Notes for Week 8/7/19 - 8/15/19

Alden Green

August 12, 2019

Suppose we observe samples $X = (x_1, \ldots, x_n) \in \mathcal{X} \subseteq \mathbb{R}^d$. For a given $r > 0$, the r-neighborhood graph $G = (V, E)$ over X is defined to be the undirected graph with vertex set $V = [n]$, and edge set $E = \{(i,j) : \|i - j\|_2 \leq r\}$. Let $A = (A_{ij})_{i,j=1}^n$ be the adjacency matrix of G, with entries A_{ij} equal to 1 if $(i,j) \in E$ and 0 otherwise. Let D be the edge incidence matrix of G, with ℓth row $D_\ell = (0, \ldots, -1, \ldots, 1, \ldots)$ with a -1 in the ith entry, and 1 in the jth entry, and 0 elsewhere, provided that the ℓth edge $e_\ell = (i,j)$ and $i < j$. Let $L = D^T D$ be the graph Laplacian matrix of G.

Let $\mathbb{Z}_+^d$ denote the set of all ordered d-tuples of nonnegative integers. For $\alpha \in \mathbb{Z}_+^d, \alpha = (\alpha_1, \ldots, \alpha_d)$, denote $|\alpha| = \sum_{i=1}^d \alpha_i$, and by $\mathcal{D}^\alpha f$ the partial derivative

$$\mathcal{D}^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}.$$

The Sobolev space $W^{s,p}(\mathcal{X})$ consists of all functions f such that for each multiindex α with $|\alpha| \leq k$, $D^\alpha f$ exists in the weak sense and belongs to $L^p(\mathcal{X})$. For $f \in W^{s,p}(\mathcal{X})$, we define the Sobolev norm of f to be

$$\|f\|_{W^{s,p}(\mathcal{X})} = \left(\sum_{|\alpha| \leq s} \int_{\mathcal{X}} |D^\alpha f(x)|^p\right)^{1/p}. \tag{1}$$

We will focus our attentions on $W^{s,2}(\mathcal{X})$.

For a function $f : X \to \mathbb{R}$, we can define a notion of smoothness of f over G. Letting $\widetilde{f} = (f(x_1), \ldots, f(x_n)) \in \mathbb{R}^d$, introduce the smoothness functional

$$S(f) := \widetilde{f}^T L^s \widetilde{f} \tag{2}$$

Our objective is to establish a relation between (1) and (2). In particular, suppose the samples x_i are sampled i.i.d from a probability measure P over $\mathcal{X}$. Our goal is to establish that, for an appropriate scaling C_n, with high probability

$$S(f) \leq C_n \|f\|_{W^{s,2}(\mathcal{X})} \tag{3}$$

1 Expectation of $S(f)$.

For $y \in \mathcal{X}$, define the partial difference operator D_y by $D_y f(x) = (f(y) - f(x))\eta(x,y)$, where $\eta(x,y) = \mathbf{1}(\|x - y\| \leq r)$. For $\alpha = (\alpha_1, \ldots, \alpha_q) \in [n]^q$, define the iterated difference operator D_α to be

$$D_\alpha f(x) = D_{x_{\alpha_1}}(D_{x_{\alpha_2}}(\cdots D_{x_{\alpha_q}} f))(x) \tag{4}$$

We begin by re-expressing the smoothness functional $S(f)$ in terms of sums of the iterated difference operator, similarly to [Sadhanala et al., 2017].

Lemma 1. *For s even, letting $q = s/2$, we have*

$$S(f) = \sum_{i=1}^{n} \left(\sum_{\alpha \in [n]^q} D_\alpha f(x_i) \right)^2 \tag{5}$$

For s odd, letting $q = (s-1)/2$, we have

$$S(f) = \frac{1}{2} \sum_{i,j=1}^{n} \left(\sum_{\alpha \in [n]^q} D_{x_i} D_\alpha f(x_j)) \right)^2 \tag{6}$$

Proof. Note that for any function $g : V \to \mathbb{R}$, $(-Lg)(x_i) = \sum_{j=1}^{n} D_{x_j} g(x_i)$. To see this, note that Dg is a length m vector with ℓth entry $(Dg)_\ell = g_i - g_j$ provided that $e_\ell = (i, j)$ and $i < j$. Therefore, as $Lg = \sum_{\ell=1}^{m} D_\ell (Dg)_\ell$, we have that $Lg = \sum_{j>i}(g_i - g_j)\mathbf{1}((i,j) \in E) - \sum_{j<i}(g_j - g_i)\mathbf{1}((j,i) \in E) = \sum_{j=1}^{n}(g_i - g_j)\eta_r(x_i, x_j) = \sum_{j=1}^{n} D_{x_i} g(x_j)$. The statement follows since $-D_x f(y) = D_y f(x)$.

Therefore, when s is even, letting $q = s/2$, we have

$$\begin{aligned} f^T L^s f &= \sum_{i=1}^{n} ((-L)^q f(x_i))^2 \\ &= \sum_{i=1}^{n} \left(\sum_{j_1=1}^{n} D_{x_{j_1}} L^{q-1} f(x_i) \right)^2 . \end{aligned}$$

If $q = 1$, this suffices. Otherwise, if $q \geq 2$, as D_y is a linear operator, we obtain

$$\begin{aligned} \sum_{i=1}^{n} \left(\sum_{j_1=1}^{n} D_{x_{j_1}} L^{q-1} f(x_i) \right)^2 &= \sum_{i=1}^{n} \left(\sum_{j_1=1}^{n} D_{x_{j_1}} \sum_{j_2=1}^{n} D_{x_{j_2}} L^{q-2} f(x_i) \right)^2 \\ &= \sum_{i=1}^{n} \left(\sum_{j_1 j_2=1}^{n} D_{x_{j_1}} D_{x_{j_2}} L^{q-2} f(x_i) \right)^2 \end{aligned}$$

and recursively, we arrive at the desired result.

When s is odd, letting $q = (s-1)/2$ we have that

$$\begin{aligned} f^T L^s f &= \sum_{i<j}^{n} \big((L^q f(x_i) - L^q f(x_j) \big)^2 \mathbf{1}\big((x_i, x_j) \in E \big) \\ &= \frac{1}{2} \sum_{i,j=1}^{n} \big((-L)^q f(x_i) - (-L)^q f(x_j) \big)^2 \eta(x_i, x_j). \end{aligned}$$

By similar reasoning to above, we obtain

$$f^T L^s f = \frac{1}{2} \sum_{i,j=1}^{n} \left(\sum_{\alpha \in [n]^q} D_\alpha f(x_i) - D_\alpha f(x_j) \right)^2 \eta(x_i, x_j)$$

Then, since $\eta = \eta^2$, we can replace $\eta(x_i, x_j)$ with $\eta^2(x_i, x_j)$ in the previous display, and obtain

$$\begin{aligned} f^T L^s f &= \frac{1}{2} \sum_{i,j=1}^{n} \left(\sum_{\alpha \in [n]^q} D_\alpha f(x_i) - D_\alpha f(x_j) \right)^2 \eta^2(x_i, x_j) \\ &= \frac{1}{2} \sum_{i,j=1}^{n} \left(\sum_{\alpha \in [n]^q} (D_\alpha f(x_i) - D_\alpha f(x_j)) \eta(x_i, x_j) \right)^2 \end{aligned}$$

and (6) follows. □

1.1 Expectation when $s = 2$.

Let P be absolutely continuous with density function p over $\mathbb{R}^d$.

Lemma 2. *Suppose $f \in W^{2,2}(\mathcal{X})$, $p \in C^1(\mathcal{X})$, and $\partial X \in C^2$. Then there exists a constant c which does not depend on n or f such that*

$$\mathbb{E}(f^T L^2 f) \le c \left(n^2 r^{d+2} \sum_{|\alpha|=1} \|\mathcal{D}^\alpha f\|_2^2 + n^3 r^{2d+4} \sum_{|\alpha| \le 2} \|\mathcal{D}^\alpha f\|_2^2 \right)$$

Proof. Using Lemma 1, we may rewrite

$$\begin{aligned} f^T L^2 f &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} D_{x_j} f(x_i) \right)^2 \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} D_{x_j} f(x_i) D_{x_k} f(x_i) \\ &= \sum_{i \ne j}^{n} \big(D_{x_j} f(x_i)\big)^2 + \sum_{i \ne j \ne k}^{n} D_{x_j} f(x_i) D_{x_k} f(x_i) \end{aligned}$$

so that the expectation $\mathbb{E}(f^T L^2 f)$ becomes

$$\mathbb{E}(f^T L^2 f) = n(n-1)\mathbb{E}\left(D_{X_1} f(X_2)\right)^2 + n(n-1)(n-2)\mathbb{E}\left(D_{X_2} f(X_1) D_{X_3} f(X_2)\right).$$

The statement then follows from Lemmas 3 and 4. □

2 Additional Theory

Lemma 3. *For any function $f \in W^{2,2}(\mathcal{X})$, if $\partial X \in C^2$ and $p \in C^1(\mathcal{X}, L)$ then there exists a constant c which does not depend on f such that*

$$\mathbb{E}(D(X_2) f(X_1) D_{X_3} f(X_1)) \le c r^{2d+4} \sum_{|\alpha| \le 2} \|\mathcal{D}^\alpha f\|_{L^2}^2 \tag{7}$$

Proof. We rewrite the expectation on the right hand side of (7) as an integral:

$$\begin{aligned}\mathbb{E}(D(X_2)f(X_1)D_{X_3}f(X_1)) &= \int_{\mathcal{X}}\int_{\mathcal{X}}\int_{\mathcal{X}} D_y f(x) D_z f(x)\, dP(z)\, dP(y)\, dP(x)\\ &= \int_{\mathcal{X}} \left\{\int_{\mathcal{X}} D_y f(x)\, dP(y)\right\}^2 dP(x)\\ &\leq 2\int_{\mathcal{X}} \left\{\int_{\mathcal{X}} D_y f(x) p(x)\, dy\right\}^2 + \left\{\int_{\mathcal{X}} D_y f(x)(p(y)-p(x))\, dy\right\}^2 dP(x)\\ &\leq 2\int_{\mathbb{R}^d} \left\{\int_{\mathcal{X}} D_y g(x) p(x)\, dy\right\}^2 + \left\{\int_{\mathbb{R}^d} D_y g(x)(p(y)-p(x))\, dy\right\}^2 dP(x)\end{aligned}$$

where g is an extension of f which is equal to f almost everywhere on $\mathcal{X}$, and satisfies $\|g\|_{W^{2,2}(\mathbb{R}^d)} \leq c\|f\|_{W^{2,2}(\mathbb{R}^d)}$ for some c which does not depend on f. We will handle each term inside the integral separately. Assume without loss of generality that $g \in C^2$; otherwise, we can take $g_m \in C^2$ such that $\|g_m - g\|_{W^{2,2}(\mathbb{R}^d)} \to 0$. For the first term, we use Lemma 5 to obtain

$$\left\{\int_{\mathcal{X}} D_y g(x) p(x)\, dy\right\}^2 dP(x) \leq \lambda_{\max}^3 \frac{d^2 r^{4+2d}}{\nu_d} \sum_{|\alpha|=2} \int_{B(0,1)} \int_0^1 (1-t)^2 \left(\mathcal{D}^\alpha g(x+trz)\right)^2 dt\, dz,$$

and therefore,

$$\begin{aligned}\int_{\mathbb{R}^d} \left\{\int_{\mathcal{X}} D_y g(x) p(x)\, dy\right\}^2 &\leq \lambda_{\max}^3 \frac{d^2 r^{4+2d}}{\nu_d} \sum_{|\alpha|=2} \int_{\mathbb{R}^d}\int_{B(0,1)} \int_0^1 (1-t)^2 \left(\mathcal{D}^\alpha g(x+trz)\right)^2 dt\, dz\, dx\\ &= \lambda_{\max}^3 \frac{d^2 r^{4+2d}}{\nu_d} \sum_{|\alpha|=2} \int_{B(0,1)} \int_0^1 (1-t)^2 \int_{\mathbb{R}^d} \left(\mathcal{D}^\alpha g(x+trz)\right)^2 dx\, dt\, dz\\ &= \frac{\lambda_{\max}^3 d^2 r^{4+2d}}{3} \sum_{|\alpha|=2} \|\mathcal{D}^\alpha g\|_{L^2}^2.\end{aligned}$$

For the second summand, since $p(x) \in C^2(\mathcal{X}; L)$, using reasoning similar to the proof of Lemma we have

$$\begin{aligned}\int_{\mathbb{R}^d} \left\{\int_{\mathbb{R}^d} D_y g(x)(p(y)-p(x))\, dy\right\}^2 dP(x) &\leq \lambda_{\max} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} (D_y g(x))^2\, dy\right)\left(\int_{B(x,r)} (p(y)-p(x))^2\, dy\right) dx\\ &\leq \lambda_{\max} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} (D_y g(x))^2\, dy\right)\left(\int_{B(x,r)} L^2\|x-y\|^2\, dy\right) dx\\ &\leq \lambda_{\max}\nu_d r^{d+2} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} (D_y g(x))^2\, dy\, dx\\ &\leq c\nu_d \lambda_{\max}\nu_d r^{2d+4} \sum_{|\alpha|=1} \|\mathcal{D}^\alpha f\|_2^2.\end{aligned}$$

□

Lemma 4. *For any function $f \in W^{1,2}(\mathcal{X})$, if $\partial\mathcal{X} \in C^1$, then there exists a constant c which does not depend on f such that*

$$\mathbb{E}\left(D_{X_1} f(X_2)\right)^2 \leq c r^{d+2} \sum_{|\alpha|=1} \|\mathcal{D}^\alpha f\|_2^2$$

Proof. We write $\mathbb{E}\left(D_{X_1} f(X_2)\right)^2$ as an integral,

$$\mathbb{E}\left(D_{X_1} f(X_2)\right)^2 = \int_{\mathcal{X}} \int_{\mathcal{X}} (f(x) - f(y))^2 \eta(y, x)\, dP(y)\, dP(x) \leq \lambda_{\max}^2 \int_{\mathcal{X}} \int_{\mathcal{X}} (f(x) - f(y))^2 \eta(y, x)\, dy\, dx$$

Now, as $\partial\mathcal{X} \in C^1$, there exists a function $g : \mathbb{R}^d \to \mathbb{R}$ such that $g(x) = f(x)$, almost everywhere in $\mathcal{X}$, and $\|g\|_{W^{1,2}(\mathbb{R}^d)} \leq c\|f\|_{W^{1,2}(\mathcal{X})}$. Since $g = f$ almost everywhere in $\mathcal{X}$, we have

$$\int_{\mathcal{X}} \int_{\mathcal{X}} (f(x) - f(y))^2\, dy\, dx \leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (g(x) - g(y))^2 \eta(y, x)\, dy\, dx. \tag{8}$$

We may assume without loss of generality that $g \in C^\infty$ (otherwise take approximations $g_m \in C^\infty(\mathbb{R}^d)$, $g_m \to g \in W^{s,2}(\mathbb{R}^d)$.) We expand

$$\begin{aligned}
\int_{\mathbb{R}^d} (g(x) - g(y))^2\, dy &= \int_{\mathbb{R}^d} \left(\int_0^1 \frac{d}{dt} g(x + t(y - x))\, dt \right)^2 dy \\
&= \int_{\mathbb{R}^d} \left(\int_0^1 \nabla g(x + t(y - x)) \cdot (y - x)\, dt \right)^2 dy \\
&\leq \int_{\mathbb{R}^d} \int_0^1 \|\nabla g(x + t(y - x))\|^2 \|y - x\|^2\, dt\, dy
\end{aligned}$$

Therefore,

$$\begin{aligned}
\int_{\mathbb{R}^d} (g(x) - g(y))^2 \eta(y, x)\, dy\, dx &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_0^1 \|\nabla g(x + t(y - x))\|^2 \|y - x\|^2\, dt \eta(y, x)\, dy\, dx \\
&\leq r^2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_0^1 \|\nabla g(x + t(y - x))\|^2\, dt \eta(y, x)\, dy\, dx \\
&= r^{2+d} \int_{\mathbb{R}^d} \int_{B(0,1)} \int_0^1 \|\nabla g(x + trz)\|^2\, dt\, dz\, dx \qquad (z = (y - x)/r) \\
&= r^{2+d} \int_{B(0,1)} \int_0^1 \int_{\mathbb{R}^d} \|\nabla g(x + trz)\|^2\, dx\, dt\, dz \\
&= \nu_d r^{2+d} \sum_{|\alpha|=1} \|\mathcal{D}^\alpha g\|_{L^2}^2
\end{aligned}$$

□

Lemma 5. *For any function $g \in C^2(\mathbb{R}^d)$, and any $x \in \mathbb{R}^d$*

$$\left(\int_{B(x,r)} g(y) - g(x)\, dy \right)^2 \leq \frac{d^2 r^{4+2d}}{\nu_d} \sum_{|\alpha|=2} \int_{B(0,1)} \int_0^1 (1 - t)^2 \left(\mathcal{D}^\alpha f(x + trz)\right)^2\, dt\, dz.$$

Proof. As $g \in C^2(\mathbb{R}^d)$, taking a Taylor expansion of g around x, we obtain

$$g(y) = g(x) + \sum_{|\alpha|=1} \mathcal{D}^\alpha g(x)(y - x)^\alpha + \sum_{|\alpha|=2} (y - x)^\alpha \int_0^1 (1 - t)\mathcal{D}^\alpha g(x + t(y - x))\, dt.$$

We note that, by standard facts of the uniform distribution

$$\int_{B(x,r)} (x - y)^\alpha = 0, \quad \text{for all } |\alpha| = 1$$

and we therefore have

$$
\begin{aligned}
\left(\int_{B(x,r)} g(y)-g(x)\,dy\right)^2 &= \left(\int_{B(x,r)} \sum_{|\alpha|=2} (y-x)^\alpha \int_0^1 (1-t)\mathcal{D}^\alpha g(x+t(y-x))\,dt\right)^2 \\
&\leq \left(\int_{B(x,r)} \left(\sum_{|\alpha|=2} (y-x)^{2\alpha}\right)^{1/2} \left(\sum_{|\alpha|=2} \left\{\int_0^1 (1-t)\mathcal{D}^\alpha g(x+t(y-x))\,dt\right\}^2\right)^{1/2} dy\right)^2 \\
&\leq d^2 r^4 \left(\int_{B(x,r)} \left(\sum_{|\alpha|=2} \int_0^1 (1-t)^2 \left[\mathcal{D}^\alpha g(x+t(y-x))\right]^2\, dt\right)^{1/2} dy\right)^2 \\
&= d^2 r^{4+2d} \left(\int_{B(0,1)} \left(\sum_{|\alpha|=2} \int_0^1 (1-t)^2 \left[\mathcal{D}^\alpha g(x+trz)\right]^2\, dt\right)^{1/2} dz\right)^2 \\
&\leq \frac{d^2 r^{4+2d}}{\nu_d} \int_{B(0,1)} \sum_{|\alpha|=2} \int_0^1 (1-t)^2 \left[\mathcal{D}^\alpha g(x+trz)\right]^2\, dt\,dz
\end{aligned}
$$

□

REFERENCES

Veeranjaneyulu Sadhanala, Yu-Xiang Wang, James L Sharpnack, and Ryan J Tibshirani. Higher-order total variation classes on grids: Minimax theory and trend filtering methods. In *Advances in Neural Information Processing Systems*, pages 5800–5810, 2017.