Argonne NATIONAL LABORATORY

Interior-Point Methods and Globalization

GIAN Short Course on Optimization: Applications, Algorithms, and Computation

Sven Leyffer

Argonne National Laboratory

September 12-24, 2016

U.S. DEPARTMENT OF ENERGY

Outline

1. Interior-Point Methods
   - Primal-Dual Interior-Point Methods
   - Barrier Interior-Point Methods

2. Classical Augmented Lagrangian Methods
   - Linearly Constrained Lagrangian Methods
   - Bound-Constrained Lagrangian (BCL) Methods.
   - Theory of Augmented Lagrangian Methods

More Methods for Nonlinear Optimization

Nonlinear Program (NLP) of the form

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & c(x) = 0 \\ & x \geq 0, \end{aligned}$$

where

- objective $f : \mathbb{R}^n \to \mathbb{R}$ twice continuously differentiable
- constraints $c : \mathbb{R}^n \to \mathbb{R}^m$ twice continuously differentiable
- y multipliers of $c(x) = 0$
- $z \geq 0$ multipliers of $x \geq 0$.

... can reformulate more general NLPs easily

... solvers accept more general format

Interior-point methods (IPMs)

IPMs are alternative to active-set methods (SLP, SQP, etc)

- Class of perturbed Newton methods
- Postpone decision of which constraints are active until end
  - SQP et al. are active-set methods
  - SQP et al. have estimate of active set at every iteration
    ... from active set of LP or QP subproblem
- Best IPM are primal-dual methods, use perturbed KKT system

Interior-point methods (IPMs)

IPMs are alternative to active-set methods (SLP, SQP, etc)

- Class of perturbed Newton methods
- Postpone decision of which constraints are active until end
  - SQP et al. are active-set methods
  - SQP et al. have estimate of active set at every iteration
    ... from active set of LP or QP subproblem
- Best IPM are primal-dual methods, use perturbed KKT system

NLP Problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & c(x) = 0 \\ & x \geq 0, \end{aligned}$$

KKT conditions

$$\begin{aligned} \nabla f(x) - \nabla c(x)^T y - z &= 0 \\ c(x) &= 0 \\ Xz &= 0 \end{aligned}$$

where $X = \text{diag}(x)$, $x \geq 0, z \geq 0$

Without $x, z \geq 0$, KKT is a system of equations!

Perturbed KKT System

$$\underset{x}{\text{minimize}} \; f(x) \quad \text{subject to } c(x) = 0, \; x \geq 0$$

Perturbation to KKT system

- Assume that given $x^{(0)} > 0$ and $z^{(0)} > 0$
- Seek an algorithm that maintains $x^{(k)} > 0$ and $z^{(k)} > 0$

$\Rightarrow$ Perturb complementarity , $Xz = 0$, in KKT system

Primal-Dual System

$$0 = F_\mu(x, y, z) = \begin{pmatrix} \nabla f(x) - \nabla c(x)^T y - z \\ c(x) \\ Xz - \mu e \end{pmatrix},$$

where $\mu > 0$ is the barrier parameter

Interior-Point Methods

Primal-dual interior-point methods

- Start at "interior" iterate $x^{(0)}, z^{(0)} > 0$ (near analytic center)
- Generate sequence of interior iterates $x^{(k)}, z^{(k)} > 0$
- Approximately solve primal-dual system, decreasing μ
- Polynomial-time algorithms for convex NLPs; e.g. [Nesterov and Nemirovskii, 1994]

Relationship to classical barrier methods

- Primal dual system related to minimization of barrier function

$$f(x;\mu) := f(x) - \mu \sum_{i=1}^{n} \log(x_i) \quad \text{subject to } c(x) = 0$$

- Log-barrier $\log(x_i)$ as approach boundary: $x_i \to 0$
- Can show that minimizers, $x^{\mu} \to x^*$ as $\mu \to 0$

... more later, for now let's look at solving the primal-dual system

Solving the Primal-Dual System

Apply Newton's method to primal-dual system:

$$0 = F_\mu(x,y,z) = \begin{pmatrix} \nabla f(x) - \nabla c(x)^T y - z \\ c(x) \\ Xz - \mu e \end{pmatrix},$$

Around iterate $x^{(k)}$ get Newton (linear) system:

$$\begin{bmatrix} H^{(k)} & -A^{(k)} & -I \\ A^{(k)^T} & 0 & 0 \\ Z^{(k)} & 0 & X^{(k)} \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = -F_\mu(x^{(k)}, y^{(k)}, z^{(k)}),$$

where $H^{(k)} \approx \nabla^2 \mathcal{L}^{(k)}$ and ensure $x^{(k+1)}, z^{(k+1)} > 0$

Algorithm: Primal-Dual Interior-Point Method (IPM)

Given $(x^{(0)}, y^{(0)}, z^{(0)})$, with $(x^{(0)}, z^{(0)}) > 0$

Choose barrier parameter μ_0, $0 < \sigma < 1$, and $\epsilon_k \searrow 0$

repeat

  Set $(x^{(k,0)}, y^{(k,0)}, z^{(k,0)}) = (x^{(k)}, y^{(k)}, z^{(k)})$, $l = 0$.

  repeat

    Approx. solve Newton system: $(x^{(k,l+1)}, y^{(k,l+1)}, z^{(k,l+1)})$

    Set $l = l + 1$

  until $\|F_{\mu_k}(x^{(k,l)}, y^{(k,l)}, z^{(k,l)})\| \le \epsilon_k$;

  Reduce barrier parameter $\mu_{k+1} = \sigma \mu_k$, set $k = k + 1$.

until $x^{(k)}, y^{(k)}, z^{(k)}$ *optimal*;

Remark (Structure of Interior-Point Methods (IPMs))

IPMs have inner (approx. Newton) & outer loop (barrier, $\mu \searrow 0$)

Relationship to Barrier Methods

Primal-dual IPMs related to classical barrier methods [Fiacco and McCormick, 1990].

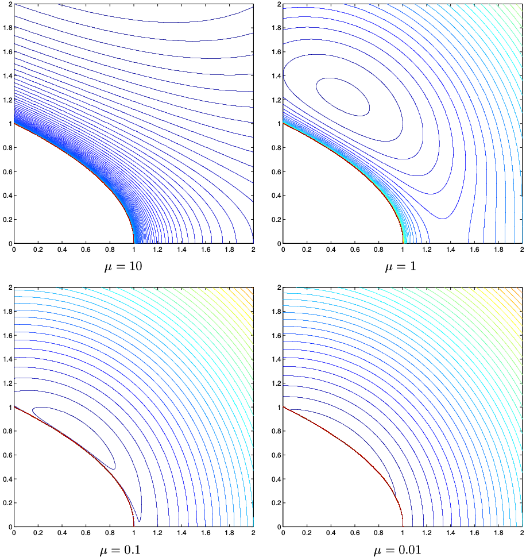
- Popular methods in 1960s
- Lost popularity with rise of SQP due to ill-conditioning
- Renewed interest in 1980s due to polynomial-time properties
- Good references on IPMs: [Wright, 1992, Forsgren et al., 2002, Nemirovski and Todd, 2008]

Barrier Problem:

$$\underset{x}{\text{minimize}} \; f(x) - \mu \sum_{i=1}^{n} \log(x_i) \quad \text{subject to } c(x) = 0,$$

for decreasing barrier parameters $\mu \searrow 0$

Illustration of Barrier Methods



Contours barrier for $\min_x \; x_1^2 + x_2^2$ subject to $x_1 + x_2^2 \geq 1$

Relationship to Barrier Methods

Barrier Problem: for barrier parameters $\mu \searrow 0$

$$\underset{x}{\text{minimize}} \;\; f(x) - \mu \sum_{i=1}^{n} \log(x_i) \quad \text{subject to } c(x) = 0,$$

First-Order Conditions of Barrier Problem:

$$\nabla f(x) - \mu X^{-1} e - A(x) y = 0 \;\text{ and }\; c(x) = 0.$$

Newton's method applied to FO conditions of barrier problem:

$$\begin{bmatrix} H^{(k)} + \mu {X^{(k)}}^{-2} & -A^{(k)} \\ {A^{(k)}}^T & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} g^{(k)} - \mu {X^{(k)}}^{-1} e - A^{(k)} y^{(k)} \\ c^{(k)} \end{pmatrix}.$$

... show how this relates to primal-dual Newton

Relationship to Barrier Methods

Newton's method applied to FO conditions of barrier problem:

$$\begin{bmatrix} H^{(k)} + \mu X^{(k)^{-2}} & -A^{(k)} \\ A^{(k)^T} & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} g^{(k)} - \mu X^{(k)^{-1}} e - A^{(k)} y^{(k)} \\ c^{(k)} \end{pmatrix}.$$

First-order multipliers: $Z(x^{(k)}) := \mu X^{(k)^{-1}} \Leftrightarrow Z(x^{(k)}) X^{(k)} = \mu e$

$$\begin{bmatrix} H^{(k)} + Z(x^{(k)}) X^{(k)^{-1}} & -A^{(k)} \\ A^{(k)^T} & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} g^{(k)} - \mu X^{(k)^{-1}} e - A^{(k)} y^{(k)} \\ c^{(k)} \end{pmatrix}$$

equivalent to primal-dual Newton system ...

Relationship between Barrier Methods & Primal-Dual IPMs

Consider primal-dual Newton system

$$\begin{bmatrix} H^{(k)} & -A^{(k)} & -I \\ A^{(k)^T} & 0 & 0 \\ Z^{(k)} & 0 & X^{(k)} \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = - \begin{pmatrix} \nabla g^{(k)} - A^{(k)^T} y^{(k)} - z^{(k)} \\ c^{(k)} \\ X^{(k)} z^{(k)} - \mu e \end{pmatrix}$$

... from last equation eliminate Δz

$$\Delta z = -X^{(k)^{-1}} Z^{(k)} \Delta x - Z^{(k)} e - \mu X^{(k)^{-1}} e.$$

then get

$$\begin{bmatrix} H^{(k)} + Z(x^{(k)}) X^{(k)^{-1}} & -A^{(k)} \\ A^{(k)^T} & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} g^{(k)} - \mu X^{(k)^{-1}} e - A^{(k)} y^{(k)} \\ c^{(k)} \end{pmatrix}$$

Difference between Classical & Primal-Dual IPM

Multipliers $Z^{(k)}$ are not free in barrier methods but set as $Z(x^{(k)}) = \mu X^{(k)^{-1}}$

Convergence of Barrier Methods

(NLP) $\underset{x}{\text{minimize}} \; f(x) \quad \text{subject to } c(x) = 0 \quad x \geq 0$

Barrier Method solves

$$\underset{x}{\text{minimize}} \; f(x) - \mu \sum_{i=1}^{n} \log(x_i) \quad \text{subject to } c(x) = 0,$$

... for $\mu \searrow 0$

Theorem (Convergence of Barrier Methods [Wright, 1992])

If there exists compact set of isolated local minimizers of (NLP) with at least one point in closure of strictly feasible set, then it follows that barrier methods converge to local minimum.

- No guarantee that have converging subsequence in $\{x^{(k)}\}$
- No convergence to local min, if we do not find global min of barrier

Outline

Augmented Lagrangian Methods

$$(\text{NLP}) \quad \begin{cases} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & c(x) = 0 \\ & x \geq 0, \end{cases}$$

Augmented Lagrangian related to quadratic penalty methods:

Given penalty $\rho_k \nearrow \infty$

repeat

Solve quadratic penalty problem:

$$x^\rho \leftarrow \underset{x}{\text{argmin}} \; f(x) + \rho_k \|c(x)\|_2^2 \quad \text{subject to } x \geq 0$$

Increase ρ

until x^ρ *optimal*;

- Inefficient (solve many bound constrained NLPs)
- Solution only in limit, i.e. $\rho \to \infty$

Augmented Lagrangian Methods

$$\text{(NLP)} \qquad \begin{cases} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & c(x) = 0 \\ & x \geq 0, \end{cases}$$

Recall Lagrangian function: $\mathcal{L}(x, y, z) = f(x) - y^T c(x) - z^T x$

Augmented Lagrangian allows convergence for finite value of ρ

$$\mathcal{L}(x, y, \rho) = f(x) - y^T c(x) + \frac{\rho}{2} \|c(x)\|_2^2,$$

where $\rho > 0$ is penalty parameter.

Classical Augmented Lagrangian Methods

1. Linearly constrained augmented Lagrangian
2. Bound constrained augmented Lagrangian

Linearly Constrained Lagrangian Methods

Successively minimize shifted augmented Lagrangian

$$\overline{\mathcal{L}}(x, y, \rho) = f(x) - y^T p^{(k)}(x) + \frac{\rho}{2}\|p^{(k)}(x)\|_2^2,$$

subject linearized constraints.
Here $p^{(k)}(x)$ are higher-order nonlinear terms at $x^{(k)}$:

$$p^{(k)}(x) = c(x) - c^{(k)} - A^{(k)^T}(x - x^{(k)}).$$

Gives approximate subproblem:

$$(\text{LCL}) \qquad \begin{cases} \underset{x}{\text{minimize}} \;\; \overline{\mathcal{L}}(x, y^{(k)}, \rho_k) \\ \text{subject to } c^{(k)} + A^{(k)^T}(x - x^{(k)}) = 0, \\ \qquad\qquad\quad x \geq 0. \end{cases}$$

Linearly Constrained Lagrangian Methods

Linearly Constrained Lagrangian (LCL) subproblem (e.g. minos)

$$\text{(LCL)} \qquad \begin{cases} \underset{x}{\text{minimize}} & \overline{\mathcal{L}}(x, y^{(k)}, \rho_k) \\ \text{subject to} & c^{(k)} + A^{(k)^T}(x - x^{(k)}) = 0, \\ & x \geq 0. \end{cases}$$

$c^{(k)} + A_k^T(x - x^{(k)}) = 0 \Rightarrow$ equivalent to min. Lagrangian

- Solve sequence of approx. subproblems for fixed penalty, ρ_k
- Update multipliers by first-order multiplier update rule:

$$y^{(k+1)} = y^{(k)} - \rho_k c(x^{(k+1)})$$

  where $x^{(k+1)}$ solves (LCL).
- Augmented Lagrangian methods iterate on the dual variables

Bound-Constrained Lagrangian (BCL) Methods

Approximately minimize the augmented Lagrangian,

$$\text{(BCL)} \quad \underset{x}{\text{minimize}} \ \mathcal{L}(x, y^{(k)}, \rho_k) \quad \text{subject to} x \geq 0.$$

Advantages

- Have fast methods for bound-constrained optimization
  e.g. projected gradient CG method described earlier
- Potential for parallel linear algebra
  ... good preconditioners are an issue

Global Convergence of BCL

- Forcing sequences: $\omega_k \searrow 0$, and $\eta_k \searrow 0$
- ω_k controls accuracy of approx. (BCL) solve
- η_k controls convergence to feasibility

Bound-Constrained Lagrangian (BCL) Methods

Given $(x^{(0)}, y^{(0)})$, and penalty parameter ρ_0

repeat

  Set $x^{(k,0)} = x^{(k)}$, $\rho_{k,0} = \rho_k$, $l = 0$, and `success = false`.

  repeat

    Find ω_k-optimal solution $x^{(k,l+1)}$ of

$$\underset{x}{\text{minimize}} \ \mathcal{L}(x, y^{(k)}, \rho_{k,l}) \quad \text{subject to } x \geq 0$$

    if $\|c(x^{(k,l+1)})\| \leq \eta_k$ **then**

      FO multiplier update: $y^{(k+1)} = y^{(k)} - \rho_{k,l} c(x^{(k,l+1)})$

      Set $\rho_{k+1} = \rho_{k,l}$, and `success = true`.

    else

      Increase penalty: $\rho_{k,l+1} = 10\rho_{k,l}$; set $l = l + 1$.

    end

  until `success = true`;

  Set $x^{(k+1)} + x^{(k,l+1)}$, and $k = k + 1$.

until $x^{(k)}, y^{(k)}$ *is optimal*;

Bound-Constrained Lagrangian (BCL) Methods

- Method has inner and outer loop:
  - Inner loop updates penalty parameter to get suff. large
  - Outer loop iterates on dual $y^{(k)}$ variables
    ... $x^{(k,0)} = x^{(k)}$ only initial guess for approx. BCL solve
- Solve each (BCL) subproblem with trust-region algorithm
  ... e.g. projected-gradient with conjugate gradient steps
- Implemented in `LANCELOT` package (open source)

Theory of Augmented Lagrangian Methods

Theorem (Global Convergence of BCL [Conn et al., 1991])

If $\{x^{(k)}\}$ bounded and if constraint Jacobian has full rank for all limit points, then BCL converges from any starting point.

Can show algorithm is R-linearly convergent and q-linearly, if

Theorem (Convergence Rates [Bertsekas, 1996])

Assume that

1. $y^{(k)}$ *updated as* $y^{(k+1)} = y^{(k)} + \rho_k c(x^{(k)})$,
2. $\{\rho^{(k)}\}$ *sequence such that* $\rho_{k+1} \geq \rho_k \ \forall \ k > 0$,
3. x^* *is strict local minimum & regular with multipliers* y^*,
4. $s^T \nabla^2 \mathcal{L}(x^*, y^*) s > 0$ *for all* $s \neq 0$ *with* $\nabla c(x^*)^T s = 0$,

then BCL converges to (x^*, y^*) *Q-linearly, if* $\{\rho_k\}$ *bounded, and superlinearly otherwise.*

Summary

Presented two families of methods

Interior-Point Methods

- Follow path defined by perturbed KKT conds
- Apply Newton's method to perturbed KKT conds
  ... solve (sparse) linear system $\Rightarrow$ suitable for large NLPs
- Related to classical barrier methods
  ... primal-dual methods avoid ill-conditioning

Augmented Lagrangian Methods

- Minimize augmented Lagrangian (add penalty $\rho\|c(x)\|_2^2$)
- Linearly constrained augmented Lagrangian
- Bound constrained augmented Lagrangian

Bertsekas, D. (1996).
Constrained Optimization and Lagrange Multiplier Methods.
Athena Scientific, Belmont, Mass.

Conn, A. R., Gould, N. I. M., and Toint, P. L. (1991).
A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds.
SIAM Journal of Numerical Analysis, 28(2):545–572.

Fiacco, A. V. and McCormick, G. P. (1990).
Nonlinear Programming: Sequential Unconstrained Minimization Techniques.
Number 4 in Classics in Applied Mathematics. SIAM.
Reprint of the original book published in 1968 by Wiley, New York.

Forsgren, A., Gill, P. E., and Wright, M. H. (2002).
Interior methods for nonlinear optimization.
SIAM Review, 4(4):525–597.

Nemirovski, A. and Todd, M. (2008).
Interior-point methods for optimization.
Acta Numerica, 17:181–234.

Nesterov, Y. and Nemirovskii, A. (1994).
Interior Point Polynomial Algorithms in Convex Programming.
Number 13 in Studies in Applied Mathematics. SIAM.

Wright, M. (1992).
Interior methods for constrained optimization.
Acta Numerica, 1:341–407.