Lecture 4: Outline

- Convexity and implications.
- Quadratic problems.
- Gradient methods.
- Step-size selection.
- Line search methods.
Convexity

**Proposition**

Let $f : C \to \mathbb{R}$ be convex over the convex set $C \subset \mathbb{R}^n$.

- A local min of $f$ over $C$ is also a global min over $C$. If $f$ is strictly convex, $\exists$ at most one global min.
- If $f$ is convex and $C$ is open, the condition
  \[ \nabla f(x^*) = 0 \]
  is necessary and sufficient for $x^* \in C$ to be a global min of $f$ over $C$.

**Quadratic problems: example**

\[ f(x, y) = \frac{1}{2}(ax^2 + by^2) - x. \]
Gradient Methods

- Generic gradient method:
  \[ x^{k+1} = x^k + \alpha^k d^k \]
such that if \( \nabla f(x^k) \neq 0 \) then \( d^k \) is chosen so that \( \nabla f(x^k)'d^k < 0 \) (descent direction).

- An interesting class of gradient methods:
  \[ x^{k+1} = x^k - \alpha^k D^k \nabla f(x^k) \].

Gradient Methods: Variants

\[ x^{k+1} = x^k - \alpha^k D^k \nabla f(x^k) \].

- Steepest descent: \( D^k = I \).
- Newton’s method: \( D^k = (\nabla^2 f(x^k))^{-1} \).
- Diagonally scaled steepest descent:
  \[ D^k = \text{diag} \left( \left( \frac{\partial^2 f(x^k)}{\partial x_1^2} \right)^{-1}, \ldots, \left( \frac{\partial^2 f(x^k)}{\partial x_n^2} \right)^{-1} \right) \).
- Modified Newton’s method: \( D^k = (\nabla^2 f(x^0))^{-1} \).
- Discretized Newton’s method: \( D^k = (H(x^k))^{-1} \), where
  \( H(x^k) \) is a finite-difference based approximation of the Hessian.
Stesize Selection: Minimization rules

- **Minimization rule:**
  \[ f(x^k + \alpha^k d^k) = \min_{\alpha \geq 0} f(x^k + \alpha d^k). \]

- **Limited minimization rule:**
  \[ f(x^k + \alpha^k d^k) = \min_{\alpha \in [0, s]} f(x^k + \alpha d^k). \]

These problems are solved by line search methods.

Constant Stesize

**Constant stepsize:** \( \alpha^k = s. \)

Idea: Try constant stepsize and if no cost reduction reduce \( s \) until you achieve cost reduction in every iteration.

Example:
Stepsizes (cont.)

- **Diminishing stepsize:**
  \[
  \alpha^k \to 0 \quad \text{with} \quad \sum_{k=0}^{\infty} \alpha^k = \infty.
  \]

- **Armijo rule:** Fix $s$ the “initial” stepsize, $\beta \in (0, 1)$, $\sigma \in (0, 1)$. Try $\alpha^k = \beta^m s$, $m = 0, 1, \ldots$ until you find $m_k$ such that
  \[
  f(x^k) - f(x^k + \beta^{m_k} s d^k) \geq -\sigma \beta^{m_k} s \nabla f(x^k)' d^k
  \]