An Inexact Newton Method for Optimization

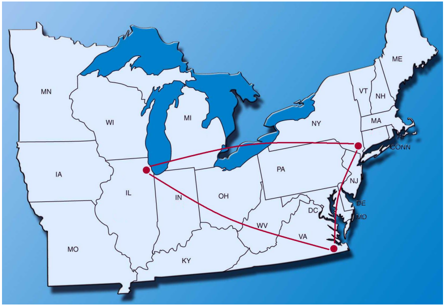
Frank E. Curtis

New York University

Lehigh Industrial Engineering Seminar, February 13, 2009

Brief biography

New York State
↓
College of William and Mary (B.S.)
↓
Northwestern University (M.S. & Ph.D.)
↓
Courant Institute (Postdoc)



Research: Nonlinear Optimization, algorithms and theory

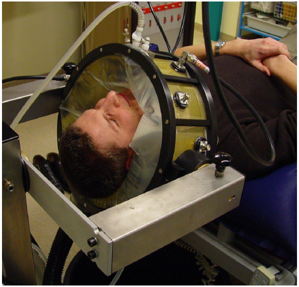
- Large-scale optimization (PDE-constrained problems, today's talk)
- Methods with fast detection of infeasibility (MINLP problems)
- Global convergence mechanisms

Outline

Outline

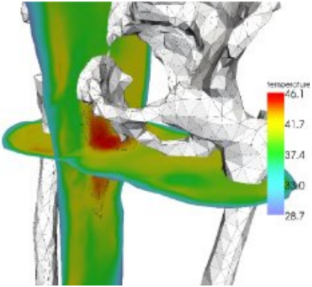
Hyperthermia treatment

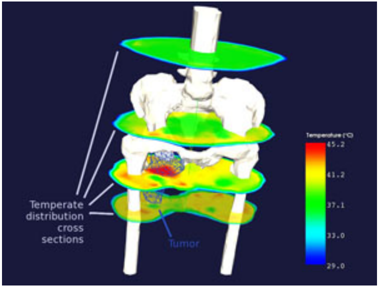
- Regional hyperthermia is a cancer therapy that aims at heating large and deeply seated tumors by means of radio wave adsorption
- Results in the killing of tumor cells and makes them more susceptible to other accompanying therapies; e.g., chemotherapy



Hyperthermia treatment planning

- Computer modeling can be used to help plan the therapy for each patient, and it opens the door for numerical optimization
- The goal is to heat the tumor to the target temperature of 43°C while minimizing damage to nearby cells





Hyperthermia treatment as an optimization problem

The problem is to

$$\min_{y,u} \int_\Omega (y - y_t)^2 dV \quad \text{where} \quad y_t = \begin{cases} 37 & \text{in } \Omega \backslash \Omega_0 \\ 43 & \text{in } \Omega_0 \end{cases}$$

subject to the bio-heat transfer equation (Pennes (1948))

$$-\underbrace{\nabla \cdot (\kappa \nabla y)}_{\text{thermal conductivity}} + \underbrace{\omega(y)\pi(y - y_b)}_{\text{effects of blood flow}} = \underbrace{\frac{\sigma}{2}\left|\sum_i u_i E_i\right|^2}_{\text{electromagnetic field}}, \quad \text{in } \Omega$$

and the bound constraints

$$\begin{aligned} 37.0 \leq y \leq 37.5, &\quad \text{on } \partial\Omega \\ 41.0 \leq y \leq 45.0, &\quad \text{in } \Omega_0 \end{aligned}$$

where Ω_0 is the tumor domain

Large-scale optimization

- Consider

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & f(x) \\ \text{s.t. } & c_{\mathcal{E}}(x) = 0 \\ & c_{\mathcal{I}}(x) \geq 0 \end{aligned}$$

  where $f : \mathbb{R}^n \to \mathbb{R}$, $c_{\mathcal{E}} : \mathbb{R}^n \to \mathbb{R}^p$ and $c_{\mathcal{I}} : \mathbb{R}^n \to \mathbb{R}^q$ are smooth functions
- The best contemporary methods are limited by problem size; e.g.,
  - sequential quadratic programming (small to moderate sizes)
  - interior-point methods (moderate to large sizes)
- We want the fast solution of problems with millions of variables

Challenges in large-scale optimization

- Computational issues:
  - Large matrices may not be stored
  - Large matrices may not be factored
- Algorithmic issues:
  - The problem may be nonconvex
  - The problem may be ill-conditioned
- Computational/Algorithmic issues:
  - No matrix factorizations makes difficulties more difficult

Main contributions

- ALGORITHMS: Inexact Newton methods for constrained optimization, broadening the potential application of fast optimization algorithms
- THEORY: Global convergence and the potential for fast local convergence
- SOFTWARE: new release of `Ipopt` (Wächter) with `Pardiso` (Schenk)
- ARTICLES:
  - "An Inexact SQP Method for Equality Constrained Optimization," *SIAM Journal on Optimization*, 19(1):351–369, with R. H. Byrd and J. Nocedal
  - "An Inexact Newton Method for Nonconvex Equality Constrained Optimization," *Mathematical Programming, Series A*, to appear, with R. H. Byrd and J. Nocedal
  - "A Matrix-free Algorithm for Equality Constrained Optimization Problems with Rank-Deficient Jacobians," *SIAM Journal on Optimization*, to appear, with J. Nocedal and A. Wächter
  - "An Interior-Point Algorithm for Large-Scale Nonlinear Optimization with Inexact Step Computations," submitted to *SIAM Journal on Scientific Computing*, with O. Schenk and A. Wächter

Outline

Equality constrained optimization

- Consider

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t. } c(x) = 0$$

- The Lagrangian is

$$\mathcal{L}(x, \lambda) \triangleq f(x) + \lambda^T c(x)$$

so the first-order optimality conditions are

$$\nabla \mathcal{L}(x, \lambda) = \begin{bmatrix} \nabla f(x) + \nabla c(x)\lambda \\ c(x) \end{bmatrix} \triangleq \mathcal{F}(x, \lambda) = 0$$

Inexact Newton methods

- Solve

$$\mathcal{F}(x,\lambda) = 0 \quad \text{or} \quad \min\ \varphi(x,\lambda) \triangleq \tfrac{1}{2}\|\mathcal{F}(x,\lambda)\|^2$$

- Inexact Newton methods compute

$$\nabla\mathcal{F}(x_k,\lambda_k)d_k = -\mathcal{F}(x_k,\lambda_k) + r_k$$

requiring (Dembo, Eisenstat, Steihaug (1982))

$$\|r_k\| \le \kappa\|\mathcal{F}(x_k,\lambda_k)\|, \quad \kappa \in (0,1)$$

A naïve Newton method for optimization

- Consider the problem

$$\min \; f(x) = x_1 + x_2, \quad \text{s.t. } c(x) = x_1^2 + x_2^2 - 1 = 0$$

that has the first-order optimality conditions

$$\mathcal{F}(x,\lambda) = \begin{bmatrix} 1 + 2x_1\lambda \\ 1 + 2x_2\lambda \\ x_1^2 + x_2^2 - 1 \end{bmatrix} = 0$$

- A Newton method applied to this problem yields

k	$\frac{1}{2}\|\mathcal{F}(x_k,\lambda_k)\|^2$
0	+3.5358e+00
1	+2.9081e-02
2	+4.8884e-04
3	+7.9028e-08
4	+2.1235e-15

A naïve Newton method for optimization

- Consider the problem

$$\min \; f(x) = x_1 + x_2, \quad \text{s.t. } c(x) = x_1^2 + x_2^2 - 1 = 0$$

that has the first-order optimality conditions

$$\mathcal{F}(x, \lambda) = \begin{bmatrix} 1 + 2x_1\lambda \\ 1 + 2x_2\lambda \\ x_1^2 + x_2^2 - 1 \end{bmatrix} = 0$$

- A Newton method applied to this problem yields

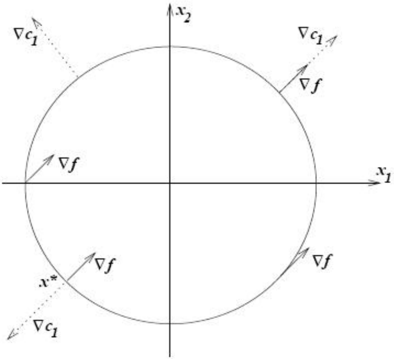
k	$\frac{1}{2}\|\mathcal{F}(x_k, \lambda_k)\|^2$
0	+3.5358e+00
1	+2.9081e-02
2	+4.8884e-04
3	+7.9028e-08
4	+2.1235e-15

k	$f(x_k)$	$\|c(x_k)\|$
0	+1.3660e+00	+1.1102e-16
1	+1.3995e+00	+8.3734e-03
2	+1.4358e+00	+3.0890e-02
3	+1.4143e+00	+2.4321e-04
4	+1.4142e+00	+1.7258e-08

A naïve Newton method for optimization fails easily

- Consider the problem

$$\min \; f(x) = x_1 + x_2, \quad \text{s.t. } c(x) = x_1^2 + x_2^2 - 1 = 0$$



Merit function

- Simply minimizing

$$\varphi(x, \lambda) = \tfrac{1}{2}\|\mathcal{F}(x, \lambda)\|^2 = \tfrac{1}{2}\left\| \begin{bmatrix} \nabla f(x) + \nabla c(x)\lambda \\ c(x) \end{bmatrix} \right\|^2$$

  is generally inappropriate for optimization

- We use the merit function

$$\phi(x; \pi) \triangleq f(x) + \pi\|c(x)\|$$

  where π is a penalty parameter

Algorithm 0: Newton method for optimization

(Assume the problem is convex and regular)
for $k = 0, 1, 2, \ldots$

- Solve the primal-dual (Newton) equations

$$\begin{bmatrix} H(x_k, \lambda_k) & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = - \begin{bmatrix} \nabla f(x_k) + \nabla c(x_k)\lambda_k \\ c(x_k) \end{bmatrix}$$

- Increase π, if necessary, so that $\pi_k \geq \|\lambda_k + \delta_k\|$ (yields $D\phi_k(d_k; \pi_k) \ll 0$)
- Backtrack from $\alpha_k \leftarrow 1$ to satisfy the Armijo condition

$$\phi(x_k + \alpha_k d_k; \pi_k) \leq \phi(x_k; \pi_k) + \eta \alpha_k D\phi_k(d_k; \pi_k)$$

- Update iterate $(x_{k+1}, \lambda_{k+1}) \leftarrow (x_k, \lambda_k) + \alpha_k(d_k, \delta_k)$

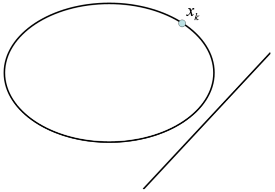
Newton methods and sequential quadratic programming

If $H(x_k, \lambda_k)$ is positive definite on the null space of $\nabla c(x_k)^T$, then

$$\begin{bmatrix} H(x_k, \lambda_k) & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} d \\ \delta \end{bmatrix} = - \begin{bmatrix} \nabla f(x_k) + \nabla c(x_k)\lambda_k \\ c(x_k) \end{bmatrix}$$

is equivalent to

$$\min_{d \in \mathbb{R}^n} \; f(x_k) + \nabla f(x_k)^T d + \tfrac{1}{2} d^T H(x_k, \lambda_k) d$$
$$\text{s.t. } c(x_k) + \nabla c(x_k)^T d = 0$$



Minimizing a penalty function

Consider the penalty function for

$$\min\ (x-1)^2,\ \text{s.t.}\ x = 0 \quad \text{i.e.} \quad \phi(x;\pi) = (x-1)^2 + \pi|x|$$

for different values of the penalty parameter π

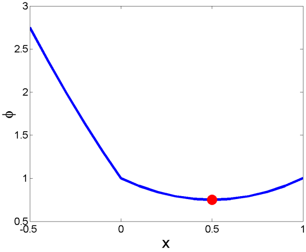


Figure: $\pi = 1$

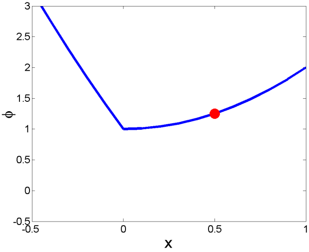


Figure: $\pi = 2$

Convergence of Algorithm 0

Assumption

The sequence $\{(x_k, \lambda_k)\}$ is contained in a convex set Ω over which f, c, and their first derivatives are bounded and Lipschitz continuous. Also,

- *(Regularity) $\nabla c(x_k)^T$ has full row rank with singular values bounded below by a positive constant*
- *(Convexity) $u^T H(x_k, \lambda_k) u \geq \mu \|u\|^2$ for $\mu > 0$ for all $u \in \mathbb{R}^n$ satisfying $u \neq 0$ and $\nabla c(x_k)^T u = 0$*

Theorem

(Han (1977)) The sequence $\{(x_k, \lambda_k)\}$ yields the limit

$$\lim_{k \to \infty} \left\| \begin{bmatrix} \nabla f(x_k) + \nabla c(x_k)\lambda_k \\ c(x_k) \end{bmatrix} \right\| = 0$$

Incorporating inexactness

- Iterative as opposed to direct methods
- Compute

$$\begin{bmatrix} H(x_k,\lambda_k) & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k) + \nabla c(x_k)\lambda_k \\ c(x_k) \end{bmatrix} + \begin{bmatrix} \rho_k \\ r_k \end{bmatrix}$$

satisfying

$$\left\| \begin{bmatrix} \rho_k \\ r_k \end{bmatrix} \right\| \leq \kappa \left\| \begin{bmatrix} \nabla f(x_k) + \nabla c(x_k)\lambda_k \\ c(x_k) \end{bmatrix} \right\|, \quad \kappa \in (0,1)$$

- If κ is not sufficiently small (e.g., 10^{-3} vs. 10^{-12}), then d_k may be an ascent direction for our merit function; i.e.,

$$D\phi_k(d_k;\pi_k) > 0 \quad \text{for all } \pi_k \geq \pi_{k-1}$$

Model reductions

- Define the model of $\phi(x;\pi)$:

$$m(d;\pi) \triangleq f(x) + \nabla f(x)^T d + \pi(\|c(x) + \nabla c(x)^T d\|)$$

- d_k is acceptable if

$$\begin{aligned}\Delta m(d_k;\pi_k) &\triangleq m(0;\pi_k) - m(d_k;\pi_k) \\ &= -\nabla f(x_k)^T d_k + \pi_k(\|c(x_k)\| - \|c(x_k) + \nabla c(x_k)^T d_k\|) \gg 0\end{aligned}$$

- This ensures $D\phi_k(d_k;\pi_k) \ll 0$ (and more)

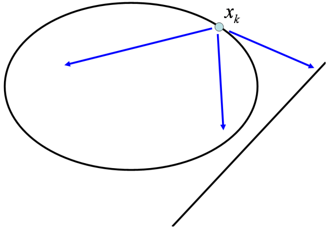
Termination test 1

The search direction (d_k, δ_k) is acceptable if

$$\left\| \begin{bmatrix} \rho_k \\ r_k \end{bmatrix} \right\| \leq \kappa \left\| \begin{bmatrix} \nabla f(x_k) + \nabla c(x_k)\lambda_k \\ c(x_k) \end{bmatrix} \right\|, \quad \kappa \in (0,1)$$

and if for $\pi_k = \pi_{k-1}$ and some $\sigma \in (0,1)$ we have

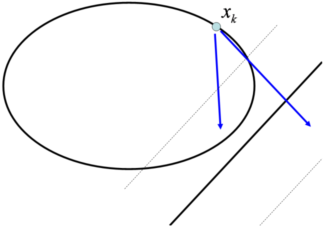
$$\Delta m(d_k;\pi_k) \geq \underbrace{\max\{\tfrac{1}{2} d_k^T H(x_k,\lambda_k) d_k, 0\} + \sigma\pi_k \max\{\|c(x_k)\|, \|r_k\| - \|c(x_k)\|\}}_{\geq 0 \text{ for any } d}$$



Termination test 2

The search direction (d_k, δ_k) is acceptable if

$$\begin{aligned} \|\rho_k\| &\leq \beta\|c(x_k)\|, \quad \beta > 0 \\ \text{and} \quad \|r_k\| &\leq \epsilon\|c(x_k)\|, \quad \epsilon \in (0,1) \end{aligned}$$



Increasing the penalty parameter π then yields

$$\Delta m(d_k; \pi_k) \geq \underbrace{\max\{\tfrac{1}{2} d_k^T H(x_k, \lambda_k) d_k, 0\} + \sigma\pi_k\|c(x_k)\|}_{\geq 0 \text{ for any } d}$$

Algorithm 1: Inexact Newton for optimization

(Byrd, Curtis, Nocedal (2008))

for $k = 0, 1, 2, \ldots$

- Iteratively solve

$$\begin{bmatrix} H(x_k,\lambda_k) & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = - \begin{bmatrix} \nabla f(x_k) + \nabla c(x_k)\lambda_k \\ c(x_k) \end{bmatrix}$$

  until termination test 1 or 2 is satisfied

- If only termination test 2 is satisfied, increase π so

$$\pi_k \geq \max\left\{ \pi_{k-1}, \frac{\nabla f(x_k)^T d_k + \max\{\frac{1}{2} d_k^T H(x_k,\lambda_k) d_k, 0\}}{(1-\tau)(\|c(x_k)\| - \|r_k\|)} \right\}$$

- Backtrack from $\alpha_k \leftarrow 1$ to satisfy

$$\phi(x_k + \alpha_k d_k; \pi_k) \leq \phi(x_k; \pi_k) - \eta \alpha_k \Delta m(d_k; \pi_k)$$

- Update iterate $(x_{k+1}, \lambda_{k+1}) \leftarrow (x_k, \lambda_k) + \alpha_k (d_k, \delta_k)$

Convergence of Algorithm 1

Assumption

The sequence $\{(x_k, \lambda_k)\}$ is contained in a convex set Ω over which f, c, and their first derivatives are bounded and Lipschitz continuous. Also,

- *(Regularity) $\nabla c(x_k)^T$ has full row rank with singular values bounded below by a positive constant*
- *(Convexity) $u^T H(x_k, \lambda_k) u \geq \mu \|u\|^2$ for $\mu > 0$ for all $u \in \mathbb{R}^n$ satisfying $u \neq 0$ and $\nabla c(x_k)^T u = 0$*

Theorem

(Byrd, Curtis, Nocedal (2008)) The sequence $\{(x_k, \lambda_k)\}$ yields the limit

$$\lim_{k \to \infty} \left\| \begin{bmatrix} \nabla f(x_k) + \nabla c(x_k) \lambda_k \\ c(x_k) \end{bmatrix} \right\| = 0$$

Handling nonconvexity and rank deficiency

- There are two assumptions we aim to drop:
  - *(Regularity)* $\nabla c(x_k)^T$ *has full row rank with singular values bounded below by a positive constant*
  - *(Convexity)* $u^T H(x_k, \lambda_k) u \geq \mu \|u\|^2$ *for* $\mu > 0$ *for all* $u \in \mathbb{R}^n$ *satisfying* $u \neq 0$ *and* $\nabla c(x_k)^T u = 0$

  e.g., the problem is not regular if it is infeasible, and it is not convex if there are maximizers and/or saddle points
- Without them, Algorithm 1 may stall or may not be well-defined

No factorizations means no clue

- We might not store or factor

$$\begin{bmatrix} H(x_k, \lambda_k) & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix}$$

so we might not know if the problem is nonconvex or ill-conditioned

- Common practice is to perturb the matrix to be

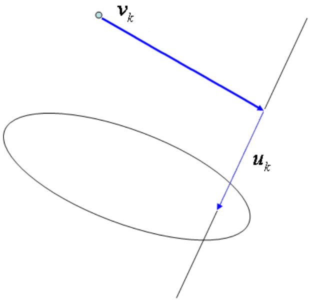
$$\begin{bmatrix} H(x_k, \lambda_k) + \xi_1 I & \nabla c(x_k) \\ \nabla c(x_k)^T & -\xi_2 I \end{bmatrix}$$

where ξ_1 convexifies the model and ξ_2 regularizes the constraints

- Poor choices of ξ_1 and ξ_2 can have terrible consequences in the algorithm

Our approach for global convergence

- Decompose the direction d_k into a normal component (toward the constraints) and a tangential component (toward optimality)



- Without convexity, we do not guarantee a minimizer, but our merit function biases the method to avoid maximizers and saddle points

Normal component computation

- (Approximately) solve

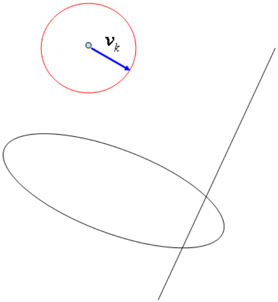
$$\begin{aligned} &\min\ \tfrac{1}{2}\|c(x_k) + \nabla c(x_k)^T v\|^2 \\ &\text{s.t. } \|v\| \le \omega \|(\nabla c(x_k)) c(x_k)\| \end{aligned}$$

  for some $\omega > 0$

- We only require Cauchy decrease:

$$\begin{aligned} &\|c(x_k)\| - \|c(x_k) + \nabla c(x_k)^T v_k\| \\ &\quad \ge \epsilon_v (\|c(x_k)\| - \|c(x_k) + \alpha \nabla c(x_k)^T \tilde{v}_k\|) \end{aligned}$$

  for $\epsilon_v \in (0, 1)$, where $\tilde{v}_k = -(\nabla c(x_k)) c(x_k)$ is the direction of steepest descent



Tangential component computation (idea #1)

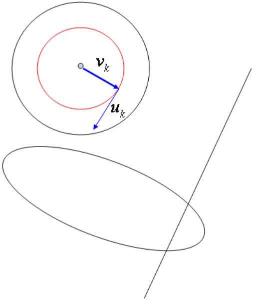
- Standard practice is to then (approximately) solve

$$\min \ (\nabla f(x_k) + H(x_k, \lambda_k) v_k)^T u + \tfrac{1}{2} u^T H(x_k, \lambda_k) u$$
$$\text{s.t. } \nabla c(x_k)^T u = 0, \ \ \|u\| \leq \Delta_k$$

- However, maintaining

$$\nabla c(x_k)^T u \approx 0 \quad \text{and} \quad \|u\| \leq \Delta_k$$

can be expensive

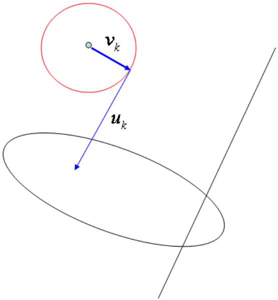


Tangential component computation

- Instead, we formulate the primal-dual system

$$\begin{bmatrix} H(x_k,\lambda_k) & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} u_k \\ \delta_k \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k) + \nabla c(x_k)\lambda_k + H(x_k,\lambda_k)v_k \\ 0 \end{bmatrix}$$

- Our ideas from before apply!



Handling nonconvexity

- Convexify the Hessian as in

$$\begin{bmatrix} H(x_k, \lambda_k) + \xi_1 I & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix}$$

  by monitoring iterates

- Hessian modification strategy: Increase ξ_1 whenever

$$\begin{aligned} \|u_k\|^2 &> \psi \|v_k\|^2, \quad \psi > 0 \\ \tfrac{1}{2} u_k^T (H(x_k, \lambda_k) + \xi_1 I) u_k &< \theta \|u_k\|^2, \quad \theta > 0 \end{aligned}$$

Inexact Newton Algorithm 2

(Curtis, Nocedal, Wächter (2009))
for $k = 0, 1, 2, \ldots$

- Approximately solve
$$\min \ \tfrac{1}{2}\|c(x_k) + \nabla c(x_k)^T v\|^2, \quad \text{s.t. } \|v\| \leq \omega \|(\nabla c(x_k))c(x_k)\|$$
to compute v_k satisfy Cauchy decrease
- Iteratively solve
$$\begin{bmatrix} H(x_k, \lambda_k) + \xi_1 I & \nabla c(x_k) \\ \nabla c(x_k)^T & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = - \begin{bmatrix} \nabla f(x_k) + \nabla c(x_k)\lambda_k \\ -\nabla c(x_k)^T v_k \end{bmatrix}$$
until termination test 1 or 2 is satisfied, increasing ξ_1 as described
- If only termination test 2 is satisfied, increase π so
$$\pi_k \geq \max\left\{ \pi_{k-1}, \frac{\nabla f(x_k)^T d_k + \max\{\frac{1}{2} u_k^T (H(x_k, \lambda_k) + \xi_1 I) u_k, \theta \|u_k\|^2\}}{(1-\tau)(\|c(x_k)\| - \|c(x_k) + \nabla c(x_k)^T d_k\|)} \right\}$$
- Backtrack from $\alpha_k \leftarrow 1$ to satisfy
$$\phi(x_k + \alpha_k d_k; \pi_k) \leq \phi(x_k; \pi_k) - \eta \alpha_k \Delta m(d_k; \pi_k)$$
- Update iterate $(x_{k+1}, \lambda_{k+1}) \leftarrow (x_k, \lambda_k) + \alpha_k (d_k, \delta_k)$

Convergence of Algorithm 2

Assumption

The sequence $\{(x_k, \lambda_k)\}$ is contained in a convex set Ω over which f, c, and their first derivatives are bounded and Lipschitz continuous

Theorem

(Curtis, Nocedal, Wächter (2009)) If all limit points of $\{\nabla c(x_k)^T\}$ have full row rank, then the sequence $\{(x_k, \lambda_k)\}$ yields the limit

$$\lim_{k\to\infty} \left\| \begin{bmatrix} \nabla f(x_k) + \nabla c(x_k)\lambda_k \\ c(x_k) \end{bmatrix} \right\| = 0.$$

Otherwise,

$$\lim_{k\to\infty} \|(\nabla c(x_k))c(x_k)\| = 0$$

and if $\{\pi_k\}$ is bounded, then

$$\lim_{k\to\infty} \|\nabla f(x_k) + \nabla c(x_k)\lambda_k\| = 0$$

Handling inequalities

- Interior point methods are attractive for large applications
- Line-search interior point methods that enforce

$$c(x_k) + \nabla c(x_k)^T d_k = 0$$

may fail to converge globally (Wächter, Biegler (2000))

- Fortunately, the trust region subproblem we use to regularize the constraints also saves us from this type of failure!

Algorithm 2 (Interior-point version)

- Apply Algorithm 2 to the logarithmic-barrier subproblem

$$\min \; f(x) - \mu \sum_{i=1}^{q} \ln s^i, \quad \text{s.t. } c_{\mathcal{E}}(x) = 0, \; c_{\mathcal{I}}(x) - s = 0$$

  for $\mu \to 0$

- Define

$$\begin{bmatrix} H(x_k, \lambda_{\mathcal{E},k}, \lambda_{\mathcal{I},k}) & 0 & \nabla c_{\mathcal{E}}(x_k) & \nabla c_{\mathcal{I}}(x_k) \\ 0 & \mu I & 0 & -S_k \\ \nabla c_{\mathcal{E}}(x_k)^T & 0 & 0 & 0 \\ \nabla c_{\mathcal{I}}(x_k)^T & -S_k & 0 & 0 \end{bmatrix} \begin{bmatrix} d_k^x \\ d_k^s \\ \delta_{\mathcal{E},k} \\ \delta_{\mathcal{I},k} \end{bmatrix}$$

  so that the iterate update has

$$\begin{bmatrix} x_{k+1} \\ s_{k+1} \end{bmatrix} \leftarrow \begin{bmatrix} x_k \\ s_k \end{bmatrix} + \alpha_k \begin{bmatrix} d_k^x \\ S_k d_k^s \end{bmatrix}$$

- Incorporate a fraction-to-the-boundary rule in the line search and a slack reset in the algorithm to maintain $s \geq \max\{0, c_{\mathcal{I}}(x)\}$

Convergence of Algorithm 2 (Interior-point)

Assumption

The sequence $\{(x_k, \lambda_{\mathcal{E},k}, \lambda_{\mathcal{I},k})\}$ is contained in a convex set Ω over which f, $c_{\mathcal{E}}$, $c_{\mathcal{I}}$, and their first derivatives are bounded and Lipschitz continuous

Theorem

(Curtis, Schenk, Wächter (2009))

- *For a given μ, Algorithm 2 yields the same limits as in the equality constrained case*
- *If Algorithm 2 yields a sufficiently accurate solution to the barrier subproblem for each $\{\mu_j\} \to 0$ and if the linear independence constraint qualification (LICQ) holds at a limit point $\bar{x}$ of $\{x_j\}$, then there exist Lagrange multipliers $\bar{\lambda}$ such that the first-order optimality conditions of the nonlinear program are satisfied*

Outline

Implementation details

- Incorporated in `IPOPT` software package (Wächter)
- Linear systems solved with `PARDISO` (Schenk)
  - Symmetric quasi-minimum residual method (Freund (1994))
- PDE-constrained model problems
  - 3D grid $\Omega = [0,1] \times [0,1] \times [0,1]$
  - Equidistant Cartesian grid with N grid points
  - 7-point stencil for discretization

Boundary control problem

$$
\begin{aligned}
\min\ & \tfrac{1}{2}\int_\Omega (y(x)-y_t(x))^2 dx, && // \ y_t(x) = 3 + 10x_1(x_1-1)x_2(x_2-1)\sin(2\pi x_3) \\
\text{s.t.}\ & -\nabla\cdot(e^{y(x)}\cdot\nabla y(x)) = 20, \quad \text{in } \Omega && \\
& y(x) = u(x), \quad \text{on } \partial\Omega, && // \ u(x) \text{ defined on } \partial\Omega \\
& 2.5 \le u(x) \le 3.5, \quad \text{on } \partial\Omega &&
\end{aligned}
$$

N	n	p	q	# nnz	f^*	# iter	CPU sec
20	8000	5832	4336	95561	1.3368e-2	12	33.4
30	27000	21952	10096	339871	1.3039e-2	12	139.4
40	64000	54872	18256	827181	1.2924e-2	12	406.0
50	125000	110592	28816	1641491	1.2871e-2	12	935.6
60	216000	195112	41776	2866801	1.2843e-2	13	1987.2
(direct) 40	64000	54872	18256	827181	1.2924e-2	10	3196.3

Hyperthermia Treatment Planning

$$\min \ \frac{1}{2}\int_\Omega (y(x) - y_t(x))^2 dx, \qquad // \ y_t(x) = \begin{cases} 37 & \text{in } \Omega\backslash\Omega_0 \\ 43 & \text{in } \Omega_0 \end{cases}$$

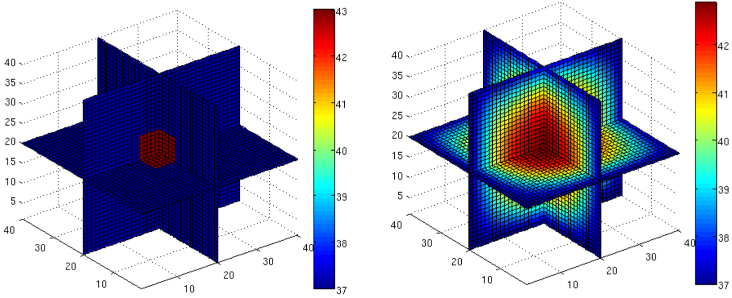
$$\text{s.t.} \ -\Delta y(x) - 10(y(x) - 37) = u^* M(x) u, \quad \text{in } \Omega \quad // \begin{cases} u_j = a_j e^{i\phi_j} \\ M_{jk}(x) = < E_j(x), E_k(x) > \\ E_j = \sin(j x_1 x_2 x_3 \pi) \end{cases}$$

$$37.0 \leq y(x) \leq 37.5, \quad \text{on } \partial\Omega$$

$$42.0 \leq y(x) \leq 44.0, \quad \text{in } \Omega_0, \qquad // \ \Omega_0 = [3/8, 5/8]^3$$

N	n	p	q	# nnz	f^*	# iter	CPU sec
10	1020	512	1070	20701	2.3037	40	15.0
20	8020	5832	4626	212411	2.3619	62	564.7
30	27020	21952	10822	779121	2.3843	146	4716.5
40	64020	54872	20958	1924831	2.6460	83	9579.7
(direct) 30	27020	21952	10822	779121	2.3719	91	10952.4

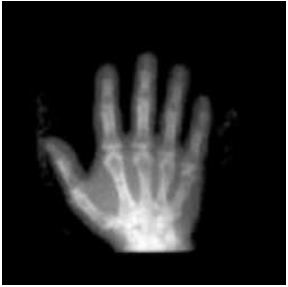
Sample solution for $N = 40$

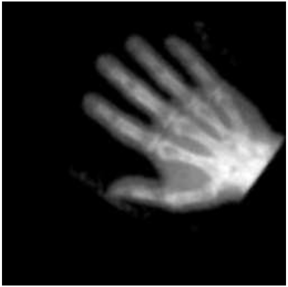


Numerical experiments (currently underway)

Joint with Andreas Wächter (IBM) and Olaf Schenk (U. of Basel)

- Hyperthermia treatment planning with real patient geometry (with Matthias Christen, U. of Basel)
- Image registration (with Stefan Heldmann, U. of Lübeck)





Outline

Conclusion and final remarks

- Inexact Newton method for optimization with theoretical foundation
- Convergence guarantees are as good as exact methods, sometimes better
- Numerical experiments are promising so far, and more to come