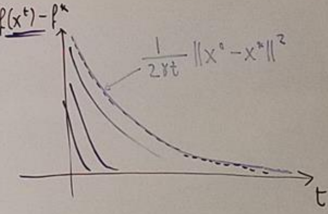
Convergence analysis of the Gradient Method (GM) when applied to convex costs.

Theorem Let $f \in \mathcal{F}_L^1(\mathbb{R}^n)$ with finite minimum $f^* = f(x^*)$ attained at $x^* \in \mathbb{R}^n$.

Then, for any $0 \le \gamma \le 1/L$, the iterates of GM satisfy

$$\left(f(x^t) - f(x^*) \right) \le \frac{1}{2\gamma t} \| x^0 - x^* \|^2 .$$

This means that the $\{f(x^t)\}_{t \in \mathbb{N}_0}$ converges sublinearly to f^*.



"worst case analysis": $f(x^t) - f^* = \frac{1}{2\gamma t}\|x^0 - x^*\|^2 \rightsquigarrow f(x^{t+1}) - f^* = \frac{1}{2\gamma(t+1)}\|x^0 - x^*\|^2$

$$\frac{f(x^{t+1}) - f^*}{f(x^t) - f^*} = \dots = \frac{t}{t+1} \xrightarrow{t \to \infty} 1 \qquad (p = 1,\ \beta = 1 \to \text{sublinear convergence})$$

Proof (Do compare with proof of GM for general f in Lecture 5)

Two key bounds on $f \in \mathcal{F}_L^1(\mathbb{R}^n)$:

- Lipschitz continuity of ∇f yields (Lemma in Lecture 6):

$$f(z) \le f(x) + (\nabla f)(x)^T (z - x) + \frac{L}{2}\|z - x\|^2 \quad \forall z, x \in \mathbb{R}^n \qquad (1)$$

- From convexity of f, we have:

$$f(z) \ge f(x) + (\nabla f)(x)^T (z - x) \quad \forall z, x \in \mathbb{R}^n . \qquad (2)$$

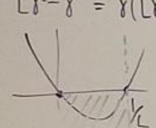
Consider a single iteration of GM: $\underline{x^+ = x - \gamma (\nabla f)(x)}$ with $0 \le \gamma \le 1/L$. Then from (1) with $z = x^+$

$$\begin{aligned}
f(x^+) &\le f(x) + (\nabla f)(x)^T (x^+ - x) + \frac{L}{2}\|x^+ - x\|^2 \\
&= f(x) + (\nabla f)(x)^T \left(-\gamma (\nabla f)(x) \right) + \frac{L}{2}\| -\gamma (\nabla f)(x) \|^2 \\
&= f(x) - \gamma \|(\nabla f)(x)\|^2 + \underbrace{\frac{L\gamma^2}{2}}_{\le \frac{\gamma}{2} \text{ with } (*)} \|(\nabla f)(x)\|^2 \\
&\le f(x) - \frac{\gamma}{2}\|(\nabla f)(x)\|^2 .
\end{aligned}$$

$x^+ - x = -\gamma(\nabla f)(x)$

$L\gamma^2 \le \gamma$ for $\gamma \le \frac{1}{L}$

(*) why? $L\gamma^2 - \gamma = \gamma(L\gamma - 1)$



$$f(x^+) \le f(x) - \frac{\gamma}{2}\|(\nabla f)(x)\|^2 \rightsquigarrow f(x^+) \le f(x)$$

Now apply (2) with $z = x^*$: $f(x^*) \geq f(x) + (\nabla f)(x)^T(x^* - x) \Rightarrow f(x) \leq f(x^*) + (\nabla f)(x)^T(x - x^*)$

(f is convex!)

$$
\begin{aligned}
f(x^+) &\leq f(x) - \frac{\gamma}{2}\|(\nabla f)(x)\|^2 \\
&\leq f(x^*) + (\nabla f)(x)^T(x - x^*) - \frac{\gamma}{2}\|(\nabla f)(x)\|^2 \\
&= f(x^*) + \frac{1}{2\gamma}\Big(\|x - x^*\|^2 \underbrace{- \|x - x^*\|^2 + 2\gamma(\nabla f)(x)^T(x - x^*) - \gamma^2\|(\nabla f)(x)\|^2}_{\text{Completion of squares}}\Big) \\
&= f(x^*) + \frac{1}{2\gamma}\Big(\|x - x^*\|^2 - \|\underbrace{(x - x^*) - \gamma(\nabla f)(x)}_{= x^+ - x^* \text{ by def. of } x^+ = x - \gamma(\nabla f)(x)}\|^2\Big) \\
&= f(x^*) + \frac{1}{2\gamma}\Big(\|x - x^*\|^2 - \|x^+ - x^*\|^2\Big).
\end{aligned}
$$

$-(a-b)^2 = -a^2 + 2ab - b^2$

Consider $x = x^{k-1}$, $x^+ = x^k$ and add the inequalities for $k = 1, 2, \dots, t$:

$$\sum_{k=1}^{t}\left(f(x^k) - f(x^*)\right) \leq \frac{1}{2\gamma}\sum_{k=1}^{t}\left(\|x^{k-1} - x^*\|^2 - \|x^k - x^*\|^2\right) \underset{\substack{\uparrow \\ \text{telescoping sum}}}{=} \frac{1}{2\gamma}\left(\|x^0 - x^*\|^2 - \|x^t - x^*\|^2\right) \leq \frac{1}{2\gamma}\|x^0 - x^*\|^2.$$

We saw $f(x^+) \leq f(x)$, so that $f(x^k)$ is nonincreasing, which implies:

$$\underbrace{\frac{1}{t}\sum_{k=1}^{t}\left(f(x^t) - f(x^*)\right)}_{f(x^t) - f(x^*)} \leq \frac{1}{t}\sum_{k=1}^{t}\left(f(x^k) - f(x^*)\right) \leq \frac{1}{2\gamma t}\|x^0 - x^*\|^2$$