AMSC 607 / CMSC 764 Advanced Numerical Optimization

Fall 2010

UNIT 4: Special Topics

Constraint Reduction

Dianne P. O'Leary

Adaptive Constraint Reduction for Linear Programming Problems

---

Motivation

Many constrained optimization problems have a great number of (dual) constraints.

For example, if we have a model matrix $\mathbf{H}$ and a data vector $\mathbf{g}$, we might try to determine the model parameters $\mathbf{u}$ to solve

$$\min_{\mathbf{u}} \|\mathbf{H}\mathbf{u} - \mathbf{g}\|_\infty$$

which can be formulated as

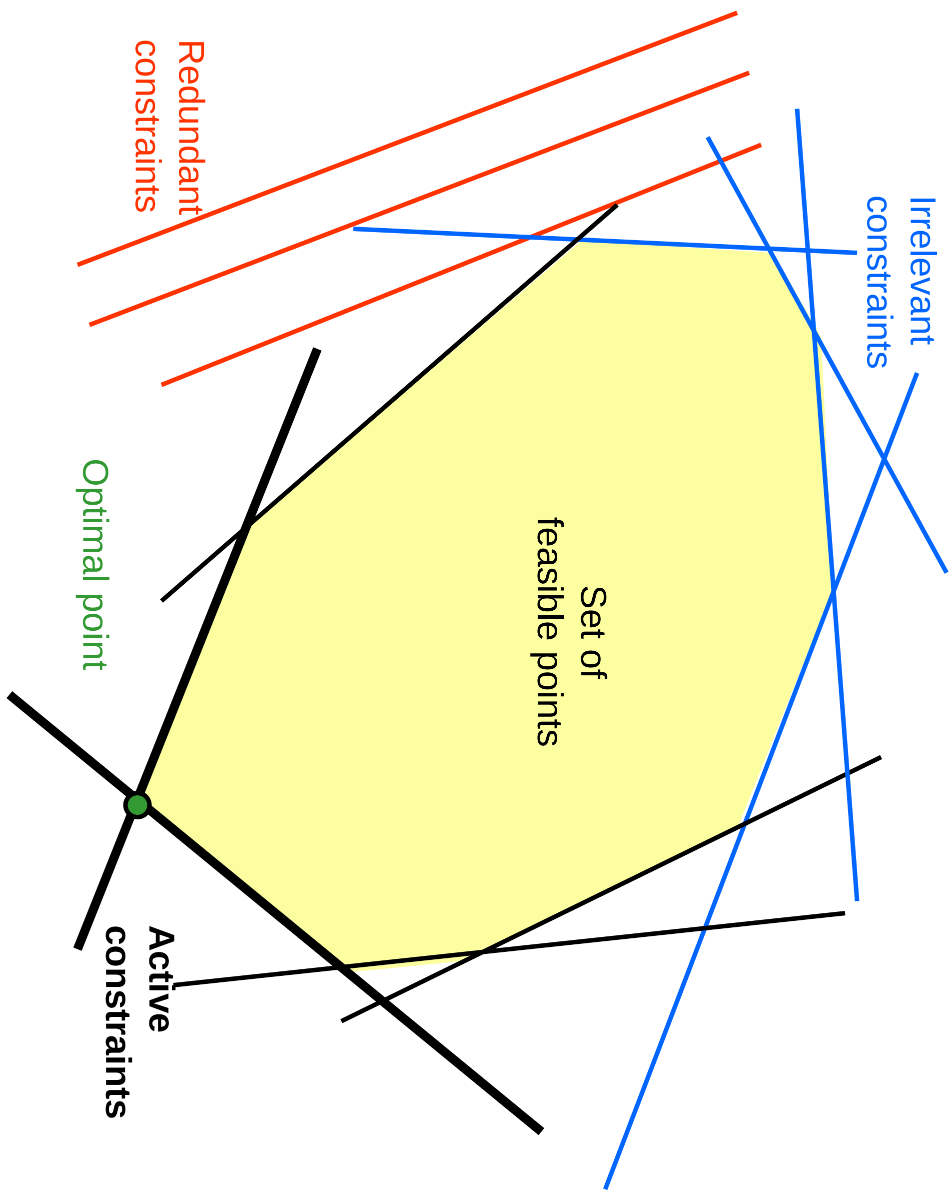
$$\min_{\mathbf{u},t} t$$

subject to

$$-t\mathbf{e} \leq \mathbf{H}\mathbf{u} - \mathbf{g} \leq t\mathbf{e}.$$

Usually there are a small number of parameters $\mathbf{u}$ but a large number of data values $\mathbf{g}$.

How many constraints will be active at the solution?


Redundant
constraints
Irrelevant
constraints
Optimal point
Set of
feasible points
Active
constraints

The Bones of an Interior Point Method (IPM) for Constrained Optimization

---

Given: An initial guess at the solution variables (primal variables) and the Lagrange multipliers (dual variables)
while no convergence
  Use Newton's method, applied to the optimality conditions or a perturbed variant of them, to generate a search direction.
  Update the primal and dual variables, taking a step in the Newton direction.
end

Main work: Formation of the matrix for Newton's method.

"Reduction"

It can make sense to adaptively reduce:

- far-away (dual) constraints.
- (primal) variables that are near zero.
- small terms in the matrix defining the Newton direction.

Linear Programming: The Idea

Primal Problem	Dual Problem
$\min \mathbf{c}^T\mathbf{x}$	$\max \mathbf{b}^T\mathbf{y}$
$\mathbf{A}\mathbf{x} = \mathbf{b},$	$\mathbf{A}^T\mathbf{y} \leq \mathbf{c}.$
$\mathbf{x} \geq \mathbf{0}.$	

- $\mathbf{A}$ is $m \times n$, with $n \gg m$.
- $\mathbf{A}$ is normalized so that each column has norm 1: $\|\mathbf{a}_j\| = 1$.

The KKT system of optimality conditions ($\tau = 0$):

$$\begin{aligned} \mathbf{A}^T\mathbf{y} + \mathbf{s} - \mathbf{c} &= \mathbf{0}, \\ \mathbf{A}\mathbf{x} - \mathbf{b} &= \mathbf{0}, \\ \mathbf{X}\mathbf{s} - \tau\mathbf{e} &= \mathbf{0}, \\ (\mathbf{x}, \mathbf{s}) &\geq \mathbf{0}, \end{aligned}$$

where $\mathbf{X} = \mathrm{diag}(\mathbf{x})$ and $\mathbf{e}$ is the vector of ones.

Interior Point Methods

IPMs construct a sequence of iterates, using Newton's method, converging to a solution to the KKT system.

The heart of the algorithm is generation of the search direction (Schur-complement form):

$$\begin{aligned} \mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^T\Delta\mathbf{y} &= \mathbf{g} - \mathbf{A}\mathbf{S}^{-1}(\mathbf{h} - \mathbf{X}\mathbf{f}), \\ \Delta\mathbf{s} &= \mathbf{f} - \mathbf{A}^T\Delta\mathbf{y}, \\ \Delta\mathbf{x} &= \mathbf{S}^{-1}(\mathbf{h} - \mathbf{X}\Delta\mathbf{s}) \end{aligned}$$

This computation involves solving a linear system involving the matrix

$$\mathbf{M} \equiv \mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^T = \sum_{j=1}^{n} \frac{x_j}{s_j}\mathbf{a}_j\mathbf{a}_j^T,$$

but all terms are **not** created equal.

Structure in M

$$\mathbf{M} \equiv \mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^T = \sum_{j=1}^{n} \frac{x_j}{s_j}\mathbf{a}_j\mathbf{a}_j^T,$$

Considerations:

- As we converge to a vertex, at least m of the slack variables $s_j \to 0$, and these terms dominate.
- It is well-known that an approximate Newton direction is sufficient for IPM convergence, so we can drop the small terms (carefully).
- Related to column generation algorithms for simplex method, but no "backtracking" or "minor iterations".
- Related to IPM of Dantzig and Ye, Tone, ..., but no backtracking needed.

Terms to include

We need a solid rule for which terms to include so that we can guarantee convergence globally, and guarantee a quadratic rate of convergence locally.

- Include all terms for which the slack variable $s_j < \epsilon$, or at least k of these, where k is a bound on the max number constraints active at a vertex.
- Include enough terms to make $\mathbf{M}$ full rank.

Linear Programming: Results

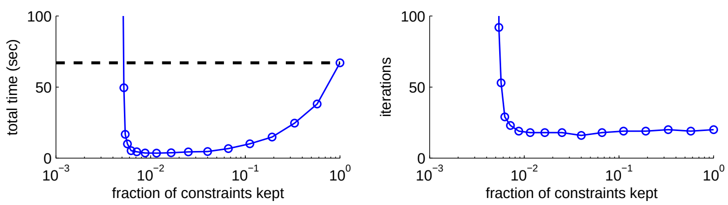
- Assume:
  - $\mathbf{A}$ has full row rank.
  - The dual strictly-feasible set is nonempty.
  - Constraint qualification.

  Then from every feasible starting point ($\mathbf{x}^0 > \mathbf{0}$, $\mathbf{s}^0 > \mathbf{0}$) if the sequence $\{\mathbf{y}^k\}$ for our reduced version of the primal-dual Mehrotra Predictor-Corrector IPM is bounded, it converges to a solution.

- Assume:
  - $\mathbf{A}$ has full row rank.
  - The dual solution set is a singleton.
  - Constraint qualification.

  Then the sequence $(\mathbf{x}^k, \mathbf{s}^k, \mathbf{y}^k)$ converges locally q-quadratically.

- Numerical experiments on random problems, common test problems, Chebyshev approximation problems, filter design problems, etc.
  - Helpful to add a random or meshed set of terms to the $\mathbf{M}$ approximation.
  - Typical behavior (random problem) 200×40000:



Convex Quadratic Programming

As in linear programming, many constrained optimization problems have a great number of (dual) constraints.

For example, in our data fitting problem above

$$\min_{\mathbf{u},t} t$$

$$-t\mathbf{e} \leq \mathbf{H}\mathbf{u} - \mathbf{g} \leq t\mathbf{e},$$

the solution might be very sensitive to noise in the measurements. In that case, we might need to regularize the solution by adding a term to the objective function:

$$\min_{\mathbf{u},t} t + \frac{1}{2}\alpha\|\mathbf{u}\|_2^2$$

subject to

$$-t\mathbf{e} \leq \mathbf{Hu} - \mathbf{g} \leq t\mathbf{e},$$

where α is a small positive number.

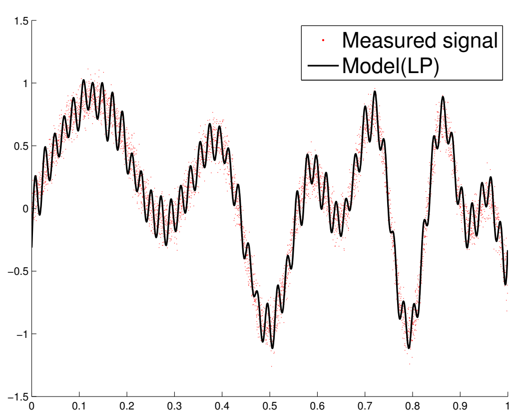
This results in a convex quadrative programming problem.

Convex Quadratic Programming Results

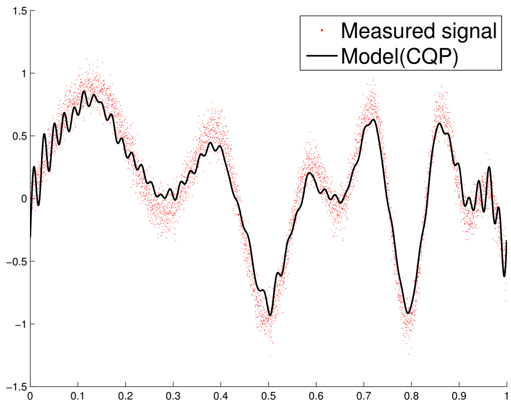
- The approximation to $\mathbf{M}$ now contains the Hessian matrix for the objective function plus the important terms in the summation.
- Proved global convergence and local q-quadratic convergence for our reduced version of the primal-dual affine-scaling IPM under assumptions similar to the LP case.
- Used adaptive choice of important terms – more included at early iterations and fewer in later ones.
- Used constraint relaxation to cope with infeasible starting points, with similar convergence properties.

Application: Semi-infinite QP: Array Pattern Synthesis

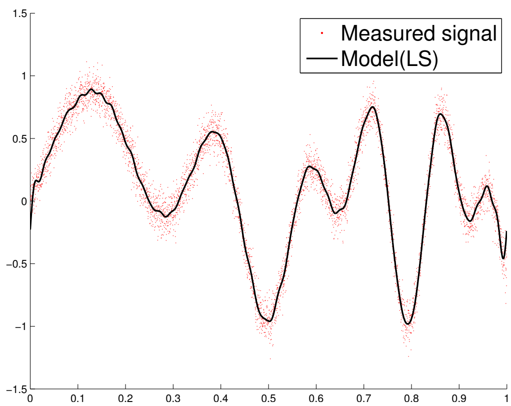
For m sensors distributed around a circle, minimize the sidelobe energy of the array response by adjusting the weights for the sensors. We can develop approximations to this problem by linear programming, the regularization method (qp) above, or least squares (qp).



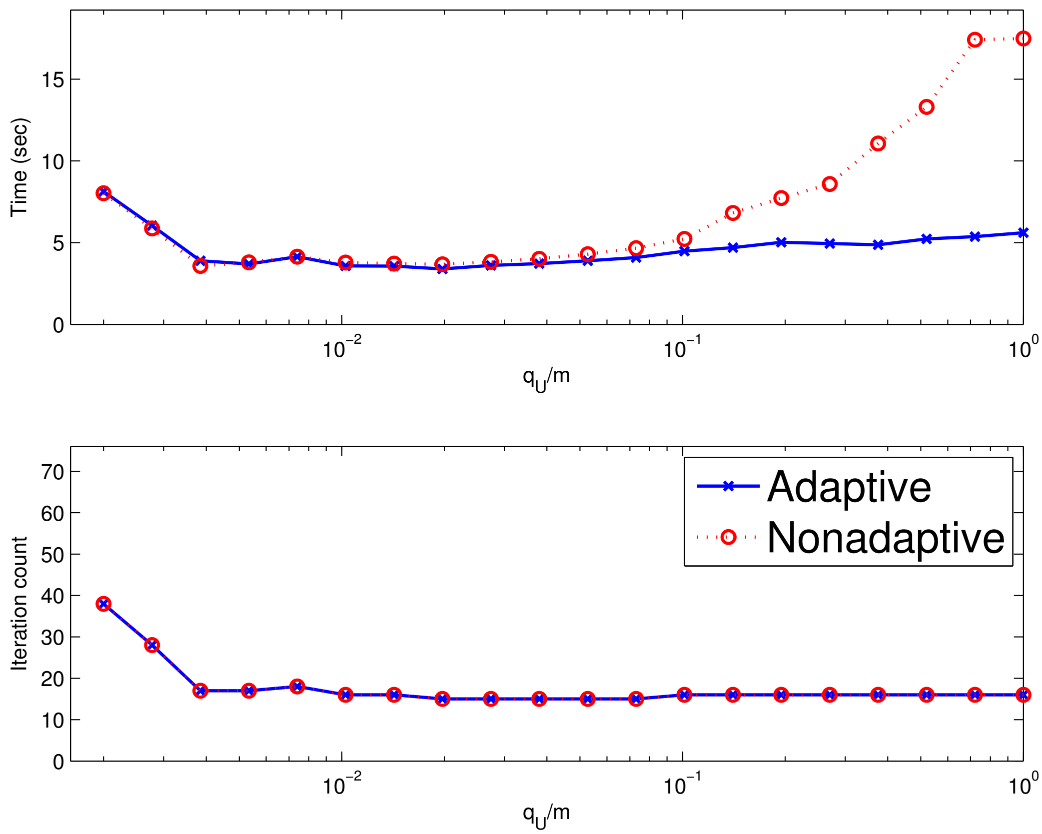
LP solution.



Regularized QP solution.



Least squares solution.



Conclusions and Future Work

- Constraint reduction is a powerful technique when used in state-of-the-art IPM algorithms.
- Global convergence and a local quadratic rate of convergence has been established for linear and quadratic optimization problems.
- Extension to semi-definite programming has begun (October 2009).
- The inexact barrier evaluation framework of Schurr, O'Leary, and Tits gives a framework for approaching the proof of polynomial time complexity.

Some References

Luke B. Winternitz, Stacey O. Nicholls, André L. Tits, Dianne P. O'Leary, "A Constraint-Reduced Variant of Mehrotras Predictor-Corrector Algorithm," submitted.

Jin Hyuk Jung, Dianne P. O'Leary, and André L. Tits, "Adaptive Constraint Reduction for Training Support Vector Machines," Electron. Trans. on Numer. Analysis, 31, 2008, 156-177.

Jin Hyuk Jung, Dianne P. O'Leary, and André L. Tits, "Adaptive Constraint Reduction for Convex Quadratic Programming," *Computational Optimization and Applics.*, March 2010

Simon P. Schurr, Dianne P. O'Leary, and André L. Tits, "A Polynomial-Time Interior-Point Method for Conic Optimization, with Inexact Barrier Evaluations," SIAM Journal on Optimization, 20:1, 2009, 548-571.