Subgradient Methods for Constrained Problems

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

Prof. S. Boyd, EE364b, Stanford University
Projected subgradient method

solves constrained optimization problem

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in C,
\end{align*}
\]

where \( f : \mathbb{R}^n \rightarrow \mathbb{R} \), \( C \subseteq \mathbb{R}^n \) are convex

**projected subgradient method** is given by

\[
x^{(k+1)} = P(x^{(k)} - \alpha_k g^{(k)}),
\]

\( P \) is (Euclidean) projection on \( C \), 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 $x^*$
**Linear equality constraints**

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

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

\[
P(z) = z - A^T(AA^T)^{-1}(Az - b) = (I - A^T(AA^T)^{-1}A)z + A^T(AA^T)^{-1}b
\]

projected subgradient update is

\[
x^{(k+1)} = P(x^{(k)} - \alpha_k g^{(k)}) = x^{(k)} - \alpha_k P(g^{(k)})
\]

\[
= x^{(k)} - \alpha_k (I - A^T(AA^T)^{-1}A)g^{(k)}
\]

\[
= x^{(k)} - \alpha_k P_N(A)(g^{(k)})
\]
Example: Least $l_1$-norm

minimize $\|x\|_1$
subject to $Ax = b$

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

projected subgradient update is

$$x^{(k+1)} = x^{(k)} - \alpha_k (I - A^T (AA^T)^{-1} A) \text{sign}(x^{(k)})$$
problem instance with $n = 1000$, $m = 50$, step size $\alpha_k = 0.1/k$, $f^* \approx 3.2$
Projected subgradient for dual problem

(convex) primal:

\[
\begin{align*}
&\text{minimize} & f_0(x) \\
&\text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m
\end{align*}
\]

solve dual problem

\[
\begin{align*}
&\text{maximize} & g(\lambda) \\
&\text{subject to} & \lambda \succeq 0
\end{align*}
\]

via projected subgradient method:

\[
\lambda^{(k+1)} = \left( \lambda^{(k)} - \alpha_k h \right)_+, \quad h \in \partial(-g)(\lambda^{(k)})
\]
Subgradient of negative dual function

assume \( f_0 \) is strictly convex, and denote, for \( \lambda \geq 0 \),

\[
x^*(\lambda) = \arg\min_z (f_0(z) + \lambda_1 f_1(z) + \cdots + \lambda_m f_m(z))
\]

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:

\[
x^{(k)} = x^*(\lambda^{(k)}), \quad \lambda^{(k+1)}_i = \left(\lambda^{(k)}_i + \alpha_k f_i(x^{(k)})\right)_+
\]
• primal iterates $x^{(k)}$ are not feasible, but become feasible in limit (sometimes can find feasible, suboptimal $\bar{x}^{(k)}$ from $x^{(k)}$)

• dual function values $g(\lambda^{(k)})$ converge to $f^* = f_0(x^*)$

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 \((P > 0)\) over unit box:

\[
\begin{align*}
& \text{minimize} \quad (1/2)x^TPx - q^Tx \\
& \text{subject to} \quad x_i^2 \leq 1, \quad i = 1, \ldots, n
\end{align*}
\]

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

\[
x^{(k)} = (P + \text{diag}(2\lambda^{(k)}))^{-1}q, \quad \lambda_i^{(k+1)} = \left(\lambda_i^{(k)} + \alpha_k((x_i^{(k)})^2 - 1)\right)_+
\]
problem instance with $n = 50$, fixed step size $\alpha = 0.1$, $f^* \approx -5.3$; $	ilde{x}^{(k)}$ is a nearby feasible point for $x^{(k)}$. 

![Graph showing upper and lower bounds over iterations $k$.]
Subgradient method for constrained optimization

solves constrained optimization problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m,
\end{align*}
\]

where \( f_i : \mathbb{R}^n \rightarrow \mathbb{R} \) are convex

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

\[
g^{(k)} \in \left\{ \begin{array}{ll}
\partial f_0(x) & f_i(x) \leq 0, \quad i = 1, \ldots, m, \\
\partial f_j(x) & f_j(x) > 0
\end{array} \right.
\]

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 $x^*$; Slater’s condition holds
• $\|g^{(k)}\|_2 \leq G; \|x^{(1)} - x^*\|_2 \leq R$

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

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