COS 324, Precept #5: Extra Topics in Convex Analysis

October 13, 2017

1 Overview

Today, we'll talk about some additional concepts in convex analysis, in order to reinforce what we learned in class.

2 Convex Projection

In class, we defined projection onto a closed convex set $\mathcal{K}$:

$$\Pi_{\mathcal{K}}[x] = \operatorname*{argmin}_{y \in \mathcal{K}} \|x - y\|.$$

But if we hadn't seen that definition, we might have come up with a different one, which we get by "dropping a perpendicular": find a point x' on the boundary of $\mathcal{K}$, which is orthogonal to the tangent plane. As it turns out, this is an equivalent way to define projection. We'll derive the following property of projection:

$$\langle x - \Pi_{\mathcal{K}}[x], x' - \Pi_{\mathcal{K}}[x] \rangle \leq 0 \quad \forall x' \in \mathcal{K}.$$

Hopefully it is not too hard to convince ourselves that this captures the same idea: "the point x' for which x and $\mathcal{K}$ lie on opposite sides". We use this form because we don't want to define the "boundary" or "tangent plane".

Here is an informal proof: suppose this were not true. Then, there's some $y^* \in \mathcal{K}$ for which $x, \Pi[x], y^*$ form an acute angle. Since $\mathcal{K}$ is convex, the whole line segment between $\Pi[x], y^*$ is in $\mathcal{K}$. Thus there are points in $\mathcal{K}$ closer to x than $\Pi[x]$, a contradiction.

With this intuition, we can conclude (draw pictures):

- $\mathcal{K}$ is the unit n-sphere. Then, $\Pi_{\mathcal{K}}[x]$ is simply $x/\|x\|$ if $\|x\| > 1$, x otherwise.
- $\mathcal{K}$ is the unit n-hypercube. Then, $\Pi_{\mathcal{K}}[x]$ is the coordinate-wise "clipping" of x.

As it turns out, the problem of constrained convex optimization can be solved by *projected gradient descent*: alternate between GD and projection onto $\mathcal{K}$. Now we know how to implement this projection step for two important cases of convex projection.

3 Positive Semidefiniteness

Let's review the theory of positive (semi-)definite matrices. This is an incredibly useful property– in short, it allows us to manipulate a certain family of matrices like positive real numbers.

Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then, the following are all equivalent:

1. $M \succcurlyeq 0$ (M is positive semidefinite, psd for short).
2. For all $x \in \mathbb{R}^n$, $x^T M x \geq 0$.
3. $f(x) = x^T M x$ is convex.
4. There is some matrix A for which $f(x) = \|Ax\|^2$.
5. There is some matrix A for which $M = A^T A$. (M is a Gram matrix)
6. All eigenvalues of M are non-negative.

These are not hard to prove, if you're given the spectral theorem.

In convex analysis, psd matrices show up as the Hessian matrices of convex functions. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Recall that $g_x(y) = f(x) + \langle \nabla f(x), y - x\rangle$ is the linear approximation of f at y.

Then, the Hessian appears in the *second-order* Taylor approximation:

$$g'_x(y) = f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2}(y-x)^T \nabla^2 f(x)(y-x).$$

When f is convex, $\nabla^2 f(x)$ is psd for all x. In light of the condition (4), this tells us that the second-order Taylor approximation for f *looks like a paraboloid at all points*.

Furthermore when $\alpha I \preccurlyeq \nabla^2 f(x) \preccurlyeq \beta I$, we call f α-strongly convex and β-smooth. (Recall that $A \succcurlyeq B$ if $A - B \succcurlyeq 0$.) In light of condition 6, this means that all eigenvalues of the Hessian at all points lie in $[\alpha, \beta]$.

In the geometric view, this is a restriction on the shape of the approximating paraboloid guaranteed by convexity: in no direction can it be too steep or too shallow. As it turns out, gradient descent converges very rapidly (exponentially small error) on these kinds of functions. Indeed β/α takes the same role as the *condition number* from last precept.

4 Kissing Spheres: A Cautionary Tale

We'll often be thinking about convex sets and functions in high dimension: think of a bag-of-words feature vector ($d \approx 10^4$), or even the vector of someone's genome ($d \approx 10^9$). Sometimes, it suffices to draw a 2D picture to capture the intuition we want, but sometimes we must be careful, as high-dimensional space has some non-intuitive properties.

Consider placing a unit sphere at each corner of the n-dimensional hypercube $(\pm 1, \pm 1, \ldots)$. (Draw the 2D version.) Now, fit as large a sphere as possible, centered at the origin. Is this sphere smaller than the unit spheres in every dimension?

If you think about it, it's true in dimensions 1, 2, and 3. But a moment's thought shows that the radius of this sphere is $\sqrt{n} - 1$. So, this inner sphere not only outgrows the unit spheres– it gets arbitrarily large!

For this reason, it is often a perverse habit of high-dimensional geometers to draw a hypersphere like a spiky object. But this discards the symmetries, and is not a perfect picture either. So, let this be a word of warning when reasoning about the geometry of n-dimensional convex sets and functions.