AM 221: Advanced Optimization Spring 2016

Prof. Yaron Singer *Lecture 9 — February 24th*

1 Overview

In the previous lecture we reviewed results from multivariate calculus in preparation for our journey into convex optimization. In this lecture we present the gradient descent algorithm for minimizing a convex function and analyze its convergence properties.

2 The Gradient Descent Algorithm

From the previous lecture, we know that in order to minimize a convex function, we need to find a stationary point. As we will see in this lecture as well as the upcoming ones, there are different methods and heuristics to find a stationary point. One possible approach is to start at an arbitrary point, and move along the gradient at that point towards the next point, and repeat until (hopefully) converging to a stationary point. We illustrate this in the figure below.

Direction and step size. In general, one can consider a search for a stationary point as having two components: the direction and the step size. The direction decides which direction we search next, and the step size determines how far we go in that particular direction. Such methods can be generally described as starting at some arbitrary point $\mathbf{x}^{(0)}$ and then at every step $k \geq 0$ iteratively moving at direction $\Delta\mathbf{x}^{(k)}$ by step size t_k to the next point $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t_k \cdot \Delta\mathbf{x}^{(k)}$. In gradient descent, the direction we search is the negative gradient at the point, i.e. $\Delta\mathbf{x} = -\nabla f(\mathbf{x})$. Thus, the iterative search of gradient descent can be described through the following recursive rule:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - t_k \nabla f(\mathbf{x}^{(k)})$$

Choosing a step size. Given that the search for a stationary point is currently at a certain point $\mathbf{x}^{(k)}$, how should we choose our step size t_k? Since our objective is to minimize the function, one reasonable approach is to choose the step size in manner that will minimize the value of the new point, i.e. find the step size that minimizes $f(\mathbf{x}^{(k+1)})$. Since $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - t\nabla f(\mathbf{x}^{(k)})$ the step size $t_k^\star$ of this approach is:

$$t_k^\star = \text{argmin}_{t \geq 0}\, f(\mathbf{x}^{(k)} - t\nabla f(\mathbf{x}^{(k)}))$$

For now we will assume that $t_k^\star$ can be computed analytically, and later revisit this assumption.

The algorithm. Formally, given a desired precision $\epsilon > 0$, we define the gradient descent as described below.