ORF 522: Lecture 18

Nonlinear Optimization: Algorithms and Models

Robert J. Vanderbei

November 29, 2012

Slides last edited at 3:02pm on Thursday 29th November, 2012

Operations Research and Financial Engineering, Princeton University

http://www.princeton.edu/∼rvdb

Outline

- Algorithm
  - Basic Paradigm
  - Step-Length Control
  - Diagonal Perturbation
- Convex Problems
  - Minimal Surfaces
  - Digital Audio Filters
- Nonconvex Problems
  - Celestial Mechanics
  - Putting on an Uneven Green
  - Goddard Rocket Problem

The Interior-Point Algorithm

Introduce Slack Variables

- Start with an optimization problem—for now, the simplest NLP:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h_i(x) \geq 0, \qquad i = 1, \ldots, m \end{array}$$

- Introduce slack variables to make all inequality constraints into nonnegativities:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) - w = 0, \\ & \qquad\quad w \geq 0 \end{array}$$

Associated Log-Barrier Problem

- Replace nonnegativity constraints with *logarithmic barrier terms* in the objective:

$$
\begin{aligned}
\text{minimize} \quad & f(x) - \mu \sum_{i=1}^{m} \log(w_i) \\
\text{subject to} \quad & h(x) - w = 0
\end{aligned}
$$

First-Order Optimality Conditions

- Incorporate the equality constraints into the objective using *Lagrange multipliers*:

$$L(x, w, y) = f(x) - \mu \sum_{i=1}^{m} \log(w_i) - y^T (h(x) - w)$$

- Set all derivatives to zero:

$$\begin{aligned} \nabla f(x) - \nabla h(x)^T y &= 0 \\ -\mu W^{-1} e + y &= 0 \\ h(x) - w &= 0 \end{aligned}$$

Symmetrize Complementarity Conditions

- Rewrite system:

$$
\begin{aligned}
\nabla f(x) - \nabla h(x)^T y &= 0 \\
WYe &= \mu e \\
h(x) - w &= 0
\end{aligned}
$$

Apply Newton's Method

- Apply Newton's method to compute *search directions*, Δx, Δw, Δy:

$$\begin{bmatrix} H(x,y) & 0 & -A(x)^T \\ 0 & Y & W \\ A(x) & -I & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta w \\ \Delta y \end{bmatrix} = \begin{bmatrix} -\nabla f(x) + A(x)^T y \\ \mu e - WYe \\ -h(x) + w \end{bmatrix}.$$

Here,

$$H(x,y) = \nabla^2 f(x) - \sum_{i=1}^{m} y_i \nabla^2 h_i(x)$$

and

$$A(x) = \nabla h(x)$$

- Note: $H(x,y)$ is positive semidefinite if f is convex, each h_i is concave, and each $y_i \geq 0$.

Reduced KKT System

- Use second equation to solve for Δw. Result is the *reduced KKT system*:

$$\begin{bmatrix} -H(x,y) & A^T(x) \\ A(x) & WY^{-1} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \nabla f(x) - A^T(x)y \\ -h(x) + \mu Y^{-1} e \end{bmatrix}$$

- Iterate:

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} \Delta x^{(k)}$$
$$w^{(k+1)} = w^{(k)} + \alpha^{(k)} \Delta w^{(k)}$$
$$y^{(k+1)} = y^{(k)} + \alpha^{(k)} \Delta y^{(k)}$$

Convex vs. Nonconvex Optimization Probs

Nonlinear Programming (NLP) Problem:

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & h_i(x) = 0, & i \in \mathcal{E}, \\ & h_i(x) \geq 0, & i \in \mathcal{I}. \end{array}$$

NLP is *convex* if

- h_i's in equality constraints are affine;
- h_i's in inequality constraints are concave;
- f is convex;

NLP is *smooth* if

- All are twice continuously differentiable.

Modifications for Convex Optimization

For convex *nonquadratic* optimization, it does not suffice to choose the steplength α simply to maintain positivity of nonnegative variables.

- Consider, e.g., minimizing
$$f(x) = (1 + x^2)^{1/2}.$$

- The iterates can be computed explicitly:
$$x^{(k+1)} = -(x^{(k)})^3$$

- Converges if and only if $|x| \leq 1$.

- Reason: away from 0, function is too linear.

Step-Length Control

A *filter-type* method is used to guide the choice of steplength α.

Define the *dual normal matrix*:

$$N(x, y, w) = H(x, y) + A^T(x)W^{-1}YA(x).$$

Theorem *Suppose that $N(x, y, w)$ is positive definite.*

1. *If current solution is primal infeasible, then $(\Delta x, \Delta w)$ is a descent direction for the infeasibility $\|h(x) - w\|$.*
2. *If current solution is primal feasible, then $(\Delta x, \Delta w)$ is a descent direction for the barrier function.*

Shorten α until $(\Delta x, \Delta w)$ is produces a decrease in either the infeasibility or the barrier function.

Nonconvex Optimization: Diagonal Perturbation

- If $H(x,y)$ is not positive semidefinite then $N(x,y,w)$ *might* fail to be positive definite.
- In such a case, we lose the descent properties given in previous theorem.
- To regain those properties, we perturb the Hessian: $\tilde{H}(x,y) = H(x,y) + \lambda I$.
- And compute search directions using $\tilde{H}$ instead of H.

Notation: let $\tilde{N}$ denote the dual normal matrix associated with $\tilde{H}$.

Theorem *If $\tilde{N}$ is positive definite, then $(\Delta x, \Delta w, \Delta y)$ is a descent direction for*

1. *the primal infeasibility, $\|h(x) - w\|$ and*
2. *the noncomplementarity, $w^T y$.*

Notes:

- *Not necessarily* a descent direction for *dual infeasibility*.

- A *line search* is performed to find a value of λ within a factor of 2 of the smallest permissible value.

Nonconvex Optimization: Jamming

Theorem *If the problem is convex and and the current solution is not optimal and ..., then for any slack variable, say w_i, we have $w_i = 0$ implies $\Delta w_i \geq 0$.*

- To paraphrase: for convex problems, as slack variables get small they tend to get large again. This is an antijamming theorem.
- An example of Wächter and Biegler shows that for nonconvex problems, jamming really can occur.
- Recent modification:
  - if a slack variable gets small and
  - its component of the step direction contributes to making a very short step,
  - then increase this slack variable to the average size of the variables the "mainstream" slack variables.
- This modification corrects all examples of jamming that we know about.

Modifications for General Problem Formulations

- Bounds, ranges, and free variables are all treated implicitly as described in *Linear Programming: Foundations and Extensions (LP:F&E)*.
- Net result is following reduced KKT system:

$$\begin{bmatrix} -(H(x,y)+D) & A^T(x) \\ A(x) & E \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \Phi_1 \\ \Phi_2 \end{bmatrix}$$

- Here, D and E are *positive definite* diagonal matrices.
- Note that D helps reduce frequency of diagonal perturbation.
- Choice of barrier parameter μ and initial solution, if none is provided, is described in the book.
- Stopping rules, matrix reordering heuristics, etc. are as described in *LP:F&E*.