Contents

Frank-Wolfe 1

Convergence Property 1

Further. . . 2

Frank-Wolfe

The frank-wolfe attempts to bypass the projection step in the project or proximal method for the constrained problem. It's also called the conditional gradient method, and solves the following linearization formulation,

$$s_t = \arg\min_{s\in\mathcal{D}} \nabla f(x_t)^\top s,$$
$$x_{t+1} = x_t + \gamma_t(s_t - x_t).$$

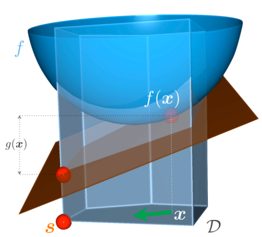

Figure 1: The illustration from Jaggi. It depicts the mechanism of Frank-Wolfe: find a extreme in the domain and move the current point towards the extreme.

Typically, finding the extreme in the domain $\mathcal{D}$ is simpler than the projection process. The application in some cases, e.g., l_1, l_p, trace norm, can be found in ryantibs slides.

Convergence Property

The vanilla convergence property depends on the assumption

$$M = \max_{\gamma,x,s,y} \frac{2}{\gamma^2}(f(y) - f(x) - \nabla f(x)^\top (y - x)),$$

which is no stronger than the smoothness assumption.

From the assumption, we can derive

$$f(x_{t+1}) \le f(x_t) - \gamma g(x_t) + \frac{\gamma_t^2}{2} M,$$

where $\gamma_t = \frac{1}{t+2}$ and $g(x_t) = \nabla f(x_t)^\top (x_t - s_t)$ is the upper bound of the dual gap.

Using $g(x_t) \geq f(x_t) - f(x^*)$ and $h_t := f(x_t) - f(x^*)$, we have

$$h_{t+1} \leq (1 - \gamma_t) h_t + \frac{\gamma_t^2}{2} M,$$

which can induce that $h_t \leq \frac{2M}{t+1}$.

Further...

More variants of FW and problems such as zig-zag can be found in On the Global Linear Convergence of Frank-Wolfe Optimization Variants.