Nonlinear Optimization: Algorithms and Models

Robert J. Vanderbei

December 2, 2010

Slides last edited on December 2, 2010

Operations Research and Financial Engineering
Princeton University
Princeton, NJ 08544
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:

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

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:

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

Convex vs. Nonconvex Optimization Probs

Nonlinear Programming (NLP)

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

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 a descent direction for 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, $\|h(x) - w\|$;*
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.
- A recent 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 paper.
- Stopping rules, matrix reordering heuristics, etc. are as described in *LP:F&E*.

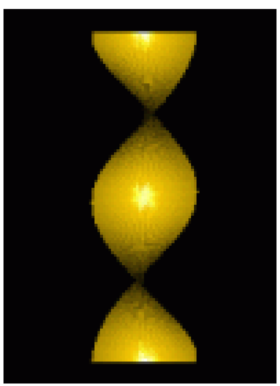
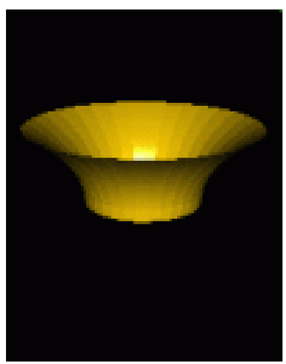
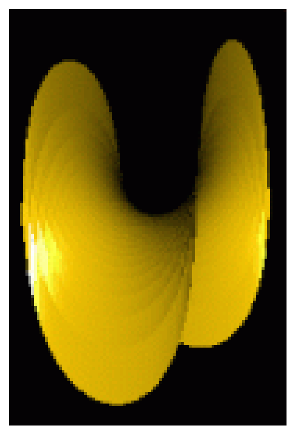
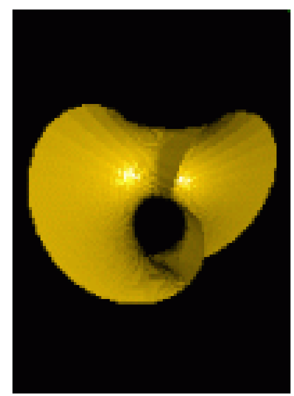
Examples: Convex Optimization Models

Minimal Surfaces

- Given: a domain D in R^2 and an embedding $\mathbf{x} = (x_1, x_2, x_3)$ of its boundary ∂D in R^3;
- Find: an embedding of the entire domain into R^3 that is consistent with the boundary embedding and has minimal surface area:

$$\begin{array}{ll} \text{minimize} & \displaystyle\iint_D \left\| \frac{\partial \mathbf{x}}{\partial s} \times \frac{\partial \mathbf{x}}{\partial t} \right\| ds dt \\ \text{subject to} & \mathbf{x}(s,t) \text{ fixed for } (s,t) \in \partial D \\ & x_1(s,t) \text{ fixed for } (s,t) \in D \\ & x_2(s,t) \text{ fixed for } (s,t) \in D \end{array}$$

The specific problems coded below take D to be either a square or an annulus.

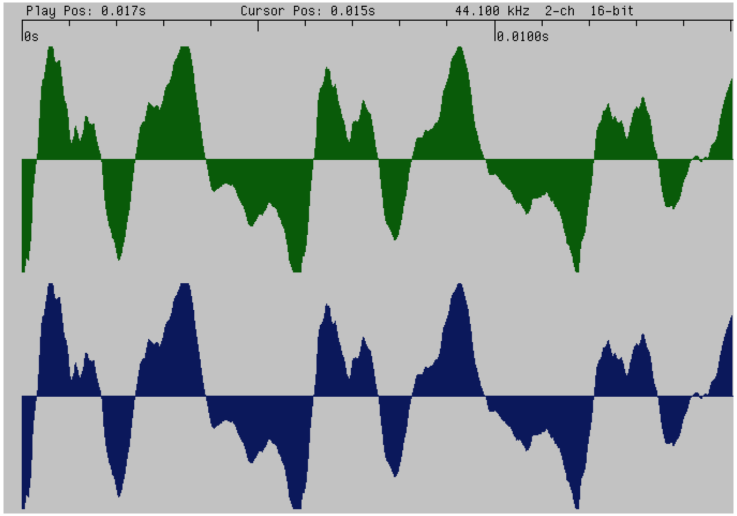
Specific Example

Scherk.mod with D discretized into a 64×64 grid gives the following results:

constraints	0
variables	3844
time (secs)	
LOQO	5.1
LANCELOT	4.0
SNOPT	*

Finite Impulse Response (FIR) Filter Design

- Audio is stored digitally in a computer as a stream of short integers: u_k, $k \in \mathbb{Z}$.
- When the music is played, these integers are used to drive the displacement of the speaker from its resting position.
- For CD quality sound, 44100 short integers get played per second per channel.



0	-32768	8	-23681	16	12111
1	-32768	9	-18449	17	17311
2	-32768	10	-11025	18	21311
3	-30753	11	-6913	19	23055
4	-28865	12	-4337	20	23519
5	-29105	13	-1329	21	25247
6	-29201	14	1743	22	27535
7	-26513	15	6223	23	29471

FIR Filter Design—Continued

- A *finite impulse response (FIR) filter* takes as input a digital signal and convolves this signal with a finite set of fixed numbers $h_{-n}, \ldots, h_n$ to produce a filtered output signal:

$$y_k = \sum_{i=-n}^{n} h_i u_{k-i}.$$

- Sparing the details, the output power at frequency ν is given by

$$|H(\nu)|$$

  where

$$H(\nu) = \sum_{k=-n}^{n} h_k e^{2\pi i k \nu},$$

- Similarly, the mean squared deviation from a flat frequency response over a frequency range, say $\mathcal{L} \subset [0, 1]$, is given by

$$\frac{1}{|\mathcal{L}|} \int_{\mathcal{L}} |H(\nu) - 1|^2 \, d\nu$$

Filter Design: Woofer, Midrange, Tweeter

$$\begin{array}{lll} \text{minimize} & \rho & \\ \text{subject to} & \int_0^1 \left(H_w(\nu) + H_m(\nu) + H_t(\nu) - 1\right)^2 d\nu \le \epsilon & \\ & \left(\frac{1}{|W|}\int_W H_w^2(\nu)d\nu\right)^{1/2} \le \rho & W = [.2, .8] \\ & \left(\frac{1}{|M|}\int_M H_m^2(\nu)d\nu\right)^{1/2} \le \rho & M = [.4, .6] \cup [.9, .1] \\ & \left(\frac{1}{|T|}\int_T H_t^2(\nu)d\nu\right)^{1/2} \le \rho & T = [.7, .3] \end{array}$$

where

$$H_i(\nu) = h_i(0) + 2\sum_{k=1}^{n-1} h_i(k)\cos(2\pi k\nu), \quad i = W, M, T$$

$h_i(k)$ = filter coefficients, i.e., **decision variables**

Specific Example

$$\text{filter length: } n = 14$$
$$\text{integral discretization: } N = 1000$$

constraints	4
variables	43
time (secs)	
LOQO	79
MINOS	164
LANCELOT	3401
SNOPT	35

Ref: J.O. Coleman, U.S. Naval Research Laboratory,

CISS98 paper available:

Click *here* for demo

Examples: Nonconvex Optimization Models

Celestial Mechanics—Periodic Orbits

- Find periodic orbits for the planar gravitational n-body problem.
- Minimize action:

$$\int_0^{2\pi} (K(t) - P(t))dt,$$

- where $K(t)$ is kinetic energy,

$$K(t) = \frac{1}{2}\sum_i \left(\dot{x}_i^2(t) + \dot{y}_i^2(t)\right),$$

- and $P(t)$ is potential energy,

$$P(t) = -\sum_{i<j} \frac{1}{\sqrt{(x_i(t) - x_j(t))^2 + (y_i(t) - y_j(t))^2}}.$$

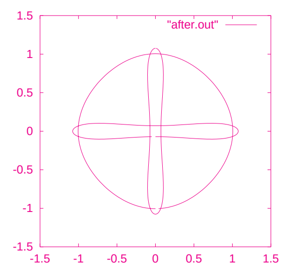
- Subject to periodicity constraints:

$$x_i(2\pi) = x_i(0), \qquad y_i(2\pi) = y_i(0).$$

Specific Example

Orbits.mod with $n = 3$ and $(0, 2\pi)$ discretized into a 160 pieces gives the following results:

constraints	0	
variables	960	
time (secs)		
LOQO	1.1	
LANCELOT	8.7	
SNOPT	287	(no change for last 80% of iterations)



Putting on an Uneven Green

Given:

- $z(x, y)$ elevation of the green.
- Starting position of the ball (x_0, y_0).
- Position of hole (x_f, y_f).
- Coefficient of friction μ.

Find: initial velocity vector so that ball will roll to the hole and arrive with minimal speed.

Variables:

- $u(t) = (x(t), y(t), z(t))$—position as a function of time t.
- $v(t) = (v_x(t), v_y(t), v_z(t))$—velocity.
- $a(t) = (a_x(t), a_y(t), a_z(t))$—acceleration.
- T—time at which ball arrives at hole.

Putting—Two Approaches

- Problem can be formulated with two decision variables:

$$v_x(0) \quad \text{and} \quad v_y(0)$$

  and two constraints:

$$x(T) = x_f \quad \text{and} \quad y(T) = y_f.$$

  In this case, $x(T)$, $y(T)$, and the objective function are complicated functions of the two variables that can only be computed by integrating the appropriate differential equation.

- A discretization of the complete trajectory (including position, velocity, and acceleration) can be taken as variables and the physical laws encoded in the differential equation can be written as constraints.

To implement the first approach, one would need an ode integrator that provides, in addition to the quantities being sought, first and possibly second derivatives of those quantities with respect to the decision variables.
The modern trend is to follow the second approach.

Putting—Continued

Objective:

$$\text{minimize } v_x(T)^2 + v_y(T)^2.$$

Constraints:

$$\begin{aligned} v &= \dot{u} \\ a &= \dot{v} \\ ma &= N + F - mge_z \\ u(0) &= u_0 \qquad u(T) \;=\; u_f, \end{aligned}$$

where

- m is the mass of the golf ball.
- g is the acceleration due to gravity.
- e_z is a unit vector in the positive z direction.

and ...

Putting—Continued

- $N = (N_x, N_y, N_z)$ is the normal force:

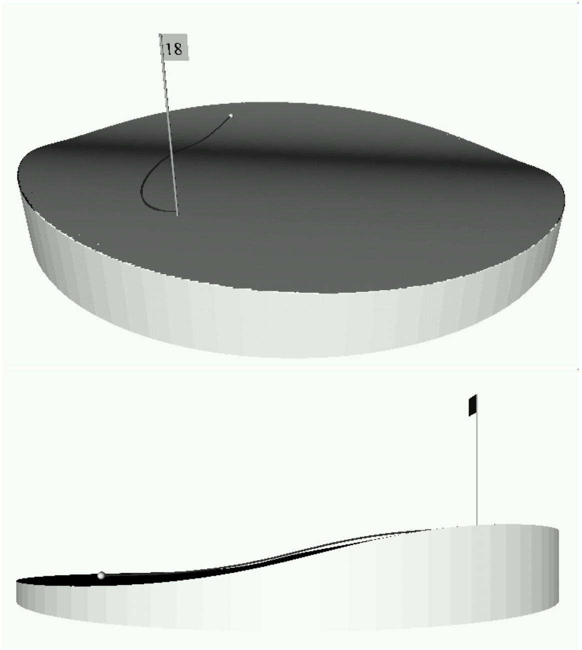
$$\begin{aligned} N_z &= m\frac{g - a_x(t)\frac{\partial z}{\partial x} - a_y(t)\frac{\partial z}{\partial y} + a_z(t)}{(\frac{\partial z}{\partial x})^2 + (\frac{\partial z}{\partial y})^2 + 1} \\ N_x &= -\frac{\partial z}{\partial x} N_z \\ N_y &= -\frac{\partial z}{\partial y} N_z. \end{aligned}$$

- F is the force due to friction:

$$F = -\mu \|N\| \frac{v}{\|v\|}.$$

Putting—Specific Example

- Discretize continuous time into $n = 200$ discrete time points.
- Use finite differences to approximate the derivatives.



constraints	597
variables	399
time (secs)	
LOQO	14.1
LANCELOT	> 600.0
SNOPT	4.1

Goddard Rocket Problem

Objective:

$$\text{maximize } h(T);$$

Constraints:

$$\begin{aligned}
v &= \dot{h} \\
a &= \dot{v} \\
\theta &= -c\dot{m} \\
ma &= (\theta - \sigma v^2 e^{-h/h_0}) - gm \\
0 &\leq \theta \leq \theta_{\max} \\
m &\geq m_{\min} \\
h(0) &= 0 \qquad v(0) = 0 \qquad m(0) = 3
\end{aligned}$$

where

- $\theta =$ Thrust , $m =$ mass
- $\theta_{\max}$, g, σ, c, and h_0 are given constants
- h, v, a, T_h, and m are functions of time $0 \leq t \leq T$.

Goddard Rocket Problem—Solution

constraints	399
variables	599
time (secs)	
LOQO	5.2
LANCELOT	(IL)
SNOPT	(IL)