Fast optimization for subgraph problems

October 16, 2015

1 Problem formulation

Let $G = (V, E)$ denote our underlying undirected unweighted graph with n nodes and let $w_{ij} = 1$ if $(i, j) \in E$ and zero otherwise. Denote the $n \times n$ edge Laplacian matrix as $L_{ij} = e_{ii} + e_{jj} - e_{ij} - e_{ji}$ for two nodes i and j, where e_{ij} is a zero matrix with a one in index i, j. Let $A \cdot B = \text{Tr}(AB)$ denote the inner product in the space of symmetric PSD matrices.

Let $G_S = (S, E_S)$ define a subgraph of graph G. Consider the following optimization problem with n-dimensional node objective vector c,

$$\max_{S \subset V : G_S \text{ connected}} \frac{\sum_{i \in S} c_i}{\sqrt{|S|}}. \tag{1}$$

Defining indicator vector $u \in \{0, 1\}^n$ such that $u_i = 1_{\{i \in S\}}$, above problem can be written as

$$\max_{u \in \{0,1\}^n, G_{\{i : u_i = 1\}} \text{ connected}} \frac{\sum_i c_i u_i}{\sqrt{\sum_i u_i}}, \tag{2}$$

and squaring the objective function and noting $u_i^2 = u_i$ we have

$$\max_{u \in \{0,1\}^n, G_{\{i : u_i = 1\}} \text{ connected}} \frac{\sum_{i,j} c_i c_j u_i u_j}{\sum_i u_i^2}. \tag{3}$$

We now relax each element u_i to a vector $v_i \in \mathcal{V}$ such that scalar multiplication is transformed to vector product and we have $\langle v_i, v_j \rangle = 1$ if $i, j \in S$ and zero otherwise. We then have

$$\max_{v_i \in \mathcal{V}, G_{\{i : |v_i| > 0\}} \text{ connected}} \frac{\sum_{i,j} c_i c_j \langle v_i, v_j \rangle}{\sum_i |v_i|^2}. \tag{4}$$

We can now use the Gram matrix $M = V^T V \succeq 0$ instead of v_i's and write

$$\max_{M \succeq 0, G_{\{i : |M_{ii}| > 0\}} \text{ connected}} \frac{C \cdot M}{I \cdot M}, \tag{5}$$

where $C = cc^T$. Note that we could also relax the problem directly such that $M = uu^T$.

Given a symmetric PSD matrix $M \in \mathbb{R}^{n \times n}$, $M \succeq 0$, we write

$$Q(M) = \sum_{i,j} (w_{ij} - \gamma) M_{ij} L_{ij} \succeq 0$$

as a constraint on the γ-connectivity of the subgraph induced by M (e.g. if $M = uu^T = 1_S 1_S^T$ for subgraph S) on underlying graph G. (To do: include Theorem 3 in [4] for the integer problem.)

Fixing $I \cdot M = 1$, we consider the following relaxed optimization problem (5):

$$\max_{M \in \Delta_n} \quad C \cdot M \quad \text{s.t.} \quad Q(M) \succeq 0, \tag{6}$$

where $\Delta_n = \{ X \in \mathbb{R}^{n \times n} : X \succeq 0, \mathrm{Tr}(X) = I_n \cdot X = 1 \}$ denotes the "spectrahedron" of $n \times n$ symmetric PSD matrices with unit trace and $C = cc^T$.

1.1 Additional constraints

We can also consider additional constraints that tighten the relaxation w.r.t. the integer problem. The first constraint we consider is an element-wise nonnegativity constraint $M \geq 0$ to help with the connectivity constraint. If we don't have such a constraint, it seems like a SDP solution may pick out disconnected components by giving positive values to one component and negative value to the other. This may complicate the iterative scheme, as we have to keep around these new n^2 constraints.

Secondly, note that ideally, we have $M = uu'$, where $u \in \{0, 1\}^n$ is K-sparse. Note that this implies

$$\left(\sum_i u_i \right)^2 \leq K \left(\sum_i u_i^2 \right),$$

which can be seen easily by induction over K. Expanding above inequality, we have

$$\sum_{i,j} u_i u_j \leq K \sum_i u_i^2.$$

This in turn corresponds to the following constraint on M

$$\sum_{i,j} M_{ij} \leq K \sum_i M_{ii},$$

and considering that we have $\mathrm{Tr}(M) = 1$, this is equivalent to $J \cdot M \leq K$, where $J \triangleq 1_n 1_n'$. Then a new formulation with the additional constraints is

$$\max_{M \in \Delta_n} \quad C \cdot M \quad \text{s.t.} \quad Q(M) \succeq 0, \quad M \geq 0, \quad J \cdot M \geq 0. \tag{7}$$

2 Non-smooth optimization with mirror descent

Mirror descent is an optimization procedure that generalizes subgradient methods to non-Euclidean spaces (see section 5.3 of [1] or [2]). For an optimization problem $\min_{x \in X} f(x)$, it tries to minimize the local linearization of the function while trying to stay close to the previous point using differentiable mirror map function $\omega(\cdot)$ to measure locality. This function must be 1-strongly convex with respect to a norm $|| \cdot ||$.

Mirror descent is given by the recurrence

$$x_0 = \arg\min_{x \in X} \omega(x), \quad x_{t+1} = \text{Prox}_{x_t}(\gamma_t f'(x_t)),$$

where $f'(x_t)$ is a subgradient of f at x_t, γ_t are step sizes and the proximity operator is defined as

$$\text{Prox}_x(\psi) = \arg\min_{y \in X} \omega(y) + \langle \psi - \omega'(x), y \rangle.$$

This proximity operator aims to move in the negative direction to ψ, while staying close to the original point x.

With the above steps, letting $x^T = \frac{\sum_{t=1}^{T} \gamma_t x_t}{\sum_{t=1}^{T} \gamma_t}$ and choosing step sizes appropriately, it is shown in Theorem 5.3.1 of [1] that

$$f(x^T) - \min_{x \in X} f(x) \leq \frac{\Omega L(f)}{\sqrt{T}},$$

where $L(f)$ is the Lipschitz constant of f w.r.t. the considered norm in X and Ω is related to the radius of X w.r.t. $\omega(\cdot)$ (e.g. $\Omega \leq \sqrt{2(\max \omega(\cdot) - \min \omega(\cdot))}$).

In the spectrahedron setup (for minimization $\min_{x \in \Delta_n} f(x)$), we will use the negative von Neumann entropy of a matrix, $\omega(x) = \sum_{i=1}^{n} \lambda_i \log \lambda_i$ as the mirror map, where λ_i are the eigenvalues of x. Notice that this map is 1-strongly convex with respect to the ℓ_1 norm of the eigenvalues, i.e., to the matrix trace norm. Working out the proximal mapping, this gives us the following multiplicative update rule (see part 2 of [2]):

$$M_{t+1} \propto \exp\left(\log M_t - \gamma_t f'(M_t)\right), \tag{8}$$

with matrix exponential and logarithm, $M_0 = \frac{1}{n} I_n$ and the right-hand side is normalized to unit trace to obtain M_{t+1}. $L(f)$ is the Lipschitz constant of f w.r.t. the matrix trace norm. We can also show that $\Omega = O(\sqrt{\log n})$.

Note that (8) is written for the general minimization problem, while we consider the maximization problem, so our update step would instead be

$$M_{t+1} \propto \exp\left(\sum_{\tau=1}^{t} \alpha_\tau f'(M_\tau)\right)$$

for some weights α_τ, where we also unrolled the recursion.

3

3 Subgradients

We consider the following Lagrangian of optimization problem (6) (where $h(x)$ is the hinge function such that $h(x) = x$ for $x \geq 0$ and zero otherwise):

$$\max_{M \in \Delta_n} f(M), \quad f(M) = C \cdot M + \min_{Y \succeq 0} h(Y \cdot Q(M)). \tag{9}$$

Note that $f(M) = C \cdot M$ if $Q(M) \succeq 0$. We do not consider the additional constraints in this section yet. TODO: formalize and state the gradient and updates for the additional constraints. Can do Lagrangian without the hinge.

We need to compute subgradients of $f(M)$ as formulated in (9) in order to compute mirror descent updates as stated above. We first state Danskin's Theorem [3] that considers saddle problems of the form (9).

Theorem 3.1 (Danskin's Theorem). *Let $f(x) = \max_z \phi(x, z)$, where $\phi(\cdot, z)$ is a convex function for all z. Define $Z_0(x) = \{z' : \phi(x, z') = \max_z \phi(x, z)\}$ to be the set of maximizers z given a point x. Then, under certain regularity conditions the subdifferential of f at x is given by*

$$\partial f(x) = \text{conv} \left\{ \partial \phi(x, z) : z \in Z_0(x) \right\}.$$

This way we see that for a given M, by finding a minimizer Y of the hinge $h(Y \cdot Q(M))$ we can obtain a subgradient for f at M. In particular, we have:

$$\frac{\partial Q(M)}{\partial M_{ij}} = (w_{ij} - \gamma) L_{ij}, \quad \frac{\partial (Y \cdot Q(M))}{\partial M_{ij}} = (w_{ij} - \gamma) Y \cdot L_{ij},$$

Therefore we can write the following expression for the subgradient of f at M.

$$\partial f(M) = \begin{cases} \{C\}, & Q(M) \succeq 0 \\ \text{conv} \left\{ C + \sum_{i,j} (w_{ij} - \gamma) (Y \cdot L_{ij}) e_{ij} : Y \in \arg \min_{Y' \succeq 0} \{h(Y' \cdot Q(M)\} \right\}, & Q(M) \not\succeq 0. \end{cases}$$

This can be further simplified by considering the definiton of the hinge function $h(\cdot)$. In particular, we have:

$$Y \in \arg \min_{Y' \succeq 0} \{h(Y' \cdot Q(M)\} = Y' \succeq 0 : Q(M) \cdot Y' \leq 0..$$

In words, any Y' which highlights a constraint broken by M is a valid subgradient.

4 Rounding

Instead of finding subgradients and solving mirror descent in the spectrahedron Δ_n to obtain a final M^T and rounding that to a cut S^T at the end of the algorithm, we may want to incorporate the rounding of the solution inside every iteration of the mirror descent algorithm. This might give us more intuitive iteration steps.

In every iteration t, we would round the matrix M_t to a cut S_t. We would then check whether $Q(S_t) \succeq 0$, if so, we would apply mirror descent with subgradient C at S_t to obtain M_{t+1}. If not, we would find a cut Y such that $Y \cdot Q(S_t)$ is minimized, which we would intuitively expect to be the cut that cuts set S_t with a value less than γ. We would then expect the mirror descent step that uses this Y to give us a new iterate M_{t+1} that has higher internal conductance by weighting the edges that are cut more.

In order to do this rounding at every step, we need to come up with a rounding scheme to go from matrix M_t to cut S_t, and also show that we can find a minimizer Y through solving a cut problem to obtain a subgradient.

Note that a simple rounding where $S = supp(diag(M))$ may not always be enough to give a connected solution. The reason for this can be that the solution M is possibly a mixture of connected solutions, so the union is not necessarily connected.

A better rounding strategy is random hyperplane rounding [5], where we do

$$w = M(v/\|v\|), \quad v \sim \mathcal{N}(0, I_n)$$

for random unit vector v, possibly multiple times. Then we can threshold these w to obtain multiple clusters, which we expect to be connected if the formulation is valid. For nodes i that are not selected, we expect that $w_i \approx 0$. Relaxation and rounding leading to an approximately optimal solution for the problem in (1) needs proof.

5 Other notes

- We may consider putting a slack term $Q(M) \succeq -\frac{\gamma}{2} I_n$ instead of $Q(M) \succeq 0$ in (9), which may lead to a better conditioning of the problem (?)

- Another thing to try is defining $Q(M) = \sum_{i<j} (w_{ij} - \gamma) M_{ij}^2 L_{ij}$ where we squared M elementwise. This might give a better relaxation w.r.t. the integer problem, but may not be convex (?)

- We can also consider an objective with a smooth soft-min penalty instead of the constraint $Q(M) \succeq 0$ such that the objective is

$$f(M) = C \cdot M - \frac{1}{t} \log \left(I \cdot e^{-tQ(M)} \right).$$

Writing $Q(M) = L \circ M$ where $\circ$ is the tensor-matrix product, we then have

$$\nabla f(M) = \frac{L \circ e^{-tL \circ M}}{I \cdot e^{-tL \circ M}},$$

for which the trace is

$$I \cdot \nabla f(M) = \frac{L \cdot e^{-tL \circ M}}{I \cdot e^{-tL \circ M}},$$

which is similar to the Rayleigh quotient for L.

References

[1] Aharon Ben-Tal and Arkadi Nemirovski, *Lectures on modern convex optimization*, `http://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt.pdf`.

[2] Sebastian Bubeck, *Mirror descent*, `https://blogs.princeton.edu/imabandit/2013/04/16/orf523-mirror-descent-part-iii/`.

[3] *Danskin's theorem*, `http://en.wikipedia.org/wiki/Danskin\%27s_theorem`.

[4] Jing Qian, Venkatesh Saligrama *Efficient minimax signal detection on graphs*, `http://arxiv.org/abs/1411.6203`.

[5] Jerome Le Ny, *Rounding techniques for semidefinite relaxations*, `http://www.professeurs.polymtl.ca/jerome.le-ny/docs/reports/SDProunding.pdf`.