Constraint Reduction for Linear and Convex Optimization


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:

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
Background: Primal-Dual Interior Point (PDIP) Methods

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

\[
\begin{align*}
\text{(P)} & \quad \min c^T x \\
& \quad \text{s.t. } Ax = b \\
& \quad \quad x \geq 0
\end{align*}
\]

\[
\begin{align*}
\text{(D)} & \quad \max b^T y \\
& \quad \text{s.t. } A^T y \leq c \\
& \quad \quad \text{(or s.t. } A^T y + s = c, \ s \geq 0) \\
\end{align*}
\]

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.

\[
\\begin{align*}
A^T y + s &= c, \\
A x &= b, \\
X s &= \tau e, \\
(x, s) &\geq 0.
\end{align*}
\]

\[
\begin{bmatrix} 0 & A^T & I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} c - A^T y - s \\ b - A x \\ \sigma \mu e - X s \end{bmatrix},
\]

where \( X := \text{diag}(x) > 0 \), \( S := \text{diag}(s) > 0 \), \( \tau = \sigma \mu \), \( \mu = x^T s / n > 0 \), \( \sigma \in [0, 1] \).
Background: Cost of PDIP iteration

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

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

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

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

- The dominant cost is that of forming the “normal matrix”

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

When \(A\) is dense, the work per iteration is approximately \(nm^2\) 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.

\[
\begin{align*}
\text{max } & \quad b^T y \\
\text{s.t. } & \quad A^T y \leq c
\end{align*}
\]

\[
\begin{align*}
\text{max } & \quad b^T y \\
\text{s.t. } & \quad A_Q^T y \leq c_Q \\
& (A_Q := [a_{i_1}, a_{i_2}, \cdots], i_j \in Q)
\end{align*}
\]

- 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

\[
\begin{align*}
\min & \quad c_Q^T x_Q \\
\text{s.t.} & \quad A_Q x_Q = b \\
& \quad x_Q \geq 0
\end{align*}
\]

\[
\max b^T y
\]

\[
\text{s.t.} \quad A_Q^T y \leq c_Q
\]

i.e., solve

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

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

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

The dominant cost is that of forming the reduced “normal matrix”

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

When \( A \) is dense, the cost is reduced from \( nm^2 \) to \( |Q|m^2 \) flops.
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
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 $\Delta y$ 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 $s^k_i$, include all those with $s^k_i < \epsilon$, 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 $s^k_i / s^{k-1}_i$ instead of $s^k_i$.)
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

$$\begin{align*}
\min & \quad c_Q^T x_Q \\
\text{s.t.} & \quad A_Q x_Q = b \\
& \quad x_Q \geq 0 \\
\max & \quad b^T y \\
\text{s.t.} & \quad A_Q^T y \leq c_Q
\end{align*}$$

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

Enforcing $\text{rank}(A_Q) = m$ 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

$$\begin{align*}
\max & \quad b^T y - \frac{\delta_k}{2} \| y - y_k \|_2^2 \\
\text{s.t.} & \quad A_Q^T y \leq c_Q
\end{align*}$$
Aggressive Approach: Regularized reduced PDIP

At $k$th iteration, choose $Q$ and $\delta_k$, and compute PDIP step for the regularized dual (and associated primal)

\[
\max b^T y - \frac{\delta_k}{2} \| y - y_k \|^2 \\
\text{s.t. } A_Q^T y \leq c
\]

\[
\min c_Q^T x_Q + \frac{1}{2\delta_k} \| r \|^2 + r^T y_k \\
\text{s.t. } A_Q x_Q + r = b \\
x_Q \geq 0 \\
\text{with vars } x_Q, r.
\]

The regularized “augmented” system is

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

and the regularized “normal-equations” are

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

**Theorem:** *Without need for rank$(A_Q) = m$ at each iteration*, a variant of the regularized reduced PDIP method with special choice of $\delta_k$ (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 $\delta$

Regularized $\Delta y(\delta)$ satisfies

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

What happens as $\delta \to 0$?

Using a spectral decomposition of the normal matrix

$$A_Q S_Q^{-1} X_Q A_Q^T = V \Sigma V^T = (V_1 \quad V_2) \left( \begin{array}{cc} \Sigma_1 & 0 \\ 0 & 0 \end{array} \right) \left( \begin{array}{c} V_1^T \\ V_2^T \end{array} \right),$$

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

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

- If $V_2^T b = 0$, i.e., $b \in \mathcal{R}(A_Q)$, then $\Delta y(\delta) \to V_1 \Sigma_1^{-1} V_1^T b$, the least norm solution to the normal equations. (E.g., this is so in the non-degenerate case: $\text{rank} (A_Q) = m$.)

- While, if $V_2^T b \neq 0$, then the second term dominates and $\delta \Delta y(\delta) \to V_2 V_2^T b$, the projection of $b$ onto $\mathcal{N}(A_Q^T) (= \mathcal{R}(A_Q)^\perp)$.
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 \( \text{rank}(A_Q) = m \) 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”.

\[ \text{He, Jung, Laiu, Park, Winternitz, Absil, O'Leary, Tits (2014)} \]
A significant disadvantage: the need for a strictly dual-feasible initial point.

Analysis relies crucially on the property $b^T \Delta y > 0$.

A remedy: introduce an $\ell_1$ penalty function:

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

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

Alternatively, an $\ell_\infty$ penalty function can be used:

$$\min_{y, z} -b^T y + \rho z \quad \text{s.t.} \quad A^T y \leq c + ze, \quad z \geq 0$$
Aggressive Approach: Exactness of Penalty Function

Let \( x^* \) and \( y^* \) be the solution of the original primal and dual problems respectively. Let \( y_\rho^* \) and \( z_\rho^* \) denote the solution of the penalized problem. If

\[
\rho > \|x^*\|_\infty,
\]

then

\[
y_\rho^* = y^*, \; z_\rho^* = 0.
\]

- \( \ell_1 \) penalty fcn is exact, i.e., \( \rho \) need not go to \( \infty \).
- But \( x^* \) is not known \textit{a priori}.

Choice of \( \rho \) is challenging:
- If \( \rho \) is too large, the cost function \( b^T y \) is too strongly deemphasized, resulting in slower convergence to the solution.
- If \( \rho \) 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 $\rho$ 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)

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

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 $\rho$ 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:
  \[
  \max b^T y - \frac{1}{2} y^T H y \quad \text{s.t.} \quad A^T y \leq c.
  \]
  where \( H \in \mathbb{R}^{m \times m}, H^T = H \succeq 0 \), 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 \( \ell_1 \) (or \( \ell_\infty \)) 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:** \( z_0 = A^T y_0 - c + s_0, \quad u_0^i = (x_0^T s_0)/z_0^i, \) for \( i = 1, \ldots, n, \) and \( \rho_0 = \| x_0 + u_0 \|_{\infty}. \)
Aggressive Approach: Numerical Results: Randomly Generated Problems

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

**Figure:** CPU time and iterations with the $\ell_1$ 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.

$$\min t$$

s.t. $$W(\omega) \left| H(e^{j\omega}) - H_d(e^{j\omega}) \right| \leq t, \forall \omega \in \Omega_{approx}$$

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

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

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

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
Aggressive Approach: Linear Phase FIR Filter Design

Under (Type II) linear phase symmetry constraints

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

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

with \( h_k = \alpha_k + j\beta_k; \)

\[ \min t \]

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

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

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

**rmnpc** = reduced regularized Mehrotra predictor-corrector using special constraint-selection rule

| prob            | alg     | status | time | iter | max $|Q_k|$ | mean $|Q_k|$ |
|-----------------|---------|--------|------|------|---------|---------|
| Linear phase FIR | rps-pp  | succ   | 2.61 | 629  | 65      | 65.0    |
|                 | mpc     | succ   | 19.91| 25   | 39372   | 39372.0 |
|                 | rmnpc   | 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|
|                 | rmnpc   | 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|
|                 | rmnpc   | 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|
|                 | rmnpc   | 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 $\alpha$ and scalar function $\varphi$.

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)

M₁ model with moments $u = [1 \ u_1]^T$

- w/o CR
- CR
$L_2$ Entropy-Based Moment Closure: Preliminary Results (Iteration Count)

M₁ model with moments $u=[1 \ u_1]^T$

- Blue circles: w/o CR
- Red stars: CR

Iterations vs $u_1$ range from -1 to 1.
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

![Graph showing the comparison of Time and Iterations for different problems using Standard MPC (No reduction), One-sided distance, and Ωe methods.](image-url)
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
SDP in Standard Form

Primal SDP: \( \min \mathbf{C} \cdot \mathbf{X} \) s.t. \( \mathbf{A}_i \cdot \mathbf{X} = b_i \) for \( i = 1, \ldots, m \), \( \mathbf{X} \succeq 0 \),

Dual SDP: \( \max \mathbf{b}^T \mathbf{y} \) s.t. \( \sum_{i=1}^{m} y_i \mathbf{A}_i + \mathbf{S} = \mathbf{C} \), \( \mathbf{S} \succeq 0 \),

where \( \mathbf{C} \in S^n \), \( \mathbf{A}_i \in S^n \), \( \mathbf{X} \in S^n \), and \( \mathbf{S} \in S^n \).

Conditions of Optimality:

\[ \mathbf{A}_i \cdot \mathbf{X} = b \text{ for } i = 1, \ldots, m, \]
\[ \sum_{i=1}^{m} y_i \mathbf{A}_i + \mathbf{S} = \mathbf{C}, \]
\[ \mathbf{X} \mathbf{S} = 0, \mathbf{X} \succeq 0, \mathbf{S} \succeq 0. \]

[Note: whenever \( \mathbf{X} \succeq 0 \) and \( \mathbf{S} \succeq 0 \), \( \mathbf{X} \mathbf{S} = 0 \) iff \( \mathbf{X} \cdot \mathbf{S} = 0 \).]
Normal System for PDIP (Newton) Direction

\[ M \Delta y = g, \]
\[ \Delta s = r_d - A^T \Delta y, \]
\[ \Delta x = (X \otimes S^{-1})(A^T \Delta y - r_d) + (I \otimes S^{-1})r_c \]

where

\[ A = [ \text{vec}(A_1), \ldots, \text{vec}(A_m) ]^T, \]
\[ M = A(X \otimes S^{-1})A^T, \]
\[ g = r_p + A(X \otimes S^{-1})r_d - A(I \otimes S^{-1})r_c. \]

with

\[ r_{pi} = b_i - A_i \cdot X \text{ for } i = 1, \ldots, m, \]
\[ r_d = \text{vec} \left( C - S - \sum_{i=1}^{m} y_i A_i \right), \]
\[ r_c = \text{vec}(\mu I - XS), \]
In many applications, $\mathbf{A}_i$ and $\mathbf{C}$ are block-diagonal,

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

yielding

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

where $\mathcal{A}_j = [\text{vec}(\mathbf{A}_{1j}), \ldots, \text{vec}(\mathbf{A}_{mj})]^T$. 
More Robust Approach: Constraint-Reduction Scheme

Replace $M$ with

$$\hat{M}(Q) = \sum_{j \in Q} A_j (X_j \otimes S_j^{-1}) A_j^T,$$

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

$$\left\| X_{Q^c} \left( \sum_{i=1}^{m} \Delta y_i A_i, Q^c \right) \right\|_F \leq q \left\| \begin{bmatrix} X_Q (\sum_{i=1}^{m} \Delta y_i A_i, Q) & 0 \\ 0 & 0 \end{bmatrix} + XR_d + R_c \right\|_F$$

where

$$R_d = C - S - \sum_{i=1}^{m} y_i A_i \quad (= \text{mat}(r_c))$$

$$R_c = \bar{\mu} l - XS \quad (= \text{mat}(r_d))$$

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 $A_i$’s and $C$ 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

$$\|X_Q^c A_Q^{Tc} \Delta y\|_2 \leq q \left\|r_c - Xr_d + \begin{bmatrix} X_Q A_Q^T \Delta y \\ 0 \end{bmatrix} \right\|_2,$$

where $\Delta y$ solves

$$M^{(Q)} \Delta y = r_p - AS^{-1} (r_c - Xr_d),$$

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

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

Then

$$
\max\{X^k \cdot S^k, \|r^k_p\|, \|r^k_d\|\} < \epsilon
$$

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

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
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.)