Is Lagrangian (SDP) Relaxation Best for Quadratic Optimization?

Henry Wolkowicz
University of Waterloo

OUTLINE

- Quadratically constrained quadratic program, QQP.
- Special case of Max-Cut, MC, problem:
  equivalent bounds to Lagrangian (SDP) relaxation
  redundant constraints - strengthened SDP bound
- SDP equivalent to Lagrangian relaxation for QQPs.
- Theme (or conjecture): *Lagrangian relaxation is best*
- zero duality gap for nonconvex problems

We work with quadratically constrained quadratic programs

QQPs

$$q_i(x) := x^T Q_i x + 2 g_i^T x + \alpha_i, \quad Q_i = Q_i^T$$

$$(\mathrm{QQP}) \quad \begin{array}{rcl} q^* := & \min & q_0(x) \\ & \text{subject to} & q_k(x) \leq 0 \\ & & k \in \mathcal{I} := \{1, \ldots, m\} \\ & & x \in \Re^n \end{array}$$

Lagrangian of QQP

$$L(x, \lambda) := q_0(x) + \sum_{k \in \mathcal{I}} \lambda_k q_k(x),$$

$\lambda = (\lambda_k) \geq 0$ are nonnegative Lagrange multipliers.

Lagrange multipliers - two uses.

1. Classical: if a CQ holds at optimum $\bar{x}$, then the KKT necessary conditions for optimality hold

$$\begin{aligned} \bar{\lambda} \geq 0, \nabla L(\bar{x}, \bar{\lambda}) &= 0 \\ \bar{x} &\text{ feasible} \\ \bar{\lambda}_k q_k(\bar{x}) &= 0, \forall k \in \mathcal{I} \end{aligned}$$

If the Lagrangian at $\bar{\lambda}$ is also convex in x, then KKT sufficient for optimality.

2. Lagrangian Relaxation (duality):

$$q^* \geq d^* := \max_{\lambda \geq 0} \min_x q_0(x) + \sum_{k \in \mathcal{I}} \lambda_k q_k(x).$$

Zero duality gap holds if $q^* = d^*$.

Strong duality holds if $q^* = d^*$ and also d^* is attained.

Moreover, d^* can be efficiently evaluated using SDP.

Questions:
- Can one do better than the Lagrangian relaxation?
- When does a zero duality gap hold?
- What redundant constraints can be added to close (or reduce) the duality gap?

Tractable Relaxations of Max-Cut

$$(MCQ) \quad \mu^* := \max_{x\in\mathcal{F}} q_0(x) \quad (:= x^TQx - 2c^Tx).$$

where $\mathcal{F} = \{\pm 1\}^n$

---

perturbing diagonal of Q on $\mathcal{F}$:

$$\begin{array}{rcl} q_u(x) & := & x^T(Q + \mathrm{Diag}\,(\mathrm{u}))\mathrm{x} - 2\mathrm{c}^\mathrm{T}\mathrm{x} - \mathrm{u}^\mathrm{T}\mathrm{e} \\ & = & q_0(x), \quad \forall x \in \mathcal{F}. \end{array}$$

Bound 0: trivial bound from diagonal perturbations:

$$\mu^* \leq f_0(u) := \max_x q_u(x).$$

The function f_0 can take on the value $+\infty$. Let

$$S := \left\{ u : u^T e = 0, Q + \text{Diag}\,(\text{u}) \preceq 0 \right\}.$$

Then:

$$\boxed{\begin{array}{c} \mu^* \leq B_0 := \min_u f_0(u) \\ \left(= \min_{u^T e = 0} f_0(u), \text{ if } S \neq \emptyset \right). \end{array}}$$

Using the hidden semidefinite constraint:

$$\mu^* \leq B_0 = \min_{Q + \text{Diag}\,(\text{u}) \preceq 0} f_0(u).$$

Bound 1: relax the feasible set to the sphere of radius $\sqrt{n}$ (tractable trust region subproblem, TRS):

$$\mu^* \leq f_1(u) := \max_{||x||^2=n} \; q_u(x)$$

and

$$\boxed{\mu^* \leq B_1 := \min_u f_1(u).}$$

The inner maximization problem is called a trust region subproblem and is tractable. This bound provides the central tool in the proofs of equivalence.

Bound 2: box constraint:

$$\mu^* \leq f_2(u) := \max_{|x_i| \leq 1} \; q_u(x).$$

add the semidefinite constraint to make bound tractable.

$$\mu^* \leq \min_u f_2(u)$$

and

$$\boxed{\mu^* \leq B_2 := \min_{Q + \text{Diag}\,(\text{u}) \preceq 0} f_2(u).}$$

Bound B_1^c: lift to eigenvalue bound:

$$Q^c := \begin{bmatrix} 0 & -c^T \\ -c & Q \end{bmatrix}$$

$$q_u^c(y) := y^T(Q^c + \mathrm{diag}\,(\mathrm{u}))\mathrm{y} - \mathrm{u}^{\mathsf{T}}\mathrm{e}$$

$$\mu^* \leq f_1^c(u) := \max_{||y||^2=n+1} q_u^c(y)$$

where

$$\max_{||y||^2=n+1} q_u^c(y) = (n+1)\lambda_{\max}(Q^c+\mathrm{diag}\,(\mathrm{u}))-\mathrm{u}^{\mathsf{T}}\mathrm{e}$$

$$\boxed{\mu^* \leq B_1^c := \min_u f_1^c(u).}$$

Similarly, we get equivalent bounds B_0^c and homogenized bounds for the other models.

Bound B_3: SDP bound:

After homogenization ($c = 0$), use

$$x^T Q x = \text{trace}\, \mathrm{x}^{\mathrm{T}} \mathrm{Qx} = \text{trace}\, \mathrm{Qxx}^{\mathrm{T}}$$

and, for $x \in \mathcal{F}$, $y_{ij} = x_i x_j$ defines a symmetric, rank one, positive semidefinite matrix Y with diagonal elements 1. Relax the rank one condition.

$$\boxed{\begin{array}{lcc} B_3 := & \max & \text{trace}\, \mathrm{QY} \\ & \text{subject to} & \text{diag}\,(\mathrm{Y}) = \mathrm{e} \\ & & Y \succeq 0. \end{array}}$$

Summary: ref. Polyak-Rendl-Wolkowicz (PRW) 1995. (without restrictions e.g. $u^T e = 0$)

$$\boxed{\begin{array}{rcl} B_0 & = & \min\limits_u \max\limits_x q_u(x) \\ B_1 & = & \min\limits_u \max\limits_{x^T x = n} q_u(x) \\ B_2 & = & \min\limits_u \max\limits_{-1 \le x_i \le 1} q_u(x) \\ B_3 & = & \max\{\mathrm{trace}\, \mathrm{Q}^{\mathrm{c}}\mathrm{Y} : \mathrm{diag}\,(\mathrm{Y}) = \mathrm{e},\ \mathrm{Y} \succeq 0.\} \\ B_1^c & = & \min\limits_u \max\limits_{y^T y = n+1} q_u^c(y) \end{array}}$$

Now replace ±1 constraints with $x_i^2 = 1, \forall i$.

$$\begin{array}{lll} (P_E) & \text{max} & q_0(x) = x^T Q x - 2c^T x \\ & \text{subject to} & x_i^2 = 1, \quad i = 1, \cdots, n. \end{array}$$

B_L denotes Lagrangian relaxation bound.

Following our theme:

Theorem B_L equals all above bounds.

ref PRW/95.
(The proofs come from exploiting strong Lagrangian duality of TRS)

A Strengthened SDP Bound for MC ref Anjos-Wolkowicz/99.

(second lifting - this is motivated by the strong duality results that follow below and are from ref. Anstreicher-Wolkowicz/98 where redundant constraints using XX^T are shown to close the duality gap.)

first lifting procedure

$$X = xx^T$$

implies second lifting procedure

$$X^2 = xx^T xx^T = nX.$$

and provides an equivalent quadratic matrix model for MC

$$
\begin{array}{rll}
\mu^* := & \max & \operatorname{trace} QX \\
& \text{s.t.} & \operatorname{diag}(X) = e \\
& & X^2 - nX = 0 \\
& & X \succeq 0.
\end{array}
$$

Note: $X^2 = nX$, $\operatorname{trace} X = n$ implies X is rank one. But, we cannot solve this nonconvex problem in general, i.e. we have to look at the Lagrangian relaxation. (We can also add the constraint $X \circ X = E$, where E is the matrix of ones, and $\circ$ denotes the Hadamard matrix product.)

notation (interesting operators and adjoints):

- $S \in \mathcal{S}_n, \quad t(n) = \frac{n(n+1)}{2}$
- $s = \mathrm{svec}\,(S) \in \Re^{t(n)}$ vector formed (columnwise) from S ignoring strictly lower triang.
- $S = \mathrm{sMat}\,(s)$ inverse of svec
- $\mathrm{hMat}\,(v)$ is adjoint of svec , off-diagonal terms are multiplied by a half
- $\mathrm{dsvec}\,(S)$ is adjoint of sMat , operator like svec but off diagonal elements are multiplied by 2
- $\mathrm{sdiag}\,(s) := \mathrm{diag}\,(\mathrm{sMat}\,(\mathrm{s}))$
- $\mathrm{vsMat}\,(s) = \mathrm{vec}\,(\mathrm{sMat}\,(s))$

$$\mathrm{vsMat}^*(s) = \mathrm{dsvec}\,\left(\left(\mathrm{Mat}\,(v) + \mathrm{Mat}\,(v)^T\right)/2\right).$$

As above - use equivalent program

$$\mathrm{MC_O} \qquad \begin{array}{rll} \mu^* = & \max & \mathrm{trace}\, QX \\ & \text{s.t.¡} & \mathrm{diag}\,(X) = e \\ & & X^2 - nX = 0. \end{array}$$

Replace objective with $\frac{1}{n} QX^2$ linear constraint with $||\mathrm{diag}\,(X) - e||^2 = 0$ and homogenize the problem. Use $X = \mathrm{sMat}\,(x)$.

$$\mathrm{MC_O} \qquad \begin{array}{rll} \mu^* = & \max & \frac{1}{n}\mathrm{trace}\, Q\ (\mathrm{sMat}\,(x))^2 \\ & \text{s.t.} & \mathrm{sdiag}\,(x)^T \mathrm{sdiag}\,(x) \\ & & \quad - 2e^T \mathrm{sdiag}\,(x) y_0 + n = 0 \\ & & \mathrm{sMat}\,(x)^2 - n\,\mathrm{sMat}\,(x) y_0 = 0 \\ & & 1 - y_0^2 = 0. \end{array}$$

(Also include the Hadmard product constraint.)

The dual of this SDP is the strengthened SDP relaxation of MC:

$$\begin{array}{rcl} \mu^*_{MCSDP3} = & \max & \operatorname{trace} H_c Y \\ & \text{s.t.} & \mathcal{H}_1^*(Y) = n \\ & & \mathcal{H}_2^*(Y) = 0 \\ & & \mathcal{H}_3^*(Y) = 0 \\ & & \mathcal{H}_4^*(Y) = 1 \\ & & Y \succeq 0. \end{array}$$

Further: restrict $Y_{0j} = \operatorname{diag}(Y)_j = 1$ corresponding to diagonal of X.

The details of this derivation and numerical tests are contained in a preliminary research report with URL:

http://orion.math.uwaterloo.ca:80/~hwolkowi /henry/reports/ABSTRACTS.html#strengthMC

We now continue to look at various QQPs and the properties related to duality gaps.

Convex Quadratic Program

$$\text{CQP} \qquad \mu^* := \begin{array}{ll} \min & q_0(x) \\ \text{s.t.} & q_k(x) \le 0, \quad k = 1, \ldots m, \end{array}$$

where all $q_i(x)$ are convex quadratic functions. The dual is

$$\text{DCQP} \qquad \nu^* := \max_{\lambda \ge 0} \min_{x} \; q_0(x) + \sum_{k=1}^{m} \lambda_k q_k(x).$$

•(KKT) conditions are sufficient for global optimality

•if the primal value of CQP is bounded then it is attained and there is no duality gap, (e.g. ref Peterson-Ecker or Terlaky).

• the dual may not be attained

Nonconvex Cases

1. Rayleigh Quotient

$A = A^T \in \mathcal{S}_n$

$$\lambda_{\text{min}} = \text{min}\{x^T A x : x^T x = 1\}.$$

tractable (nonconvex) problem

no (Lagrangian) duality gap even though this is a nonconvex problem

$$\begin{array}{rcl} \mu^* & := & \max\limits_{\lambda} \min\limits_{x} \; x^T A x - \lambda(x^T x - 1) \\ & = & \max\limits_{A - \lambda I \succeq 0} \min\limits_{x} \; x^T (A - \lambda I) x + \lambda \\ & = & \max\limits_{A - \lambda I \succeq 0} \; \lambda \\ & = & \lambda_{\text{min}} \end{array}$$

2. Trust Region Subproblem

$$\begin{array}{rll} \mu^* := \min & q_0(x) & \\ \text{s.t.} & x^T x - \delta^2 \leq 0 \text{ (or } = 0). \end{array}$$

(more general nonconvex constraints $\alpha \leq q(x) \leq \beta$.)
for "$\leq$," the Lagrangian dual is:

$$\text{DTRS} \qquad \nu^* := \max_{\lambda \geq 0} \min_x \; q_0(x) + \lambda(x^T x - \delta^2).$$

strong duality holds and equivalent to (ref Stern-Wolkowicz/95) the (concave) nonlinear semidefinite program

$$\begin{array}{lrl} \text{DTRS} & \nu^* := \max & g_0^T (Q + \lambda I)^\dagger g_0 - \lambda \delta^2 \\ & \text{s.t.} & Q + \lambda I \succeq 0 \\ & & \lambda \geq 0. \end{array}$$

Again our theme holds: if we have a tractable problem then strong duality holds. (tractable:- ref e.g. More-Sorensen/83) short proof of strong duality (e.g. A. Lewis)

WLOG TRS is nonconvex; smallest eigenvalue of Q_0, denoted γ, is negative:

$$\begin{aligned}
\mu^* &= \min_{x^Tx\leq\delta^2} && x^T(Q_0-\gamma I)x-2c_0^tx+\gamma x^Tx \\
&= \min_{x^Tx=\delta^2} && x^T(Q_0-\gamma I)x-2c_0^tx+\gamma x^tx, \\
&= \min_{x^Tx=\delta^2} && x^T(Q_0-\gamma I)x-2c_0^tx+\gamma\delta^2 \\
&= \min_{x^Tx\leq\delta^2} && x^T(Q_0-\gamma I)x-2c_0^tx+\gamma\delta^2, \\
&= \max_{\lambda\geq 0}\min_{x} && x^T(Q_0-\gamma I)x-2c_0^tx+\lambda(x^Tx-\delta^2)+ \\
&= \max_{\lambda\geq 0}\min_{x} && x^TQ_0x-2c_0^tx+(\lambda-\gamma)(x^Tx-\delta^2) \\
&\leq \max_{\lambda\geq\gamma}\min_{x} && x^TQ_0x-2c_0^tx+(\lambda-\gamma)(x^Tx-\delta^2) \\
&= \nu^*\leq\mu^*.
\end{aligned}$$

3. Two Trust Region Subproblem TTRS consists in minimizing a (possibly nonconvex) quadratic function subject to a norm and a least squares constraint.

ref in SQP methods by Celis-Dennis-Tapia.

TTRS can have a nonzero duality gap, ref Peng-Yuan.

if objective not convex, then the primal may not be attained, ref Luo-Zhang.

TRS can have at most one local and nonglobal optimum, ref Martinez.

Still an open problem whether TTRS is an NP-hard or a polynomial time problem.

Note: a quadratic constraint can be written as

$$y^t P y \leq \delta$$

after the homogenization. This is lifted to

$$\text{trace}\,\mathsf{PY} \leq \delta.$$

Second lifting:

$$YPY \preceq \delta Y.$$

This strictly strengthens the SDP relaxation.

Orthogonally Constrained Programs with Zero Duality Gaps

permutation matrices are a subset of orthogonal matrices

$$\Pi \subset \mathcal{O} := \{X : XX^T = I\}$$

(Stiefel manifold ref e.g. Edelman,Arias,Smith)

A and B $n \times n$ symmetric matrices

$$\begin{array}{llll} \mathrm{QQP}_{\mathrm{O}} & \mu^* := & \min & \operatorname{trace} \mathrm{AXBX}^{\mathrm{T}} \\ & & \text{s.t.} & XX^T = I. \end{array}$$

Tractable problem; can be solved using classical Hoffman-Wielandt inequality. Provides bounds for e.g. QAP.

Proof using first application of Lagrange multipliers:

$$L(X,S) = \text{trace}\, \mathsf{AXBX}^{\mathsf{T}} + \mathsf{S}(\mathsf{XX}^{\mathsf{T}} - \mathrm{I})$$

$$0 = \langle \nabla L(X,S), h \rangle = 2\text{trace}\, \mathsf{AXBh}^{\mathsf{T}} + \mathsf{SXIh}^{\mathsf{T}} \quad \forall \mathsf{h}.$$

Therefore $AXBX^T = -S = -S^T$, i.e. A and XBX^T are mutually diagonalizable. The optimal value is then

$$\mu^* = \sum_i \lambda_i(A)\lambda_{n-i+1}(B)$$

But, Lagrangian dual can have a duality gap, ref Zhao,Karisch,Rendl,Wolkowicz

Lagrangian of (P)

$L(X,S) = \mathrm{trace\,AXBX^t} + \mathrm{trace\,SXX^t} - \mathrm{trace\,S}.$

Lagrangian dual is

$$\nu^* = \max_S \min_X L(X,S).$$

The hidden constraint (Hessian is psd) yields the dual

$$(D) \quad \begin{array}{ccc} \mu^D = & \max\limits_{\hat S = \hat S^t} & -\mathrm{trace\,\hat S} \\ & \text{subject to} & (B \otimes A + I \otimes \hat S) \succeq 0. \end{array}$$

A simple Example

Consider the following 2 × 2 example

$$A := \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \qquad B := \begin{pmatrix} 3 & 0 \\ 0 & 4 \end{pmatrix}.$$

$\mu^* = 10$

$$B \otimes A = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix}.$$

But $s_{11} \geq -3$ and $s_{22} \geq -6$; to maximize (D), equality must hold, and therefore $-\mathrm{trace}\,\hat{\mathsf{S}} = 9$ in the optimum.

However, add redundant constraint $X^T X = I$. Get $T \otimes I$ in Hessian and $-\mathrm{trace}\,\mathsf{T}$ in objective function. $s_{11} \geq -3 - t_{11}$ $s_{11} \geq -4 - t_{22}$ $s_{22} \geq -6 - t_{11}$ $s_{22} \geq -8 - t_{22}$ So $t_{22} = -1$. Duality gap is closed.

Summary: constraints $XX^T = I$ and $X^TX = I$ are equivalent. Add redundant constraints.

$$\mathrm{QQP_{OO}} \qquad \begin{array}{rll} \mu^O := & \min & \mathrm{trace}\, \mathrm{AXBX}^{\mathrm{T}} \\ & \text{s.t.} & XX^T = I,\ X^TX = I. \end{array}$$

dual problem is

$$\begin{array}{rll} & \mathrm{DQQP_{OO}} & \\ \mu^O \geq \mu^D := & \max & \mathrm{trace}\, \mathrm{S} + \mathrm{trace}\, \mathrm{T} \\ & \text{s.t.} & (I \otimes S) + (T \otimes I) \preceq (B \otimes A) \\ & & S = S^T,\ T = T^T. \end{array}$$

Theorem (Anstreicher-Wolkowicz/98) Strong duality holds for $\mathrm{QQP_{OO}}$ and $\mathrm{DQQP_{OO}}$, i.e. $\mu^D = \mu^O$ and both primal and dual are attained. ∎

Theme again holds.

Other applications:

Weighted Sums of Eigenvalues;

Graph Partitioning Problem;

TRS like constraints $\{X : XX^T \preceq I\}$ (ref Anstreicher-Wolkowicz-Xin-Yuan/99) (extension of Hoffman-Wielandt inequality)

Concluding Remarks:

In each case of a tractable bound for a nonconvex problem, the structure allows for redundant constraints to be added to close the Lagrangian duality gap.

What is the correct question about “best”? Can we close or reduce the duality gap in general?