37 Projected Gradient Descent

We consider a simple iterative procedure for solving a constrainted optimization

$$\text{minimize} \qquad f(x) \qquad \text{over} \qquad x \in \mathcal{X} \tag{88}$$

where f is a convex function twice differentiable function and where $\mathcal{X} \subset \mathbb{R}^p$ is some non-empty closed convex set, eg. $\{x \geq 0 : Ax = b\}$. Like in Section 36 we want to follow the steepest descent direction $x_{t+1} = x_t - \eta \nabla f(x_t)$. However, such points need not belong to $\mathcal{X}$. Thus we consider the *projection* of $x \in \mathbb{R}^p$ on to $\mathcal{X}$:

$$P_{\mathcal{X}}(x) = \operatorname*{argmin}_{y \in \mathcal{X}} ||x - y||^2$$

and then, from x_0, we define the *projected gradient descent* by

$$x_{t+1} = P_{\mathcal{X}}(x_t - \eta_t \nabla f(x_t)), \qquad t \in \mathbb{N} \tag{89}$$

- After projecting it need not be true that $f(x_{t+1}) < f(x_t)$. Thus we adjusted the step-size $\eta_t > 0$ and our proof will study the distance between x_t and the optimal solution x^* rather than the gap $f(x_t) - f(x^*)$.

The key observation is that making a projection cannot increase distances

Lemma 11.

$$||P_{\mathcal{X}}(x) - P_{\mathcal{X}}(y)|| \leq ||x - y||$$

Proof. For all $x' \in \mathcal{X}$, we must have $(x' - P_{\mathcal{X}}(x)) \cdot (x - P_{\mathcal{X}}(x)) \leq 0$, i.e. the plane $\mathcal{H} = \{z : (z - P_{\mathcal{X}}(x)) \cdot (x - P_{\mathcal{X}}(x)) = 0\}$ separates x from $\mathcal{X}$. If this were not true, then we would have a contradiction, in particular, there would be a point on the line joining x' and $P_{\mathcal{X}}(x)$ that is closer to x. We thus have

$$(P_{\mathcal{X}}(y) - P_{\mathcal{X}}(x)) \cdot (x - P_{\mathcal{X}}(x)) \leq 0 \quad \text{and} \quad (P_{\mathcal{X}}(x) - P_{\mathcal{X}}(y)) \cdot (y - P_{\mathcal{X}}(y)) \leq 0.$$

Adding together and then applying Cauchy-Schwartz implies

$$||P_{\mathcal{X}}(y) - P_{\mathcal{X}}(x))||^2 \leq (y - x) \cdot (P_{\mathcal{X}}(y) - P_{\mathcal{X}}(x)) \leq ||y - x|| \, ||P_{\mathcal{X}}(y) - P_{\mathcal{X}}(x))||,$$

as required. □

Theorem 33. *If the gradient of iterates are bounded, $K = \max_{t \in \mathbb{N}}\{||\nabla f(x_t)||^2\} < \infty$ then*

$$f(x^*) - \min_{0 \leq t \leq T} f(x_t) \leq \frac{\sum_{t=0}^T \eta_t^2}{\sum_{t=0}^T \eta_t} \max_{0 \leq t \leq T} ||\nabla f(x_t)||^2 + \frac{||x^* - x_0||^2}{\sum_{t=0}^T \eta_t}. \tag{90}$$

Hence if $\sum_{t=0}^\infty \eta_t = \infty$, $\sum_{t=0}^\infty \eta_t^2 < \infty$ and $\max_{0 \leq t \leq T} ||\nabla f(x_t)||^2 < \infty$ then $\lim_{t \to \infty} f(x^*) - \min_{0 \leq t \leq T} f(x_t) = 0$.

Proof.

$$\begin{aligned} ||x_{t+1} - x^*||^2 &= ||P_{\mathcal{X}}(x_t - \eta_t \nabla f(x_t)) - x^*||^2 \\ &\underbrace{\leq}_{\text{by Lemma 11}} ||x_t - x^* - \eta_t \nabla f(x_t)||^2 = ||x_t - x^*||^2 + \eta_t \underbrace{\nabla f(x_t)\,(x^* - x_t)}_{\substack{\leq f(x^*) - f(x_t), \\ \text{by convexity}}} + \eta_t^2 ||\nabla f(x_t)||^2 \\ &\leq ||x_t - x^*||^2 + \eta_t\,(f(x^*) - f(x_t)) + \eta_t^2 ||\nabla f(x_t)||^2 \end{aligned}$$

Summing the above expression yields

$$\begin{aligned} 0 \leq ||x_{T+1} - x^*||^2 &\leq ||x_0 - x^*||^2 + \sum_{t=0}^T \eta_t\,(f(x^*) - f(x_t)) + \sum_{t=0}^T \eta_t^2 ||\nabla f(x_t)||^2 \\ &\leq ||x_0 - x^*||^2 + \left(f(x^*) - \min_{0 \leq t \leq T} f(x_t)\right) \sum_{t=0}^T \eta_t + \max_{0 \leq t \leq T} ||\nabla f(x_t)||^2 \sum_{t=0}^T \eta_t^2. \end{aligned}$$

Rearranging the above gives the required result. □