FTML practical session 5: 2023/04/07

1 LINE SEARCH FOR LEAST SQUARES

We first note that

$$\begin{aligned}\nabla_\theta f(\alpha(\gamma)) &= H\alpha(\gamma) - \frac{1}{n}X^T y \\ &= H(\theta_t - \gamma\nabla_\theta f(\theta_t)) - \frac{1}{n}X^T y \\ &= \nabla_\theta f(\theta_t) - \gamma H \nabla_\theta f(\theta_t)\end{aligned} \tag{1}$$

Let us derivate $g(\gamma) = f(\alpha(\gamma))$ with respect to γ. By a composition :

$$\begin{aligned}g'(\gamma) &= \langle \nabla_\theta f(\alpha(\gamma)), -\alpha'(\gamma)\rangle \\ &= -\langle \nabla_\theta f(\theta_t) - \gamma H\nabla_\theta f(\theta_t), \nabla_\theta f(\theta_t)\rangle \\ &= -||\nabla_\theta f(\theta_t)||^2 + \gamma\langle H\nabla_\theta f(\theta_t), \nabla_\theta f(\theta_t)\rangle\end{aligned} \tag{2}$$

In order to cancel the derivative, we must have that

$$\gamma^* = \frac{||\nabla_{\theta_t} f||^2}{\langle H\nabla_\theta f(\theta_t), \nabla_\theta f(\theta_t)\rangle} \tag{3}$$

We note that this is correct if $\nabla_\theta f(\theta_t) \neq 0$. If $\nabla_\theta f(\theta_t) = 0$, this means that $\theta_t = \eta^*$, as f is convex.

This computation may then be done at each iteration.

An important remark is that if we note $\theta^*_{t+1} = \theta_t - \gamma^* \nabla_\theta f(\theta_t) = \alpha(\gamma^*)$, then equations 1 and 2 shows that

$$\langle \nabla_\theta f(\theta^*_{t+1}), \nabla_\theta f(\theta_t)\rangle = 0 \tag{4}$$

Two optimal directions of the gradient updates are **orthogonal**. Importantly, this is true in the general case, not only for least-squares.

1.0.1 *Backtracking line search*

In many practical situations, it is not possible to compute explicitely the optimal step γ^*. Or it could be possible, but too expensive computationnally.

In such situations, it is possible to compute an approximation of γ^*, for instance using **backtracking line search.** This method attempts to find a good γ by trying several decreasing values until a sufficient decrease in f after the gradient update is obtained.

https://en.wikipedia.org/wiki/Backtracking_line_search

RÉFÉRENCES