UBC CPSC 536N: Sparse Approximations	Winter 2013

Lecture 19 — March 18, 2013

Prof. Nick Harvey — *Scribe: Samira Samadi*

1 Sparsifiers

Given an undirected graph $G = (V, E)$, we want to find sparse subgraph H of G with few edges and weights on edges of H such that H is approximately equal to (sparse copy of) G. In general, G could also be weighted, but for notational convenience we focus on the unweighted case. Let $n = |V|$ and $m = |E|$.

There are two notions for sparsification:

1. Cut sparsifiers:

$$(1-\epsilon)|\delta_G(U)| \leq |\delta_H(U)| \leq (1+\epsilon)|\delta_G(U)| \quad \forall U \in V$$
$$\iff$$
$$(1-\epsilon)x^\mathsf{T} L_G x \leq x^\mathsf{T} L_H x \leq (1+\epsilon)x^\mathsf{T} L_G x \quad \forall x \in \{0,1\}^n$$

2. Spectral sparsifiers:

$$(1-\epsilon)L_G \preccurlyeq L_H \preccurlyeq (1+\epsilon)L_G$$
$$\iff$$
$$(1-\epsilon)x^\mathsf{T} L_G x \leq x^\mathsf{T} L_H x \leq (1+\epsilon)x^\mathsf{T} L_G x \quad \forall x \in \mathbb{R}^n$$

1.1 History

Here is an abbreviated history of some graph sparsification results.

Paper	Type	Sparsity (# edges)	Running time
Benczur-Karger 96'	Cut sparsifier	$O(\frac{n}{\epsilon^2} \cdot \log n)$	$O(\frac{m}{\epsilon^2} \cdot \log^3 n)$, Randomized
Spielman-Srivastava 08'	Spectral sparsifirer	$O(\frac{n}{\epsilon^2} \cdot \log n)$	$O(\frac{m}{\epsilon^2} \cdot \log^{50} n)$, Randomized
Batson-Spielman-Srivastava	Spectral sparsifier	$O(\frac{n}{\epsilon^2})$	$O(\frac{n^3}{\epsilon^2} m)$, Deterministic

We are going to explain the algorithm that gets the last result.

2 Algorithm

2.1 Reduction

As usual, we apply the reduction (Lemma 6) described in Lecture 14. This yields vectors $\{ w_e \,:\, e \in E\} \subset \mathbb{R}^{n-1}$ with $\sum_{e \in E} w_e w_e^\mathsf{T} = I$ and satisfying the key condition that

$$l \cdot I \preccurlyeq \sum_{e \in E} s_e w_e w_e^T \preccurlyeq u \cdot I \qquad \iff \qquad l \cdot L_G \preccurlyeq \underbrace{\sum_{e \in E} s_e y_e y_e^T}_{L_H} \preccurlyeq u \cdot L_G \qquad \forall s \in \mathbb{R}^E_{\geq 0}. \quad (2.1)$$

For the remainder of the lecture we are going to work in this setup so let $p := n - 1$.

Theorem 2.1. Fix $\epsilon \in (0, \frac{1}{4}]$ and $w_1, \cdots, w_m \in \mathbb{R}^p$ with $\sum_{i=1}^m w_i w_i^\mathsf{T} = I$. There exists $F \subseteq [m]$ and $s : F \to \mathbb{R}_{\geq 0}$ such that

$$(1-\epsilon)I \preccurlyeq \sum_{i \in F} s_i \, w_i w_i^\mathsf{T} \preccurlyeq (1+5\epsilon)I \tag{2.2}$$

and $|F| \leq \frac{p}{\epsilon^2}$.

Note that (2.2) is the same as

$$\begin{cases} \lambda_{max}(\sum_{i \in F} s_i \, w_i w_i^\mathsf{T}) \leq 1 + 5\epsilon \\ \lambda_{min}(\sum_{i \in F} s_i \, w_i w_i^\mathsf{T}) \geq 1 - \epsilon \end{cases}$$

Once we've got this result about these vectors, with reduction we apply it to the Laplacian of the graph and we get the desired result.

2.2 Algorithm

- Parameters:

	lower bound	upper bound
Starting point	$l_0 := -\epsilon$	$u_0 := \epsilon$
Increase per iteration	$\delta_L := \dfrac{\epsilon^2}{p}$	$\delta_U := (1+4\epsilon)\,\dfrac{\epsilon^2}{p}$
Bound on potential	$\epsilon_L := \dfrac{p}{\epsilon}$	$\epsilon_U := \dfrac{p}{\epsilon}$

- Number of iterations: $T := p/\epsilon^2$

- Potential functions:

  $$\Phi_l(A) = \mathrm{tr}[(A - lI)^{-1}] = \textstyle\sum_{k=1}^p (\lambda_k - l)^{-1}$$
  $$\Phi^u(A) = \mathrm{tr}[(uI - A)^{-1}] = \textstyle\sum_{k=1}^m (u - \lambda_k)^{-1}$$

  where $\{\lambda_k\}$ are eigenvalues of A.

- Variables: $A \leftarrow 0$, $F \leftarrow \emptyset$, $s \leftarrow 0$, $l \leftarrow l_0$, $u \leftarrow u_0$

- **Pseudocode:**

  For $j = 1, \cdots, T$ {

  - Invariants:

    $$\begin{array}{llll} (a) & \lambda_{min}(A) > l, & (b) & \Phi_l(A) \leq \epsilon_L \\ (c) & \lambda_{max}(A) < u, & (d) & \Phi^u(A) \leq \epsilon_U \end{array}$$

  - Say index $i \in [m]$ and scalar $t > 0$ are good if

    $$\begin{array}{lll} (a) & \lambda_{min}(A + t\, w_i w_i^\mathsf{T}) > l + \delta_L, & (b)\Phi_{l+\delta_L}(A + t\, w_i w_i^\mathsf{T}) \leq \Phi_l(A) \\ (c) & \lambda_{max}(A + t\, w_i w_i^\mathsf{T}) < u + \delta_U, & (d)\Phi^{u+\delta_U}(A + t\, w_i w_i^\mathsf{T}) \leq \Phi^u(A) \end{array}$$

  - Find good i and t.

- $F \leftarrow F \cup \{i\}$, $s(i) \leftarrow s(i) + t$, $A \leftarrow A + t\, w_i w_i^\mathsf{T}$, $l \leftarrow l + \delta_L$, $u \leftarrow u + \delta_U$

}

End.

Initially its obvious that invariants (a) and (c) are satisfied. Also $\Phi_{l_0}(0) = \mathrm{tr}((-l_0\, I)^{-1}) = \mathrm{tr}(\frac{1}{\epsilon} I) = \frac{p}{\epsilon} = \epsilon_L$ so (b) is true. (d) is similar.

At the end of the algorithm:

$$l = l_0 + T\,\delta_L = -\epsilon + (\frac{p}{\epsilon^2})\,(\frac{\epsilon^2}{p}) = 1 - \epsilon \quad \Rightarrow \quad \lambda_{min}(A) = \lambda_{min}(\sum_{i \in F} s_i\, w_i w_i^\mathsf{T}) \geq 1 - \epsilon$$

$$u = u_0 + T\delta_U = \epsilon + \frac{p}{\epsilon^2}(1 + 4\epsilon)\frac{\epsilon^2}{p} = 1 + 5\epsilon \quad \Rightarrow \quad \lambda_{max}(A) = \lambda_{max}(\sum_{i \in F} s_i\, w_i w_i^\mathsf{T}) \leq 1 + 5\epsilon$$

2.3 Why does there exist good index i and scalar t?

From Lecture 16, assume $\lambda_{max}(A) < u$, $M_u = ((u+\delta u)I - A)^{-1}$, $N_U = \dfrac{M^2 u}{\Phi^u(A) - \Phi^{u+\delta}(A)} + M_u$

Lemma 2.2. If $v^\mathsf{T} N_U v \leq \frac{1}{t}$ then $\Phi^{u+\delta}(A + t\, vv^\mathsf{T}) \leq \phi^u(A)$ and $\lambda_{max}(A + t\, vv^\mathsf{T}) < u + \delta u$

Lemma 2.3. $\mathrm{tr}(N_U) \leq \frac{1}{\delta_U} + \Phi^u(A)$

From Lecture 18, assume $\lambda_{min}(A) > l$, $M_L = (A - (l + \delta_L)I)^{-1}$, $N_L = \dfrac{{M_L}^2}{\Phi_{l+\delta}(A) - \Phi_l(A)} - M_L$ and $\Phi_l(A) \leq \frac{1}{\delta_L} \Rightarrow \lambda_{min}(A) > l + \delta_L$

Lemma 2.4. If $v^\mathsf{T} N_L v \geq \frac{1}{t}$ then $\Phi_{l+\delta_L}(A) \leq \Phi_l(A)$ and $\lambda_{min}(A + t\, vv^\mathsf{T}) > l + \delta_L$

Lemma 2.5. $\mathrm{tr}(N_L) \geq \frac{1}{\delta_L} - \epsilon_L$

Proposition 2.6. $\forall x \in [0, 1]$, $\dfrac{1}{1+x} \leq 1 - \dfrac{x}{2}$

We use this proposition to prove the following claim:

Claim 2.7. $\exists i \in [m]$ such that $w_i^\mathsf{T} N_U w_i \leq w_i^\mathsf{T} N_L w_i$

Proof. As in lecture 16 and lecture 18,

$$\sum_{i=1}^{m} w_i^\mathsf{T} N_U w_i \;=\; \sum_{i=1}^{m} \mathrm{tr}(w_i w_i^\mathsf{T} N_U) \;=\; \mathrm{tr}(N_U) \;=\; \sum_i w_i^\mathsf{T} N_L w_i \;=\; \mathrm{tr}(N_L).$$

Also using Lemma 2.3 and Proposition 2.6,

$$\mathrm{tr}(N_U) \overset{\text{Lem 2.3}}{\leq} \frac{1}{\delta_U} + \epsilon_U \;=\; (1+4\epsilon)^{-1}\frac{p}{\epsilon^2} + \frac{p}{\epsilon} \overset{\text{Prop 2.6}}{\leq} \frac{p}{\epsilon^2}[(1-2\epsilon)+\epsilon] \;=\; \frac{p}{\epsilon^2} - \frac{p}{\epsilon} \;=\; \frac{1}{\delta_L} - \epsilon_L \;\leq\; \mathrm{tr}(N_L).$$

So, using the hypothesis that $\sum_i w_i w_i^\mathsf{T} = I$, we have $\sum_i w_i^\mathsf{T} N_U w_i \leq \sum_i w_i^\mathsf{T} N_L w_i$. Thus $\exists i$ such that $w_i^\mathsf{T} N_U w_i \leq w_i^\mathsf{T} N_L w_i$. $\square$

Corollary 2.8. There exists a good $i \in [m]$ and scalar t.

Proof. Fix any i with $w_i^\mathsf{T} N_U w_i \leq w_i^\mathsf{T} N_L w_i$. Set $t = \dfrac{1}{w_i^\mathsf{T} N_U w_i}$ so $w_i^\mathsf{T} N_U w_i \leq \dfrac{1}{t} \leq w_i^\mathsf{T} N_L w_i$. By lemmas 2.2 and 2.5, i and t are good. $\square$

2.4 Some remarks

1. Constants can be slightly improved. We can get

$$(1 - 2\epsilon + \epsilon^2) L_G \preccurlyeq L_H \preccurlyeq (1 + 2\epsilon + \epsilon^2) L_G$$

2. Theorem is asymptotically optimal: Let G be complete graph. For any $\epsilon \gg \dfrac{1}{\sqrt{n}}$ and any weighted subgraph H, with $(1-\epsilon)L_G \preccurlyeq L_H \preccurlyeq (1+\epsilon)L_G$ then $|F| = \Omega(\dfrac{n}{\epsilon^2})$

3. Main weakness: Weights:

   - For thin forests: we produced integral forest.
   - For expanders: we produced integral expander.

   But in today's sparsifier construction, the edges can have arbitrary (non-integral) weights.

   With the thin forests and expanders results, we had only one potential function, so an averaging argument allowed us to get integral weights. Today we have two potential functions, and averaging argument allows us to satisfy both simultaneously, but it does not also allow us to get integrality. Intuitively, i and t are "two degrees of freedom", and we also have "two constraints" which are the two potential functions. We cannot achieve integrality, because that would be a third constraint.

We remark that today's algorithm trivially extends to vectors $w_i \in \mathbb{C}^p$ instead of $w_i \in \mathbb{R}^p$. Suppose we could improve the algorithm to get uniform weights.

Claim 2.9. Assume $\|w_i\| \leq \alpha \quad \forall i$ (some constant α). Suppose algorithm can ensure $s_i = \beta \, \forall i$ and $I \preccurlyeq \sum_{i \in F} \beta \, w_i {w_i}^* \preccurlyeq \kappa \, I$ where $\beta > \kappa$. Then the Kadison-Singer Conjecture (1959), a major open problem in operator algebra is true.