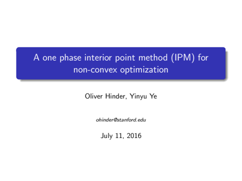
A one phase IPM for non-convex optimization



1. Inspired by the success of one phase algorithms for convex optimization.
2. Simplify non-convex IPM by eliminating two phases.
3. Eliminate unnecessary assumptions for convergence theory.
4. Improve reliability of non-convex interior point algorithms, particularly on infeasible problems.

2016-07-11

└Outline

Outline

Problem we wish to solve:

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

Assume:

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

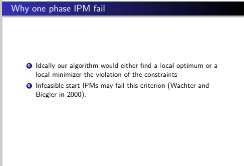
One phase IPMs for convex optimization

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

1. Infeasible start IPM (Mehortra ...)
2. Homogenous algorithm (Ye ...)

State of the art for convex optimization

In a one phase algorithm we use one fixed newton system at each step corresponding to the homotopy equations, but change the right hand side to decide the target reduction in primal, dual and complementary to get quick convergence. Furthermore, same set of primal and dual variables are used at each iteration.



Our solution:

1. Use inequality constraints
2. Adjust the target rate of reduction of primal and complementary
3. Add dummy constraints s.t. farkas certificate of infeasibility $\rightarrow$ minimizer of L1-norm of constraint violation

- Two phases (IPOPT)
  - Each phase has a different variables
  - Awkward convergence assumptions
  - Poor performance on infeasible problems
- Compute two directions (KNITRO)
- Penalty (or big-M) method (Curtis, 2012)

1. Two phases (IPOPT): main phase and feasibility restoration phase
   1.1 Each phase has a different variables: each time initialize FRP start from scratch, the dual variables = zero, new introduced primal variables have ad hoc values.
   1.2 Awkward convergence assumptions: to prove convergence it is assumed constraints are linearly independent in a neighborhood of the feasible region.
   1.3 Poor performance on infeasible problems:
2. At each iteration compute two directions (KNITRO) using two different newton systems
   2.1 One direction aims to minimize L2 norm of constraint violation, other step aims to improve optimality
   2.2 Not 'pure' IPM in feasibility search, requires factorizing two different newton systems at each step.

Local infeasibility certificates

Farkas certificate, there exists $y \geq 0$, $y \neq 0$, $y_i a_i(x) \geq 0$ such that x^* is a stationary point of:

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

L1 constraint violation, x^* is a stationary point of:

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

we are moving complementary and primal feasibility at same rate - idea that is common in convex codes (dual)

Direction computation

$$H^k d_x + (A^k)^T d_y = -(\nabla c(x) + y^T \nabla a(x))$$
$$A^k d_x + d_s = -\eta_k (a(x^k) + s^k)$$
$$S^k d_y + Y^k d_s = (1 - \eta_k)\mu_k e - Y^k s^k$$

$$H^k = \nabla_x^2 L(x^k, y^k) + \delta_k I$$
$$A^k = \nabla a(x^k)$$

we are moving complementary and primal feasibility at same rate - idea that is common in convex codes

One phase non-convex algorithm

for $k \leftarrow 1, \ldots, \infty$ **do**
  Factor newton system using IPOPT strategy to compute δ_k

  if $\|\nabla c(x^k) + (y^k)^T \nabla a(x^k)\| < \mu_k$ **then**
    Aggressive step:
    Compute direction with $\eta_k \in (0, 1]$
    Take largest step while maintaining complementary
  else
    Stabilization step:
    Compute direction with $\eta_k = 0$
    Line search on direction
  end if

end for

Iterate update

Backtrack on α and update iterates:

$$
\begin{aligned}
x^{k+1} &= x^k + \alpha d_x^k \\
y^{k+1} &= y^k + \alpha d_y^k \\
s^{k+1} &= (1 - \alpha_k \eta_k)(a(x^k) + s^k) - a(x^{k+1}) \\
\mu_{k+1} &= \mu_k + (1 - \eta_k \alpha_k)\mu_k
\end{aligned}
$$

Only accept iterates that approximately satisfy complementary:

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

2016-07-11

Merit function line search

Backtrack on α until sufficient progress has been made on the merit function:

$$\psi^k(x) = c(x) - \mu^k \sum_i \log\left(a_i(x^k) + s_i^k - a_i(x)\right)$$

Termination criterion

Declare problem first-order ϵ-local optimal if:

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

Declare problem first-order ϵ-local infeasible if:

$$\frac{1}{y^T(a(x)+s)}\left(\mu + \|y^T \nabla a(x)\|\right) < \epsilon$$

If δ_k and η_k are judiciously chosen then it possible to show the following result:

Theorem

For any $\epsilon > 0$, after a finite number of steps the algorithm has found either a first-order ϵ-local optimum or a first-order ϵ-local infeasibility certificate.

A one phase IPM for non-convex optimization
└─Theory
└─Global convergence
└─Conjectures

2016-07-11

Conjecture

Assume (x^*, s^*) is a limit point of the algorithm and there exists finite dual multipliers at (x^*, s^*) then for some subsequence of the dual multipliers $y^{\pi(k)} \to y^*$.

Conjecture

After a polynomial number of steps a simple variant of the algorithm has found either a first-order ϵ-local optimum or a first-order ϵ-local infeasibility certificate.

2016-07-11

Local convergence

Assume:

1. Converging towards a KKT solution that satisfies strict complementary.
2. The sufficient conditions for local optimality are satisfied at the KKT solution.

If η_k is judiciously chosen then it is possible to show:

Theorem

If assumptions (1-3) hold, the algorithm converges Q-quadratically to a local optimality certificate.

How to return L1 minimizer of constraint violation

We wish to solve:

$$\begin{aligned} &\min c(x) \\ &a(x) \leq 0 \end{aligned}$$

This problem is trivially equivalent to:

$$\begin{aligned} &\min c(x) \\ &a(x) \leq z \\ &e^T z \leq 0 \\ &z \geq 0 \end{aligned}$$

The local Farkas infeasibility certificate to this problem is an certificate of L1 minimization of the constraint violation.

1. Use fraction to boundary rule
2. Choose η_k on aggressive steps using predictor-corrector technique
3. Search on negative eigenvector to guarantee second-order necessary conditions are satified
4. Filter to improve stabilization step performance

Scaled KKT merit function

Scaled KKT residuals is an alternate merit function:

$$\frac{\|\nabla c(x) + y^T \nabla a(x)\|}{\|y\|}$$

Accept step that makes progress on either:

1. Classic merit function
2. Scaled KKT residuals

Using a filter to prevent cycling

Empirical results

Table: Results on 14 selected CUTEst problems

	One phase	IPOPT (2004)	KNITRO (2004)
Mean iterations	71	152	740
Standard error	5	25	85

1. Second-order corrections?