6FMAI19 Nonlinear Optimization	Spring, 2022

Lecture #6 — Feb 23, 2022

Lecturer: Yura Malitsky *Scribe: Yannick Strocka*

1 Projected Subgradient Method

1.1 Update Rule

In order to discuss the analysis of the projected subgradient method, the definition of the subgradient is introduced.

Definition 1. *Consider a convex function $f\colon C \to \mathbb{R}$ on the convex and closed set $C \subset \mathbb{R}^n$. The quantity u is a subgradient of the function f if*

$$f(y) \geq f(x) + \langle u, y - x\rangle \quad \forall y \in C. \tag{1}$$

Now consider the optimization problem

$$\min_{x\in C} f(x) \tag{2}$$

with the function $f\colon \mathbb{R}^n \to \bar{\mathbb{R}}$ convex and the convex and closed set C. The projected subgradient method has the following update rule

$$\begin{cases} g_k \in \partial f(x_k) \\ x_{k+1} = P_C\left(x_k - \alpha_k g_k\right) \end{cases} \tag{3}$$

whereas P_C denoted the projection operator to the map C. The objective now is to develop a convergence analysis of the method.

1.2 Analysis

Now consider a minimizer $x^* \in \operatorname{argmin}_{x\in C} f(x)$. It holds $x^* = P_C x^*$. Apart from that, the projection operator has the property

$$\left\|P_C x - P_C y\right\| \leq \left\|x - y\right\|. \tag{4}$$

Since convergence properties are the figure of merit the squared error $\left\|x_{k+1} - x^*\right\|^2$ will be estimated. It holds

$$\begin{aligned} \left\|x_{k+1} - x^*\right\|^2 &= \left\|P_C\left(x_k - \alpha_k g_k\right) - P_C x^*\right\|^2 && (5)\\ &\leq \left\|x_k - \alpha_k g_k - x^*\right\|^2 && (6)\\ &= \left\|x_k - x^*\right\|^2 - 2\alpha_k\langle g_k, x_k - x^*\rangle + \alpha_k^2\left\|g_k\right\|^2 && (7)\\ &\leq \left\|x_k - x^*\right\|^2 - 2\alpha_k\left(f(x_k) - f(x^*)\right) + \alpha_k^2\left\|g_k\right\|^2 && (8) \end{aligned}$$

whereas the definition of the subgradient was used in the last step. Let $f_* := f(x^*)$. Rearranging the inequality yields

$$2\alpha_k\left(f(x_k) - f_*\right) \leq \left\|x_k - x^*\right\|^2 - \left\|x_{k+1} - x^*\right\|^2 + \alpha_k^2\left\|g_k\right\|^2. \tag{9}$$

Similarly to the previous lectures it is possible to sum up such that the result is

$$\sum_{i=0}^{k} \alpha_i \left(f(x_i) - f_*\right) \leq \frac{1}{2} \left\| x_0 - x^* \right\|^2 + \frac{1}{2} \sum_{i=0}^{k} \alpha_i^2 \left\| g_i \right\|^2 . \tag{10}$$

Consider the smallest term in the sum on the left hand side. It holds

$$\left(\min_{0 \leq i \leq k} \left(f(x_i) - f_*\right) \right) \sum_{i=0}^{k} \alpha_i \leq \frac{1}{2} \left\| x_0 - x^* \right\|^2 + \frac{1}{2} \sum_{i=0}^{k} \alpha_i^2 \left\| g_i \right\|^2 . \tag{11}$$

Let $i_m \in \mathbb{N}$ be chosen such that $G := \|g_{i_m}\| \geq \|g_i\| \, \forall i \in [0, k]$. Then it holds

$$\min_{0 \leq i \leq k} \left(f(x_i) - f_*\right) \leq \frac{\left\| x_0 - x^* \right\|^2 + G^2 \sum_{i=0}^{k} \alpha_i^2}{2 \sum_{i=0}^{k} \alpha_i}. \tag{12}$$

Since fast convergence should be achieved the right hand side should converge to zero as fast as possible. That means that a step size α_i must be chosen such that the quantity in the denominator $\sum_{i=0}^{k} \alpha_i$ grows as fast as possible. Lets consider $\alpha_k = \frac{c}{k^p}$ for some $c \geq 0$ and $p \in \mathbb{R}$ and analyze some cases for p. For $p = 1$ it holds

$$\sum_{i=0}^{k} \alpha_i \to \infty \quad \text{for} \quad k \to \infty. \tag{13}$$

The convergence speed is crucial. Therefore, the fact that the sum converges to infinity is not enough. How fast it tends to infinity is relevant as well. Basic knowledge about harmonic series tell us that the speed is $\log(k)$. This is heuristically a slowly growing function. So it is worth it to try out another value for p. Lets choose $p = \frac{1}{2}$. For this value of p it is possible to prove that the series goes to infinity in $\sqrt{k}$ which is a faster growing function than $\log(k)$. This implies

$$\min_{i} \left(f(x_i) - f_*\right) \leq \frac{\|x_0 - x^*\| + c \log(k)}{2\sqrt{k}} = O\left(\frac{\log(k)}{\sqrt{k}}\right) \tag{14}$$

and in order to achieve an accuracy ϵ about $O\left(\frac{1}{\epsilon^2}\right)$ iterations are needed. Furthermore, we need an optimal choice for the constant c. The set C is bounded by assumption. Therefore, there exists a constant R such that $\operatorname{diam}(C) \leq R$. This implies $\left\| x_{k+1} - x^* \right\|^2 \leq R^2$ and

$$\begin{aligned}
f(x_k) - f_* &\leq \frac{1}{2\alpha_k} \left\| x_k - x^* \right\|^2 - \frac{1}{2\alpha_k} \left\| x_{k+1} - x^* \right\|^2 + \frac{\alpha_k}{2} G^2 && (15)\\
&= \frac{1}{2\alpha_k} \left\| x_k - x^* \right\|^2 - \frac{1}{2\alpha_{k+1}} \left\| x_{k+1} - x^* \right\|^2 && (16)\\
&+ \frac{1}{2} \left\| x_{k+1} - x^* \right\|^2 \left(\frac{1}{\alpha_{k+1}} - \frac{1}{\alpha_k} \right) + \frac{\alpha_k}{2} G^2 && (17)\\
&\leq \frac{1}{2\alpha_k} \left\| x_k - x^* \right\|^2 - \frac{1}{2\alpha_{k+1}} \left\| x_{k+1} - x^* \right\|^2 && (18)\\
&+ \frac{R^2}{2} \left(\frac{1}{\alpha_{k+1}} - \frac{1}{\alpha_k} \right) + \frac{\alpha_k}{2} G^2 && (19)
\end{aligned}$$

Now it is possible again to sum up and get a compact result. That is

$$\sum_{i=0}^{k}(f(x_i)-f_*) \leq \frac{1}{2\alpha_0}\|x_0-x^*\|^2+\frac{R^2}{2}\left(\frac{1}{\alpha_{k+1}}-\frac{1}{\alpha_0}\right)+\frac{G^2}{2}\sum_{i=0}^{k}\alpha_i \tag{20}$$

$$\leq \frac{R^2}{2\alpha_{k+1}}+\frac{G^2}{2}\sum_{i=0}^{k}\alpha_i \tag{21}$$

$$\leq \frac{R^2\sqrt{k+1}}{2c}+\frac{G^2}{2}c\sqrt{k+1}. \tag{22}$$

This finally results in

$$\min_i (f(x_i)-f_*) \leq \frac{R^2}{2c\sqrt{k+1}}+\frac{G^2c}{2\sqrt{k+1}} = O\left(\frac{1}{\sqrt{k}}\right) \quad \text{for} \quad c=\frac{R}{G}. \tag{23}$$

For this value of c it holds consequently

$$\min_i (f(x_i)-f_*) \leq \sqrt{\frac{RG}{k+1}}. \tag{24}$$