A one phase interior point method (IPM) for non-convex optimization

Oliver Hinder, Yinyu Ye

Stanford University

ohinder@stanford.edu

October 13, 2016

Outline

Problem we wish to solve:

$$\min f(x)$$
$$a(x) \leq 0$$

Assume:

1. The constraints and objective are C^2
2. The set $a(x) \leq \theta$ is bounded for any $\theta \in \mathbb{R}^m$

Local infeasibility certificate(s)

1. *First-order local L1-infeasibility certificate*, if x^* is a first-order local optimum of:

$$\min e^T (a(x))^+$$

With $e^T (a(x^*))^+ > 0$.

2. *First-order local farkas infeasibility certificate*, if there exists $y \geq 0$ such that x^* is a first-order local optimum of:

$$\min y^T a(x)^+$$

With $y^T a(x^*)^+ > 0$.

One phase IPMs for conic optimization

Either declare problem infeasible (primal or dual) or optimal in one phase:

1. Homogenous algorithm
   [Ye et al., 1994, Andersen and Ye, 1999]
2. Infeasible start IPM
   [Lustig, 1990, Mehrotra, 1992, Todd, 2002]

Failure of infeasible start IPM for non-convex optimization

[Wächter and Biegler, 2000] showed that if we apply an infeasible start IPM to:

$$\begin{aligned} \min x \\ x^2 - s_1 - 1 = 0 \\ x - s_2 - 1/2 = 0 \\ s_1, s_2 \geq 0 \end{aligned}$$

Fails to converge to either a local optimum or infeasibility certificate

Modifications to the infeasible start IPM

1. Two phases (IPOPT)
2. Compute two directions (KNITRO)
3. Penalty (or big-M) method e.g. [Chen, 2006, Curtris, 2012]

Modifications to the infeasible start IPM

1. Two phases (IPOPT)
2. Compute two directions (KNITRO)
3. Penalty (or big-M) method e.g. [Chen, 2006, Curtris, 2012]

Modifications to the infeasible start IPM

1. Two phases (IPOPT)
2. Compute two directions (KNITRO)
3. Penalty (or big-M) method e.g. [Chen, 2006, Curtris, 2012]

Modifications to the infeasible start IPM

1. Two phases (IPOPT)
2. Compute two directions (KNITRO)
3. Penalty (or big-M) method e.g. [Chen, 2006, Curtris, 2012]

It doesn't need to be this complicated!

One phase barrier problem

$$\min f(x)$$
$$a(x) \leq 0$$

1. Add a slack variable
2. Add a log barrier term
3. Keep the constraint violation the same
4. Reduce the constraint violation and μ^k by $\eta^k \in [0, 1]$
5. Add proximal term

One phase barrier problem

$$\min f(x)$$
$$a(x) + s = 0$$
$$s \geq 0$$

1. Add a slack variable
2. Add a log barrier term
3. Keep the constraint violation the same
4. Reduce the constraint violation and μ^k by $\eta^k \in [0, 1]$
5. Add proximal term

One phase barrier problem

$$\min f(x) - \mu^k \sum_i \log s_i$$

$$a(x) + s = 0$$

$$s \geq 0$$

1. Add a slack variable
2. Add a log barrier term
3. Keep the constraint violation the same
4. Reduce the constraint violation and μ^k by $\eta^k \in [0, 1]$
5. Add proximal term

One phase barrier problem

$$\min f(x) - \mu^k \sum_i \log s_i$$

$$a(x) + s = (a(x^k) + s^k)$$

$$s \geq 0$$

1. Add a slack variable
2. Add a log barrier term
3. Keep the constraint violation the same
4. Reduce the constraint violation and μ^k by $\eta^k \in [0,1]$
5. Add proximal term

One phase barrier problem

$$
\begin{aligned}
&\min f(x) - (1-\eta^k)\mu^k \sum_i \log s_i \\
&a(x) + s = (1-\eta^k)(a(x^k) + s^k) \\
&s \geq 0
\end{aligned}
$$

1. Add a slack variable
2. Add a log barrier term
3. Keep the constraint violation the same
4. Reduce the constraint violation and μ^k by $\eta^k \in [0,1]$
5. Add proximal term

One phase barrier problem

$$\min f(x) - (1-\eta^k)\mu^k \sum_i \log s_i + \frac{\delta^k}{2}\|x - x^k\|^2$$
$$a(x) + s = (1-\eta^k)(a(x^k) + s^k)$$
$$s \geq 0$$

1. Add a slack variable
2. Add a log barrier term
3. Keep the constraint violation the same
4. Reduce the constraint violation and μ^k by $\eta^k \in [0, 1]$
5. Add proximal term

Primal-dual direction computation

$$\begin{aligned}\nabla f(x) + \delta^k (x - x^k) + y^T \nabla a(x) &= 0 \\ a(x) + s &= (1 - \eta^k)(a(x^k) + s^k) \\ s_i y_i &= (1 - \eta^k)\mu^k \\ s, y &\geq 0\end{aligned}$$

1. Form KKT system for barrier problem
2. Linearize the KKT system
3. Factorize matrix and compute direction d^k

Primal-dual direction computation

$$\begin{bmatrix} H^k & \nabla a(x^k)^T & 0 \\ \nabla a(x^k) & 0 & I \\ 0 & S^k & Y^k \end{bmatrix} \begin{bmatrix} d_x^k \\ d_y^k \\ d_s^k \end{bmatrix} = \begin{bmatrix} -(\nabla f(x^k) + \nabla a(x^k)^T y^k) \\ -\eta^k (a(x^k) + s^k) \\ (1 - \eta^k)\mu^k e - Y^k s^k \end{bmatrix}$$

$$H^k = \nabla_x^2 L(x^k, y^k) + \delta^k I$$

1. Form KKT system for barrier problem
2. Linearize the KKT system
3. Factorize matrix and compute direction d^k

Primal-dual direction computation

$$\begin{bmatrix} H^k & \nabla a(x^k)^T & 0 \\ \nabla a(x^k) & 0 & I \\ 0 & S^k & Y^k \end{bmatrix} \begin{bmatrix} d_x^k \\ d_y^k \\ d_s^k \end{bmatrix} = \begin{bmatrix} -(\nabla f(x^k) + \nabla a(x^k)^T y^k) \\ -\eta^k (a(x^k) + s^k) \\ (1-\eta^k)\mu^k e - Y^k s^k \end{bmatrix}$$

$$H^k = \nabla_x^2 L(x^k, y^k) + \delta^k I$$

1. Form KKT system for barrier problem
2. Linearize the KKT system
3. Factorize matrix and compute direction d^k

Iterate update

For step size $\alpha^k \in [0, 1]$ update iterates as follows:

$$
\begin{aligned}
x^{k+1} &= x^k + \alpha^k d_x^k \\
y^{k+1} &= y^k + \alpha^k d_y^k \\
s^{k+1} &= (1 - \alpha^k \eta^k)(a(x^k) + s^k) - a(x^{k+1}) \\
\mu^{k+1} &= (1 - \alpha^k \eta^k)\mu^k
\end{aligned}
$$

Only accept iterates that approximately satisfy complementary:

$$
\beta \mu^{k+1} \leq s_i^{k+1} y_i^{k+1} - \mu^{k+1} \leq \frac{1}{\beta} \mu^{k+1}
$$

For some constant $\beta \in (0, 1)$.

One phase non-convex algorithm

for $k \leftarrow 1, \ldots, \infty$ **do**

if $\|\nabla f(x^k) + (y^k)^T \nabla a(x^k)\| < \mu^k$ **then**

Aggressive step:

Compute direction with $\eta^k \in (0, 1]$

Take largest step while maintaining complementary

else

Stabilization step:

Compute direction with $\eta^k = 0$

Backtracking line search on direction using merit function

end if

end for

Termination criterion

First-order ϵ-locally optimal if:

$$\left\|\nabla f(x) + y^T \nabla a(x)\right\| + \|a(x) + s\| + \mu < \epsilon$$

First-order ϵ-locally farkas infeasible if:

$$\frac{\mu + \|y^T \nabla a(x)\|}{y^T (a(x) + s)} < \epsilon$$

Why poly-time is hard for interior point methods

Suppose we are given a feasible starting point to the following problem:

$$\min_x f(x) - \mu^k \sum_i \log(-a(x))$$

1. Log barrier is $O(1/\delta^2)$ Lipschitz continuous on the set $\{x : a(x) \leq -\delta\}$ for any $\delta > 0$
2. Bounding δ using the log barrier merit function:

$$\delta = \Omega\left(\exp(-1/\mu^k)\right)$$

Implies exponential runtime in μ^k for gradient descent or newton's method!!!

Global poly-time convergence proof sketch

Prove that there exists $\delta^k = O(\|y^k\|)$ such that:

$$\delta^k \|d_x^k\| = O\left(1 + \sqrt{\|y^k\| \mu^k}\right)$$

i.e. independent of r.h.s. This implies:

$$\|KKT^k\| = O\left(1 + \sqrt{\|y^k\| \mu^k}\right)$$

Which implies assuming the algorithm has not produced an infeasibility certificate that:

$$s_i^k = \Omega((\mu^k \epsilon)^2)$$

Global poly-time convergence

1. Differentiability and boundedness (defined earlier)
2. Slack variable s^0 is chosen carefully e.g. such that $a_i(x^0) + s_i^0 > 0$
3. That δ^k and η^k are judiciously chosen

Theorem

If (1-3) hold, then for any $\epsilon > 0$, after at most $\tilde{O}(\epsilon^{-17})$ steps the algorithm has found either a first-order ϵ-local optimum or a first-order ϵ-local farkas infeasibility certificate.

Convergence to a KKT point

Conjecture

Let $(x^{\pi(k)}, s^{\pi(k)}) \rightarrow (x^*, s^*)$ and suppose the set Y^* of dual multipliers that satisfy the KKT conditions is non-empty then:

1. The sequence $y^{\pi(k)}$ is bounded above
2. The sequence $y^{\pi(k)}$ is maximal complementary

Local convergence

1. Differentiability and boundedness (defined earlier)
2. *** Converging towards a KKT solution that satisfies strict complementary.
3. The sufficient conditions for local optimality are satisfied at this KKT solution.
4. That $\delta^k = 0$ and η^k is judiciously chosen

Theorem

If (1-4) hold, either the sequence of odd or even iterates converge Q-quadratically to a local optimum.

First-order L1-infeasibility certificate

$$\min f(x)$$
$$a(x) \leq 0$$

This problem is trivially equivalent to:

$$\min f(x)$$
$$a(x) \leq z$$
$$e^T z \leq 0$$
$$z \geq 0$$

Farkas infeasibility certificate $\rightarrow$ L1-infeasibility certificate

Empirical results with CUTEst

Table: Results on 30 selected CUTEst problems, number of variables between 100 and 1000

	One phase	IPOPT (2004)	KNITRO (2004)
Median #iterations	71 (59)	41	53
SEGFAULTS	4	0	0
MAXIT	3	3	3

Conclusions

1. Infeasible start IPM only needs minor modifications to guarantee convergence
2. For global and super-linear convergence: no constraint qualifications
3. Early computational results promising - more work needed

The End

References I

Andreas Wächter, Lorenz T. Biegler (2000)
Failure of global convergence for a class of interior point methods for nonlinear programming. *Mathematical Programming*.

Yinyu Ye, Michael J. Todd, and Shinji Mizuno (1994)
An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm *Mathematics of Operations Research*.

Erling D. Andersen and Yinyu Ye (1999)
On a homogeneous algorithm for the monotone complementarity problem. *Mathematical Programming*.

Irvin J Lustig (1990)
Feasibility issues in a primal-dual interior-point method for linear programming. *Mathematical Programming*.

References II

Mehrotra Sanjay (1992)
On the implementation of a primal-dual interior point method. *SIAM Journal on optimization*.

Michael J. Todd (2002)
Detecting infeasibility in infeasible-interior-point methods for optimization. *Foundations of Computational Mathematics*.

Frank E. Curtis (2012)
A penalty-interior-point algorithm for nonlinear constrained optimization. *Mathematical Programming Computation*.

Lifeng Chen and Donald Goldfarb (2006)
A penalty-interior-point algorithm for nonlinear constrained optimization. *Mathematical Programming*.