# Subgradient Methods for Constrained Problems

- projected subgradient method
- projected subgradient for dual
- subgradient method for constrained optimization

# Projected subgradient method

solves constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathcal{C}, \end{array}$$

where $f : \mathbf{R}^n \to \mathbf{R}$, $\mathcal{C} \subseteq \mathbf{R}^n$ are convex

**projected subgradient method** is given by

$$x^{(k+1)} = \Pi(x^{(k)} - \alpha_k g^{(k)}),$$

$\Pi$ is (Euclidean) projection on $\mathcal{C}$, and $g^{(k)} \in \partial f(x^{(k)})$

same convergence results:

- for constant step size, converges to neighborhood of optimal (for $f$ differentiable and $h$ small enough, converges)

- for diminishing nonsummable step sizes, converges

**key idea**: projection does not increase distance to $x^\star$

# Linear equality constraints

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

projection of $z$ onto $\{x \mid Ax = b\}$ is

$$\begin{aligned} \Pi(z) &= z - A^T(AA^T)^{-1}(Az - b) \\ &= (I - A^T(AA^T)^{-1}A)z + A^T(AA^T)^{-1}b \end{aligned}$$

projected subgradient update is (using $Ax^{(k)} = b$)

$$\begin{aligned} x^{(k+1)} &= \Pi(x^{(k)} - \alpha_k g^{(k)}) \\ &= x^{(k)} - \alpha_k(I - A^T(AA^T)^{-1}A)g^{(k)} \\ &= x^{(k)} - \alpha_k \Pi_{\mathcal{N}(A)}(g^{(k)}) \end{aligned}$$

# Example: Least $l_1$-norm

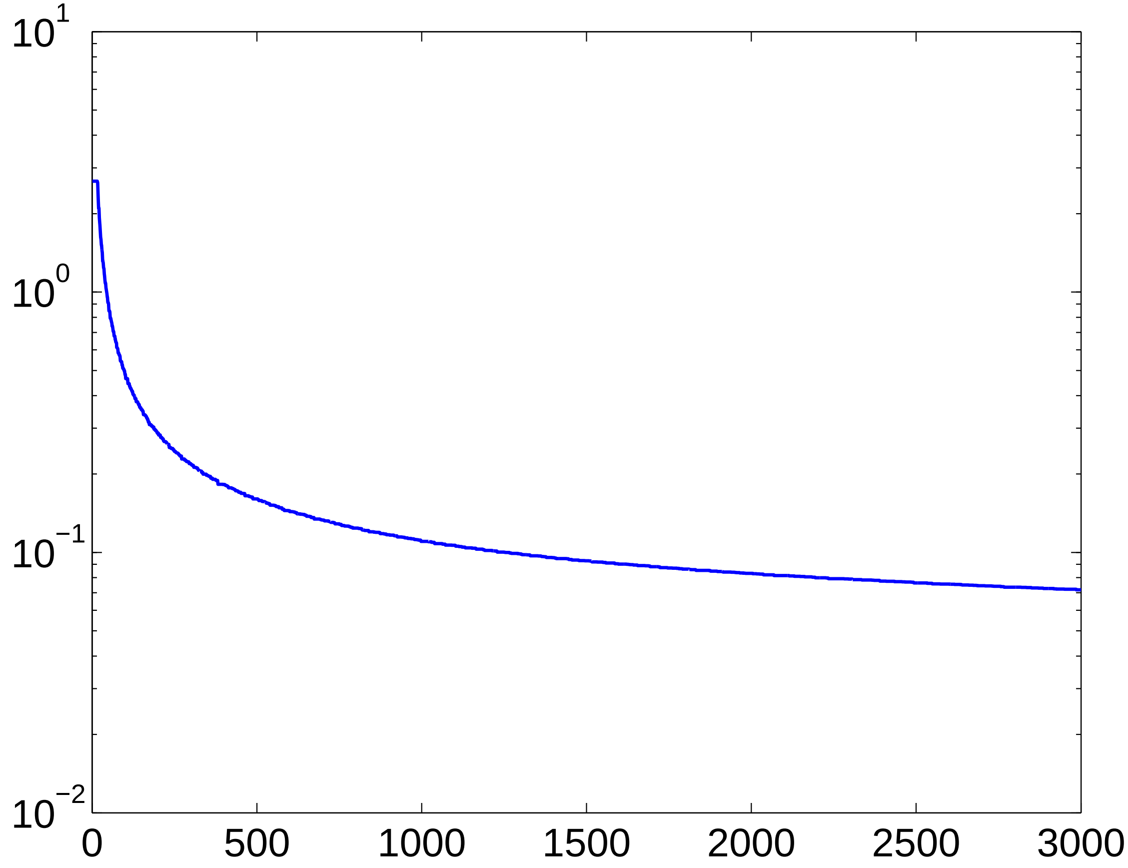
$$
\begin{array}{ll}
\text{minimize} & \|x\|_1 \\
\text{subject to} & Ax = b
\end{array}
$$

subgradient of objective is $g = \mathbf{sign}(x)$

projected subgradient update is

$$
x^{(k+1)} = x^{(k)} - \alpha_k (I - A^T(AA^T)^{-1}A)\,\mathbf{sign}(x^{(k)})
$$

problem instance with $n = 1000$, $m = 50$, step size $\alpha_k = 0.1/k$, $f^\star \approx 3.2$



# Projected subgradient for dual problem

(convex) primal:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \end{array}$$

solve dual problem

$$\begin{array}{ll} \text{maximize} & g(\lambda) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

via projected subgradient method:

$$\lambda^{(k+1)} = \left(\lambda^{(k)} - \alpha_k h\right)_+, \qquad h \in \partial(-g)(\lambda^{(k)})$$

# Subgradient of negative dual function

assume $f_0$ is strictly convex, and denote, for $\lambda \succeq 0$,

$$x^*(\lambda) = \operatorname*{argmin}_z \left(f_0(z) + \lambda_1 f_1(z) + \cdots + \lambda_m f_m(z)\right)$$

so $g(\lambda) = f_0(x^*(\lambda)) + \lambda_1 f_1(x^*(\lambda)) + \cdots + \lambda_m f_m(x^*(\lambda))$

a subgradient of $-g$ at $\lambda$ is given by $h_i = -f_i(x^*(\lambda))$

projected subgradient method for dual:

$$x^{(k)} = x^*(\lambda^{(k)}), \qquad \lambda_i^{(k+1)} = \left(\lambda_i^{(k)} + \alpha_k f_i(x^{(k)})\right)_+$$

- primal iterates $x^{(k)}$ are not feasible, but become feasible in limit (sometimes can find feasible, suboptimal $\tilde{x}^{(k)}$ from $x^{(k)}$)
- dual function values $g(\lambda^{(k)})$ converge to $f^\star = f_0(x^\star)$

interpretation:

- $\lambda_i$ is price for 'resource' $f_i(x)$
- price update $\lambda_i^{(k+1)} = \left(\lambda_i^{(k)} + \alpha_k f_i(x^{(k)})\right)_+$
  - increase price $\lambda_i$ if resource $i$ is over-utilized ($i.e.$, $f_i(x) > 0$)
  - decrease price $\lambda_i$ if resource $i$ is under-utilized ($i.e.$, $f_i(x) < 0$)
  - but never let prices get negative

# Example

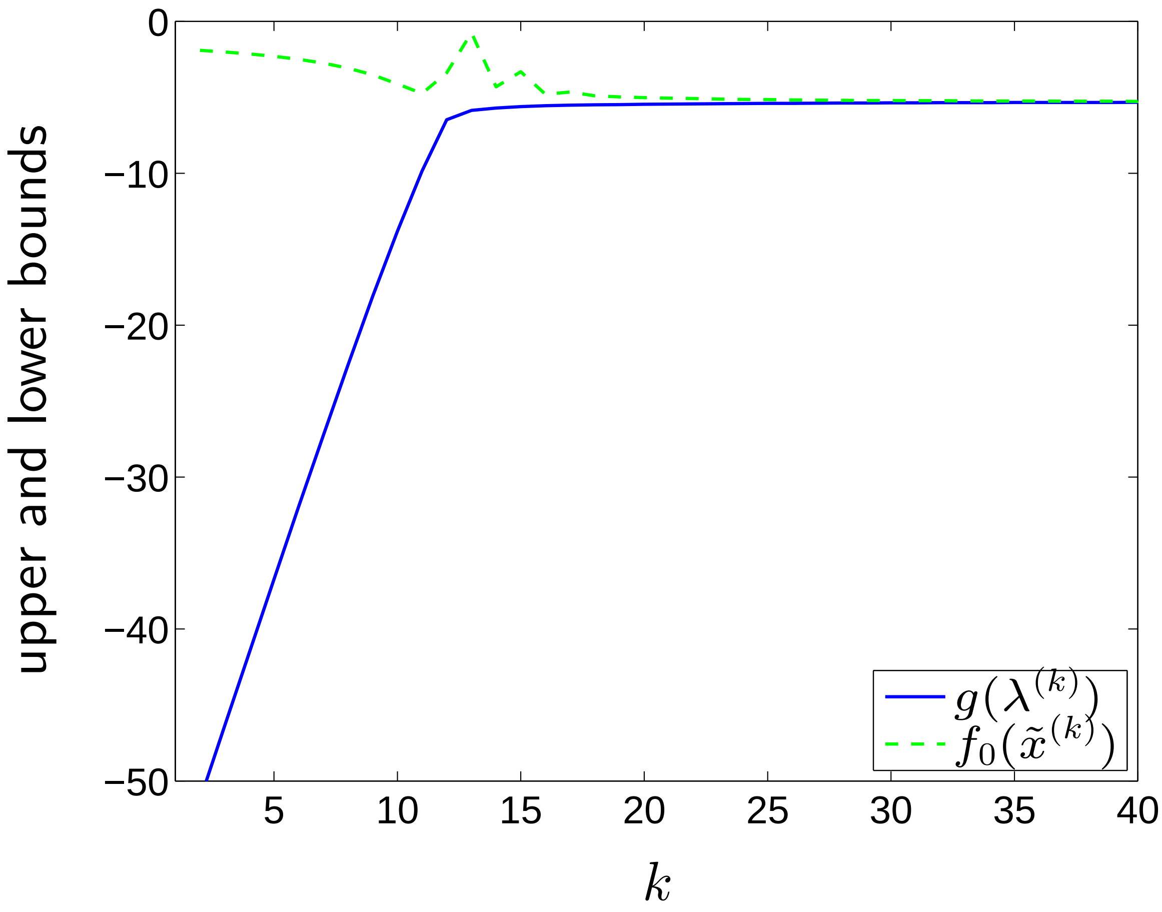
minimize strictly convex quadratic ($P \succ 0$) over unit box:

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x - q^T x \\ \text{subject to} & x_i^2 \leq 1, \quad i = 1, \ldots, n \end{array}$$

- $L(x, \lambda) = (1/2)x^T(P + \mathbf{diag}(2\lambda))x - q^T x - \mathbf{1}^T \lambda$
- $x^*(\lambda) = (P + \mathbf{diag}(2\lambda))^{-1} q$
- projected subgradient for dual:

$$x^{(k)} = (P + \mathbf{diag}(2\lambda^{(k)}))^{-1} q, \quad \lambda_i^{(k+1)} = \left(\lambda_i^{(k)} + \alpha_k((x_i^{(k)})^2 - 1)\right)_+$$

problem instance with $n = 50$, fixed step size $\alpha = 0.1$, $f^\star \approx -5.3$; $\tilde{x}^{(k)}$ is a nearby feasible point for $x^{(k)}$



# Subgradient method for constrained optimization

solves constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m, \end{array}$$

where $f_i : \mathbf{R}^n \to \mathbf{R}$ are convex

same update $x^{(k+1)} = x^{(k)} - \alpha_k g^{(k)}$, but we have

$$g^{(k)} \in \begin{cases} \partial f_0(x) & f_i(x) \leq 0, \quad i = 1, \ldots, m, \\ \partial f_j(x) & f_j(x) > 0 \end{cases}$$

define $f_{\text{best}}^{(k)} = \min\{f_0(x^{(i)}) \mid x^{(i)} \text{ feasible}, \ i = 1, \ldots, k\}$

# Convergence

assumptions:

- there exists an optimal $x^\star$; Slater's condition holds
- $\|g^{(k)}\|_2 \le G$; $\|x^{(1)} - x^\star\|_2 \le R$

**typical result**: for $\alpha_k > 0$, $\alpha_k \to 0$, $\sum_{i=1}^\infty \alpha_i = \infty$, we have $f_{\rm best}^{(k)} \to f^\star$

# Example: Inequality form LP

LP with $n = 20$ variables, $m = 200$ inequalities, $f^\star \approx -3.4$;
$\alpha_k = 1/k$ for optimality step, Polyak's step size for feasibility step