Notes on Optimization

Vincenzo Gargano

February 20, 2023

Contents

Chapter 1

Lecture 5.2 : How simple function are made and how to optimize them

1.1 Scalar product, norm, distance and balls

Topologically balls are define as follows: What are points close to another point.

Ball, center $x \in \mathbb{R}^n$, radius $r > 0$: $\mathcal{B}(x, r) := y \in \mathbb{R}^n : ||y - x|| \leq r$ We want to work with multivariate functions, someone need to figure out things geometrically and some others that want to see algebraically.
Some function $f : D \to \mathbb{R}$, D is the domain $dom(f)$, but may not be all $\mathbb{R}^n$. For minimization $f : \mathbb{R}^n \mapsto \overline{\mathbb{R}}, f(x) = \infty$ for $x \notin D$. So we assume $f(x)$ a very large.

Here is how the professor see the high variable fucntions, the graph of a function in 1-dimension we can think to the epigraph of the function as all that is above the function:

[Epigraph] $epi(f) = (v, x) \in \mathbb{R}^{n+1} : v \geq f(x)$

So the epigraph lives in a $n + 1$ dimension with respect to our original function, more complicated in higher dimension...

You can slice the function at a certain heigh and plot the level set, because we are minimizing we want to use sublevels set, because if you find a value v the sublevel set will contain surely the minimum global.
Level set are the level set of the kind of norms, the 2-norm, an intresting property of the p-norm are some specimens, like 1-norm, whose level set are like diamond in a plane, and the infinity norm that is a square in the graph, as p grow the level set becomes larger and more smooth then a ssquare, when p shrink at less than 1 values the levels set becomes concave, that is not good, and the zero norm, (usually describes how many elements in the vector are non-zero). The level set corresponds to the cross of the plane, without 0 (useful for feature selection).

1.1.1 Picturing Multivariate Functions: Tomography

For more than 4 variables is impossible, we need alternative way called Tomography. So just this $f : \mathbb{R}^n \mapsto \mathbb{R}, x \in \mathbb{R}^n$,

$$\phi_{x,d}(\alpha) = f(x + \alpha d) : \mathbb{R} \mapsto \mathbb{R}$$

(1.1)

The tomography of a multivariable function is a function in one variable, and it's specified by giving two vector, a point to start (where we are), and the direction d, so you look to your function and slice the space with a vertical plane that passes from x and goes towards d, another way to visualize is to consider what happen when you fix all the variable other than one and then you have a one dimensional variable and this a tomography along one of the vectors of the canonical basis.

Let's look to simple function (linear) and their property

- $f(\gamma x) = \gamma f(x)$
- $f(x + z) = f(x) + f(z), \forall x, \gamma, z$

So linear function has no minima unless b is zero then all points are minimums, but talking about more intresting stuff.

1.1.2 Quadratic Functions

Quadratic are simple but more intresting!
Enough complicated to be useful. Fixed $Q \in \mathbb{R}^{n \times n}, (nQ_i \in \mathbb{R}^n), q \in \mathbb{R}^n$, should we be scared for big n number...

$$f(x) = \frac{1}{2}x^T Qx + qx \tag{1.2}$$

With quadratic term in diagonal and linear terms in non-diagonal and a linear term, this function is not linear. W.l.o.g Q is symmetric.

Proof. $x^T Qx = [(x^T Qx) + (x^T Qx)]/2] = x^T [(Q + Q^T)/2]x$ □

Symmetric function like $f(x) = f(-x)$, centered at 0 Tomography $\phi(\alpha) = f(\alpha d) = \alpha^2(d^T Qd)$ homogeneous quadratic univariate, sign and steepness depend on $d^T Qd$

Depending on the function if our direction is collinear to the eigenvectors some different things can happen (parabola more steep or less steep, sometimes even completely flat!), the same for parabola that is upside-down.
Also you can have different type of level sets (hyperboloids, ellipsoids or those from degenerate ellipsoids that are linear in one dimension). Shape of levels set will tells us if algorithm are fast to converge! Identity as perfect circles as levels sets.
Remember the curvature of our parabola is the eigenvalue.
For a function with all eigenvalues positive the only minimum is in zero, however you turn the parabola is pointing upwards so as soon you move from zero you increase!

If all are non negative but some are zero the minimum is still zero, but if you look along the eigevectors that have 0 as λ then you get still zero along that direction (parabola becomes flat for certain d), in these cases minimum in zero and in other place, and the function has no maximum because goes to $+\infty$, and if you reverse the sign the maximum is in zero and has no minimum, and if you have at least 1 positive and 1 negative λ the function has either no maximum no minimum, along a direction point downwards and another upwards!
Sometimes we can have a look at eigenvalues to see if our functions has a minimum! Once we know this we can rapidly know the cases what to do, with a quadratic function at least.

Chapter 2

Lecture 6.1: From optimality condition of quadratic function to gradients

Under simplifing assumption: (no zero λ_i), and we can immediately draw conclusion without our Q, is a fundamental trick: i know something on the easy case now i have a more complicated case i use my trick and my case in now simple again. (Changing the space or kernel trick). I can define a vector like $z = x - \hat{x}$ and $\hat{x} = -Q^{-1}q$, so $z = 0$ when x is equal to $\hat{x}$