Next Week

Tuesday — Subgradient Methods

Thursday — Stochastic Methods

Friday — Midterm posted

Generic Nonsmooth Optimization (Convex)

Not assuming prox is computable
(not structured)

HW3 Q3, looks SVM

$$\min_{\omega} \underbrace{\sum \max\{0, 1-y_i x_i^T \omega\}} + \frac{\lambda}{2}\|\omega\|_2^2$$

Easy to compute a subgradient

Hard to compute hard prox

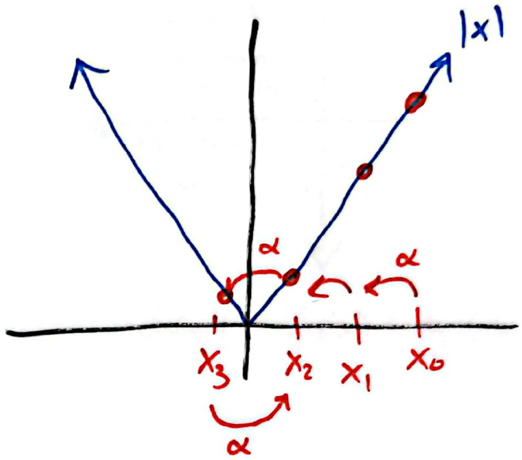
HW3, Q4, Subgradient Method

$$x_{k+1} = x_k - \alpha_k g_k, \quad g_k \in \partial f(x_k)$$

This quickly goes "wrong" (by our standards from smooth opt).

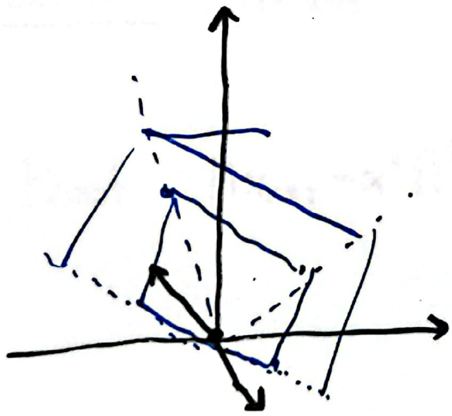
No guarantee of convergence for constant α.

Why? $f(x) = |x|$

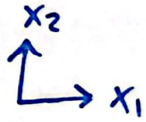


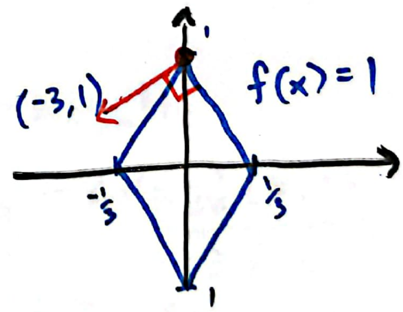
No guarantee of descent.

Why? $f(x) = 3|x_1| + |x_2|$



Contour Plot



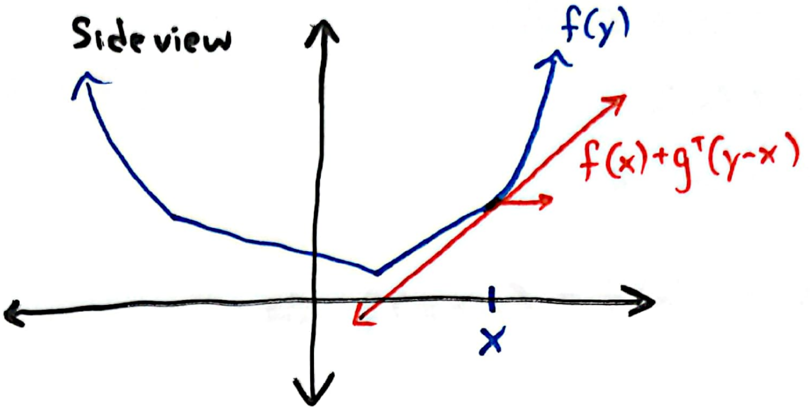


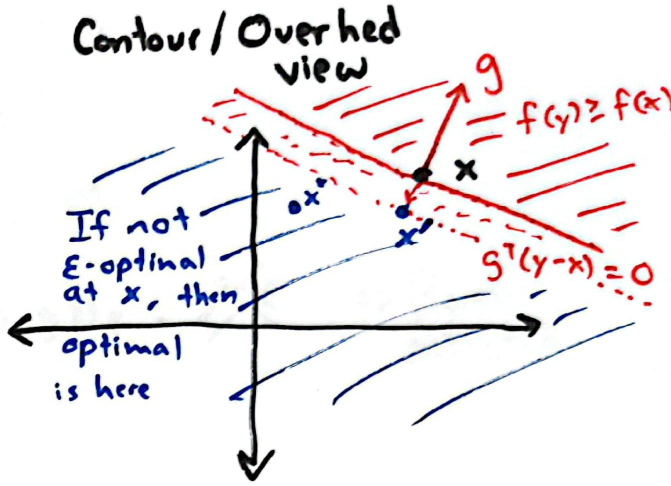
$$\partial f(0,1) = \partial(3|x_1|)(0,1) + \partial(|x_2|)(0,1)$$

$$= 3\,\partial(|x_1|)(0,1) + \partial(|x_2|)(0,1)$$

$$= \begin{bmatrix} 3[-1,1] \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

In a picture, two ways to think about $g \in \partial f(x)$





$$f(y) \geq f(x) + g^T(y-x)$$

For $x_{k+1} = x_k - \alpha_k g_k$, $g_k \in \partial f(x_k)$

If $g^T(y-x) > -\varepsilon$, then $f(y) \geq f(x) - \varepsilon$.

For any convex f,

<u>Lemma</u> $$\|x_{k+1} - x^*\|_2^2 \leq \|x_k - x^*\|_2^2 - 2\alpha_k(f(x_k) - f(x^*)) + \alpha_k^2 \|g_k\|_2^2.$$

Proof.
$$\begin{aligned} \|x_{k+1} - x^*\|_2^2 &= \|x_k - x^* - \alpha_k g_k\|_2^2 \\ &= \|x_k - x^*\|_2^2 - 2\alpha_k \underbrace{g_k^T(x_k - x^*)}_{\geq f(x_k) - f(x^*)} + \alpha_k^2 \|g_k\|^2 \\ &\leq \|x_k - x^*\|_2^2 - 2\alpha_k(f(x_k) - f(x^*)) + \alpha_k^2\|g_k\|^2 \end{aligned}$$

<u>Lemma</u>. If f is M-Lipschitz, then all $g \in \partial f(x)$, $\|g\|_2 \leq M$.

Also true.

Proof. Suppose to the contrary, some $\|g\|_2 > M$, $g \in \partial f(x)$

Look at $y = x + g$.

$$
\begin{aligned}
f(y) &\geq f(x) + g^T (y - x) \\
&= f(x) + \|g\|_2^2 \\
&> f(x) + \|g\|_2 \cdot M
\end{aligned}
$$

$\Rightarrow f(x+g) - f(x) > M \|g\|_2$. $\times\!\!\!\times$ $\Rightarrow$ Bound subgrad. $\square$