Notes 17: Graph sparsification

1. Graph sparsification

Problem 1.1. Given an undirected, connected graph $G = (V, E_G, w_G)$ with positive edge weights $w_G : E_G \to \mathbb{R}_+$, find a sparse subgraph $H = (V, E_H, w_H)$ (with possibly different weights w_H) that approximates G, so that they have similar cut value across every cut.

In fact, we will solve this problem with a stronger guarantee: H will spectrally approximate G, not just have similar cut values.

Definition 1.2. Suppose G and H are graphs on the same set of vertices. H ε-approximates G if

$$(1 - \varepsilon)L_G \preccurlyeq L_H \preccurlyeq (1 + \varepsilon)L_G \; . \tag{1}$$

If G is the complete graph on n vertices with self-loops, then graphs H that approximates G are exactly expanders in Notes14.

If H approximates G in this spectral sense, then H and G must have similar values across every cut. Recall that quadratic forms of the Laplacian are closely related to cuts. For any subset $S \subseteq V$, the total weight of edges across the cut is given by

$$w_G(S, \overline{S}) = \mathbb{1}_S{}^\top L_G \mathbb{1}_S \; .$$

Therefore, if H ε-approximates G, then simultaneously for any $S \subseteq V$,

$$(1 - \varepsilon)\mathbb{1}_S{}^\top L_G \mathbb{1}_S \leqslant \mathbb{1}_S{}^\top L_H \mathbb{1}_S \leqslant (1 + \varepsilon)\mathbb{1}_S{}^\top L_G \mathbb{1}_S \; ,$$

that is,

$$(1 - \varepsilon)w_G(S, \overline{S}) \leqslant w_H(S, \overline{S}) \leqslant (1 + \varepsilon)w_G(S, \overline{S}) \; .$$

Our sparse graph H will contain only edges from G, so $E_H \subseteq E_G$. But these edges can have new edge weights w_H suitably rescaled from the original weights w_G.

2. Isotropic position

Suppose H ε-approximates G, so their Laplacians are close as in Eq. (1). If we "divide Eq. (1) by L_G", or rather, left and right multiply every term by $L_G^{+/2}$, we get

$$(1 - \varepsilon)\Pi \preccurlyeq L_G^{+/2} L_H L_G^{+/2} \preccurlyeq (1 + \varepsilon)\Pi \; , \tag{2}$$

where $\Pi = L_G^{+/2} L_G L_G^{+/2}$ is the orthogonal projection to the span of L_G. The normalized condition Eq. (2) is equivalent to the original one Eq. (1) since L_H and L_G share the same nullspace (spanned by $\mathbb{1}$).

Under this normalization, L_G can be seen as the second moment matrix of some vectors in isotropic position.

Definition 2.1. A set of vectors $\{u_e\}_{e \in E}$ in a vector space U are in isotropic position if its second moment matrix is the identity matrix in U:

$$\sum_{e \in E} u_e u_e^\top = I \; .$$

This condition means the second moment is the same in every direction:

$$x^\top \left(\sum_{e \in E} u_e u_e^\top \right) x = x^\top x = \|x\|^2 \qquad \text{for every } x \in U \text{, independent of the direction of } x.$$

If $\{u_e\}_{e \in E}$ represents high dimensional data with mean 0, then a set of data in isotropic position has covariance being the identity matrix, so the projected covariance in every direction is the same.

How does

$$L_G = \sum_{(a,b) \in E} w_e (\mathbb{1}_a - \mathbb{1}_b)(\mathbb{1}_a - \mathbb{1}_b)^\top$$

represent vectors in isotropic position? If we set $v_e = \sqrt{w_e}(\mathbb{1}_a - \mathbb{1}_b)$ for edge $e = (a, b)$, and $u_e = L_G^{+/2} v_e$, then

$$\sum_{e\in E} u_e u_e^\top = L_G^{+/2}\left(\sum_{e\in E} v_e v_e^\top\right) L_G^{+/2} = L_G^{+/2} L_G L_G^{+/2} = \Pi ,$$

which is essentially the identity operator on the subspace U orthogonal to $\mathbb{1}$. Π also zeros out vector parallel to $\mathbb{1}$. If we regard u_e as vectors in U (an $(n-1)$-dimensional vector space), not just vectors in $\mathbb{R}^V$ (an n-dimensional vector space containing U), then $\{u_e\}_{e\in E}$ are in isotropic position.

The original problem of finding sparse subgraph H to approximate G now reduces to the following problem:

Problem 2.2 (Isotropic sampling)**.** Given a set vectors $\{u_e\}_{e\in E}$ in isotropic position, obtain a new collection $\{\tilde{u}_{e'}\}_{e'\in E'}$ of vectors, so that every new vector $u_{e'}$ is a rescaled vector u_e in the original collection:

$$\text{for every } e' \in E', \text{ there is } \alpha_{e'} > 0, e \in E, \text{ such that} \quad \tilde{u}_{e'} = \alpha_{e'} u_e .$$

We want $|E'|$ to be as small as possible, and

$$(1-\varepsilon)I \preccurlyeq \sum_{e\in E'} \tilde{u}_e \tilde{u}_e^\top \preccurlyeq (1+\varepsilon)I ,$$

i.e. the new collection $\{\tilde{u}_e\}_{e\in E'}$ is ε-close to be in isotropic position.

3. Sampling by squared norm

Here is an algorithm for the isotropic sampling problem given vectors $\{u_e\}_{e\in E}$ in a d-dimensional vector space U.

Sampling by squared norm

Let $\quad Z = \sum_{e\in E} \|u_e\|^2 \quad$ and $\quad T = 4(d \log d)/\varepsilon^2$
For $e' = 1, \ldots, T$
  Choose $e \in E$ with probability $p_e = \|u_e\|^2/Z$
  Add $\tilde{u}_{e'} = u_e/\sqrt{Tp_e}$ to the output collection

In other words, we sample $u_{e'}$ independently with repetition as some vector u_e scaled. Any u_e is chosen with probability proportional to its squared norm $\|u_e\|^2$. If u_e is chosen, we scale it down by the factor $\sqrt{Tp_e}$.

Why scale factor $1/\sqrt{Tp_e}$? So that the second moment matrix of $\{\tilde{u}_{e'}\}_{e'\in E}$ has the correct expectation. Note that since $\tilde{u}_{e'}$ are random, their second moment matrix $\sum_{e'\in E'} \tilde{u}_{e'} \tilde{u}_{e'}^\top$ is a random matrix. We will study the expectation of this random matrix, and its deviation from expectation. For any fixed e', when sampling the e'-th vector $\tilde{u}_{e'}$,

$$\mathop{\mathbb{E}}_{\tilde{u}_{e'}}\left[\tilde{u}_{e'} \tilde{u}_{e'}^\top\right] = \sum_{e\in E} p_e \left(\frac{u_e}{\sqrt{Tp_e}}\right)\left(\frac{u_e}{\sqrt{Tp_e}}\right)^\top = \frac{1}{T}\sum_{e\in E} u_e u_e^\top = \frac{I}{T} ,$$

and the second moment matrix of all T vectors has expectation

$$\mathop{\mathbb{E}}_{\{\tilde{u}_{e'}\}}\left[\sum_{e'\in E'} \tilde{u}_{e'} \tilde{u}_{e'}^\top\right] = T \mathop{\mathbb{E}}_{\tilde{u}_{e'}}\left[\tilde{u}_{e'} \tilde{u}_{e'}^\top\right] = T\frac{I}{T} = I .$$

4. Matrix Chernoff bounds

We will need to show that a sum of independent random second moment matrices is close to its expectation with high probability. This was proved by Tropp in *"User-Friendly Tail Bounds for Sums of Random Matrices"* (Corollary 5.2 there).

Theorem 4.1 (Tropp)**.** *Let $X_1, \ldots, X_m$ be independent random d-dimensional symmetric positive semidefinite matrices so that $\|X_i\| \leqslant R$ almost surely. Let $X = \sum_{1\leqslant i\leqslant m} X_i$ and $\mu_{\min}$ and $\mu_{\max}$ be the smallest and largest eigenvalues of*

$$\mathbb{E}[X] = \sum_{1\leqslant i\leqslant m} \mathbb{E}[X_i] .$$

Then

$$\mathbb{P}\left[\lambda_{\min}(X) \leqslant (1-\varepsilon)\mu_{\min}\right] \leqslant d\exp(-\varepsilon^2\mu_{\min}/2R) \qquad \textit{for } 0<\varepsilon<1 \ ,$$
$$\mathbb{P}\left[\lambda_{\max}(X) \geqslant (1+\varepsilon)\mu_{\max}\right] \leqslant d\exp(-\varepsilon^2\mu_{\max}/3R) \qquad \textit{for } 0<\varepsilon<1 \ .$$

5. Concentration

We will apply Matrix Chernoff with $X_{e'} = \tilde{u}_{e'}\tilde{u}_{e'}^\top$.

We choose $e \in E$ with probability proportional to $\|u_e\|^2$ in order to minimize the norm of $X_{e'}$:

$$\|X_{e'}\| \leqslant \max_{e\in E}\left\|\left(\frac{u_e}{\sqrt{Tp_e}}\right)\left(\frac{u_e}{\sqrt{Tp_e}}\right)^\top\right\| = \max_{e\in E}\left\|\frac{u_e}{\sqrt{Tp_e}}\right\|^2 = \max_{e\in E}\frac{\|u_e\|^2}{Tp_e} = \frac{Z}{T} \ .$$

The point is that, in the last equality, every term inside the maximum is Z/T, independent of $e\in E$. This ensures the best possible bound $R = Z/T$ for the norm of $X_{e'}$. That's why sampling probabilities p_e are proportional to $\|u_e\|^2$.

In fact, the normalization constant Z is simply $d = \dim U$. Indeed,

$$Z = \sum_{e\in E} u_e^\top u_e = \sum_{e\in E}\operatorname{Tr}\left(u_e u_e^\top\right) = \operatorname{Tr}\left(\sum_{e\in E} u_e u_e^\top\right) = \operatorname{Tr}(I) = \dim U \ .$$

By Matrix Chernoff with $X = \sum_{e'\in E'} X_{e'} = \sum_{e'\in E'}\tilde{u}_{e'}\tilde{u}_{e'}^\top$,

$$\mathbb{P}[X \not\preccurlyeq (1+\varepsilon)I] \leqslant d\exp(-\varepsilon^2/3R) = d\exp(-(4/3)\log d) = d^{-1/3} \ .$$
$$\mathbb{P}[X \not\succcurlyeq (1-\varepsilon)I] \leqslant d\exp(-\varepsilon^2/2R) = d\exp(-(4/2)\log d) = d^{-1} \ .$$

Therefore, with overwhelming probability for large d, the second moment matrix is ε-close to the identity, so the output vectors $\{\tilde{u}_{e'}\}_{e'\in E'}$ are ε-close to be in isotropic position. This completes the analysis of the sampling algorithm.

6. Variants

The above sampling algorithm outputs a collection with $O(d(\log d)/\varepsilon^2)$ vectors. The $\Omega(d\log d)$ dependence on d is unavoidable for randomized algorithms with independent samples: A special case is the input $\{u_e\}_{e\in E}$ consists of standard basis vectors. In this case Coupon collector tells us $\Omega(d\log d)$ samples are required to see all vectors.

Batson–Spielman–Srivastava came up with a deterministic algorithm (without random sampling) to solve the isotropic sampling problem that outputs a collection with $O(d/\varepsilon^2)$ vectors.

7. Effective resistance

Back to our original question of graph sparsification. The resulting algorithm (proposed by Spielman–Srivastava) gives us a subgraph H with $O(n(\log n)/\varepsilon^2)$ edges that ε-approximate given any graph G. H is very sparse even when G is dense.

What is the sampling probability p_e for edge $e=(a,b)$? It is proportional to $\|u_e\|^2$, where $u_e = L^{+/2}\sqrt{w_e}(\mathbb{1}_a - \mathbb{1}_b)$. Therefore

$$\|u_{(a,b)}\|^2 = w_{a,b}\|L^{+/2}(\mathbb{1}_a-\mathbb{1}_b)\|^2 = w_{a,b}R_{\text{eff}}(a,b) \ .$$

If input graph G is unweighted, then we are sampling an edge with probability proportional to the effective resistance between its endpoints.

We previously showed that $Z = \sum_{e\in E}\|u_e\|^2 = \dim U$. In the context of graphs,

$$\sum_{(a,b)\in E} w_{a,b}R_{\text{eff}}(a,b) = n-1 \ .$$

This result has a combinatorial meaning: One can consider sampling a random spanning tree of G, with probability proportional to the product of edge weights in the tree. Turns out $w_{a,b}R_{\text{eff}}(a,b)$ is exactly the probability that an edge (a,b) appears in this random spanning tree. And above calculations say that the expected number of edges in the random spanning tree is $n-1$.