Constraint Reduction for Linear and Convex Optimization

Meiyun He, Jin Jung, Paul Laiu, Sungwoo Park, Luke Winternitz, P.-A. Absil, Dianne O'Leary, André Tits

Presenter: André Tits

University of Maryland, College Park

March 11, 2014

Outline

1. Constraint Reduction for LP: Basic Ideas
2. Constraint Reduction for LP: An Aggressive Approach
   - Selection of Working (Q) Set, and Convergence Properties
   - Addressing “Rank Degeneracy”
   - Allowing Infeasible Starting Points
   - Extension to Convex Quadratic Optimization (CQP)
   - Numerical Results and Applications
3. Constraint Reduction for SDP: A More Robust, Polynomial-Time Approach
   - Block-Structured SDP
   - Constraint-Reduction Scheme
   - Special Case: LP
   - Polynomial Convergence
4. Discussion

This talk is an overview of work carried out in our research group over the past few years. For more details, see:

- Tits, Absil, Bill Woessner, “Constraint Reduction for Linear Programs with Many Inequality Constraints”, SIOPT 2006.
- Jung, O’Leary, Tits, “Adaptive Constraint Reduction for Training Support Vector Machines”, ETNA 2008.
- Jung, O’Leary, Tits, “Adaptive Constraint Reduction for Convex Quadratic Programming”, COAP 2012.
- Winternitz, Stacey Nicholls, Tits, O’Leary, “A Constraint-Reduced Variant of Mehrotra’s Predictor-Corrector Algorithm”, COAP 2012.
- He, Tits, “Infeasible Constraint-Reduced Interior-Point Methods for Linear Optimization”, GOMS 2012.
- Winternitz, Tits, Absil, “Addressing rank degeneracy in constraint-reduced interior-point methods for linear optimization”, JOTA, 2014.
- Park, O’Leary “A Polynomial Time Constraint Reduced Algorithm for Semidefinite Optimization Problems”, submitted for publication, 2013.

Outline

Background: Primal-Dual Interior Point (PDIP) Methods

- Consider the standard-form primal and dual linear program (LP)

$$\text{(P)} \quad \begin{array}{c} \min c^{\mathrm{T}} x \\ \text{s.t. } Ax = b \\ x \geq 0 \end{array} \quad \Bigg| \quad \text{(D)} \quad \begin{array}{c} \max b^{\mathrm{T}} y \\ \text{s.t. } A^{\mathrm{T}} y \leq c \\ (\text{or s.t. } A^{\mathrm{T}} y + s = c,\ s \geq 0) \end{array}$$

where $A \in \mathbb{R}^{m \times n}$.

- PDIP search direction: Newton direction for perturbed version of the equalities in the Karush-Kuhn-Tucker (KKT) conditions.

$$\left.\begin{array}{r} A^{\mathrm{T}} y + s = c, \\ Ax = b, \\ Xs = \tau e, \end{array}\right\} \quad \xrightarrow{\text{Newton}} \quad \begin{bmatrix} 0 & A^{\mathrm{T}} & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} c - A^{\mathrm{T}} y - s \\ b - Ax \\ \sigma \mu e - Xs \end{bmatrix},$$
$$(x, s) \geq 0.$$

where $X := diag(x) > 0$, $S := diag(s) > 0$, $\tau = \sigma\mu$, $\mu = x^{\mathrm{T}} s / n > 0$, $\sigma \in [0, 1]$.

Background: Cost of PDIP iteration

- Commonly, the Newton-KKT system is reduced (by block gaussian elimination) to the symmetric indefinite "augmented" system

$$\begin{bmatrix} -X^{-1}S & A^{\mathrm{T}} \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} \star \\ \star \end{bmatrix},$$

an $(n+m) \times (n+m)$ linear system; or, further reduced to the positive definite "normal equations"

$$M\Delta y = [\star], \text{ where } M := AS^{-1}XA^{\mathrm{T}}.$$

- The dominant cost is that of forming the "normal matrix"

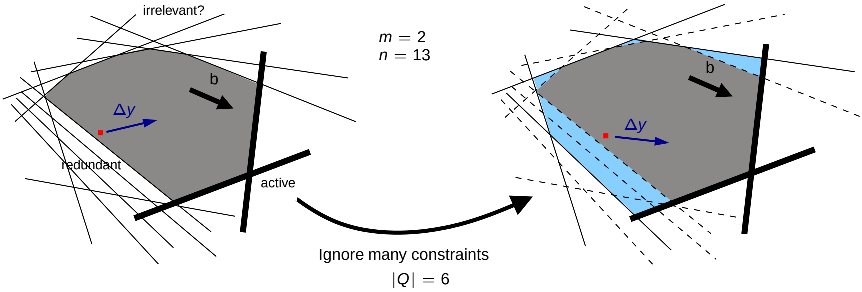
$$M = AS^{-1}XA^{\mathrm{T}} = \sum_{i=1}^{n} \frac{x_i}{s_i} a_i a_i^{\mathrm{T}}.$$

When A is dense, the work per iteration is approximately

$$\boxed{nm^2 \text{ flops}}.$$

Constraint Reduction for LP: Basic Intuition

We expect many constraints are redundant or somehow not very relevant. We could try to guess, at each iteration, a good set Q to "pay attention to" and ignore the rest.



$$\max b^{\mathrm{T}} y$$
$$\text{s.t. } A^{\mathrm{T}} y \leq c$$

$$\max\ b^{\mathrm{T}} y$$
$$\text{s.t. } A_Q^{\mathrm{T}} y \leq c_Q$$
$$(A_Q := [a_{i_1}, a_{i_2}, \cdots],\ i_j \in Q)$$

- Some prior work in 1990's, Dantzig and Ye [1991], Tone [1993], Den Hertog et al. [1994], for basic classes of dual interior-point algorithms.
- Our work focuses on primal-dual interior-point methods (PDIP).

Constraint Reduction: Basic Scheme

- Given a small set Q of constraints deemed critical at the current iteration, compute a PDIP search direction for

$$\begin{array}{c} \min c_Q^{\mathrm{T}} x_Q \\ s.t.\ A_Q x_Q = b \\ x_Q \geq 0 \end{array} \qquad\qquad \begin{array}{c} \max b^{\mathrm{T}} y \\ s.t.\ A_Q^{\mathrm{T}} y \leq c_Q \end{array}$$

i.e., solve

$$\begin{bmatrix} 0 & A_Q^{\mathrm{T}} & I \\ A_Q & 0 & 0 \\ S_Q & 0 & X_Q \end{bmatrix} \begin{bmatrix} \Delta x_Q \\ \Delta y \\ \Delta s_Q \end{bmatrix} = \begin{bmatrix} * \\ * \\ * \end{bmatrix}.$$

This system can be reduced (by block Gaussian elimination) to the "normal equations"

$$M^{(Q)} \Delta y = [*], \text{ where } M^{(Q)} := A_Q S_Q^{-1} X_Q A_Q^{\mathrm{T}}.$$

The dominant cost is that of forming the reduced "normal matrix"

$$M^{(Q)} = A_Q S_Q^{-1} X_Q A_Q^{\mathrm{T}} := \sum_{i \in Q} \frac{x_i}{s_i} a_i a_i^{\mathrm{T}}.$$

When A is dense, the cost is reduced from nm^2 to $\boxed{|Q| m^2}$ flops.

Outline

Aggressive Approach: Selection of Working (Q) Set

[Given a dual-feasible initial point, a dual-feasible sequence is generated.]

- Key requirements for working set Q_k at iteration k:
  - A_Q must have full row rank, in order for Δy to be well defined.
  - IF the sequence $\{y^k\}$ converges to some limit y', THEN, for k large enough, all constraints that are active at y' must be contained in Q.
- Sufficient rule to satisfy these requirements:
  Let M be an upper bound to the number of active constraints at any feasible y, and let $\epsilon > 0$. Among the M smallest slacks s_i^k, include all those with $s_i^k < \epsilon$, subject to A_Q full row rank.
- Possibly augment Q with heuristics addressing the class of problems or application under consideration.
- Reduced "normal" matrix $M^{(Q)}$ need not be close to unreduced matrix M.
- (Ongoing investigation: sort the constraints by s_i^k / s_i^{k-1} instead of s_i^k.)

Aggressive Approach: Convergence Properties

If

- Problem is primal-dual strictly feasible
- A has full row rank

Then y^k converges to y^*, a stationary point.

If, in addition,

- A linear-independence condition holds [Conjecture: This condition is not needed]

Then y^k converges to y^*, a dual solution.

If further

- The dual solution set is a singleton

Then (x^k, y^k) converges q-quadratically to the unique PD solution.

Aggressive Approach: Addressing “Rank Degeneracy”

If A_Q is rank deficient, it means the reduced primal-dual problem

$$\begin{array}{c} \min c_Q^{\mathrm{T}} x_Q \\ s.t.\ A_Q x_Q = b \\ x_Q \geq 0 \end{array} \qquad \begin{array}{c} \max b^{\mathrm{T}} y \\ s.t.\ A_Q^{\mathrm{T}} y \leq c_Q \end{array}$$

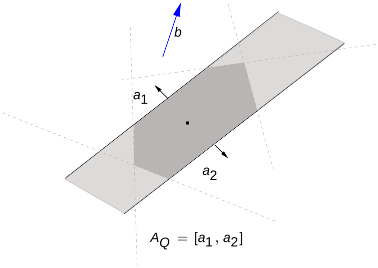
is degenerate, and the reduced PDIP search direction is not well-defined.

Enforcing $\mathrm{rank}(A_Q) = m$ may require significant effort or make $|Q|$ larger than desired:

- Add constraints until the condition holds.

OR

- More systematic linear-algebra methods to ensure a good basis is obtained.



Instead we propose dealing with the degeneracy by the regularization

$$\begin{array}{c} \max\ b^{\mathrm{T}} y - \dfrac{\delta_k}{2} \|y - y_k\|_2^2 \\ s.t.\ A_Q^{\mathrm{T}} y \leq c_Q \end{array}$$

Aggressive Approach: Regularized reduced PDIP

At kth iteration, choose Q and δ_k, and compute PDIP step for the regularized dual (and associated primal)

$$\max b^{\mathrm{T}}y - \frac{\delta_k}{2}\|y - y_k\|^2 \quad \text{s.t. } A_Q^{\mathrm{T}}y \le c$$

$$\min c_Q^{\mathrm{T}}x_Q + \frac{1}{2\delta_k}\|r\|^2 + r^{\mathrm{T}}y_k \quad \text{s.t. } A_Q x_Q + r = b, \quad x_Q \ge 0 \quad \text{with vars } x_Q, r.$$

The regularized "augmented" system is

$$\begin{pmatrix} -X_Q^{-1}S_Q & A_Q^{\mathrm{T}} \\ A_Q & \delta_k I \end{pmatrix} \begin{pmatrix} \Delta x_Q \\ \Delta y \end{pmatrix} = \begin{pmatrix} s \\ b - A_Q x_Q \end{pmatrix},$$

and the regularized "normal-equations" are

$$(A_Q S_Q^{-1} X_Q A_Q^{\mathrm{T}} + \delta_k I)\Delta y = b,$$

Theorem: *Without need for rank$(A_Q) = m$ at each iteration*, a variant of the regularized reduced PDIP method with special choice of δ_k (that has $\delta_k \to 0$ appropriately fast as the solution is approached) converges globally with local quadratic rate.

Aggressive Approach: Regularization in the limit of small δ

Regularized $\Delta y(\delta)$ satisfies

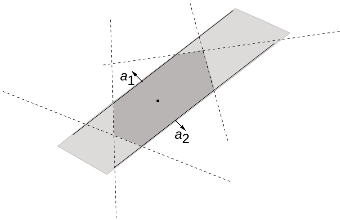
$$(A_Q S_Q^{-1} X_Q A_Q^{\mathrm{T}} + \delta I)\Delta y(\delta) = b$$

What happens as $\delta \to 0$?

Using a spectral decomposition of the normal matrix

$$A_Q S_Q^{-1} X_Q A_Q^{\mathrm{T}} = V \Sigma V^{\mathrm{T}} = \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^{\mathrm{T}} \\ V_2^{\mathrm{T}} \end{pmatrix},$$

with $\mathcal{R}(V_1) = \mathcal{R}(A_Q)$ and $\mathcal{R}(V_2) = \mathcal{N}(A_Q^{\mathrm{T}})$, we get



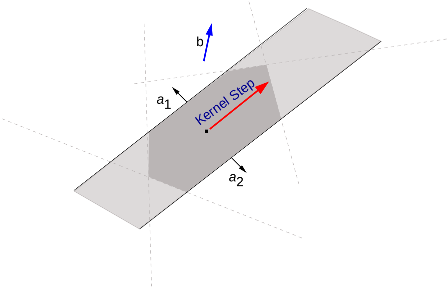
$$\Delta y(\delta) = V_1(\Sigma_1 + \delta I)^{-1} V_1^{\mathrm{T}} b + \delta^{-1} V_2 V_2^{\mathrm{T}} b.$$

- If $V_2^{\mathrm{T}} b = 0$, i.e., $b \in \mathcal{R}(A_Q)$, then $\Delta y(\delta) \to V_1 \Sigma_1^{-1} V_1^{\mathrm{T}} b$, the least norm solution to the normal equations. (E.g., this is so in the non-degenerate case: $\mathrm{rank}(A_Q) = m$.)
- While, if $V_2^{\mathrm{T}} b \neq 0$, then the second term dominates and $\delta \Delta y(\delta) \to V_2 V_2^{\mathrm{T}} b$, the projection of b onto $\mathcal{N}(A_Q^{\mathrm{T}})$ $(= \mathcal{R}(A_Q)^{\perp})$.

Aggressive Approach: Kernel-step constraint-reduced PDIP

- Regular step: If $b \in \mathcal{R}(A_Q)$ then use least norm solution to $A_Q S_Q^{-1} X_Q A_Q^T \Delta y = b$,
- Kernel step: Otherwise take a long step along the projection of b onto $\mathcal{N}(A_Q^T)$.

- We proposed and analyzed an algorithm based on this step within our general constraint-reduced PDIP framework.
- **Theorem:** *Without need for $rank(A_Q) = m$ at each iteration*, a variant of the kernel-step reduced PDIP method converges globally with local quadratic rate. Furthermore, only finitely many kernel-steps are taken.



It turns out that the total number of kernel steps can be related to a suitably defined "degree of degeneracy".

Aggressive Approach: Infeasible Starting Point

- A significant disadvantage: the need for a strictly dual-feasible initial point.
  - Analysis relies crucially on the property $b^{\mathrm{T}}\Delta y > 0$.
- A remedy: introduce an ℓ_1 penalty function:

$$\min_{y,z} \; -b^T y + \rho \sum_i z_i \qquad \text{s.t.} \quad A^T y \leq c + z, \quad z \geq 0$$

  where $\rho > 0$ is the penalty parameter.

- Alternatively, an ℓ_∞ penalty function can be used:

$$\min_{y,z} \; -b^T y + \rho z \qquad \text{s.t.} \quad A^T y \leq c + ze, \quad z \geq 0$$

Aggressive Approach: Exactness of Penalty Function

Let x^* and y^* be the solution of the original primal and dual problems respectively. Let y^*_ρ and z^*_ρ denote the solution of the penalized problem. If

$$\rho > \|x^*\|_\infty,$$

then

$$y^*_\rho = y^*,\, z^*_\rho = 0.$$

- ℓ_1 penalty fcn is exact, i.e., ρ need not go to ∞.
- But x^* is not known *a priori*.

Choice of ρ is challenging:

- If ρ is too large, the cost function $b^{\mathrm{T}}y$ is too strongly deemphasized, resulting in slower convergence to the solution.
- If ρ is too small,
    - the penalized problem is unbounded

  or
    - the solution of the penalized problem is infeasible for the original problem.

Aggressive Approach: Adaptive Adjustment of Penalty Parameter

Begin with ρ relatively small. Let $\sigma > 1$, $\gamma_i > 0$, $i = 1, 2, 3, 4$ be given.

Update: At every iteration of the optimization process, set $\rho^+ = \sigma \rho$ when

EITHER ($\{z_k\}$ seems to be unbounded)

$$\|z\|_\infty \geq \gamma_1 \rho$$

OR (sequence seems to converge to an infeasible KKT point)

$$\|[\Delta y; \Delta z]\| \leq \frac{\gamma_2}{\rho} \ \text{ AND } \ \tilde{x}_Q \geq -\gamma_3 e \ \text{ AND } \ \tilde{u}_Q \not\geq \gamma_4 e$$

where $\tilde{x} = x + \Delta x$, $\tilde{u} = u + \Delta u$ and where u is the primal variables (i.e., KKT multiplier) associated to "$z \geq 0$".

Theorem: Under mild assumptions it is guaranteed that ρ is increased at most finitely many times, and that the iterates converge quadratically to the solution.

Aggressive Approach: Extension to Convex Quadratic Programming (CQP)

- Problem:
$$\max \; b^{\mathrm{T}}y - \frac{1}{2}y^{\mathrm{T}}Hy \quad \text{s.t.} \quad A^{\mathrm{T}}y \leq c.$$
where $H \in \mathbb{R}^{m\times m}$, $H^{\mathrm{T}} = H \succeq 0$, with $[H, A]$ full row rank.
- PDIP iteration extends readily.
- Q-selection rule also extends. However, the number of constraints active at the solution may be significantly smaller than the number m of variables.
- The ℓ_1 (or ℓ_∞) penalization scheme readily extends.

Aggressive Approach: Numerical Results: Randomly Generated Problems

Parameters and initial conditions

- Parameters in the penalty adjustment scheme: $\sigma = 10$, $\gamma_1 = 10$, $\gamma_2 = 1$, $\gamma_3 = \gamma_4 = 100$.
- Typical infeasible initial points x_0, y_0, s_0 generated as in MPC algorithm [Mehrotra, 1992];
- Other initial values: $z_0 = A^T y_0 - c + s_0$, $u_0^i = (x_0^T s_0)/z_0^i$, for $i = 1, \cdots, n$, and $\rho_0 = \|x_0 + u_0\|_\infty$.

Aggressive Approach: Numerical Results: Randomly Generated Problems

- $A \sim \mathcal{N}(0,1)$; $b \sim \mathcal{N}(0,1)$; $c := A^T\bar{y} + \bar{s}$, with $\bar{y} \sim \mathcal{N}(0,1)$ and $\bar{s} \sim \mathcal{U}(0,1)$.
- $m = 100$ and $n = 20000$.

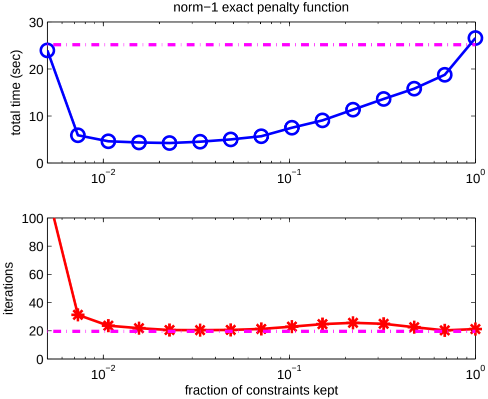


Figure: CPU time and iterations with the ℓ_1 exact penalty function

Aggressive Approach: Some Successful Applications

- LP: Digital Filter Design for GPS Application (NASA)
- CQP: L_2 Entropy-Based Moment Closure
- CQP: Support-Vector Machine
- CQP: Model-Predictive Control

Aggressive Approach: Digital Filter Design (NASA)

N-"tap" FIR filter frequency response

$$H(e^{j\omega}) = \sum_{k=0}^{N-1} h_k e^{-j\omega k}, \omega \in [-\pi, \pi].$$

Chebyshev approximation with side constraints gives optimality criterion that matches a natural approach to filter specification.

$$\min \ t$$

$$\text{s.t. } W(\omega)\left|H(e^{j\omega}) - H_{\mathrm{d}}(e^{j\omega})\right| \le t, \forall\ \omega \in \Omega_{\text{approx}}$$

$$\alpha(\omega) \le \left|H(e^{j\omega})\right| \le \beta(\omega), \forall\ \omega \in \Omega_{\text{side}}$$

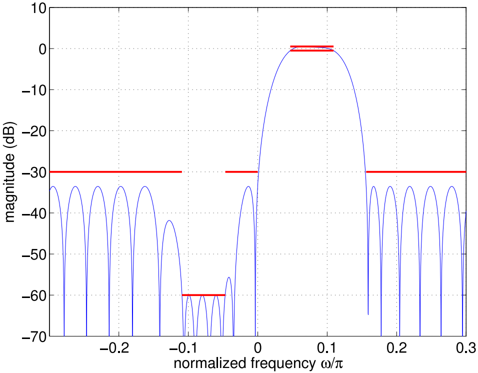
This is not an LP (since $H(e^{j\omega})$ is complex), but it can be rewritten as one:

- Impose linear phase symmetry constraints
- Design the filter "power-spectrum", then perform spectral factorization
- Introduce auxiliary semi-infinite variable

We proposed an effective constraint selection rule for this problem class:

- $M \ge m$ most active, plus
- All grid points on a coarse $O(m)$ discretization grid, plus
- All local minimizers of "slack function" (local maximizers of error).

Aggressive Approach: Linear Phase FIR Filter Design



Under (Type II) linear phase symmetry constraints

$$H(e^{j\omega}) = A(e^{j\omega})e^{j\omega\eta}$$

$$A(e^{j\omega}) = \sum_{k=0}^{\frac{N}{2}-1} \alpha_k 2\cos\omega(k-\tau) + \beta_k 2\sin\omega(k-\tau),$$

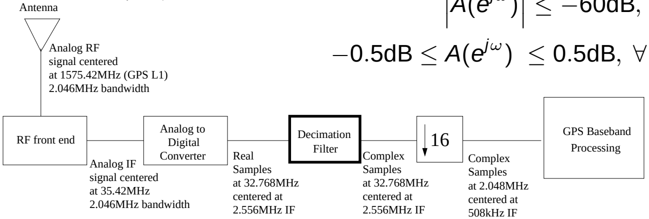
with $h_k = \alpha_k + j\beta_k$;

$$\min \ t$$

$$\text{s.t.} \ \left|A(e^{j\omega}) - 0\right| \leq t, \ \forall\, \omega \in \Omega_{\text{stop}}$$

$$\left|A(e^{j\omega})\right| \leq -60\text{dB}, \ \forall\, \omega \in \Omega_{\text{image}},$$

$$-0.5\text{dB} \leq A(e^{j\omega}) \ \leq 0.5\text{dB}, \ \forall\, \omega \in \Omega_{\text{pass}},$$



Aggressive Approach: Numerical Results on Filter Design

rps-pp=revised primal simplex with partial pricing.
(3m random columns priced, avoids $O(mn)$ work)
mpc=unreduced unregularized Mehrotra predictor-corrector
rmpc=reduced regularized Mehrotra predictor-corrector using special constraint-selection rule

prob	alg	status	time	iter	max $\lvert Q_k \rvert$	mean $\lvert Q_k \rvert$
Linear phase FIR	rps-pp	succ	2.61	629	65	65.0
	mpc	succ	19.91	25	39372	39372.0
	rmpc	succ	5.48	40	1985	1947.3
Phase Noise Filter	rps-pp	fail	Inf	Inf	252	252.0
	mpc	succ	696.47	33	163840	163840.0
	rmpc	succ	112.57	63	7907	7772.4
Linear Predictor	rps-pp	succ	10.14	2672	26	26.0
	mpc	succ	35.72	31	105050	105050.0
	rmpc	succ	12.50	49	1963	1704.5
Antenna Array	rps-pp	succ	130.93	8042	99	99.0
	mpc	succ	299.61	32	272250	272250.0
	rmpc	succ	42.68	35	9769	8598.3

Additional tests showed that our methods typically outperform prior constraint-reduced interior-point algorithms.

Aggressive Approach: CQP: Application to L_2 Entropy-Based Moment Closure

Nonnegative-L_2-entropy-based moment closure constructs an “ansatz” of the underlying distribution given a finite set of moments, by solving

$$\text{minimize} \int f(\mu)^2 d\mu \quad \text{s.t. } f \geq 0 \text{ and} \int m(\mu) f(\mu) d\mu = u,$$

where f is a trial distribution, u is a vector of known moments, and m is a vector of polynomials that define the moments.

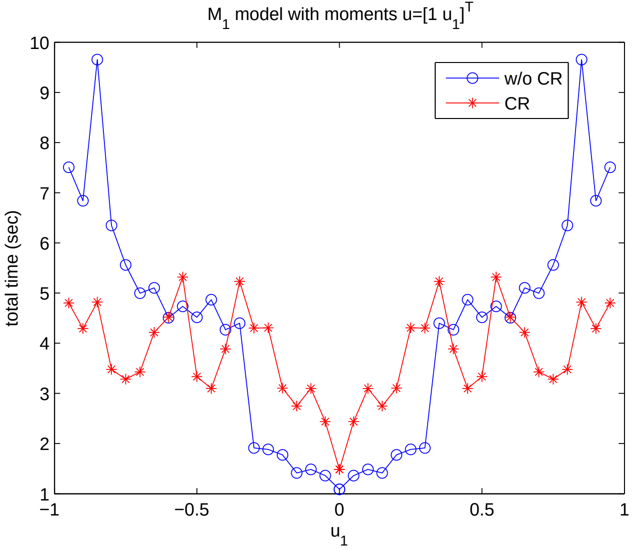
The dual problem can be expressed as

$$\text{minimize } \frac{1}{2} \int \varphi(\mu)^2 d\mu - u^T \alpha \quad \text{s.t. } \alpha^T m(\mu) \leq \varphi(\mu) \text{ for all } \mu,$$

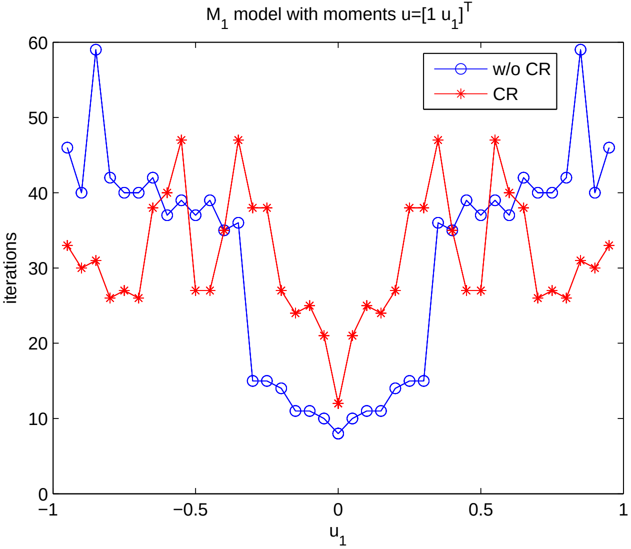
where minimization is with respect to vector α and scalar function φ.

After fine discretization this yields a CQP with many inequality constraints, for which, on “hard” problems, only a small percentage of the constraints are active at the solution: a clear candidate for constraint reduction.

Aggressive Approach for CQP: L_2 Entropy-Based Moment Closure: Preliminary Results (Total Time)



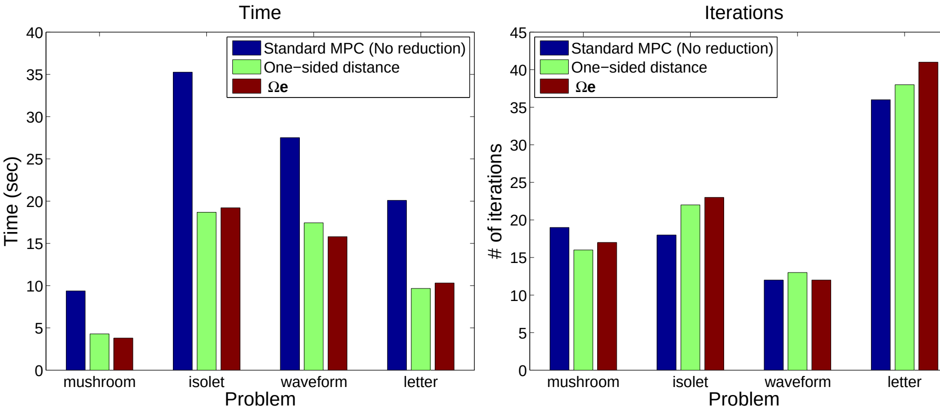
L_2 Entropy-Based Moment Closure: Preliminary Results (Iteration Count)



Aggressive Approach for CQP: Support-Vector Machine

Problems from [Gertz-Griffin, 2006]:

- “Mushroom”: space dimension = 276, # of patterns = 8124
- “Isolet”: space dimension = 617, # of patterns = 7797
- “Waveform”: space dimension = 861, # of patterns = 5000
- “Letter”: space dimension = 153, # of patterns = 20,000



Outline

SDP in Standard Form

$$\text{Primal SDP: } \min_{\boldsymbol{X}} \boldsymbol{C} \bullet \boldsymbol{X} \quad \text{s.t. } \boldsymbol{A}_i \bullet \boldsymbol{X} = b_i \text{ for } i = 1, \ldots, m, \quad \boldsymbol{X} \succeq \boldsymbol{0},$$

$$\text{Dual SDP: } \max_{\boldsymbol{y}, \boldsymbol{S}} \boldsymbol{b}^T \boldsymbol{y} \quad \text{s.t. } \sum_{i=1}^{m} y_i \boldsymbol{A}_i + \boldsymbol{S} = \boldsymbol{C}, \quad \boldsymbol{S} \succeq \boldsymbol{0},$$

where $\boldsymbol{C} \in \mathcal{S}^n$, $\boldsymbol{A}_i \in \mathcal{S}^n$, $\boldsymbol{X} \in \mathcal{S}^n$, and $\boldsymbol{S} \in \mathcal{S}^n$.

Conditions of Optimality:

$$\begin{aligned} \boldsymbol{A}_i \bullet \boldsymbol{X} = \boldsymbol{b} \quad & \text{for } i = 1, \ldots, m, \\ \sum_{i=1}^{m} y_i \boldsymbol{A}_i + \boldsymbol{S} &= \boldsymbol{C}, \\ \boldsymbol{X}\boldsymbol{S} = 0, \quad \boldsymbol{X} \succeq \boldsymbol{0}, & \quad \boldsymbol{S} \succeq \boldsymbol{0}. \end{aligned}$$

[Note: whenever $\boldsymbol{X} \succeq \boldsymbol{0}$ and $\boldsymbol{S} \succeq \boldsymbol{0}$, $\quad \boldsymbol{X}\boldsymbol{S} = 0 \quad \text{iff} \quad \boldsymbol{X} \bullet \boldsymbol{S} = 0$.]

Normal System for PDIP (Newton) Direction

$$\boldsymbol{M}\Delta\boldsymbol{y} = \boldsymbol{g},$$
$$\Delta\boldsymbol{s} = \boldsymbol{r}_d - \mathcal{A}^T\Delta\boldsymbol{y},$$
$$\Delta\boldsymbol{x} = (\boldsymbol{X}\otimes\boldsymbol{S}^{-1})(\mathcal{A}^T\Delta\boldsymbol{y} - \boldsymbol{r}_d) + (\boldsymbol{I}\otimes\boldsymbol{S}^{-1})\boldsymbol{r}_c$$

where

$$\mathcal{A} = [\,\mathrm{vec}\,(\boldsymbol{A}_i)\,, \ldots,\ \mathrm{vec}\,(\boldsymbol{A}_m)]^T,$$
$$\boldsymbol{M} = \mathcal{A}(\boldsymbol{X}\otimes\boldsymbol{S}^{-1})\mathcal{A}^T,$$
$$\boldsymbol{g} = \boldsymbol{r}_p + \mathcal{A}(\boldsymbol{X}\otimes\boldsymbol{S}^{-1})\boldsymbol{r}_d - \mathcal{A}(\boldsymbol{I}\otimes\boldsymbol{S}^{-1})\boldsymbol{r}_c.$$

with

$$r_{pi} = b_i - \boldsymbol{A}_i \bullet \boldsymbol{X} \quad \text{for } i = 1, \ldots, m,$$
$$\boldsymbol{r}_d = \mathrm{vec}\left(\boldsymbol{C} - \boldsymbol{S} - \sum_{i=1}^{m} y_i\boldsymbol{A}_i\right),$$
$$\boldsymbol{r}_c = \mathrm{vec}\,(\overline{\mu}\boldsymbol{I} - \boldsymbol{X}\boldsymbol{S}),$$

Block-Structured SDP

In many applications, $\boldsymbol{A}_i$ and $\boldsymbol{C}$ are block-diagonal,

$$\boldsymbol{A}_i = \begin{bmatrix} \boldsymbol{A}_{i1} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{A}_{ip} \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_1 & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{C}_p \end{bmatrix},$$

yielding

$$\boldsymbol{M} = \mathcal{A}(\boldsymbol{X} \otimes \boldsymbol{S}^{-1})\mathcal{A}^T = \sum_{j=1}^{p} \mathcal{A}_j(\boldsymbol{X}_j \otimes \boldsymbol{S}_j^{-1})\mathcal{A}_j^T,$$

where $\mathcal{A}_j = [\,\mathrm{vec}\,(\boldsymbol{A}_{1j}), \ldots, \,\mathrm{vec}\,(\boldsymbol{A}_{mj})]^T$.

More Robust Approach: Constraint-Reduction Scheme

Replace $\boldsymbol{M}$ with

$$\widehat{\boldsymbol{M}}(Q) = \sum_{j \in Q} \mathcal{A}_j (\boldsymbol{X}_j \otimes \boldsymbol{S}_j^{-1}) \mathcal{A}_j^T,$$

where Q is a "small" subset of $\{1, \ldots, p\}$ such that, for prescribed $q \in (0, 1)$,

$$\left\| \boldsymbol{X}_{Q^c} \left(\sum_{i=1}^m \Delta y_i \boldsymbol{A}_{i,Q^c} \right) \right\|_F \leq q \left\| \begin{bmatrix} \boldsymbol{X}_Q \left(\sum_{i=1}^m \Delta y_i \boldsymbol{A}_{i,Q} \right) & 0 \\ 0 & 0 \end{bmatrix} + \boldsymbol{X}\boldsymbol{R}_d + \boldsymbol{R}_c \right\|_F$$

where

$$\boldsymbol{R}_d = \boldsymbol{C} - \boldsymbol{S} - \sum_{i=1}^m y_i \boldsymbol{A}_i \quad (= \mathrm{mat}(\mathbf{r}_c))$$

$$\boldsymbol{R}_c = \bar{\mu} \boldsymbol{I} - \boldsymbol{X}\boldsymbol{S} \quad (= \mathrm{mat}(\mathbf{r}_d))$$

Important: The chosen value of q is linked to the step size rule. The price to be paid for more aggressive constraint reduction (q closer to 1) is a shorter step.

More Robust Approach: Special Case: LP

When the $\boldsymbol{A}_i$'s and $\boldsymbol{C}$ are *scalar*-diagonal, the SDP becomes our LP in standard form, with the following constraint reduction rule:

$$M^{(Q)} = A_Q S_Q^{-1} X_Q A_Q^T,$$

where $Q \in \{1, \ldots, n\}$ must satisfy

$$\|X_{Q^c} A_{Q^c}^T \Delta y\|_2 \leq q \left\| r_c - X r_d + \begin{bmatrix} X_Q A_Q^T \Delta y \\ 0 \end{bmatrix} \right\|_2,$$

where Δy solves

$$M^{(Q)} \Delta y = r_p - A S^{-1} (r_c - X r_d),$$

with

$$r_p := b - Ax, \quad r_d := c - s - A^T y, \quad r_c := \bar{\mu} e - Xs.$$

More Robust Approach: Polynomial Convergence

After adding an appropriate “corrector” direction to the “predictor” (affine-scaling) direction just discussed, and incorporating an appropriate steplength rule (along the resulting direction), an overall algorithm is obtained that was proved to be polynomially convergent. Specifically, let

$$\epsilon_0 = \max\{\boldsymbol{X}^0 \bullet \boldsymbol{S}^0, \|\boldsymbol{r}_p^0\|, \|\boldsymbol{r}_d^0\|\}.$$

Then

$$\max\{\boldsymbol{X}^k \bullet \boldsymbol{S}^k, \|\boldsymbol{r}_p^k\|, \|\boldsymbol{r}_d^k\|\} < \epsilon$$

after a number k of iterations no larger than

$$O(n \ln(\epsilon_0/\epsilon)).$$

This algorithm is an adaptation of an (“unreduced”) scheme due to Potra and Sheng (1998).

Outline

Discussion

- Two approaches to constraint reduction were presented:
  1. A rather aggressive approach, w/ the following properties:
     - dual feasible; infeasible initial points are handled by incorporating an exact penalty function scheme;
     - no guarantee of polynomial time;
     - constraint-reduced search direction potentially remote from (at times better than) the "unreduced" direction;
     - extends to QP, and even to NLP.
  2. A more robust approach, w/ the following properties:
     - targets SDP (which includes CQP, LP,...);
     - no requirement of initial feasibility;
     - polynomial complexity;
     - constraint-reduced search direction close to the "unreduced" direction.
- Promising numerical results were reported with the former. (Numerical implementation of the latter is underway.)