Fast optimization for subgraph problems

September 25, 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. 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 $M = \frac{1}{n} 1_S 1_S^T$ for subgraph S) on underlying graph G.

We consider the following optimization problem:

$$\max_{M \in \Delta_n} \quad C \cdot M \quad \text{s.t.} \quad Q(M) \succeq 0, \tag{1}$$

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

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

$$\max_{M \in \Delta_n} f(M), \quad f(M) = C \cdot M + \min_{Y \succeq 0} h(Y \cdot Q(M)). \tag{2}$$

Note that $f(M) = C \cdot M$ if $Q(M) \succeq 0$.

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

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

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

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

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]):

$$M_{t+1} \propto \exp\left(\log M_t - \gamma_t f'(M_t)\right), \tag{3}$$

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 (3) 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 Subgradients

We need to compute subgradients of $f(M)$ as formulated in (2) in order to compute mirror descent updates as stated above. We first state Danskin's Theorem [3] that considers saddle problems of the form (2).

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.

$$\partial f(M) = \begin{cases} \{C\}, & Q(M) \succeq 0 \\ \text{conv} \left\{ C + \sum_{i,j} (w_{ij} - \gamma) \left(Y \cdot L_{ij} \right) e_{ij} : Y \in \arg\min_{Y' \succeq 0} \{h(Y' \cdot Q(M))\} \right\}, & Q(M) \nsucceq 0. \end{cases}$$

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 Some fixes to be included

Here are a number of things that need to be modified or dealt with in the current approach. Feel free to add more:

1. Formulation of the relaxation: if the goal is to relax the objective: $\max_{S \text{ connected}} \frac{\sum_{i \in S} c_i}{|S|^2}$, as it is in oneof the applications, then we probably need to do a binary search over the possible sizes of S. The reason is that the constraint $I \cdot M = 1$ and the objective $C \cdot M = 1$ are homogeneous, so they are suitable for optimizing $\frac{\sum_{i \in S} c_i}{|S|}$, which is trivially optimal at the maximum valued vertex.

2. I am a bit worried about the constraint $Q(M) \succeq 0$. It seems to me that it will work fine if we also impose $M_{ij} \geq 0$ for all i, j. If we don't do this, 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. Lorenzo will think more about this.

5 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.

6 Other notes

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

7 More refined formulation

We refine the formulation in Section 1, considering point 2 in Section 4, i.e., the fact that only $Q(M) \succeq 0$ might not give us a connected set. 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

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

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 is

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

where $M \in \Delta_n$ implies $M \succeq 0$ and $I \cdot M = 1$.

7.1 Optimization

Can be formalized in the mirror descent framework as above, with two additional dual variables corresponding to $J \cdot M \leq K$ and $M \geq 0$. TODO: formalize and state the gradient and updates.

7.2 Issues

It seems that the above formulation is not enough for connectedness, which can be seen in the simulations in constr.m. However adding constraints found in [4] does enforce connectedness.

For comparison purposes, in [4] the optimization formulation is

$$\max_{M \text{ symmetric}} \quad C \cdot M \quad \text{s.t.} \quad Q(M) \succeq 0, \quad 1 \geq M \geq 0, \quad M_{pp} = 1$$

$$M_{ij} \leq M_{ii} \, \forall i, j, \quad M_{ii} \leq M_{ip} \, \forall i, \quad I \cdot M \leq K$$

so the PSD constraint is relaxed but box constraints are added along with anchor constraints for an anchor node p and non-diagonal constraints (which is stronger than PSD).

It is stated in Theorem 3 in [4] that above constraints are satisfied for some γ and anchor p if and only if subgraph corresponding to M ($S = supp(diag(M))$) is connected. It might be worthwhile to try to establish a if and only if relationship between the proposed formulation (without anchor nodes) and the condition stated in the Theorem, i.e. with our constraints, there exists some p such that anchor constraints are satisfied and vice versa.

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://papers.nips.cc/paper/5525-efficient-minimax-signal-detection-on-graphs`.