13.1

Linear constrained NLP algorithms

Katta G. Murty, IOE 611 Lecture slides

The Frank-Wolfe Method (1956)

One of the first algos. developed for constrained NLP. Consider:

$$\min \theta(x)$$
$$\text{s. to} \quad A_{i.}x \begin{cases} = b_i, & i = 1 \text{ to } m \\ \geq b_i, & i = m+1 \text{ to } m+p \end{cases}$$

Let K denote set of feasible sols.

Assumptions: We assume that K has at least one extreme point. Also, for each $\bar{x} \in K$, assume that $\nabla\theta(\bar{x})x$ is bounded below over K.

The Method: Initiate with any $x^0 \in K$.

When x^r is current pt.:

Step 1: Solve the LP: $\min \nabla\theta(x^r)x$ over $x \in K$.

If x^r is optimal to this LP, then x^r is optimal to original

NLP if $\theta(x)$ is convex, and it is a KKT pt. to original NLP whether $\theta(x)$ is convex or not. Terminate.

Otherwise, let z^r be an opt. extreme pt. sol. for this LP. Go to Step 2.

Step 2: So, $y^r = z^r - x^r$ is a feasible descent direction at x^r. Do a line search to find min $\theta(x^r + \lambda y^r)$ over $0 \leq \lambda \leq 1$.

If λ_r is the step length, $x^{r+1} = x^r + \lambda_r y^r$ is next pt., go to next iteration with it.

Theorem: If method does not terminate finitely, it generates a descent sequence s. th. every limit pt. of this sequence is a KKT pt.

Theorem: If $\theta(x)$ is convex, and when x^r is current pt. $\nabla\theta(x^r)(x^r - z^r) \leq \epsilon$, then x^r is ϵ-opt. to original NLP.

Work in each step is an LP and a line search. Too much. Also method has slow convergence. Practical only if LP in each step can be solved by a highly efficient special method.

Traffic Assignment Application: **Input:** $G = (\mathcal{N}, \mathcal{A})$, city's street network, directed.

(s_u, t_u) an O-D pair with estimated volume V^u vehicles/unit time, $u = 1$ to g.

Arc travel time functions: For each arc (i, j), $c_{ij}(f_{ij})$ = travel time for travelling arc (i, j) if f_{ij} is traffic flow on this arc/unit time. $c_{ij}(f_{ij})$ is $\uparrow +\infty$ with f_{ij}.

Desired Output: How will traffic distribute itself? i.e., find flows $f^u = (f^u_{ij}) : u = 1$ to g which minimizes total travel time of all vehicles.

Can be formulated as a multicommodity flow to min $\Sigma\,\Sigma\, c_{ij}(f^u_{ij})$.

FW is suitable to solve this because LPs in each iteration become shortest chain problems for which there are very efficient special algos.

The Gradient Projection Method:

J B Rosen (1960).

Theorem: Consider: min $\theta(x)$ s. to $Dx = d$ where $D_{m \times n}$ has rank m.

Let $(\bar{x}, \bar{\pi})$ be an opt. pair for this problem, and suppose i is s. that $\bar{\pi}_i < 0$. Then there exists a descent feasible direction for the problem:

$$\min \theta(x)$$
$$\text{s. to} \quad D_{t.}x \quad \begin{cases} = b_t, & t = 1 \text{ to } m, t \neq i \\ \geq b_i, & \text{for } t = i \end{cases}$$

at $\bar{x}$ which moves off the constraint $A_{i.}x = b_i$.

The G. P. method generates a descent sequence $\{x^r\}$ of feasible points beginning with an initial feasible sol. x^0.

In each step, instead of solving an LP to get a descent feasble direction at current pt., it obtains it by projecting the negative gradient direction on the subspace of active constraints at current pt.

When x^r is current pt., let $B(x^r)$ denote the index set of active inequality constraints at it.

If there are no active constraints at x^r, choose $y^r = -(\nabla\theta(x^r))^T$.

If there are active constraints at x^r, let A_r denote the matrix with rows $A_{i.}, i \in \{1, \ldots, m\} \cup B(x^r)$.

Assume that A_r is of full row rank, otherwise delete some dependent row vectors from A_r until it becomes of full row rank

.

Projection matrix corresponding to active subspace is $P_r = I - A_r^T(A_r A_r^T)^{-1}A_r$

Projection of $-(\nabla\theta(x^r))^T$ is $\eta^r = -P_r(\nabla\theta(x^r))^T$. η^r is a positive multiple of opt. sol. of: $\quad \min \nabla\theta(x^r)y \quad$ s. to $\quad A_r y = 0$ and $y^t y \leq 1$.

If $\eta^r \neq 0$, it is a descent direction at x^r, find $\bar{\lambda}$, the maximum step length in this direction in the feasible region. Then solve the line search problem: min $\theta(x^r + \lambda\eta^r)$, $0 \leq \lambda \leq \bar{\lambda}$, and if λ_r is the opt. step length for it, take $x^{r+1} = x^r + \lambda_r \eta^r$ and go to the next step.

If $\eta^r = 0$, let $\beta^r = (A_r A_r^T)^{-1} A_r (\nabla\theta(x^r))^T$. Augment $(\beta^r)^T$ into a row vector of order $m+p$ by inserting in it 0's for all $i \in \{m+1, \ldots, m+p\} \backslash B(x^r)$, and call it π^r.

Then $\nabla\theta(x^r) = \pi^r A$. So, if $\pi_i^r \geq 0 \forall i \in \{m+1, \ldots, m+p\}$, x^r, π^r together satisfy the KKT conds, terminate.

If $\pi_i^r < 0$ for some $i \in \{m+1, \ldots, m+p\}$, identify the most negative among $\pi_{m+1}^r, \ldots, \pi_{m+p}^r$, and if it is π_t^r, delete the row $A_{t.}$ from the active constraint matrix A_r and repeat the whole process with the new matrix.

How to update the projection matrix?

To delete a row from A_r

Suppose row $A_{t.}$ is the sth row in A_r. To delete it from A_r, let $\hat{A}$ denote the resulting matrix.

In $(A_r A_r^T)^{-1}$ interchange the last row and sth row, and then the last col. and sth col. After these interchanges, suppose this inverse is $\begin{pmatrix} E & u \\ u^T & \delta \end{pmatrix}$.

Then $(\hat{A}\hat{A}^T)^{-1} = E - \frac{uu^T}{\delta}$.

To add a row to A_r

Let P_r be the projection matrix corresponding to A_r. Suppose we want to incude the new row vector $A_{t.}$ in A_r. It will be included as last row, let resulting matrix be $\tilde{A}$.

Let $\gamma = A_{t.} P_r (A_{t.})^T$. If $\gamma = 0$, $A_{t.}$ is linearly dependent on rows in A_r, and hence cannot be included in A_r, i.e., continue method with same A_r as active constraint row matrix.

If $\gamma \neq 0$, then $(\tilde{A}\tilde{A}^T)^{-1} = \begin{pmatrix} F & u \\ u^T & 1/\gamma \end{pmatrix}$

where $w = (A_r A_r^T)^{-1} A_r (A_{t.})^T, u = -(w/\gamma)$,

$F = (A_r A_r^T)^{-1} + \frac{ww^T}{\gamma}$.

Show that the Simplex algo. for LP can be viewed as a G. P. method.

Primal Active Set Methods

They handle inequalities using techniques for solving linear equality constrained problems iteratively. They guess the active inequalities at Optimum and apply equality constrained methods treating these inequalities as eqs. Modifications to this active set are made using the Lagrange multiplier vectors, based on theorem discussed earlier.

Since objective function nonlinear, no. of active constraints may be m_1 $(0 \leq m_1 \leq n)$ (in simplex algo. for LP it is n).

$\mathcal{A}$ = index set of working active set. $\{1, \ldots, m\} \subset \mathcal{A}$ always, and $\{A_{i.} : i \in \mathcal{A}\}$ is held l.i. Method adjusts $\mathcal{A}$ to identify correct active constraints at optimum.

Initially $\mathcal{A}$ = active constraints at x^0, or a maximal l.i. subset of them.

When current pt. is x^r and working active set is $\mathcal{A}$, **degeneracy** occurs if a constraint not in $\mathcal{A}$ is active at x^r. In this case, step lenghts choosen later may be 0, and algo. can cycle by returning to a previous active set in sequence.

Step 1: Find descent direction at x^r for EP (equality problem treating all constraints in $\mathcal{A}$ as eqs. and ignoring others)

If x^r satisfies term. conds. for this EP, let β^r be the Lagrange multiplier vector for it . If $\beta^r \geq 0$, $\forall i \in \mathcal{A} \cap \{m+1, \ldots, m+p\}$, augment β^r into π^r by inserting 0's $\forall i \notin \mathcal{A}$. Then x^r, π^r is a KKT pair for original problem, terminate.

If $\beta_i^r < 0$ for some $i \in \mathcal{A} \cap \{m+1, \ldots, m+p\}$, let β_t^r be the most negative among them, delete t from $\mathcal{A}$, get the new EP and repeat.

If x^r does not satisfy term. conds. for EP, let η^r be the search direction at x^r for the EP. Fine $\bar{\lambda}$, the max. step length that keeps $x^r + \lambda\eta^r$ feasible to original problem. Do a line search to: min $\theta(x^r + \lambda\eta^r)$, $0 \leq \lambda \leq \bar{\lambda}$. Let λ_r be opt. step length for this problem.

If $\lambda_r < \bar{\lambda}$, leave $\mathcal{A}$ as it is, and with $x^{r+1} = x^r + \lambda_r\eta^r$, go to next iteration.

If $\lambda_r = \bar{\lambda}$, a new constraint becomes active. It is the i which attains the min in definition of $\bar{\lambda}$, include it in $\mathcal{A}$, and with $x^{r+1} = x^r + \lambda_r \eta^r$, go to next iteration.

The Reduced gradient Method

P. Wolfe (1963). Consider problem in form: $\min \theta(x)$ s. to $Ax = b$, $\ell \leq x \leq u$; where $A_{m \times n}$ has rank m.

Let $\bar{x}$ be current feasible sol. and B a basis for A (usually the one corresponding to the largest components in $\bar{x}$), with $(B \vdots D)$ the basic, nonbasic partition of A. $\bar{x} = (\bar{x}_B, \bar{x}_D)$. So, $\bar{x}_B = B^{-1}(b - D\bar{x}_D)$.

Problem can be transformed into one in space of independent variables x_D only. The **reduced gradient** at $\bar{x}$ in this space is:

$$\bar{c}_D = (\nabla_{x_D}\theta(\bar{x}) - (\nabla_{x_B}\theta(\bar{x}))B^{-1}D$$

Define the direction $\bar{y}_D = (\bar{y}_j)$ in the space of independent variables to be:

$$\bar{y}_j = \begin{cases} -\bar{c}_j & \text{if either } \bar{c}_j < 0 \ \& \ \bar{x}_j < u_j; \text{ or } \bar{c}_j > 0 \ \& \ \bar{x}_j > \ell_j \\ 0 & \text{if above conds. not met} \end{cases}$$

If $\bar{y}_D = 0$, $\bar{x}$ is a KKT pt., terminate.

If $\bar{y}_D \neq 0$, $\bar{c}_D\bar{y}_D < 0$, so $\bar{y}_D$ is a descent direction at $\bar{x}_D$ in the space of independent variables, it is the negative reduced gradient direction. Define $\bar{y}_B = -B^{-1}D\bar{y}_D$ and let $\bar{y} = (\bar{y}_B, \bar{y}_D)$. $\bar{y}$ is the search direction at $\bar{x}$. $A\bar{y} = 0$, so equality conds. continue to hold when we move in this direction at $\bar{x}$.

Find $\bar{\lambda}$ = max step length that you can move in this direction at $\bar{x}$ while continuing to satisfy the bounds on vars.

If $\bar{\lambda} > 0$, do a line search to: min $\theta(\bar{x} + \lambda\bar{y})$, $0 \leq \lambda \leq \bar{\lambda}$. Let λ_1 be opt. step length for this problem. Repeat with $\bar{x} + \lambda_1\bar{y}$ as new current pt.

If $\bar{\lambda} = 0$ (this happens due to degeneracy), $\bar{y}$ is a descent but not feasible direction at $\bar{x}$. Identify active constraints at $\bar{x}$, and carry out a G. P. step. Let $\bar{y}_p$ be the orthogonal projection of $\bar{y}$ in the subspace of active constraints at $\bar{x}$. Now carry out a line search step in the direction $\bar{y}_p$ instead of $\bar{y}$, and go to next step.

In actual implementations, they normally partition the nonbasic variables into *superbasic*, and other variables. The superbasic are the most attractive nonbasic variables at this stage to change, based on their reduced gradient coeffs. In defining $\bar{y}_D$, $\bar{y}_j$ is fixed at 0 for other nonbasic variables, and defined as above only for superbasic variables, and the rest of the step is carried out exactly as above. By proper selection of superbasic variables, this strategy was observed to improve the performance of the algo.