Exploiting Sparsity of SDPs (Semidefinite Programs) and Their Applications to POPs (Polynomial Optimization Problems)

Masakazu Kojima, Tokyo Institute of Technology

The 3rd Sino-Japanese Optimization Meeting

November 2, 2005

Singapore

The mian purpose is to show how important exploiting sparsity is in solving SDPs and the applications to POPs.

Outline

1. SDP (semidefinite program) and its dual
2. Primal-dual IPM (Interior-Point Method)
3. Various types of structured sparsities
4. Numerical results: structured sparsities + parallel
5. POPs (Polynomial Optimization Problems)
6. Rough sketch of SDP relaxation of POPs
7. Exploiting structured sparsity
8. Numerical results on POPs
9. Summary and concluding remarks

Sparsity of SSPs is based on joint works with
K. Fujisawa, M. Fukuda, K. Murota and K. Nakata

Sparse SDP relaxation is based on joint works with
S. Kim, M. Muramatsu and H. Waki

Outline

1. SDP (semidefinite program) and its dual
2. Primal-dual IPM
3. Various types of structured sparsities
4. Numerical results: structured sparsities + parallel
5. POPs (Polynomial Optimization Problems)
6. Rough sketch of SDP relaxation of POPs
7. Exploiting structured sparsity
8. Numerical results on POPs
9. Summary and concluding remarks

$$\mathcal{P} : \min \quad A_0 \bullet X \quad \text{sub.to} \quad A_p \bullet X = b_p \ (1 \leq p \leq m), \ \mathcal{S}^n \ni X \succeq O$$
$$\mathcal{D} : \max \quad \sum_{p=1}^m b_p y_p \quad \text{sub.to} \quad \sum_{p=1}^m A_p y_p + S = A_0, \ \mathcal{S}^n \ni S \succeq O$$

$$\begin{aligned}
\mathcal{S}^n \quad &: \quad \text{the set of } n \times n \text{ symmetric matrices} \\
X, \ S \ &\in \ \mathcal{S}^n, \ y_p \in R \ (1 \leq p \leq m) \ : \ \text{variables} \\
A_0, \ A_p \ &\in \ \mathcal{S}^n, \ b_p \in R \ (1 \leq p \leq m) \ : \ \text{given data} \\
U \bullet V \ &= \ \sum_{i=1}^n \sum_{j=1}^n U_{ij} V_{ij} \ \text{ for every } U, \ V \in R^{n \times n} \\
X \succeq O \ &\Leftrightarrow \ X \in \mathcal{S}^n \ \text{ is positive semidefinite}
\end{aligned}$$

Important features — SDP can be large-scale easily

- $n \times n$ matrix variables $X, \ S \in \mathcal{S}^n$, each of which involves $n(n+1)/2$ real variables; for example, $n = 2000 \Rightarrow n(n+1)/2 \approx 2$ million.
- m linear equality constraints in $\mathcal{P}$ or m A_p's $\in \mathcal{S}^n$.

$\Downarrow$

◊ Exploit sparsity and structured sparsity.
◊ Enormous computational power$\Rightarrow$ parallel computation.

$$\mathcal{P}: \min \quad A_0 \bullet X \quad \text{sub.to} \quad A_p \bullet X = b_p \ (1 \leq p \leq m), \ \mathcal{S}^n \ni X \succeq O$$
$$\mathcal{D}: \max \quad \sum_{p=1}^m b_p y_p \quad \text{sub.to} \quad \sum_{p=1}^m A_p y_p + S = A_0, \ \mathcal{S}^n \ni S \succeq O$$

Outline

1. SDP (semidefinite program) and its dual
2. Primal-dual IPM (Interior-Point Method)
3. Various types of structured sparsities
4. Numerical results: structured sparsities + parallel
5. POPs (Polynomial Optimization Problems)
6. Rough sketch of SDP relaxation of POPs
7. Exploiting structured sparsity
8. Numerical results on POPs
9. Concluding remarks

$\mathcal{P}$: min $A_0 \bullet X$ sub.to $A_p \bullet X = b_p \ (1 \leq p \leq m), \ \mathcal{S}^n \ni X \succeq O$
$\mathcal{D}$: max $\sum_{p=1}^m b_p y_p$ sub.to $\sum_{p=1}^m A_p y_p + S = A_0, \ \mathcal{S}^n \ni S \succeq O$

Generic primal-dual IPM on a single CPU $\Rightarrow$ SDPA

Step 0: Choose $(X, y, S) = (X^0, y^0, S^0)$; $X^0 \succ O$ and $S^0 \succ O$. $k = 1$.
Step 1: Compute a search direction (dX, dy, dS). $\Rightarrow$ $Bdy = r$
Step 2: Choose α_p and α_d;
$X^{k+1} = X^k + \alpha_p dX \succ O, \ S^{k+1} = S^k + \alpha_d dS \succ O, \ y^{k+1} = y^k + \alpha_d dy.$
Step 3: Let $k = k + 1$. Go to Step 1.

B : $m \times m$ dense in general, computed from $A_1, \ldots, A_m, X, S$.

Major time consumption (second) on a single cpu implemention.

part	control11	theta6	maxG51
Elements of B	463.2	78.3	1.5
Cholesky fact. of B	31.7	209.8	3.0
dX	1.8	1.8	47.3
Other dense mat. comp.	1.0	4.1	86.5
Others	7.2	5.13	1.8
Total	505.2	292.3	140.2

$\mathcal{P}$: min $A_0 \bullet X$ sub.to $A_p \bullet X = b_p \ (1 \leq p \leq m), \ \mathcal{S}^n \ni X \succeq O$
$\mathcal{D}$: max $\sum_{p=1}^m b_p y_p$ sub.to $\sum_{p=1}^m A_p y_p + S = A_0, \ \mathcal{S}^n \ni S \succeq O$

B : $m \times m$ dense in general, computed from $A_1, \ldots, A_m$, X, S.

$$B_{pq} = X A_p S^{-1} \bullet A_q \ (1 \leq p \leq q \leq m).$$

Suppose that p is fixed.
How do we compute B_{pq} $(p \leq q \leq m)$ in large sclale & sparse cases?

X : dense, S^{-1} : dense,

$A_1, \ldots, A_m$: a few dense (or mildly dense), most sparse,

$f_q \equiv$ the number of nonzeros in A_q $(p \leq q \leq m)$.

Three formula for computing B_{pq} $(p \leq q \leq m)$

(Fujisawa-Kojima-Nakata '97)

	Formula $\mathcal{F}_1$ (for dense)	# of $\times$
1.	$F = A_p S^{-1}$	nf_p
2.	$G = XF$	n^3
3.	$B_{pq} = G \bullet A_q$	f_q $(p \leq q \leq m)$
Total	B_{pq} $(p \leq q \leq m)$	$nf_p + n^3 + \sum_{q=p}^{m} f_q$

$\mathcal{P}$: min $A_0 \bullet X$ sub.to $A_p \bullet X = b_p \ (1 \leq p \leq m)$, $\mathcal{S}^n \ni X \succeq O$
$\mathcal{D}$: max $\sum_{p=1}^m b_p y_p$ sub.to $\sum_{p=1}^m A_p y_p + S = A_0$, $\mathcal{S}^n \ni S \succeq O$

B : $m \times m$ dense in general, computed from $A_1, \ldots, A_m$, X, S.

$$B_{pq} = X A_p S^{-1} \bullet A_q \ (1 \leq p \leq q \leq m).$$

Suppose that p is fixed.
How do we compute B_{pq} $(p \leq q \leq m)$ in large sclale & sparse cases?

X : dense, S^{-1} : dense,

$A_1, \ldots, A_m$: a few dense (or mildly dense), most sparse,

$f_q \equiv$ the number of nonzeros in A_q $(p \leq q \leq m)$.

	Formula $\mathcal{F}_2$ (for mildly dense)	# of ×
1.	$F = A_p S^{-1}$	$n f_p$
2.	$B_{pq} = \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} [A_q]_{\alpha\beta} \left(\sum_{\gamma=1}^{n} X_{\alpha\gamma} F_{\gamma\beta} \right)$	$(n+1) f_q \ (p \leq q \leq m)$
Total	$B_{pq} \ (p \leq q \leq m)$	$n f_p + (n+1) \sum_{q=p}^{m} f_q$

$\mathcal{P}$: min $A_0 \bullet X$ sub.to $A_p \bullet X = b_p$ $(1 \leq p \leq m)$, $\mathcal{S}^n \ni X \succeq O$
$\mathcal{D}$: max $\sum_{p=1}^m b_p y_p$ sub.to $\sum_{p=1}^m A_p y_p + S = A_0$, $\mathcal{S}^n \ni S \succeq O$

B : $m \times m$ dense in general, computed from $A_1, \ldots, A_m$, X, S.

$$B_{pq} = X A_p S^{-1} \bullet A_q \ (1 \leq p \leq q \leq m).$$

Suppose that p is fixed.
How do we compute B_{pq} $(p \leq q \leq m)$ in large sclale & sparse cases?

X : dense, S^{-1} : dense,

$A_1, \ldots, A_m$: a few dense (or mildly dense), most sparse,

$f_q \equiv$ the number of nonzeros in A_q $(p \leq q \leq m)$.

Formula $\mathcal{F}_3$ (for sparse)	# of ×
$B_{pq} = \sum_{\gamma=1}^n \sum_{\epsilon=1}^n [A_q]_{\gamma\epsilon} \left(\sum_{\alpha=1}^n \sum_{\beta=1}^n X_{\gamma\alpha} [A_p]_{\alpha\beta} [S^{-1}]_{\beta\epsilon} \right)$	$(2f_p + 1) f_q$ $(p \leq q \leq m)$
B_{pq} $(p \leq q \leq m)$	$(2f_p + 1) \sum_{q=p}^m f_q$

$$\mathcal{P}: \min \quad A_0 \bullet X \quad \text{sub.to} \quad A_p \bullet X = b_p \ (1 \leq p \leq m), \ \mathcal{S}^n \ni X \succeq O$$
$$\mathcal{D}: \max \quad \sum_{p=1}^m b_p y_p \ \text{sub.to} \quad \sum_{p=1}^m A_p y_p + S = A_0, \ \mathcal{S}^n \ni S \succeq O$$

B : $m \times m$ dense in general, computed from $A_1, \ldots, A_m$, X, S.

$$B_{pq} = X A_p S^{-1} \bullet A_q \ (1 \leq p \leq q \leq m).$$

Suppose that p is fixed.
How do we compute B_{pq} $(p \leq q \leq m)$ in large sclale & sparse cases?

X : dense, S^{-1} : dense,

$A_1, \ldots, A_m$: a few dense (or mildly dense), most sparse,

$f_q \equiv$ the number of nonzeros in A_q $(p \leq q \leq m)$.

		Typical cases $p = 1$, $m = n$	
Formula	# of $\times$ for B_{pq} $(p \leq q \leq n)$	$f_q = n^2$	$f_q = 2$
$\mathcal{F}_1$ (for dense)	$n f_p + n^3 + \sum_{q=p}^m f_q$	$O(n^3)$	$O(n^3)$
$\mathcal{F}_2$ (for mildly dense)	$n f_p + (n+1) \sum_{q=p}^m f_q$	$O(n^4)$	$O(n^2)$
$\mathcal{F}_3$ (for sparse)	$(2 f_p + 1) \sum_{q=p}^m f_q$	$O(n^5)$	$O(n)$

$\mathcal{P}$: min $A_0 \bullet X$ sub.to $A_p \bullet X = b_p \ (1 \le p \le m),\ \mathcal{S}^n \ni X \succeq O$
$\mathcal{D}$: max $\sum_{p=1}^m b_p y_p$ sub.to $\sum_{p=1}^m A_p y_p + S = A_0,\ \mathcal{S}^n \ni S \succeq O$

B : $m \times m$ dense in general, computed from $A_1, \ldots, A_m$, X, S.

$$B_{pq} = X A_p S^{-1} \bullet A_q \ (1 \le p \le q \le m).$$

Suppose that p is fixed.
How do we compute B_{pq} $(p \le q \le m)$ in large sclale & sparse cases?

X : dense, S^{-1} : dense,

$A_1, \ldots, A_m$: a few dense (or mildly dense), most sparse,

$f_q \equiv$ the number of nonzeros in A_q $(p \le q \le m)$.

Numerical evaluation of Formula $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$

			cpu time / iteration second			Their suitable combination
problem	m	n	$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	used in SDPA
QAP	1021	101	61.3	29.5	-	4.5
GP	501	500	7247.2	52.0	6341.6	29.3
MC	944	300	2472.2	43.0	1.4	1.3

$\mathcal{P}$: min $A_0 \bullet X$ sub.to $A_p \bullet X = b_p \ (1 \leq p \leq m)$, $\mathcal{S}^n \ni X \succeq O$
$\mathcal{D}$: max $\sum_{p=1}^m b_p y_p$ sub.to $\sum_{p=1}^m A_p y_p + S = A_0$, $\mathcal{S}^n \ni S \succeq O$

B : $m \times m$ dense in general, computed from $A_1, \ldots, A_m$, X, S.

$$B_{pq} = X A_p S^{-1} \bullet A_q \ (1 \leq p \leq q \leq m).$$

Suppose that p is fixed.
How do we compute B_{pq} $(p \leq q \leq m)$ in large sclale & sparse cases?

X : dense, S^{-1} : dense,

$A_1, \ldots, A_m$: a few dense (or mildly dense), most sparse,

$f_q \equiv$ the number of nonzeros in A_q $(p \leq q \leq m)$.

X : dense, S^{-1} : dense

In some cases, $S = A_0 - \sum_{p=1}^m A_p y_p$ is sparse and X^{-1} can be sparse.

Use S and X^{-1} instead of S^{-1} and X!

⇒ SDPARA-C (the positive definite matrix completion technique)

⇒ Later

$$\mathcal{P}: \min \quad A_0 \bullet X \quad \text{sub.to} \quad A_p \bullet X = b_p \ (1 \leq p \leq m), \ \mathcal{S}^n \ni X \succeq O$$
$$\mathcal{D}: \max \quad \sum_{p=1}^m b_p y_p \quad \text{sub.to} \quad \sum_{p=1}^m A_p y_p + S = A_0, \ \mathcal{S}^n \ni S \succeq O$$

Outline

1. SDP (semidefinite program) and its dual
2. Primal-dual IPM
3. Various types of structured sparsities
4. Numerical results: structured sparsities + parallel
5. POPs (Polynomial Optimization Problems)
6. Rough sketch of SDP relaxation of POPs
7. Exploiting structured sparsity
8. Numerical results on POPs
9. Summary and concluding remarks

$\mathcal{P}$: min $A_0 \bullet X$ sub.to $A_p \bullet X = b_p \ (1 \leq p \leq m), \ \mathcal{S}^n \ni X \succeq O$
$\mathcal{D}$: max $\sum_{p=1}^m b_p y_p$ sub.to $\sum_{p=1}^m A_p y_p + S = A_0, \ \mathcal{S}^n \ni S \succeq O$

Structured sparsity

The aggregate sparsity pattern $\widehat{A}$: a symbolic $n \times n$ matrix:

$$\widehat{A}_{ij} = \begin{cases} \star \text{ if the } (i,j)\text{th element of } A_p \text{ is nonzero for } \exists p = 0,\ldots,m, \\ 0 \text{ otherwise,} \end{cases}$$

where $\star$ denotes a nonzero number.

Example: $m = 1$

$$A_0 = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ A_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 3 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 2 \end{pmatrix} \ \Rightarrow \ \widehat{A} = \begin{pmatrix} \star & \star & 0 & \star \\ \star & \star & \star & 0 \\ 0 & \star & \star & 0 \\ \star & 0 & 0 & \star \end{pmatrix}.$$

Next — three types of structured sparsity

$\mathcal{P}$: min $A_0 \bullet X$ sub.to $A_p \bullet X = b_p$ $(1 \leq p \leq m)$, $\mathcal{S}^n \ni X \succeq O$
$\mathcal{D}$: max $\sum_{p=1}^m b_p y_p$ sub.to $\sum_{p=1}^m A_p y_p + S = A_0$, $\mathcal{S}^n \ni S \succeq O$

Structured sparsity

The aggregate sparsity pattern $\widehat{A}$: a symbolic $n \times n$ matrix:

$$\widehat{A}_{ij} = \begin{cases} \star \text{ if the } (i,j)\text{th element of } A_p \text{ is nonzero for } \exists p = 0, \ldots, m, \\ 0 \text{ otherwise,} \end{cases}$$

where $\star$ denotes a nonzero number.

Structured sparsity-1 : $\widehat{A}$ is block-diagonal.

Then X, S have the same diagonal block structure as $\widehat{A}$.

$$\widehat{A} = \begin{pmatrix} B_1 & O & O \\ O & B_2 & O \\ O & O & B_3 \end{pmatrix}, \; B_i : \text{symmetric.}$$

Example: CH_3N : an SDP from quantum chemistry, Fukuda et al. 2005.
$m = 20,709$, $n = 12,802$, "the number of blocks in $\widehat{A}$" $= 22$,
the largest bl.size $= 3,211 \times 3,211$, the average bl.size $= 583 \times 583$.

$$\mathcal{P}: \min \quad A_0 \bullet X \quad \text{sub.to} \quad A_p \bullet X = b_p \ (1 \leq p \leq m), \ \mathcal{S}^n \ni X \succeq O$$
$$\mathcal{D}: \max \quad \sum_{p=1}^m b_p y_p \ \text{sub.to} \ \sum_{p=1}^m A_p y_p + S = A_0, \ \mathcal{S}^n \ni S \succeq O$$

Structured sparsity

The aggregate sparsity pattern $\widehat{A}$: a symbolic $n \times n$ matrix:

$$\widehat{A}_{ij} = \begin{cases} \star \text{ if the } (i,j)\text{th element of } A_p \text{ is nonzero for } \exists p = 0, \ldots, m, \\ 0 \text{ otherwise,} \end{cases}$$

where $\star$ denotes a nonzero number.

Structured sparsity-2 : $\widehat{A}$ has a sparse Cholesky factorization.

"a small bandwidth" "a small bandwidth + bordered"

$$\widehat{A} = \begin{pmatrix} \star & \star & O & O & O \\ \star & \star & \star & O & O \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ O & O & \star & \star & \star \\ O & O & \cdots & \star & \star \end{pmatrix}, \quad \widehat{A} = \begin{pmatrix} \star & \star & O & O & \star \\ \star & \star & \star & O & \star \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ O & O & \star & \star & \star \\ \star & \star & \cdots & \star & \star \end{pmatrix}, \quad \star : \text{bl.matrix} \neq O$$

- S : the same sparsity pattern as $\widehat{A}$. • X : fully dense.
- X^{-1} : the same sparsity pattern as $\widehat{A}$ $\Rightarrow$ Use X^{-1} instead X (the positive definite matrix completion used in SDPARA-C)

$$\mathcal{P}: \min \ A_0 \bullet X \quad \text{sub.to} \ A_p \bullet X = b_p \ (1 \leq p \leq m), \ \mathcal{S}^n \ni X \succeq O$$
$$\mathcal{D}: \max \ \sum_{p=1}^m b_p y_p \ \text{sub.to} \ \sum_{p=1}^m A_p y_p + S = A_0, \ \mathcal{S}^n \ni S \succeq O$$

Structured sparsity

The aggregate sparsity pattern $\widehat{A}$: a symbolic $n \times n$ matrix:

$$\widehat{A}_{ij} = \begin{cases} \star \text{ if the } (i,j)\text{th element of } A_p \text{ is nonzero for } \exists p = 0, \ldots, m, \\ 0 \text{ otherwise,} \end{cases}$$

where $\star$ denotes a nonzero number.

Structured sparsity-3 : block-diagonal $\widehat{A}$ + blockwise orthogonality,
for most pairs (p, q) $1 \leq p < q \leq m$,
A_p and A_q do not share nonzero blocks; hence $A_p \bullet A_q = 0$.
$\Rightarrow$ the Schur complement matrix B used in PDIPM becomes sparse.

$$A_1 = \begin{pmatrix} A_{11} & O & O \\ O & O & O \\ O & O & O \end{pmatrix}, \ A_2 = \begin{pmatrix} O & O & O \\ O & A_{22} & O \\ O & O & O \end{pmatrix}, \ A_3 = \begin{pmatrix} O & O & O \\ O & O & O \\ O & O & A_{33} \end{pmatrix}.$$

- An engineering application, Ben-Tal and Nemirovskii 1999.
- A sparse SDP relaxation of poly. opt. problem, Waki et al. 2005.
- Incorporated in SDPT3 and SeDuMi but not in SDPA.

$$\mathcal{P}: \min \quad A_0 \bullet X \quad \text{sub.to} \quad A_p \bullet X = b_p \ (1 \leq p \leq m), \ \mathcal{S}^n \ni X \succeq O$$
$$\mathcal{D}: \max \quad \sum_{p=1}^m b_p y_p \quad \text{sub.to} \quad \sum_{p=1}^m A_p y_p + S = A_0, \ \mathcal{S}^n \ni S \succeq O$$

Outline

1. SDP (semidefinite program) and its dual
2. Primal-dual IPM
3. Various types of structured sparsities
4. Numerical results: structured sparsities + parallel
5. POPs (Polynomial Optimization Problems)
6. Rough sketch of SDP relaxation of POPs
7. Exploiting structured sparsity
8. Numerical results on POPs
9. Summary and concluding remarks

$$\mathcal{P}: \min \quad A_0 \bullet X \quad \text{sub.to} \quad A_p \bullet X = b_p \ (1 \leq p \leq m), \ \mathcal{S}^n \ni X \succeq O$$
$$\mathcal{D}: \max \quad \sum_{p=1}^m b_p y_p \ \text{sub.to} \quad \sum_{p=1}^m A_p y_p + S = A_0, \ \mathcal{S}^n \ni S \succeq O$$

SDPs from quantum chemistry, Fukuda et al. 2005.

atoms/molecules	m	n	#blocks	the sizes of largest blocks
O	7230	5990	22	[1450, 1450, 450, ...]
HF	15018	10146	22	[2520, 2520, 792, ...]
CH_3N	20709	12802	22	[3211, 3211, 1014, ...]

number of processors		16	64	128	256
O	elements of B	10100.3	2720.4	1205.9	694.2
	Chol.fact. of B	218.2	87.3	68.2	106.2
	total	14250.6	4453.3	3281.1	2951.6
HF	elements of B	*	*	13076.1	6833.0
	Chol.fact. of B	*	*	520.2	671.0
	total	*	*	26797.1	20780.7
CH_3N	elements of B	*	*	34188.9	18003.3
	Chol.fact. of B	*	*	1008.9	1309.9
	total	*	*	57034.8	45488.9

$\mathcal{P}$: min $A_0 \bullet X$ sub.to $A_p \bullet X = b_p \ (1 \leq p \leq m),\ \mathcal{S}^n \ni X \succeq O$
$\mathcal{D}$: max $\sum_{p=1}^m b_p y_p$ sub.to $\sum_{p=1}^m A_p y_p + S = A_0,\ \mathcal{S}^n \ni S \succeq O$

Large-size SDPs by SDPARA-C (64 CPUs)

3 types of test Problems:

(a) SDP relaxations of randomly generated max. cut problems on lattice graphs with size 10×1000, 10×2000 and 10×4000.

(b) SDP relaxations of randomly generated max. clique problems on lattice graphs with size 10×500, 10×1000 and 10×2000.

(c) Randomly generated norm minimization problems

$$\text{min.} \left\| F_0 - \sum_{i=1}^{10} F_i y_i \right\| \ \text{sub.to} \ \ y_i \in \mathbb{R} \ (i = 1, 2, \ldots, 10)$$

where $F_i : 10 \times 9990,\ 10 \times 19990$ or 10×39990 and $\|G\|$ = the square root of the max. eigenvalue of $G^T G$.

In all cases, the aggregate sparsity pattern consists of one block and is very sparse.

$$\mathcal{P}: \min \quad A_0 \bullet X \quad \text{sub.to} \quad A_p \bullet X = b_p \ (1 \leq p \leq m), \ \mathcal{S}^n \ni X \succeq O$$
$$\mathcal{D}: \max \quad \sum_{p=1}^m b_p y_p \ \text{sub.to} \quad \sum_{p=1}^m A_p y_p + S = A_0, \ \mathcal{S}^n \ni S \succeq O$$

Large-size SDPs by SDPARA-C (64 CPUs)

	Problem	n	m	time (s)	memory (MB)
(a)	Cut(10×1000)	10000	10000	274.3	126
	Cut(10×2000)	20000	20000	1328.2	276
	Cut(10×4000)	40000	40000	7462.0	720
(b)	Clique(10×500)	5000	9491	639.5	119
	Clique(10×1000)	10000	18991	3033.2	259
	Clique(10×2000)	20000	37991	15329.0	669
(c)	Norm(10×9990)	10000	11	409.5	164
	Norm(10×19990)	20000	11	1800.9	304
	Norm(10×39990)	40000	11	7706.0	583

Outline

1. SDP (semidefinite program) and its dual
2. Primal-dual IPM
3. Various types of structured sparsities
4. Numerical results: structured sparsities + parallel
5. POPs (Polynomial Optimization Problems)
6. Rough sketch of SDP relaxation of POPs
7. Exploiting structured sparsity
8. Numerical results on POPs
9. Summary and concluding remarks

$\mathbb{R}^n$: the n-dim Euclidean space.

$x = (x_1, \ldots, x_n) \in \mathbb{R}^n$: a vector variable.

$f_p(x)$: a multivariate polynomial in $x \in \mathbb{R}^n$ $(p = 0, 1, \ldots, m)$.

POP: min $f_0(x)$ sub.to $f_p(x) \geq 0$ $(p = 1, \ldots, m)$.

Example: $n = 3$

$$
\begin{array}{rl}
\min & f_0(x) \equiv x_1^3 - 2x_1x_2^2 + x_1^2x_2x_3 - 4x_3^2 \\
\text{sub.to} & f_1(x) \equiv -x_1^2 + 5x_2x_3 + 1 \geq 0, \\
& f_2(x) \equiv x_1^2 - 3x_1x_2x_3 + 2x_3 + 2 \geq 0, \\
& f_3(x) \equiv -x_1^2 - x_2^2 - x_3^2 + 1 \geq 0, \\
& x_1(x_1 - 1) = 0 \text{ (0-1 integer)}, \\
& x_2 \geq 0, \; x_3 \geq 0, \; x_2x_3 = 0 \text{ (complementarity)}.
\end{array}
$$

- Various problems can be described as POPs.

- A unified theoretical model for global optimization in nonlinear and combinatorial optimization problems.

$$\text{POP:} \quad \min \ f_0(x) \quad \text{sub.to} \quad f_p(x) \geq 0 \ (p = 1, \ldots, m).$$

Some Examples: Unconstrained cases.

Minimize the genalized Rosenbrock funcion

$$f_0(x) = \sum_{i=1}^{n-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_{i+1})^2.$$

- $\bar{x}$: a global minimizer $\Leftrightarrow$ an $\bar{x}$ and an exact lower bound ζ such that $f(\bar{x}) = \zeta \leq f(x)$ for every x.
- How to exploit sparsity of polynomials
  $\Rightarrow$ the sparsity pattern of the Hessian matrix of $f(x)$

POP: min $f_0(x)$ sub.to $f_p(x) \geq 0$ $(p = 1, \ldots, m)$.

Some Examples: Constrained case 2

alkyl.gms : a benchmark problem from globallib

$$
\begin{array}{ll}
\text{min} & -6.3x_5x_8 + 5.04x_2 + 0.35x_3 + x_4 + 3.36x_6 \\
\text{sub.to} & -0.820x_2 + x_5 - 0.820x_6 = 0, \\
& 0.98x_4 - x_7(0.01x_5x_{10} + x_4) = 0, \\
& -x_2x_9 + 10x_3 + x_6 = 0, \\
& x_5x_{12} - x_2(1.12 + 0.132x_9 - 0.0067x_9^2) = 0, \\
& x_8x_{13} - 0.01x_9(1.098 - 0.038x_9) - 0.325x_7 = 0.574, \\
& x_{10}x_{14} + 22.2x_{11} = 35.82, \\
& x_1x_{11} - 3x_8 = -1.33, \\
& \text{lbd}_i \leq x_i \leq \text{ubd}_i \ (i = 1, 2, \ldots, 14).
\end{array}
$$

- How to exploit sparsity of polynomials

the sparsity pattern of the Hessian matrices of $f_0(x)$

+

the set of variables involved in $f_p(x)$ $(p = 1, 2, \ldots, m)$

For example, $0.98x_4 - x_7(0.01x_5x_{10} + x_4)$ involves x_4, x_5, x_7, x_{10}.

$$\text{POP:} \quad \min\ f_0(x) \quad \text{sub.to} \quad f_p(x) \geq 0\ (p = 1, \ldots, m).$$

Outline

1. SDP (semidefinite program) and its dual
2. Primal-dual IPM
3. Various types of structured sparsities
4. Numerical results: structured sparsities + parallel
5. POPs (Polynomial Optimization Problems)
6. Rough sketch of SDP relaxation of POPs
7. Exploiting structured sparsity
8. Numerical results on POPs
9. Summary and concluding remarks

POP: $\min\ f_0(x)$ sub.to $f_p(x) \geq 0\ (p = 1, \ldots, m)$.

SDP relaxation (Lasserre 2001) from a practical point of view.

(a) Linearization $\Longrightarrow$ relaxation.

(b) Strengthening the relaxation by valid poly. matrix inequalities (before (a)) $\Longrightarrow$ a poly. SDP equiv. to POP.

Represent a polynomial f as $f(x) = \sum_{\alpha \in \mathcal{G}} c(\alpha) x^{\alpha}$, where

$$\mathcal{G} = \text{a finite subset of } \mathbb{Z}^n_+ \equiv \{\alpha \in \mathbb{R}^n : \alpha_i \text{ is an integer} \geq 0\},$$
$$x^{\alpha} = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \text{ for } \forall x \in \mathbb{R}^n \text{ and } \forall \alpha \in \mathbb{Z}^n_+.$$

Replacing each x^{α} by a single variable $y_{\alpha} \in \mathbb{R}$, we have the linearization of $f(x)$: $F(y) = F((y_{\alpha} : \alpha \in \mathcal{G})) = \sum_{\alpha \in \mathcal{G}} c(\alpha) y_{\alpha}$.

Example

$$\begin{aligned} f(x_1, x_2) &= 2x_1 - 3x_1^2 + 4x_1 x_2^3 \\ &= 2x^{(1,0)} - 3x^{(2,0)} + 4x^{(1,3)} \end{aligned}$$

$\Downarrow$ (a) Linearization

$$F(y_{(1,0)}, y_{(2,0)}, y_{(1,3)}) = 2y_{(1,0)} - 3y_{(2,0)} + 4y_{(1,3)}.$$

POP: min $f_0(x)$ sub.to $f_p(x) \geq 0$ $(p = 1, \ldots, m)$.

SDP relaxation (Lasserre 2001) from a practical point of view.

(a) Linearization $\Longrightarrow$ relaxation.

(b) Strengthening the relaxation by valid poly. matrix inequalities (before (a)) $\Longrightarrow$ a poly. SDP equiv. to POP.

For $\forall$ finite $\mathcal{G} \subset \mathbb{Z}^n_+$, let $u(x;\mathcal{G})$ denote a column vector of x^α $(\alpha \in \mathcal{G})$. Then

(i) rank 1 sym.matrix $u(x;\mathcal{G})u(x;\mathcal{G})^T \succeq O$ for $\forall x \in \mathbb{R}^n$.

(ii) $f_p(x)u(x;\mathcal{G})u(x;\mathcal{G})^T \succeq O$ if $f_p(x) \geq 0$.

Example of (ii). $n = 2$. $\mathcal{G} = \{(0,0),(1,0)\}$.

$$(1-x_1x_2)\begin{pmatrix} 1 \\ x_1 \end{pmatrix}\begin{pmatrix} 1 \\ x_1 \end{pmatrix}^T \succeq O \quad \Leftrightarrow \begin{pmatrix} 1-x_1x_2 & x_1 - x_1^2x_2 \\ x_1 - x_1^2x_2 & x_1^2 - x_1^3x_2 \end{pmatrix} \succeq O$$

$\Updownarrow$ $\qquad\qquad$ $\Downarrow$ (a) Linearization

$1 - x_1x_2 \geq 0$

$\Downarrow$ (a) Linearization

$1 - y_{(1,1)} \geq 0$

$$\Leftarrow \begin{pmatrix} 1-y_{(1,1)} & y_{(1,0)} - y_{(2,1)} \\ y_{(1,0)} - y_{(2,1)} & y_{(2,0)} - y_{(3,1)} \end{pmatrix} \succeq O$$

LMI is stronger!

POP: $\min\ f_0(x)$ sub.to $f_p(x) \geq 0\ (p = 1, \ldots, m)$.

SDP relaxation (Lasserre 2001) from a practical point of view.

(a) Linearization $\Longrightarrow$ relaxation.

(b) Strengthening the relaxation by valid poly. matrix inequalities (before (a)) $\Longrightarrow$ a poly. SDP equiv. to POP.

For $\forall$ finite $\mathcal{G} \subset \mathbb{Z}^n_+$, let $u(x;\mathcal{G})$ denote a column vector of x^α $(\alpha \in \mathcal{G})$. Then

(i) rank 1 sym.matrix $u(x;\mathcal{G})u(x;\mathcal{G})^T \succeq O$ for $\forall x \in \mathbb{R}^n$.

(ii) $f_p(x)u(x;\mathcal{G})u(x;\mathcal{G})^T \succeq O$ if $f_p(x) \geq 0$.

Let $\mathcal{G}_p$ $(p = 1, \ldots, q > m)$ be finite subsets of $\mathbb{Z}^n_+$.

Polynomial SDP($\mathcal{G}_p$'s)

$$
\begin{array}{lll}
\min & f_0(x) & \\
\text{sub.to} & f_p(x)u(x,\mathcal{G}_p)u(x,\mathcal{G}_p)^T \succeq O\ (p = 1, \ldots, m) & \Leftarrow \text{(ii)} \\
& u(x,\mathcal{G}_p)u(x,\mathcal{G}_p)^T \succeq O\ (p = m+1, \ldots, q) & \Leftarrow \text{(i)}
\end{array}
$$

Apply (a) $\Rightarrow$ Linear SDP($\mathcal{G}_p$'s) = SDP relaxation of POP

Exploiting sparsity

$\Rightarrow$ How to choose sparse $\mathcal{G}_p$'s depending on sparsity of $f_p(x)$

relaxation order r = the max. degree of poly. in $u(x, \mathcal{G}_p)$

$$\text{POP: } \min\ f_0(x) \quad \text{sub.to} \quad f_p(x) \geq 0 \ (p = 1, \ldots, m).$$

Outline

1. SDP (semidefinite program) and its dual
2. Primal-dual IPM
3. Various types of structured sparsities
4. Numerical results: structured sparsities + parallel
5. POPs (Polynomial Optimization Problems)
6. Rough sketch of SDP relaxation of POPs
7. Exploiting structured sparsity
8. Numerical results on POPs
9. Summary and concluding remarks

POP: $\min_{x \in \mathbb{R}^n} \; f_0(x)$

G. Rosenbrock func: $f_0(x) = \sum_{i=1}^{n-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_{i+1})^2.$

Dense relaxation = Linearization of

$$\min \; f_0(x) \text{ s.t. } u(x,\mathcal{G})u(x,\mathcal{G})^T \succeq O,$$

where $u(x,\mathcal{G}) = (1, x_1, \ldots, x_n, x_1^2, x_1x_2, \ldots, x_2^2, x_2x_3, \ldots, x_n^2)^T$
the col. vector of all monomials in $x_1, \ldots, x_n$ with deg. ≤ 2.

- relaxation order $r = 2$ (the max. degree of poly. in $u(x,\mathcal{G})$).
- The size of $u(x,\mathcal{G})u(x,\mathcal{G})^T = \binom{n+2}{2}$; $\geq 20,000$ if n=200.

POP: $\min_{x \in \mathbb{R}^n} \; f_0(x)$

H : the sparsity pattern of the Hessian matrix of f_0

$$H_{ij} = \begin{cases} \star \text{ if } i = j \text{ or } \partial^2 f_0(x)/\partial x_i \partial x_j \not\equiv 0, \\ 0 \text{ otherwise.} \end{cases}$$

$\exists$ sparse Cholesky fact. of H.

G. Rosenbrock func: $f_0(x) = \sum_{i=1}^{n-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_{i+1})^2.$

- The Hessian matrix is sparse (tridiagonal).

Sparse relaxation = Linearization of

$$\min \; f_0(x) \text{ s.t. } \begin{pmatrix} 1 \\ x_i \\ x_{i+1} \\ x_i^2 \\ x_i x_{i+1} \\ x_{i+1}^2 \end{pmatrix} \begin{pmatrix} 1 \\ x_i \\ x_{i+1} \\ x_i^2 \\ x_i x_{i+1} \\ x_{i+1}^2 \end{pmatrix}^T \succeq O \; (i = 1, \ldots, n-1)$$

- relaxation order $r = 2$ (the max. degree of poly. in $u(x, \mathcal{G})$).
- Much smaller than Dense relaxation; the size is linear in n.

$$\text{POP:} \quad \min \ f_0(x) \quad \text{sub.to} \quad f_p(x) \geq 0 \ (p = 1, \ldots, m).$$

Outline

1. SDP (semidefinite program) and its dual
2. Primal-dual IPM
3. Various types of structured sparsities
4. Numerical results: structured sparsities + parallel
5. POPs (Polynomial Optimization Problems)
6. Rough sketch of SDP relaxation of POPs
7. Exploiting structured sparsity
8. Numerical results on POPs
9. Summary and concluding remarks

Numerical results on POPs

Software

- MATLAB for constructing sparse and dense SDP relaxation problems
- SeDuMi to solve SDPs.

Hardware

- 2.4GHz Xeon cpu with 6.0GB memory.

G.Rosenbrock function:

$$f(x) = \sum_{i=2}^{n} \left(100(x_i - x_{i-1}^2)^2 + (1 - x_i)^2\right)$$

- **Two minimizers on $\mathbb{R}^n$: $x_1 = \pm 1$, $x_i = 1$ $(i \geq 2)$.**
- **Sparse can not handle multiple minimizers effectively.**
- **Perturb the function or add $x_1 \geq 0 \Rightarrow$ unique minimizer.**
- **relaxation order $r = 2$ (the max. degree of poly. in $u(x, \mathcal{G})$).**

cpu in sec.				cpu in sec.	
Sparse	ϵ_{obj}	n	ϵ_{obj}	Sparse	Dense
0.2	5.1e-04	10	2.5e-08	0.2	10.6
0.3	1.8e-03	15	6.5e-08	0.2	756.6
4.6	5.9e-03	400	2.5e-06	3.7	—
8.6	8.3e-03	800	5.5e-06	6.8	—

$$\epsilon_{\text{obj}} = \frac{|\text{the lower bound for opt. value} - \text{the approx. opt. value}|}{\max\{1, |\text{the lower bound for opt. value}|\}}.$$

G.Rosenbrock function:

$$f(x) = \sum_{i=2}^{n} \left(100(x_i - x_{i-1}^2)^2 + (1 - x_i)^2\right)$$

- Two minimizers on $\mathbb{R}^n$: $x_1 = \pm 1$, $x_i = 1$ $(i \geq 2)$.
- Sparse can not handle multiple minimizers effectively.
- Perturb the function or add $x_1 \geq 0 \Rightarrow$ unique minimizer.
- relaxation order $r = 2$ (the max. degree of poly. in $u(x, \mathcal{G})$).

cpu in sec.				cpu in sec.	
Sparse	ϵ_{obj}	n	ϵ_{obj}	Sparse	Dense
0.2	5.1e-04	10	2.5e-08	0.2	10.6
0.3	1.8e-03	15	6.5e-08	0.2	756.6
4.6	5.9e-03	400	2.5e-06	3.7	—
8.6	8.3e-03	800	5.5e-06	6.8	—

When $n = 800$, SDP relaxation problem:

- $A_p : 4794 \times 4794$ $(p = 1, 2, \ldots, 7, 988)$ $\Rightarrow$ $B : 7,988 \times 7,988$.
- Each A_p consiss of 799 diagonal blocks with the size 6×6 matrices.
- $A_p \bullet A_q = 0$ for most pairs $(p, q) \Rightarrow$ a sparse Chol. fact. of B.

An optimal control problem from Coleman et al. 1995

$$\left.\begin{array}{l} \min \ \frac{1}{M}\sum_{i=1}^{M-1}(y_i^2 + x_i^2) \\ \text{s.t.} \quad y_{i+1} = y_i + \frac{1}{M}(y_i^2 - x_i), \quad (i = 1, \ldots, M-1), \quad y_1 = 1. \end{array}\right\}$$

Numerical results on sparse relaxation ($r = 2$)

M	# of variables	ϵ_{obj}	ϵ_{feas}	cpu
600	1198	3.4e-08	2.2e-10	3.4
700	1398	2.5e-08	8.1e-10	3.3
800	1598	5.9e-08	1.6e-10	3.8
900	1798	1.4e-07	6.8e-10	4.5
1000	1998	6.3e-08	2.7e-10	5.0

$$\epsilon_{\text{obj}} = \frac{|\text{the lower bound for opt. value} - \text{the approx. opt. value}|}{\max\{1, |\text{the lower bound for opt. value}|\}},$$

ϵ_{feas} = the maximum error in the equality constraints,

cpu : cpu time in sec. to solve an SDP relaxation problem.

alkyl.gms : a benchmark problem from globallib

$$
\begin{array}{ll}
\min & -6.3x_5x_8 + 5.04x_2 + 0.35x_3 + x_4 + 3.36x_6 \\
\text{sub.to} & -0.820x_2 + x_5 - 0.820x_6 = 0, \\
& 0.98x_4 - x_7(0.01x_5x_{10} + x_4) = 0, \\
& -x_2x_9 + 10x_3 + x_6 = 0, \\
& x_5x_{12} - x_2(1.12 + 0.132x_9 - 0.0067x_9^2) = 0, \\
& x_8x_{13} - 0.01x_9(1.098 - 0.038x_9) - 0.325x_7 = 0.574, \\
& x_{10}x_{14} + 22.2x_{11} = 35.82, \\
& x_1x_{11} - 3x_8 = -1.33, \\
& \text{lbd}_i \leq x_i \leq \text{ubd}_i \ (i = 1, 2, \ldots, 14).
\end{array}
$$

			Sparse			Dense		
problem	n	r	ϵ_{obj}	ϵ_{feas}	cpu	ϵ_{obj}	ϵ_{feas}	cpu
alkyl	14	2	4.1e-03	2.7e-01	0.9	6.3e-06	1.8e-02	17.6
alkyl	14	3	5.6e-10	2.0e-08	6.9	—	—	—

$r =$ relaxation order,

$$\epsilon_{\text{obj}} = \frac{|\text{the lower bound for opt. value} - \text{the approx. opt. value}|}{\max\{1, |\text{the lower bound for opt. value}|\}},$$

$\epsilon_{\text{feas}} =$ the maximum error in the equality constraints,

cpu : cpu time in sec. to solve an SDP relaxation problem.

Some other benchmark problems from globallib

			Sparse			Dense		
problem	n	r	ϵ_{obj}	ϵ_{feas}	cpu	ϵ_{obj}	ϵ_{feas}	cpu
ex3_1_1	8	3	6.3e-09	4.7e-04	3.3	0.7e-08	2.5e-03	211.4
ex5_4_2	8	3	8.1e-07	3.2e-02	5.5	0.7e-08	2.5e-03	597.8
st_e07	10	2	0.0e+00	8.1e-05	0.4	0.0e+00	8.8e-06	3.0
ex2_1_3	13	2	5.1e-09	3.5e-09	0.5	1.6e-09	1.5e-09	7.7
ex9_1_1	13	2	0.0	4.5e-06	1.5	0.0	9.2e-07	7.7
ex9_2_3	16	2	0.0e+00	5.7e-06	2.3	0.0e+00	7.5e-06	49.7
ex2_1_8	24	2	1.0e-05	0.0e+00	304.6	3.4e-06	0.0e+00	1946.6
ex5_2_2_c2	9	2	1.0e-02	7.2e+01	2.1	1.3e-04	2.7e-01	3.5
ex5_2_2_c2	9	3	5.8e-04	8.9e-01	332.9	-	-	-

- ex5_2_2_c2 ($r = 2$) — Dense is better.
- Sparse attains approx. opt. solutions with the same quality as Dense except ex5_2_2_c2.
- Sparse is much faster than Dense in large dim. and higher relaxation order cases.

Outline

1. SDP (semidefinite program) and its dual
2. Primal-dual IPM
3. Various types of structured sparsities
4. Numerical results: structured sparsities + parallel
5. POPs (Polynomial Optimization Problems)
6. Rough sketch of SDP relaxation of POPs
7. Exploiting structured sparsity
8. Numerical results on POPs
9. Summary and concluding remarks

SDP:
$(\mathcal{P})$ min $A_0 \bullet X$ sub.to $A_p \bullet X = b_p \ (1 \leq p \leq m), \ \mathcal{S}^n \ni X \succeq O$
$(\mathcal{D})$ max $\sum_{p=1}^m b_p y_p$ sub.to $\sum_{p=1}^m A_p y_p + S = A_0, \ \mathcal{S}^n \ni S \succeq O$

POP: min $f_0(x)$ sub.to $f_p(x) \geq 0 \ (1 \leq p \leq m)$.

Exploiting sparsity in SDPs

- Computing $B_{pq} = X A_p S^{-1} \bullet A_q \ (1 \leq p \leq q \leq m)$ in three formula $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$.
- Structured sparsity using the aggregated sparsity pattern $\widehat{A}$ over A_p $(1 \leq p \leq m)$.
- Numerical results on exploiting sparsity + parallel computation.

Exploiting sparsity in Lasserre's SDP relaxation of POPs

- Although the sparse SDP relaxation does not guarantee the global convergence and it is weaker than the original dense SDP relaxation, it is very powerful in practice.

SDP:

$$(\mathcal{P})\ \min\ A_0 \bullet X \quad \text{sub.to}\quad A_p \bullet X = b_p\ (1 \le p \le m),\ \mathcal{S}^n \ni X \succeq O$$

$$(\mathcal{D})\ \max\ \sum_{p=1}^m b_p y_p \quad \text{sub.to}\quad \sum_{p=1}^m A_p y_p + S = A_0,\ \mathcal{S}^n \ni S \succeq O$$

POP: $\min\ f_0(x)\ \text{ sub.to }\ f_p(x) \ge 0\ (1 \le p \le m).$

Some Future Works

- Solving larger scale SDPs and POPs.
  - (a) Exploiting sparsity in POPs and SDPs + parallel computation.
  - (b) Numerical stability.
- Incorporating sparse SDP relaxations into the branch-and-bound method.
- Practical implementation of a sparse SDP relaxation of polynomial SDPs and SOCPs, which were proposed in Kojima '03 and Kojima-Muramatsu '04, respectively.