Lecture Feb 3, 2010 - scribe notes

Today: completely non-systematic approach, toy version of problem to get some level of understanding

The “toy” problem

Given convex functions f_i, $i = 1$ to n, $-1 \leq f_i \leq 1$ (ie $|f_i| \leq 1$. We want to find an x s.t. $f_i(x) \leq 0$ for all i.

We now assume that we have an Oracle that can satisfy any one constraint:

given weights $w_1 ... w_n$, all weights positive, oracle will return an x s.t. $\sum w_i f_i(x) \leq 0$

1 Initial proof

Initially, we set all $w_i^{(0)} := 1$. We then modify weights: if an inequality is satisfied, its weight is lessened; if unsatisfied, it is increased. The hope is that sufficient iterations of this will solve the problem.

In other words, the iteration has the form:

$w_i^{(1+t)} = w_i^{(t)}(1 + \epsilon f_i(x^{(t)})$.

After T rounds, $x^* = \frac{1}{T}\sum\limits_t x^{(t)}$, and $T = \frac{\rho_n n}{\epsilon^2}$, $f_i(x^*) \leq \frac{\epsilon}{1-\epsilon}$.

Our potential function: $\Phi(t) = \sum w_i(t)$. Then $\Phi(0) = n$. It would be valuable to bound the growth of this, as that would ensure that the oracle cannot ignore a particular constraint for too long.

In fact, we will claim:

Claim $\Phi(t+1) \leq \Phi(t)$

Proof: $\Phi(t+1) = \sum w_i^{(t+1)} = \sum w_i^t(1 + \epsilon f_i x^t) = \Phi(t) + \epsilon \sum w_i^t f_i(x^t)) \leq \Phi(t)$

We now must show that x^* is $\epsilon - good$

Proof: $n \geq \Phi(T) \leq \max\limits_i(w_i^T) = \prod\limits_t(1+\epsilon f_i(x^t))$...now we must make this a sum.

Thought: why not have made the update rule multiplicative to begin with, so we might avoid this issue?

Answer: to show the potential function decreases, need the linear form anyhow...switch between the two is thus inevitable, so one is not easier than the other.

Returning to the proof:

Note that $(1 + \epsilon y) \geq (1 + \epsilon)^y$ for y between 0 and 1. Then $\prod\limits_t(1 + \epsilon f_i(x^t)) \geq (1 + \epsilon)^{\sum\limits_{t+} f_i(x^t)} + (1 - \epsilon)^{\sum\limits_{t-} -f_i(x^t)}$ (where t+ = positive f_i, t-= negative f_i) $\geq T\sum\limits_t \frac{t_i(x^t)}{T}[\epsilon - \epsilon^2] \geq Tf(x^*)\epsilon(1-\epsilon)$. $(1-\epsilon)f(x^*) \leq \frac{ln(n)}{T\epsilon} \leq \epsilon$, ie $T = \frac{ln}{\epsilon^2}$.

2 Generalizing the result

Let S be a convex set, $S \subset \mathbb{R}^m$, and consider given $f, g : S \to \mathbb{R}$, where all f_i convex, all g_i concave.

We now want some $x \in S$ s.t. $f_i(x) - g_i(x) \leq 0$. Let $\rho = \max_{x,i} |f_i(x) - g_i(x)|$.

As before we assume we have an oracle: $\sum w_i(f_i - g_i) \leq 0$. Then we now know that we can do this in $O(\frac{\rho ln}{\epsilon^2})$ oracle calls, and $x = \max_i [(1-\epsilon)f_i(x) - (1+\epsilon)g_i(x)] \leq \epsilon$.

3 Application to linear programming

Given some $\mathbf{A}x = \mathbf{b}$(and assuming $\mathbf{b} = 0$), we consider this as n different functions, assiging to each function A_i a weight variable w_i. The oracle then computes $\sum w_i A_i x \leq 0$. We now are examining the slack of this solution, and update accordingly.

Returning to SDPs

4 Refresher - Nonhomogenous Farkas

Statement of the theorem:

$A_i' s, B \in \mathbb{R}^{m \times m}$

$y_1 A_1 + ... + y_n A_n - B > 0$ or $\exists X \geq 0 : \forall i (A_i \cdot X = 0)$ and $B \cdot X \geq 0$.

Recall that we made sense of this by saying we were working in an m^2-dimensional vector space. Then the inequality is NOT satisfied iff our two convex sets (cone of positive semidefinite matrices, subspace generated by the linear combinations of the $A_i' s$ shifted by the B) have no intersection.

4.1 Proof of this thing

Consider the translation of this subspace back to the origin (in other words, assume $B = 0$), and assume we are not in the first case. What could this look like? Only way it could intersect the cone is if it lies on the surface (any move into the interior means that the difference between the surface and the subspace can be made arbitrarily large, contradicting the assumption that it didn't originally intersect). We then consider an X normal to the subspace. We know by def of normality that $X \cdot A_i = 0$, and for any positive semidefinite Z, $X \cdot Z \geq 0$. This may then be applied to B, as $-B$ "points in the wrong direction"/

5 Application of the toy

Given $A_1...A_n, B \in \mathbb{R}^{m\times m}$, $A_i^* = A_i$, etc, we want $y_1A_1 + ... + y_nA_n - B \succeq 0$. Suppose we have a potential solution, and wish to test it. Our test is analogous to the previous situation, as this is like checking infinitely many conditions: if $Z = y_1A_1 + ... + y_nA_n - B$, we are now saying that we want

$$\text{for all positive semidefinite } X,\ X \cdot Z \succeq 0$$

Start: $X = I$.

Assume we have an oracle that finds, on a particular X, $y_1...y_n : X \cdot (\sum(y_iA_i - B) \succeq 0$. This is the same as $\sum y_iA_i \cdot X - B \cdot X \geq 0$ (a linear inequality!)

We now must update X. The mind boggles: what does a multiplicative update on X look like?

Unboggling:

The situation we were in before with the "toy" is akin to having X be diagonal (as X analogizes to the weights in the previous problem). We now have X simply positive in general, making the situation a little more complicated. What constitutes a valid update, now?

Remember that a positive semidefinite matrix corresponds to a probability distribution of vectors. What we are doing now, in other words, is adjusting the probability distribution based on the result of the oracle, "fixing" in the direction that our current distribution fails (much in the way we did in the previous case).

5.1 Solving this dilemma

We will now output the average of the $y_i's$. Let $Y_t = \sum y_i^tA_i - B$. $\frac{1}{T}\sum Y_t$ is then the interesting value...

What should the update on X look like? The problem with updating on X: before, we could check n directions and assign penalties accordingly, but now we have infinitely many cases. Instead of looking at the current weight and assigning a penalty from there, we will generalize a different expression:

Remember from the toy that $\underset{i}{max} w_i^T = \underset{t}{\Pi}(1+\epsilon f_i(x^t)) \geq (1+\epsilon)^{\sum\limits_{t+} f_i(x^t)} + (1-\epsilon)^{\sum\limits_{t-} -f_i(x^t)}$. We now generalize: $w_i^t = \Pi(1+\epsilon)^{\sum f_i(x^t)}$.

So now: $X^t = (1+\epsilon)^{\sum\limits_t (y_i^tA_i - B)}$.

5.1.1 What is this doing?

If this is diagonal, not too hard to understand, rather like toy situation. The main issue here: the basis could keep changing. We must show that this does not mess us up.