ACTIVE-SET METHODS

References:

J. Nocedal and S.J. Wright, "*Numerical Optimization*," 2006. Chapter 16

M.S. Bazaraa, H.D. Sherali, C.M. Shetty, "*Nonlinear Programming - Theory and Algorithms*," 2006

LINEAR PROGRAM IN STANDARD FORM

- Consider the **linear program** in **standard form**

$$\begin{aligned} \min \quad & c'x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

- Assumption: $A \in \mathbb{R}^{m \times n}$ has full row rank (this implies $n \geq m$)

- In case the LP is not in standard form, remember that:
  - Inequality constraints $Ax \leq b$ can be transformed into $Ax + z = b$, $z \geq 0$
  - Variables without sign restriction can be split into $x = x^+ - x^-$, with $x^+, x^- \geq 0$
  - We can solve the dual LP (in standard form) with respect to the dual vector λ:

$$\begin{aligned} \min_x \quad & c'x \\ \text{s.t.} \quad & Ax \leq b \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \min_\lambda \quad & b'\lambda \\ \text{s.t.} \quad & A'\lambda = -c, \ \lambda \geq 0 \end{aligned}$$

  and get x^* as the optimal dual vector of the dual LP problem

BASIC SOLUTIONS

- A subset $\mathcal{B} \subseteq \{1, \ldots, n\}$ of exactly m elements is a **basis** and a vector $x \in \mathbb{R}^n$ is a **basic feasible point** (a.k.a. **basic feasible solution**) if
  - $x \geq 0$
  - $x_i = 0$ for all $i \notin \mathcal{B}$
  - the **basis matrix** $B \in \mathbb{R}^{m \times m}$ obtained by collecting the columns A_i of A indexed by $i \in \mathcal{B}$ is nonsingular

THEOREM

- *If the LP is feasible then there exists at least one basic feasible point*
- *If the LP admits optimal solutions then at least one basic feasible point is optimal*
- *If the LP is feasible and bounded then it has a basic feasible optimal solution*

BASIC SOLUTIONS

THEOREM

The basic feasible points are the vertices of the polyhedron $\{x: \ Ax = b, \ x \geq 0\}$.

DEFINITION

A basis $\mathcal{B}$ is **degenerate** if $x_i = 0$ for some $i \in \mathcal{B}$. An LP is degenerate if it has at least one degenerate basis

- The **simplex method** determines the solution of a solvable LP problem in a finite number of iterations, iterating from a vertex of the feasible set (basic feasible) point to an adjacent one

- The KKT conditions of optimality for the LP we considered are

$$
\begin{aligned}
& c + A'\nu - s = 0 \\
& Ax = b \\
& x, s \geq 0 \\
& x_i s_i = 0,\ i = 1, \ldots, n
\end{aligned}
$$

- Given a basis $\mathcal{B}$ and the corresponding basic matrix B, let $\mathcal{N} = \{1, \ldots, n\} \setminus \mathcal{B}$ and N the corresponding matrix of columns A_i indexed by $i \in \mathcal{N}$

- Let x_B be the subvector of x indexed by $\mathcal{B}$ and x_N the subvector indexed by $\mathcal{N}$, and similarly s_B, s_N, c_B, c_N

REVISED SIMPLEX METHOD

- Start from a basic point x, that is $x_N = 0$
- From $Ax = b$ we get $x_B = B^{-1}b$ (this requires solving $Bx_b = b$, e.g., by LU factorization)
- To satisfy complementarity slackness, set $s_B = 0$
- Partition the KKT condition $A'\nu - s = -c$ into

$$\begin{array}{rcl} B'\nu & = & -c_B \\ N'\nu - s_N & = & -c_N \end{array}$$

- Therefore $\nu = -B^{-T}c_B$ and $s_N = c_N - (B^{-1}N)'c_B$ (=**reduced costs**)
- The only missing KKT condition to satisfy is $s_N \geq 0$

REVISED SIMPLEX METHOD

- If $s_N \geq 0$ we have found an optimal solution x. Stop

- Otherwise, we execute a **pivoting** procedure:
  - select an index $q \in N$ such that $s_q < 0$ and make index q enter the basis $\mathcal{B}$
  - increase x_q from 0 while keeping $Ax = b$ satisfied, until another component $x_p = 0, p \in \mathcal{B}$:

$$A_q x_q + B(x_B + \Delta x_B) = b \text{ and } x_B + \Delta x_B \geq 0$$

$$\Rightarrow \quad \Delta x_B = \overbrace{B^{-1}b - x_B}^{Bx_B = b} - B^{-1}A_q x_q = -B^{-1}A_q x_q \geq -x_B$$

$$\Rightarrow \quad [\underbrace{B^{-1}A_q}_{d}]_j x_q \leq [x_B]_j, \ \forall j = 1, \ldots, m$$

  - the index $p = \arg\min_j \left\{ \frac{[x_B]_j}{d_j} \,|\, d_j > 0, \ j = 1, \ldots, m \right\}$ leaves $\mathcal{B}$

- One can prove that $c'x$ is strictly decreasing if $\mathcal{B}$ is nondegenerate

- If the LP is nondegenerate, since the number of possible basis $\mathcal{B}$ is finite the procedure terminates after a finite number of pivoting steps

REVISED SIMPLEX METHOD

- **Initialization**: a basic feasible point is obtained by solving a modified LP, for which a starting basic feasible point is obvious (this is called **phase-1 LP**)

- **degenerate steps** may be encountered in which x_q remains 0 (only $\mathcal{B}$ changes). In this case $c'x$ remains constant

- **cycling** may occur if the same basis $\mathcal{B}$ is encountered again. To prevent this, **anti-cycling strategies** are usually included in the LP solver

- The **dual simplex method** is similar to the revised simplex method. It keeps s feasible rather than x feasible during the iterations

SIMPLEX METHOD FOR LP

- Good LP solvers include a **presolver**, that attempts eliminating variables/constraints to accelerate the subsequent LP solution algorithm

- (Rare) pathological counterexamples exist in which the simplex method visits 2^n vertices, showing that its non-polynomial convergence (Klee, Minty, 1972)

- In practice, usually simplex methods converge in at most $2m$ to $3m$ iterations

- The simplex method is the ancestor of **active set methods** for solving nonlinear programs, such as QP and problems with bound constraints

ACTIVE-SET METHOD FOR NNLS

(Lawson, Hanson, 1974)

- Active-set method to solve the NNLS problem $\min_{x\geq 0} \|Ax-b\|_2^2,\ A\in\mathbb{R}^{m\times n}$

1. $\mathcal{P}\leftarrow\emptyset, x\leftarrow 0$;
2. $w\leftarrow A'(Ax-b)$;
3. **if** $w\geq 0$ **or** $\mathcal{P}=\{1,\ldots,m\}$ **then go to** Step 10;
4. $i\leftarrow\arg\min_{i\in\{1,\ldots,m\}\setminus\mathcal{P}} w_i, \mathcal{P}\leftarrow\mathcal{P}\cup\{i\}$;
5. $y_{\mathcal{P}}\leftarrow\arg\min_{x_{\mathcal{P}}}\|((A')_{\mathcal{P}})'x_{\mathcal{P}}-b\|_2^2$,
   $y_{\{1,\ldots,m\}\setminus\mathcal{P}}\leftarrow 0$;
6. **if** $y_{\mathcal{P}}\geq 0$ **then** $x\leftarrow y$ and **go to** Step 2;
7. $j\leftarrow\arg\min_{h\in\mathcal{P}:\ y_h\leq 0}\left\{\frac{x_h}{x_h-y_h}\right\}$;
8. $x\leftarrow x+\frac{x_j}{x_j-y_j}(y-x)$;
9. $\mathcal{I}\leftarrow\{h\in\mathcal{P}: x_h=0\}, \mathcal{P}\leftarrow\mathcal{P}\setminus\mathcal{I}$; **go to** Step 5;
10. **end**.

The algorithm maintains the primal vector x feasible and keeps switching the active set until the dual variable w is also feasible.

The key step 5 requires solving an unconstrained LS problem. An LDL', Cholesky, or QR factorization of $(A')_{\mathcal{P}}$ can be computed recursively

very simple to solve (750 chars in Embedded MATLAB)

NONNEGATIVE LEAST SQUARES - EXAMPLES

- Solving a **least distance problem** (LDP): (Lawson, Hanson, 1974)

$$x^* = \arg\min \|x\|_2^2 \quad \text{s.t.}\ Ax \le b \qquad \Leftrightarrow \qquad \begin{cases} y^* = \arg\min \left\| \begin{bmatrix} A' \\ b' \end{bmatrix} y + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\|_2^2 \\ \qquad \text{s.t.}\ y \ge 0 \\ x^* = -\frac{A'y^*}{1+b'y^*} \end{cases}$$

- Solving a **quadratic program** (QP) with $Q \succ 0$: (Bemporad, 2016)

$$x^* = \arg\min \tfrac{1}{2}x'Qx + c'x \quad \text{s.t.}\ Ax \le b \qquad \Leftrightarrow \qquad \begin{cases} y^* = \arg\min \left\| \begin{bmatrix} M' \\ d' \end{bmatrix} y + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\|_2^2 \\ \qquad \text{s.t.}\ y \ge 0 \\ x^* = -\frac{(C^{-1})'M'y^*}{1+d'y^*} - Q^{-1}c \end{cases}$$

where $M = A(C^{-1})', CC' = Q$ (Cholesky factorization), $d = b + AQ^{-1}c$

- The LDP/QP is infeasible if and only the residual $r^* = \begin{bmatrix} M' \\ d' \end{bmatrix} y^* + \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ of the corresponding NNLS is zero (by Farkas' lemma)

PARTIALLY NONNEGATIVE LEAST SQUARES

(Bemporad, 2015)

- Consider the **partially nonnegative least squares** (PNNLS) problem

$$\begin{array}{ll} \min_{x,u} & \|Ax + Bu - c\|_2^2 \\ \text{s.t.} & x \geq 0,\ u \text{ free} \end{array} \qquad \begin{array}{l} A \in \mathbb{R}^{m \times n} \\ B \in \mathbb{R}^{m \times p} \end{array}$$

- Let $B^{\#}$ be the **pseudoinverse** of B. In case B has full column rank then $B^{\#} = (B'B)^{-1}B'$

- The PNNLS problem can be solved as the NNLS problem

$$\begin{array}{ll} \min & \|\bar{A}x - \bar{b}\|_2^2 \\ \text{s.t.} & x \geq 0 \end{array}$$

where $\bar{A} = (I - BB^{\#})A, \bar{b} = (I - BB^{\#})c$

PARTIALLY NONNEGATIVE LEAST SQUARES - EXAMPLES

(Bemporad, 2015)

- Computing a feasible point in a polyhedron: A polyhedron $P = \{x : \; Ax \leq b\}$ is nonempty if and only if

$$\begin{array}{rl} 0 = \min_x & \|Ax + y - b\|_2^2 \\ \text{s.t.} & y \geq 0, \; x \text{ free} \end{array}$$

- Solving an LP: The following two problems are equivalent

$$\begin{array}{rl} \min & c'x \\ \text{s.t.} & Ax \leq b \end{array} \quad \Leftrightarrow \quad \begin{array}{rl} \min_x & \left\| \begin{bmatrix} b' & 0 \\ 0 & I \\ A' & 0 \end{bmatrix} \begin{bmatrix} y \\ s \end{bmatrix} + \begin{bmatrix} c' \\ A \\ 0 \end{bmatrix} x - \begin{bmatrix} 0 \\ b \\ -c \end{bmatrix} \right\|_2^2 \\ \text{s.t.} & y, s \geq 0, \; x \text{ free} \end{array}$$

which follows from the optimality conditions $A'y + c = 0$, $Ax + s = b$, and $y'(Ax - b) = 0$, where the latter is equivalent to zero duality gap $c'x = -b'y$

ACTIVE SET METHODS FOR QP

- **Active set methods** for QP are usually the best on small problems because:
  - they provide excellent quality solutions within few iterations
  - are less sensitive to preconditioning (= their behavior is more predictable)
  - they do not require advanced linear algebra libraries

  although they may be less robust than other methods in single precision arithmetic (due to divisions)

- Different active set methods for QP exist. They all work similar to the simplex method, switching the set of active constraints $A_i x = b_i$ until all the KKT conditions are satisfied (Wolfe, 1959) (Lemke, 1962) (Dantzig, 1963) (Fletcher, 1971)

- Most of these methods are equivalent, i.e., visit the same sequence of active sets, although with different linear algebra (Pang, 1983) (Best, 1984)

AN ACTIVE SET METHOD FOR QP

(Bazaraa, Sherali, Shetty, 2006, p. 732)

- We want to solve the following general strictly convex QP

$$\begin{array}{rl} \min & \frac{1}{2}x'Qx + c'x \\ \text{s.t.} & A_ix \leq b_i,\ i \in I \\ & A_ix = b_i,\ i \in E \end{array}$$

  where $I \cup E = \{1, \ldots, m\}$ and $Q = Q' \succ 0, Q \in \mathbb{R}^{n\times n}$

- Assume a feasible starting point x_0 is available (e.g., by solving a phase-1 LP)

- At iteration k, given a feasible x_k, let $I_k = \{i \in I:\ A_ix_k = b_i\}, W_k = I_k \cup E$ be the active set and consider the equality-constrained QP

$$\begin{array}{rl} \min & \frac{1}{2}x'Qx + c'x \\ \text{s.t.} & A_ix = b_i,\ i \in W_k \end{array}$$

AN ACTIVE SET METHOD FOR QP

(Bazaraa, Sherali, Shetty, 2006, p. 732)

- By shifting the coordinates to $d = x - x_k$ the equality-constrained QP becomes

$$d_k = \arg\min \quad \frac{1}{2}d'Qd + (Qx_k + c)'d \qquad \text{s.t.} \quad A_i d = 0,\ i \in W_k$$

$$\Longrightarrow \begin{bmatrix} Q & A'_{W_k} \\ A_{W_k} & 0 \end{bmatrix} \begin{bmatrix} d_k \\ v_k \end{bmatrix} = \begin{bmatrix} -Qx_k - c \\ 0 \end{bmatrix}$$

  providing the best shift from x_k within the null-space of the submatrix A_{W_k}

- If $d_k = 0$:
  - if $v_k \geq 0$ then x_k is the optimal solution, v_k the optimal dual variables corresponding to the active constraints
  - Otherwise, let $q \in W_k$ such that $(v_k)_q$ is the most negative component of v_k and update $I_{k+1} = I_k \setminus \{q\}, W_{k+1} = I_{k+1} \cup E, x_{k+1} = x_k$

AN ACTIVE SET METHOD FOR QP

(Bazaraa, Sherali, Shetty, 2006, p. 732)

- If $d_k \neq 0$:
  - if $A_i(x_k + d_k) \leq b_i$ for all $i \notin W_k$, set $x_{k+1} = x_k + d_k, W_{k+1} = W_k$
  - otherwise choose the maximum step length $\alpha_k < 1$ that maintains feasibility

$$\alpha_k = \min_{i \notin I_k : A_i d_k > 0} \left\{ \frac{b_i - A_i x_k}{A_i d_k} \right\} = \frac{b_q - A_q x_k}{A_q d_k}$$

  and set $x_{k+1} = x_k + \alpha_k d_k, I_{k+1} = I_k \cup \{q\}, W_{k+1} = I_{k+1} \cup E$

- Since at each iteration the objective function is non-increasing, the algorithm terminates in a finite number k of steps

- For more efficiency a factorization of $\begin{bmatrix} Q & A'_{W_k} \\ A_{W_k} & 0 \end{bmatrix}$ can be updated recursively

- The above active-set method maintains feasibility of x_k during the iterations. Other (often more effective) methods maintain the dual vector v_k feasible and stop when the corresponding primal solution x_k is feasible

BLOCK PIVOTING METHODS - KR ALGORITHM

(Kunisch, Rendl, 2003)

- Active set methods only add or remove one constraint at each iteration, which makes them slow for QPs with many constraints/variables

- **Block principal pivoting** methods perform instead simultaneous changes in the working-set in one iteration

- Kunisch and Rendl's (KR) method is an infeasible primal-dual method to solve box-constrained QP quite efficiently

$$\begin{array}{ll} \min & \frac{1}{2}x'Qx + c'x \\ \text{s.t.} & \ell \leq x \leq u \end{array}$$

BLOCK PIVOTING METHODS - KR ALGORITHM

- The algorithm iteratively mass-updates the sets $L, U \subseteq N, N = \{1, \ldots, n\}$ of active lower and upper bounds, starting from an arbitrary initial guess L, U:

1. $A \leftarrow L \cup U, I \leftarrow N \setminus A$
2. $\left[\begin{smallmatrix} z_L \\ z_U \end{smallmatrix}\right] \leftarrow \left[\begin{smallmatrix} \ell_L \\ u_U \end{smallmatrix}\right], z_I \leftarrow -Q_{II}^{-1}(c_I + Q_{IA} z_A)$ — solve unconstrained QP
   $\lambda_I \leftarrow 0, \lambda_A \leftarrow -c_A - Q_{AN} z$ — get λ from KKT
3. $L \leftarrow \{i \in N : z_i < \ell_i$ or $(\lambda_i < 0$ and $i \in L)\}$ — update active set
   $U \leftarrow \{i \in N : z_i > u_i$ or $(\lambda_i > 0$ and $i \in U)\}$
4. if $(L \cup U) = \emptyset$ return $z^* \leftarrow z$, else go to 1

- **Very simple** to implement and **fast** (convergence usually in ≤ 12 steps)
- Convergence is guaranteed only under **restrictive assumptions**. Variants with less restrictive conditions (but slower to execute) exist (Hungerl ander, Rendl, 2015)
- For given *parametric* QP ($c = F\theta + f, \ell = W\theta + w, u = S\theta + s, Q$ fixed) one can **exactly** map the number of iterations KR takes to converge (or cycle) as a function of the parameter $\theta \in \mathbb{R}^m$ (Cimini, Bemporad, 2019)