Algorithm Analysis of Computing an KKT Point

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye

Linearly Constrained Optimization Problem Again

$$
\text{(LCOP)} \qquad \begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{array} \quad .
$$

We assume that A has full rank and f is a differentiable but may not convex function. The KKT conditions:

$$
\begin{array}{rcl} X\mathbf{s} & = & \mathbf{0} \\ A\mathbf{x} & = & \mathbf{b} \\ -A^T\mathbf{y} + \nabla f(\mathbf{x})^T - \mathbf{s} & = & \mathbf{0} \\ (\mathbf{x}, \mathbf{s}) & \geq & \mathbf{0}. \end{array}
$$

An ϵ KKT solution: if $|x_j s_j| \leq \epsilon$ for all j.

First-Order Affine Scaling Algorithm

$$
\begin{array}{llll}
 & \text{minimize} & \nabla f(\mathbf{x})\mathbf{d}_x & \\
(\text{BOP}) & \text{subject to} & A\mathbf{d}_x = \mathbf{0} & \\
 & & \|(X)^{-1}\mathbf{d}_x\| \le \alpha < 1. &
\end{array}
$$

The subproblem is an convex problem with a close form optimal solution.

Then let $\mathbf{x}^+ = \mathbf{x} + \mathbf{d}_x$. The new solution will be in the interior of the feasible region and the objective function value will be strictly reduced for some small α. This is true when $f(\mathbf{x})$ is either convex or non-convex.

Analysis: Linear Programming Case

$$
\text{(BOP)} \quad \begin{array}{ll} \text{minimize} & \mathbf{c}^T \mathbf{d}_x \\ \text{subject to} & A\mathbf{d}_x = \mathbf{0} \\ & \|(X)^{-1}\mathbf{d}_x\| \le \alpha < 1. \end{array}
$$

Let

$$
\mathbf{p}(\mathbf{x}) = (I - XA^T(AX^2A^T)^{-1}AX)X\mathbf{c} = X(\mathbf{c} - A^T\mathbf{y}(\mathbf{x})),
$$

where

$$
\mathbf{y}(\mathbf{x}) = (AX^2A^T)^{-1}AX^2\mathbf{c}.
$$

Then, the minimizer

$$
\mathbf{d}_x^* = -\alpha \frac{X\mathbf{p}(\mathbf{x})}{\|\mathbf{p}(\mathbf{x})\|}.
$$

Objective function reduction

Let new solution $\mathbf{x}^+ = \mathbf{x} + \mathbf{d}_x^* > \mathbf{0}$. Then, $A\mathbf{x}^+ = \mathbf{b}$ and

$$\mathbf{c}^T\mathbf{x}^+ - \mathbf{c}^T\mathbf{x} = \mathbf{c}^T\mathbf{d}_x^* = -\alpha \cdot \frac{\mathbf{c}^T X\mathbf{p}(\mathbf{x})}{\|\mathbf{p}(\mathbf{x})\|} = -\alpha \cdot \|\mathbf{p}(\mathbf{x})\|.$$

Thus, unless $\|\mathbf{p}(\mathbf{x})\| = 0$, the objective function is strictly reduced.

Does $\|\mathbf{p}(\mathbf{x})\| \le \epsilon$ imply $\mathbf{x}$ is an ϵ KKT point? Recall

$$\mathbf{p}(\mathbf{x}) = (I - XA^T(AX^2A^T)^{-1}AX)X\mathbf{c} = X(\mathbf{c} - A^T\mathbf{y}(\mathbf{x})).$$

Not necessarily, unless certain non-degeneracy conditions are satisfied.

Second-Order Affine Scaling Algorithm

$$
(\text{BOP}) \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{d}_x^T \nabla^2 f(\mathbf{x}) \mathbf{d}_x + \nabla f(\mathbf{x}) \mathbf{d}_x \\ \text{subject to} & A\mathbf{d}_x = \mathbf{0} \\ & \|(X)^{-1}\mathbf{d}_x\| \le \alpha < 1. \end{array}
$$

The subproblem is an SDP problem, and can be (approximately) solved in polynomial time.

Then let $\mathbf{x}^+ = \mathbf{x} + \mathbf{d}_x$. The new solution will be in the interior of the feasible region and the objective function value will be strictly reduced for some small α. This is true when $f(\mathbf{x})$ is either convex or non-convex.

What happens if $f(\mathbf{x})$ is quadratic?

Why it is an SDP Problem

$$
\begin{array}{llll}
 & z^* := & \text{minimize} & \frac{1}{2}\mathbf{y}^T Q\mathbf{y} + \mathbf{c}^T\mathbf{y} \\
(\text{BOP}) & & \text{subject to} & A\mathbf{y} = \mathbf{0} \\
 & & & \|\mathbf{y}\|^2 \le 1.
\end{array}
$$

The problem can be relaxed to

$$
\begin{array}{llll}
 & (z^* \ge) z^{sdp} := & \text{minimize} & \frac{1}{2}Q \bullet Y + \mathbf{c}^T\mathbf{y} \\
 & & \text{subject to} & A^T A \bullet Y = 0, \\
(\text{BOP-SDP}) & & & I \bullet Y \le 1, \\
 & & & Y \succeq \mathbf{y}\mathbf{y}^T (\succeq \mathbf{0}).
\end{array}
$$

If (BQP-SDP) has a solution where the rank of Y exactly equals 1, or $Y^* = \mathbf{y}^*(\mathbf{y}^*)^T$, then $\mathbf{y}^*$ is an optimal solution to (BQP).

Theorem 1 *(BQP-SDP) always possesses a rank-one solution for any $Q, \mathbf{c}, A$, that is, the relaxation is exact $z^* = z^{sdp}$.*

Nonconvex Objective Function: Potential Function Reduction Algorithm

$$\phi(\mathbf{x}) = (n+\rho)\log(f(\mathbf{x})) - \sum_{j=1}^{n} \log x_j,$$

where $f(\mathbf{x})$ is a non-negative valued function in the feasible region. Consider

$$f(\mathbf{x}) = \sum_{j=1}^{n} x_j^p$$

for some constant $0 < p < 1$. Then,

$$\phi(x) = (n+\rho)\log\left(\sum_{j=1}^{n} x_j^p\right) - \sum_{j=1}^{n} \log x_j.$$

If we start from the analytic center $\mathbf{x}^0$ of the feasible region and can reduce the potential function value by a constant, then $f(\mathbf{x}^k) \to 0$.

Affine-Scaling for Potential reduction

Let $\mathbf{d}_x$, $A\mathbf{d}_x = \mathbf{0}$, be a vector such that $\mathbf{x}^+ = \mathbf{x} + \mathbf{d}_x > \mathbf{0}$. Then, from the concavity of $\log(f(\mathbf{x}))$, we have

$$\log(f(\mathbf{x}^+)) - \log(f(\mathbf{x})) \le \frac{1}{f(\mathbf{x})}\nabla f(\mathbf{x})^T \mathbf{d}_x.$$

On the other hand, if $\|X^{-1}\mathbf{d}_x\| \le \alpha < 1$, we have

$$\sum_{j=1}^n \log(x_j^+) - \sum_{j=1}^n \log(x_j) \le -\mathbf{e}^T X^{-1}\mathbf{d}_x + \frac{\alpha^2}{2(1-\alpha)}.$$

Thus, we have

$$\begin{array}{lc} z_x := \text{Minimize} & \left(\frac{\rho}{f(\mathbf{x})}\nabla f(\mathbf{x})^T - X^{-1}\mathbf{e}^T\right)\mathbf{d}_x \\ \text{Subject to} & A\mathbf{d}_x = \mathbf{0} \\ & \|X^{-1}\mathbf{d}_x\| \le \alpha < 1. \end{array}$$

Potential reduction Analysis

The minimal value $z_x = -\alpha \cdot \|\mathbf{p}(\mathbf{x})\|$ where

$$\mathbf{p}(\mathbf{x}) = -(I - XA^T(AX^2A^T)^{-1}AX)(\frac{\rho}{f(\mathbf{x})}X\nabla f(\mathbf{x}) - \mathbf{e})$$

$$\mathbf{e} - \frac{\rho}{f(\mathbf{x})}X(\nabla f(\mathbf{x}) - A^T\mathbf{y}(\mathbf{x})),$$

where

$$\mathbf{y}(\mathbf{x}) = (AX^2A^T)^{-1}AX(\frac{\rho}{f(\mathbf{x})}X\nabla f(\mathbf{x}) - \mathbf{e}).$$

If case $\|\mathbf{p}(\mathbf{x})\| \geq 1$, then the minimal objective value of the subproblem is less than $-\alpha$ so that

$$\phi(\mathbf{x}^+) - \phi(\mathbf{x}) < -\alpha + \frac{\alpha^2}{2(1-\alpha)}.$$

Thus, the potential value is reduced by 0.25 after setting $\alpha = 1/2$.

Time bound

After k iterations in this case, we have

$$\phi(\mathbf{x}^k) - \phi(\mathbf{x}^0) \le -\frac{k}{4}.$$

Since

$$\sum_{j=1}^n \log x_j^k < \sum_{j=1}^n \log x_j^0,$$

we have

$$(n+\rho)\log(f(\mathbf{x}^k)) - (n+\rho)\log(f(\mathbf{x}^0)) < -\frac{k}{4}.$$

Thus, after $O((n+\rho)\log\frac{f(\mathbf{x}^0)}{\epsilon})$ iterations, we would have $f(\mathbf{x}^k) \le \epsilon$.

The Process May Halt

What happens if $\|\mathbf{p}(\mathbf{x})\| < 1$?

$$\mathbf{p}(\mathbf{x}) = \mathbf{e} - \frac{\rho}{f(\mathbf{x})} X(\nabla f(\mathbf{x}) - A^T \mathbf{y}(\mathbf{x})),$$

we must have

$$X(\nabla f(\mathbf{x}) - A^T \mathbf{y}(\mathbf{x})) \geq \mathbf{0}; \quad X(\nabla f(\mathbf{x}) - A^T \mathbf{y}(\mathbf{x})) \leq \frac{2f(x)}{\rho} \mathbf{e}.$$

The first condition indicates that $\mathbf{y}(\mathbf{x})$ is a valid Lagrange multiplier vector. From the second inequality, by choosing $\rho \geq \frac{2f(x^0)}{\epsilon}$, we have

$$|x_j(\nabla f(\mathbf{x}) - A^T \mathbf{y}(\mathbf{x}))_j| \leq \epsilon,$$

which implies that $\mathbf{x}$ is an ϵ KKT solution.

FPTAS Time Complexity

Concluding the analysis above, we have the following result.

Theorem 2 *The interior point algorithm returns an* ϵ *KKT solution or global minimizer in no more than* $O(\frac{f(\mathbf{x}^0)}{\epsilon}\log\frac{f(\mathbf{x}^0)}{\epsilon})$ *iterations by setting* $\rho = \frac{2f(x^0)}{\epsilon}$.

This type of algorithm is called fully polynomial time approximation scheme.

The second order affine-scaling will make the KKT solution satisfy the second order optimality condition in the same time bound.

More questions

- Could the time bound be further improved?
- Could we guarantee that the final KKT solution is a local minimizer?
- Could you prove the sparsity of the final KKT solution when this is a L_p norm minimization?