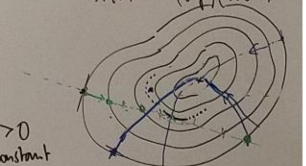
The Gradient Method (GM)

The GM is one of the most common algorithms used in optimization.
Starting from an initialization $x^0 \in \mathbb{R}^n$, the GM iteratively generates the sequence

$$x^t = x^{t-1} - \gamma_t (\nabla f)(x^{t-1}) \quad , \quad t = 1, 2, 3, \dots$$

$\dot{x}(t) = -(\nabla f)(x(t))$



where $\gamma_t > 0$ is the step size. The algorithm requires the user to select an appropriate step size.

Step-size Selection:

- Fixed step: $\gamma_t \equiv \gamma > 0$. γ small to ensure convergence, large enough for GM to be fast
- Decreasing step: $\gamma_t = \frac{C}{t}$ or $\frac{C}{\sqrt{t}}$, $C > 0$ constant
- Exact Line Search: Selecting γ_t via $\gamma_t = \arg\min_{\gamma \geq 0} \phi(\gamma) := f(x^{t-1} - \gamma (\nabla f)(x^{t-1}))$.

Recall: **Theorem:** Let $f \in C_L^1(\mathbb{R}^n)$ be the cost that has finite minimum at $x^* \in \mathbb{R}^n$. Then for any step size $\gamma \in (0, \frac{2}{L})$, the iterates generated by the gradient method satisfy $\lim_{t \to \infty} \|(\nabla f)(x^t)\| = 0$.

Lemma: For any $f \in C_L^1(\mathbb{R}^n)$, we have that $f(y) \leq f(x) + (\nabla f)(x)^T (y-x) + \frac{L}{2}\|y-x\|^2$, $\forall x, y \in \mathbb{R}^n$.

Proof: Define $\boxed{\varphi(t) := f(x + tv) \quad \text{where} \quad v := y - x.}$ Fundamental thm of Calculus: $\int_0^1 \varphi'(t)\,dt = \varphi(1) - \varphi(0) \quad | -\varphi'(0)$

$\int_0^1 (\varphi'(t) - \varphi'(0))\,dt = \varphi(1) - \varphi(0) - \varphi'(0)$

$$\Rightarrow \varphi(1) = \varphi(0) + \varphi'(0) + \int_0^1 (\varphi'(t) - \varphi'(0))\,dt$$

$$\begin{aligned}
f(y) &= f(x) + (\nabla f)(x)^T v + \int_0^1 \big((\nabla f)(\underbrace{x+tv}_{\tilde{y}}) - (\nabla f)(x)\big)^T v\,dt \\
&\leq f(x) + (\nabla f)(x)^T v + \int_0^1 \underbrace{\|(\nabla f)(x+tv) - (\nabla f)(x)\|}_{\leq L\|\tilde{y} - x\| = L\|tv\| = Lt\|v\|} \|v\|\,dt \qquad \text{Cauchy Schwarz} \\
&\leq f(x) + (\nabla f)(x)^T v + \int_0^1 Lt\|v\|^2\,dt \\
&= f(x) + (\nabla f)(x)^T v + \frac{L}{2}\|v\|^2 \longleftarrow \text{recall } v = y - x
\end{aligned}$$

In the proof of the theorem, we managed to show:

$$\frac{\gamma}{2} \sum_{i=1}^{t} \|(\nabla f)(x^{i-1})\|^2 \leq f(x^0) - f(x^1) + \dots + f(x^{t-1}) - f(x^t) = \dots = f(x^0) - f(x^t) \leftarrow \text{finite}$$

We concluded that $\lim_{i \to \infty} \|(\nabla f)(x^i)\| = 0$, which proves the theorem. We can, in addition, say a little more. Define

$$g_t = \min_{1 \leq i \leq t} \|(\nabla f)(x^{i-1})\|^2.$$

Then

$$\frac{\gamma}{2} t g_t \leq \frac{\gamma}{2} \sum_{i=1}^{t} \|(\nabla f)(x^{i-1})\|^2 \leq f(x^0) - f(x^t)$$

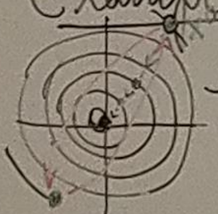
$$\Rightarrow g_t \leq \frac{2}{\gamma t}(f(x^0) - f(x^t)) \Rightarrow \min_{1 \leq i \leq t} \|(\nabla f)(x^{i-1})\|^2 \leq \sqrt{\frac{2}{\gamma t} \underbrace{(f(x^0) - f(x^t))}_{\text{const.}}} \leq \varepsilon \text{, e.g. } 10^{-6}$$

$$\Rightarrow \frac{2}{\gamma t}\text{const.} \leq \varepsilon^2 \Rightarrow t \geq \frac{2\,\text{const.}}{\gamma} \varepsilon^2.$$

Thus, GM finds an ε-approximate stationary point in in $O(\frac{1}{\varepsilon^2})$ iterations. $(\|(\nabla f)(x)\| \leq \varepsilon) \dots \varepsilon$ small

Examples: Fixed step

$\|(\nabla f)(x) - (\nabla f)(y)\| \le \underset{?}{L}\, \|x-y\| \quad (L=2)$



$f(x_1,x_2) = x_1^2 + x_2^2 \quad , (\nabla f)(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} = 2x \quad , \quad x^0 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$

$$x^1 = x^0 - \overset{=1}{\gamma}(\nabla f)(x^0)$$

(i) $\gamma = 1$:

$$= \begin{pmatrix} 3 \\ 4 \end{pmatrix} - 1 \times \begin{pmatrix} 6 \\ 8 \end{pmatrix} = \begin{pmatrix} -3 \\ -4 \end{pmatrix}$$

bad
oscillating!

(ii) $\gamma = \frac{1}{2}$:

$$x^1 = \begin{pmatrix} 3 \\ 4 \end{pmatrix} - \frac{1}{2} \times \begin{pmatrix} 6 \\ 8 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

works
(maybe too well)

(iii) $\gamma = \frac{1}{4}$:

$$x^1 = \begin{pmatrix} 3 \\ 4 \end{pmatrix} - \frac{1}{4} \times \begin{pmatrix} 6 \\ 8 \end{pmatrix} = \begin{pmatrix} 3/2 \\ 2 \end{pmatrix}$$

$$x^2 = \begin{pmatrix} 3/2 \\ 2 \end{pmatrix} - \frac{1}{4} \times \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 3/4 \\ 1 \end{pmatrix}$$

$$x^3 = \begin{pmatrix} 3/4 \\ 1 \end{pmatrix} - \frac{1}{4} \times \begin{pmatrix} 3/2 \\ 2 \end{pmatrix} = \begin{pmatrix} 3/8 \\ 1/2 \end{pmatrix}$$

Exact Line Search:

quadratic cost with ellipses as level sets