MS&E 310 Course Project III: First-Order Potential or Barrier Reduction for Linear Programming

Yinyu Ye

November 5, 2017

1 Convex optimization over the simplex constraint

We consider the following optimization problem over the simplex:

$$\begin{array}{ll} \text{Minimize} & f(\mathbf{x}) \\ \text{Subject To} & \mathbf{e}^T\mathbf{x} = n;\ \mathbf{x} \geq \mathbf{0}, \end{array} \tag{1}$$

where $\mathbf{e}$ is the vector of all ones. This problem is to minimize a nonlinear function with a Simplex constraint. Such a problem in considered in [7], where function $f(\mathbf{x})$ does not need to be convex and a FPTAS algorithm was developed for computing an approximate KKT point of general quadratic programming. The following algorithm and analysis resemble those in [7].

We assume that $f(\mathbf{x})$ is a convex function in $\mathbf{x} \in R^n$ and $f(\mathbf{x}^*) = 0$ where $\mathbf{x}^*$ is a minimizer of the problem. Furthermore, we make a standard Lipschitz assumption such that

$$f(\mathbf{x}+\mathbf{d}) - f(\mathbf{x}) \leq \nabla f(\mathbf{x})^T\mathbf{d} + \frac{\gamma}{2}\|\mathbf{d}\|^2,$$

where positive γ is the Lipschitz parameter. Note that any homogeneous linear feasibility problem, e.g., the canonical Karmarkar form in [2]:

$$\begin{aligned} A\mathbf{x} &= \mathbf{0}; \\ \mathbf{e}^T\mathbf{x} &= n; \\ \mathbf{x} &\geq \mathbf{0}. \end{aligned}$$

can be formulated as the model with $f(\mathbf{x}) = \frac{1}{2}\|A\mathbf{x}\|^2$ and γ as the half of the largest eigenvalue of matrix A^TA.

Furthermore, any linear programming problem in the standard form and its dual

$$\begin{array}{llll} \text{Minimize} & \mathbf{c}^T\mathbf{x} & \text{Maximize} & \mathbf{b}^T\mathbf{y} \\ \text{Subject to} & A\mathbf{x} = \mathbf{b};\ \mathbf{x} \geq \mathbf{0}; & \text{Subject to} & A^T\mathbf{y} + \mathbf{s} = \mathbf{c};\ \mathbf{s} \geq \mathbf{0} \end{array}$$

can be represented as a homogeneous linear feasibility problem (Ye et al. [5]):

$$\begin{aligned} A\mathbf{x} - \mathbf{b}\tau &= 0; \\ -A^T\mathbf{y} - \mathbf{s} + \mathbf{c}\tau &= 0; \\ \mathbf{b}^T\mathbf{y} - \mathbf{c}^T\mathbf{x} - \kappa &= 0; \\ \mathbf{e}^T\mathbf{x} + \mathbf{e}^T\mathbf{s} + \tau + \kappa &= 2n + 2; \\ (\mathbf{x}, \mathbf{s}, \tau, \kappa) &\geq \mathbf{0}. \end{aligned}$$

We consider the potential function (e.g., see [2, 4, 1, 6])

$$\phi(\mathbf{x}) = \rho \ln(f(\mathbf{x})) - \sum_j \ln(x_j),$$

(alternatively, one may consider barrier function $b_\mu(\mathbf{x}) = f(\mathbf{x}) - \mu \sum_j \ln(x_j)$ for a small fixed μ)

where $\rho \geq n$ over the simplex. Clearly, if we start from $\mathbf{x}^0 = \mathbf{e}$, the analytic center of the simplex, and generate a sequence of points $\mathbf{x}^k$, $k = 1, ...,$, whose potential value is strictly decreased, then when

$$\phi(\mathbf{x}^k) - \phi(\mathbf{x}^0) \leq -\rho \ln(1/\epsilon),$$

we must have

$$\rho \ln(f(\mathbf{x}^k)) - \rho \ln(f(\mathbf{x}^0)) \leq -\rho \ln(1/\epsilon)$$

or

$$\frac{f(\mathbf{x}^k)}{f(\mathbf{x}^0)} \leq \epsilon.$$

This is because on the simplex

$$\sum_j \ln(x_j^k) \leq \sum_j \ln(x_j^0), \forall k = 1, ...$$

We now describe a first order steepest descent potential reduction algorithm in the next section.

2 Steepest-Descent Potential Reduction and Complexity Analysis

Note that the gradient vector of the potential function of $x > 0$ is

$$\nabla\phi(\mathbf{x}) = \frac{\rho}{f(\mathbf{x})} \nabla f(\mathbf{x}) - X^{-1}\mathbf{e}.$$

where in this note X denotes the diagonal matrix whose diagonal entries are elements of vector $\mathbf{x}$.

The following lemma is well known in the literature of interior-point algorithms ([2, 1, 6]):

Lemma 1. *Let* $\mathbf{x} > \mathbf{0}$ *and* $\|X^{-1}\mathbf{d}\| \leq \beta < 1$*. Then*

$$-\sum_j \ln(x_j + d_j) + \sum_j \ln(x_j) \leq -\mathbf{e}^T X^{-1}\mathbf{d} + \frac{\beta^2}{2(1-\beta)}.$$

Lemma 2. *For any* $\mathbf{x} > \mathbf{0}$ *and* $\mathbf{x} \neq \mathbf{x}^*$*, a matrix* $A \in R^{m \times n}$ *with* $A\mathbf{x} = A\mathbf{x}^*$*, and a vector* $\bar{\lambda} \in R^m$*, consider vector*

$$\mathbf{p}(\mathbf{x}) = X\left(\nabla\phi(\mathbf{x}) - A^T\bar{\lambda}\right).$$

Then,

$$\|\mathbf{p}(\mathbf{x})\| \geq 1.$$

Proof. First,

$$\mathbf{p}(\mathbf{x}) = X\left(\frac{\rho}{f(\mathbf{x})}\nabla f(\mathbf{x}) - X^{-1}\mathbf{e} - A^T\bar{\lambda}\right) = \frac{\rho}{f(\mathbf{x})}X\left(\nabla f(\mathbf{x}) - \frac{f(\mathbf{x})}{\rho}A^T\bar{\lambda}\right) - \mathbf{e}.$$

If any entry of $(\nabla f(\mathbf{x}) - \frac{f(\mathbf{x})}{\rho}A^T\bar{\lambda})$ is equal or less than 0, then $\|\mathbf{p}(\mathbf{x})\| \geq \|\mathbf{p}(\mathbf{x})\|_\infty \geq 1$. On the other hand, if $\left(\nabla f(\mathbf{x}) - \frac{f(\mathbf{x})}{\rho}A^T\bar{\lambda}\right) > \mathbf{0}$, we have $\left(\nabla f(\mathbf{x}) - \frac{f(\mathbf{x})}{\rho}A^T\bar{\lambda}\right)^T \mathbf{x}^* \geq 0$. Then, from convexity and $A\mathbf{x} = A\mathbf{x}^*$,

$$f(\mathbf{x}^*) - f(\mathbf{x}) \geq \nabla f(\mathbf{x})^T(\mathbf{x}^* - \mathbf{x}) = \left(\nabla f(\mathbf{x}) - \frac{f(\mathbf{x})}{\rho}A^T\bar{\lambda}\right)^T(\mathbf{x}^* - \mathbf{x}).$$

Thus, from $f(\mathbf{x}^*) = 0$

$$f(\mathbf{x}) \leq \left(\nabla f(\mathbf{x}) - \frac{f(\mathbf{x})}{\rho}A^T\bar{\lambda}\right)^T\mathbf{x}.$$

Furthermore,

$$\begin{aligned}
\|\mathbf{p}(\mathbf{x})\|^2 &= \tfrac{\rho^2}{f(\mathbf{x})^2}\|X\left(\nabla f(\mathbf{x}) - \tfrac{f(\mathbf{x})}{\rho}A^T\bar{\lambda}\right)\|^2 - 2\tfrac{\rho}{f(\mathbf{x})}\left(\nabla f(\mathbf{x}) - \tfrac{f(\mathbf{x})}{\rho}A^T\bar{\lambda}\right)^T\mathbf{x} + n \\
&\geq \tfrac{\rho^2}{n\cdot f(\mathbf{x})^2}\|X\left(\nabla f(\mathbf{x}) - \tfrac{f(\mathbf{x})}{\rho}A^T\bar{\lambda}\right)\|_1^2 - 2\tfrac{\rho}{f(\mathbf{x})}\left(\nabla f(\mathbf{x}) - \tfrac{f(\mathbf{x})}{\rho}A^T\bar{\lambda}\right)^T\mathbf{x} + n \\
&\geq \tfrac{\rho^2}{n}\left(\tfrac{\left(\nabla f(\mathbf{x}) - \frac{f(\mathbf{x})}{\rho}A^T\bar{\lambda}\right)^T\mathbf{x}}{f(\mathbf{x})}\right)^2 - 2\rho\left(\tfrac{\left(\nabla f(\mathbf{x}) - \frac{f(\mathbf{x})}{\rho}A^T\bar{\lambda}\right)^T x}{f(\mathbf{x})}\right) + n \\
&= \tfrac{(\rho z)^2}{n} - 2\rho z + n = \tfrac{1}{n}(\rho z - n)^2,
\end{aligned}$$

where

$$z = \frac{\left(\nabla f(\mathbf{x}) - \frac{f(\mathbf{x})}{\rho}A^T\bar{\lambda}\right)^T\mathbf{x}}{f(\mathbf{x})} \geq 1.$$

The above quadratic function of z has the minimizer at $z = 1$ if $\rho \geq n$, so that

$$\frac{1}{n}(\rho z - n)^2 \geq \frac{1}{n}(\rho - n)^2 \geq 1$$

for $\rho \geq n + \sqrt{n}$. □

For any given $\mathbf{x} > \mathbf{0}$ in the simplex and any $\mathbf{d}$ with $\mathbf{e}^T\mathbf{d} = 0$,

$$f(\mathbf{x}+\mathbf{d}) - f(\mathbf{x}) \leq \nabla f(\mathbf{x})^T\mathbf{d} + \frac{\gamma}{2}\|\mathbf{d}\|^2 \leq \nabla f(\mathbf{x})^T\mathbf{d} + \frac{\gamma}{2}\|XX^{-1}\mathbf{d}\|^2 \leq \nabla f(\mathbf{x})^T\mathbf{d} + \frac{\gamma}{2}\|X^{-1}\mathbf{d}\|^2,$$

where the last inequality is due to $\|X\| \le 1$. Let $\|X^{-1}\mathbf{d}\| = \beta < 1$ and $\mathbf{x}^+ = \mathbf{x} + \mathbf{d} = X(\mathbf{e} + X^{-1}\mathbf{d}) > \mathbf{0}$. Then, from Lemma 1

$$\begin{aligned}\phi(\mathbf{x}^+) - \phi(\mathbf{x}) &\le \rho \ln\left(1 + \tfrac{\nabla f(\mathbf{x})^T \mathbf{d} + \frac{\gamma}{2}\|X^{-1}\mathbf{d}\|^2}{f(\mathbf{x})}\right) - \mathbf{e}^T X^{-1}\mathbf{d} + \tfrac{\beta^2}{2(1-\beta)} \\ &\le \rho \tfrac{\nabla f(\mathbf{x})^T \mathbf{d} + \frac{\gamma}{2}\|X^{-1}\mathbf{d}\|^2}{f(\mathbf{x})} - \mathbf{e}^T X^{-1}\mathbf{d} + \tfrac{\beta^2}{2(1-\beta)} \\ &= \nabla\phi(\mathbf{x})^T \mathbf{d} + \tfrac{\rho\gamma}{2f(\mathbf{x})}\beta^2 + \tfrac{\beta^2}{2(1-\beta)}.\end{aligned}$$

The first order steepest descent potential reduction algorithm would update $\mathbf{x}$ by solving

$$\begin{array}{ll}\text{Minimize} & \nabla\phi(\mathbf{x})^T \mathbf{d} \\ \text{Subject to} & \mathbf{e}^T\mathbf{d} = 0,\ \|X^{-1}\mathbf{d}\| \le \beta;\end{array} \tag{2}$$

or

$$\begin{array}{ll}\text{Minimize} & \nabla\phi(\mathbf{x})^T X\mathbf{d}' \\ \text{Subject to} & \mathbf{e}^T X\mathbf{d}' = 0,\ \|\mathbf{d}'\| \le \beta;\end{array}$$

where parameter $\beta < 1$ is yet to be determined.

Let the scaled gradient projection vector

$$\mathbf{p}(\mathbf{x}) = \left(I - \frac{1}{\|\mathbf{x}\|^2} X\mathbf{e}\mathbf{e}^T X\right) X\nabla\phi(\mathbf{x}) = X\left(\frac{\rho}{f(\mathbf{x})}\left(\nabla f(\mathbf{x}) - \mathbf{e}\cdot\lambda(\mathbf{x})\right)\right) - \mathbf{e},$$

where

$$\lambda(\mathbf{x}) = \frac{\mathbf{e}^T X^2 \nabla\phi(\mathbf{x}) \cdot f(\mathbf{x})}{\|\mathbf{x}\|^2 \cdot \rho}.$$

Then the minimizer of problem (2) would be

$$\mathbf{d} = -\frac{\beta}{\|p(\mathbf{x})\|} X\mathbf{p}(\mathbf{x}),$$

and

$$\nabla\phi(\mathbf{x})^T \mathbf{d} = -\frac{\beta}{\|\mathbf{p}(\mathbf{x})\|}\|\mathbf{p}(\mathbf{x})\|^2 = -\beta\|\mathbf{p}(\mathbf{x})\| \le -\beta,$$

since $\|\mathbf{p}(\mathbf{x})\| \ge 1$ based on Lemma 2.

Thus,

$$\phi(\mathbf{x}^+) - \phi(\mathbf{x}) \le -\beta + \frac{\rho\gamma}{2f(\mathbf{x})}\beta^2 + \frac{\beta^2}{2(1-\beta)}$$

For $\beta \le 1/2$, the above quantity is less than

$$-\beta + \left(2 + \frac{\rho\gamma}{f(\mathbf{x})}\right)\beta^2/2.$$

Thus, one can choose β to minimize the quantity at

$$\beta = \frac{1}{2 + \frac{\rho\gamma}{f(\mathbf{x})}} \le 1/2$$

so that

$$\phi(\mathbf{x}^+) - \phi(\mathbf{x}) \le \frac{-f(\mathbf{x})}{2(f(\mathbf{x}) + 2\rho\gamma)}.$$

One can see that the larger value of $f(\mathbf{x})$, the greater reduction of the potential function.

Starting from $x^0 = \frac{1}{n}e$, we iteratively generate x^k, $k = 1, ...,$ such that

$$\phi(\mathbf{x}^{k+1}) - \phi(\mathbf{x}^k) \le \frac{-f(\mathbf{x}^k)}{2(f(\mathbf{x}^k) + 2\rho\gamma)} \le \frac{-f(\mathbf{x}^k)}{2(f(\mathbf{x}^0) + 2\rho\gamma)} \le \frac{-f(\mathbf{x}^k)}{4\max\{f(\mathbf{x}^0), 2\rho\gamma\}}.$$

The second inequality is due to $f(\mathbf{x}^k) < f(\mathbf{x}^0)$ from $\phi(\mathbf{x}^k) < \phi(\mathbf{x}^0)$ for all $k \ge 1$ and $\mathbf{x}^0$ is the analytic center of the simplex.

Thus, if $\frac{f(\mathbf{x}^k)}{f(\mathbf{x}^0)} \ge \epsilon$ for $1 \le k \le K$, we must have

$$\phi(\mathbf{x}^0) - \phi(\mathbf{x}^K) \le \rho \ln(\frac{1}{\epsilon}),$$

so that

$$\sum_{k=1}^{K} \frac{f(\mathbf{x}^k)}{4\max\{f(\mathbf{x}^0), 2\rho\gamma\}} \le \rho \ln(\frac{1}{\epsilon})$$

or

$$K\epsilon f(\mathbf{x}^0) \le 4\max\{f(\mathbf{x}^0), 2\rho\gamma\}\rho \ln(\frac{1}{\epsilon}).$$

Note that $\rho = n + \sqrt{n} \le 2n$. We conclude

Theorem 3. *The steepest descent potential reduction algorithm generates a $\mathbf{x}^k$ with $f(\mathbf{x}^k)/f(\mathbf{x}^0) \le \epsilon$ in no more than*

$$4(n + \sqrt{n})\frac{\max\{1, 2(n + \sqrt{n})\gamma/f(\mathbf{x}^0)\}}{\epsilon} \ln(\frac{1}{\epsilon})$$

steps.

3 Extension, Implementation and Possible Further Analysis

Question 1: Develop a similar analysis for solving

$$\begin{array}{ll} \text{Minimize} & f(\mathbf{x}) \\ \text{Subject To} & 0 \le x_j \le 2, \ \forall j = 1, ..., n, \end{array} \tag{3}$$

where we start $\mathbf{x}^0 = \mathbf{e}$, the analytic center of the BOX constraint.

Question 2: Implement the algorithm and perform numerical tests to solve for

$$f(\mathbf{x}) = \frac{1}{2}\|A\mathbf{x}\|^2$$

either in (1), or (3), or both.

Question 3: Implement the algorithm and perform numerical tests to solve for

$$f(\mathbf{x}) = \frac{1}{2}\|(AA^T)^{-1/2}A\mathbf{x}\|^2,$$

and compare the performance with that in Question 2. This can be viewed as one-time preconditioning.

Question 4: Test your implementation on homogeneous and self LP models for various linear programs (feasible or infeasible), where you may eliminate free variables $\mathbf{y}$ from the formulation.

4 Extension to MDP

Consider the MDP problem

$$\begin{array}{lrcl} \text{maximize}_{\mathbf{y}} & \sum_{i=1}^m y_i & & \\ \text{subject to} & y_1 - \gamma \mathbf{p}_j^T \mathbf{y} & \leq & c_j, \; j \in \mathcal{A}_1 \\ & & \vdots & \\ & y_i - \gamma \mathbf{p}_j^T \mathbf{y} & \leq & c_j, \; j \in \mathcal{A}_i \\ & & \vdots & \\ & y_m - \gamma \mathbf{p}_j^T \mathbf{y} & \leq & c_j, \; j \in \mathcal{A}_m. \end{array}$$

One can construct a potential/barrier function for a small fixed μ as

$$b_\mu(\mathbf{y}) = -\mathbf{e}^T \mathbf{y} - \mu \sum_j \log(c_j - y_i + \gamma \mathbf{p}_j^T \mathbf{y}),$$

or

$$\psi(\mathbf{y}) = \rho \log(z - \mathbf{e}^T \mathbf{y}) - \sum_j \log(c_j - y_i + \gamma \mathbf{p}_j^T \mathbf{y})$$

where $\rho \geq n$ and z is a upper bound on the maximal value of the MDP problem..

Question 5: The problem becomes a unconstrained problem when start $\mathbf{y}^0 = -\Delta \mathbf{e}$ (in the interior of the feasible region) for a big enough Δ. You may apply the (stochastic) steepest descent method, the conjugate gradient method, the BFGS method, or any deep-learning method, etc, and do numerical experiments. The stochastic gradient would be sample some log terms in the summation and sum up their gradient vectors.

References

[1] C. C. Gonzaga, Polynomial affine algorithms for linear programming, *Math. Programming* 49 (1990) 7–21.

[2] N. Karmarkar, A new polynomial-time algorithm for linear programming, *Combinatorica* 4 (1984) 373-395.

[3] S. Mehrotra. On the implementation of a primal–dual interior point method. *SIAM J. Optimization*, 2(4):575–601, 1992.

[4] M. J. Todd and Y. Ye, A centered projective algorithm for linear programming, *Math. Oper. Res.* 15 (1990) 508-529.

[5] Y. Ye, M. J. Todd, and S. Mizuno, An $O(\sqrt{n}L)$ - iteration homogeneous and self-dual linear programming algorithm, *Math. Oper. Res.* 19 (1994) 53–67.

[6] Y. Ye, An $O(n^3L)$ potential reduction algorithm for linear programming, *Math. Programming* 50 (1991) 239–258.

[7] Y. Ye, On the complexity of approximating a KKT point of quadratic programming, *Math. Programming* 80 (1998) 195-211.