Supplementary Material for Poisson Learning: Graph Based Semi-Supervised Learning At Very Low Label Rates

A. Proofs

We provide the proofs by section.

A.1. Proofs for Section 2.1

We recall $X_0^x, X_1^x, X_2^x, \ldots$ is a random walk on X starting at $X_0^x = x$ with transition probabilities

$$\mathbb{P}(X_k^x = x_j \mid X_{k-1}^x = x_i) = \frac{w_{ij}}{d_i}.$$

Before giving the proof of Theorem 2.1, we recall some properties of random walks and Markov chains. The random walk described above induces a Markov chain with state space X. Since the graph is connected and X is finite, the Markov chain is *positive recurrent*. We also assume the Markov chain is *aperiodic*. This implies the distribution of the random walker converges to the invariant distribution of the Markov chain as $k \to \infty$. In particular, choose any initial distribution $p_0 \in \ell^2(X)$ such that $\sum_{i=1}^n p_0(x_i) = 1$ and $p_0 \geq 0$, and define

$$p_{k+1}(x_i) = \sum_{j=1}^n \frac{w_{ij}}{d_j} p_k(x_j). \tag{A.1}$$

Then p_k is the distribution of the random walker after k steps. Since the Markov chain is positive recurrent and aperiodic we have that

$$\lim_{k\to\infty} p_k(x_i) = \pi(x_i)$$

for all i, where

$$\pi(x_i) = \frac{d_i}{\sum_{i=1}^n d_i}$$

is the invariant distribution of the Markov chain. It is simple to check that if $p_0 \in \ell^2(X)$ is any function (i.e., not necesarily a probability distribution), and we define p_k by the iteration (A.1), then

$$\lim_{k\to\infty} p_k(x_i) = \pi(x_i) \sum_{j=1}^n p_0(x_j). \tag{A.2}$$

We now give the proof of Theorem 2.1.

Proof of Theorem 2.1. Define the normalized Green's function

$$G_T(x_i, x_j) = \frac{1}{d_i} \mathbb{E}\left[\sum_{k=0}^T \mathbb{1}_{\{X_k^{x_j} = x_i\}}\right] = \frac{1}{d_i} \sum_{k=0}^T \mathbb{P}(X_k^{x_j} = x_i).$$

Then we have

$$\begin{aligned} d_i G_T(x_i, x_j) &= \delta_{ij} + \sum_{k=1}^{T}\sum_{\ell=1}^{n} \frac{w_{\ell i}}{d_\ell}\mathbb{P}(X_{k-1}^{x_j} = x_\ell) \\ &= \delta_{ij} + \sum_{\ell=1}^{n} \frac{w_{\ell i}}{d_\ell}\sum_{k=1}^{T}\mathbb{P}(X_{k-1}^{x_j} = x_\ell) \\ &= \delta_{ij} + \sum_{\ell=1}^{n} \frac{w_{\ell i}}{d_\ell}\sum_{k=0}^{T-1}\mathbb{P}(X_{k}^{x_j} = x_\ell) \\ &= \delta_{ij} + \sum_{\ell=1}^{n} w_{\ell i} G_{T-1}(x_\ell, x_j). \end{aligned}$$

Therefore we have

$$d_i(G_T(x_i,x_j) - G_{T-1}(x_i,x_j)) + \mathcal{L}G_{T-1}(x_i,x_j) = \delta_{ij},$$

where the Laplacian $\mathcal{L}$ is applied to the first variable of G_{T-1} while the second variable is fixed (i.e. $\mathcal{L}G_{T-1}(x_i,x_j) = [\mathcal{L}G_{T-1}(\cdot,x_j)]_{x_i}$). Since

$$u_T(x_i) = \sum_{j=1}^{m}(y_j - \overline{y}_{\mathrm{u}})G_T(x_i,x_j)$$

we have

$$d_i(u_T(x_i) - u_{T-1}(x_i)) + \mathcal{L}u_{T-1}(x_i) = \sum_{j=1}^{m}(y_j - \overline{y}_{\mathrm{u}})\delta_{ij}.$$

Summing both sides over $i = 1, \dots, n$ we find that

$$(u_T)_{d,X} = \sum_{i=1}^{n} d_i u_T(x_i) = \sum_{i=1}^{n} d_i u_{T-1}(x_i) = (u_{T-1})_{d,X},$$

where $d = (d_1, d_2, \dots, d_n)$ is the vector of degrees. Therefore $(u_T)_{d,X} = (u_{T-1})_{d,X} = \cdots = (u_0)_{d,X}$. Noting that

$$d_i u_0(x_i) = \sum_{j=1}^{m}(y_j - \overline{y}_{\mathrm{u}})\delta_{ij},$$

we have $(u_0)_{d,X} = 0$, and so $(u_T)_{d,X} = 0$ for all $T \geq 0$. Let $u \in \ell^2(X)$ be the solution of

$$\mathcal{L}u(x_i) = \sum_{j=1}^{m}(y_j - \overline{y}_{\mathrm{u}})\delta_{ij}$$

satisfying $(u)_{d,X} = 0$. Define $v_T(x_i) = d_i(u_T(x_i) - u(x_i))$. We then check that v_T satisfies

$$v_T(x_i) = \sum_{j=1}^{n} \frac{w_{ij}}{d_j} v_{T-1}(x_j),$$

and $(v_T)_X = 0$. Since the random walk is aperiodic and the graph is connected, we have by (A.2) that $\lim_{T\to\infty} v_T(x_i) = \pi(x_i)(v_0)_X = 0$, which completes the proof. □

A.2. Proofs for Section 2.2

We first review some additional calculus on graphs. The *graph divergence* of a vector field V is defined as

$$\mathrm{div}\, V(x_i) = \sum_{j=1}^{n} w_{ij} V(x_i, x_j).$$

The divergence is the negative adjoint of the gradient; that is, for every vector field $V \in \ell^2(X^2)$ and function $u \in \ell^2(X)$

$$(\nabla u, V)_{\ell^2(X^2)} = -(u, \operatorname{div} V)_{\ell^2(X)}. \tag{A.3}$$

We also define $\|u\|_{\ell^p(X)}^p = \sum_{i=1}^n |u(x_i)|^p$ and

$$\|V\|_{\ell^p(X^2)}^p = \frac{1}{2}\sum_{i,j=1}^n w_{ij}|V(x_i,x_j)|^p,$$

where $|\cdot|$ is the Euclidean norm on $\mathbb{R}^k$.

The graph Laplacian $\mathcal{L}u$ of a function $u \in \ell^2(X)$ is defined as negative of the composition of gradient and divergence

$$\mathcal{L}u(x_i) = -\operatorname{div}(\nabla u)(x_i) = \sum_{j=1}^n w_{ij}(u(x_i) - u(x_j)).$$

The operator $\mathcal{L}$ is the *unnormalized* graph Laplacian. Using (A.3) we have

$$(\mathcal{L}u, v)_{\ell^2(X)} = (-\operatorname{div}\nabla u, v)_{\ell^2(X)} = (\nabla u, \nabla v)_{\ell^2(X^2)}.$$

In particular $(\mathcal{L}u, v)_{\ell^2(X)} = (u, \mathcal{L}v)_{\ell^2(X)}$, and so the graph Laplacian $\mathcal{L}$ is self-adjoint as an operator $\mathcal{L} : \ell^2(X) \to \ell^2(X)$. We also note that

$$(\mathcal{L}u, u)_{\ell^2(X)} = (\nabla u, \nabla u)_{\ell^2(X^2)} = \|\nabla u\|_{\ell^2(X^2)}^2,$$

that is, $\mathcal{L}$ is positive semi-definite.

The variational interpretation of Poisson learning can be directly extended to ℓ^p versions, so we proceed in generality here. For a function $u : X \to \mathbb{R}^k$ and a positive vector $a \in \mathbb{R}^n$ (meaning $a_i > 0$ for all $i = 1, \dots, n$) we define the weighted mean value

$$(u)_{a,X} := \frac{1}{\sum_{i=1}^n a_i}\sum_{i=1}^n a_i u(x_i).$$

We define the space of weighted mean-zero functions

$$\ell^p_{a,0}(X) = \{u \in \ell^p(X) \, : \, (u)_{a,X} = 0\}.$$

For $p \geq 1$ and $\mu > 0$ we consider the variational problem

$$\min_{u \in \ell^p_{a,0}(X)} \left\{ \frac{1}{p}\|\nabla u\|_{\ell^p(X^2)}^p - \mu\sum_{j=1}^m (y_j - \overline{y}_{\mathrm{u}}) \cdot u(x_j) \right\} \tag{A.4}$$

where $\overline{y}_{\mathrm{u}} = \frac{1}{m}\sum_{j=1}^m y_j$. This generalizes the variational problem (2.8) for Poisson learning, and the theorem below generalizes Theorem 2.3.

Theorem A.1. *Assume G is connected. For any $p > 1$, positive $a \in \mathbb{R}^n$, and $\mu \geq 0$, there exists a unique solution $u \in \ell^p_{a,0}(X)$ of (A.4). Furthermore, the minimizer u satisfies the graph p-Laplace equation*

$$-\operatorname{div}(|\nabla u|^{p-2}\nabla u)(x_i) = \mu\sum_{j=1}^m (y_j - \overline{y}_{\mathrm{u}})\delta_{ij}. \tag{A.5}$$

We give the proof of Theorem A.1 below, after some remarks and other results.

Remark A.2. When $p = 1$, solutions of (A.4) may not exist for all $\mu \geq 0$, since the variational problem (A.4) may not be bounded from below. We can show that there exists $C > 0$ such that if $\mu < C$, the variational problem is bounded from below and our argument for existence in Theorem A.1 goes through.

It turns out that $\mu > 0$ is a redundant parameter when $p > 1$.

Lemma A.3. *Let $p > 1$ and for $\mu > 0$ let u_μ be the solution of* (A.4)*. Then,* $u_\mu = \mu^{1/(p-1)} u_1$.

It follows from Lemma A.3 that when $p > 1$, the fidelity parameter $\mu > 0$ is *completely irrelevant* for classification problems, since the identity $u_\mu = \mu^{1/(p-1)} u_1$ implies that the label decision (2.2) gives the same labeling for *any* value of $\mu > 0$. Hence, in Poisson learning with $p > 1$ we always take $\mu = 1$. This remark is false for $p = 1$.

Before proving Theorem A.1 we first record a Poincaré inequality. The proof is standard but we include it for completeness. We can prove the Poincaré inequality for non-negative vectors $a \in \mathbb{R}^n$, meaning that $a_i \geq 0$ for every $i = 1, \dots, n$ as long as $\sum_{i=1}^n a_i > 0$.

Proposition A.4. *Assume G is connected, $a \in \mathbb{R}^d$ is non-negative with $\sum_{i=1}^n a_i > 0$, and $p \geq 1$. There exists $\lambda_p > 0$ such that*

$$\lambda_p \|u - (u)_{a,X}\|_{\ell^p(X)} \leq \|\nabla u\|_{\ell^p(X^2)}, \tag{A.6}$$

for all $u \in \ell^p(X)$.

Proof. Define

$$\lambda_p = \min_{\substack{u \in \ell^p(X) \\ u \not\equiv (u)_{a,X}}} \frac{\|\nabla u\|_{\ell^p(X^2)}}{\|u - (u)_{a,X}\|_{\ell^p(X)}}.$$

Then clearly (A.6) holds for this choice of λ_p, and $\lambda_p \geq 0$. If $\lambda_p = 0$, then there exists a sequence $u_k \in \ell^p(X)$ with $u_k \not\equiv (u_k)_{a,X}$ such that

$$\frac{\|\nabla u_k\|_{\ell^p(X^2)}}{\|u_k - (u)_{a,X}\|_{\ell^p(X)}} \leq \frac{1}{k}.$$

We may assume that $(u_k)_{a,X} = 0$ and $\|u_k\|_{\ell^p(X)} = 1$, and so

$$\|\nabla u_k\|_{\ell^p(X^2)} \leq \frac{1}{k}. \tag{A.7}$$

Since $|u_k(x)| \leq \|u_k\|_{\ell^p(X)} = 1$, the sequence u_k is uniformly bounded and by the Bolzano-Weierstrauss Theorem there exists a subsequence u_{k_j} such that $u_{k_j}(x_i)$ is a convergent sequence in $\mathbb{R}^k$ for every i. We denote $u(x_i) = \lim_{j\to\infty} u_{k_j}(x_i)$. Since $\|u_{k_j}\|_{\ell^p(X)} = 1$ we have $\|u\|_{\ell^p(X)} = 1$, and thus $u \not\equiv 0$. Similarly, since $(u_k)_{a,X} = 0$ we have $(u)_{a,X} = 0$ as well. On the other hand it follows from (A.7) that $\|\nabla u\|_{\ell^p(X^2)} = 0$, and so

$$w_{ij}(u(x_i) - u(x_j)) = 0 \quad \text{for all } i, j.$$

It follows that $u(x_i) = u(x_j)$ whenever $w_{ij} > 0$. Since the graph is connected, it follows that u is constant. Since $(u)_{a,X} = 0$ we must have $u \equiv 0$, which is a contradiction, since $\|u\|_{\ell^p(X)} = 1$. Therefore $\lambda_p > 0$, which completes the proof. □

We can now prove Theorem A.1.

Proof of Theorem A.1. Let us write

$$I_p(u) = \frac{1}{p}\|\nabla u\|^p_{\ell^p(X^2)} - \mu \sum_{j=1}^m (y_j - \overline{y}_{\mathrm{u}}) \cdot u(x_j). \tag{A.8}$$

By Proposition A.4 we have

$$I_p(u) \geq \frac{1}{p}\lambda_p^p \|u\|^p_{\ell^p(X)} - \mu \sum_{j=1}^m (y_j - \overline{y}_{\mathrm{u}}) \cdot u(x_j)$$

for $u \in \ell^p_{a,0}(X)$. By Hölder's inequality we have

$$\begin{aligned}\sum_{j=1}^m (y_j - \overline{y}_{\mathrm{u}}) \cdot u(x_j) &\leq \sum_{j=1}^m |y_j - \overline{y}_{\mathrm{u}}||u(x_j)| \\ &\leq \left(\sum_{j=1}^m |y_j - \overline{y}_{\mathrm{u}}|^q\right)^{1/q} \left(\sum_{j=1}^m |u(x_j)|^p\right)^{1/p} \\ &\leq \left(\sum_{j=1}^m |y_j - \overline{y}_{\mathrm{u}}|^q\right)^{1/q} \|u\|_{\ell^p(X)},\end{aligned}$$

where $q = p/(p-1)$. Letting $C = \left(\sum_{j=1}^m |y_j - \overline{y}_{\mathrm{u}}|^q\right)^{1/q}$ we have

$$I_p(u) \geq \frac{1}{p}\lambda_p^p \|u\|^p_{\ell^p(X)} - C\mu\|u\|_{\ell^p(X)}. \tag{A.9}$$

Since $p > 1$, we see that I_p is bounded below.

Let $u_k \in \ell^p_{a,0}(X)$ be a minimizing sequence, that is, we take u_k so that

$$-\infty < \inf_{u \in \ell^p_{a,0}(X)} I_p(u) = \lim_{k\to\infty} I_p(u_k).$$

By (A.9) we have that

$$\frac{1}{p}\lambda_p^p \|u_k\|^p_{\ell^p(X)} - C\mu\|u_k\|_{\ell^p(X)} \leq \inf_{u \in \ell^p_{a,0}(X)} I_p(u) + 1,$$

for k sufficiently large. Since $p > 1$, it follows that there exists $M > 0$ such that $\|u_k\|_{\ell^p(X)} \leq M$ for all $k \geq 1$. Since $|u_k(x_i)| \leq \|u_k\|_{\ell^p(X)} \leq M$ for all $i = 1, \ldots, n$, we can apply the Bolzano-Weierstrauss Theorem to extract a subsequence u_{k_j} such that $u_{k_j}(x_i)$ is a convergent sequence in $\mathbb{R}^k$ for all $i = 1, \ldots, n$. We denote by $u^*(x_i)$ the limit of $u_{k_j}(x_i)$ for all i. By continuity of I_p we have

$$\inf_{u \in \ell^p_{a,0}(X)} I_p(u) = \lim_{j\to\infty} I_p(u_{k_j}) = I_p(u^*),$$

and $(u^*)_{a,X} = 0$. This shows that there exists a solution of (A.4).

We now show that any solution of (A.4) satisfies $-\mathrm{div}\left(|\nabla u|^{p-2}\nabla u\right) = \mu f$. The proof follows from taking a variation. Let $v \in \ell^p_{a,0}(X)$ and consider $g(t) := I_p(u + tv)$, where I_p is defined in (A.8). Then g has a minimum at $t = 0$ and hence $g'(0) = 0$. We now compute

$$\begin{aligned}g'(0) &= \frac{d}{dt}\Big|_{t=0} \left\{\frac{1}{p}\|\nabla u + t\nabla v\|^p_{\ell^p(X^2)} - \mu\sum_{j=1}^m (y_j - \overline{y}_{\mathrm{u}}) \cdot (u(x_j) + tv(x_j))\right\} \\ &= \frac{1}{2p}\sum_{i,j=1}^n w_{ij}\frac{d}{dt}\Big|_{t=0} |\nabla u(x_i,x_j) + t\nabla v(x_i,x_j)|^p - \mu\sum_{j=1}^m (y_j - \overline{y}_{\mathrm{u}}) \cdot v(x_j) \\ &= \frac{1}{2}\sum_{i,j=1}^n w_{ij}|\nabla u(x_i,x_j)|^{p-2}\nabla u(x_i,x_j) \cdot \nabla v(x_i,x_j) - \mu\sum_{j=1}^m (y_j - \overline{y}_{\mathrm{u}}) \cdot v(x_j) \\ &= (|\nabla u|^{p-2}\nabla u, \nabla v)_{\ell^2(X^2)} - \mu\sum_{j=1}^m (y_j - \overline{y}_{\mathrm{u}}) \cdot v(x_j) \\ &= (-\mathrm{div}\,(|\nabla u|^{p-2}\nabla u), v)_{\ell^2(X)} - \mu\sum_{j=1}^m (y_j - \overline{y}_{\mathrm{u}}) \cdot v(x_j) \\ &= (-\mathrm{div}\,(|\nabla u|^{p-2}\nabla u) - \mu f, v)_{\ell^2(X)},\end{aligned}$$

where

$$f(x_i) = \sum_{j=1}^{m} (y_j - \overline{y}_{\mathrm{u}})\delta_{ij}.$$

We choose

$$v(x_i) = \frac{1}{a_i}\left(-\text{div}\,\left(|\nabla u|^{p-2}\nabla u\right)(x_i) - \mu f(x_i)\right)$$

then

$$(v)_{a,X} = \sum_{i=1}^{n}\left(-\text{div}\,\left(|\nabla u|^{p-2}\nabla u\right)(x_i) - \mu f(x_i)\right) = 0$$

so $v \in \ell^p_{a,0}(X)$. Moreover, for this choice of v,

$$0 = g'(0) = \sum_{i=1}^{n} \frac{1}{a_i}\left|\text{div}\,\left(|\nabla u|^{p-2}\nabla u\right)(x_i) + \mu f(x_i)\right|^2 \geq \frac{1}{\max a_i}\left\|\text{div}\,\left(|\nabla u|^{p-2}\nabla u\right)(x_i) + \mu f(x_i)\right\|^2_{\ell^2(X)}.$$

So, $-\text{div}\,\left(|\nabla u|^{p-2}\nabla u\right) = \mu f$ as required.

To prove uniqueness, let $u, v \in \ell^p_{a,0}(X)$ be minimizers of (A.4). Then u and v satisfy (A.5) which we write as

$$-\text{div}\,(|\nabla u|^{p-2}\nabla u) = \mu f.$$

Applying Lemma A.5 (below) we find that $\|u - v\|_{\ell^p(X)} = 0$ and so $u = v$. □

In the above proof we used a quantitive error estimate which is of interest in its own right. The estimate was on equations of the form

$$-\text{div}\,(|\nabla u|^{p-2}\nabla u) = f$$

when $f \in \ell^p_0(X)$, where we use the notation: if $a \in \mathbb{R}^n$ is a constant vector (without loss of generality the vector of ones) then we write $(u)_X = (u)_{a,X} = \frac{1}{n}\sum_{i=1}^{n} u(x_i)$ and $\ell^p_0(X) = \{u \in \ell^p(X) \,:\, (u)_X = 0\}$.

Lemma A.5. *Let $p > 1$, $a \in \mathbb{R}^n$ be non-negative, and assume $u, v \in \ell^p_{a,0}(X)$ satisfy*

$$-\text{div}\,(|\nabla u|^{p-2}\nabla u)(x_i) = f(x_i)$$

and

$$-\text{div}\,(|\nabla v|^{p-2}\nabla v)(x_i) = g(x_i)$$

for all $i = 1, \dots, n$, where $f, g \in \ell^p_0(X)$. Then,

$$\|u - v\|_{\ell^p(X)} \leq \begin{cases} C\lambda_p^{-q}\|f - g\|^{1/(p-1)}_{\ell^q(X)} & \text{if } p \geq 2 \\ C\lambda_p^{-2}\left(\|\nabla u\|_{\ell^p(X)} + \|\nabla v\|_{\ell^p(X)}\right)^{2-p}\|f - g\|_{\ell^2(X)} & \text{if } 1 < p < 2 \end{cases}$$

where C is a constant depending only on p and $q = \frac{p}{p-1}$.

Remark A.6. If $-\text{div}\,(|\nabla u|^{p-2}\nabla u) = f$ then we can write $\left(|\nabla u|^{p-2}\nabla u, \nabla\varphi\right)_{\ell^2(X^2)} = (f, \varphi)_{\ell^2(X)}$ for any $\varphi \in \ell^2(X)$. Choosing $\varphi = u$ implies $\|\nabla u\|^p_{\ell^p(X^2)} = (f, u)_{\ell^2(X)} \leq \|f\|_{\ell^q(X)}\|u\|_{\ell^p(X)}$ so we could write the bound for $p \in (1,2)$ in the above lemma without $\|\nabla u\|_{\ell^p(X)}$ and $\|\nabla v\|_{\ell^p(X)}$ on the right hand side.

Proof. For $p \geq 2$ we use the identity

$$|a - b|^p \leq C(|a|^{p-2}a - |b|^{p-2}b) \cdot (a - b)$$

for vectors $a, b \in \mathbb{R}^k$ for some constant C depending only on p (which can be found in Lemma 4.4 Chapter I (DiBenedetto, 1993)) to obtain

$$\begin{aligned}
\|\nabla u - \nabla v\|_{\ell^p(X^2)}^p &= \frac{1}{2}\sum_{i,j=1}^n w_{ij}|\nabla u(x_i,x_j) - \nabla v(x_i,x_j)|^p \\
&\leq C\sum_{i,j=1}^n w_{ij}\left(|\nabla u(x_i,x_j)|^{p-2}\nabla u(x_i,x_j) - |\nabla v(x_i,x_j)|^{p-2}\nabla v(x_i,x_j)\right)\cdot(\nabla u(x_i,x_j) - \nabla v(x_i,x_j)) \\
&= C(|\nabla u|^{p-2}\nabla u - |\nabla v|^{p-2}\nabla v, \nabla(u-v))_{\ell^2(X^2)} \\
&= C(-\text{div}\,(|\nabla u|^{p-2}\nabla u) + \text{div}\,(|\nabla v|^{p-2}\nabla v), u-v)_{\ell^2(X)} \\
&= C(f-g, u-v)_{\ell^2(X)} \\
&\leq C\|f-g\|_{\ell^q(X)}\|u-v\|_{\ell^p(X)},
\end{aligned}$$

where in the last line we used Hölder's inequality, $\frac{1}{p} + \frac{1}{q} = 1$, and the value of C may change from line-to-line. By Proposition A.4 we have

$$\lambda_p^p\|u-v\|_{\ell^p(X)}^p \leq \|\nabla u - \nabla v\|_{\ell^p(X^2)}^p \leq C\|f-g\|_{\ell^q(X)}\|u-v\|_{\ell^p(X)}.$$

Therefore we deduce

$$\|u-v\|_{\ell^p(X)} \leq C\lambda_p^{-q}\|f-g\|_{\ell^q(X)}^{1/(p-1)}.$$

Now for $1 < p < 2$ we follow the proof of Lemma 4.4 in Chapter I (DiBenedetto, 1993) to infer

$$\begin{aligned}
\left(|a|^{p-2}a - |b|^{p-2}b\right)\cdot(a-b) = &\int_0^1 |sa+(1-s)b|^{p-2}\,|a-b|^2\,\mathrm{d}s \\
&+ (p-2)\int_0^1 |sa+(1-s)b|^{p-4}\,|(sa+(1-s)b)\cdot(a-b)|^2\,\mathrm{d}s
\end{aligned}$$

for any $a, b \in \mathbb{R}^k$. Hence, by the Cauchy Schwarz inequality,

$$\begin{aligned}
\left(|a|^{p-2}a - |b|^{p-2}b\right)\cdot(a-b) &\geq (p-1)\int_0^1 |sa+(1-s)b|^{p-2}\,|a-b|^2\,\mathrm{d}s \\
&\geq (p-1)|a-b|^2\int_0^1 \frac{1}{(s|a|+(1-s)|b|)^{2-p}}\,\mathrm{d}s \\
&\geq \frac{(p-1)|a-b|^2}{(|a|+|b|)^{2-p}}.
\end{aligned}$$

In the sequel we make use of the inequality

$$\left(|a|^{p-2}a - |b|^{p-2}b\right)\cdot(a-b) \geq \frac{C|a-b|^2}{(|a|+|b|)^{2-p}}.$$

By Hölder's inequality and the above inequality we have (where again the constant C may change from line-to-line)

$$
\begin{aligned}
\|\nabla u-\nabla v\|_{\ell^p(X^2)}^p &= \frac{1}{2}\sum_{i,j=1}^n w_{ij}|\nabla u(x_i,x_j)-\nabla v(x_i,x_j)|^p \\
&\leq \left(\frac{1}{2}\sum_{i,j=1}^n \frac{w_{ij}|\nabla u(x_i,x_j)-\nabla v(x_i,x_j)|^2}{(|\nabla u(x_i,x_j)|+|\nabla v(x_i,x_j)|)^{2-p}}\right)^{\frac{p}{2}}\left(\frac{1}{2}\sum_{i,j=1}^n w_{ij}\left(|\nabla u(x_i,x_j)|+|\nabla v(x_i,x_j)|\right)^p\right)^{\frac{2-p}{2}} \\
&\leq C\left(\sum_{i,j=1}^n w_{ij}\big(|\nabla u(x_i,x_j)|^{p-2}\nabla u(x_i,x_j)-|\nabla v(x_i,x_j)|^{p-2}\nabla v(x_i,x_j)\big)\cdot(\nabla u(x_i,x_j)-\nabla v(x_i,x_j))\right)^{\frac{p}{2}} \\
&\qquad\times\left(\|\nabla u\|_{\ell^p(X^2)}+\|\nabla v\|_{\ell^p(X^2)}\right)^{\frac{(2-p)p}{2}} \\
&= C\left(|\nabla u|^{p-2}\nabla u-|\nabla v|^{p-2}\nabla v,\nabla(u-v)\right)_{\ell^2(X^2)}^{\frac{p}{2}}\left(\|\nabla u\|_{\ell^p(X^2)}+\|\nabla v\|_{\ell^p(X^2)}\right)^{\frac{(2-p)p}{2}} \\
&= C\left(-\mathrm{div}\left(|\nabla u|^{p-2}\nabla u\right)+\mathrm{div}\left(|\nabla v|^{p-2}\nabla v\right),u-v\right)_{\ell^2(X)}^{\frac{p}{2}}\left(\|\nabla u\|_{\ell^p(X^2)}+\|\nabla v\|_{\ell^p(X^2)}\right)^{\frac{(2-p)p}{2}} \\
&= C\left(f-g,u-v\right)_{\ell^2(X)}^{\frac{p}{2}}\left(\|\nabla u\|_{\ell^p(X^2)}+\|\nabla v\|_{\ell^p(X^2)}\right)^{\frac{(2-p)p}{2}} \\
&\leq C\,\|f-g\|_{\ell^2(X)}^{\frac{p}{2}}\,\|u-v\|_{\ell^2(X)}^{\frac{p}{2}}\left(\|\nabla u\|_{\ell^p(X^2)}+\|\nabla v\|_{\ell^p(X^2)}\right)^{\frac{(2-p)p}{2}}.
\end{aligned}
$$

Combining the above with Proposition A.4 we have

$$
\lambda_p^p\|u-v\|_{\ell^p(X)}^{\frac{p}{2}}\leq C\,\|f-g\|_{\ell^2(X)}^{\frac{p}{2}}\left(\|\nabla u\|_{\ell^p(X^2)}+\|\nabla v\|_{\ell^p(X^2)}\right)^{\frac{(2-p)p}{2}}
$$

which implies the result. □

The final proof from Section 2.2 is Lemma A.3.

Proof of Lemma A.3. Let us write

$$
I_{p,\mu}(u)=\frac{1}{p}\|\nabla u\|_{\ell^p(X^2)}^p-\mu\sum_{j=1}^m(y_j-\overline{y}_{\mathrm{u}})\cdot u(x_j).
$$

We note that

$$
I_{p,\mu}(\mu^{1/(p-1)}u)=\mu^{p/(p-1)}I_{p,1}(u).
$$

Therefore

$$
I_{p,\mu}(u_\mu)=\mu^{p/(p-1)}I_{p,1}(u_\mu\mu^{-1/(p-1)})\geq\mu^{p/(p-1)}I_{p,1}(u_1).
$$

On the other hand

$$
\mu^{p/(p-1)}I_{p,1}(u_1)=I_{p,\mu}(\mu^{1/(p-1)}u_1)\geq I_{p,\mu}(u_\mu)
$$

Therefore

$$
I_{p,\mu}(\mu^{1/(p-1)}u_1)=I_{p,\mu}(u_\mu).
$$

By uniqueness in Theorem A.1 we have $u_\mu=\mu^{1/(p-1)}u_1$, which completes the proof. □

A.3. Proofs for Section 2.4

We now turn our attention to the Ginzburg–Landau approximation of the graph cut problem (2.11).

Proof of Theorem 2.4. Let us redefine GL_τ in a more general form,

$$
\mathrm{GL}_\tau(u)=\frac{1}{2}\|\nabla u\|_{\ell^2(X^2)}^2+\frac{1}{\tau}\sum_{i=1}^n V(u(x_i))
$$

where $V : \mathbb{R}^k \to [0, +\infty)$ is continuous and $V(t) = 0$ if and only if $t \in S_k$. Of course, the choice of $V(t) = \prod_{j=1}^k |t - \mathbf{e}_j|^2$ satisfies these assumptions. We let

$$\mathcal{E}_\tau(u) = \begin{cases} \mathrm{GL}_\tau(u) - \mu \sum_{j=1}^m (y_j - \overline{y}_\mathrm{u}) \cdot u(x_j) & \text{if } (u)_X = b \\ +\infty & \text{else,} \end{cases}$$

$$\mathcal{E}_0(u) = \begin{cases} \frac{1}{2}\|\nabla u\|^2_{\ell^2(X^2)} - \mu \sum_{j=1}^m (y_j - \overline{y}_\mathrm{u}) \cdot u(x_j) & \text{if } (u)_X = b \text{ and } u : X \to S_k \\ +\infty & \text{else.} \end{cases}$$

The theorem can be restated as showing that minimisers u_τ of $\mathcal{E}_\tau$ contain convergent subsequences, and any convergent subsequence converges to a minimiser of $\mathcal{E}_0$. We divide the proof into two steps, in the first step we show that the sequence of minimisers $\{u_\tau\}_{\tau>0}$ is precompact, in the second step we show that any convergent subsequence is converging to a minimiser of $\mathcal{E}_0$.

1. Compactness. We first show that any sequence $\{u'_\tau\}_{\tau>0}$ and $M \in \mathbb{R}$ satisfying $\sup_{\tau>0} \mathcal{E}_\tau(u'_\tau) \le M$ is precompact. By Proposition A.4 and the Cauchy–Schwarz inequality

$$\begin{aligned} M &\ge \frac{\lambda_2^2}{2}\|u'_\tau - b\|^2_{\ell^2(X)} + \underbrace{\frac{1}{\tau}\sum_{i=1}^n V(u'_\tau(x_i))}_{\ge 0} - \mu \underbrace{\sqrt{\sum_{j=1}^m (y_j - \overline{y}_\mathrm{u})^2}}_{=:C} \|u'_\tau\|_{\ell^2(X)} \\ &\ge \frac{\lambda_2^2}{2}\|u'_\tau - b\|^2_{\ell^2(X)} - C\mu\|u'_\tau - b\|_{\ell^2(X)} - C\mu\|b\|_{\ell^2(X)}. \end{aligned}$$

Hence,

$$\|u'_\tau - b\|_{\ell^2(X)} \le \frac{C\mu}{\lambda_2^2}\left(1 + \sqrt{1 + \frac{2\lambda_2^2(M + C\mu\|b\|_{\ell^2(X)})}{C^2\mu^2}}\right) =: \widetilde{C}$$

so $\{\mu'_\tau\}_{\tau>0}$ is bounded in $\ell^2(X)$ and therefore, by the Bolzano–Weierstrass Theorem, precompact.

To show that minimisers $\{u_\tau\}_{\tau>0}$ are precompact it is enough to show that there exists $M \in \mathbb{R}$ such that $\sup_{\tau>0} \mathcal{E}_\tau(u_\tau) \le M$. This follows easily as we take $u \in \ell^2(X)$ satisfying $\sum_{i=1}^n u(x_i) = b$ and $u(x_i) \in S_k$ for all $i = 1, 2, \ldots, n$ as a candidate. We have

$$\mathcal{E}_\tau(u_\tau) \le \mathcal{E}_\tau(u) = \frac{1}{2}\|\nabla u\|^2_{\ell^2(X^2)} - \mu\sum_{j=1}^m (y_j - \overline{y}_\mathrm{u}) \cdot u(x_j) =: M.$$

Now we have shown that there exists convergent subsequences we show that any limit must be a minimiser of $\mathcal{E}_0$.

2. Converging Subsequences. Let u_0 be a cluster point of $\{u_\tau\}_{\tau>0}$, i.e. there exists a subsequence such that $u_{\tau_m} \to u_0$ as $m \to \infty$. Pick any $v \in \ell^2(X)$ with $\mathcal{E}_0(v) < +\infty$. We will show

(a) $\mathcal{E}_\tau(v) = \mathcal{E}_0(v)$,

(b) $\liminf_{\tau\to 0} \mathcal{E}_\tau(u_\tau) \ge \mathcal{E}_0(u_0)$.

Assuming (a) and (b) hold then, by (a),

$$\mathcal{E}_0(v) = \mathcal{E}_{\tau_m}(v) \ge \mathcal{E}_{\tau_m}(u_{\tau_m}).$$

Taking the limit as $m \to \infty$, and applying (b) we have

$$\mathcal{E}_0(v) \ge \liminf_{m\to\infty} \mathcal{E}_{\tau_m}(u_{\tau_m}) \ge \mathcal{E}_0(u_0).$$

It follows that for all $v \in \ell^2(X)$ we have $\mathcal{E}_0(u_0) \le \mathcal{E}_0(v)$, hence u_0 is a minimiser of $\mathcal{E}_0$.

To show (a), we easily notice that

$$\mathcal{E}_\tau(v) = \frac{1}{2}\|\nabla v\|^2_{\ell^2(X^2)} + \frac{1}{\tau}\sum_{i=1}^n \underbrace{V(v(x_i))}_{=0} - \mu\sum_{j=1}^m (y_j - \overline{y}_\mathrm{u}) \cdot v(x_j) = \mathcal{E}_0(v).$$

For (b) we without loss of generality assume that $u_\tau \to u_0$ and

$$\liminf_{\tau\to 0} \mathcal{E}_\tau(u_\tau) = \lim_{\tau \to 0} \mathcal{E}_\tau(u_\tau) < +\infty.$$

As $\sum_{i=1}^n u_\tau(x_i) = b$ for all $\tau > 0$ and $u_\tau(x_i) \to u_0(x_i)$ for every $i \in \{1,\dots,n\}$ then $(u_0)_X = \sum_{i=1}^n u_0(x_i) = b$. And since $V(u_\tau(x_i)) \leq \tau \mathcal{E}_\tau(u_\tau) \to 0$ then we have $V(u_0(x_i)) = 0$, hence $u_0(x_i) \in S_k$. Now,

$$\mathcal{E}_\tau(u_\tau) = \underbrace{\frac{1}{2}\|\nabla u_\tau\|^2_{\ell^2(X^2)}}_{\to \frac{1}{2}\|\nabla u_0\|^2_{\ell^2(X^2)}} + \frac{1}{\tau}\sum_{i=1}^n \underbrace{V(u_\tau(x_i))}_{\geq 0} - \underbrace{\mu \sum_{j=1}^m (y_j - \overline{y}_{\mathrm{u}}) \cdot u_\tau(x_j)}_{\to \sum_{j=1}^m (y_j - \overline{y}_{\mathrm{u}})\cdot u_0(x_j)}.$$

So $\liminf_{\tau\to 0} \mathcal{E}_\tau(u_\tau) \geq \mathcal{E}_0(u_0)$ as required. □

Remark A.7. If (a) and (b) in the proof of Theorem 2.4 are strengthened to

(a′) for all $v \in \ell^2(X)$ there exists $v_\tau \to v$ such that $\lim_{\tau\to 0}\mathcal{E}_\tau(v_\tau) = \mathcal{E}_0(v)$,

(b′) for all $v \in \ell^2(X)$ and for all $v_\tau \to v$ then $\liminf_{\tau \to 0} \mathcal{E}_\tau(v_\tau) \geq \mathcal{E}_0(v)$

then one says that $\mathcal{E}_\tau$ Γ-converges to $\mathcal{E}_0$ (and one can show that (a′) and (b′) hold in our case with a small modification of the above proof). The notion of Γ-convergence is fundamental in the calculus of variations and is considered the variational form of convergence as it implies (when combined with a compactness result) the convergence of minimisers.

B. Continuum limits

We briefly discuss continuum limits for the Poisson learing problem (2.3). We take $p = 2$ for simplicity, but the arguments extend similarly to other values of $p \geq 1$. In order to analyze continuum limits of graph-based learning algorithms, we make the *manifold assumption*, and assume G is a random geometric graph sampled from an underlying manifold. To be precise, we assume the vertices of the graph corresponding to unlabeled points $x_1, \dots, x_n$ are a sequence of *i.i.d.* random variables drawn from a d-dimensional compact, closed, and connected manifold $\mathcal{M}$ embedded in $\mathbb{R}^D$, where $d \ll D$. We assume the probability distribution of the random variables has the form $\mathrm{d}\mu = \rho \mathrm{dVol}_{\mathcal{M}}$, where $\mathrm{Vol}_{\mathcal{M}}$ is the volume form on the manifold, and ρ is a smooth density. For the labeled vertices in the graph, we take a fixed finite set of points $\Gamma \subset \mathcal{M}$. The vertices of the random geometric graph are

$$X_n := \{x_1, \dots, x_n\} \cup \Gamma.$$

We define the edge weights in the graph by

$$w_{xy} = \eta_\varepsilon\left(|x-y|\right),$$

where $\varepsilon > 0$ is the length scale on which we connect neighbors, $|x-y|$ is Euclidean distance in $\mathbb{R}^D$, and $\eta : [0,\infty) \to [0,\infty)$ is smooth with compact support, and $\eta_\varepsilon(t) = \frac{1}{\varepsilon^d}\eta\left(\frac{t}{\varepsilon}\right)$. We denote the solution of the Poisson learning problem (2.3) for this random geometric graph by $u_{n,\varepsilon}(x)$.

The normalized graph Laplacian is given by

$$\mathcal{L}_{n,\varepsilon}u(x) = \frac{2}{\sigma_\eta n \varepsilon^2}\sum_{y \in X_n} \eta_\varepsilon(|x-y|)(u(x) - u(y)),$$

where $\sigma_\eta = \int_{\mathbb{R}^d} |z_1|^2 \eta(|z|)\, dz$. It is well-known (see, e.g., (Hein et al., 2007)), that $\mathcal{L}_{n,\varepsilon}$ is consistent with the (negative of) the weighted Laplace-Beltrami operator

$$\Delta_\rho := -\rho^{-1}\mathrm{div}_{\mathcal{M}}(\rho^2 \nabla_{\mathcal{M}} u),$$

where $\mathrm{div}_{\mathcal{M}}$ is the manifold divergence and $\nabla_{\mathcal{M}}$ is the manifold gradient. We write $\mathrm{div} = \mathrm{div}_{\mathcal{M}}$ and $\nabla = \nabla_{\mathcal{M}}$ now for convenience. In particular, for any $u \in C^3(\mathcal{M})$ we have

$$|\mathcal{L}_{n,\varepsilon}u(x) - \Delta_\rho u(x)| \leq C(\|u\|_{C^3(\mathcal{M})} + 1)(\lambda + \varepsilon)$$

holds for all $x \in X_n$ with probability at least $1 - Cn\exp\left(-cn\varepsilon^{d+2}\lambda^2\right)$ for any $0 < \lambda \leq 1$, where $C, c > 0$ are constants.

Using the normalised graph Laplacian in the Poisson learning problem (2.3) we write

$$\mathcal{L}_{n,\varepsilon} u_{n,\varepsilon}(x) = n \sum_{y\in\Gamma}^{m} (g(y) - \overline{y}_{\mathrm{u}})\delta_{x=y} \quad \text{for } x \in X_n, \tag{B.1}$$

where $g(y) \in \mathbb{R}$ denotes the label associated to $y \in \Gamma$ and $\overline{y}_{\mathrm{u}} = \frac{1}{|\Gamma|}\sum_{x\in\Gamma} g(x)$. We restrict to the scalar case (binary classification) for now. Note that the normalisation plays no role in the classification problem (2.2). To see what should happen in the continuum, as $n \to \infty$ and $\varepsilon \to 0$, we multiply both sides of (B.1) by a smooth test function $\varphi \in C^\infty(\mathcal{M})$, sum over $x \in X$, and divide by n to obtain

$$\frac{1}{n}(\mathcal{L}_{n,\varepsilon} u_{n,\varepsilon}, \varphi)_{\ell^2(X)} = \sum_{y\in\Gamma} (g(y) - \overline{y}_{\mathrm{u}})\varphi(y). \tag{B.2}$$

Since $\mathcal{L}_{n,\varepsilon}$ is self-adjoint (symmetric), we have

$$(\mathcal{L}_{n,\varepsilon} u_{n,\varepsilon}, \varphi)_{\ell^2(X)} = (u_{n,\varepsilon}, \mathcal{L}_{n,\varepsilon}\varphi)_{\ell^2(X)} = (u_{n,\varepsilon}, \Delta_\rho \varphi)_{\ell^2(X)} + O\left((\lambda + \varepsilon)\|u_{n,\varepsilon}\|_{\ell^1(X)}\right).$$

We also note that

$$\sum_{y\in\Gamma} (g(y) - \overline{y}_{\mathrm{u}})\varphi(y) = \int_{\mathcal{M}} \sum_{y\in\Gamma} (g(y) - \overline{y}_{\mathrm{u}})\delta_y(x)\varphi(x)\, \mathrm{dVol}_{\mathcal{M}}(x),$$

where δ_y is Dirac-Delta distribution centered at $y \in \mathcal{M}$, which has the property that

$$\int_{\mathcal{M}} \delta_y(x)\varphi(x)\, \mathrm{dVol}_{\mathcal{M}}(x) = \varphi(y)$$

for every smooth $\varphi \in C^\infty(\mathcal{M})$. Combining these observations with (B.2) we see that

$$\frac{1}{n}(u_{n,\varepsilon}, \Delta_\rho \varphi)_{\ell^2(X)} + O\left(\frac{(\lambda+\varepsilon)}{n}\|u_{n,\varepsilon}\|_{\ell^1(X)}\right) = \int_{\mathcal{M}} \sum_{y\in\Gamma} (g(y) - \overline{y}_{\mathrm{u}})\delta_y(x)\varphi(x)\, \mathrm{dVol}_{\mathcal{M}}(x).$$

If we can extend $u_{n,\varepsilon}$ to a function on $\mathcal{M}$ in a suitable way, then the law of large numbers would yield

$$\frac{1}{n}(u_{n,\varepsilon}, \Delta_\rho \varphi)_{\ell^2(X)} \approx \int_{\mathcal{M}} u_{n,\varepsilon}(x)\rho(x)\Delta_\rho\varphi(x)\, \mathrm{dVol}_{\mathcal{M}}(x).$$

Hence, if $u_{n,\varepsilon} \to u$ as $n \to \infty$ and $\varepsilon \to 0$ in a sufficiently strong sense, then the function $u : \mathcal{M} \to \mathbb{R}$ would satisfy

$$-\int_{\mathcal{M}} u \,\mathrm{div}\,(\rho^2 \nabla\varphi)\, \mathrm{dVol}_{\mathcal{M}} = \int_{\mathcal{M}} \sum_{y\in\Gamma} (g(y) - \overline{y}_{\mathrm{u}})\delta_y(x)\varphi(x)\, \mathrm{dVol}_{\mathcal{M}}(x)$$

for every smooth $\varphi \in C^\infty(\mathcal{M})$. If $u \in C^2(\mathcal{M})$, then we can integrate by parts on the left hand side to find that

$$-\int_{\mathcal{M}} \varphi \,\mathrm{div}\,(\rho^2 \nabla u)\, \mathrm{dVol}_{\mathcal{M}} = \int_{\mathcal{M}} \sum_{y\in\Gamma} (g(y) - \overline{y}_{\mathrm{u}})\delta_y(x)\varphi(x)\, \mathrm{dVol}_{\mathcal{M}}(x)$$

Since φ is arbitrary, this would show that u is the solution of the Poisson problem

$$-\mathrm{div}\,\left(\rho^2 \nabla u\right) = \sum_{y\in\Gamma} (g(y) - \overline{y}_{\mathrm{u}})\delta_y \quad \text{on } \mathcal{M}. \tag{B.3}$$

We conjecture that the solutions $u_{n,\varepsilon}$ converge to the solution of (B.3) as $n \to \infty$ and $\varepsilon \to 0$ with probability one.

Conjecture B.1. Assume ρ is smooth. Assume that $n \to \infty$ and $\varepsilon = \varepsilon_n \to 0$ so that

$$\lim_{n\to\infty} \frac{n\varepsilon^{d+2}}{\log n} = \infty.$$

Let $u \in C^\infty(\mathcal{M} \setminus \Gamma)$ be the solution of the Poisson equation (B.3) and $u_{n,\varepsilon}$ solve the graph Poisson problem (B.1). Then with probability one

$$\lim_{n\to\infty} \max_{\substack{x\in X_n \\ \mathrm{dist}(x,\Gamma)>\delta}} |u_{n,\varepsilon}(x) - u(x)| = 0$$

for all $\delta > 0$.

The conjecture states that $u_{n,\varepsilon}$ converges to u uniformly as long as one stays a positive distance away from the source points Γ, where the solution u is singular. We expect the conjecture to be true, since similar results are known to hold when the source term on the right hand side is a smooth function f. The fact that the right hand side in (B.3) is highly singular, involving delta-mass concentration, raises difficult technical problems that will require new insights that are far beyond the scope of this paper.

Remark B.2. If Conjecture B.1 is true, it shows that Poisson learning is consistent with a well-posed continuum PDE for arbitrarily low label rates. This is in stark contrast to Laplace learning, which does not have a well-posed continuum limit unless the number of labels grows to ∞ as $n \to \infty$ sufficiently fast. This partially explains the superior performance of Poisson learning for low label rate problems.

References

DiBenedetto, E. *Degenerate parabolic equations*. Springer Science & Business Media, 1993.

Hein, M., Audibert, J.-Y., and Luxburg, U. v. Graph Laplacians and their convergence on random neighborhood graphs. *Journal of Machine Learning Research*, 8(Jun):1325–1368, 2007.