Row by row methods for semidefinite programming

Donald Goldfarb[1]

[1] Joint work with Zaiwen Wen, Shiqian Ma and Katya Scheinberg
Department of Industrial Engineering and Operations Research
Columbia University

Tapia 70 Conference
May 29, 2009

Semidefinite programming (SDP)

- $X \in S^n$, the space of real symmetric $n \times n$ matrices
- $b \in \mathbb{R}^m$, $C \in S^n$ and $A^{(i)} \in S^n$ are problem parameters
- The inequality $X \succeq 0$ means X is positive semidefinite
- Inner product: $\langle C, X \rangle := \sum_{j=1}^{n} \sum_{k=1}^{n} C_{j,k} X_{j,k}$

Optimization problem

$$\begin{aligned} \min_{X \in S^n} \quad & \langle C, X \rangle \\ \text{s.t.} \quad & \left\langle A^{(i)}, X \right\rangle = b_i, \quad i = 1, \cdots, m, \\ & X \succeq 0 \end{aligned}$$

Overview: cracking positive semidefiniteness $X \succeq 0$

- Solving a sequence of barrier functions
  - $\log\det X^{-1}$ is a self-concordant barrier function
  - Interior point methods (primal, dual, primal-dual): potential reduction algorithms, path-following methods
  - Other penalty and barrier functions
- Maximum eigenvalue function
  - Spectral bundle method
- Eigenvalue decomposition
  - Newton-CG augmented Lagrangian method
  - Boundary point method
- Change of variables $X = RR^\top$
  - Nonlinear programming approaches via low-rank factorization

Expressing $X \succ 0$ by Schur complement

- Assume $X \in S^n$ is partitioned as $\begin{pmatrix} \xi & y^\top \\ y & B \end{pmatrix}$, where $\xi \in \mathbb{R}$, $y \in \mathbb{R}^{n-1}$ and $B \in S^{n-1}$ is nonsingular
- Factorization:

$$X = \begin{pmatrix} 1 & y^\top B^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \xi - y^\top B^{-1} y & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} 1 & 0 \\ B^{-1} y & I \end{pmatrix}$$

Positive definiteness and Schur complement:

$$X \succ 0 \iff B \succ 0 \text{ and } (X/B) := \xi - y^\top B^{-1} y > 0$$

- Cholesky factorization: $B := LL^\top$.
- $\xi - y^\top B^{-1} y > 0 \iff \|L^{-1} y\|^2 \leq \xi$ (second-order cone)

Constructing SOC constraint row by row

- Given $X^k \succ 0$, we can fix the principal submatrix

$$B := X^k_{1^c,1^c} = \begin{pmatrix} X^k_{2,2} & \cdots & X^k_{2,\,n} \\ \cdots & \cdots & \cdots \\ X^k_{n,\,2} & \cdots & X^k_{n,n} \end{pmatrix}$$

and let $\xi := X_{1,1}$ and $y := X_{1^c,1} := (X_{1,2}, \cdots, X_{1,n})^\top$

- The variable X now is $\begin{pmatrix} \xi & y^\top \\ y & B \end{pmatrix} := \begin{pmatrix} \xi & y^\top \\ y & X^k_{1^c,1^c} \end{pmatrix}$
- SOC constraint: $\xi - y^\top B^{-1} y \geq \nu$ for $\nu > 0$
- In general: $\xi := X_{i,i}$, $y := X_{i^c,i}$ and $B := X^k_{i^c,i^c}$

Solving RBR subproblem

SDP		SOCP restriction
$$\begin{aligned} \min_{X \in S^n} \quad & \langle C, X \rangle \\ \text{s.t.} \quad & \mathcal{A}(X) = b, \\ & X \succeq 0, \end{aligned}$$	$\Longrightarrow$	$$\begin{aligned} \min_{[\xi; y] \in \mathbb{R}^n} \quad & \widetilde{c}^\top [\xi; y] \\ \text{s.t.} \quad & \widetilde{A}\, [\xi; y] = \widetilde{b}, \\ & \xi - y^\top B^{-1} y \geq \nu, \end{aligned}$$

where $\nu > 0$ and

$$\widetilde{c} := \begin{pmatrix} C_{i,i} \\ 2C_{i^c,i} \end{pmatrix}, \quad \widetilde{A} := \begin{pmatrix} A^{(1)}_{i,i} & 2A^{(1)}_{i,i^c} \\ \cdots & \cdots \\ A^{(m)}_{i,i} & 2A^{(m)}_{i,i^c} \end{pmatrix} \text{ and } \widetilde{b} := \begin{pmatrix} b_1 - \left\langle A^{(1)}_{i^c,i^c}, B \right\rangle \\ \cdots \\ b_m - \left\langle A^{(m)}_{i^c,i^c}, B \right\rangle \end{pmatrix}$$

Row-by-Row (RBR) algorithm prototype

Algorithm 1: A row-by-row (RBR) method prototype

Set $X^1 \succ 0$, $\nu \geq 0$ and $k := 1$.
while *not converge* **do**
  for $i = 1, \cdots, n$ **do**
    Solve the SOCP subproblem for i-th row/column.
    Update $X^k_{i,i} := \xi$, $X^k_{i^c,i} := y$ and $X^k_{i,i^c} := y^\top$.
  Set $X^{k+1} := X^k$ and $k := k + 1$.

Application: the maxcut SDP relaxation

The RBR subproblem for SDP with only diagonal element constraints:

$$\begin{aligned} \min_{X \in S^n} \quad & \langle C, X \rangle \\ \text{s.t.} \quad & X_{i,i} = 1, \\ & X \succeq 0, \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \min_{y \in \mathbb{R}^{n-1}} \quad & \widehat{c}^\top y \\ \text{s.t.} \quad & 1 - y^\top B^{-1} y \geq \nu \end{aligned}$$

Closed-form solution of the RBR subproblem

If $\gamma := \widehat{c}^\top B \widehat{c} > 0$, the solution of the RBR subproblem is

$$y = -\sqrt{\frac{1-\nu}{\gamma}} B \widehat{c}.$$

Otherwise, the solution is $y = 0$.

Interpretation in terms of log-barrier approach

- Consider the logarithmic barrier problem

$$\begin{aligned} \min_{X \in S^n} \quad & \langle C, X \rangle - \sigma \log \det X \\ \text{s.t.} \quad & X_{ii} = 1, \forall i = 1, \cdots, n, \quad X \succeq 0 \end{aligned}$$

- Key: $\det(X) = \det(B)(1 - y^\top B^{-1} y)$
- The RBR subproblem is:

$$\min_{y \in \mathbb{R}^{n-1}} \quad \widehat{c}^\top y - \sigma \log(1 - y^\top B^{-1} y)$$

whose solution is $y = -\frac{\sqrt{\sigma^2+\gamma}-\sigma}{\gamma} B\widehat{c}$, where $\gamma := \widehat{c}^\top B\widehat{c}$.

- Equal to the pure RBR method if $\nu = 2\sigma \frac{\sqrt{\sigma^2+\gamma}-\sigma}{\gamma}$

Convergence for general function

Consider the RBR method for solving

$$(P) \qquad \begin{array}{ll} \min\limits_{X \in S^n} & f(X) - \sigma \log\det X \\ \text{s.t.} & L \le X \le U, \quad X \succ 0 \end{array}$$

- $f(X)$ is a convex function of X
- $L, U \in S^n$ are constant matrices and $L \le X \le U$ means that $L_{i,j} \le X_{i,j} \le U_{i,j}$ for all $i, j = 1, \cdots, n$

Theorem

Let $\{X^k\}$ be a sequence generated by the row-by-row method for solving (P). Then every limit point of $\{X^k\}$ is a global minimizer of (P).

Failure on a SDP with general linear constraints

Consider the SDP

$$\begin{aligned} \min \quad & X_{11} + X_{22} - \log\det(X) \\ \text{s.t.} \quad & X_{11} + X_{22} \geq 4, \quad X \succeq 0. \end{aligned}$$

- Initial point: $\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$ and optimal solution: $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.
- The RBR subproblems are

$$\begin{aligned} \min \quad & X_{11} - \log(3X_{11} - X_{12}^2), \text{ s.t. } X_{11} \geq 1, \\ \min \quad & X_{22} - \log(X_{22} - X_{12}^2), \quad \text{s.t. } X_{22} \geq 3 \end{aligned}$$

- Optimal solutions of subproblems are, respectively, $X_{11} = 1$, $X_{12} = 0$ and $X_{12} = 0$, $X_{22} = 3$

Row-by-row augmented Lagrangian method

We consider the following augmented Lagrangian approach:

- Given $\mu > 0$, we start from $X^1 \succ 0$ and $b^1 := b$.
- Solve the quadratic penalty function

$$X^k := \arg\min_X \langle C, X\rangle + \frac{1}{2\mu}\|\mathcal{A}(X) - b^k\|_2^2, \text{ s.t. } X \succeq 0,$$

and update $b^{k+1} := b + \frac{\mu}{\mu^k}\left(b^k - \mathcal{A}(X^k)\right)$.

The RBR subproblem

$$\begin{aligned} \min_{(\xi;y)\in\mathbb{R}^n} \quad & \widetilde{c}^\top \begin{pmatrix}\xi \\ y\end{pmatrix} + \frac{1}{2\mu^k}\left\|\widetilde{A}\begin{pmatrix}\xi \\ y\end{pmatrix} - \widetilde{b}\right\|_2^2 \\ \text{s.t.} \quad & \xi - y^\top B^{-1} y \geq \nu. \end{aligned}$$

Convergence Properties

- Solving the RBR subproblem essentially corresponds to minimizing the unconstrained function obtained by subtracting $\sigma \log(\xi - y^\top B^{-1} y)$ from the objective function
- Convergence of the RBR method for minimizing this log-barrier function
- The convergence of our augmented Lagrangian framework follows from the standard theory for the augmented Lagrangian method for minimizing a strictly convex function subject to linear equality constraints; see Bertsekas, Rockafellar and etc.

Application: the maxcut SDP relaxation

Since $\left\| \widetilde{A} \begin{pmatrix} \xi \\ y \end{pmatrix} - \widetilde{b} \right\|^2 = (\xi - b_i^k)^2$, we have:

The RBR subproblem

$$\begin{array}{ll} \min_{X \in S^n} & \langle C, X \rangle \\ \text{s.t.} & X_{i,i} = 1, \\ & X \succeq 0, \end{array} \quad \Longrightarrow \quad \begin{array}{ll} \min_{(\xi; y) \in \mathbb{R}^n} & c\xi + \widehat{c}^\top y + \frac{1}{2\mu^k}(\xi - b_i^k)^2 \\ \text{s.t.} & \xi - y^\top B^{-1} y \geq \nu. \end{array}$$

- If $\widehat{c} \neq 0$, the solution is: $\xi = b_i^k + \mu^k(\lambda - c)$ and $y = -\frac{1}{2\lambda} B\widehat{c}$, where λ is the unique real root of the cubic equation:

$$\varphi(\lambda) := 4\mu^k \lambda^3 + 4(b_i^k - \mu^k c - \nu)\lambda^2 - \gamma = 0.$$

Application: nuclear-norm matrix completion

Given a matrix $M \in \mathbb{R}^{p \times q}$ and an index set

$$\Omega \subseteq \{(i,j) \mid i \in \{1,\cdots,p\}, j \in \{1,\cdots,q\}\},$$

the nuclear norm matrix completion problem is

$$\begin{array}{ll} \min_{W \in \mathbb{R}^{p \times q}} & \|W\|_* \\ s.t. & W_{ij} = M_{ij}, \ \forall\, (i,j) \in \Omega, \end{array}$$

which is equivalent to the SDP problem

$$\begin{array}{ll} \min_{X \in S^n} & Tr(X) \\ s.t. & X := \begin{bmatrix} X^{(1)} & W \\ W^\top & X^{(2)} \end{bmatrix} \succeq 0 \\ & W_{ij} = M_{ij}, \ \forall\, (i,j) \in \Omega. \end{array}$$

Application: nuclear-norm matrix completion

- Partition the vector y into the subvectors $\widehat{y}$ and $\widetilde{y}$, whose elements are in and not in the set Ω, respectively.
- The residual is: $\left\|\widetilde{A}\begin{pmatrix}\xi\\ y\end{pmatrix} - \widetilde{b}\right\| = \|\widehat{y} - \widetilde{b}\|$

The RBR subproblem

$$\begin{aligned} \min_{(\xi;y)\in\mathbb{R}^n} \quad & \xi + \frac{1}{2\mu^k}\left\|\widehat{y} - \widetilde{b}\right\|_2^2, \\ \text{s.t.} \quad & \xi - y^\top B^{-1} y \geq \nu, \end{aligned} \qquad B = \begin{pmatrix} X^k_{\alpha,\alpha} & X^k_{\alpha,\beta} \\ X^k_{\beta,\alpha} & X^k_{\beta,\beta} \end{pmatrix},$$

whose optimal solution is

$$\begin{cases} \xi = \dfrac{1}{2\mu^k}\widehat{y}^\top(\widetilde{b} - \widehat{y}) + \nu, \\ \widehat{y} = \left(2\mu^k I + X^k_{\alpha,\alpha}\right)^{-1} X^k_{\alpha,\alpha}\widetilde{b}, \quad \widetilde{y} = \dfrac{1}{2\mu^k} X^k_{\beta,\alpha}(\widetilde{b} - \widehat{y}). \end{cases}$$

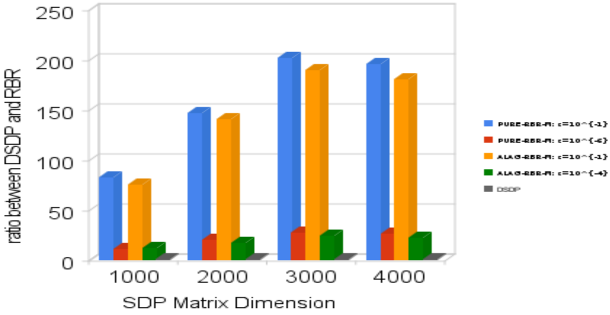
Numerical Results

- The maxcut SDP relaxation
  - The test problems whose size ranging from $n = 1000$ to $n = 4000$ are based on graphs generated by "rudy"
  - Two variants: PURE-RBR-M and ALAG-RBR-M
- The nuclear-norm matrix completion problem
  - Gaussian random matrices M_L and M_R and set $M = M_L M_R^\top$
  - Sample a subset Ω of m entries uniformly at random
  - Sampling ratio (SR): $m/(pq)$
  - Ratio "FR": $r(p+q-r)/m < 1$
- Codes were written in C Language MEX-files in MATLAB (Release 7.3.0) and all experiments were performed on a Dell Precision 670 workstation with an Intel Xeon 3.4GHZ CPU and 6GB of RAM.

Numerical Results: the maxcut SDP relaxation

Table: Average ratio of DSDP CPU time to RBR CPU time

	PURE-RBR-M		ALAG-RBR-M		DSDP
n	$\epsilon = 10^{-3}$	$\epsilon = 10^{-6}$	$\epsilon = 10^{-1}$	$\epsilon = 10^{-4}$	
1000	82.2	11.4	75.9	12.1	1
2000	146.9	20.7	140.4	17.8	1
3000	201.8	27.1	190.3	24.1	1
4000	196.0	26.2	180.2	22.8	1
rel.err in obj	10^{-3}	10^{-5}	10^{-3}	10^{-5}	



Numerical Results: the maxcut SDP relaxation

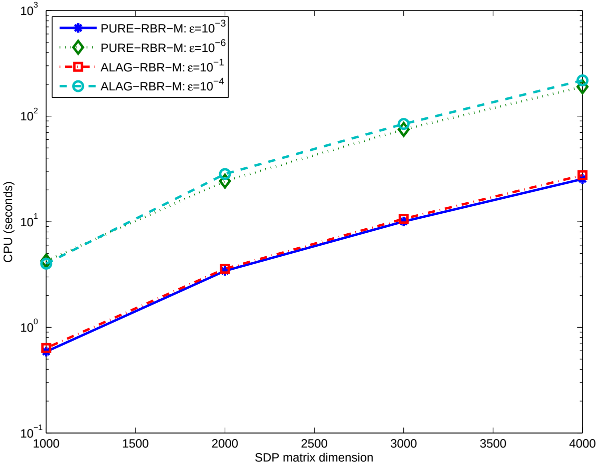


Figure: Relationship between CPU time and SDP matrix dimension for the maxcut SDP relaxation

Numerical Results: the maxcut SDP relaxation

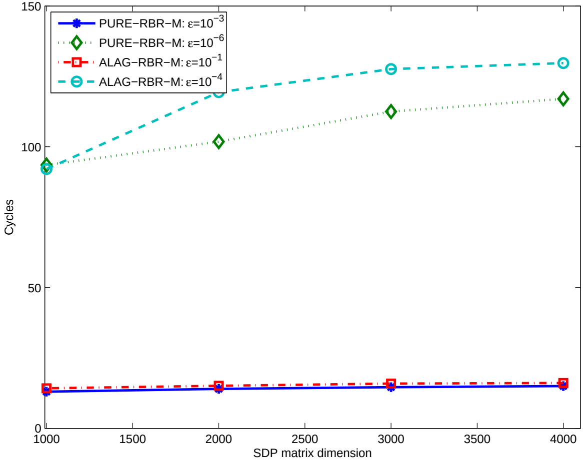


Figure: Relationship between cycles and SDP matrix dimension for the maxcut SDP relaxation

Numerical Results: nuclear-norm matrix completion

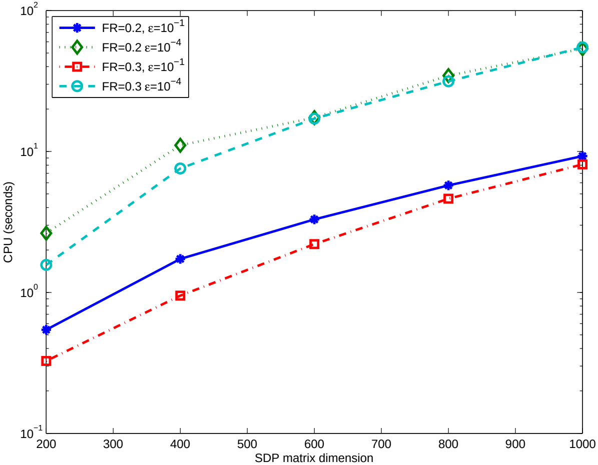


Figure: Relationship between CPU time and SDP matrix dimension for SDP matrix completion

Numerical Results: nuclear-norm matrix completion

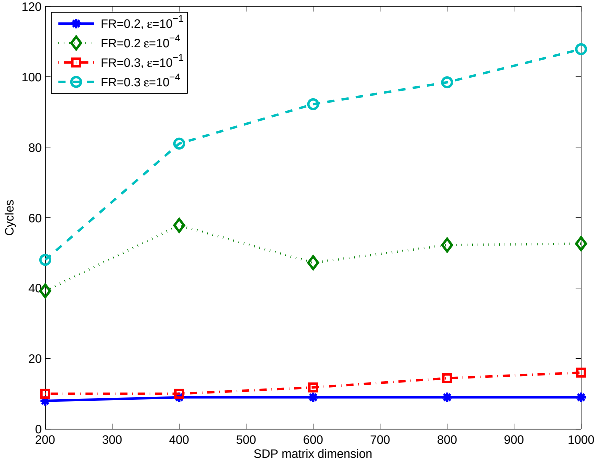


Figure: Relationship between cycles and SDP matrix dimension for SDP matrix completion