THE FRANK-WOLFE Algorithm

Gonghan Xu

The optimization problem

$$\min_{x \in P} f(x)$$

P is a compact convex set. f(x) is convex.

Naïve Gradient Descent

Algorithm 1 Vanilla Gradient Descent Algorithm

```
Initialize iterate x_0 ∈ P
Set some positive step size α and maximum number of iterations N
for t = 0 to N − 1 do
    x_{t+1} = x_t − α · ∇f(x_t)
end for
return x_N
```

Initialize iterate $x_0 \in P$

Set some positive step size α and maximum number of iterations N

for $t = 0$ to $N - 1$ **do**

$\quad x_{t+1} = x_t - \alpha \cdot \nabla f(x_t)$

end for

return x_N

The Frank-Wolfe Algorithm

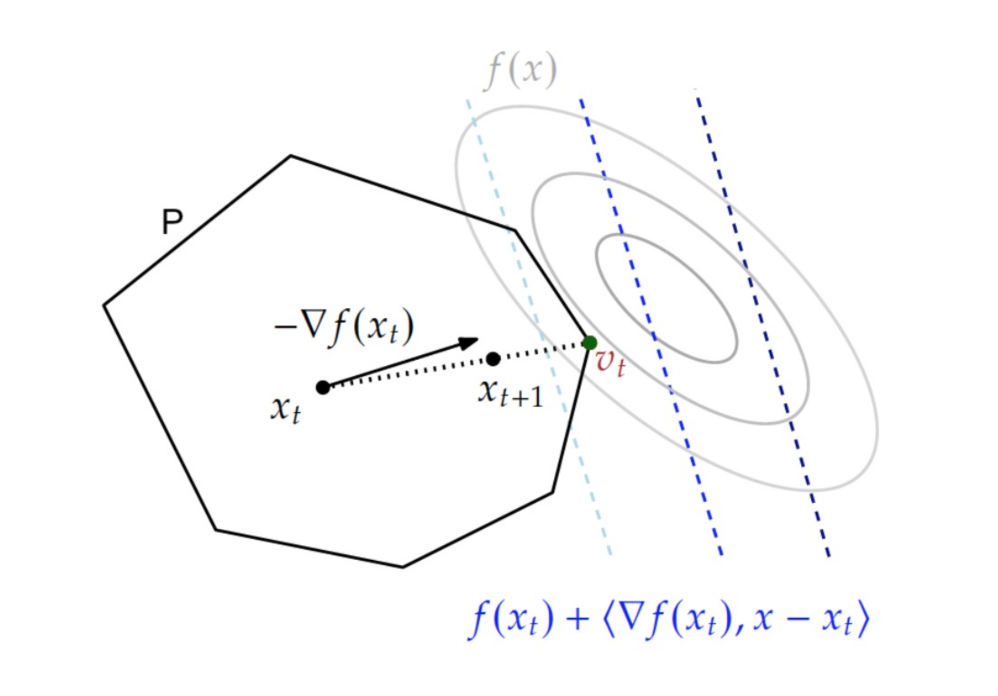
The target point: $v_t = \text{argmax}_{v \in P} \langle -\nabla f(x), v \rangle$

Descent step: $x_{t+1} = x_t + \alpha(v_t - x_t)$

Algorithm 2 Frank-Wolfe Algorithm

Initialize iterate $x_0 \in P$
Set maximum number of iterations N
for $t = 0$ to $N - 1$ **do**
    $v_t = \text{argmin}_{v \in P} \langle \nabla f(x_t), v \rangle$
    Set step size $\alpha_t \in [0, 1]$ based on some rules
    $x_{t+1} = (1 - \alpha_t) x_t + \alpha_t v_t$
end for
return x_N

The Illustration



The Frank-Wolfe Gap

$$f(x) - f(x^*) \leqslant \langle \nabla f(x), x - x^* \rangle \;\leqslant \max_{v \in P} \langle \nabla f(x), x - v \rangle$$

Suboptimality

Frank-Wolfe Gap g(x)

$$\begin{aligned} g(x_t) &\equiv \max_{v \in P} \langle \nabla f(x_t), x_t - v \rangle \\ &= \langle \nabla f(x_t), x_t \rangle - \min_{v \in P} \langle \nabla f(x_t), v \rangle \\ &= \langle \nabla f(x_t), x_t \rangle - \langle \nabla f(x_t), v_t \rangle \end{aligned}$$

Towards Convergence

$$f(y) - f(x) \leqslant \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|_2^2$$

$$f(x_{t+1}) - f(x^*) \leqslant (1 - \alpha_t)[f(x_t) - f(x^*)] + \alpha_t^2 \frac{LD^2}{2}$$

$$f(x_t) - f(x^*) \leqslant \frac{2LD^2}{t + 2}$$

Theorem 1 (Suboptimality convergence of the Frank-Wolfe algorithm) *Let f be an L-smooth convex function and let P be a compact convex set with diameter D, then with open loop step size $\alpha_t = \frac{2}{t+2}$, the suboptimality of the iterates in Algorithm 2 satisfies*

$$f(x_t) - f(x^*) \leqslant \frac{2LD^2}{t+2}$$

Proved by induction!

Step Size Choices

- 1. Constant step size
- 2. Open loop step size: 2/t+2
- 3. Short step: $$\alpha_t = \min\{\frac{\langle \nabla f(x_t), x_t - v_t \rangle}{L \|x_t - v_t\|_2^2}, 1\}$$
- 4. Line search: $$\alpha_t = \text{argmin}_{\alpha_t \in [0,1]} f(x_t + \alpha_t(v_t - x_t))$$

Experiment on Lasso Problem

$$\min_{\|\beta\|_1 \leqslant \lambda} \|X\beta - y\|_2^2$$