perp}(\overline{z_{1:n}}), \bar A\theta\rangle\right| \\ &\le M^{1/2}\left|\|\mathrm{proj}_V(\overline{z_{1:n}})\|_2\|\bar A\theta\|_2 + \langle \mathrm{proj}_{V^\perp}(\overline{z_{1:n}}), \bar A\theta\rangle\right| \\ &\le M^{1/2}\left(\|\mathrm{proj}_V(\overline{z_{1:n}})\|_2\|\theta\|_2 + |\langle \mathrm{proj}_{V^\perp}(\overline{z_{1:n}}), \bar A\theta\rangle|\right|,\end{aligned}$$

where $\overline{z_{1:n}} \in \mathbb{R}^{\bar N}$ has independent $\mathcal{N}(0,\sigma^2)$ entries and $V = \mathrm{null}(T)$. Substitute back in to obtain,

$$\begin{aligned}\|\hat\theta - \theta_0\|_2^2 \le 2M^{1/2}&(\|\mathrm{proj}_V(\overline{z_{1:n}})\|_2\|\hat\theta-\theta_0\|_2 + |\langle \mathrm{proj}_{V^\perp}(\overline{z_{1:n}}), \bar A(\hat\theta - \theta_0)\rangle|) \\ &+ 2\|z_{1:n}\|_\infty\Phi(D,T,A)(\|D\theta_0\|_1 + \|D\hat\theta\|_1) + \lambda(\|D\theta_0\|_1 - \|D\hat\theta\|_1),\end{aligned}$$

and consequently,

$$\begin{aligned}&\|\hat\theta - \theta_0\|_2(\|\hat\theta-\theta_0\|_2 - 2M^{1/2}\|\mathrm{proj}_V(\overline{z_{1:n}})\|_2) \\ &\quad\le 2M^{1/2}|\langle \mathrm{proj}_{V^\perp}(\overline{z_{1:n}}), \bar A(\hat\theta-\theta_0)\rangle| + 2\|z_{1:n}\|_\infty\Phi(D,T,A)(\|D\theta_0\|_1 + \|D\hat\theta\|_1) + \lambda(\|D\theta_0\|_1 - \|D\hat\theta\|_1)\end{aligned}$$

Case 1. $\|\hat\theta - \theta_0\|_2 \le 4M^{1/2}\|\mathrm{proj}_V(\overline{z_{1:n}})\|_2$.
Case 2. $\|\hat\theta - \theta_0\|_2 > 4M^{1/2}\|\mathrm{proj}_V(\overline{z_{1:n}})\|_2$. Then,

$$\|\hat\theta-\theta_0\|_2^2 \le 4M^{1/2}|\langle \mathrm{proj}_{V^\perp}(\overline{z_{1:n}}), \bar A(\hat\theta-\theta_0)\rangle| + 4\|z_{1:n}\|_\infty\Phi(D,T,A)(\|D\theta_0\|_1 + \|D\hat\theta\|_1) + \lambda(\|D\theta_0\|_1 - \|D\hat\theta\|_1).$$

We then bound,

$$\begin{aligned}|\langle \mathrm{proj}_{V^\perp}(\overline{z_{1:n}}), \bar A(\hat\theta-\theta_0)\rangle| &= \left|\left\langle \mathrm{proj}_{V^\perp}(\overline{z_{1:n}}), \frac{\bar A(\hat\theta-\theta_0)}{\|T\bar A(\hat\theta-\theta_0)\|_1}\right\rangle \|T\bar A(\hat\theta-\theta_0)\|_1\right| \\ &\le \sup_{\bar A\theta \in V^\perp : \|T\bar A\theta\|_1 \le 1} |\langle \overline{z_{1:n}}, \bar A\theta\rangle| \|T\bar A(\hat\theta-\theta_0)\|_1 \\ &\le \sup_{\bar A\theta \in V^\perp : \|T\bar A\theta\|_1 \le 1} |\langle \overline{z_{1:n}}, \bar A\theta\rangle| \Phi(D,T,A)(\|D\hat\theta\|_1 + \|D\theta_0\|_1),\end{aligned}$$

where the last inequality follows by (S.66). Conditioning on $z_{1:n} \in \mathcal{Z}_\lambda$, we find that under Case 2,

$$\begin{aligned}\|\hat\theta-\theta_0\|_2^2 &\le \frac{\lambda}{2}(\|D\hat\theta\|_1 + \|D\theta_0\|_1) + \frac{\lambda}{2}(\|D\hat\theta\|_1 + \|D\theta_0\|_1) + \lambda(\|D\theta_0\|_1 - \|D\hat\theta\|_1) \\ &\le 2\lambda\|D\theta_0\|_1.\end{aligned}$$

Therefore, conditioning on $z_{1:n} \in \mathcal{Z}_\lambda$ and combining Case 1 and Case 2, we obtain that

$$\|\hat\theta-\theta_0\|_2^2 \le 16M\|\mathrm{proj}_V(\overline{z_{1:n}})\|_2^2 + 2\lambda\|D\theta_0\|_1.$$

Upper bound in the "bad case". On the "bad event" that $z_{1:n} \notin \mathcal{Z}_\lambda$, we apply Hölder directly to the basic inequality to bound,

$$\|\hat{\theta} - \theta_0\|_2^2 \leq 2\|z_{1:n}\|_2 \|\hat{\theta} - \theta_0\|_2 + \lambda \|D\theta_0\|_1,$$

and rearrange to obtain,

$$\|\hat{\theta} - \theta_0\|_2^2 \leq 16\|z_{1:n}\|_2^2 + 2\lambda\|D\theta_0\|_1. \tag{S.70}$$

Combining the "good case" and "bad case" upper bounds.

$$\begin{aligned}
\frac{1}{n}\mathbb{E}\|\hat{\theta} - \theta_0\|_2^2 &= \frac{1}{n}\mathbb{E}\left[\|\hat{\theta} - \theta_0\|_2^2 \mathbf{1}\{z_{1:n} \in \mathcal{Z}_\lambda\} + \|\hat{\theta} - \theta_0\|_2^2 \mathbf{1}\{z_{1:n} \notin \mathcal{Z}_\lambda\}\right] \\
&\leq \frac{1}{n}\left[\mathbb{E}\left[16M\|\mathrm{proj}_V(\overline{z_{1:n}})\|_2^2 + 2\lambda\|D\theta_0\|_1\right] + \mathbb{E}\left[(16\|z_{1:n}\|_2^2 + 2\lambda\|D\theta_0\|_1)\mathbf{1}\{z_{1:n} \notin \mathcal{Z}_\lambda\}\right]\right] \\
&\leq \frac{1}{n}\left[16M\dim(V) + 4\lambda\|D\theta_0\|_1 + \sqrt{\mathbb{E}[\|z_{1:n}\|_2^4]} \cdot \mathbb{P}[z_{1:n} \notin \mathcal{Z}_\lambda]\right] \\
&\leq \frac{1}{n}\left[16M\dim(V) + 4\lambda\|D\theta_0\|_1 + \sqrt{3}n \cdot \mathbb{P}[z_{1:n} \notin \mathcal{Z}_\lambda]\right]
\end{aligned}$$

It remains to bound the probability of the bad case,

$$\begin{aligned}
\mathbb{P}\{z_{1:n} \notin \mathcal{Z}_\lambda\} &\leq \mathbb{P}\left\{M^{1/2} \sup_{\bar{A}\theta \in V^\perp : \|T\bar{A}\theta\|_1 \leq 1} |\langle \overline{z_{1:n}}, \bar{A}\theta\rangle| \geq \lambda/8\Phi(D,T,A)\right\} + \mathbb{P}\left\{\|z_{1:n}\|_\infty \geq \lambda/8\Phi(D,T,A)\right\} \\
&\leq \mathbb{P}\{M^{1/2}\Phi(D,T,A)\|(T^+)^\top \overline{z_{1:n}}\|_\infty \geq \lambda/8\} + \mathbb{P}\{\Phi(D,T,A)\|z_{1:n}\|_\infty \geq \lambda/8\}.
\end{aligned}$$

Standard results on the maxima of Gaussians provide that,

$$\mathbb{P}\left\{M^{1/2}\Phi_1(D,T,A)\|(T^+)^\top \overline{z_{1:n}}\|_\infty \geq M^{1/2}\Phi_1(D,T,A) \cdot \sigma\sqrt{\log(2n^2/\delta) \cdot \sum_{k=2}^{\bar{N}} \frac{\|u_k\|_\infty^2}{\xi_k^2}}\right\} \leq \delta,$$

$$\mathbb{P}\left\{\Phi_2(D,T,A)\|z_{1:n}\|_\infty \geq \Phi_2(D,T,A) \cdot \sigma\sqrt{\log(2n^2/\delta)}\right\} \leq \delta.$$

Recalling the choice of penalty parameter,

$$\lambda > \max\left\{8M^{1/2}\Phi_1(D,T,A) \cdot \sigma\sqrt{\log(2n^4)\sum_{k=2}^{\bar{N}} \frac{\|u_k\|_\infty^2}{\xi_k^2}}, \Phi_2(D,T,A) \cdot \sigma\sqrt{2\log(n)}\right\},$$

we conclude that

$$\mathbb{P}\{z_{1:n} \notin \mathcal{Z}_\lambda\} \leq \frac{2}{n^2},$$

completing the proof. □

We now state a well-known result controlling certain functionals of the lattice difference operator. These quantities have been analyzed by others studying graph total variation denoising on lattices, e.g., Hutter and Rigollet (2016) and Sadhanala et al. (2017).

Lemma S.11. *Let T be the edge incidence operator of the d-dimensional lattice graph N elements per direction. Denote $n = N^d$. The left singular vectors of T satisfy an incoherence condition,*

$$\|u_j\|_\infty \leq \frac{C_d}{\sqrt{n}}, \qquad j = 1, \ldots, n,$$

for some $C_d > 0$, and its singular values satisfy an asymptotic scaling,

$$c_d(j/n)^{1/d} \leq \xi_j \leq C_d(j/n)^{1/d}, \qquad j = 2, \ldots, n,$$

for some $0 < c_d < C_d$. Consequently,

$$\sum_{j=2}^{n} \frac{\|u_j\|_\infty^2}{\xi_j^2} = C_d \begin{cases} \log n & d = 2, \\ 1 & d > 2. \end{cases} \tag{S.71}$$

F Embeddings for random graphs

We begin by providing a result that controls the number of sample points that fall into each cell of a lattice mesh.

Lemma S.12. *Suppose $x_1, \dots, x_n$ are sampled from a distribution P supported on $(0,1)^d$ with density p such that $0 < p_{\min} < p(x) < p_{\max} < 1$ for all $x \in (0,1)^d$. Form a partition of $(0,1)^d$ using an equally spaced mesh with $N = C_1(p_{\min} n/\log^\alpha n)^{1/d}$, $\alpha > 1$, along each dimension. Let $\mathcal{C}_\ell$ denote the ℓth cell of the mesh, and let $|\mathcal{C}_\ell|$ denote its empirical content. Then for all $x_{1:n} \in \mathscr{X}_1$, with $\mathbb{P}\{x_{1:n} \in \mathscr{X}_1\} \geq 1 - 2/n^4$,*

$$\max_\ell |\mathcal{C}_\ell| \leq C_3 \log^\alpha n, \tag{S.72}$$

$$\min_\ell |\mathcal{C}_\ell| \geq c_4 \log^\alpha n, \tag{S.73}$$

for n sufficiently large, where $C_3, c_4 > 0$ depend only on $p_{\min}, p_{\max}, d$.

Proof. From standard concentration bounds (e.g., Von Luxburg et al., 2014; Proposition 27) on a random variable $m \sim \text{Bin}(n, p)$, for all $\delta \in (0, 1]$,

$$\mathbb{P}\{m \geq (1+\delta)np\} \leq \exp\{-\frac{1}{3}\delta^2 np\},$$
$$\mathbb{P}\{m \leq (1-\delta)np\} \leq \exp\{-\frac{1}{3}\delta^2 np\}.$$

Apply these bounds with $p = \mathbb{P}\{x \in \mathcal{C}_\ell\}$ to obtain that,

$$\mathbb{P}\left\{\max_\ell |\mathcal{C}_\ell| \geq (1+\delta) C_1^d \frac{p_{\max}}{p_{\min}} \log^\alpha n\right\} \leq N^d \exp\left\{-\frac{1}{3}\delta^2 C_1^d \log^\alpha n\right\},$$
$$\mathbb{P}\left\{\min_\ell |\mathcal{C}_\ell| \leq (1-\delta) C_1^d \log^\alpha n\right\} \leq N^d \exp\left\{-\frac{1}{3}\delta^2 C_1^d \log^\alpha n\right\},$$

for all $\delta \in (0, 1)$. Setting the RHS to $1/n^4$,

$$\begin{aligned}
\frac{C_1^d p_{\min} n}{\log^\alpha n} \exp\{-\frac{1}{3}\delta^2 C_1^{-d} \log^\alpha n\} &\leq \frac{1}{n^4} \\
\log(C_1^d p_{\min}) - \log(\log^\alpha n) - \frac{1}{3}\delta^2 C_1^{-d} \log^\alpha n &\leq -5 \log n \\
\Rightarrow \frac{1}{3}\delta^2 C_1^{-d} \log^\alpha n &\geq 5 \log n + \log(C_1^d p_{\min}) - \log(\log^\alpha n) \\
\delta^2 &\geq 3C_1^d \left(5 \log^{1-\alpha} n + \frac{\log(C_1^d p_{\min})}{\log^\alpha n} - \frac{\log(\log^\alpha n)}{\log^\alpha n}\right) \\
\delta &\geq C_2 \log^{(1-\alpha)/2} n,
\end{aligned}$$

for some $C_2 > 0$ for all n sufficiently large. Therefore deduce that,

$$\mathbb{P}\left\{\max_\ell |\mathcal{C}_\ell| \geq C_1^d \frac{p_{\max}}{p_{\min}} \log^\alpha n + C_1^d C_2 \frac{p_{\max}}{p_{\min}} \log^{(1+\alpha)/2} n\right\} \leq \frac{1}{n^4},$$
$$\mathbb{P}\left\{\min_\ell |\mathcal{C}_\ell| \leq C_1^d \log^\alpha n - C_1^d C_2 \log^{(1+\alpha)/2} n\right\} \leq \frac{1}{n^4}.$$

Recall that $\alpha > 1$ by assumption, and choose $C_3, c_4 > 0$ with n sufficiently large to obtain the claim. □

The following lemma establishes embeddings from certain random graphs into a coarser lattice graph.

Lemma S.13. *Partition the domain $(0,1)^d$ using an equally spaced mesh with $N = C_1(p_{\min} n/\log^\alpha n)^{1/d}$ elements per direction. Suppose that $x_{1:n} \in \mathscr{X}_1$, with $x_{1:n}$ re-indexed such that*

$$x_1, \dots, x_{|\mathcal{C}_1|} \in \mathcal{C}_1,$$

$$x_{|\mathcal{C}_1|+1}, \dots, x_{|\mathcal{C}_1|+|\mathcal{C}_2|} \in \mathcal{C}_2,$$
$$\vdots$$
$$x_{\sum_{\ell=1}^{N^d-1}|\mathcal{C}_\ell|+1}, \dots, x_{N^d} \in \mathcal{C}_{N^d}.$$

Consider the averaging operators,

$$A = \begin{bmatrix} |\mathcal{C}_1|^{-1}\mathbf{1}_{|\mathcal{C}_1|}\mathbf{1}_{|\mathcal{C}_1|}^\top & 0 & \dots & 0 \\ 0 & |\mathcal{C}_2|^{-1}\mathbf{1}_{|\mathcal{C}_2|}\mathbf{1}_{|\mathcal{C}_2|}^\top & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |\mathcal{C}_{N^d}|^{-1}\mathbf{1}_{|\mathcal{C}_{N^d}|}\mathbf{1}_{|\mathcal{C}_{N^d}|}^\top \end{bmatrix}, \tag{S.74}$$

and the lattice difference operator T based on the graph

$$G_T = (\{1, \dots, N^d\}, E_T), \tag{S.75}$$

where $(i,j) \in E_T$ if the midpoints of $\mathcal{C}_i, \mathcal{C}_j$ are $1/N$ apart. Also, let $\bar{A} \in \mathbb{R}^{N^d \times n}$ be the matrix obtained by dropping the redundant rows of A.

- *Build the Voronoi graph from $x_{1:n}$, and let $\tilde{D}^{\mathrm{V}}$ denote the edge incidence operator with edge set E^{V} and edge weights $\tilde{w}_{ij}^{\mathrm{V}} = \max\{c_0 n^{-(d-1)/d}, w_{ij}^{\mathrm{V}}\}$ for each i, j. Further condition on the set $\mathscr{X}_2$ such that the result of Lemma S.14 holds with probability $1 - 1/n^4$ (equivalently, the set that the result of Lemma S.15 holds with probability $1 - 1/n^4$). Then there exists a constant $C_6 > 0$ such that for all $\theta \in \mathbb{R}^n$,*

$$\|T\bar{A}\theta\|_1 \le C_6 n^{(d-1)/d} \|\tilde{D}^{\mathrm{V}}\theta\|_1. \tag{S.76}$$
$$\|(I - A)\theta\|_1 \le C_6 (\log n)^\alpha n^{(d-1)/d} \|\tilde{D}^{\mathrm{V}}\theta\|_1, \tag{S.77}$$

- *Build the Voronoi graph from $x_{1:n}$, and let $\check{D}^{\mathrm{V}}$ denote the edge incidence operator with edge set E^{V} and edge weights $\check{w}_{ij}^{\mathrm{V}} = 1$ for each i, j such that $w_{i,j}^{\mathrm{V}} > 0$. Further condition on the set $\mathscr{X}_2$ such that the result of Lemma S.14 holds with probability $1 - 1/n^4$. Then there exists a constant $C_7 > 0$ such that for all $\theta \in \mathbb{R}^n$,*

$$\|T\bar{A}\theta\|_1 \le C_7 \|\check{D}^{\mathrm{V}}\theta\|_1. \tag{S.78}$$
$$\|(I - A)\theta\|_1 \le C_7 (\log n)^\alpha \|\check{D}^{\mathrm{V}}\theta\|_1, \tag{S.79}$$

- *Build the ε-neighborhood graph from $x_{1:n}$, with $\varepsilon \ge 2\sqrt{d}/N$. Then with the constant c_4 from Lemma S.12, it holds that for all $\theta \in \mathbb{R}^n$,*

$$\|T\bar{A}\theta\|_1 \le \frac{1}{c_4^2 \log^{2\alpha} n} \|D^\varepsilon \theta\|_1. \tag{S.80}$$
$$\|(I - A)\theta\|_1 \le \frac{2}{c_4 \log^\alpha n} \|D^\varepsilon \theta\|_1, \tag{S.81}$$

- *Build the k-nearest neighbors graph from $x_{1:n}$, with $k \ge C_5 \log^3 n$. Further condition on the set $\mathscr{X}_2$ such that the result of Lemma S.14 holds with probability $1 - 1/n^4$. Then with the constant c_4 from Lemma S.12, it holds for all $\theta \in \mathbb{R}^n$,*

$$\|T\bar{A}\theta\|_1 \le \frac{1}{c_4^2 \log^{2\alpha} n} \|D^k \theta\|_1. \tag{S.82}$$
$$\|(I - A)\theta\|_1 \le \frac{2}{c_4 \log^\alpha n} \|D^k \theta\|_1, \tag{S.83}$$

Proof. **ε-neighborhood graph.** First, we prove (S.80) and (S.81). For the former, observe that

$$\|T\bar{A}\theta\|_1 = \sum_{(k,\ell)\in E_T} \left| |\mathcal{C}_k|^{-1} \sum_{i\in\mathcal{C}_k} \theta_i - |\mathcal{C}_\ell|^{-1} \sum_{j\in\mathcal{C}_\ell} \theta_j \right|$$

$$\le \sum_{(k,\ell)\in E_T} \frac{1}{|\mathcal{C}_k||\mathcal{C}_\ell|} \sum_{i\in|\mathcal{C}_k|, j\in|\mathcal{C}_\ell|} |\theta_i - \theta_j|$$
$$\le \frac{1}{c_4^2 \log^{2\alpha} n} \sum_{(k,\ell)\in E_T} \sum_{i\in\mathcal{C}_k, j\in\mathcal{C}_\ell} |\theta_i - \theta_j|$$
$$\le \frac{1}{c_4^2 \log^{2\alpha} n} \|D^\varepsilon \theta\|_1,$$

as $\varepsilon = 2\sqrt{d}/N$. For the latter, similarly deduce that

$$\begin{aligned}
\|(I-A)\theta\|_1 &= \sum_{i=1}^n \left| \theta_i - |\mathcal{C}(i)|^{-1} \sum_{j\in\mathcal{C}(i)} \theta_j \right| \\
&\le \sum_{i=1}^n |\mathcal{C}(i)|^{-1} \left| \sum_{j\in\mathcal{C}(i)} \theta_j - \theta_i \right| \\
&\le \sum_{i=1}^n |\mathcal{C}(i)|^{-1} \sum_{j\in\mathcal{C}(i)} |\theta_i - \theta_j| \\
&= \sum_{\ell=1}^{N^d} |\mathcal{C}_\ell|^{-1} \sum_{i\in\mathcal{C}_\ell} \sum_{j\in\mathcal{C}_\ell} |\theta_i - \theta_j| \\
&\le \frac{2}{c_4 \log^\alpha n} \sum_{\ell=1}^{N^d} \sum_{i<j\in\mathcal{C}_\ell} |\theta_i - \theta_j| \\
&\le \frac{2}{c_4 \log^\alpha n} \|D^\varepsilon \theta\|_1.
\end{aligned}$$

k-nearest neighbors graph. Recall that we have conditioned on the set $\mathscr{X}_2$ such that the result of Lemma S.14 holds. In particular, (S.86) gives that

$$\min_{i=1,\dots,n} \varepsilon_k(x_i) \ge C \left(\frac{k}{n}\right)^{1/d},$$

where $\varepsilon_k(x_i) := \|x_i - x_{(k)}(x_i)\|_2$. The results (S.82) and (S.83) then follow by observing that on the event $\mathscr{X}_2$, the k-nearest neighbors graph with $k \ge C_5 \log^3 n$ dominates the ε-neighborhood graph with $\varepsilon = 2\sqrt{d}/N$.

Voronoi adjacency graph. We will prove the results (S.78) and (S.79) by providing a graph comparison inequality between the ε-neighborhood graph with $\varepsilon = 2\sqrt{d}/N$ and the Voronoi adjacency graph. The results (S.76), (S.77) follow from the inequality $\|\check{D}^{\mathrm{V}}\theta\|_1 \le c_0^{-1} n^{(d-1)/d} \|\tilde{D}^{\mathrm{V}}\theta\|_1$ for all $\theta \in \mathbb{R}^n$.

Intuition and outline. The central goal of this proof is to show that

$$\|D^\varepsilon \theta\|_1 \le C(n) \|\check{D}^{\mathrm{V}}\theta\|_1,$$

for all $\theta \in \mathbb{R}^n$, where $C(n)$ is at most polylogarithmic in n. This will be accomplished by

(i) verifying that for any $\{x_i, x_j\} \in E^\varepsilon$, there exists a path $\{x_i, x_{k_1}\}, \{x_{k_1}, x_{k_2}\}, \dots, \{x_{k_{ij}}, x_j\} \in E^{\mathrm{V}}$, and

(ii) showing that if one uses the shortest path in the Voronoi adjacency graph G^{V} to connect each $\{x_i, x_j\} \in E^\varepsilon$, then no one edge is used more than $C_9 \log^{2\alpha} n$ times, where C_9 is a positive constant and $\alpha > 1$ may be chosen.

Step (i). Consider x_i, x_j such that $\{x_i, x_j\} \in E^\varepsilon$. We will show the existence of a path between x_i and x_j in G^{V} and also characterize some properties of the path for step (ii).

By definition, $\|x_i - x_j\| \leq \varepsilon$. Denote

$$x_{ij} := \frac{x_i + x_j}{2},$$
$$r_{ij} := \|x_i - x_{ij}\|.$$

Consider the subgraph $G^{ij} = (V^{ij}, E^{ij})$, where

$$V^{ij} := \{V_k : V_k \cap B(x_{ij}, r_{ij}) \neq \emptyset\},$$
$$E^{ij} := \{\{V_k, V_\ell\} : V_k, V_\ell \in V^{ij}, \mathcal{H}^{d-1}(\partial V_k \cap \partial V_\ell) > 0\},$$

where $B(x_{ij}, r_{ij})$ is the closed ball centered at x_{ij} with radius r_{ij}. By construction, $x_i, x_j \in V^{ij}$, and by Lemma S.16, G^{ij} is connected. Therefore a path between x_i and x_j exists in the graph G^{ij} (one can use, e.g., breadth-first search or Dijkstra's algorithm to find such a path).

Step (ii). For any $\{x_i, x_j\} \in E^\varepsilon \setminus E^{\mathrm{V}}$, we create a path in G^{V} as prescribed in step (i). With these paths created, we upper bound the number of times any edge in E^{V} is used. We do so by uniformly bounding above the number of times a vertex x_k appears in these paths (and since each edge involves two vertices, this immediately yields an upper bound on the number of times an edge appears in these paths). We split this into two substeps:

(a) first, we derive a necessary condition for x_k to appear in the path between x_i and x_j;

(b) then, we will upper bound the number of possible pairs x_i, x_j such that this necessary condition is satisfied.

Step (ii a). For x_k to appear in the path between x_i and x_j as designed in step (i), it is necessary for $V_k \in V^{ij}$. Consider $x \in V_k \cap B(x_{ij}, r_{ij})$. Since x belongs to the Voronoi cell V_k,

$$\|x - x_k\| < \min\{\|x - x_i\|, \|x - x_j\|\},$$

but since x also lies in $B(x_{ij}, r_{ij})$,

$$\|x - x_{ij}\| < r_{ij}.$$

It follows that,

$$\begin{aligned}
\|x_k - x_{ij}\| &\leq \|x - x_k\| + \|x - x_{ij}\| \\
&\leq \|x - x_i\| + \|x - x_{ij}\| \\
&\leq \|x - x_{ij}\| + \|x_i - x_{ij}\| + \|x - x_{ij}\| \\
&\leq 3r_{ij},
\end{aligned}$$

thus if $V_k \in V^{ij}$, then it is necessary for $x_k \in B(x_{ij}, 3r_{ij})$.

Step (ii b). Recalling $\varepsilon = 2\sqrt{d}/N$, where $N = C_1(p_{\min} n / \log^\alpha n)^{1/d}$, we have a uniform upper bound of

$$\max_{\{x_i, x_j\} \in E^\varepsilon} r_{ij} \leq C_8 \left(\frac{\log^\alpha n}{n}\right)^{1/d},$$

for some $C_8 > 0$. Thus, we conclude that for an edge of x_k to be involved in a path between x_i and x_j, it is necessary for

$$x_{ij} \in B(x_k, 3C_8(\log n/n)^{1/d}),$$

or more loosely,

$$x_i, x_j \in B(x_k, 4C_8(\log n/n)^{1/d}),$$

recalling that $r_{ij} = \|x_{ij} - x_i\| = \|x_{ij} - x_j\|$ and the uniform upper bound on r_{ij}. Therefore, the number of paths in which any x_k may appear is bounded above,

$$(nP_n(\cdot, 4C_8(\log^\alpha n/n)^{1/d}))^2 \leq C_9 \log^{2\alpha} n,$$

where the final inequality is obtained by (S.85). □

G Auxiliary lemmas and proofs

G.1 Useful concentration results

The following is an immediate consequence of the well-known fact that the set of balls B in $\mathbb{R}^d$ has VC dimension $d+1$, e.g., Lemma 16 of (Chaudhuri and Dasgupta, 2010).

Lemma S.14. *Suppose $x_1, \ldots, x_n$ are drawn from P satisfying Assumption A1. There exist constants C_1-C_5 depending only on d, $p_{\min}$, and $p_{\max}$ such that the following statements hold: with probability at least $1-\delta$, for any $z \in \Omega$,*

$$\{|B(z,r) \cap \{x_1, \ldots, x_n\}| = 0\} \implies \left\{ r < C_1 \left(\frac{\log n + \log(1/\delta)}{n} \right)^{1/d} \right\}, \tag{S.84}$$

and

$$\left\{ r < C_2 \left(\frac{k - C_3(d \log n + \log(1/\delta) + \sqrt{k(d \log n + \log(1/\delta))})}{n} \right)^{1/d} \right\} \implies \{|B(z,r) \cap \{x_1, \ldots, x_n\}| < k\}. \tag{S.85}$$

In particular, if $k \geq C_4(\log(1/\delta))^2 \log n$, then

$$\{|B(z,r) \cap \{x_1, \ldots, x_n\}| \geq k\} \implies \left\{ r \geq C_5 \left(\frac{k}{n} \right)^{1/d} \right\}. \tag{S.86}$$

G.2 Properties of the Voronoi diagram

G.2.1 High probability control of cell geometry

The following lemma shows that with high probability, no Voronoi cell is very large. Let $r(V_i) := \max\{\|x - x_i\| : x \in V_i\}$ be the radius of the Voronoi cell V_i.

Lemma S.15. *Suppose $x_1, \ldots, x_n$ are drawn from P satisfying Assumption A1. There exist constants C_1 and C_2 such that the following statement holds: for any $\delta \in (0,1)$, with probability at least $1-\delta$*

$$\max_{i=1,\ldots,n} r(V_i) \leq C_1 \left(\frac{\log n + \log(1/\delta)}{n} \right)^{1/d}, \tag{S.87}$$

and

$$\max_{i=1,\ldots,n} \mu(V_i) \leq C_2 \left(\frac{\log n + \log(1/\delta)}{n} \right). \tag{S.88}$$

Proof. If $x \in V_i$, then $|B(x, \frac{1}{2}\|x - x_i\|) \cap \{x_1, \ldots, x_n\}| = 0$. (Note that the same holds true if $\frac{1}{2}$ is replaced with any $a \in [0,1)$). Taking x to be such that $\|x - x_i\| = r(V_i)$, it follows by Lemma S.14 that

$$\frac{1}{2} r(V_i) = \frac{1}{2}\|x - x_i\| \leq C \left(\frac{\log n + \log(1/\delta)}{n} \right)^{1/d},$$

with probability at least $1-\delta$. Multiplying both sides by 2 and taking a maximum over $i = 1, \ldots, n$ gives (S.87). The upper bound (S.88) on the maximum Lebesgue measure of V_i follows immediately, since $V_i \subseteq B(x, r(V_i))$. □

G.2.2 Connectedness of the Voronoi adjacency graph

The following lemma relates graph theoretic connectedness to a kind of topological connectedness that excludes connectedness using sets of $\mathcal{H}^{d-1}$-measure zero.

Lemma S.16. *Let $\Omega \subset \mathbb{R}^d$ be open such that there does* not *exist any set $S \subsetneq \Omega$ with $\mathcal{H}^{d-1}(S) = 0$ such that $\Omega \setminus S$ is disconnected. Let $\{V_1, \ldots, V_m\}$ denote an open polyhedral partition of Ω. Then the graph $G = (\{V_1, \ldots, V_m\}, E)$, where*

$$E = \{\{V_i, V_j\} : \mathcal{H}^{d-1}(\partial V_i \cap \partial V_j) > 0\},$$

is connected.

Proof. Assume by way of contradiction that G is disconnected. Therefore there exists sets of vertices $\mathcal{C}_1, \mathcal{C}_2$ such that

$$\mathcal{H}^{d-1}(\bar{V}_i \cap \bar{V}_j) = 0, \tag{S.89}$$

for all $V_i \in \mathcal{C}_1$, $V_j \in \mathcal{C}_2$. Next, define

$$\Omega_1 := (\cup_{V_i \in \mathcal{C}_1} \bar{V}_i)^\circ,$$
$$\Omega_2 := (\cup_{V_j \in \mathcal{C}_2} \bar{V}_j)^\circ,$$

such that $\{\Omega_1, \Omega_2\}$ constitutes an open partition of Ω. Let

$$\begin{aligned} S &:= \Omega \setminus (\Omega_1 \cup \Omega_2) && \text{(S.90)} \\ &= \Omega \cap ((\partial\Omega_1 \cap \partial\Omega_2) \cup ((\Omega_1^c)^\circ \cap (\Omega_2^c)^\circ)) \\ &= \Omega \cap ((\partial\Omega_1 \cap \partial\Omega_2) \cup (\Omega_2 \cap \Omega_1)) \\ &= \Omega \cap \partial\Omega_1 \cap \partial\Omega_2. && \text{(S.91)} \end{aligned}$$

From (S.89) and (S.91) we see that $\mathcal{H}^{d-1}(S) = 0$. On the other hand, (S.90) yields that $\Omega \setminus S = \Omega_1 \cup \Omega_2$ is disconnected (Ω_1, Ω_2 are open and disjoint). □

G.2.3 Analysis of the Voronoi kernel

Recall that in the proof of Theorem 1, we compare Voronoi TV to a U-statistic involving the kernel function

$$H_{\mathrm{Vor}}(x, y) = \mathbb{E}[\mathcal{H}^{d-1}(\partial V_{x_1} \cap \partial V_{x_2}) | x_1 = x, x_2 = y] = \int_{L\cap\Omega} (1 - p_x(z))^{(n-2)}\, dz.$$

The following lemma shows that this kernel function is close to a spherically symmetric kernel.

Lemma S.17. *Suppose $x_1, \ldots, x_n$ are sampled from distribution P satisfying A1. There exist constants C_1-$C_4 > 0$ such that for $h = h_n = C_1(3\log n/n)^{1/d}$, the following statements hold.*

- *For any $x, y \in \Omega_h$,*

$$H_{\mathrm{Vor}}(x, y) = \frac{\eta_{d-2}}{(np(x))^{\frac{d-1}{d}}} K_{\mathrm{Vor}}\left(\frac{\|y - x\|}{\varepsilon_{(1)}}\right) + O\left(\frac{1}{n^3} + \frac{(\log n)^2}{n} 1\{\|x - y\| \le C_2(\log n/n)^{1/d}\}\right) \tag{S.92}$$

- *For any $x, y \in \Omega$,*

$$H_{\mathrm{Vor}}(x, y) \le \frac{C_3}{n^{(d-1)/d}} K_{\mathrm{Vor}}\left(\frac{\|y - x\|}{C_4 n^{1/d}}\right). \tag{S.93}$$

Proof of (S.92). We now replace the integral above with one involving an exponential function that can be more easily evaluated. Then we evaluate this latter integral.

Step 1: Reduction to easier integral. Let $\Omega_x = \{z \in \Omega : \mathrm{dist}(z, \partial\Omega) > \|z - x\|\}$. (Note that $L \cap \Omega_x = L \cap \Omega_y$.) Separate the integral into two parts,

$$\int_{L\cap\Omega} \Big(1 - p_x(z)\Big)^{(n-2)} dz = \int_{L\cap\Omega_x} \Big(1 - p_x(z)\Big)^{(n-2)} dz + \int_{L\cap(\Omega\setminus\Omega_x)} \Big(1 - p_x(z)\Big)^{(n-2)} dz.$$

We start by showing that the second term above is negligible for $x, y \in \Omega_h$. For any $z \in \Omega \setminus \Omega_x$, it follows by the triangle inequality that

$$\mathrm{dist}(x, \partial\Omega) \le \|x - z\| + \mathrm{dist}(z, \Omega) \le 2\|x - z\|.$$

Since $x \in \Omega_h$, it follows that $p_x(z) \ge (p_{\min}/2d)\|z - x\|^d \ge (p_{\min}/2^{d+1}d)(\mathrm{dist}(x, \partial\Omega))^d \ge (p_{\min}/2^{d+1}d)h^d$. Integrating over $z \in \Omega \setminus \Omega_x$ implies an upper bound on the second term,

$$\int_{L\cap(\Omega\setminus\Omega_x)} \Big(1 - p_x(z)\Big)^{(n-2)} dz \le \int_{L\cap(\Omega\setminus\Omega_x)} \exp(-(n-2)p_x(z))\, dz$$

$$
\begin{aligned}
&= O(\exp(-(p_{\min}/2^{d+2}d)nh^d)) \\
&= O(\frac{1}{n^3}),
\end{aligned}
$$

with the last line following upon choosing $C_1 \geq (p_{\min}/2^{d+2}d)^{-1/d}$ in the definition of h.

On the other hand, if $z \in \Omega_x$ then $B(z, \|z-x\|) \subset \Omega$. Consequently, letting $\widetilde{p}_x(z) := p(x)\mu_d\|z-x\|^d$, it follows by the Lipschitz property of p that

$$
|p_x(z) - \widetilde{p}_x(z)| \leq \int_{B(z,\|z-x\|)} |p(z) - p(x)| \, dz \leq C\mu_d\|z-x\|^{d+1},
$$

and

$$
|\exp\bigl(-np_x(z)\bigr) - \exp\bigl(-n\widetilde{p}_x(z)\bigr)| \leq C\mu_d\|z-x\|^{d+1}n.
$$

Additionally recall that $\exp(-np) \geq (1-p)^n \geq \exp(-np)(1-np^2)$ for any $|p| < 1$. Combining these facts, we conclude that

$$
\begin{aligned}
\int_{L\cap\Omega_x} \Bigl(1 - p_x(z)\Bigr)^n \, dz &= \int_{L\cap\Omega_x} \exp(-np_x(z))\bigl(1 + O(np_x(z)^2)\bigr) \, dz \\
&= \int_{L\cap\Omega_x} \exp(-n\widetilde{p}_x(z))\Bigl(1 + O(n\|z-x\|^{2d}) + O(n\|z-x\|^{d+1})\Bigr) \, dz \\
&\overset{(i)}{=} \int_{L\cap\Omega_x} \exp(-np(x)\mu_d\|x-z\|^d) \, dz + O\Bigl(\frac{1}{n^3} + \frac{1}{n}1\bigl\{\|x-y\| \leq C_2(\log n/n)^{1/d}\bigr\}\Bigr) \\
&\overset{(ii)}{=} \int_L \exp\bigl(-np(x)\mu_d\|x-z\|^d\bigr) \, dz + O\Bigl(\frac{1}{n^3} + \frac{(\log n)^2}{n}1\bigl\{\|x-y\| \leq C_2(\log n/n)^{1/d}\bigr\}\Bigr).
\end{aligned}
\tag{S.94}
$$

We prove the last two equalities, which control the remainder terms, after completing our analysis of the leading order term.

Step 2: Leading order term. Let $r = \|x-y\|/2$. Due to rotational symmetry, we may as well take $x = re_1, y = -re_1$, in which case the integral becomes

$$
\begin{aligned}
\int_L \exp\bigl(-np(x)\mu_d\|x-z\|^d\bigr) \, dz &= \int_{\{0\}\times\mathbb{R}^{d-1}} \exp\bigl(-np(x)\mu_d\|re_1 - z\|^d\bigr) \, dz \\
&= \int_{\mathbb{R}^{d-1}} \exp\bigl(-np(x)\mu_d(r^2 + \|z\|^2)^{d/2}\bigr) \, dz,
\end{aligned}
$$

with the latter equality following from the Pythagorean theorem. Converting to polar coordinates, we see that

$$
\begin{aligned}
\int_{\mathbb{R}^{d-1}} \exp\bigl(-np(x)\mu_d(r^2 + \|z\|^2)^{d/2}\bigr) \, dz &= \int_0^\infty \int_{\mathbb{S}^{d-2}} \exp\bigl(-np(x)\mu_d(r^2+t^2)^{d/2}\bigr)t^{d-2} \, d\theta \, dt \\
&= \eta_{d-2} \int_0^\infty \exp\bigl(-np(x)\mu_d(r^2+t^2)^{d/2}\bigr)t^{d-2} \, dt \\
&= \frac{\eta_{d-2}}{(np(x))^{\frac{d-1}{d}}} \int_0^\infty \exp\Bigl(-\mu_d\bigl\{\bigl(r(np(x))^{1/d}\bigr)^2 + s^2\bigr\}^{d/2}\Bigr)s^{d-2} \, ds, \\
&= \frac{\eta_{d-2}}{(np(x))^{\frac{d-1}{d}}} K_{\mathrm{Vor}}\Bigl(\frac{\|y-x\|}{\varepsilon_{(1)}}\Bigr),
\end{aligned}
$$

with the second to last equality following by substituting $s = t/(np(x))^{-1/d}$.

Controlling remainder terms. We complete the proof of (S.92) by establishing (i) and (ii) in (S.94).

Proof of (i). Take $\varepsilon_0 = (4 \log n/\mu_d p_{\min} n)^{1/d}$, and note that if $\|z-x\| \geq \varepsilon_0$ then $\exp\bigl(-\mu_d n p(x)\|z-x\|^d\bigr) \leq \frac{1}{n^4}$. Recalling the definition of $\widetilde{p}_x(z)$, we have

$$
\begin{aligned}
n \int_{L\cap\Omega_x} \exp\bigl(-n\widetilde{p}_x(z)\bigr)\|z-x\|^{d+1}\,dz &= n\int_{L\cap\Omega_x} \exp\bigl(-\mu_d n p(x)\|z-x\|^d\bigr)\|z-x\|^{d+1}\,dz \\
&\leq n \int_{L\cap B(x,\varepsilon_0)} \exp\bigl(-\mu_d n p_{\min}\|z-x\|^d\bigr)\|z-x\|^{d+1}\,dz + \frac{\mathcal{H}^{d-1}(L\cap\Omega)}{n^3} \\
&\leq n\varepsilon_0^{d+1} \int_{L\cap B(x,\varepsilon_0)} \exp\bigl(-\mu_d n p_{\min}\|z-x\|^d\bigr)\,dz + \frac{\mathcal{H}^{d-1}(L\cap\Omega)}{n^3} \\
&\leq n\varepsilon_0^{d+1}\mathcal{H}^{d-1}(L\cap B(x,\varepsilon_0)) + \frac{\mathcal{H}^{d-1}(L\cap\Omega)}{n^3}.
\end{aligned} \tag{S.95}
$$

For any x, y we have $\mathcal{H}^{d-1}(L\cap B(x,\varepsilon_0)) \leq \mu_{d-1}\varepsilon_0^{d-1}$. If additionally $\|x-y\|/2 > \varepsilon_0$ then $L\cap B(x,\varepsilon_0) = \emptyset$, and so $\mathcal{H}^{d-1}(L\cap B(x,\varepsilon_0)) = 0$. Compactly, these estimates can be written as

$$
\mathcal{H}^{d-1}(L\cap B(x,\varepsilon_0)) \leq \mu_{d-1}1\{\|x-y\| \leq 2\varepsilon_0\}\varepsilon_0^{d-1}.
$$

Plugging this back into (S.95), we conclude that

$$
\begin{aligned}
n\int_{L\cap\Omega_x} \exp\bigl(-n\widetilde{p}_x(z)\bigr)\|z-x\|^{d+1}\,dz &\leq n\varepsilon_0^{2d}1\{\|x-y\|\leq 2\varepsilon_0\} + \frac{\mathcal{H}^{d-1}(L\cap\Omega)}{n^3} \\
&\leq C\Bigl(\frac{(\log n)^2}{n}1\{\|x-y\| \leq C_2(\log n/n)^{1/d}\} + \frac{1}{n^3}\Bigr),
\end{aligned}
$$

for $C_2 = 2(4/(p_{\min}\mu_d))^{1/d}$. This is precisely the claim.

Proof of (ii). Recall the fact established previously, that if $z \in L\setminus\Omega_x$ then $\|z-x\| \geq h/2$. Therefore,

$$
\begin{aligned}
\int_{L\setminus\Omega_x} \exp\bigl(-n\widetilde{p}_x(z)\bigr)\,dz &\leq \int_{L\setminus\Omega_x} \exp\bigl(-\mu_d n p_{\min}\|z-x\|^d\bigr)\,dz \\
&\leq \int_{L\setminus B(x,2)} \exp\bigl(-\mu_d n p_{\min}\|z-x\|^d\bigr)\,dz + \int_{(L\cap B(x,2)\setminus\Omega_x)} \exp\bigl(-\mu_d p_{\min} n(h/2)^d\bigr)\,dz \\
&\leq \int_{L\setminus B(x,2)} \exp\bigl(-\mu_d n p_{\min}\|z-x\|^d\bigr)\,dz + \frac{\mathcal{H}^{d-1}(L\cap B(x,2))}{n^3},
\end{aligned}
$$

with the last inequality following upon choosing $C_1 \geq 2/(\mu_d p_{\min})^{1/d}$ in the definition of h. The remaining integral is exponentially small in n, proving the upper bound (ii). □

Proof of (S.93). Note immediately that

$$
H_{\mathrm{Vor}}(x,y) \leq \int_{L\cap\Omega} \exp(-np_x(z))\,dz \leq \int_L \exp(-n\mu_d p_{\min}\|x-z\|^d/2d)\,dz.
$$

We have already analyzed this integral in the proof of (S.92), with the analysis implying that

$$
\int_L \exp(-n\mu_d p_{\min}\|x-z\|^d/2d)\,dz = \frac{\eta_{d-2}(2d)^{\frac{d-1}{d}}}{(np_{\min})^{(d-1)/d}} K_{\mathrm{Vor}}\Bigl(\frac{\|y-x\|}{(2dn/p_{\min})^{1/d}}\Bigr).
$$

This is exactly (S.93) with $C_3 = \eta_{d-2}(2d/p_{\min})^{(d-1)/d}$ and $C_4 = (2d/p_{\min})^{1/d}$. □

G.2.4 Compact kernel approximation

The kernel function $H_{\mathrm{Vor}}(x,y)$ is not compactly supported, and in our analysis it will frequently be convenient to approximate it by a compactly supported kernel. The following lemma does the trick. Let $\varepsilon_0 := (\log n/n)^{1/d}$.

Lemma S.18. *Let $x, y \in \Omega$, and $L = \{z : \|x - z\| = \|y - x\|\}$. For any $a, c > 0$, there exists a constant $C > 0$ depending only on a, c and d such that*

$$\int_{L\cap\Omega} \exp(-cn\|x - z\|^d)\, dz \le C\left(\frac{1\{\|x - y\| \le C\varepsilon_0\}}{n^{(d-1)/d}} + \frac{1}{n^a}\right) \tag{S.96}$$

where $\varepsilon_0 := (\log n/n)^{1/d}$.

Proof. Let $\widetilde{\varepsilon}_0 = C_1\varepsilon_0$ for $C_1 = (a/c)^{1/d}$. The key is that if $\|x - z\| \ge \widetilde{\varepsilon}_0$, then

$$\exp(-cn\|x - z\|^d) \le \frac{1}{n^a}.$$

Now suppose $\|y - x\| > 2\widetilde{\varepsilon}_0$. Then $\|x - z\| \ge \widetilde{\varepsilon}_0$ for all $z \in L$, and

$$\int_{L\cap\Omega} \exp(-cn\|x - z\|^d)\, dz \le \frac{\mathcal{H}^{d-1}(L\cap\Omega)}{n^a}.$$

It follows that

$$\begin{aligned}
\int_{L\cap\Omega} \exp(-cn\|x - z\|^d)\, dz &\le 1\{\|y - x\| \le 2\widetilde{\varepsilon}_0\} \int_{L\cap\Omega} \exp(-cn\|x - z\|^d)\, dz + \frac{\mathcal{H}^{d-1}(L\cap\Omega)}{n^a} \\
&\le 1\{\|y - x\| \le 2\widetilde{\varepsilon}_0\} \int_{B_{d-1}((x+y)/2, \widetilde{\varepsilon}_0)} \exp(-cn\|x - z\|^d)\, dz + 2\frac{\mathcal{H}^{d-1}(L\cap\Omega)}{n^a} \\
&\le \frac{1\{\|y - x\| \le 2\widetilde{\varepsilon}_0\}}{n^{(d-1)/d}} \int_{\mathbb{R}^{d-1}} \exp(-\|z\|^d)\, dz + 2\frac{\mathcal{H}^{d-1}(L\cap\Omega)}{n^a} \\
&\le C_2\left(\frac{1\{\|y - x\| \le 2\widetilde{\varepsilon}_0\}}{n^{(d-1)/d}} + \frac{1}{n^a}\right).
\end{aligned}$$

for $C_2 = \max\{\int_{\mathbb{R}^{d-1}} \exp(-\|z\|^d)\, dz, 2\mathcal{H}^{d-1}(L\cap\Omega)\}$. Equation (S.96) follows upon taking $C = \max\{2C_1, C_2\}$. $\square$

G.3 Proofs of technical lemmas for Theorem 5

G.3.1 Proofs of Lemmas S.7 and S.8

Proof of Lemma S.7. For each $k \in \mathcal{K}(\ell)$ and $\mathbf{i} \in \mathcal{I}$, it follows from Hölder's inequality that

$$|\theta^{\mathbf{i}}_{\ell k}(f)| \le \|f\|_{L^\infty(\Omega)} \|\Psi^{\mathbf{i}}_{\ell k}\|_{L^1(\Omega)} = \|f\|_{L^\infty(\Omega)} 2^{-\ell d/2},$$

and taking supremum over $k \in \mathcal{K}(\ell)$ and $\mathbf{i} \in \mathcal{I}$ gives the result. $\square$

Proof of Lemma S.8. The proof hinges on an application of an integration by parts identity (S.97), valid for all $f \in C^1(\Omega)$. We thus first derive (S.57) for all $f \in C^1(\Omega)$, before returning to complete the full proof.

Now, taking $f \in C^1(\Omega)$, a simple calculation verifies that for each $i = 1, \ldots, d$, and all $\Psi^{\mathbf{i}}_{\ell k}$ such that $\mathbf{i} \in \mathcal{I}$, we have

$$\int_\Omega f(x)\Psi^{\mathbf{i}}_{\ell k}(x)\, dx = -\int_\Omega D_i f I_i \Psi^{\mathbf{i}}_{\ell k}(x)\, dx. \tag{S.97}$$

Here $I_1 f(x) = \int_{-\infty}^{x_1} f((t, x_2, \ldots, x_d))\, dt$ is the partial integral operator in the 1st coordinate, and I_i are defined likewise. Now, we introduce some notation: for all $x, y \in \Omega$, and for each $i = 1, \ldots, d$, take

$$K^i_\ell(x, y) = \sum_{k\in\mathcal{K}(\ell), \mathbf{i}\in\mathcal{I}_i} \Psi^{\mathbf{i}}_{\ell,k}(x)\Psi^{\mathbf{i}}_{\ell,k}(y), \quad \Lambda^i_\ell(x, y) = \sum_{k\in\mathcal{K}(\ell), \mathbf{i}\in\mathcal{I}_i} \Psi^{\mathbf{i}}_{\ell,k}(x) I_i \Psi^{\mathbf{i}}_{\ell,k}(y).$$

By definition, K^i_ℓ is the integral operator such that

$$P_{\ell,i} f(x) := \sum_{k\in\mathcal{K}(\ell), \mathbf{i}\in\mathcal{I}_i} \theta^{\mathbf{i}}_{k\ell}(f)\Psi^{\mathbf{i}}_{\ell k}(x) = \int f(y) K^i_\ell(x, y)\, dy,$$

and we may use the integration by parts identity (S.97) to obtain

$$P_{\ell,i}f = -\int D_i f(y) \Lambda_\ell^i(\cdot, y)\, dy. \tag{S.98}$$

Taking absolute value, integrating over Ω and applying Fubini's theorem, we determine that

$$\begin{aligned}
\|P_{\ell,i}f\|_{L^1(\Omega)} &\leq \int_\Omega \int_\Omega |D_i f(y)||\Lambda_\ell^i(x,y)|\, dx\, dy \\
&\leq \|D_i f\|_{L^1(\Omega)} \cdot \sup_{y\in\Omega} \|\Lambda_\ell^i(\cdot, y)\|_{L^1(\Omega)} \\
&\leq \|D_i f\|_{L^1(\Omega)} \cdot \sup_{y\in\Omega} \sum_{k\in\mathcal{K}(\ell), \mathbf{i}\in\mathcal{I}_i} \|\Psi_{\ell k}^{\mathbf{i}}\|_{L^1(\Omega)} \cdot |I_i \Psi_{\ell,k}^{\mathbf{i}}(y)| \\
&\leq 2^d 2^{-\ell} \|D_i f\|_{L^1(\Omega)}.
\end{aligned}$$

Now, for each $i = 1, \ldots, d$, take

$$\theta_{\ell\cdot}^i(f) = \big(\theta_{\ell k}^{\mathbf{i}}(f) : k \in \mathcal{K}(\ell), \mathbf{i} \in \mathcal{I}_i\big),$$

where $\mathcal{I}_i \subset \mathcal{I}$ contains all indices $\mathbf{i} \in \mathcal{I}$ for which $\mathbf{i}_j = 0$ for all $j < i$, and $\mathbf{i}_i = 1$.

Using the $L^2(\Omega)$ orthogonality property of the Haar basis and applying Hölder's inequality gives

$$\|\theta_{\ell\cdot}^i(f)\|_1 = \|\theta_{\ell\cdot}^i(P_{\ell,i}f)\|_1 \leq \|P_{\ell,i}f\|_{L^1(\Omega)} \cdot \|\sum_{k\in\mathcal{K}, \mathbf{i}\in\mathcal{I}_i} \Psi_{\ell k}^{\mathbf{i}}\|_{L^\infty(\Omega)} \leq 2^{\ell d/2} \|P_{\ell,i}f\|_{L^1(\Omega)}.$$

and summing up over $i = 1, \ldots, d$ gives the desired upper bound on $\|\theta_{\ell\cdot}(f)\|_1$.

Finally, a density argument will imply the same result holds for any $f \in \mathrm{BV}(\Omega)$. In particular, there exists a sequence $\{f_n\} \subset C^1(\Omega)$ for which

$$\lim_{n\to\infty} \|f_n - f\|_{L^1(\Omega)} \to 0, \quad \lim_{n\to\infty} \mathrm{TV}(f_n; \Omega) = \mathrm{TV}(f; \Omega), \tag{S.99}$$

see, e.g., Evans and Gariepy (2015, Theorem 5.3); consequently

$$\|\theta_{k\cdot}(f)\|_1 = \lim_{n\to\infty} \|\theta_{k\cdot}(f_n)\|_1 \leq d2^d 2^{\ell(1-d/2)} \cdot \lim_{n\to\infty} \mathrm{TV}(f_n; \Omega) = d2^d 2^{\ell(1-d/2)} \mathrm{TV}(f; \Omega),$$

and the proof is complete upon taking $C_1 = d2^d$. □

G.3.2 Proof of Proposition S.2

The proof of Proposition S.2 follows in spirit the analysis of (Donoho et al., 1995). First we upper bound the ℓ^2-loss by the sum of two terms, a modulus of continuity and tail width, then we proceed to separately upper bound each term. The primary difference between our situation and that of (Donoho et al., 1995) is that we are working with respect to intersections of Besov bodies rather than Besov bodies themselves.

For the rest of this proof, take $\widehat{\theta} = \widehat{\theta}^{(\lambda,\ell^*)}$ and $\theta_0 = \theta(f_0)$.

Step 1: Upper bound involving modulus of continuity and tail width. We are going to establish that

$$\|\widehat{\theta} - \theta_0\|_2 \leq \Omega\Big(\Theta_\infty^{0,\infty}(2M) \cap \Theta_\infty^{1,1}(2L), \epsilon_n + \lambda\Big) + \Delta\Big(\Theta_\infty^{0,\infty}(M) \cap \Theta_\infty^{1,1}(L), \ell^*\Big), \tag{S.100}$$

where $\Omega(\Theta, \epsilon)$ is the modulus of continuity

$$\Omega(\Theta, \epsilon) := \sup_{\theta,\theta\in\Theta} \big\{\|\theta - \theta'\|_2 : \|\theta - \theta'\|_\infty \leq \epsilon\big\}, \tag{S.101}$$

and $\Delta(\Theta, \ell)$ is the tail width

$$\Delta(\Theta, \ell) := \sup_{\theta\in\Theta} \|\theta - \theta_{\leq\ell}\|_2. \tag{S.102}$$

Observe that by the triangle inequality,

$$\|\widehat{\theta} - \theta_0\|_2 \leq \|\widehat{\theta} - \theta_{0,\leq \ell^*}\|_2 + \|\theta_0 - \theta_{0,\leq \ell^*}\|_2.$$

The second term on the right hand side of the previous expression is upper bounded by $\Delta\big(\Theta_\infty^{0,\infty}(M) \cap \Theta_\infty^{1,1}(L), \ell^*\big)$. We turn now to upper bounding the first term by the modulus of continuity. Observe that for each index, we are in one of two situations: either

$$|\widetilde{\theta}^{\mathbf{i}}_{\ell,k}| < \lambda \Longrightarrow |\widehat{\theta}^{\mathbf{i}}_{\ell k}| = 0 \quad \text{ and } |\widehat{\theta}^{\mathbf{i}}_{\ell k} - \theta^{\mathbf{i}}_{0,\ell k}| = |\theta^{\mathbf{i}}_{0,\ell k}| \leq \lambda + \epsilon_n,$$

or

$$|\widetilde{\theta}^{\mathbf{i}}_{\ell,k}| \geq \lambda \Longrightarrow |\theta^{\mathbf{i}}_{0,\ell k}| \geq |\widetilde{\theta}^{\mathbf{i}}_{\ell,k}| - \epsilon_n \geq \frac{1}{2}|\widetilde{\theta}_{\ell k}| = \frac{1}{2}|\widehat{\theta}_{\ell k}| \text{ and } |\widehat{\theta}_{\ell k} - \theta^{\mathbf{i}}_{0,\ell k}| \leq \epsilon_n.$$

It follows that for every $\ell \in \mathbb{N} \cup \{0\}$, $k \in \mathcal{K}(\ell)$, we have $|\widehat{\theta}_{\ell k}| \leq 2|\theta^{\mathbf{i}}_{0,\ell k}|$, and so $\widehat{\theta} \in \Theta_\infty^{0,\infty}(2M) \cap \Theta_\infty^{1,1}(2L)$. Moreover, the above calculations also confirm that $\|\widehat{\theta} - \theta_{0,\leq \ell^*}\|_\infty \leq \lambda + \epsilon_n$. Thus

$$\|\widehat{\theta} - \theta_{0,\leq \ell^*}\|_2 \leq \Omega\big(\Theta_\infty^{0,\infty}(2M) \cap \Theta_\infty^{1,1}(2L), \epsilon_n + \lambda),$$

establishing (S.100).

Step 2: Tail width. Fix $\theta \in \Theta_\infty^{0,\infty}(M) \cap \Theta_\infty^{1,1}(L)$. For each $\ell \in \mathbb{N} \cup \{0\}$ we have

$$\|\theta_{\ell\cdot}\|_2^2 \leq \|\theta_{\ell\cdot}\|_1 \|\theta_{\ell\cdot}\|_\infty \leq C_1 2^{-\ell} LM,$$

with the first inequality following from Hölder, and the second inequality from Lemmas S.7 and S.8. Summing over $\ell = \ell^* + 1, \ell^* + 2, \ldots$ gives

$$\|\theta - \theta_{\leq \ell^*}\|_2^2 = \sum_{\ell = \ell^* + 1}^{\infty} \|\theta_{\ell\cdot}\|_2^2 \leq 2C_1 LM 2^{-\ell^*},$$

and it follows that

$$\Delta(\Theta_\infty^{0,\infty}(M) \cap \Theta_\infty^{1,1}(L), \ell^*) \leq \sqrt{2C_1 LM} 2^{-\ell^*/2}. \tag{S.103}$$

Step 3: Modulus of continuity. Fix $\theta, \theta' \in \Theta_\infty^{0,\infty}(2M) \cap \Theta_\infty^{1,1}(2L)$ such that $\|\theta - \theta'\|_\infty \leq \epsilon$. For each $\ell \in \mathbb{N} \cup \{0\}$ we have

$$\begin{aligned} \|\theta_{\ell\cdot} - \theta'_{\ell\cdot}\|_2^2 &\leq \min\{2^{\ell d}\|\theta_{\ell\cdot} - \theta'_{\ell\cdot}\|_\infty^2, \|\theta_{\ell\cdot} - \theta'_{\ell\cdot}\|_\infty \|\theta_{\ell\cdot} - \theta'_{\ell\cdot}\|_1\} \\ &\leq \min\{2^{\ell d}\epsilon^2, 2C_1 \epsilon L 2^{\ell(d/2-1)}, 4C_1 ML 2^{-\ell}\}, \end{aligned} \tag{S.104}$$

with the final inequality following from the triangle inequality, i.e. $\|\theta - \theta'\| \leq \|\theta\| + \|\theta'\|$, and Lemmas S.7 and S.8.

The three upper bounds in (S.104) divide $\mathbb{N} \cup \{0\}$ into three zones, based on the indices ℓ for which each upper bound is tightest.

- The first *dense* zone contains $\ell = 0, \ldots, \lceil \log_2(2C_1 L/\epsilon) \cdot \frac{2}{2+d} \rceil =: N_1$. In the dense zone, the extremal vectors θ are dense, i.e. everywhere non-zero.
- The second *intermediate* zone contains $\ell = N_1 + 1, \ldots, \lceil \log_2(2M/\epsilon) \cdot \frac{2}{d} \rceil =: N_2$. In the intermediate zone, the extremal vectors θ are neither dense nor sparse.
- The third *sparse* zone contains $\ell = N_2 + 1, \ldots$. In the sparse zone, the extremal vectors are sparse, i.e. they have exactly one non-zero entry.

Summing over $\ell \in \mathbb{N} \cup \{0\}$ gives

$$\begin{aligned} \|\theta - \theta'\|_2^2 &= \sum_{\ell=0}^{\infty} \|\theta_{\ell\cdot} - \theta'_{\ell\cdot}\|_2^2 \\ &\leq \epsilon^2 \sum_{\ell=0}^{N_1} 2^{\ell d} + 2C_1 L\epsilon \sum_{\ell=N_1+1}^{N_2} 2^{-\ell(1-d/2)} + 4MC_1 L \sum_{\ell=N_2+1}^{\infty} 2^{-\ell}, \end{aligned} \tag{S.105}$$

where we use the convention that $\sum_{\ell=N_1+1}^{N_2} \cdot = 0$ if $N_1 \geq N_2$. There are three terms on the right hand side of (S.105), and we derive upper bounds on each.

- For the first term, recalling that $\sum_{\ell=0}^{N} a^\ell = \frac{a^{N+1}-1}{a-1}$ for any $a > 1$, and noting that $\lceil b \rceil \le b+1$, we have

$$\epsilon^2 \sum_{\ell=0}^{N_1} 2^{\ell d} = \epsilon^2 \Big(\frac{(2^d)^{N_1+1}-1}{2^d-1} \Big) \le 4^d \epsilon^2 \Big(\frac{2C_1 L}{\epsilon} \Big)^{2/d} = 4^d (2C_1)^{2/d} L^{2/d} \epsilon^{4/(2+d)}.$$

- For the second term, we obtain separate upper bounds depending on whether $d = 2$ or $d \ge 3$: when $d = 2$,

$$2C_1 L\epsilon \sum_{\ell=N_1+1}^{N_2} 2^{-\ell(1-d/2)} = 2C_1 L\epsilon (N_2 - (N_1+1))_+ \le 2C_1 L\epsilon N_2 \le 2C_1 L\epsilon \big(\log_2(2M/\epsilon) + 1\big),$$

  and for $d \ge 3$,

$$2C_1 L\epsilon \sum_{\ell=N_1+1}^{N_2} 2^{-\ell(1-d/2)} \le 2C_1 L\epsilon 2^{(N_2+1)(d/2-1)} \le 2^d C_1 L\epsilon \Big(\frac{2M}{\epsilon} \Big)^{1-2/d} = 2^{d+1-2/d} C_1 L M (M^{-1}\epsilon)^{2/d}.$$

- For the third term,

$$4MC_1 L \sum_{\ell=N_2+1}^{\infty} 2^{-\ell} = 8MC_1 L 2^{-(N_2+1)} \le 4MC_1 L \Big(\frac{\epsilon}{2M} \Big)^{2/d} = \frac{4C_1}{2^{2/d}} L (M^{-1}\epsilon)^{2/d}.$$

Combining these upper bounds, we conclude that for an appropriate choice of constant $C_2 := 3 \max\{4^d (2C_1)^{2/d}, 2C_1, 2^{d+1-2/dC_1}, 4C_1 2^{2/d}\}$,

$$\|\theta - \theta'\|_2^2 \le C_2 \cdot \begin{cases} L\epsilon \max\{1, 1/M, \log_2(M/\epsilon)\}, & \text{if } d = 2 \\ L^{2/d} \epsilon^{4/(2+d)} + LM \Big(\frac{\epsilon}{M} \Big)^{2/d}, & \text{if } d \ge 3. \end{cases}$$

Since this holds for all θ, θ', it follows that the modulus of continuity is likewise upper bounded, i.e.

$$\Big\{ \Omega\Big(\Theta_\infty^{0,\infty}(2M) \cap \Theta_\infty^{1,1}(2L), \epsilon_n + \lambda \Big) \Big\}^2 \le C_3 \cdot \begin{cases} L\lambda \max\{1, 1/M, \log_2(M/\lambda)\}, & \text{if } d = 2 \\ L^{2/d} \lambda^{4/(2+d)} + LM \Big(\frac{\lambda}{M} \Big)^{2/d}, & \text{if } d \ge 3, \end{cases} \tag{S.106}$$

where $C_3 := 2^{1+2/d} C_2$, and we recall the assumption $\lambda \ge 2\epsilon_n$, which implies $\epsilon_n + \lambda \le 2\lambda$. Combining (S.100), (S.103) and (S.106) gives the desired upper bound (S.59). □

G.3.3 Proof of Lemma S.9

We are going to show that

$$\mathbb{P}(|\widetilde{\theta}_{\ell k}^{\mathbf{i}}(y_{1:n}) - \theta_{\ell k}^{\mathbf{i}}(f_0)| \ge \delta_n) \le \frac{3\delta}{n}. \tag{S.107}$$

From (S.107), taking a union bound over all $\ell = 0, \dots, \log_2(n)/d$, $k \in \mathcal{K}(\ell)$ and $\mathbf{i} \in \mathcal{I}$ implies the claim with $C_4 := 2^{(d+2)}$, noting that $|\mathcal{I}| = 2^d - 1$ and so

$$\sum_{\ell=0}^{\frac{1}{d}\log_2(n)} |\mathcal{I}||\mathcal{K}(\ell)| \le 2^d \sum_{\ell=0}^{\frac{1}{d}\log_2(n)} 2^\ell \le 2^{(d+2)} n.$$

It remains to show (S.107). For ease of notation, in the remainder of this proof we write $\widetilde{\theta}(\cdot) = \widetilde{\theta}_{\ell k}^{\mathbf{i}}(\cdot)$ and $\theta(\cdot) = \theta_{\ell k}^{\mathbf{i}}(\cdot)$. Decomposing $y_i = f_0(x_i) + z_i$, we have

$$\big|\widetilde{\theta}(y_{1:n}) - \theta(f_0)\big| \le \big|\widetilde{\theta}(y_{1:n}) - \widetilde{\theta}(f_0)\big| + \big|\widetilde{\theta}(f_0) - \theta(f_0)\big|, \tag{S.108}$$

and we now proceed to give high-probability upper bounds on each term in (S.108). To do so, recall Bernstein's inequality: if $x_1, \ldots, x_n$ are independent, zero-mean random variables such that $|x_i| \le b$ and $\mathbb{E}[x_i^2] \le \sigma^2$, for all $i = 1, \ldots, n$, then

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} x_i\right| \ge t\right) \le 2\exp\left(-\frac{\frac{1}{2}nt^2}{\sigma^2 + \frac{1}{3}bt}\right). \tag{S.109}$$

Term 1 in (S.108): response noise. To upper bound $|\widetilde{\theta}(y_{1:n}) - \widetilde{\theta}(f_0)| = |\widetilde{\theta}(z_{1:n})|$, we condition on the event

$$\mathcal{Z} = \big\{\max_{i=1,\ldots,n} z_i \le \sqrt{4\log(2n/\delta)}\big\},$$

which occurs with probability at least $1 - \delta/n$. Note that the following statements hold conditional on $\mathcal{Z}$:

1. The noise variables z_i are conditionally independent, $z_i \perp z_j|\mathcal{Z}$, and have conditional mean $\mathbb{E}[z_i|\mathcal{Z}] = 0$.
2. The conditional variance of $z_i\Psi_{\ell k}^{\mathbf{i}}(x_i)$ is upper bounded,

   $$\mathrm{Var}\big(z_i\Psi_{\ell k}^{\mathbf{i}}(x_i)|\mathcal{Z}\big) \le \mathbb{E}\big[z_i^2\big(\Psi_{\ell k}^{\mathbf{i}}(x_i)\big)^2|\mathcal{Z}\big] \le 2\log(2n/\delta)\mathbb{E}\big[\big(\Psi_{\ell k}^{\mathbf{i}}(x_i)\big)^2|\mathcal{Z}\big] = 2\log(2n/\delta),$$

   with the last equality following from the $L^2(\Omega)$ normalization of $\Psi_{\ell k}^{\mathbf{i}}$, along with the independence of x_i and z_i.
3. For each $i = 1, \ldots, n$,

   $$|z_i\Psi_{\ell k}^{\mathbf{i}}(x_i)| \le \sqrt{2\log(2n/\delta)}2^{\ell d/2} \le \sqrt{2n\log(2n/\delta)}.$$

We may therefore apply Bernstein's inequality (S.109) conditional on $\mathcal{Z}$, and conclude that for $\delta_{1,n} = 4\log^{3/2}(2n/\delta)/\sqrt{n}$,

$$\mathbb{P}\Big(|\widetilde{\theta}(z_{1:n})| \ge \delta_{1,n}\Big) \le \mathbb{P}(\mathcal{Z}^c) + \mathbb{P}(|\widetilde{\theta}(z_{1:n})| \ge \delta_{1,n}|\mathcal{Z}) \le \frac{2\delta}{n}. \tag{S.110}$$

Term 2 in (S.108): empirical coefficient. To upper bound $|\widetilde{\theta}(f_0)-\theta(f_0)|$, observe that the random variables $\Psi_{\ell k}(x_i)f_0(x_i)-\theta(f_0)$ for $i = 1, \ldots, n$ are mean-zero and independent. Additionally,

$$\mathrm{Var}\big(f_0(x_i)\Psi_{\ell k}^{\mathbf{i}}(x_i)\big) \le M^2$$

and

$$|f_0(x)\Psi_{\ell k}^{\mathbf{i}}(x)| \le 2^{\ell d/2}M.$$

Applying Bernstein's inequality (S.109) again, unconditionally this time, we conclude that for $\delta_{2,n} = \sqrt{12}M\sqrt{\log(2n/\delta)}/\sqrt{n}$,

$$\mathbb{P}\Big(|\widetilde{\theta}(f_0) - \theta(f_0)| \ge \delta_{2,n}\Big) \le \frac{\delta}{n}. \tag{S.111}$$

Together (S.110) and (S.111) imply (S.107), noting that $\delta_n = \delta_{1,n} + \delta_{2,n}$. □

G.3.4 Proof of Lemma S.10

Set $f(\delta) = \sum_{k=1}^{K} A_k \log^{a_k}(b_k/\delta)$. We use the identity

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > t)\,dt \le f(1) + \int_{f(1)}^\infty \mathbb{P}(X > t)\,dt. \tag{S.112}$$

Note $f^{-1}(f(1)) = 1$, $f^{-1}(\infty) = 0$ and $f'(\delta) = \sum_{k=1}^{K} A_k a_k(\log b/\delta)^{a_k-1}/\delta$. U-substitution with $t = f(\delta)$ gives

$$\begin{aligned}\mathbb{E}[X] &= -\int_0^1 \mathbb{P}(X > f(\delta))f'(\delta)\,d\delta \\ &= \sum_{k=1}^{K} a_k A_k \int_0^1 \mathbb{P}(X > f(\delta))\frac{(\log b_k/\delta)^{a_k-1}}{\delta}\,d\delta\end{aligned}$$

$$
\begin{aligned}
&\le B \sum_{k=1}^K a_k A_k \Big(\int_0^1 (\log b_k/\delta)^{a_k - 1} \, d\delta \Big) \\
&\le B \sum_{k=1}^K a_k A_k \max\{2^{a_k}, 1\} \Big((\log b_k)^{a_k-1} + \int_0^1 (\log 1/\delta)^{a_k-1} \, d\delta \Big) \\
&= B \sum_{k=1}^K a_k A_k \max\{2^{a_k}, 1\} \Big((\log b_k)^{a_k-1} + \Gamma(a_k) \Big),
\end{aligned}
$$

where the last inequality follows by the algebraic fact $(x+y)^a \le 2^a(x^a + y^a)$ for all $a > 0$, and the last equality comes from substituting $h = \log(1/\delta)$. Combining this with (S.112), we conclude that

$$
\begin{aligned}
\mathbb{E}[X] &\le \sum_{k=1}^K A_k (\log b_k)^{a_k} + B \sum_{k=1}^K a_k A_k \max\{2^{a_k}, 1\} \Big((\log b_k)^{a_k-1} + \Gamma(a_k) \Big) \\
&\le C_5 \sum_{k=1}^K A_k (\log b_k)^{a_k}
\end{aligned}
$$

for $C_5 := 2B \max_{k=1,\ldots,K} \{a_k 2^{a_k}\}$. □

References

Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, 2011.

Kamalika Chaudhuri and Sanjoy Dasgupta. Rates of convergence for the cluster tree. In *Advances in Neural Information Processing Systems*, 2010.

Albert Cohen, Wolfgang Dahmen, Ingrid Daubechies, and Ronald A. DeVore. Harmonic analysis of the space BV. *Revista Matemática Iberoamericana*, 19(1):235–263, 2003.

Miguel del Álamo, Housen Li, and Axel Munk. Frame-constrained total variation regularization for white noise regression. *Annals of Statistics*, 49(3), 2021.

David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, and Dominique Picard. Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society: Series B*, 57(2):301–337, 1995.

Lawrence C. Evans. *Partial Differental Equations*. American Mathematical Society: Graduate Studies in Mathematics, 2010. Second edition.

Lawrence C. Evans and Ronald F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, 2015. Revised edition.

Nicolás García Trillos and Dejan Slepčev. Continuum limit of total variation on point clouds. *Archive for Rational Mechanics and Analysis*, 220(1):193–241, 2016.

Nicolás García Trillos, Dejan Slepčev, and James Von Brecht. Estimating perimeter using graph cuts. *Advances in Applied Probability*, 49(4):1067–1090, 2017.

Evarist Giné and Richard Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, 2021.

Alden Green, Sivaraman Balakrishnan, and Ryan J. Tibshirani. Minimax optimal regression over Sobolev spaces via Laplacian regularization on neighborhood graphs. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2021a.

Alden Green, Sivaraman Balakrishnan, and Ryan J. Tibshirani. Minimax optimal regression over Sobolev spaces via Laplacian eigenmaps on neighborhood graphs. arXiv:2111.07394, 2021b.

Wassily Hoeffding. A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19 (3):293 – 325, 1948.

Jan-Christian Hutter and Philippe Rigollet. Optimal rates for total variation denoising. In *Proceedings of the Annual Conference on Learning Theory*, 2016.

Frederik Riis Mikkelsen and Niels Richard Hansen. Degrees of freedom for piecewise Lipschitz estimators. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, 54(2):819–841, 2018.

Oscar Hernan Madrid Padilla, James Sharpnack, James G. Scott, and Ryan J. Tibshirani. The DFS fused lasso: Linear-time denoising over general graphs. *Journal of Machine Learning Research*, 18:176–1, 2018.

Oscar Hernan Madrid Padilla, James Sharpnack, Yanzhen Chen, and Daniela Witten. Adaptive non-parametric regression with the k-nn fused lasso. *Biometrika*, 107(2):293–310, 2020.

Veeranjaneyulu Sadhanala, Yu-Xiang Wang, and Ryan J. Tibshirani. Total variation classes beyond 1d: Minimax rates, and the limitations of linear smoothers. In *Advances in Neural Information Processing Systems*, 2016.

Veeranjaneyulu Sadhanala, Yu-Xiang Wang, James Sharpnack, and Ryan J. Tibshirani. Higher-total variation classes on grids: Minimax theory and trend filtering methods. In *Advances in Neural Information Processing Systems*, 2017.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009.

Ulrike Von Luxburg, Agnes Radl, and Matthias Hein. Hitting and commute times in large random neighborhood graphs. *Journal of Machine Learning Research*, 15(1):1751–1798, 2014.

Yu-Xiang Wang, James Sharpnack, Alex Smola, and Ryan J. Tibshirani. Trend filtering on graphs. *Journal of Machine Learning Research*, 17(105):1–41, 2016.