Solving unconstrained binary quadratic optimization problems by Lasserre hierarchy and an interior-point method

Michal Kočvara

University of Birmingham

jointly with Soodeh Habibi and Michael Stingl

Modern Techniques of Very Large Scale Optimization, Edinburgh, May 2022

Poema
Polynomial Optimization, Efficiency through Moments and Algebra

Linear Semidefinite Optimization, notation

Primal problem

$$\max_{X\in\mathbb{S}^m,\, x_{\text{lin}}\in\mathbb{R}^m} \quad C\bullet X + d^\top x_{\text{lin}}$$

subject to

$$A_i \bullet X + (D^\top x_{\text{lin}})_i = b_i\,, \quad i = 1,\ldots,n$$

$$X \succeq 0,\; x_{\text{lin}} \geq 0$$

Dual problem

$$\min_{y\in\mathbb{R}^n,\, S\in\mathbb{S}^m,\, s_{\text{lin}}\in\mathbb{R}^m} \quad c^\top y$$

subject to

$$\sum_{i=1}^{n} y_i A_i + S = C, \quad S \succeq 0$$

$$Dy + s_{\text{lin}} = d, \quad s_{\text{lin}} \geq 0$$

Here: $A_i, C \in \mathbb{R}^{m\times m}$, symmetric, $X \bullet Y = \text{trace}(X^T Y)$

Unconstrained BQP

Find a global minimum of the non-convex binary problem

$$\boxed{\min_{x\in\mathbb{R}^s} x^\top Qx \quad \text{subject to} \quad x_i \in \mathcal{B}\,, \quad i = 1,\ldots,s} \qquad \text{(BQP)}$$

$Q \in \mathbb{R}^{s\times s}$ symmetric, $\mathcal{B}$ either $\{0,1\}$ or $\{-1,1\}$.

We do not assume any sparsity in Q, it is a generally dense matrix.

Technique: Hierarchy of convex conic relaxations

Kim-Kojima (2017): *"BQP instances ... can serve as challenging problems for developing conic relaxation methods"* (MK: *"and/or SDP software"*).

Unconstrained BQP

To find a global optimum, we use Lasserre hierarchy of semidefinite optimization (SDP) problems—relaxations—of growing dimension.

The SDP relaxations have the form

$$\boxed{\begin{aligned}&\min_{y\in\mathbb{R}^n} q^\top y\\ &\text{subject to } M(y) := \sum_{i=1}^n y_i M_i - M_0 \succeq 0\,.\end{aligned}} \qquad \text{(BQP-rel)}$$

Here M is a *moment matrix*, a (generally) dense matrix of a very specific form. (For $\omega = 1$, we have $q = \text{svec}(Q)$.)

In particular, if the solution of (BQP) is unique and the order of the relaxation is high enough, then rank M(y*) $= 2$.

Dimensions of the relaxations

variables	matrix size
n	$\sum_{i=1}^{\omega} \binom{s}{i}$

ω... relaxation order ($\omega = 1, 2, \ldots$)
s... dimension of BQP
$\frac{s^4}{24} \lessapprox n \leq 2^s - 1$

For instance, for $s = 9$:

ω	vars	matrix size
2	255	46
3	465	130
4	510	256
5	511	382

Dimensions of the relaxations

The theoretical lower bound on ω to get exact solution is $\lceil s/2 \rceil$ (Laurent 2003 and Fawzi et al. 2016).

This is confirmed by (specially constructed) examples!

This gives

s	ω	matrix size
21	11	784 625
31	16	759 852 346
41	21	$7.5 \cdot 10^{11}$
51	26	$7.5 \cdot 10^{14}$

So problems with $s > 20$ seem unsolvable by this approach.

Introducing Loraine

Loraine — LOw-RAnk INtErior point method

Loraine uses a primal-dual predictor-corrector interior-point method together with iterative solution of the resulting linear systems.
The iterative solver is a preconditioned Krylov-type method with a preconditioner utilizing low rank of the solution.

Implemented in Matlab (Julia version on the way)

Proved to be very efficient for SDP problems with very-low-rank solutions.

Only efficient under assumptions.

Loraine assumptions

Recall:

$$(P)\ \max_{X,\, x_{\text{lin}}} C \bullet X$$
$$\text{s.t. } A_i \bullet X + (D^\top x_{\text{lin}})_i = b_i \quad \forall i$$
$$X \succeq 0, \quad x_{\text{lin}} \geq 0$$

$$(D)\ \min_{y,\, S,\, s_{\text{lin}}} c^\top y$$
$$\text{s.t. } \sum_{i=1}^{n} y_i A_i - C = S, \quad S \succeq 0$$
$$Dy + s_{\text{lin}} = d, \quad s_{\text{lin}} \geq 0$$

We assume that the solution X^* has very low rank and develop a preconditioner based on this.

Loraine assumptions

Recall:

$$(P)\ \max_{X,\, x_{\text{lin}}} C \bullet X$$
$$\text{s.t. } A_i \bullet X + (D^\top x_{\text{lin}})_i = b_i \quad \forall i$$
$$X \succeq 0, \quad x_{\text{lin}} \geq 0$$

$$(D)\ \min_{y,\, S,\, s_{\text{lin}}} c^\top y$$
$$\text{s.t. } \sum_{i=1}^{n} y_i A_i - C = S, \quad S \succeq 0$$
$$Dy + s_{\text{lin}} = d, \quad s_{\text{lin}} \geq 0$$

Further assumptions:

- Slater condition + strict complementarity
- Sparsity: Define the matrix

$$\mathcal{A} = [\text{svec}\, A_1, \ldots, \text{svec}\, A_n].$$

  We assume that matrix-vector products with $\mathcal{A}$ and $\mathcal{A}^T$ may each be applied in $O(n)$ flops and memory.
- "Sparsity" of D: The inverse $(D^\top D)^{-1}$ and matrix-vector product with $(D^\top D)^{-1}$ may each be computed in $\mathcal{O}(n)$ flops and memory.

Low-rank preconditioner for Interior-Point method

In each iteration of the (primal-dual, predictor-corrector) interior-point method we have to solve two systems of linear equations in variable y:

$$((Hy)_i =)\ A_i \bullet [W(\sum_{j=1}^{n} y_j A_j)W] = r_i \quad \text{for } i = 1, \ldots, n\,.$$

Critical observation: If the solution X^* is low-rank, W will be low-rank.

Hence $W = W_0 + UU^T$ and

$$H = \mathcal{A}^T(W_0 \otimes W_0)\mathcal{A} + \underbrace{\mathcal{A}^T(U \otimes Z)}_{V}\underbrace{(U \otimes Z)^T\mathcal{A}}_{V^T}$$

Preconditioner

$$\boxed{\mathcal{H}_\alpha = \left(\sum_{i=1}^{p} \tau_i^2 I + D^T X_{\text{lin}} S_{\text{lin}}^{-1} D\right) + \widetilde{V}\widetilde{V}^T\,.}$$

Loraine and relaxed BQP?

Recall:

$$(P)\ \max_{X,\,x_{\rm lin}} C \bullet X \quad \text{s.t. } A_i \bullet X + (D^\top x_{\rm lin})_i = b_i \ \ \forall i, \quad X \succeq 0,\ \ x_{\rm lin} \geq 0$$

$$(D)\ \min_{y,\,S,\,s_{\rm lin}} c^\top y \quad \text{s.t. } \sum_{i=1}^{n} y_i A_i - C = S,\ \ S \succeq 0, \quad Dy + s_{\rm lin} = d,\ \ s_{\rm lin} \geq 0$$

Loraine assumes that the solution X^* has very low rank.

The SDP relaxation of BQP has the form of (D) with low-rank solution S, just the oposite of our assumption!

Using additional variables and equality constraints, we will reformulate it as (P) with low-rank solution X.

Re-writing the BQP relaxation

The dual problem to

$$\boxed{\begin{aligned} &\min_{y\in\mathbb{R}^n} q^\top y \\ &\text{subject to } M(y) := \sum_{i=1}^n y_i M_i - M_0 \succeq 0\,. \end{aligned}} \qquad \text{(BQP-rel)}$$

can be written as

$$\boxed{\begin{aligned} &\max_{z\in\mathbb{R}^{\tilde{n}}} (\mathrm{svec}(I))^\top z \\ &\text{subject to } \mathrm{smat}(z) \succeq 0 \\ &\qquad\qquad \mathbf{M}z = \tilde{q}\,, \end{aligned}} \qquad \text{(BQP-rel-dual)}$$

where $\widetilde{n} = m(m+1)/2$, $\mathbf{M} = (\mathrm{svec}(M_1), \ldots, \mathrm{svec}(M_n))^T \in \mathbb{R}^{n\times\widetilde{n}}$.

Now the dual solution to (BQP-rel-dual) has rank two, just as Loraine needs.

Handling linear equalities

Problem

$$\boxed{\begin{aligned} &\max_{z\in\mathbb{R}^{\tilde{n}}} (\mathrm{svec}(I))^\top z \\ &\text{subject to } \mathrm{smat}(z) \succeq 0 \\ &\qquad\qquad\quad \mathbf{M}z = \tilde{q} \end{aligned}}$$

(BQP-rel-dual)

is now in the right form.

But what about the (many) linear equality constraints?

Interior-point methods do not like them.

Treat them by ℓ_1 penalty:

$$\boxed{\begin{aligned} &\max_{z\in\mathbb{R}^{\tilde{n}}} (\mathrm{svec}(I))^\top z + \mu\|\mathbf{M}z - \tilde{q}\|_1 \\ &\text{subject to } \mathrm{smat}(z) \succeq 0\,, \end{aligned}}$$

with a penalty parameter $\mu > 0$.

Handling linear equalities

Introduce two new variables, $r \in \mathbb{R}^n$, $s \in \mathbb{R}^n$, satisfying

$$\mathbf{M}z - \tilde{q} = r - s, \quad r \geq 0,\ s \geq 0\,.$$

Using the identity $r = \mathbf{M}z - \tilde{q} + s$ to eliminate variable r, we arrive at our final problem

$$\max_{z\in\mathbb{R}^{\tilde{n}},\, s\in\mathbb{R}^n} (\mathrm{svec}(I))^\top z + \mu \sum_{i=1}^{n} ((\mathbf{M}z - \tilde{q})_i + 2s_i) \quad \text{(BQP-rel-final)}$$

$$\begin{aligned} \text{subject to } & \mathrm{smat}(z) \succeq 0 \\ & \mathbf{M}z - \tilde{q} + s \geq 0 \\ & s \geq 0 \end{aligned}$$

Problem (BQP-rel-final) is now in the required form.

Sparsity assumption

Recall the sparsity assumptions:

- Sparsity: Define the matrix

$$\mathcal{A} = [\operatorname{svec} A_1, \ldots, \operatorname{svec} A_n].$$

  We assume that matrix-vector products with $\mathcal{A}$ and $\mathcal{A}^T$ may each be applied in $O(n)$ flops and memory.

Every matrix A_i contains at most two nonzero elements, hence this is trivially satisfied.

- "Sparsity" of D: The inverse $(D^\top D)^{-1}$ and matrix-vector product with $(D^\top D)^{-1}$ may each be computed in $\mathcal{O}(n)$ flops and memory.

Lemma: There exists a permutation matrix $P \in \mathbb{R}^{\widetilde{n} \times \widetilde{n}}$ such that $P\mathbf{M}^T\mathbf{M}P^T$ is a block diagonal matrix with small full blocks. In particular, $\mathbf{M}^T\mathbf{M}$ is a sparse chordal matrix.

Numerical experiments

We solve

$$\min_{x\in\mathbb{R}^s} x^\top Q x \quad \text{subject to} \quad x_i \in \{-1,1\}\,, \quad i=1,\dots,s \qquad \text{(BQP)}$$

with randomly generated full-rank Q:

```
q = randn(s,1); Q = q*q';
for k=1:s-1
   if ceil(k/2)*2 == k
      q = randn(s,1); Q = Q - q*q';
   else
      q = randn(s,1); Q = Q + q*q';
   end
end
```

Conjecture: For Q constructed as above, relaxation order $\omega = 2$ is sufficient to get exact solution of (BQP).

Numerical experiments, problem sizes

$$\min_{x\in\mathbb{R}^s} x^\top Qx \quad \text{subject to} \quad x_i \in \{-1,1\}\,, \quad i=1,\dots,s \qquad \text{(BQP)}$$

Problems of growing dimension, from $s = 10$ to $s = 50$, relaxation order $\omega = 2$.

BQP size	problem (BQP-rel) variables	problem (BQP-rel) matrix size	problem (BQP-rel-final) variables	problem (BQP-rel-final) matrix size	problem (BQP-rel-final) lin. con.
10	385	56	1 981	56	770
15	1 940	121	9 321	121	3 880
20	6 195	211	28 561	211	12 390
25	15 275	326	68 576	326	30 550
30	31 930	466	140 741	466	63 860
35	59 535	631	258 904	631	119 070
40	102 090	821	439 521	821	204 180
45	164 220	1 036	701 386	1 036	328 440
50	251 175	1 276	1 065 901	1 276	502 350

Numerical experiments, results

Randomly generated BQP problems, relaxation order $\omega = 2$.

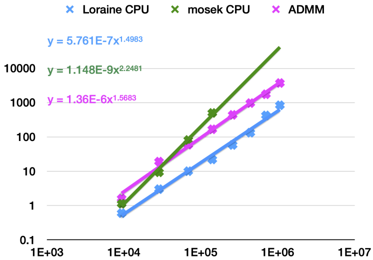
	(BQP-rel-final)			(BQP-rel)		
BQP	Loraine			Mosek	ADMM	
size	iter	CG iter	CPU	CPU	iter	CPU
10	10	256	0.1	0.27	995	0.38
15	10	725	1.3	1.13	1036	1.62
20	11	423	2.4	9.3	3836	19
25	13	294	5.5	81	5888	59
30	13	326	14	496	8906	166
35	15	530	43	mem	10903	436
40	14	730	106		13258	960
45	16	896	230		16731	1737
50	16	1114	431		21589	3689

iMac with 3.6 GHz 8-Core Intel Core i9 and 40 GB RAM

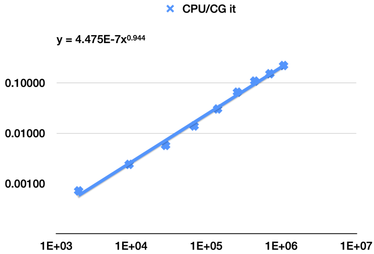
Numerical Results

BQP problems:

Mosek (direct solver) vs Loraine (iterative solver) vs ADMM



(a) Mosek vs Loraine vs ADMM CPU time



(b) Loraine: time per CG iterations

~ THE END ~

Loraine preprint available: arXiv:2105.08529

S. Habibi, A. Kavand, M. Kočvara and M. Stingl:
Barrier and penalty methods for low-rank semidefinite programming with application to truss topology design