Duality and Electromagnetic Inverse Design in Linear Media

Sean Molesky,[1,*] Pengning Chao,[2,*] and Alessio Amaolo[3]

[1] *Department of Engineering Physics, Polytechnique Montréal, Montréal, Québec H3T 1J4, CAN*
[2] *Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[3] *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA*

...

§INTRODUCTION

Definitions and Assumptions

A complex *quadratic* function is a function of the form

$$f : \mathbb{C}^n \to \mathbb{R} \ni \mathbf{x} \mapsto 2\Re\left[\mathbf{s}^\dagger \mathbf{x}\right] - \mathbf{x}^\dagger \mathbf{A} \mathbf{x} + c,$$

with $\mathbf{A} : \mathbb{C}^n \to \mathbb{C}^n$ Hermitian, $\mathbf{s}^\dagger \in \mathbb{C}^{n\dagger}$ and $c \in \mathbb{R}$. A complex quadratic function is *positive definite* or *compact* (resp. *positive semi-definite*) if $\mathbf{A} \succ 0$ ($\mathbf{A} \succeq 0$.) A *quadractically constrained concave program* (*QCCP*) is a optimization problem $\mathtt{P}$ of the form

$$\begin{aligned} &\max_{\mathbf{x} \in F_{\boldsymbol{\kappa}}} f_o\left(\boldsymbol{x}\right) \qquad (1)\\ &\ni \left(\forall j \in J\right) f_j\left(\mathbf{x}\right) \geq 0, \end{aligned}$$

wherein f_o is a concave function, J is some finite indexing set, and there exists a collection of scalars $\boldsymbol{\kappa} = \left\{\kappa_j\right\}_{j \in J}$ such that the non-negative domain of

$$f_{\boldsymbol{\kappa}}\left(\mathbf{x}\right) = \sum_{j \in J} \kappa_j f_j\left(\boldsymbol{x}\right),$$

with $\kappa_j \geq 0$ for all $j \in J$, is compact. Any quadratic function formed as a sum of the "basis" quadratic constraint functions of a QCCP $\mathtt{P}$,

$$f_{\boldsymbol{\phi}}\left(\boldsymbol{x}\right) = 2\Re\left[\boldsymbol{s}^\dagger_{\boldsymbol{\phi}} \mathbf{x}\right] - \mathbf{x}^\dagger \mathbf{A}_{\boldsymbol{\phi}} \mathbf{x} + v_{\boldsymbol{\phi}} = \sum_{j \in J} \phi_j f_j\left(\mathbf{x}\right),$$

is a *composite constraint*. $F_{\boldsymbol{\phi}} = \left\{\boldsymbol{x} \in \mathbb{C}^n \mid f_{\boldsymbol{\phi}}\left(\boldsymbol{x}\right) \geq 0\right\}$ is used to denote the feasible set of $f_{\boldsymbol{\phi}}$—so that that the $F_{\boldsymbol{\kappa}}$ set appearing in Eq. (1) is compact and convex [1].

Hereafter, by "QCCP", we mean the subset of QCCPs that are *feasible*—$F_{\mathtt{P}} = \cap_{j \in J} F_j \neq \emptyset$—with achievable objective values greater than zero. With $\mathbf{s}_{\boldsymbol{\phi}} = \sum_Q \phi_j \mathbf{s}_j$, $\mathbf{s}_{\boldsymbol{\psi}} = \mathbf{s}_o + \mathbf{s}_{\boldsymbol{\phi}}$, $\mathbf{A}_{\boldsymbol{\phi}} = \sum_Q \phi_j \mathbf{A}_j$, $\mathbf{A}_{\boldsymbol{\psi}} = \mathbf{A}_o + \mathbf{A}_{\boldsymbol{\phi}}$, and $v_{\boldsymbol{\phi}} = \sum_Q \phi_j c_j$, the *Lagrangian* of (1) is

$$\mathcal{L}\left(\boldsymbol{\phi}, \mathbf{x}\right) = 2\Re\left[\mathbf{s}^\dagger_{\boldsymbol{\psi}} \mathbf{x}\right] - \mathbf{x}^\dagger \mathbf{A}_{\boldsymbol{\psi}} \mathbf{x} + v_{\boldsymbol{\phi}}. \qquad (2)$$

The *dual* program of a QCCP $\mathtt{P}$ is

$$\mathtt{D}\left(\mathtt{P}\right) = \inf_{\boldsymbol{\phi} \in \Phi_{\mathtt{P}}} \max_{\mathbf{x} \in F_{\kappa}} \mathcal{L}\left(\boldsymbol{\phi}, \mathbf{x}\right) = \inf_{\boldsymbol{\phi} \in \Phi_{\mathtt{P}}} \mathcal{D}\left(\boldsymbol{\phi}\right), \qquad (3)$$

with $\mathcal{D}\left(\boldsymbol{\phi}\right) = \max_{\mathbf{x} \in F_{\kappa}} \mathcal{L}\left(\boldsymbol{\phi}, \mathbf{x}\right)$ the *dual* function of the Lagrangian, $\Phi_{\mathtt{P}} = \{\boldsymbol{\phi} \in \mathbb{H}^J \mid \mathbf{A}_{\boldsymbol{\phi}} \succeq 0\}$, and $\mathbb{H}$ the positive half space $\mathbb{H} = \{r \in \mathbb{R} \mid r \geq 0\}$—$\Phi_{\mathtt{P}}$ is the closure of $\{\boldsymbol{\phi} \in \mathbb{H}^J \mid \mathbf{A}_{\boldsymbol{\phi}} \succ 0\}$. Since the supremum of a collection of convex lower semi-continuous functions is both convex and lower semi-continuous, the dual function of any QCCP is convex and lower semi-continuous. A QCCP is *strongly dual* if $\mathtt{D}\left(\mathtt{P}\right) = \mathtt{P}$—the value of the dual program coincides with the value of the *primal* program. $\circledast$ superscripts are used to denote optimal points. For example, $\boldsymbol{\phi}_{\circledast}$ would be used to denote a minimizer of $\mathcal{D}\left(\boldsymbol{\phi}\right)$.

Note that the definition of $\mathcal{D}\left(\boldsymbol{\phi}\right)$ given above is unconventional: typically $\max_{\mathbf{x} \in F_{\kappa}}$ is replaced by $\sup_{\mathbf{x} \in \mathbb{C}^n}$. However, as along as a compact constraint exists, this alteration does not result in any (meaningful) practical differences. The convexity of $\Phi_{\mathtt{P}}$ implies that it possible to come arbitrarily close to the minimum of $\mathcal{D}\left(\boldsymbol{\phi}\right)$ along a path $p : \mathbb{I} \to \Phi_{\mathtt{P}}$, where $\mathbb{I}$ is the unit interval, such that $\mathbf{A}_{p(t)} \succ 0$. The Hessian of $\mathcal{D}\left(\boldsymbol{\phi}\right)$ is well defined along p, and if

$$\mathbf{x}_{\circledast} = \mathbf{A}^{-1}_{\boldsymbol{\psi}} \mathbf{s}_{\boldsymbol{\psi}} \notin F_{\boldsymbol{\kappa}},$$

there is a $\delta > 0$ such that increasing $\boldsymbol{\psi}$ along $\boldsymbol{\kappa}$ will result in $\mathbf{A}_{\boldsymbol{\psi} + \delta\boldsymbol{\kappa}} \succ \mathbf{A}_{\boldsymbol{\psi}}$ and $\mathcal{D}\left(\boldsymbol{\psi} + \delta\boldsymbol{\kappa}\right) < \mathcal{D}\left(\boldsymbol{\psi}\right)$ [2]. That is, if the standard dual is minimized with sufficiently fine precision, the restriction $\boldsymbol{x} \in F_{\boldsymbol{\kappa}}$ is immaterial. The $\max_{x \in F_{\boldsymbol{\kappa}}}$ alteration is technically helpful as it guarantees the continuity of the dual, which is used in is used in lemma 5, c.f. Appendix.

Augmenting the dimension of primal optimization variable $\mathbf{x}$ by one ($\mathbf{x} \to \tilde{\mathbf{x}}$), and introducing the auxiliary constraint $\tilde{x}^\dagger_{n+1} \tilde{x}_{n+1} = 1$, any quadratic function can be converted into the homogeneous form

$$f_j\left(\mathbf{x}\right) = \tilde{\mathbf{x}}^\dagger \begin{bmatrix} -\mathbf{A}_j & \mathbf{s}_j \\ \mathbf{s}^\dagger_j & v_j \end{bmatrix} \tilde{\mathbf{x}} = \tilde{\mathbf{x}}^\dagger \mathbf{H}_j \tilde{\mathbf{x}}. \qquad (4)$$

Accordingly, any QCCP $\mathtt{P}$ can be converted into a homogeneous QCCP—a QCCP wherein every imposed constraint is strictly quadratic. Via this connection, the dual of a QCCP is found to be equivalent to a semi-definite program relaxation, c.f. Appendix.

The remainder of the article investigates the application of the following result to QCCPs.

Sion's minimax theorem [3, 4]. Let K be a compact convex subset of a topological vector space, $\Phi_{\mathtt{P}}$ a convex subset of a topological vector space, and $\mathcal{L}$ a real valued function on $\Phi_{\mathtt{P}} \times K$. Suppose (a) that $\mathcal{L}(\#, \mathbf{x})$ is lower semicontinuous and quasi-convex on $\Phi_{\mathtt{P}}$ for each $\mathbf{x} \in F_\kappa$, and (b) that $\mathcal{L}(\boldsymbol{\phi}, \#)$ is upper semicontinuous and quasi-concave on K for each $\boldsymbol{\phi} \in \Phi_{\mathtt{P}}$. Then,

$$\max_{\mathbf{x}\in F_\kappa} \inf_{\boldsymbol{\phi}\in\Phi_{\mathtt{P}}} \mathcal{L}(\boldsymbol{\phi}, \mathbf{x}) = \inf_{\boldsymbol{\phi}\in\Phi_{\mathtt{P}}} \max_{\mathbf{x}\in F_\kappa} \mathcal{L}(\boldsymbol{\phi}, \mathbf{x}). \tag{5}$$

Mirroring the definition of $\mathcal{D}(\boldsymbol{\phi})$, we take

$$\mathcal{S}(\mathbf{x}) = \inf_{\boldsymbol{\phi}\in\Phi_{\mathtt{P}}} \mathcal{L}(\boldsymbol{\phi}, \mathbf{x}) \tag{6}$$

to be the *Sion* function of $\mathtt{P}$, and $\mathtt{S}(\mathtt{P}) = \max_{\boldsymbol{x}\in F_\kappa} \mathcal{S}(\mathbf{x})$ to be the Sion program of $\mathtt{P}$. Since the infimum of a collection of concave upper semi-continuous functions retains these properties, the Sion function of any QCCP is quasi-concave upper semi-continuous.

In discussing the behaviour of Sion functions it is often useful to consider the convex hull of $F_{\mathtt{P}}$—$C_{\mathtt{P}}$—and the Sion set $S_{\mathtt{P}} = \{\boldsymbol{x} \in F_{\boldsymbol{\kappa}} \mid \mathcal{S}(\boldsymbol{x}) \neq -\infty\}$. For example, simplifying to the case that f_o is simply linear, if $\boldsymbol{x} \in F_{\mathtt{P}}$ then $\mathcal{S}(\boldsymbol{x}) = f_o(\boldsymbol{x})$—every constraint is satisfied and so $\boldsymbol{\phi} = \mathbf{0}$ is the unique minimizer of $\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{x})$. Hence, it follows that $\mathcal{S}(\boldsymbol{x}) = f_o(\boldsymbol{x})$ if $\boldsymbol{x} \in C_{\mathtt{P}}$.

Lemmas

Lemma 1—If f_o is linear, then $\boldsymbol{y} \in C_{\mathtt{P}} \Rightarrow \mathcal{S}(\boldsymbol{y}) = f_o(\boldsymbol{y})$.

Corollary—$C_{\mathtt{P}} \subseteq S_{\mathtt{P}}$.

is Ja Jb a partition

t_j indexing

Proof of lemma 1. Let $\{\boldsymbol{x}_j\}_{j\in J} \subset F_{\mathtt{P}} \ni \boldsymbol{y} = \sum_{j\in J} t_k \boldsymbol{x}_j$, with $(\forall j \in J)\ 0 < t_j < 1$ and $\sum_{j\in J} t_j = 1$. Split J into J_a and J_b, so that $\boldsymbol{y}_a = \sum_{j\in J_a} t_k \boldsymbol{x}_j$, $\boldsymbol{y}_b = \sum_{j\in J_b} t_j \boldsymbol{x}_j$, $n_a = \sum_{j\in J_a} t_j$ and $n_b = \sum_{j\in J_b} t_j$. Since $\mathcal{S}$ is concave $\mathcal{S}(\boldsymbol{y}) \geq n_a\ \mathcal{S}(\boldsymbol{y}_a/n_a) + n_b\ \mathcal{S}(\boldsymbol{y}_b/n_b)$. Setting J_a equal to a single element, this result specializes to $\mathcal{S}(\boldsymbol{y}) \geq t_j\ \mathcal{S}(\boldsymbol{x}_j/n_a) + n_b\ \mathcal{S}(\boldsymbol{y}_b/n_b) = t_j f_o(\boldsymbol{x}_j) + n_b\ \mathcal{S}(\boldsymbol{y}_b/n_b)$. Therefore, $\mathcal{S}(\boldsymbol{y}) \geq \sum_{j\in J} t_j\ \mathcal{S}(\boldsymbol{x}_j) \geq \sum_{j\in J} t_j f_o(\boldsymbol{x}_j) = f_o(\boldsymbol{y})$. Because $\mathcal{S}(\boldsymbol{y}) \leq f_o(\boldsymbol{y})$, $\mathcal{S}(\boldsymbol{y}) = f_o(\boldsymbol{y})$.

convex hull assumption here?

The corollary follows as $\mathcal{S}(\boldsymbol{y}) \geq \sum_{j\in J} t_j\ \mathcal{S}(\boldsymbol{x}_j) \geq \sum_{j\in J} t_j f_o(\boldsymbol{x}_k)$ holds in generality. □

Rem.—The inclusion stated in the corollary of lemma 1 is typically strict, and differences between $S_{\mathtt{P}}$ and $C_{\mathtt{P}}$ are an important cause of duality gaps. In particular, the maximum of a linear function on $C_{\mathtt{P}}$ is equivalent to its maximum on $F_{\mathtt{P}}$, but its maximum on $S_{\mathtt{P}}$ may be greater.

Lemma 2—Eq. (5) applies to Eq.(1):

$$\inf_{\boldsymbol{\phi}\in\Phi_{\mathtt{P}}} \mathcal{D}(\boldsymbol{\phi}) = \max_{\mathbf{x}\in F_\kappa} \mathcal{S}(\boldsymbol{x}).$$

Proof of lemma 2. As $\boldsymbol{\phi} \in \Phi_{\mathtt{P}} \Rightarrow A_{\boldsymbol{\psi}} \succeq 0$, $\mathcal{L}(\boldsymbol{\phi}, \#) = \mathcal{L}_{\boldsymbol{\phi}}(\mathbf{x})$ is a concave continuous function on C, Ref. [5], that is bounded from above by $\mathbf{s}_{\boldsymbol{\psi}}^\dagger \mathbf{A}_{\boldsymbol{\psi}}^{-1} \mathbf{s}_{\boldsymbol{\psi}} + v_{\boldsymbol{\phi}}$. Suppose $\mathbf{x}, \mathbf{y} \in C$. For any $t \in (0,1)$

What is C

$$\begin{aligned}
&t\mathcal{L}_{\boldsymbol{\phi}}(\mathbf{x}) + (1-t)\,\mathcal{L}_{\boldsymbol{\phi}}(\mathbf{y}) = 2\Re\left[\mathbf{s}_{\boldsymbol{\psi}}^\dagger (t\mathbf{x} + (1-t)\,\mathbf{y})\right] - \\
&t\ \mathbf{x}^\dagger \mathbf{A}_{\boldsymbol{\psi}} \mathbf{x} - (1-t)\,\mathbf{y}^\dagger \mathbf{A}_{\boldsymbol{\psi}} \mathbf{y} + v_{\boldsymbol{\phi}} = \\
&\mathcal{L}_{\boldsymbol{\phi}}\left[t\mathbf{x} + (1-t)\,\mathbf{y}\right] - t\,(1-t)\,(\mathbf{x}-\mathbf{y})^\dagger\, \mathbf{A}_{\boldsymbol{\psi}}\,(\mathbf{x}-\mathbf{y}) \Rightarrow \\
&\mathcal{L}_{\boldsymbol{\phi}}\left[t\mathbf{x} + (1-t)\,\mathbf{y}\right] \geq t\mathcal{L}_{\boldsymbol{\phi}}(\mathbf{x}) + (1-t)\,\mathcal{L}_{\boldsymbol{\phi}}(\mathbf{y}) \quad \Rightarrow \\
&t\mathbf{x} + (1-t)\,\mathbf{y} \in C.
\end{aligned}$$

This is a subspace of H^J, not functions

$\Phi_{\mathtt{P}}$ convex as $\mathbf{A}_{\boldsymbol{\theta}} \succeq 0$ and $\mathbf{A}_{\boldsymbol{\phi}} \succeq 0$ imply

$$\mathbf{x}^\dagger \left[t\mathbf{A}_{\boldsymbol{\phi}} + (1-t)\,\mathbf{A}_{\boldsymbol{\theta}}\right]\mathbf{x} \geq 0$$

$\Rightarrow t\mathbf{A}_{\boldsymbol{\phi}} + (1-t)\,\mathbf{A}_{\boldsymbol{\theta}} \in \Phi_{\mathtt{P}}$. $\mathcal{L}(\#, \boldsymbol{x}) = \mathcal{L}_{\boldsymbol{x}}(\boldsymbol{\phi})$ is convex and continuous on $\Phi_{\mathtt{P}}$ since it is affine. □

Lemma 3—Let $\boldsymbol{\phi}_{\circledast}$ minimize $\inf_{\boldsymbol{\phi}\in\Phi_{\mathtt{P}}} \max_{\mathbf{x}\in F_\kappa} \mathcal{L}(\boldsymbol{\phi}, \mathbf{x})$, and take $\mathbf{x}_{\circledast} = \mathbf{A}_{\boldsymbol{\psi}_{\circledast}}^{-1} \mathbf{s}_{\boldsymbol{\psi}_{\circledast}}$—with $\mathbf{A}_{\boldsymbol{\psi}_{\circledast}}^{-1}$ the pseudo inverse of $\mathbf{A}_{\boldsymbol{\psi}_{\circledast}}$ in the event that $\mathbf{A}_{\boldsymbol{\psi}_{\circledast}} \not\succ 0$. If $\tilde{\mathbf{x}}_{\circledast}$ is a solution of $\max_{\mathbf{x}\in F_\kappa} \inf_{\boldsymbol{\phi}\in\Phi_{\mathtt{P}}} \mathcal{L}(\boldsymbol{\phi}, \mathbf{x})$, then there exists $\mathbf{k} \in \ker \mathbf{A}_{\boldsymbol{\phi}_{\circledast}} \ni \tilde{\mathbf{x}}_{\circledast} = \mathbf{x}_{\circledast} + \mathbf{k}$.

Corollary—If $\mathbf{x}_{\circledast a}$ and $\mathbf{x}_{\circledast b}$ are associated with the dual minimizers $\boldsymbol{\phi}_{\circledast a}$ and $\boldsymbol{\phi}_{\circledast b}$ as above, then $\mathbf{x}_{\circledast a} = \mathbf{x}_{\circledast b}$ on the complement of $\ker \mathbf{A}_{\boldsymbol{\psi}_{\circledast a}} \cup \ker \mathbf{A}_{\boldsymbol{\psi}_{\circledast b}}$. If $\{\boldsymbol{\phi}_{\circledast j}\}_{j\in J}$ is a collection of dual minimizers, then two solutions of $\max_{\mathbf{x}\in F_\kappa} \inf_{\boldsymbol{\phi}\in\Phi_{\mathtt{P}}} \mathcal{L}(\boldsymbol{\phi}, \mathbf{x})$ can differ only an element of

$$\bigcap_{j\in J} \ker \mathbf{A}_{\boldsymbol{\psi}_{\circledast j}}.$$

Proof of lemma 3. Let $\boldsymbol{\phi}_{\circledast}$ be the minimizer of $\inf_{\boldsymbol{\phi}\in\Phi_{\mathtt{P}}} \mathcal{D}(\boldsymbol{\phi})$ such that $\mathbf{A}_{\boldsymbol{\psi}_{\circledast}} \succ 0$. $\mathbf{x}_{\circledast}$ as defined above is then the unique solution of $\max_{\mathbf{x}\in F_\kappa} \mathcal{L}(\boldsymbol{\phi}_{\circledast}, \mathbf{x})$ as determined by the vanishing of the gradient in $\mathbf{x}$. By Sion's minimax theorem, there is then a $\tilde{\mathbf{x}}_{\circledast} \in C$ such that $\inf_{\boldsymbol{\phi}\in\Phi_{\mathtt{P}}} \mathcal{L}(\boldsymbol{\phi}, \tilde{\mathbf{x}}_{\circledast}) = \mathcal{L}(\boldsymbol{\phi}_{\circledast}, \mathbf{x}_{\circledast})$. Because $\mathbf{x}_{\circledast}$ is selected by $\max_{\mathbf{x}\in\mathtt{C}} \mathcal{L}(\boldsymbol{\phi}_{\circledast}, \mathbf{x})$,

Negative semi-def A is not discussed

$$\mathcal{L}(\boldsymbol{\phi}_{\circledast}, \mathbf{x}_{\circledast}) = \inf_{\boldsymbol{\phi}\in\Phi_{\mathtt{P}}} \mathcal{L}(\boldsymbol{\phi}, \tilde{\mathbf{x}}_{\circledast}) \leq \mathcal{L}(\boldsymbol{\phi}_{\circledast}, \tilde{\mathbf{x}}_{\circledast}) \leq \mathcal{L}(\boldsymbol{\phi}_{\circledast}, \mathbf{x}_{\circledast}).$$

Therefore, $\mathcal{L}(\boldsymbol{\phi}_{\circledast}, \tilde{\mathbf{x}}_{\circledast}) = \mathcal{L}(\boldsymbol{\phi}_{\circledast}, \mathbf{x}_{\circledast}) \Rightarrow \tilde{\mathbf{x}}_{\circledast} = \mathbf{x}_{\circledast} + \mathbf{k}$ with $\mathbf{k} \in \ker \mathbf{A}_{\boldsymbol{\phi}_{\circledast}}$.

The corollary follows as the same argument applies to any dual minimizer and Sion maximizer. □

Lemma 4—If $\mathbf{x}_{\circledast}$ is maximizes $\max_{\mathbf{x}\in F_\kappa} \inf_{\boldsymbol{\phi}\in\Phi_{\mathtt{P}}} \mathcal{L}(\boldsymbol{\phi}, \mathbf{x})$, and $f_{\boldsymbol{\gamma}}$ is a positive definite equality constraint such that $f_{\boldsymbol{\gamma}}(\mathbf{x}_{\circledast}) = 0$, then $f_j(\mathbf{x}_{\circledast}) \geq 0$ for all $j \in J$.

Proof of lemma 4. Suppose that $f_j(\mathbf{x}_{\circledast}) = -k < 0$. There is then a $\delta > 0$ such that $f_j(\mathbf{x}_{\circledast}) + \delta f_{\boldsymbol{\gamma}} = -k$ and $\mathbf{A}_j + \delta \mathbf{A}_{\boldsymbol{\gamma}} \succeq 0$. Let $\boldsymbol{\zeta} = \phi_j + \boldsymbol{\gamma} \in \Phi_{\mathtt{P}}$, and set "only" here?

$m = \Re\left[\boldsymbol{s}_o^\dagger \mathbf{x}_\circledast\right]$. Then, for any $n > m$, $\mathcal{L}\left(n\boldsymbol{\zeta}/\kappa, \mathbf{x}_\circledast\right) < 0$, contradicting the assumption that $\mathtt{P}$ possesses feasible solutions with objective values greater than zero. $\square$

Rem.—Lemma 4 restricts the possible constraint violation of an optimal $\boldsymbol{x}_\circledast$ based on the tightness of any positive definite constraint. Suppose that $f_j\left(\boldsymbol{x}_\circledast\right) = -v$ and $f_\gamma\left(\boldsymbol{x}_\circledast\right) = u$, with $\mathbf{A}_{\boldsymbol{\gamma}} \succ 0$ and $u, v > 0$. Take n to be the most negative eigenvalue of $\mathbf{A}_j$ and m to be the smallest eigenvalue of $\mathbf{A}_\gamma$. Because $f_j + |n/m|\, f_\gamma$ is then a positive semi-definite constraint, the definition of $\boldsymbol{x}_\circledast$ requires that $|n/m|\, u \geq v$. This same argument also implies that $S_{\mathtt{P}}$ coincides with $F_{\mathtt{P}}$ on ∂F_κ.

Ambiguous notation

Lemma 5—Let $\Phi_{\mathtt{P}_1} = \{\boldsymbol{\phi} \in \Phi_{\mathtt{P}} \mid \|\boldsymbol{\phi}\| = 1\}$. If there exists a $\delta > 0$ such that $\forall \boldsymbol{\phi}_1 \in \Phi_{\mathtt{P}_1}$

$$\max_{\boldsymbol{x} \in C} f_{\boldsymbol{\phi}_1}\left(\mathbf{x}\right) > \delta,$$

there is a finite number $b > 0$ such that, letting $\Phi_{\mathtt{P}_b} = \{\boldsymbol{\phi} \in \Phi_{\mathtt{P}} \mid \|\boldsymbol{\phi}\| \leq b\}$, $\inf_{\boldsymbol{\phi} \in \Phi_{\mathtt{P}}} \mathcal{D}\left(\boldsymbol{\phi}\right) = \min_{\boldsymbol{\phi} \in \Phi_{\mathtt{P}_b}} \mathcal{D}\left(\boldsymbol{\phi}\right)$.

Corollary—Under the conditions of the lemma 5,

$$\max_{\mathbf{x} \in F_\kappa} \min_{\boldsymbol{\phi} \in \Phi_{\mathtt{P}_b}} \mathcal{L}\left(\boldsymbol{\phi}, \mathbf{x}\right) = \max_{\mathbf{x} \in F_\kappa} \inf_{\boldsymbol{\phi} \in \Phi_{\mathtt{P}}} \mathcal{L}\left(\boldsymbol{\phi}, \mathbf{x}\right). \qquad (7)$$

That is, letting

$$\mathcal{S}_b\left(\boldsymbol{x}\right) = \min_{\boldsymbol{\phi} \in \Phi_{\mathtt{P}_b}} \mathcal{L}\left(\boldsymbol{\phi}, \mathbf{x}\right),$$

$\max_{x \in C} \mathcal{S}\left(\boldsymbol{x}\right) = \max_{x \in C} \mathcal{S}_b\left(\boldsymbol{x}\right)$ with $\mathcal{S}_b\left(\boldsymbol{x}\right)$ continuous and concave.

typo?

Proof of lemma 5. Define $m > 0$ via the equality $f_{\boldsymbol{\kappa}}\left(m\mathbf{s}_o\right) = 0$. $2\Re\left[\boldsymbol{s}_o^\dagger \mathbf{x}\right]$ is then bounded from below over C by $-l = -2m\|\boldsymbol{s}_o\|^2$. As $\mathbf{0} \in \Phi_{\mathtt{P}}$, and $\Phi_{\mathtt{P}}$ is convex, if $\boldsymbol{\phi} \in \Phi_{\mathtt{P}}$ then $\boldsymbol{\phi}/\|\boldsymbol{\phi}\| \in \Phi_{\mathtt{P}_1}$. Set $u = \max_{\boldsymbol{x} \in C} 2 + f_{\boldsymbol{\kappa}}\left(\mathbf{x}\right)$, and $b = \left(u + l\right)/\delta$. If $\|\boldsymbol{\phi}\| \geq b$, then $\max_x \mathcal{D}\left(\boldsymbol{\phi}\right) \geq u$. Therefore, $\inf_{\boldsymbol{\phi} \in \Phi_{\mathtt{P}}} \mathcal{D}\left(\boldsymbol{\phi}\right) = \min_{\boldsymbol{\phi} \in \Phi_{\mathtt{P}_b}} \mathcal{D}\left(\boldsymbol{\phi}\right)$.

Because $\Phi_{\mathtt{P}_b}$ is convex, $\Phi_{\mathtt{P}_b}$ can be substituted for $\Phi_{\mathtt{P}}$ in lemma 2 without alteration. Since $\Phi_{\mathtt{P}_b}$ is closed and bounded, it is compact, and so there $\exists \boldsymbol{\phi}_\circledast \in \Phi_{\mathtt{P}_b}$ such that $\mathcal{D}\left(\boldsymbol{\phi}_\circledast\right) = \inf_{\boldsymbol{\phi} \in \Phi_{\mathtt{P}}} \mathcal{D}\left(\boldsymbol{\phi}\right)$. $\mathcal{S}_b\left(\boldsymbol{x}\right)$ is continuous by lemma A (Appendix). $\square$

Rem.—The conditions of Lemma 5 are satisfied, for example, if $\boldsymbol{\phi} \in \Phi_{\mathtt{P}_1} \Rightarrow v_{\boldsymbol{\phi}} > 0 \vee \|\mathbf{s}_{\boldsymbol{\phi}}\| > 0$. This follows from the fact that $\Phi_{\mathtt{P}_1}$ is compact, $\boldsymbol{\phi} \in \Phi_{\mathtt{P}_1}$ implies that the maximum eigenvalue of $\mathbf{A}_{\boldsymbol{\phi}}$ is bounded, and that $v_{\boldsymbol{\phi}}$ and $\|\mathbf{s}_{\boldsymbol{\phi}}\|$ are continuous functions of $\boldsymbol{\phi}$.

Supposing that $\mathtt{P}$ is a QCCP in standard form, see appendix, let τ be a global relaxation parameter for all constraints but the (compact) unit ball, $f_j\left(\mathbf{x}\right) \geq 0 \to f_j\left(\mathbf{x}\right) \geq -\tau$. Define $\mathtt{P}\left(\tau\right)$, and all other related program objects like $\mathtt{D}\left(\mathtt{P}\left(\tau\right)\right)$ and $S_{\mathtt{P}}\left(\tau\right)$, to result from the τ relaxed constraint set.

Lemma 6—If $\mathtt{P}$ is a QCCP in standard form, the intersection of $S_{\mathtt{P}}\left(\tau\right)$ with the boundary of the unit ball in $\mathbb{C}^n$ is path connected $\forall \tau \geq 0$.

Proof. Take $\mathbf{H}_j$ to be the operator define the j^{th} constraint of $\mathtt{P}$. Let $\mathbf{z} \in \mathbb{C}^n \ni \|\mathbf{z}\| = 1$. Recalling that $\mathbf{H}_j$ is Hermitian, it may be freely assumed that $\mathbf{H}_j$ is diagonal. If $\mathbf{z}^\dagger \mathbf{H}_j \mathbf{z} \geq \tau$, there must be a (unit normalized) eigenvector $\mathbf{h}_\alpha$ of $\mathbf{H}_j$ such that the eigenvalue of $\mathbf{h}_\alpha$ is greater or equal to $-\tau$ and $\mathbf{h}_\alpha^\dagger \mathbf{z} \neq \mathbf{0}$. Making use of the phase freedom of $\mathbb{C}$, by simultaneously increasing the magnitude of the projection of $\mathbf{z}$ along $\mathbf{h}_\alpha$ and decreasing the magnitude of the projection of $\mathbf{z}$ along any eigenvector $\mathbf{h}_\beta$ of $\mathbf{H}_j$ with eigenvalue less than $-\tau$, a path connecting $\mathbf{h}_\alpha$ and $\mathbf{z}$ can be traced through the intersection of $S_{\mathtt{P}}\left(\tau\right)$ and the boundary of the unit ball. An analogous argument implies that the intersection of the boundary of the ball and the subspace spanned by all eigenvectors of $\mathbf{H}_j$ with eigenvalues greater than $-\tau$ is path connected. As $\mathtt{P}$ is assumed to be feasible, $F_j^\partial\left(\tau\right) = \left\{\mathbf{z} \in \mathbb{C}^n \mid \|\mathbf{z}\| = 1 \wedge \mathbf{z}^\dagger \mathbf{H}_j \mathbf{z} \geq -\tau\right\} \neq \emptyset$ and $F_j^\partial \cap F_k^\partial \neq \emptyset$. Therefore, the intersection of $F_{\mathtt{P}}\left(\tau\right)$ with the boundary of the unit ball is path connected. $S_{\mathtt{P}}\left(\tau\right)$ coincides with $F_{\mathtt{P}}\left(\tau\right)$ on the boundary of the unit ball by the proof of lemma 4. $\square$

Discuss case the boundary of $S_{\mathtt{P}}$ is determined by a finite number of sets $\to$ relation to S-lemma.

§PROPOSAL

The results presented in the previous section are assuredly only a small part of what can be understood about duality via Sion's theorem. Nevertheless, as we now aim to show, they are already a useful guide for QCCP heuristics. For simplicity we focus on the specialization to linear objectives, i.e. compact quadratically constrained linear programs (compact QCLPs). We believe that much of what we propose can be adapted to larger problem classes. However, the exact alterations that should be made are not evident, c.f. Appendix. The algorithm presented in the next subsection is based on three main ideas.

(a) In a compact QCLP, the Sion set $S_{\mathtt{P}}$ is determined wholly by the imposed constraints. Therefore, given a feasible $\mathtt{P}$, it is always possible to modify the objective vector $\boldsymbol{s}_o$ in such a way that the problem becomes strongly dual.

(b) The smaller the separation between the dual solution $\mathbf{x}_\circledast$ and the boundary ∂F_κ is, the closer strong duality is to being achieved, lemma 4. Hence, supposing

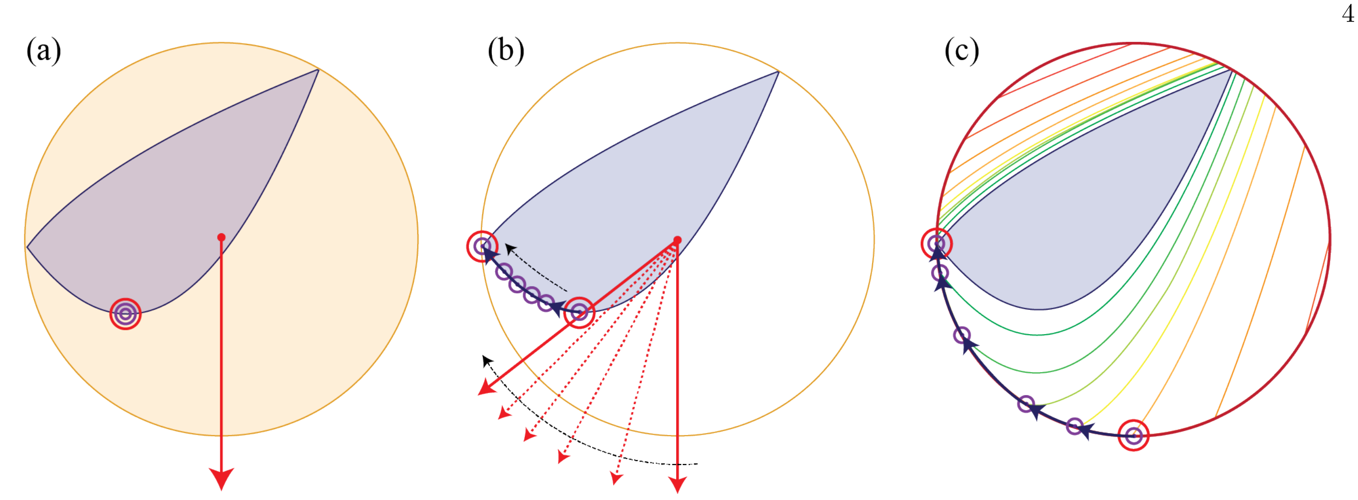

FIG. 1. **Schematic example of protocols.** The figure illustrates the mechanics of the *scrape* and *deflate* protocols, defined in the proposed algorithm, for a simple linear program $\mathtt{P}$ in $\mathbb{R}^2$ subject to the constraints $4+4x_a-3x_b-4x_b^2=0$ and $1-x_a^2-x_b^2=0$ (orange disk). In this case the Sion set, the tear-drop shaped purple region, is simple to compute as every admissible constraint is effectively describe by a single Lagrange multiplier value. (a) Supposing an objective vector of $\mathbf{s}_o=-\langle 0,1\rangle$, $\max_{\mathbf{x}\in F_\kappa}\mathcal{S}(\mathbf{x})$ occurs at $\mathbf{x}=\langle 0,-1/3\rangle$—the point emphasized by the bullseye. Consequently there is a duality gap of $1/3$ compared to $\mathtt{P}$: $\mathtt{P}$ has two feasible points, the intersection points of the Sion set with the orange circle, with the best objective value being achieved at $\mathbf{x}=\langle 0,0\rangle$. (b) In the middle panel, the location of the $\mathbf{x}_\circledast$ solution of $\mathtt{D}(\mathtt{P}(0,\mathbf{y}))$ is shown for five (scrape) applications of $\mathbf{y}^{(n)}=\mathbf{y}^{(o)}/2+\mathbf{x}_\circledast^{(o)}/2$. As the objective vector drifts, $\mathbf{x}_\circledast$ is observed to move towards the boundary of the orange disk. After the fifth modification of $\mathbf{y}$ strong duality is achieved. (c) With $\tau=5$—giving the constraint set $4+4x_a-3x_b-4x_b^2\geq-\tau$, $-4-4x_a+3x_b+4x_b^2\geq-\tau$, and $1-x_a^2-x_b^2=0$—$S_{\mathtt{P}}(\tau)$ coincides with the unit disk. The rightmost panel shows how deflation and redefinition of the objective vector, $\mathbf{y}^{(n)}=\mathbf{x}_\circledast^{(o)}$, leads to $\mathbf{x}_\circledast$ tracking the intersection of $S_{\mathtt{P}}(\tau)$ with the circle.

a standard form in which the compact constraint F_κ is a ball, a reasonable criterion for modifying $\boldsymbol{s}_o$ is to ensure that $\mathbf{x}_\circledast$ remains near the boundary, i.e. $\|\mathbf{x}_\circledast\|\approx 1$.

(c) As the (inequality) constraints of a compact QCLP are relaxed $S_{\mathtt{P}}$ approaches the convex hull $C_{\mathtt{P}}$, lemma 1 and lemma 4, and on $C_{\mathtt{P}}$ strong duality holds. Consequently, by introducing a global relaxation parameter τ on all constraints but the ball, any compact QCLP can be relaxed into a strongly dual form.

Our heuristic combines these elements by alternating between the (b) and (c) approaches to strong duality through a sequence of relaxed sub-programs. Pictorially, Fig. X, a hierarchy of "inflated" Sion sets are used to "scrape" the objective towards a strongly dual direction.

Algorithm

Take $\mathtt{P}$ to be a feasible compact QCLP, which is not strongly dual, converted into standard form (see Appendix). Let δ be an acceptable constraint tolerance. Define $\mathtt{P}(\tau,\mathbf{y})$ to be the program resulting from $\mathtt{P}$ by asserting the relaxation parameter τ and the objective vector $\mathbf{y}$, and use g, γ and σ to respectively denote reduction and scraping parameters in the unit interval, e.g. $g=1/2$ and $\gamma=\sigma=1/10$. In the context of variable updates, let (n), (o), and (δ) superscripts refer to new, old, and difference values respectively. For example, $\mathbf{x}_\circledast^{(n)}=\mathbf{x}_\circledast^{(o)}+\mathbf{x}_\circledast^{(\delta)}$. Recall that, by lemma 4, $\|\mathbf{x}_\circledast\|=1$ implies strong duality.

1. *Inflate*—Set $\mathbf{y}=\mathbf{s}_o$. Increase τ until solving $\mathtt{D}(\mathtt{P}(\tau,\mathbf{y}))$ yields $\|\mathbf{x}_\circledast\|\geq 1-\delta$.

2. *Deflate*—Decrease τ by some small amount $\tau^{(\delta)}$—in practice linked to $\mathbf{x}_\circledast^{(\delta)}$. Solve $\mathtt{D}\left(\mathtt{P}\left(\tau^{(n)},\mathbf{y}\right)\right)$. If $\tau<\delta$ and $\|\mathbf{x}_\circledast\|>1-\delta$, $\mathbf{x}_\circledast$ is a heuristic for $\mathtt{P}$.

3. *Scrape*—(Success) If $\|\mathbf{x}_\circledast\|\geq 1-\delta$, set

$$\mathbf{y}^{(n)}=(1-\gamma)\,\mathbf{y}^{(o)}+\gamma\mathbf{x}_\circledast^{(o)}$$

and return to 2. (Failure) Otherwise, replace τ^δ with $g\tau^\delta$ and resolve $\mathtt{D}\left(\mathtt{P}\left(\tau^{(n)},\mathbf{y}\right)\right)$. Repeat until either $\|\mathbf{x}_\circledast\|\geq 1-\delta$, or a predefined reduction limit is reached. In the second case, proceed to sub-protocol (a).

(a) Let $\mathbf{y}^{(n)}=\mathbf{y}^{(o)}+\sigma\mathbf{x}_\circledast^{(o)}$ and solve $\mathtt{D}\left(\mathtt{P}\left(\tau,\mathbf{y}^{(n)}\right)\right)$. Repeat until either $\|\mathbf{x}_\circledast\|>1-\delta$, or $\mathbf{x}_\circledast^{(n)}=\mathbf{x}_\circledast^{(o)}$. In the first case take $\mathbf{y}^{(n)}=(1-\gamma)\,\mathbf{y}^{(o)}+\gamma\mathbf{x}_\circledast^{(o)}$ and return to 2.; in the second move to sub-protocol (b).

(b) Reset τ and $\mathbf{y}$ to the respective values observed after the last successful scrape step. Draw a random vector $\mathbf{q}\in\mathbb{C}^n$. Replace $\mathbf{y}$ with $\mathbf{y}+\epsilon\mathbf{q}$, where $\epsilon\ll\|\mathbf{y}\|/\|\mathbf{q}\|$. Solve $\mathtt{D}\left(\mathtt{P}\left(\tau^{(o)},\mathbf{y}^{(n)}\right)\right)$. If $\|\mathbf{x}_\circledast\|\geq 1-\delta$ move to 2.

(Success). If not, increase ϵ and repeat sub-protocol (b).

In carrying out the deflate and scrape steps, with idealized precision and control parameters, the protocol tracks the intersection of the boundary of the ball with the boundary of $S_{\mathtt{P}}$ nearest the direction defined by the original objective vector. The γ parameter is used to balance between this goal and the computational expedient of maintaining strong duality without resorting to the corrective (a) and (b) sub-protocols. Effectively, $\gamma > 0$ causes objective vector to creep towards the current solution, allowing for faster deflation. In contrast, by setting $\gamma = 0$, deviation between $\mathbf{y}$ and $\mathbf{s}_o$ is forestalled for as long as possible. Our belief is that the $\gamma = 0$ choice generally leads to greater fidelity between $\mathbf{y}$ and $\mathbf{s}_o$, but requires great numerical precision as some deflate step will eventually fail.

In the event of a loss of contact with the compact boundary, sub-protocol (a) increases the magnitude of $\mathbf{x}_{\circledast}$ by performing additional scraping steps. Because $\Re\left[\mathbf{y}^{(n)\dagger}\mathbf{x}_{\circledast}^{(n)}\right] \geq \Re\left[\mathbf{y}^{(n)\dagger}\mathbf{x}_{\circledast}^{(o)}\right]$, $\Re\left[\mathbf{y}^{(n)\dagger}\mathbf{x}_{\circledast}^{(\delta)}\right] \geq 0$. Because $\mathbf{x}_{\circledast}^{(o)}$ has the largest real projection along $\mathbf{y}^{(o)}$ of any point in $S_{\mathtt{P}}(\tau)$, $\Re\left[\mathbf{y}^{(o)\dagger}\mathbf{x}_{\circledast}^{(\delta)}\right] \leq 0$. Therefore, $\Re\left[\mathbf{y}^{(\delta)\dagger}\mathbf{x}_{\circledast}^{(\delta)}\right] \geq 0$ and

$$\begin{aligned}\|\mathbf{x}_{\circledast}^{(n)}\|^2 &= \|\mathbf{x}_{\circledast}^{(o)}\|^2 + 2\Re\left[\mathbf{x}_{\circledast}^{(o)\dagger}\mathbf{x}_{\circledast}^{(\delta)}\right] + \|\mathbf{x}_{\circledast}^{(\delta)}\|^2 \\ &= \|\mathbf{x}_{\circledast}^{(o)}\|^2 + 2\Re\left[\mathbf{y}^{(\delta)\dagger}\mathbf{x}_{\circledast}^{(\delta)}\right]/\gamma + \|\mathbf{x}_{\circledast}^{(\delta)}\|^2 \geq \|\mathbf{x}_{\circledast}^{(o)}\|^2.\end{aligned}$$

The second (more serious) failure mode alluded to by sub-protocol (b) corresponds to the seemingly improbable event that in decreasing τ the boundary of $S_{\mathtt{P}}(\tau)$, as it first crosses into the ball at $\mathbf{y}$, is exactly perpendicular to $\mathbf{y}$. For such cases, $S_{\mathtt{P}}(\tau)$ does not clearly indicate a preferred direction of movement, and so we resort to random choice. If certain feasible points are known, sub-protocol (b) may be partially informed by the objective value at the best known feasible point through the addition of a linear projection constraints that is not relaxed by τ: $\Re\left[\mathbf{s}_o^{\dagger}\mathbf{x}\right] \geq m$, with m the best known feasible value of the objective.

Reviewing Fig. 1, one may also reasonably worry about the possible formation of islands in the intersection of $S_{\mathtt{P}}$ and the boundary of the compact constraint ∂F_{κ}. That is, could one ever encounter the (unrecoverable) situation in which $\mathbf{x}_{\circledast}$ becomes trapped on a region of the boundary that vanishes with deflation? Lemma 6 rules out this possibility: $S_{\mathtt{P}}(\tau) \cap \partial F_{\kappa}$ is connected $\forall \tau \geq 0$. The appearance of multiple boundary components in $S_{\mathtt{P}}(\tau) \cap \partial F_{\kappa}$, as seen in Fig. 1, can occur only for "QCCPs" defined over $\mathbb{R}$ (instead of $\mathbb{C}$).

We do not believe that it is possible to assert that an algorithm for compact QCLPs is optimal in any meaningful way CITE. Nevertheless, strong duality ensures that the $\mathbf{x}_{\circledast}$ solution of $\mathtt{D}(\mathtt{P}(\tau, \mathbf{y}))$ is a point in $S_{\mathtt{P}}(\tau)$ with maximal real projection along $\mathbf{y}$. Hence, when a redefinition of $\mathbf{y}$ is needed to maintain strong duality, absent additional information, shifting towards $\mathbf{x}_{\circledast}$ is an obvious choice. Much like gradient descent, the guiding intuition of the proposed algorithm is that stringing together a series of (approximately) "locally" optimal choices will lead to a "nearly optimal" heuristic. However, because each of these "local" choices is made from the dual perspective, certain global characteristic of $\mathtt{P}$ are continually present.

§RESULTS

Discussion

§APPENDIX

Continuity Lemma

Lemma A—If $f : X \times C \to \mathbb{R}$ is continuous, and C is compact, then $g : X \to \mathbb{R},\ x \mapsto \sup_{c \in C} f\langle x, c\rangle$ and $h : X \to \mathbb{R},\ x \mapsto \inf_{c \in C} f\langle x, c\rangle$ are continuous.

Proof of lemma A. Recalling that continuity implies point continuity, about each $\langle x, c\rangle \in X$, $\forall \delta > 0$, there is a product neighbourhood $N_c^{\langle x,c\rangle} \times M_c^{\langle x,c\rangle}$ such that $f\left[N_x^{\langle x,c\rangle} \times M_c^{\langle x,c\rangle}\right] \subseteq (f\langle x,c\rangle - \delta, f\langle x,c\rangle + \delta)$. A finite collection of $M_c^{\langle x,c\rangle}$ subsets, indexed by $F_x = \{c_i\}$, cover C for each $x \in X$. Take $N_x = \cap_{c_i \in F_x} N_x^{\langle x,c_i\rangle}$. $\forall c \in C\ f[N_x \times \{c\}] \subseteq (f\langle x,c\rangle - \delta, f\langle x,c\rangle + \delta)$. Via the product topology, $f_x : C \to \mathbb{R},\ c \mapsto f\langle x,c\rangle$, is continuous. Hence, for any given $x \in X$, there is a $c_u \in C \ni g(x) = f\langle x, c_u\rangle$. As such, $g[N_x] \subseteq (g(x) - \delta, g(x) + \delta)$, and g is continuous. For any continuous $f : X \times C \to \mathbb{R}$, $\inf_{c \in C} f\langle x,c\rangle = \sup_{c \in C} -f\langle x,c\rangle$. $\square$

Schur Complement

Recall that if a block matrix is either upper or lower diagonal then its determinant is given by the product of the determinants of its main diagonal blocks. That is, so long as $\mathbf{A}$ is invertible, the expansion

$$\begin{bmatrix}\mathbf{A} & \mathbf{C}\\ \mathbf{B} & \mathbf{D}\end{bmatrix} = \begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{B}\mathbf{A}^{-1} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{A} & \mathbf{C}\\ \mathbf{0} & \mathbf{D}-\mathbf{B}\mathbf{A}^{-1}\mathbf{C}\end{bmatrix} = \begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{B}\mathbf{A}^{-1} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}-\mathbf{B}\mathbf{A}^{-1}\mathbf{C}\end{bmatrix}\begin{bmatrix}\mathbf{I} & \mathbf{A}^{-1}\mathbf{C}\\ \mathbf{0} & \mathbf{I}\end{bmatrix} \tag{8}$$

indicates that the total matrix is positive semi-definite iff $\mathbf{A} \succ 0$ and $\mathbf{D} - \mathbf{B}\mathbf{A}^{-1}\mathbf{C} \succ 0$. Due to this critical property, $\mathbf{D} - \mathbf{B}\mathbf{A}^{-1}\mathbf{C}$ is referred to as the Schur complement of $\mathbf{A}$. An analogous expansion can be carried out if $\mathbf{D}$ is invertible, showing that the block matrix is positive definite iff $\mathbf{D} \succ 0$ and $\mathbf{A} - \mathbf{C}\mathbf{D}^{-1}\mathbf{B} \succ 0$.

Equivalence with Semi-Definite Programs

The dual, in the wider setting of quadratically constrained quadratic programs (QCQPs), is equivalent to the semi-definite program relaxation. Taking $f_o(\boldsymbol{x})$ in Eq. (1) to be a general quadratic function, this equivalence can be realized as follows. Using the homogeneous form mentioned at the beginning of the text and the bilinear correspondence with quadratics, any quadratic function $f_j(\mathbf{x})$ be rewritten as

$$\operatorname{tr}\left[\begin{bmatrix} -\mathbf{A}_j & \mathbf{s}_j \\ \mathbf{s}_j^\dagger & v_j \end{bmatrix}\begin{bmatrix} \mathbf{x}\mathbf{x}^\dagger & \mathbf{x} \\ \mathbf{x}^\dagger & 1 \end{bmatrix}\right] \geq 0. \tag{9}$$

Letting $\mathbf{H}_k$ denote the block matrix associated with a quadratic function in the above manner, by increasing the dimension of $\mathbf{x}$ by one, $\mathbf{x} \to \tilde{\mathbf{x}}$, any QCQP $\mathtt{P}$ can be placed in the homogeneous form

$$\begin{aligned} &\max_{\tilde{\mathbf{x}} \in \mathbb{C}^{n+1}} \operatorname{tr}[\mathbf{H}_o \mathbf{X}] \qquad (10) \\ &\ni (\forall j \in J)\ \operatorname{tr}[\mathbf{H}_j \mathbf{X}] \geq 0 \\ &\wedge \tilde{x}_{n+1}^2 = 1, \end{aligned}$$

where $\mathbf{X} = \tilde{\mathbf{x}}\tilde{\mathbf{x}}^\dagger$—a dummy variable has been used to increase the dimension of the problem by one, if $\tilde{\mathbf{x}}$ is a solution of Eq. (10), then $\mathbf{x} = \langle \tilde{x}_1/\tilde{x}_{n+1}, \tilde{x}_2/\tilde{x}_{n+1}, \ldots, 1\rangle$ is a solution of the original QCQP. The dual of Eq. (10), using $\mathbf{x}_{\circledast} = \mathbf{A}_{\boldsymbol{\psi}}^{-1}\mathbf{s}_{\boldsymbol{\psi}}$ and $\boldsymbol{\psi}$ to denote the inclusion of the objective in the $\boldsymbol{\phi}$ sum notation used in the main text, is

$$\inf_{\boldsymbol{\phi} \in \Phi_{\mathtt{P}}} \mathbf{s}_{\boldsymbol{\psi}}^\dagger \mathbf{A}_{\boldsymbol{\psi}}^{-1} \mathbf{s}_{\boldsymbol{\psi}} + v_{\boldsymbol{\psi}}.$$

Via the Schur complement construction, the dual of a QCQP is thus equivalent to

$$\inf_{\alpha \in \mathbb{R}, \boldsymbol{\phi} \in \mathbb{H}^J} \alpha \ni \begin{bmatrix} -\mathbf{A}_{\boldsymbol{\psi}} & \mathbf{s}_{\boldsymbol{\psi}} \\ \mathbf{s}_{\boldsymbol{\psi}}^\dagger & v_{\boldsymbol{\psi}} - \alpha \end{bmatrix} \preceq 0. \tag{11}$$

The constraint that the block matrix appearing in Eq.(11), $\mathbf{H}_{\boldsymbol{\psi}}^{\alpha}$, is negative semi-definite is equivalent to checking that $\mathbf{b}_l^\dagger \mathbf{H}_{\boldsymbol{\psi}}^{\alpha} \mathbf{b}_l \leq 0$ for a collection of basis vectors $\{\mathbf{b}_l\}_{l \in L}$ spanning $\mathbb{C}^{n+1}$. Since the sum of a collection of such constraints can be represented by a positive semi-definite matrix $\mathbf{B}$, the Lagrangian of Eq. (11) is

$$\mathcal{L}(\alpha, \boldsymbol{\phi}, \mathbf{B}) = \alpha(1 - \mathbf{B}_{n+1,n+1}) + \operatorname{tr}[(\mathbf{H}_o + \mathbf{H}_{\boldsymbol{\phi}})\mathbf{B}].$$

If either $\mathbf{B}_{n+1,n+1} \neq 1$ or $\operatorname{tr}[\mathbf{H}_{\boldsymbol{\phi}}\mathbf{B}] < 0$, then the infimum of this form over α and $\boldsymbol{\phi}$ is $-\infty$; if $\mathbf{B}_{n+1,n+1} = 1$ and $\operatorname{tr}[\mathbf{H}_{\boldsymbol{\phi}}\mathbf{B}] \geq 0$ then the infimum is $\operatorname{tr}[\mathbf{H}_o\mathbf{B}]$. The dual of Eq. (11), accordingly, is

$$\begin{aligned} &\max_{\mathbf{B} \in \mathbb{C}^{(n+1)^2}, \mathbf{B} \succeq 0} \operatorname{tr}[\mathbf{H}_o \mathbf{B}] \qquad (12) \\ &\ni (\forall j \in J)\ \operatorname{tr}[\mathbf{H}_j \mathbf{B}] \geq 0 \wedge \mathbf{B}_{n+1,n+1} = 1. \end{aligned}$$

Interestingly, Eq.(12) is equivalent to Eq. (10) under the semi-definite relaxation that $\mathbf{X}$ is simply some positive semi-definite matrix—forgetting that the rank of $\mathbf{X}$ as written in Eq. (10) is one. As can be shown in greater generality [ARN], the duality transformation is an involution, and taking the dual of (12). The Lagrangian of Eq.(12) is

$$\mathcal{L}(\alpha, \boldsymbol{\phi}, \mathbf{Z}, \mathbf{B}) = \operatorname{tr}[(\mathbf{H}_{\boldsymbol{\psi}} + \mathbf{Z})\mathbf{B}] + \alpha(1 - \mathbf{B}_{n+1,n+1}).$$

$\mathbf{Z}$ does not help, condition becomes that $\mathbf{H}_{\boldsymbol{\psi}}$ is negative semi-definite, and value becomes α.

Standard Form for QCCPs

Let $\mathtt{P}$ be a linear QCCP with constraint indexing set J. Introducing an auxiliary "dummy" dimension, place all constraints in homogeneous form, $f_j = \mathbf{x}^\dagger \mathbf{H}_j \mathbf{x}$. Add the auxiliary constraint $\|x_{n+1}\|^2 = 1$ to the homogeneous form of the compact constraint $f_{\boldsymbol{\kappa}}$ until $\mathbf{H}_{\boldsymbol{\kappa}} \succeq \epsilon$, appending the resulting constraint $f_{\boldsymbol{\rho}}$ to the constraint set of $\mathtt{P}$. Note that any additional constraints on x_{n+1} are superfluous. If $\boldsymbol{x}_{\circledast}$ is a solution to

$$\max_{\boldsymbol{x} \in \mathbb{C}^{n+1}} 2\Re\left[\boldsymbol{s}_o^\dagger \mathbf{x}\right] \ni (\forall j \in J)\ \mathbf{x}^\dagger \mathbf{H}_j \mathbf{x} = 0 \wedge \mathbf{x}^\dagger \mathbf{H}_{\boldsymbol{\rho}} \mathbf{x} = \rho,$$

then $\boldsymbol{x}_{\circledast}/x_{n+1}$ is a solution to $\mathtt{P}$. Because $\mathbf{H}_\rho$ is positive definite, there is a unique matrix $\mathbf{H}_\rho^{1/2} \succ 0$ such that $\mathbf{H}_\rho^{1/2}\mathbf{H}_\rho^{1/2} = \mathbf{H}_\rho$. Redefining $\mathbf{x}$ as $\mathbf{x} \to \mathbf{H}_\rho^{-1/2}\mathbf{z}$, $\mathtt{P}$ becomes

$$\max_{\boldsymbol{x} \in \mathbb{C}^{n+1}} 2\Re\left[\tilde{\boldsymbol{s}}_o^\dagger \mathbf{z}\right] \ni (\forall j \in J)\ \mathbf{z}^\dagger \tilde{\mathbf{H}}_j \mathbf{z} = 0 \wedge \|\mathbf{z}\|^2 = \rho,$$

with $\tilde{\boldsymbol{s}}_o = \mathbf{H}_\rho^{1/2}\boldsymbol{s}_o$ and $\mathbf{H}_\rho^{-1/2}\tilde{\mathbf{H}}_j\mathbf{H}_\rho^{-1/2}$. Hence, any linear QCCP can be recast an equivalent linear QCCP subject to homogeneous hyperbolic constraints on a sphere. Within the main text we refer to this redefinition as "standard form".

Additional normalization may help improve numerical precision of the algorithm presented in the *Proposal* section. Specifically, once $\mathtt{P}$ has placed in standard form, each homogeneous constraint, $f_j(\mathbf{z}) = \mathbf{z}^\dagger \tilde{\mathbf{H}}_j \mathbf{z} \geq 0$, can be normalized by the largest (magnitude) eigenvalue of $\tilde{\mathbf{H}}_j$. After this rescaling setting $\tau = 1$ is guaranteed to induce strong duality in $\mathtt{P}(\tau, \mathbf{s}_o)$. As the algorithm progresses, further rescalings of the $\tilde{\mathbf{H}}_j$ matrices based on observed multiplier values, and desired constraint tightness, may also improve computational performance.

Compact QCQPs to a Compact QCLPs

In full generality, quadratically constrained quadratic programs (QCQPs) and quadratically constrained linear programs (QCLPs) are equivalent. Suppose $\mathtt{P}$ is a QCQP with $\mathbf{A}_o \neq \mathbf{0}$ and design variable $\mathbf{x}$. Increasing the dimension of $\mathbf{x}$ by one, $\mathbf{x} \mapsto \tilde{\mathbf{x}}$, $\mathtt{P}$ can be converted into a QCLP by modifying the linear part of the objective as $\mathbf{s}_o \mapsto \tilde{\mathbf{s}}_o = [\mathbf{s}, -1]$, and introducing the auxiliary constraint $\Re\left[\tilde{x}_{n+1}\right] - \tilde{x}^{\dagger}\tilde{\mathbf{A}}_o\tilde{\mathbf{x}}$, with $\tilde{\mathbf{A}}_o$ given by inserting $\mathbf{A}_o$ into the upper left-hand corner of a $(n+1)\times(n+1)$ matrix of zeros.

The requirement of compactness complicates matters. $\|\tilde{x}_{n+1}\|^2$ is clearly bounded, since $\mathbf{x}$ is bounded in $\mathtt{P}$. However, the relationship is not quadratic. Moreover, through homogeneous forms, the general determination of a constant c_{n+1} such that $\|\tilde{x}_{n+1}\|^2 \leq c_{n+1}$ does not relax or artificially tighten $\mathtt{P}$ is equivalent to the general determination of the value of a compact QCQP.

* Equal contribution, sean.molesky@polymtl.ca

[1] Glen E Bredon. *Topology and geometry*, volume 139. Springer Science & Business Media, 2013.

[2] Note1. A similar argument can be used to show that $\ker \mathbf{A}_{\boldsymbol{\phi}_{\circledast}} \neq \mathbf{0}$, then $\mathbf{s}_{\circ} \notin \ker \mathbf{A}_{\boldsymbol{\phi}_{\circledast}}$ unless $\inf_{\boldsymbol{\phi}\in\Phi_{\mathtt{P}}} \mathcal{D}\left(\boldsymbol{\phi}\right) = \mathcal{D}\left(\mathbf{0}\right)$ [6].

[3] Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.

[4] Hidetoshi Komiya. Elementary proof for sion's minimax theorem. *Kodai Mathematical Journal*, 11(1):5–7, 1988.

[5] Gerald B Folland. *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.

[6] Sean Molesky, Pengning Chao, and Alejandro W. Rodriguez. Hierarchical mean-field $\mathbb{T}$ operator bounds on electromagnetic scattering: Upper bounds on near-field radiative purcell enhancement. *Physical Review Research*, 2:043398, Dec 2020. doi: 10.1103/PhysRevResearch.2.043398.