

Conic Geometric Programming

Venkat Chandrasekaran^c and Parikshit Shah^{w *}

^c Departments of Computing and Mathematical Sciences and of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125

^w Wisconsin Institutes for Discovery
University of Wisconsin
Madison, WI 53715

October 10, 2013

Abstract

We introduce and study *conic geometric programs* (CGPs), which are convex optimization problems that unify geometric programs (GPs) and conic optimization problems such as semidefinite programs (SDPs). A CGP consists of a linear objective function that is to be minimized subject to affine constraints, convex conic constraints, and upper bound constraints on sums of exponential and affine functions. The conic constraints are the central feature of conic programs such as SDPs, while upper bounds on combined exponential/affine functions are generalizations of the types of constraints found in GPs. The dual of a CGP involves the maximization of the negative relative entropy between two nonnegative vectors jointly, subject to affine and conic constraints on the two vectors. Although CGPs contain GPs and SDPs as special instances, computing global optima of CGPs is not much harder than solving GPs and SDPs. More broadly, the CGP framework facilitates a range of new applications that fall outside the scope of SDPs and GPs. Specifically, we demonstrate the utility of CGPs in providing solutions to problems such as permanent maximization, hitting-time estimation in dynamical systems, the computation of the capacity of channels transmitting quantum information, and robust optimization formulations of GPs.

Keywords: convex optimization; semidefinite programming; permanent; robust optimization; quantum information; Von-Neumann entropy.

1 Introduction

Geometric programs (GPs) [11, 17] and semidefinite programs (SDPs) [52] are prominent classes of structured convex optimization problems that each generalize linear programs (LPs) in different ways. By virtue of convexity, GPs and SDPs possess favorable analytical and computational properties: a rich duality theory and polynomial-time algorithms for computing globally optimal solutions. Further, due to their flexible modeling capabilities, GPs and SDPs are useful in a range of problems throughout the sciences and engineering. Some prominent applications of SDPs include

*Email: venkatc@caltech.edu, pshah@discovery.wisc.edu

relaxations for combinatorial optimization [1, 21, 34, 54], ellipsoid volume optimization via determinant maximization [53], statistical estimation [49], problems in control and systems theory [9], and matrix norm optimization [44, 52]. On the other hand, applications of GPs include the computation of information-theoretic quantities [14], digital circuit gate sizing [10], chemical process control [56], matrix scaling and approximate permanent computation [37], entropy maximization problems in statistical learning [15], and power control in communication systems [13]. In this paper we describe a new class of tractable convex optimization problems called *conic geometric programs* (CGPs), which conceptually unify GPs and conic programs such as SDPs, thus facilitating a broad range of new applications.

Motivation Although SDPs and GPs provide effective solutions for many applications, natural generalizations of these applications that arise in practice fall outside the scope of GPs and SDPs.

1. **Permanent maximization** Given a collection of matrices, find the one with the largest permanent. Computing the permanent of a matrix is a well-studied problem that is thought to be computationally intractable. Therefore, we will only seek approximate solutions to the problem of permanent maximization.
2. **Robust optimization** The solution of a GP is sensitive to the input parameters of the problem. Compute GPs within a robust optimization framework so that the solutions offer robustness to variability in the problem parameters.
3. **Hitting times in dynamical systems** Given a linear dynamical system consisting of multiple modes and a region of feasible starting points, compute the smallest time required to hit a specified target set from an arbitrary feasible starting point.
4. **Quantum channel capacity** Compute the capacity of a noisy communication channel over which quantum states can be transmitted. The capacity of a channel is the maximum rate at which information can be communicated per use of the channel [30, 47].

A detailed mathematical description of each of these applications is given in Section 4. The pursuit of a unified and tractable solution framework for these questions motivates our development of CGPs in this paper. Some of these applications and close variants have been studied previously in the literature; we discuss the relevant prior work and the more powerful generalizations afforded by CGPs in Section 4. We also describe in Section 4 other problems in domains such as quantum state tomography, statistical mechanics (computation of equilibrium densities), and kernel learning in data analysis for which previously proposed methods can be viewed as special instances of CGPs.

A GP is a convex optimization problem in which a linear function of a decision variable $\mathbf{x} \in \mathbb{R}^n$ is minimized subject to affine constraints on \mathbf{x} and upper-bound constraints on *positive sums of exponentials* of affine functions of \mathbf{x} . Formally, these latter constraints are specified as follows¹:

$$\sum_{i=1}^k \mathbf{c}_i \exp([\mathbf{Q}'\mathbf{x}]_i) + d \leq 0.$$

Here $\mathbf{Q} \in \mathbb{R}^{n \times k}$, $\mathbf{c} \in \mathbb{R}^k$ is a vector with nonnegative entries (\mathbf{c}_i is the i 'th entry of \mathbf{c}), and $d \in \mathbb{R}$; all these quantities are fixed problem parameters. Note that if there are no constraints on sums-of-exponentials, then a GP reduces to an LP. Therefore, GPs include LPs as a special case. An SDP

¹GPs are sometimes described in the literature as optimization problems involving so-called *posynomial* functions, which are in general non-convex; details of the equivalence between the posynomial description of GPs and their specification in convex form can be found in [11, 17].

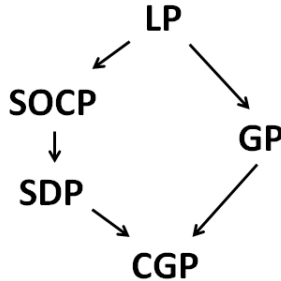


Figure 1: A graphical description of the various families of structured convex optimization problems ordered by inclusion (the arrows denote inclusion).

is a prominent instance of a *conic* optimization problem that generalizes LPs in a different manner in comparison to GPs. In a conic program, the goal is to minimize a linear function of a decision variable subject to affine constraints as well as constraints that require the decision variable to belong to a convex cone. Examples of conic optimization problems include LPs, second-order cone programs or SOCPs, and SDPs. We note here that GPs can also be viewed as conic programs for an appropriately chosen cone [20], although they are not usually described in this fashion. In an SDP the decision variable is a symmetric matrix and the conic constraint requires the matrix to be positive semidefinite. If the matrix is further constrained to be diagonal, then an SDP reduces to an LP. Thus, both SDPs and GPs are generalizations of LPs, as demonstrated in Figure 1. However, neither SDPs nor GPs contain the other as a special case, and we expand upon this point next.

While GPs and SDPs have been employed successfully across many application domains, they also have certain restrictions that limit their applicability in solving our motivating problems. For example, GPs do not in general allow for matrix decision variables on which spectral constraints can be imposed, e.g., a positive semidefiniteness condition on a symmetric matrix. On the other hand, SDPs face the basic restriction that the constraint sets must have algebraic boundaries, i.e., these are specified by the vanishing of systems of polynomial equations. Consequently, transcendental quantities such as the entropy function – which are naturally expressed via GPs – are inexpressible using SDPs. These restrictions of GPs and SDPs limit their application in domains such as quantum information in which decision variables are usually specified as positive semidefinite matrices, but the objective functions are variants of the entropy function.

Conic geometric programs These limitations inform our development in Section 2 of CGPs, which are obtained by blending the salient features of GPs and of conic programs such as SDPs. A CGP is an optimization problem in which a linear function of a decision variable $\mathbf{x} \in \mathbb{R}^n$ is minimized subject to affine constraints and two additional types of constraints. First, we allow conic constraints that require \mathbf{x} to belong to a convex cone. This addresses the restriction with GPs as the conic constraint can, for example, require a matrix decision variable to be positive semidefinite. In order to address the second restriction, namely the inexpressibility of transcendental quantities using SDPs, we consider convex constraints in \mathbf{x} that combine an exponential function and an affine function as follows:

$$\exp(\mathbf{q}'\mathbf{x} - s) + \mathbf{r}'\mathbf{x} - t \leq 0. \quad (1)$$

Here $\mathbf{q}, \mathbf{r} \in \mathbb{R}^n$ and $s, t \in \mathbb{R}$ are fixed problem parameters. Note that this constraint is more general than the pure sum-of-exponentials constraint allowed in GPs, as (1) can be used to specify upper-bound constraints on combined sums of exponential and affine functions. Duffin et al [17] mention

this generalization of GPs in their initial work (see also [11]), but such constraints are usually not considered as part of the standard form of a GP. While constraints of the type (1) may appear to be a modest extension of GPs, we will show in Section 4 that such combined exponential/affine constraints along with conic constraints significantly expand the expressive power of CGPs relative to GPs and SDPs, and consequently broaden the range of applications. In particular, constraints of the form (1) enable effective *lift-and-project* representations in which complicated convex sets in \mathbb{R}^n are specified as the projection of efficiently described sets in $\mathbb{R}^{n'}$ with $n' > n$ (see Section 3). Such lift-and-project techniques employed in the CGP framework are very useful in obtaining solutions to GPs that are robust to parameter uncertainty and in approximately maximizing the permanent over a family of matrices. In summary, as described in Figure 2, a CGP is an optimization problem in which a linear function is minimized subject to three types of constraints: combined exponential/affine constraints (1), affine constraints, and convex conic constraints. Thus, CGPs are convex optimization problems, and they contain GPs, SOCPs, and SDPs as special cases as shown in Figure 1.

The convexity of CGPs leads to favorable analytical properties, and these are especially apparent via convex duality. As described in Section 2, the dual of a CGP has a particularly appealing form: it is a convex optimization problem in which the negative of the *relative entropy* between two nonnegative vectors is jointly maximized subject to affine and conic constraints on the vectors. Recall that the relative entropy between two nonnegative vectors is a jointly convex function. The conic constraint is expressed in terms of the dual of the cone that specifies the constraint in the primal CGP. The additional expressive capability of CGPs in contrast to GPs and SDPs is also apparent via the respective dual formulations. Figure 2 provides a concise summary – see Section 2 for more details.

In addition to containing GPs and SDPs as special instances, CGPs offer significantly more expressive power than either of these families of convex programs considered separately. For example, GPs can be used to solve entropy maximization problems involving probability vectors, while CGPs can be used to solve entropy maximization problems involving quantum density matrices. Another example is a constraint in a decision variable $\mathbf{x} \in \mathbb{R}^n$ of the form:

$$\sum_{i=1}^k \mathbf{C}^{(i)} \exp([\mathbf{Q}'\mathbf{x}]_i) \preceq \mathbf{I},$$

where the $\mathbf{C}^{(i)}$ are positive semidefinite matrices, \mathbf{I} is an identity matrix of appropriate size, $\mathbf{Q} \in \mathbb{R}^{n \times k}$, and \preceq denotes the matrix semidefinite ordering. Such constraints allow us to minimize the largest eigenvalue of matrix expressions of the form $\sum_{i=1}^k \mathbf{C}^{(i)} \exp([\mathbf{Q}'\mathbf{x}]_i)$ subject to additional convex constraints on \mathbf{x} . These types of expressions cannot be handled directly via GPs or SDPs alone; see Section 3 for more examples of convex constraints that can be expressed via CGPs.

Solutions of motivating problems The enhanced expressive power of CGPs enables us to solve our motivating problems described earlier in the introduction. CGPs can be used to approximately maximize the permanent over a family of (entrywise) nonnegative matrices that is specified via affine and conic constraints, e.g., a set that is defined as a section of the cone of positive semidefinite matrices. We provide bounds on the quality of the approximation based on Van der Waerden’s conjecture, which was originally settled in the 1980s [18, 19] and for which Gurvits gave an elegant proof more recently [26]. CGPs also provide solutions to GPs that are robust to parameter uncertainty when the uncertainty sets are specified via affine and conic constraints. In this paper, we consider worst-case deterministic variability in the parameters. As our third example, we demonstrate that CGPs can compute the hitting times in linear dynamical systems consisting of multiple

		Conic programs	GPs	CGPs
Primal	<i>objective</i>	linear	linear	linear
	<i>constraints</i>	affine, conic	affine, sums of exponentials	affine, conic, combined exponential/affine
Dual	<i>objective</i>	linear	normalized entropy + linear	(negative) joint relative entropy + linear
	<i>constraints</i>	affine, conic	affine, nonnegativity	affine, conic

Figure 2: A summary of the types of objective functions and constraints in the primal and dual formulations of conic programs, GPs, and CGPs. Note that the dual formulation consists of the maximization of a concave objective function subject to convex constraints.

modes in which the initial point may belong to a set specified by affine and conic constraints, and in which the value of each mode is not known precisely (each mode may take on any value in some known range). Finally, CGPs can be used to exactly or approximately compute certain capacities associated with quantum channels; as an example, in Section 4.4 we provide a comparison between a “classical-to-quantum” capacity of a quantum channel, and the capacity of a classical channel induced by the quantum channel (i.e., one is restricted to send and receive only classical information). These and other applications of CGPs are detailed in Section 4.

Paper outline In Section 2, we define a CGP, derive the corresponding dual problem of relative entropy minimization, and discuss the complexity of solving a CGP using an interior-point method. We give examples of functions and constraints that can be expressed via CGPs in Section 3. We describe solutions to our motivating problems based on CGP in Section 4. We conclude with a brief summary and discussion of further research directions in Section 5.

Notational convention The nonnegative orthant in \mathbb{R}^k of vectors consisting of entrywise non-negative entries is denoted by \mathbb{R}_+^k and the positive orthant with strictly positive elements by \mathbb{R}_{++}^k . The space of $n \times n$ symmetric matrices is denoted by $\text{Sym}(n)$, which is a subspace of $\mathbb{R}^{n \times n}$, in particular, $\text{Sym}(n) \cong \mathbb{R}^{\binom{n+1}{2}}$. We refer throughout the paper to (closed) *convex cones*, which are sets that are closed under nonnegative linear combinations of the elements. For a vector $\mathbf{x} \in \mathbb{R}^n$ we denote the elementwise exponential by $\exp\{\mathbf{x}\} \in \mathbb{R}^n$ with curly braces so that $[\exp\{\mathbf{x}\}]_i = \exp(\mathbf{x}_i)$. Similarly the elementwise logarithm for vectors $\mathbf{x} \in \mathbb{R}_{++}^n$ is denoted by $\log\{\mathbf{x}\} \in \mathbb{R}^n$. For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we denote elementwise ordering by $\mathbf{x} \leq \mathbf{y}$ to specify that $\mathbf{x}_i \leq \mathbf{y}_i$ for $i = 1, \dots, n$. We denote positive semidefinite ordering by \preceq so that $\mathbf{M} \preceq \mathbf{N}$ implies $\mathbf{N} - \mathbf{M}$ is a positive semidefinite matrix for symmetric matrices $\mathbf{M}, \mathbf{N} \in \text{Sym}(n)$. The *dual* of a convex cone $\mathcal{K} \subseteq \mathbb{R}^n$ is denoted \mathcal{K}^* and it is the set of elements in \mathbb{R}^n that form a nonnegative inner-product with every element of \mathcal{K} :

$$\mathcal{K}^* = \{\boldsymbol{\nu} \mid \mathbf{x}'\boldsymbol{\nu} \geq 0, \forall \mathbf{x} \in \mathcal{K}\}. \quad (2)$$

Finally, for any real-valued function f with domain $\mathcal{D} \subseteq \mathbb{R}^n$, the *Fenchel* or *convex conjugate* f^* is given by:

$$f^*(\boldsymbol{\nu}) = \sup\{\mathbf{x}'\boldsymbol{\nu} - f(\mathbf{x}) \mid \mathbf{x} \in \mathcal{D}\} \quad (3)$$

for $\boldsymbol{\nu} \in \mathbb{R}^n$. For more details of these concepts from convex analysis, we refer the reader to [46].

2 Definition, Duality, and Numerical Aspects

In this section, we give a formal definition of CGPs and derive the associated dual problems. We also describe the key elements of an interior-point method for solving CGPs.

2.1 What is a CGP?

CGPs are optimization problems specified in terms of conic, affine, and combined exponential/affine constraints:

Definition 1. A CGP is a convex optimization problem in a decision variable $\mathbf{x} \in \mathbb{R}^n$ as follows:

$$\begin{aligned} \inf_{\mathbf{x} \in \mathbb{R}^n} \quad & \mathbf{p}'\mathbf{x} \\ \text{s.t.} \quad & \exp\{\mathbf{Q}'\mathbf{x} - \mathbf{u}\} + \mathbf{R}'\mathbf{x} - \mathbf{v} \leq \mathbf{0} \\ & \mathbf{S}'\mathbf{x} = \mathbf{w} \\ & \mathbf{x} \in \mathcal{K} \end{aligned} \tag{P}$$

Here $\mathbf{Q}, \mathbf{R} \in \mathbb{R}^{n \times k}$, $\mathbf{S} \in \mathbb{R}^{n \times m}$, and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^k$, $\mathbf{w} \in \mathbb{R}^m$, $\mathbf{p} \in \mathbb{R}^n$ are fixed parameters. Further, \mathcal{K} is a convex cone.

CGPs can be specialized to obtain conic optimization problems. Specifically, by considering CGPs without the combined exponential/affine constraint, we obtain optimization problems in which a linear function is minimized subject to affine constraints and a conic constraint – with appropriate choice of this cone, one can obtain conic programs such as LPs (\mathcal{K} is the nonnegative orthant), SOCPs (\mathcal{K} is the second-order or Lorentz cone), and SDPs (\mathcal{K} is the cone of positive semidefinite matrices). Similarly, CGPs can be specialized to obtain GPs as special cases. In particular, we may neglect the conic constraint (i.e., $\mathcal{K} = \mathbb{R}^n$), and we need to be able to express constraints of the form:

$$\sum_{i=1}^k \mathbf{c}_i \exp([\mathbf{Q}'\mathbf{x}]_i) + d \leq 0$$

for $\mathbf{c} \in \mathbb{R}_{++}^k$, $\mathbf{Q} \in \mathbb{R}^{n \times k}$, and $d \in \mathbb{R}$. Letting $\mathbf{u} = \log\{\mathbf{c}\} \in \mathbb{R}^k$ be the vector of the elementwise logarithms of the coefficients, this constraint is equivalent to the following system of constraints in decision variables $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and fixed parameters $\mathbf{Q} \in \mathbb{R}^{n \times k}$, $\mathbf{u} \in \mathbb{R}^k$, $d \in \mathbb{R}$:

$$\exp\{\mathbf{Q}'\mathbf{x} + \mathbf{u}\} - \mathbf{y} \leq \mathbf{0}; \quad \mathbf{1}'\mathbf{y} + d \leq 0,$$

each of which can be expressed within the CGP framework (P). Here $\mathbf{1} \in \mathbb{R}^n$ refers to the all-ones vector. In this manner, CGPs provide a common generalization of LPs, SOCPs, SDPs, and GPs.

2.2 Dual of a CGP

The dual of a CGP is specified in terms of the *relative entropy* function of two vectors $\boldsymbol{\nu}, \boldsymbol{\lambda} \in \mathbb{R}_+^k$:

$$D(\boldsymbol{\nu}, \boldsymbol{\lambda}) = \sum_{i=1}^k \nu_i \log \left(\frac{\nu_i}{\lambda_i} \right). \tag{4}$$

We adopt the convention that $D(\mathbf{0}, \mathbf{0}) = 0$. The relative entropy function is *jointly convex* in both the arguments. In order to transparently obtain the convex dual of a CGP, we first establish a relationship between the relative entropy function and the combined exponential/affine function via convex conjugacy. This result is well-known but we include it here for completeness.

Lemma 1. Let $\chi_{\text{expaff}}(\mathbf{x}, \mathbf{y}) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ denote the characteristic function of the exponential/affine as follows:

$$\chi_{\text{expaff}}(\mathbf{x}, \mathbf{y}) = \begin{cases} 0, & \text{if } \exp\{\mathbf{x}\} + \mathbf{y} \leq \mathbf{0} \\ \infty, & \text{otherwise} \end{cases}$$

Then the convex conjugate $\chi_{\text{expaff}}^*(\boldsymbol{\nu}, \boldsymbol{\lambda}) = D(\boldsymbol{\nu}, e\boldsymbol{\lambda})$ with domain $(\boldsymbol{\nu}, \boldsymbol{\lambda}) \in \mathbb{R}_+^n \times \mathbb{R}_+^n$, where e is the Euler constant.

Proof. From the definition (3) of the convex conjugate of a function we have that:

$$\begin{aligned} \chi_{\text{expaff}}^*(\boldsymbol{\nu}, \boldsymbol{\lambda}) &= \sup_{\mathbf{x}, \mathbf{y}} \mathbf{x}'\boldsymbol{\nu} + \mathbf{y}'\boldsymbol{\lambda} \text{ s.t. } \exp\{\mathbf{x}\} + \mathbf{y} \leq \mathbf{0} \\ &= \sup_{\mathbf{x}} \mathbf{x}'\boldsymbol{\nu} - \exp\{\mathbf{x}\}'\boldsymbol{\lambda}. \end{aligned} \quad (5)$$

Taking derivatives with respect to each \mathbf{x}_i and setting the result equal to 0, we have that the optimal $\hat{\mathbf{x}}$ must satisfy:

$$\boldsymbol{\nu}_i - \exp(\hat{\mathbf{x}}_i)\boldsymbol{\lambda}_i = 0$$

for each $i = 1, \dots, n$. Plugging in the optimal value $\hat{\mathbf{x}}_i = \log\left(\frac{\boldsymbol{\nu}_i}{\boldsymbol{\lambda}_i}\right)$ into (5), we have that

$$\begin{aligned} \chi_{\text{expaff}}^*(\boldsymbol{\nu}, \boldsymbol{\lambda}) &= \sum_i \boldsymbol{\nu}_i \log\left(\frac{\boldsymbol{\nu}_i}{\boldsymbol{\lambda}_i}\right) - \boldsymbol{\nu}_i \\ &= D(\boldsymbol{\nu}, e\boldsymbol{\lambda}), \end{aligned}$$

which is the claimed result. \square

We are now in a position to state the main result concerning the duals of CGPs. As CGPs are specified in terms of exponential/affine functions, it is natural to expect the relative entropy function to play a prominent role in the dual based on the conjugacy result of Lemma 1.

Proposition 2. The dual of the CGP (P) is the following convex optimization problem:

$$\begin{aligned} \sup_{\boldsymbol{\nu}, \boldsymbol{\lambda} \in \mathbb{R}_+^k, \boldsymbol{\delta} \in \mathbb{R}^m} & - [D(\boldsymbol{\nu}, e\boldsymbol{\lambda}) + \mathbf{u}'\boldsymbol{\nu} + \mathbf{v}'\boldsymbol{\lambda} + \mathbf{w}'\boldsymbol{\delta}] \\ \text{s.t.} & \quad \mathbf{p} + \mathbf{Q}\boldsymbol{\nu} + \mathbf{R}\boldsymbol{\lambda} + \mathbf{S}\boldsymbol{\delta} \in \mathcal{K}^* \end{aligned} \quad (D)$$

Here e is the Euler constant, $\mathcal{K}^* \subseteq \mathbb{R}^n$ is the dual of the cone \mathcal{K} (2), and $\mathbf{Q}, \mathbf{R}, \mathbf{S}$ are the matrices specified in the primal form of a CGP (P).

Proof. We consider a CGP in the following form with additional variables $\mathbf{y}, \mathbf{z} \in \mathbb{R}^k$ for the exponential/affine constraints in order to simplify the derivation of the dual:

$$\begin{aligned} \inf_{\mathbf{x} \in \mathbb{R}^n, \mathbf{y}, \mathbf{z} \in \mathbb{R}^k} & \quad \mathbf{p}'\mathbf{x} + \chi_{\text{expaff}}(\mathbf{y}, \mathbf{z}) \\ \text{s.t.} & \quad \mathbf{Q}'\mathbf{x} - \mathbf{u} \leq \mathbf{y} \\ & \quad \mathbf{R}'\mathbf{x} - \mathbf{v} \leq \mathbf{z} \\ & \quad \mathbf{S}'\mathbf{x} = \mathbf{w} \\ & \quad \mathbf{x} \in \mathcal{K} \end{aligned}$$

Introducing dual variables $\boldsymbol{\nu}, \boldsymbol{\lambda} \in \mathbb{R}_+^k, \boldsymbol{\delta} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathcal{K}^* \subseteq \mathbb{R}^n$, we have the following Lagrangian:

$$\begin{aligned} L(\mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\nu}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\mu}) &= \mathbf{x}'\mathbf{p} + \chi_{\text{expaff}}(\mathbf{y}, \mathbf{z}) + [\mathbf{x}'\mathbf{Q} - \mathbf{u}' - \mathbf{y}']\boldsymbol{\nu} \\ &\quad + [\mathbf{x}'\mathbf{R} - \mathbf{v}' - \mathbf{z}']\boldsymbol{\lambda} + [\mathbf{x}'\mathbf{S} - \mathbf{w}']\boldsymbol{\delta} - \mathbf{x}'\boldsymbol{\mu} \\ &= \chi_{\text{expaff}}(\mathbf{y}, \mathbf{z}) - \mathbf{y}'\boldsymbol{\nu} - \mathbf{z}'\boldsymbol{\lambda} \\ &\quad + \mathbf{x}'[\mathbf{p} + \mathbf{Q}\boldsymbol{\nu} + \mathbf{R}\boldsymbol{\lambda} + \mathbf{S}\boldsymbol{\delta} - \boldsymbol{\mu}] - \mathbf{u}'\boldsymbol{\nu} - \mathbf{v}'\boldsymbol{\lambda} - \mathbf{w}'\boldsymbol{\delta}, \end{aligned}$$

where the second expression is obtained by a rearrangement of terms. Minimizing the Lagrangian over the primal variables \mathbf{y}, \mathbf{z} , we have the following simplified expression in terms of the joint relative entropy function by appealing to Lemma 1:

$$\inf_{\mathbf{y}, \mathbf{z}} L(\mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\nu}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = -D(\boldsymbol{\nu}, e\boldsymbol{\lambda}) + \mathbf{x}'[\mathbf{p} + \mathbf{Q}\boldsymbol{\nu} + \mathbf{R}\boldsymbol{\lambda} + \mathbf{S}\boldsymbol{\delta} - \boldsymbol{\mu}] - \mathbf{u}'\boldsymbol{\nu} - \mathbf{v}'\boldsymbol{\lambda} - \mathbf{w}'\boldsymbol{\delta}. \quad (6)$$

Next we minimize over the primal variable \mathbf{x} and obtain the following condition that must be satisfied by the dual variables:

$$\mathbf{p} + \mathbf{R}\boldsymbol{\lambda} + \mathbf{Q}\boldsymbol{\nu} + \mathbf{S}\boldsymbol{\delta} - \boldsymbol{\mu} = \mathbf{0}.$$

By eliminating the dual variable $\boldsymbol{\mu}$, we are left with the following constraint:

$$\mathbf{p} + \mathbf{R}\boldsymbol{\lambda} + \mathbf{Q}\boldsymbol{\nu} + \mathbf{S}\boldsymbol{\delta} \in \mathcal{K}^*.$$

From (6) and by incorporating this constraint, we obtain the dual optimization problem (D). \square

We emphasize that the dual of a CGP is a *joint* relative entropy optimization problem in which the negative of the relative entropy between two vectors is jointly maximized subject to linear and conic constraints on the vectors. The dual of a standard GP is a *normalized entropy* maximization problem subject to linear and nonnegativity constraints (see [11, 13] for an extensive survey of GPs). Specifically, the normalized entropy of a nonnegative vector $\boldsymbol{\nu} \in \mathbb{R}_+^k$ is a concave function defined as follows [12]:

$$h_{\text{norm}}(\boldsymbol{\nu}) = - \sum_{i=1}^k \nu_i \log \left(\frac{\nu_i}{\mathbf{1}'\boldsymbol{\nu}} \right)$$

This function can be viewed as a restriction of the negative of the joint relative entropy function (4) in which each entry of the second argument $\boldsymbol{\lambda}$ is identically equal to $\lambda_i = \mathbf{1}'\boldsymbol{\nu}$. Therefore, the additional expressive power provided by CGPs in contrast to standard GPs can be understood from the dual viewpoint as the result of two generalizations – the first is the negative relative entropy function in the objective instead of the normalized entropy, and the second is the linear and general conic constraints rather than just linear and nonnegativity constraints. The dual of a CGP can also be viewed as a generalization of the dual of a conic program: although the constraints in both cases are of the same type, the dual of a conic program consists of a linear objective while the dual of a CGP consists of a joint relative entropy term as well as linear terms. We refer the reader to the table in Figure 2 for a concise summary.

2.3 Numerical Solution of CGPs

Interior-point techniques [39] provide a general methodology via *self-concordant barrier functions* to obtain polynomial-time algorithms for general families of convex programs such as LPs, SOCPs, GPs, and SDPs. We refer the reader to many excellent texts for details [6, 39, 42, 45]. Here we

highlight the types of barrier functions that may be employed in interior-point methods for CGPs, and we give runtime bounds for these algorithms.

For conic constraints, natural self-concordant barriers for cones such as the orthant, the second-order cone, and the semidefinite cone are specified by the associated *logarithmic barriers* [39, 40]. The situation is a bit more subtle for the combined exponential/affine constraints $\exp\{\mathbf{Q}'\mathbf{x} - \mathbf{u}\} + \mathbf{R}'\mathbf{x} - \mathbf{v} \leq \mathbf{0}$ in the primal form of a CGP (with $\mathbf{Q}, \mathbf{R} \in \mathbb{R}^{n \times k}$ and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^k$). Consider the following restatement of this constraint with additional variables $\mathbf{y}, \mathbf{z} \in \mathbb{R}^k$:

$$\mathbf{y} - \log\{-\mathbf{z}\} \leq \mathbf{0}, \quad \mathbf{z} \leq \mathbf{0}, \quad \mathbf{Q}'\mathbf{x} - \mathbf{u} \leq \mathbf{y}, \quad \mathbf{R}'\mathbf{x} - \mathbf{v} \leq \mathbf{z}.$$

This method of rewriting exponential constraints is also employed in interior-point solvers for GPs [39]. The redundancy in these constraints is useful because the following logarithmic barrier function for these constraints is self-concordant [39]:

$$\sum_{i=1}^k \left(\log(-\mathbf{z}_i) + \log(\log(-\mathbf{z}_i) - \mathbf{y}_i) + \log(\mathbf{y}_i + \mathbf{u}_i - [\mathbf{Q}'\mathbf{x}]_i) + \log(\mathbf{z}_i + \mathbf{v}_i - [\mathbf{R}'\mathbf{x}]_i) \right).$$

One can directly write down the usual logarithmic barrier for the constraint $\exp\{\mathbf{Q}'\mathbf{x} - \mathbf{u}\} + \mathbf{R}'\mathbf{x} - \mathbf{v} \leq \mathbf{0}$ without the additional variables \mathbf{y}, \mathbf{z} , but this barrier is not known to be self-concordant. The runtime analysis of interior-point procedures relies on the self-concordance assumption, and therefore one resorts to the above reformulation [39]. An upper bound on the number of Newton steps that must be computed to obtain an ϵ -accurate solution of a CGP is given by [39]:

$$\# \text{ Newton steps} = \mathcal{O} \left(\sqrt{k + \vartheta_{\mathcal{K}}} \log \left(\frac{k + \vartheta_{\mathcal{K}}}{\epsilon} \right) \right).$$

Here $\vartheta_{\mathcal{K}}$ is a “complexity parameter” of the barrier associated with the cone \mathcal{K} [39]. If $\mathcal{K} = \mathbb{R}_+^n$ is the nonnegative orthant in the primal CGP (P), then $\vartheta_{\mathcal{K}} = n$ and we have the following bound:

$$\# \text{ total operations for CGP with orthant constraint} = \mathcal{O} \left((k + n)^{3.5} \log \left(\frac{k + n}{\epsilon} \right) \right).$$

Similarly, if $\mathcal{K} \subset \text{Sym}(\ell)$ is the cone of positive semidefinite matrices in (P) (note that $n = \binom{\ell+1}{2}$), then $\vartheta_{\mathcal{K}} = \ell$ and we have:

$$\# \text{ total operations for CGP with semidefinite constraint} = \mathcal{O} \left((k + \ell)^{3.5} \ell^3 \log \left(\frac{k + \ell}{\epsilon} \right) \right).$$

Note that these bounds hold for arbitrary CGPs. The types of CGPs that arise in practical settings frequently possess additional structure such as sparsity or symmetry, which can be exploited to significantly reduce the overall runtime.

3 The Expressive Power of CGPs

A central question with any structured family of convex programs is the class of convex sets that can be represented within that family. For example, all polytopes are representable within the framework of LPs. The problem of characterizing the set of SDP-representable convex sets is more challenging and it is an active research topic [8, 23, 29]. As we have seen in the previous section, CGPs can be specialized to obtain traditional GPs as well as conic optimization problems such as SOCPs and SDPs. In this section, we describe the expressive power of the CGP formulation in representing convex sets and functions that cannot be represented purely via GPs or via SDPs. The examples discussed below play a prominent role in Section 4 where we demonstrate the utility of

CGPs in a range of problem domains. We denote the sum of the top- r eigenvalues of $\mathbf{M} \in \text{Sym}(n)$ by $s_r(\mathbf{M})$. In this section, we will exploit the fact that the sublevel sets of this function are SDP-representable [6]:

$$\begin{aligned} s_r(\mathbf{M}) \leq t \\ \Updownarrow \\ \exists \mathbf{Z} \in \text{Sym}(n), \alpha \in \mathbb{R} \text{ s.t. } \mathbf{Z} \succeq \mathbf{0}, \mathbf{M} - \mathbf{Z} \preceq \alpha \mathbf{I}, r\alpha + \text{trace}(\mathbf{Z}) \leq t. \end{aligned} \quad (7)$$

3.1 Basics of Lift-and-Project Representations

At an elementary level, the collection of CGP-representable sets in \mathbb{R}^n includes those that can be represented as the intersection of sets specified by combined exponential/affine inequalities, affine constraints, and conic constraints, all in a decision variable $\mathbf{x} \in \mathbb{R}^n$. However, such direct CGP representations in \mathbb{R}^n can be somewhat limited in their expressive power, and therefore it is of interest to consider *lifted* CGP representations in which a convex set in \mathbb{R}^n is defined as the *projection* of a CGP-representable set in $\mathbb{R}^{n'}$ for $n' > n$. Such lift-and-project CGP representations are extremely useful if n' is not much larger than n and if the CGP representation in $\mathbb{R}^{n'}$ is tractable to compute. Here we give a simple example of a lift-and-project representation to illustrate its power. The remainder of Section 3 is devoted to more sophisticated examples.

The logarithm function has the following variational representation for $\alpha > 0$, which can be easily verified:

$$\log(\alpha) = \inf_{\beta \in \mathbb{R}} e^{-\beta-1} \alpha + \beta. \quad (8)$$

Here β plays the role of the additional variable or “lifting dimension.” The appearance of β in this nonlinear expression with exponential and affine terms is somewhat non-traditional; in comparison, in conic lift-and-project representations the additional variables appear in linear and conic constraints. This simple observation yields a number of interesting CGP lift-and-project representations.

Logarithm of sum-of-exponentials Expressions involving a combination of an affine function, a sum-of-exponentials, and a logarithm of a sum-of-exponentials can be upper-bounded by appealing to (8):

$$\begin{aligned} \log \left(\sum_{i=1}^k \mathbf{c}_i \exp([\mathbf{Q}'\mathbf{x}]_i) \right) + \sum_{j=1}^{\bar{k}} \bar{\mathbf{c}}_j \exp([\bar{\mathbf{Q}}'\mathbf{x}]_j) + \mathbf{r}'\mathbf{x} - t \leq 0 \\ \Updownarrow \\ \exists \beta \text{ s.t. } \sum_{i=1}^k \mathbf{c}_i \exp([\mathbf{Q}'\mathbf{x}]_i - \beta - 1) + \sum_{j=1}^{\bar{k}} \bar{\mathbf{c}}_j \exp([\bar{\mathbf{Q}}'\mathbf{x}]_j) + \mathbf{r}'\mathbf{x} - t + \beta \leq 0 \end{aligned}$$

Here $\mathbf{Q} \in \mathbb{R}^{n \times k}$, $\bar{\mathbf{Q}} \in \mathbb{R}^{n \times \bar{k}}$, $\mathbf{c} \in \mathbb{R}_+^k$, $\bar{\mathbf{c}} \in \mathbb{R}_+^{\bar{k}}$, $\mathbf{r} \in \mathbb{R}^n$, $t \in \mathbb{R}$ are fixed parameters. The latter constraint can be specified in the CGP framework (P), with the additional variable β playing the role of the lifting dimension. We note that upper bounds on a pure sum-of-exponentials function, or equivalently, on a logarithm of a sum-of-exponentials are considered within the framework of GPs [11, 13]. However, CGPs can be used to bound composite expressions involving both these types of functions based on the additional flexibility provided by constraints on combined exponential/affine functions.

Product of sum-of-exponentials The reformulation (8) of the logarithm is also useful in providing CGP representations of *products* of expressions, each of which is sum-of-exponentials. Specifically, consider the following expressions in a decision variable $\mathbf{x} \in \mathbb{R}^n$ for $j = 1, \dots, p$:

$$f_j(\mathbf{x}) = \sum_{i=1}^{k_j} \mathbf{c}_i^{(j)} \exp\left(\left[\mathbf{Q}^{(j)'} \mathbf{x}\right]_i\right). \quad (9)$$

As usual, $\mathbf{c}^{(j)} \in \mathbb{R}_+^{k_j}$, $\mathbf{Q}^{(j)} \in \mathbb{R}^{n \times k_j}$ for $j = 1, \dots, p$. Suppose we wish to represent the constraint:

$$\prod_{j=1}^p f_j(\mathbf{x}) \leq t. \quad (10)$$

This expression can equivalently be written as:

$$\sum_{j=1}^p \log(f_j(\mathbf{x})) \leq \log(t).$$

By appealing to (8) and (9), we have that the constraint (10) is equivalent to:

$$\exists \boldsymbol{\beta} \in \mathbb{R}^p, \delta \in \mathbb{R} \text{ s.t. } \sum_{j=1}^p \sum_{i=1}^{k_j} \mathbf{c}_i^{(j)} \exp\left(\left[\mathbf{Q}^{(j)'} \mathbf{x}\right]_i - \beta_j - 1\right) + \mathbf{1}'\boldsymbol{\beta} \leq \delta \text{ and } \exp(\delta) \leq t. \quad (11)$$

The reason we introduce the extra variable δ is to allow for potentially additional convex constraints on t , as constraints of the form (10) may be embedded inside larger convex programs in which t may be a decision variable. Here again $\boldsymbol{\beta} \in \mathbb{R}^p$ plays the role of the lifting dimension. Using these techniques one can also bound composite expressions involving a sum of exponentials, a logarithm of a sum of exponentials, and a product of sums of exponentials.

3.2 Sums of Exponentials with General Coefficients

As our next collection of examples, we consider constraints on sums of exponentials in which the coefficients may be more general than positive scalars.

Matrix-weighted sum-of-exponentials Consider the following constraint in decision variable $\mathbf{x} \in \mathbb{R}^n$ with $\mathbf{Q} \in \mathbb{R}^{n \times k}$, $\mathbf{C}^{(i)} \in \text{Sym}(\ell)$ for $i = 1, \dots, k$, $\mathbf{R}^{(j)} \in \text{Sym}(\ell)$ for $j = 1, \dots, n$, $\mathbf{S} \in \text{Sym}(\ell)$, and where each $\mathbf{C}^{(i)} \succeq \mathbf{0}$:

$$\sum_{i=1}^k \mathbf{C}^{(i)} \exp([\mathbf{Q}'\mathbf{x}]_i) + \sum_{j=1}^n \mathbf{R}^{(j)} \mathbf{x}_j + \mathbf{S} \preceq \mathbf{0}.$$

This constraint can equivalently be written in terms of decision variables $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^k$ as:

$$\sum_{i=1}^k \mathbf{C}^{(i)} \mathbf{y}_i + \sum_{j=1}^n \mathbf{R}^{(j)} \mathbf{x}_j + \mathbf{S} \preceq \mathbf{0}; \quad \exp\{\mathbf{Q}'\mathbf{x}\} - \mathbf{y} \leq \mathbf{0}.$$

These latter constraints can be expressed within the framework of CGP. The first constraint here is a *linear matrix inequality* and an SDP consists purely of such types of constraints. A CGP allows for additional exponential/affine constraints of the second type, thus broadening the scope of SDPs. As an instance of this construction, one can impose upper-bound constraints on sums of the top- r eigenvalues of matrix-valued expressions such as $\sum_{i=1}^k \mathbf{C}^{(i)} \exp([\mathbf{Q}'\mathbf{x}]_i)$ with $\mathbf{C}^{(i)} \succeq \mathbf{0}$ by applying (7).

Coefficients from general cones Generalizing the previous example, we recall that the cone of positive semidefinite matrices is a *proper* cone [6] – these are cones in Euclidean space that are closed, convex, have a non-empty interior, and contain no lines. Other examples include the nonnegative orthant, the second-order cone, the cone of polynomials that can be expressed as sums-of-squares of polynomials of a given degree, and the cone of everywhere nonnegative polynomials of a given degree. An important feature of proper cones is that they induce a *partial ordering* analogous to the positive semidefinite ordering. Specifically, for a proper cone $\mathcal{K} \subset \mathbb{R}^n$, we denote the corresponding partial order by $\preceq_{\mathcal{K}}$ so that $\mathbf{u} \preceq_{\mathcal{K}} \mathbf{v}$ if and only if $\mathbf{v} - \mathbf{u} \in \mathcal{K}$. As with the previous example, we consider the following convex constraint in a decision variable $\mathbf{x} \in \mathbb{R}^n$ with $\mathbf{Q} \in \mathbb{R}^{n \times k}$, $\mathbf{C}^{(i)} \in \mathbb{R}^{\ell}$ for $i = 1, \dots, k$, $\mathbf{R}^{(j)} \in \mathbb{R}^{\ell}$ for $j = 1, \dots, n$, $\mathbf{S} \in \mathbb{R}^{\ell}$, and where each $\mathbf{C}^{(i)} \in \mathcal{K}$ for a proper cone $\mathcal{K} \subset \mathbb{R}^{\ell}$:

$$\sum_{i=1}^k \mathbf{C}^{(i)} \exp([\mathbf{Q}'\mathbf{x}]_i) + \sum_{j=1}^n \mathbf{R}^{(j)} \mathbf{x}_j + \mathbf{S} \preceq_{\mathcal{K}} \mathbf{0}.$$

Again, introducing additional variables $\mathbf{y} \in \mathbb{R}^{\ell}$ we can rewrite this constraint as follows:

$$\sum_{i=1}^k \mathbf{C}^{(i)} \mathbf{y}_i + \sum_{j=1}^n \mathbf{R}^{(j)} \mathbf{x}_j + \mathbf{S} \preceq_{\mathcal{K}} \mathbf{0}; \quad \exp\{\mathbf{Q}'\mathbf{x}\} \leq \mathbf{y}.$$

If the cone \mathcal{K} is tractable to represent so that checking membership in the cone can be accomplished efficiently, then these constraints can be tractably specified via CGP.

Inverses via Schur complement Based on the first example with matrix-weighted sums of exponentials, we now show that CGPs can also be used to express certain constraints involving inverses based on the Schur complement. Consider the following matrix-valued expressions in a variable $\mathbf{x} \in \mathbb{R}^n$:

$$\begin{aligned} F_{11}(\mathbf{x}) &= \sum_{i=1}^k \mathbf{C}_{11}^{(i)} \exp([\mathbf{Q}'_{11}\mathbf{x}]_i) + \sum_{j=1}^n \mathbf{R}_{11}^{(j)} \mathbf{x}_j + \mathbf{S}_{11} \\ F_{22}(\mathbf{x}) &= \sum_{i=1}^{\tilde{k}} \mathbf{C}_{22}^{(i)} \exp([\mathbf{Q}'_{22}\mathbf{x}]_i) + \sum_{j=1}^n \mathbf{R}_{22}^{(j)} \mathbf{x}_j + \mathbf{S}_{22} \\ F_{12}(\mathbf{x}) &= \sum_{j=1}^n \mathbf{R}_{12}^{(j)} \mathbf{x}_j + \mathbf{S}_{12}. \end{aligned}$$

As in the first example with matrix coefficients, we have that $\mathbf{Q}_{11} \in \mathbb{R}^{n \times k}$, $\mathbf{C}_{11}^{(i)} \in \text{Sym}(\ell)$ for $i = 1, \dots, k$; $\mathbf{Q}_{22} \in \mathbb{R}^{n \times \tilde{k}}$, $\mathbf{C}_{22}^{(i)} \in \text{Sym}(\tilde{\ell})$ for $i = 1, \dots, \tilde{k}$; $\mathbf{R}_{11}^{(j)} \in \text{Sym}(\ell)$, $\mathbf{R}_{22}^{(j)} \in \text{Sym}(\tilde{\ell})$, $\mathbf{R}_{12}^{(j)} \in \mathbb{R}^{\ell \times \tilde{\ell}}$ for $j = 1, \dots, n$; $\mathbf{S}_{11} \in \text{Sym}(\ell)$, $\mathbf{S}_{22} \in \text{Sym}(\tilde{\ell})$, $\mathbf{S}_{12} \in \mathbb{R}^{\ell \times \tilde{\ell}}$; and where each $\mathbf{C}_{11}^{(i)} \succeq \mathbf{0}$ and each $\mathbf{C}_{22}^{(i)} \succeq \mathbf{0}$. Then the following set of constraints can be expressed via CGP:

$$F_{11}(\mathbf{x}) \prec \mathbf{0} \quad \text{and} \quad F_{22}(\mathbf{x}) - F_{12}(\mathbf{x})F_{11}(\mathbf{x})^{-1}F_{12}(\mathbf{x})' \prec \mathbf{0}.$$

In particular, this set of constraints is equivalent based on the Schur complement to the single constraint:

$$\begin{bmatrix} F_{11}(\mathbf{x}) & F_{12}(\mathbf{x}) \\ F_{12}'(\mathbf{x}) & F_{22}(\mathbf{x}) \end{bmatrix} \prec \mathbf{0}$$

This constraint on a symmetric matrix expression of size $(\ell + \tilde{\ell}) \times (\ell + \tilde{\ell})$ can be specified using CGP by following the first example on matrix-weighted sums of exponentials. The Schur complement method allows us to bound expressions with inverses. For instance, consider the following set of constraints for $t > 0$:

$$\sum_{j=1}^k \frac{\left(\mathbf{r}^{(j)'} \mathbf{x} - \mathbf{s}_j\right)^2}{-f_j(\mathbf{x})} < t; \quad f_j(\mathbf{x}) < 0,$$

where each $f_j(\mathbf{x})$ may be an expression involving a sum of exponentials, affine functions, products of sums of exponentials, and logarithms of sums of exponentials (these latter two types of expressions may be transformed to combined exponential/affine from Section 3.1). Here each $\mathbf{r}^{(j)} \in \mathbb{R}^n$ and $\mathbf{s} \in \mathbb{R}^k$. One can express this system of constraints within the CGP framework by applying the Schur complement procedure outlined above.

3.3 Maximum over Sums of Exponentials

CGPs provide a flexible methodology to deal with constraints on the maximum over a collection of sums of exponentials. Such constraints arise prominently in robust GP in which one seeks solutions that are robust to uncertainty in the parameters of a GP. These types of constraints also arise in applications ranging from permanent maximization to hitting-time estimation in dynamical systems – see Section 4 for details. Formally, consider a maximum over a set of sums of exponentials in a decision variable $\mathbf{x} \in \mathbb{R}^n$ of the following type:

$$\begin{aligned} \sup_{\mathbf{c} \in \mathcal{C}, \mathbf{q}^{(i)} \in \mathcal{Q}^{(i)}} \sum_{i=1}^k \mathbf{c}_i \exp\left(\mathbf{q}^{(i)'} \mathbf{x}\right) &\leq t & (12) \\ \updownarrow \\ \forall \mathbf{c} \in \mathcal{C}, \mathbf{q}^{(i)} \in \mathcal{Q}^{(i)} \sum_{i=1}^k \mathbf{c}_i \exp\left(\mathbf{q}^{(i)'} \mathbf{x}\right) &\leq t \end{aligned}$$

Here $\mathcal{C} \subset \mathbb{R}_+^k$ and $\mathcal{Q}^{(i)} \subset \mathbb{R}^n$ for $i = 1, \dots, k$, and in the robust optimization literature these are usually referred to as parameter uncertainty sets. In principle, these constraints specify convex sets because they can be viewed as the intersection of a (possibly infinite) collection of convex constraints. When \mathcal{C} and each $\mathcal{Q}^{(i)}$ are finite sets, then constraints of the form (12) reduce to a finite collection of constraints on sums of exponentials. For the case in which \mathcal{C} and each $\mathcal{Q}^{(i)}$ are infinite, the challenge is to obtain a tractable representation for the corresponding constraint set via a small, finite number of constraints. The literature on robust optimization [4, 5] has considered this question in great detail in a number of settings; in Section 4.2 we discuss the previous state-of-the-art results for handling parameter uncertainty in sum-of-exponentials functions in the context of robust GPs, and we contrast these with the significant advance facilitated by CGPs.

In order to obtain tractable representations for constraints of the form (12), we focus here on sets \mathcal{C} and $\mathcal{Q}^{(i)}$ that are themselves tractable to represent. If these sets are convex, and in particular they have tractable conic representations, then CGPs provide a natural framework to obtain efficient representations for (12). The next result illuminates this point by appealing to convex duality:

Proposition 3. *Suppose $\mathcal{C} \subset \mathbb{R}^k$ is a convex set of the form:*

$$\mathcal{C} = \{\mathbf{c} \mid \mathbf{c} \geq \mathbf{0}, \mathbf{g} + \mathbf{F}\mathbf{c} \in \mathcal{K}\}$$

for a convex cone $\mathcal{K} \subset \mathbb{R}^\ell$ and $\mathbf{g} \in \mathbb{R}^\ell, \mathbf{F} \in \mathbb{R}^{\ell \times k}$, and $\mathcal{Q}^{(i)} \subset \mathbb{R}^n$ for $i = 1, \dots, k$ are each sets of the form:

$$\mathcal{Q}^{(i)} = \left\{ \mathbf{q}^{(i)} \mid \mathbf{h}^{(i)} + \mathbf{M}^{(i)} \mathbf{q}^{(i)} \in \mathcal{F}^{(i)} \right\}$$

Maximum over logarithms, products of sum-of-exponentials Proposition 3 is also useful for providing CGP representations of the maximum of a composite expression involving a sum of exponentials, a product of sums of exponentials, and a logarithm of sums of exponentials. By appealing to the techniques of Section 3.1, such a composite expression may be transformed to an expression formed only of a sum of exponentials and an affine function. As a result, uncertainty in the parameters of the exponentials of the original expression with products and logarithms can directly be interpreted as uncertainty in the parameters of the exponentials of the final expression involving only a sum of exponentials and an affine function. Consequently, Proposition 3 can be applied to this final expression, thus leading to a tractable formulation for parameter uncertainty in expressions with products and logarithms of sums of exponentials.

3.4 CGP-Representable Spectral Functions

An elegant result of Lewis [36], extending earlier work by Von Neumann [55], states that any convex function of a symmetric matrix that is invariant under conjugation by orthogonal matrices is a convex and permutation-invariant function of the eigenvalues. This result along with the next proposition provides a recipe for CGP representations of certain convex sets in the space of symmetric matrices.

Proposition 4. [6] *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function that is invariant under permutation of its argument, and let $g : \text{Sym}(n) \rightarrow \mathbb{R}$ be the convex function defined as $g(\mathbf{M}) = f(\lambda(\mathbf{M}))$. Here $\lambda(\mathbf{M})$ refers to the list of eigenvalues of M . Then we have that*

$$\begin{aligned} & g(\mathbf{M}) \leq t \\ & \quad \Updownarrow \\ \exists \mathbf{x} \in \mathbb{R}^n \text{ s.t.} & \quad f(\mathbf{x}) \leq t \\ & \quad \mathbf{x}_1 \geq \cdots \geq \mathbf{x}_n \\ & \quad s_r(\mathbf{M}) \leq \mathbf{x}_1 + \cdots + \mathbf{x}_r, \quad r = 1, \dots, n-1 \\ & \quad \text{trace}(\mathbf{M}) = \mathbf{x}_1 + \cdots + \mathbf{x}_n. \end{aligned}$$

By appealing to the fact that the function s_r is SDP-representable from (7), we obtain the next example of a CGP-representable set in the space of symmetric matrices.

GP-representable functions of spectra Consider any convex and permutation-invariant function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ that is representable via CGP. Then by appealing to Proposition 4, we have that the function $g : \text{Sym}(n) \rightarrow \mathbb{R}$ defined as $g(\mathbf{M}) = f(\lambda(\mathbf{M}))$ is also representable via CGP. A prominent special case of this class of CGP-representable functions is the Von-Neumann entropy of a matrix $\mathbf{M} \in \text{Sym}(n)$ with $\mathbf{M} \succeq \mathbf{0}$:

$$H_{vn}(\mathbf{M}) \triangleq -\text{trace}[\mathbf{M} \log(\mathbf{M})] = -\sum_{i=1}^n \lambda_i(\mathbf{M}) \log(\lambda_i(\mathbf{M})) = -D(\lambda(\mathbf{M}), \mathbf{1}). \quad (15)$$

Here $\lambda_i(\mathbf{M})$ refers to the i 'th largest eigenvalue of \mathbf{M} . In quantum information theory, the Von-Neumann entropy is typically considered for positive semidefinite matrices with trace equal to one (i.e., quantum density matrices), but the function itself is concave over the full cone of positive semidefinite matrices. As Von-Neumann entropy is concave and invariant under conjugation of the argument by any orthogonal matrix (i.e., it is a concave and permutation-invariant function of the eigenvalues of its argument), we can appeal to Proposition 4 to obtain a CGP representation. Specifically, as $-H_{vn}(\mathbf{M}) = D(\lambda(\mathbf{M}), \mathbf{1})$, sublevel sets of negative Von-Neumann entropy are representable via (the dual of a) CGP.

4 Applications

In the preceding sections we described the conceptual unification of GPs and of conic optimization problems such as SDPs provided by the CGP framework. In this section, we discuss the utility of CGPs in solving problems in a range of application domains; these problems fall outside the scope of GPs and of SDPs due to the limitations of each of these frameworks, as discussed in the introduction. All of the numerical experiments in this section employed the CVX modeling framework [24] and the SDPT3 solver [51].

4.1 Permanent Maximization

The permanent of a matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$ is defined as:

$$\text{perm}(\mathbf{M}) \triangleq \sum_{\sigma \in \mathfrak{S}_n} \prod_{i=1}^n M_{i, \sigma(i)}, \quad (16)$$

where \mathfrak{S}_n refers to the set of all permutations of the ordered set $[1, \dots, n]$. The permanent of a matrix is a quantity that arises in combinatorics as the sum over weighted perfect matchings in a bipartite graph [38], in geometry as the mixed volume of hyperrectangles [7], and in multivariate statistics in the computation of order statistics [2]. In this section we consider the problem of maximizing the permanent over a family of matrices. This problem has received some attention for families of positive semidefinite matrices with specified eigenvalues [16, 25], but all of these works have tended to seek analytical solutions for very special instances of this problem. Here we consider permanent maximization over general convex subsets of the set of nonnegative matrices. (Recall that SDPs are useful for maximizing the *determinant* of a symmetric positive definite matrix subject to affine and conic constraints [53].)

Problem 1. *Given a family of matrices $\mathcal{C} \in \mathbb{R}^{n \times n}$, find a matrix in \mathcal{C} with maximum permanent.*

Permanent maximization is relevant for designing a bipartite network in which the average weighted matching is to be maximized subject to additional topology constraints on the network. The question also arises in geometry in finding a configuration of a set of hyperrectangles so that their mixed volume is maximized. Even computing the permanent of a fixed matrix is a #P-hard problem and it is therefore considered to be intractable in general; accordingly a large body of research has investigated approximate computation of the permanent [3, 27, 33, 37]. The class of elementwise nonnegative matrices has particularly received attention as these most commonly arise in application domains (e.g., bipartite graphs with nonnegative edge weights). For such matrices, several *deterministic* polynomial-time approximation algorithms provide exponential-factor (in the size n of the matrix) approximations of the permanent, e.g. [37]. The approach in [37] describes a technique based on solving a GP to approximate the permanent of a nonnegative matrix. The approximation guarantees in [37] are based on the Van Der Waerden conjecture, which states that the matrix $\mathbf{M} = \frac{1}{n} \mathbf{1}\mathbf{1}'$ has the strictly smallest permanent among all $n \times n$ doubly stochastic matrices (nonnegative matrices with all row and column sums equal to one). This conjecture was proved originally in the 1980s [18, 19], and Gurvits recently gave a very simple and elegant proof of this result [26]. In what follows, we describe the approach of [37], but using the terminology developed by Gurvits in [26].

The permanent of a matrix $\mathbf{M} \in \mathbb{R}_+^{n \times n}$ can be defined in terms of a particular coefficient of the

following homogenous polynomial in $\mathbf{y} \in \mathbb{R}_+^n$:

$$p_{\mathbf{M}}(\mathbf{y}_1, \dots, \mathbf{y}_n) = \prod_{i=1}^n \left(\sum_{j=1}^n \mathbf{M}_{i,j} \mathbf{y}_j \right). \quad (17)$$

Specifically, the permanent of \mathbf{M} is the coefficient corresponding to the $\mathbf{y}_1 \cdots \mathbf{y}_n$ monomial term of $p_{\mathbf{M}}(\mathbf{y}_1, \dots, \mathbf{y}_n)$:

$$\text{perm}(\mathbf{M}) = \frac{\partial^n}{\partial \mathbf{y}_1 \cdots \partial \mathbf{y}_n} p_{\mathbf{M}}(\mathbf{y}_1, \dots, \mathbf{y}_n)$$

In his proof of Van Der Waerden's conjecture, Gurvits defines the *capacity* of a homogenous polynomial $p(\mathbf{y}_1, \dots, \mathbf{y}_n)$ of degree n defined over $\mathbf{y} \in \mathbb{R}_+^n$ as follows:

$$\begin{aligned} \text{cap}(p) &\triangleq \inf_{\mathbf{y} \in \mathbb{R}_+^n} \frac{p(\mathbf{y})}{\mathbf{y}_1 \cdots \mathbf{y}_n} \\ &= \inf_{\mathbf{y} \in \mathbb{R}_+^n, \mathbf{y}_1 \cdots \mathbf{y}_n = 1} p(\mathbf{y}). \end{aligned} \quad (18)$$

Gurvits then proves the following result:

Theorem 5. [26] *For any matrix $\mathbf{M} \in \mathbb{R}_+^{n \times n}$, we have that:*

$$\frac{n!}{n^n} \text{cap}(p_{\mathbf{M}}) \leq \text{perm}(\mathbf{M}) \leq \text{cap}(p_{\mathbf{M}}).$$

Here the polynomial $p_{\mathbf{M}}$ and its capacity $\text{cap}(p_{\mathbf{M}})$ are as defined in (17) and (18). Further if each column of \mathbf{M} has at most k nonzeros, then the factor in the lower bound can be improved from $\frac{n!}{n^n}$ to $\left(\frac{k-1}{k}\right)^{(k-1)n}$.

Gurvits in fact proves a more general statement involving so-called stable polynomials, but the above restricted version will suffice for our purposes. The upper-bound in this statement is straightforward to prove; it is the lower bound that is the key technical novelty. Thus, if one could compute the capacity of the polynomial $p_{\mathbf{M}}$ associated to a nonnegative matrix \mathbf{M} , then one can obtain an exponential-factor approximation of the permanent of \mathbf{M} as $\frac{n!}{n^n} \approx \exp(-n)$. In order to compute the capacity of $p_{\mathbf{M}}$ via GP, we apply the transformation $\mathbf{x} = \log\{\mathbf{y}\}$ in (18) and solve the following program²:

$$\begin{aligned} \text{cap}(p_{\mathbf{M}}) &= \inf_{\mathbf{x} \in \mathbb{R}^n} \prod_{i=1}^n \left(\sum_{j=1}^n \mathbf{M}_{i,j} \exp(\mathbf{x}_j) \right) \quad \text{s.t.} \quad \mathbf{1}'\mathbf{x} = 0 \\ &= \inf_{\mathbf{x} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^n, \delta \in \mathbb{R}} \exp(\delta) \quad \text{s.t.} \quad \mathbf{1}'\mathbf{x} = 0, \sum_{i,j=1}^n \mathbf{M}_{i,j} \exp(\mathbf{x}_j - \boldsymbol{\beta}_i - 1) + \mathbf{1}'\boldsymbol{\beta} \leq \delta. \end{aligned} \quad (19)$$

In obtaining the second formulation, we appeal to the expression (11) for CGP representations of upper-bound constraints on a product of sums of exponentials.

²Such a logarithmic transformation of the variables is also employed in converting a GP specified in terms of non-convex polynomial functions to a GP in convex form; see [11, 17] for more details.

Solution. Focusing on a set $\mathcal{C} \subset \mathbb{R}_+^{n \times n}$ of entrywise nonnegative matrices, our proposal to approximately maximize the permanent is to find $\mathbf{M} \in \mathcal{C}$ with maximum capacity $\text{cap}(p_{\mathbf{M}})$, which leads to the following consequence of Theorem 5:

Proposition 6. *Let $\mathcal{C} \in \mathbb{R}_+^{n \times n}$ be a set of nonnegative matrices, and consider the following two quantities:*

$$\begin{aligned}\hat{\mathbf{M}}_{\text{perm}} &= \arg \sup_{\mathbf{M} \in \mathcal{C}} \text{perm}(\mathbf{M}) \\ \hat{\mathbf{M}}_{\text{cap}} &= \arg \sup_{\mathbf{M} \in \mathcal{C}} \text{cap}(p_{\mathbf{M}})\end{aligned}$$

Then we have that

$$\frac{n!}{n^n} \text{perm}(\hat{\mathbf{M}}_{\text{perm}}) \leq \text{perm}(\hat{\mathbf{M}}_{\text{cap}}) \leq \text{perm}(\hat{\mathbf{M}}_{\text{perm}}).$$

The factor in the lower bound can be improved from $\frac{n!}{n^n}$ to $\left(\frac{k-1}{k}\right)^{(k-1)n}$ if every matrix in \mathcal{C} has at most k nonzeros in each column.

Proof. The proof follows from a direct application of Theorem 5:

$$\frac{n!}{n^n} \text{perm}(\hat{\mathbf{M}}_{\text{perm}}) \leq \frac{n!}{n^n} \text{cap}(p_{\hat{\mathbf{M}}_{\text{perm}}}) \leq \frac{n!}{n^n} \text{cap}(p_{\hat{\mathbf{M}}_{\text{cap}}}) \leq \text{perm}(\hat{\mathbf{M}}_{\text{cap}}) \leq \text{perm}(\hat{\mathbf{M}}_{\text{perm}}).$$

The first and third inequalities are a result of Theorem 5, and the second and fourth inequalities follow from the definitions of $\hat{\mathbf{M}}_{\text{cap}}$ and of $\hat{\mathbf{M}}_{\text{perm}}$ respectively. The improvement in the lower bound also follows from Theorem 5. \square

In summary, maximizing the capacity with respect to the family \mathcal{C} provides an approximation to the permanent-maximizing element of \mathcal{C} . Specifically, for sets \mathcal{C} that have tractable conic representations, the next result provides a CGP solution for capacity maximization:

Proposition 7. *Suppose that $\mathcal{C} \subset \mathbb{R}_+^{n \times n}$ is defined as:*

$$\mathcal{C} = \{\mathbf{M} \mid \mathbf{M} \geq \mathbf{0}, \mathbf{g} + \mathcal{L}(\mathbf{M}) \in \mathcal{K}\}$$

for $\mathcal{L} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^\ell$ a linear operator, for $\mathbf{g} \in \mathbb{R}^\ell$, and for $\mathcal{K} \subset \mathbb{R}^\ell$ a convex cone. Suppose further that \mathcal{C} is a compact subset, and that there exists a point that strictly satisfies the nonnegativity and conic constraints. Then the problem of maximizing capacity with respect to the set \mathcal{C} can be solved via CGP as follows:

$$\begin{aligned}\sup_{\mathbf{M} \in \mathcal{C}} \text{cap}(p_{\mathbf{M}}) &= \inf_{\mathbf{x} \in \mathbb{R}^n, \beta \in \mathbb{R}^n, \zeta \in \mathbb{R}^\ell} \exp(\mathbf{g}'\zeta + \mathbf{1}'\beta) \\ &\text{s.t. } [\mathcal{L}^\dagger(\zeta)]_{i,j} + \exp(\mathbf{x}_j - \beta_i - 1) \leq 0, \quad \forall i, j = 1, \dots, n \\ &\quad \mathbf{1}'\mathbf{x} = 0 \\ &\quad \zeta \in \mathcal{K}^*\end{aligned}$$

Here \mathcal{L}^\dagger refers to the adjoint of the operator \mathcal{L} .

Proof. By plugging in the expression (19) for the capacity of a polynomial, we have:

$$\begin{aligned}\sup_{\mathbf{M} \in \mathcal{C}} \log(\text{cap}(p_{\mathbf{M}})) &= \sup_{\mathbf{M} \in \mathcal{C}} \inf_{\mathbf{x}, \beta \in \mathbb{R}^n, \mathbf{1}'\mathbf{x}=0} \sum_{i,j=1}^n \mathbf{M}_{i,j} \exp(\mathbf{x}_j - \beta_i - 1) + \mathbf{1}'\beta \\ &= \inf_{\mathbf{x}, \beta \in \mathbb{R}^n, \mathbf{1}'\mathbf{x}=0} \sup_{\mathbf{M} \in \mathcal{C}} \sum_{i,j=1}^n \mathbf{M}_{i,j} \exp(\mathbf{x}_j - \beta_i - 1) + \mathbf{1}'\beta.\end{aligned}$$

In the second expression, the supremum and infimum are switched as the following conditions are satisfied [46]: the objective function is concave in \mathbf{M} for fixed $\mathbf{x}, \boldsymbol{\beta}$ and convex in $\mathbf{x}, \boldsymbol{\beta}$ for fixed \mathbf{M} , the constraints on \mathbf{M} and on $\mathbf{x}, \boldsymbol{\beta}$ are convex, and the constraint set \mathcal{C} for \mathbf{M} is compact. Given the particular form of \mathcal{C} and appealing to Proposition 3, we have the desired result. \square

The steps of this proof follow for the most part from the discussion in Section 3.1 on constraints on products of functions, each of which is a maximum of a sum of exponentials (see also Section 3.3). The main novelty in the proof here is the switching of the infimum and supremum, which is based on saddle-point theory from convex analysis [46].

4.2 Robust GP

As our next application, we describe the utility of CGP in addressing the problem of computing solutions to GPs that are robust to variability in the input parameters of the GP. Robust GPs arise in power control problems in communication systems as well as in robust digital circuit gate sizing, see e.g., [10].

Problem 2. *Given convex sets $\mathcal{C} \subset \mathbb{R}_+^k$ and $\mathcal{Q} \subset \mathbb{R}^{n \times k}$, obtain a tractable reformulation of the following constraint in a decision variable $\mathbf{x} \in \mathbb{R}^n$:*

$$\sup_{\mathbf{c} \in \mathcal{C}, \mathbf{Q} \in \mathcal{Q}} \sum_{i=1}^k \mathbf{c}_i \exp([\mathbf{Q}'\mathbf{x}]_i) \leq 1. \quad (20)$$

A robust GP is an optimization problem in which an affine function is minimized subject to affine constraints and constraints of the form (20). Such optimization problems in which one seeks solutions that are robust to parameter uncertainty have been extensively investigated in the field of robust optimization [4, 5], and exact, tractable reformulations of robust convex programs are available in a number of settings, e.g., for robust LPs. However, progress has been limited in the context of GPs for general convex uncertainty sets \mathcal{C}, \mathcal{Q} .

In their seminal work on robust convex optimization [5], Ben-Tal and Nemirovski obtained an exact, tractable reformulation of robust GPs in which a very restricted form of coefficient uncertainty is allowed in a sum-of-exponentials function – specifically, the set $\mathcal{Q} \subset \mathbb{R}^{n \times k}$ is a singleton set (no uncertainty) and $\mathcal{C} \subset \mathbb{R}_+^k$ must be a convex ellipsoid in \mathbb{R}_+^k specified by a quadratic form defined by an elementwise nonnegative matrix. The reformulation given in [5] for such robust GPs is itself a GP with additional variables, which can be solved efficiently.

In subsequent work, Hsiung et al [31] considered sum-of-exponentials functions with the coefficients absorbed into the exponent as follows:

$$\sup_{[\mathbf{d}, \mathbf{Q}] \in \mathcal{D}} \sum_{i=1}^k \exp([\mathbf{Q}'\mathbf{x} + \mathbf{d}]_i) \leq 1$$

Therefore, the uncertainty in the coefficients and the exponents is “coupled” via the set $\mathcal{D} \subset \mathbb{R}^{n \times (k+1)}$. For such constraints with \mathcal{D} being either a polyhedral set or an ellipsoidal set, Hsiung et al [31] obtain tractable but inexact reformulations via piecewise linear relaxations, with the reformulations again being GPs.

Solution. In order to develop a framework that allows for flexible specifications of uncertainty sets in robust GP, yet one in which we can obtain *exact, tractable* reformulations via CGP, we

consider the following general setting. Suppose the uncertainty set $\mathcal{C} \in \mathbb{R}_+^k$ is *any* convex set with a tractable conic representation, and suppose that $\mathcal{Q} \in \mathbb{R}^{n \times k}$ is decoupled columnwise as follows:

$$\mathcal{Q} = \{\mathbf{Q} \mid \mathbf{Q}^{(i)} \in \mathcal{Q}^{(i)}\}, \quad (21)$$

where $\mathbf{Q}^{(i)} \in \mathbb{R}^n$ is the i 'th column of \mathbf{Q} and $\mathcal{Q}^{(i)} \subset \mathbb{R}^n$ is *any* convex set with a tractable conic representation (this set specifies uncertainty in the i 'th column of \mathbf{Q} , independent of the other columns of \mathbf{Q}). For such forms of uncertainty, Proposition 3 provides a CGP reformulation of (20). Specifically, when the uncertainty sets \mathcal{C} and $\mathcal{Q}^{(i)}$ for $i = 1, \dots, k$ have tractable conic representations, we can obtain an *exact* reformulation of (20) that is efficiently computed via CGP.

In contrast to previous results on robust GP, note that the form of uncertainty for which we obtain an efficient and exact CGP reformulation is significantly more general than that considered in [5], in which \mathcal{C} can only be a restricted kind of ellipsoidal set and \mathcal{Q} must be a singleton set (no uncertainty). On the other hand, Hsiung et al [31] consider robust GPs in a different form and with polyhedral or ellipsoidal uncertainty that may be coupled across $\mathcal{Q}^{(i)}$, but their reformulation is inexact. A crucial distinction of our approach relative to those in [5] and in [31] is that our reformulation is a CGP, while those described in [5] and in [31] are GPs. Hence, the additional flexibility provided by CGPs via combined exponential/affine constraints and conic constraints – and the appearance of additional “lifting” variables in both these types of constraints – plays a crucial role in obtaining exact and tractable reformulations for fairly general types of uncertainty sets in robust GPs.

Example. To illustrate these distinctions concretely, consider a specific instance of a robust GP constraint. Suppose $\mathcal{C} \subset \mathbb{R}^k$ is a convex set of the form:

$$\mathcal{C} = \{\mathbf{c} \mid \mathbf{c} \geq \mathbf{0}, \mathbf{g} + \mathbf{F}\mathbf{c} \succeq \mathbf{0}\}$$

for $\mathbf{g} \in \text{Sym}(\ell)$ and $\mathbf{F} : \mathbb{R}^k \rightarrow \text{Sym}(\ell)$ a linear operator. Suppose also that each $\mathcal{Q}^{(i)} \subset \mathbb{R}^n$ for $i = 1, \dots, k$ is a set of the form:

$$\mathcal{Q}^{(i)} = \{\mathbf{q}^{(i)} \mid \mathbf{h}^{(i)} + \mathbf{M}^{(i)}\mathbf{q}^{(i)} \succeq \mathbf{0}\}$$

for $\mathbf{h}^{(i)} \in \text{Sym}(m)$ and $\mathbf{M}^{(i)} : \mathbb{R}^n \rightarrow \text{Sym}(m)$ a linear operator. With these uncertainty sets, we have from Proposition 3 that

$$\begin{aligned} \sup_{\mathbf{c} \in \mathcal{C}, \mathbf{Q}^{(i)} \in \mathcal{Q}^{(i)}} \sum_{i=1}^k \mathbf{c}_i \exp([\mathbf{Q}'\mathbf{x}]_i) &\leq 1 \\ \Updownarrow \\ \exists \boldsymbol{\zeta} \in \text{Sym}(\ell), \boldsymbol{\theta}^{(i)} \in \text{Sym}(m) \text{ for } i = 1, \dots, k \text{ s.t.} \\ \mathbf{g}'\boldsymbol{\zeta} &\leq 1, \boldsymbol{\zeta} \succeq \mathbf{0}, \\ \mathbf{M}^{(i)\dagger}(\boldsymbol{\theta}^{(i)}) + \mathbf{x} = \mathbf{0}, \exp\left(\mathbf{h}^{(i)'}\boldsymbol{\theta}^{(i)}\right) + [\mathbf{F}^\dagger(\boldsymbol{\zeta})]_i &\leq 0, \boldsymbol{\theta}^{(i)} \succeq \mathbf{0}, \text{ for } i = 1, \dots, k \end{aligned}$$

Here $\mathbf{M}^{(i)\dagger}$ and \mathbf{F}^\dagger represent the adjoint of the operators $\mathbf{M}^{(i)}$ and \mathbf{F} respectively. Note that in this formulation, the uncertainty sets \mathcal{C} and $\mathcal{Q}^{(i)}$ are essentially arbitrary SDP-representable sets. Such forms of robust GP cannot be handled by previous approaches [5, 31], but the expressive power of combined exponential/affine constraints and conic constraints enables us to obtain exact, tractable representations via CGP.

4.3 Hitting Times of Dynamical Systems

Consider a linear dynamical system with state-space equations as follows:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t), \quad (22)$$

where the state $\mathbf{x}(t) \in \mathbb{R}^n$ for $t \geq 0$. We assume throughout this section that the transition matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is diagonal; otherwise, one can always change to the appropriate *modal coordinates* given by the eigenvectors of \mathbf{A} (assuming \mathbf{A} is diagonalizable). The diagonal entries of \mathbf{A} are called the *modes* of the system. Suppose that the parameters of the dynamical system can take on a range of possible values with $\mathbf{A} \in \mathcal{A}$ and $\mathbf{x}(0) \in \mathcal{X}_{\text{initial}}$; the set \mathcal{A} specifies a set of possible modes and the set $\mathcal{X}_{\text{initial}} \subseteq \mathbb{R}^n$ specifies the set of possible initial conditions. Suppose also that we are given a *target set* $\mathcal{X}_{\text{target}} \subseteq \mathbb{R}^n$, and we wish to find the smallest time required for the system to reach a state in $\mathcal{X}_{\text{target}}$ from an arbitrary initial state in $\mathcal{X}_{\text{initial}}$. Formally, we define the *worst-case hitting time* of the dynamical system (22) to be:

$$\tau(\mathcal{X}_{\text{initial}}, \mathcal{X}_{\text{target}}, \mathcal{A}) \triangleq \inf \left\{ t \mid t \geq 0 \text{ and } \mathbf{x}(0) \exp \{ \mathbf{A}t \} \in \mathcal{X}_{\text{target}} \text{ for all } \mathbf{x}(0) \in \mathcal{X}_{\text{initial}}, \mathbf{A} \in \mathcal{A} \right\}. \quad (23)$$

Indeed, for an initial state $\mathbf{x}(0)$, the state of the system (22) at time t is given by $\mathbf{x}(t) = \mathbf{x}(0) \exp \{ \mathbf{A}t \}$; consequently, the quantity $\tau(\mathcal{X}_{\text{initial}}, \mathcal{X}_{\text{target}}, \mathcal{A})$ represents the amount of time that the worst-case trajectory of the system, taken over all initial conditions $\mathcal{X}_{\text{initial}}$ and mode values \mathcal{A} , requires to enter the target set $\mathcal{X}_{\text{target}}$.

Problem 3. *Compute the hitting time $\tau(\mathcal{X}_{\text{initial}}, \mathcal{X}_{\text{target}}, \mathcal{A})$ (23) of the linear dynamical system (22) with the possible initial states, the target set, and the possible set of modes given by $\mathcal{X}_{\text{initial}}$, $\mathcal{X}_{\text{target}}$, and \mathcal{A} respectively.*

Hitting times are of interest in system analysis and verification [43]. As an example, suppose that a system has the property that $\tau(\mathcal{X}_{\text{initial}}, \mathcal{X}_{\text{target}}, \mathcal{A}) = \infty$; this provides a proof that from certain initial states in $\mathcal{X}_{\text{initial}}$, the system never enters the target state $\mathcal{X}_{\text{target}}$. On the other hand, if $\tau(\mathcal{X}_{\text{initial}}, \mathcal{X}_{\text{target}}, \mathcal{A}) = 0$, we have a certificate that $\mathcal{X}_{\text{target}} \subseteq \mathcal{X}_{\text{initial}}$. While verification of linear systems has been extensively studied via Lyapunov and barrier certificate methods, the study of hitting times has received relatively little attention (with a few exceptions such as [57]). In particular, the approaches in [57] can lead to loose bounds as the worst-case hitting time is computed based on box-constrained outer approximations of $\mathcal{X}_{\text{initial}}$ and $\mathcal{X}_{\text{target}}$.

Solution. In this section we show that for a particular class of these problems, the hitting time can be computed *exactly* via a CGP. Specifically, we make the following assumptions regarding the structure of the dynamical system (22):

- The set of modes is given by

$$\mathcal{A} = \{ \mathbf{A} \mid \mathbf{A} \text{ diagonal with } \mathbf{A}_{j,j} \in [\ell_j, u_j] \ \forall j = 1, \dots, n \}$$

- The set of initial states $\mathcal{X}_{\text{initial}} \subseteq \mathbb{R}_+^n$ is given by a convex set with a tractable representation via affine and conic constraints. In particular, as in Proposition 3, $\mathcal{X}_{\text{initial}}$ is specified as follows:

$$\mathcal{X}_{\text{initial}} = \{ \mathbf{x} \in \mathbb{R}_+^n \mid \mathbf{g} + \mathbf{F}\mathbf{x} \in \mathcal{K} \}.$$

Here $\mathbf{g} \in \mathbb{R}^m$, $\mathbf{F} \in \mathbb{R}^{m \times n}$, and the cone $\mathcal{K} \subset \mathbb{R}^m$ is assumed to be tractable to compute.

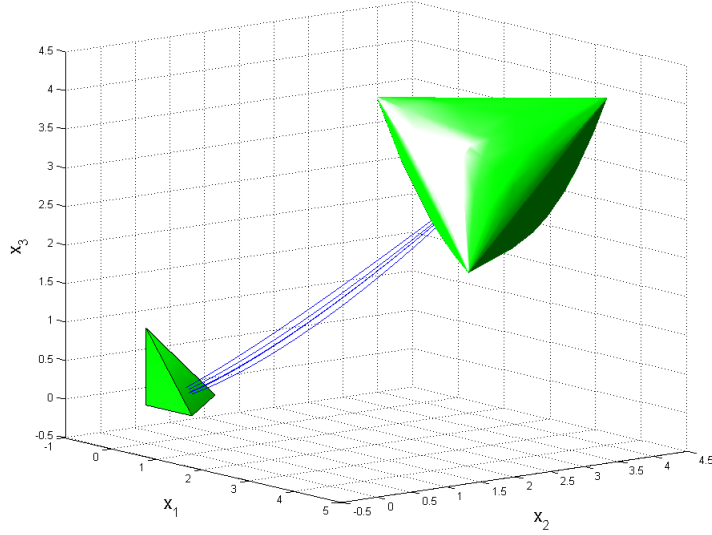


Figure 3: Some sample trajectories of a linear system from $\mathcal{X}_{\text{initial}}$ to $\mathcal{X}_{\text{target}}$ for the dynamical system described in the example in Section 4.3. The (shifted) ellipsope specifies the set of feasible starting points, while the tetrahedron specifies the target set. The system consists of three modes.

- $\mathcal{X}_{\text{target}} \subset \mathbb{R}^n$ is representable as the intersection of finitely many half spaces:

$$\mathcal{X}_{\text{target}} = \{\mathbf{x} \in \mathbb{R}_+^n \mid \mathbf{c}^{(i)'} \mathbf{x} \leq \mathbf{d}_i, i = 1, \dots, k, \text{ where each } \mathbf{c}^{(i)} \in \mathbb{R}_+^n, \mathbf{d} \in \mathbb{R}_+^k\}.$$

Under these conditions, the worst-case hitting time $\tau(\mathcal{X}_{\text{initial}}, \mathcal{X}_{\text{target}}, \mathcal{A})$ can be computed exactly via CGP. In particular, note that $\tau(\mathcal{X}_{\text{initial}}, \mathcal{X}_{\text{target}}, \mathcal{A})$ can be written as:

$$\tau(\mathcal{X}_{\text{initial}}, \mathcal{X}_{\text{target}}, \mathcal{A}) = \inf_{t \geq 0} t \quad \text{s.t.} \quad \sup_{\mathbf{x}(0) \in \mathcal{X}_{\text{initial}}, \mathbf{A} \in \mathcal{A}} \sum_{j=1}^n \mathbf{c}_j^{(i)} \mathbf{x}_j(0) \exp(\mathbf{A}_{j,j} t) \leq \mathbf{d}_i, i = 1, \dots, k. \quad (24)$$

Each constraint here can be reformulated as a collection of CGP-representable constraints via Proposition 3. Hence, for the particular problem structure we consider, the hitting time $\tau(\mathcal{X}_{\text{initial}}, \mathcal{X}_{\text{target}}, \mathcal{A})$ can be computed exactly via CGP.

Example. Consider a dynamical system with three modes that each take on values in certain ranges as follows:

$$\mathcal{A} = \{\mathbf{A} \in \mathbb{R}^{3 \times 3} \mid \mathbf{A} \text{ diagonal}, \mathbf{A}_{1,1} \in [-0.4, -0.45], \mathbf{A}_{2,2} \in [-0.5, -0.6], \mathbf{A}_{3,3} \in [-0.6, -0.7]\}.$$

The set of initial states is a shifted *ellipsope* that is contained within the nonnegative orthant:

$$\mathcal{X}_{\text{initial}} = \left\{ \mathbf{x}(0) \in \mathbb{R}_+^3 : \begin{bmatrix} 1 & \mathbf{x}_1(0) - 3 & \mathbf{x}_2(0) - 3 \\ \mathbf{x}_1(0) - 3 & 1 & \mathbf{x}_3(0) - 3 \\ \mathbf{x}_2(0) - 3 & \mathbf{x}_3(0) - 3 & 1 \end{bmatrix} \succeq 0 \right\}.$$

Finally, we let the target region be a tetrahedron:

$$\mathcal{X}_{\text{target}} = \{\mathbf{x} \in \mathbb{R}_+^3 \mid \mathbf{1}'\mathbf{x} \leq 1\}.$$

For these system attributes, a few sample trajectories are shown in Figure 3. We solve the CGP (24) and obtain that $\tau(\mathcal{X}_{\text{initial}}, \mathcal{X}_{\text{target}}, \mathcal{A}) = 7.6253$ is the worst-case hitting time.

4.4 Quantum Channel Capacity

A quantum channel transmits quantum information, and every such channel has an associated *quantum capacity* that characterizes the maximum rate at which quantum information can be reliably communicated across the channel. We refer the reader to the literature for an overview of this topic [30, 41, 47, 50]. For our purposes, we give a mathematical description of a quantum channel, and we describe methods for obtaining bounds on the associated quantum capacity via CGP. The input and the output of a quantum channel are quantum states that are specified by positive semidefinite matrices with unit trace. Such matrices are called *quantum density matrices*. Formally, a quantum channel is characterized by a linear operator that maps density matrices to density matrices; the dimensions of the input and output density matrices may, in general, be different:

$$\mathcal{L} : \text{Sym}(n) \rightarrow \text{Sym}(k)$$

For the linear operator \mathcal{L} to specify a valid quantum channel, it must map positive semidefinite matrices in $\text{Sym}(n)$ to positive semidefinite matrices in $\text{Sym}(k)$ and it must be trace-preserving. These properties ensure that density matrices are mapped to density matrices. In addition, \mathcal{L} must satisfy a property known as complete-positivity, which is required by the postulates of quantum mechanics – see [41] for more details. Any linear map $\mathcal{L} : \text{Sym}(n) \rightarrow \text{Sym}(k)$ that satisfies all these properties can be expressed via matrices³ $\mathbf{A}^{(j)} \in \mathbb{R}^{k \times n}$ as follows [28]:

$$\mathcal{L}(\rho) = \sum_j \mathbf{A}^{(j)} \rho \mathbf{A}^{(j)'} \quad \text{where} \quad \sum_j \mathbf{A}^{(j)'} \mathbf{A}^{(j)} = \mathbf{I}.$$

Letting \mathbb{S}^n denote the n -sphere, the capacity of the channel specified by $\mathcal{L} : \text{Sym}(n) \rightarrow \text{Sym}(k)$ is defined as:

$$C(\mathcal{L}) \triangleq \sup_{\mathbf{v}^{(i)} \in \mathbb{S}^n, p_i \geq 0, \sum_i p_i = 1} H_{vn} \left[\mathcal{L} \left(\sum_i p_i \mathbf{v}^{(i)} \mathbf{v}^{(i)'} \right) \right] - \sum_i p_i H_{vn} \left[\mathcal{L} \left(\mathbf{v}^{(i)} \mathbf{v}^{(i)'} \right) \right]. \quad (25)$$

Here H_{vn} is the Von-Neumann entropy (15) and the number of states $\mathbf{v}^{(i)}$ is part of the optimization.

Problem 4. *Given a quantum channel specified by linear map \mathcal{L} , compute the capacity of the channel.*

We note that there are several kinds of capacities associated with a quantum channel, depending on the protocol employed for encoding and decoding states transmitted across the channel; the version considered here is the one most commonly investigated, and we refer the reader to Shor's survey [50] for details of the other types of capacities. The quantity (25) on which we focus here is called the $C_{1,\infty}$ quantum capacity – roughly speaking, it is the capacity of a quantum channel in which the sender cannot couple inputs across multiple uses of the channel, while the receiver can jointly decode messages received over multiple uses of the channel. Shor also describes in [50] a procedure based on nonlinear programming to provide lower bounds on this quantum capacity. As the first step of this method, one fixes a finite set of states $\{\mathbf{v}^{(i)}\}_{i=1}^m$ with each $\mathbf{v}^{(i)} \in \mathbb{S}^n$ and a density matrix $\rho \in \text{Sym}(n)$, so that ρ is in the convex hull of $\{\mathbf{v}^{(i)} \mathbf{v}^{(i)'}\}_{i=1}^m$. With these quantities fixed, one can obtain a lower bound on $C(\mathcal{L})$ by solving the following *linear program*:

$$C \left(\mathcal{L}, \left\{ \mathbf{v}^{(i)} \right\}_{i=1}^m, \rho \right) = \sup_{p_i \geq 0, \sum_{i=1}^m p_i = 1, \rho = \sum_{i=1}^m p_i \mathbf{v}^{(i)} \mathbf{v}^{(i)'}} H_{vn}(\mathcal{L}(\rho)) - \sum_{i=1}^m p_i H_{vn} \left(\mathcal{L} \left(\mathbf{v}^{(i)} \mathbf{v}^{(i)'} \right) \right). \quad (26)$$

³In full generality, density matrices are trace-one, positive semidefinite *Hermitian* matrices, and $\mathbf{A}^{(j)} \in \mathbb{C}^{k \times n}$. As with SDPs, the CGP framework can handle linear matrix inequalities on Hermitian matrices, but we stick with the real case for simplicity.

In [50], Shor also suggests local heuristics to search for better sets of states and density matrices to improve the lower bound (26). It is clear that $C\left(\mathcal{L}, \{\mathbf{v}^{(i)}\}_{i=1}^m, \boldsymbol{\rho}\right)$ as computed in (26) is a lower bound on $C(\mathcal{L})$. Indeed, one can check that:

$$C(\mathcal{L}) = \sup_{\mathbf{v}^{(i)} \in \mathbb{S}^n} \sup_{\boldsymbol{\rho} \in \text{conv}\{\mathbf{v}^{(i)}\mathbf{v}^{(i)'}\}} C\left(\mathcal{L}, \{\mathbf{v}^{(i)}\}, \boldsymbol{\rho}\right). \quad (27)$$

Here the number of states is not explicitly denoted, and it is also a part of the optimization.

Solution. We describe an improvement upon the lower bound (26) by proposing a solution based on a CGP. Specifically, we observe that for a fixed set of states $\{\mathbf{v}^{(i)}\}_{i=1}^m$ the following optimization problem is a CGP:

$$\begin{aligned} C\left(\mathcal{L}, \{\mathbf{v}^{(i)}\}_{i=1}^m\right) &= \sup_{\boldsymbol{\rho} \in \text{conv}\{\mathbf{v}^{(i)}\mathbf{v}^{(i)'}\}_{i=1}^m} C\left(\mathcal{L}, \{\mathbf{v}^{(i)}\}_{i=1}^m, \boldsymbol{\rho}\right) \\ &= \sup_{p_i \geq 0, \sum_{i=1}^m p_i = 1} H_{vn} \left[\mathcal{L} \left(\sum_{i=1}^m p_i \mathbf{v}^{(i)} \mathbf{v}^{(i)'} \right) \right] - \sum_{i=1}^m p_i H_{vn} \left[\mathcal{L} \left(\mathbf{v}^{(i)} \mathbf{v}^{(i)'} \right) \right]. \end{aligned} \quad (28)$$

In particular, the optimization over density matrices $\boldsymbol{\rho} \in \text{conv}\{\mathbf{v}^{(i)}\mathbf{v}^{(i)'}\}_{i=1}^m$ in (27) can be folded into the computation of (26) at the expense of solving a CGP instead of an LP. It is easily seen from (25), (27), and (28) that for a fixed set of states $\{\mathbf{v}^{(i)}\}_{i=1}^m$:

$$C(\mathcal{L}) \geq C\left(\mathcal{L}, \{\mathbf{v}^{(i)}\}_{i=1}^m\right) \geq C\left(\mathcal{L}, \{\mathbf{v}^{(i)}\}_{i=1}^m, \boldsymbol{\rho}\right).$$

For a finite set of states $\{\mathbf{v}^{(i)}\}_{i=1}^m$, the quantity $C\left(\mathcal{L}, \{\mathbf{v}^{(i)}\}_{i=1}^m\right)$ in (28) is referred to as the *classical-to-quantum capacity* with respect to the states $\{\mathbf{v}^{(i)}\}_{i=1}^m$ [30, 47]; the reason for this name is that $C\left(\mathcal{L}, \{\mathbf{v}^{(i)}\}_{i=1}^m\right)$ is the capacity of a quantum channel in which the input is “classical” as it is restricted to be a convex combination of the finite set $\{\mathbf{v}^{(i)}\mathbf{v}^{(i)'}\}_{i=1}^m$.

One can improve upon the bound $C\left(\mathcal{L}, \{\mathbf{v}^{(i)}\}_{i=1}^m\right)$ by progressively adding more states to the collection $\{\mathbf{v}^{(i)}\}_{i=1}^m$. It is of interest to develop techniques to accomplish this in a principled manner.

Example. We consider a quantum channel given by a linear operator $\mathcal{L} : \text{Sym}(8) \rightarrow \text{Sym}(8)$ with:

$$\mathcal{L}(\boldsymbol{\rho}) = \mathbf{A}^{(1)} \boldsymbol{\rho} \mathbf{A}^{(1)'} + \epsilon^2 \mathbf{A}^{(2)} \boldsymbol{\rho} \mathbf{A}^{(2)'} + \mathbf{A}^{(3)} \boldsymbol{\rho} \mathbf{A}^{(3)'}, \quad (29)$$

where $\mathbf{A}^{(1)}$ is a random 8×8 diagonal matrix (entries chosen to be i.i.d. standard Gaussian), $\mathbf{A}^{(2)}$ is a random 8×8 matrix (entries chosen to be i.i.d. standard Gaussian), $\epsilon \in [0, 1]$, and $\mathbf{A}^{(3)}$ is the symmetric square root of $\mathbf{I} - \mathbf{A}^{(1)'} \mathbf{A}^{(1)} - \epsilon^2 \mathbf{A}^{(2)'} \mathbf{A}^{(2)}$. We scale $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ suitably so that $\mathbf{I} - \mathbf{A}^{(1)'} \mathbf{A}^{(1)} - \epsilon^2 \mathbf{A}^{(2)'} \mathbf{A}^{(2)}$ is positive definite for all $\epsilon \in [0, 1]$. For this quantum channel, we describe two numerical experiments.

In the first experiment, we compute the classical-to-quantum capacity of this channel with the input states $\mathbf{v}^{(i)} \in \mathbb{S}^8$ for $i = 1, \dots, 8$ being the 8 standard basis vectors in \mathbb{R}^8 . In other words, the input density $\boldsymbol{\rho} \in \text{Sym}(8)$ is given by a diagonal matrix. We add unit vectors in \mathbb{R}^8 at random

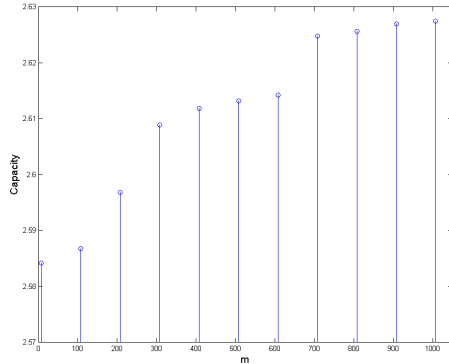


Figure 4: A sequence of increasingly tighter lower bounds on the quantum capacity of the channel specified by (29), obtained by computing classical-to-quantum capacities with increasingly larger collections of input states.

to this collection, and we plot the increase in Figure 4 in the corresponding classical-to-quantum channel capacity (28) – in each case, the capacity is evaluated at $\epsilon = 0.5$. In this manner, CGPs can be used to provide successively tighter lower bounds for the capacity of a quantum channel.

Next, we compare the classical-to-quantum capacity of the channel (29) – input states $\mathbf{v}^{(i)} \in \mathbb{S}^8$ for $i = 1, \dots, 8$ being the 8 standard basis vectors in \mathbb{R}^8 – to the capacity of a purely classical channel induced by the quantum channel (29). Specifically, let $\mathfrak{P}_{\text{diag}} : \text{Sym}(8) \rightarrow \text{Sym}(8)$ denote an operator that projects a matrix onto the space of diagonal matrices, i.e., zeros out the off-diagonal entries. Then we consider the following classical channel induced by the quantum channel (29) for a diagonal density matrix ρ :

$$\mathcal{L}_{\text{classical}}(\rho) \triangleq \mathfrak{P}_{\text{diag}}(\mathcal{L}(\rho)). \quad (30)$$

Both the inputs and outputs of this channel are diagonal density matrices, and they can therefore be interpreted as classical distributions. Figure 5 shows two plots for ϵ ranging from 0 to 1 of both the classical-to-quantum capacity of (29) and the classical capacity of the induced classical channel (30). Note that for $\epsilon = 0$ the output of the operator (29) is diagonal if the input is given by a diagonal density matrix. Therefore, the two curves coincide at $\epsilon = 0$ in Figure 5. For larger values of ϵ , the classical-to-quantum channel is greater than the capacity of the induced classical channel.

4.5 Other Applications involving Von-Neumann Entropy Optimization

Several previously proposed methods in varied application domains can be viewed as special cases of CGP. Specifically, the problem of maximizing Von-Neumann entropy of a positive semidefinite matrix subject to affine constraints on the matrix has been recognized as a useful conceptual and computational approach in a number of areas beyond the quantum communication setting of Section 4.4. We briefly discuss some of these applications here. To be clear, although the proposed solutions in these domains can be viewed as particular instances of CGP, the framework developed in our paper is substantially more general and CGPs facilitate many new applications as described above.

Quantum state tomography Quantum state tomography arises in the characterization of optical systems and in quantum computation [22]. The goal is to reconstruct a quantum state specified by a density matrix (symmetric positive semidefinite with unit trace) given partial information

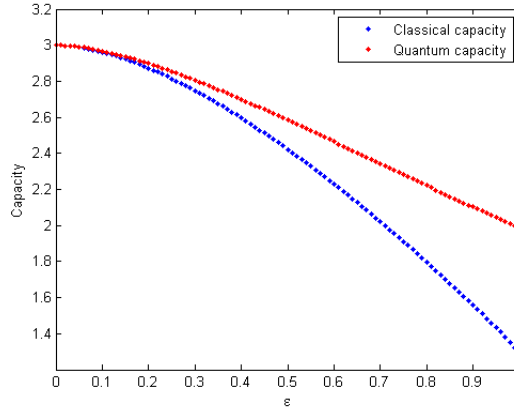


Figure 5: Comparison of a classical-to-quantum capacity of the quantum channel specified by (29) and the classical capacity of a classical channel induced by the quantum channel given by (29).

about the state. Such information is typically provided via *measurements* of the state, which can be viewed as linear functionals of the underlying density matrix. As measurements can frequently be expensive to obtain, this is a type of inverse problem in which one must choose among the many density matrices that satisfy the limited measurement information that is acquired. One proposal for tomography, based on the original work of Jaynes [32] on the maximum entropy principle, is to find the Von-Neumann-entropy-maximizing density matrix among all density matrices consistent with the measurements [22] – the rationale behind this approach is that the entropy-maximizing matrix makes the least assumptions about the quantum state beyond the constraints imposed by the acquired measurements. Using this method, the optimal reconstructed quantum state can be computed as the solution of a convex program that is a special instance of a CGP.

Equilibrium densities in statistical physics In statistical mechanics, a basic objective is to investigate the properties of a system at equilibrium given information about macroscopic attributes of the system such as energy, number of particles, volume, and pressure. In mathematical terms, the situation is quite similar to the previous setting with quantum state tomography – specifically, the system is described by a density matrix (as in quantum state tomography, this matrix is symmetric positive semidefinite with unit trace), and the macroscopic properties can be characterized as constraints on linear functionals of this density matrix. The Massieu-Planck extremum principle in statistical physics states that the system at equilibrium is given by the Von-Neumann-entropy-maximizing density matrix among all density matrices that satisfy the specified constraints [48]. As before, this equilibrium density can be computed efficiently via CGP.

Kernel learning A commonly encountered problem in machine learning is to measure similarity or affinity between entities that may not belong to a Euclidean space, e.g., text documents. A first step in such settings is to specify coordinates for the entities in a linear vector space after performing some nonlinear mapping, and then subsequently computing distances in the linear space. Kernel methods approach this problem implicitly by working directly with inner-products between pairs of entities, thus combining the nonlinear mapping and the distance computation in a single step. Therefore, given a finite collection of entities, a kernel is a symmetric positive definite matrix in which each entry specifies the inner-product between the corresponding pair of entities. This approach leads to the natural question of learning a good kernel matrix. Kulis et al [35] propose to

minimize the following *Von-Neumann relative entropy* between a kernel $\mathbf{M} \in \text{Sym}(n)$ (the decision variable) and a given reference kernel $\mathbf{M}_0 \in \text{Sym}(n)$ (specifying prior knowledge about the affinities among the n entities) with $\mathbf{M}_0 \succ \mathbf{0}$:

$$D_{vn}(\mathbf{M}, \mathbf{M}_0) = -H_{vn}(\mathbf{M}) - \text{trace}(\mathbf{M} \log(\mathbf{M}_0)).$$

This minimization is subject to the decision variable \mathbf{M} being positive definite as well as constraints on \mathbf{M} that impose bounds on distances between pairs of entities in the linear space. We refer the reader to [35] for more details, and for the virtues of minimizing the Von-Neumann relative entropy as a method for learning kernel matrices. In the context of this paper, we note that this problem can be viewed as a Von-Neumann entropy maximization problem subject to linear constraints, and it is therefore again a special instance of CGP.

In summary, we note that each of these problems can be viewed as a very special case of a CGP. As such, the approach described in Section 2.3 provides an effective method for solving these problems. Although we do not discuss the details here, it is possible to improve upon the general techniques in Section 2.3 by exploiting the particular structure of Von-Neumann entropy maximization problems subject to affine and semidefinite constraints.

5 Discussion

In this paper we introduced and investigated CGPs, a new class of structured convex optimization problems obtained by blending aspects of GPs and conic programs such as SDPs. CGPs consist of conic constraints, affine constraints, and constraints on sums of exponential and affine functions. The dual of a CGP can be expressed in a canonical manner as the joint maximization of the negative relative entropy between two nonnegative vectors, subject to affine and conic constraints. CGPs can be solved efficiently in practice, and they expand significantly upon the expressive power of GPs and SDPs. We demonstrate this point by describing a range of problems that are outside the scope of SDPs and GPs alone but for which CGPs provide effective solutions. The application domains include permanent maximization, quantum capacity computation, hitting-time estimation in dynamical systems, and robust solutions of GPs.

There are several avenues for further research suggested by this paper. It is of interest to develop efficient first-order methods to solve CGPs in order to scale to large problem instances. Such massive size problems are especially prevalent in data analysis tasks, and are of interest in settings such as kernel learning. On a related note, there exists a vast literature on exploiting the structure of a particular problem instance of an SDP or a GP – e.g., sparsity in the problem parameters – which can result in significant computational speedups in practice in the solution of these problems. A similar set of techniques would be useful and relevant in all of the applications described in this paper. Finally, we seek a deeper understanding of the geometry of CGP-representable convex sets such as their face structure, building on the extensive research in recent years on the geometry of SDP-representable convex sets [8, 23, 29].

Acknowledgements

The authors would like to thank Pablo Parrilo for many enlightening conversations, and Leonard Schulman for pointers to the literature on Von-Neumann entropy.

References

- [1] F. Alizadeh. Interior-point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5:13–51, 1995.
- [2] R. B. Bapat and M. I. Beg. Order statistics for nonidentically distributed variables and permanents. *Sankhya: The Indian Journal of Statistics, A*, 51:79–93, 1989.
- [3] A. I. Barvinok. Computing mixed discriminants, mixed volumes, and permanents. *Discrete and Computational Geometry*, 18:205–237, 1997.
- [4] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton University Press, 2009.
- [5] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23:769–805, 1998.
- [6] A. Ben-Tal and A. Nemirovskii. *Lectures on Modern Convex Optimization*. Society for Industrial and Applied Mathematics, 2001.
- [7] U. Betke. Mixed volumes of polytopes. *Archiv der Mathematik*, 58:388–391, 1992.
- [8] G. Blekherman, P. Parrilo, and R. Thomas. *Semidefinite Optimization and Convex Algebraic Geometry*. Society for Industrial and Applied Mathematics, 2013.
- [9] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. Society for Industrial and Applied Mathematics, 1994.
- [10] S. Boyd, S. J. Kim, D. Patil, and M. Horowitz. Digital circuit optimization via geometric programming. *Operations Research*, 53:899–932, 2005.
- [11] S. Boyd, S. J. Kim, L. Vandenberghe, and A. Hassibi. A tutorial on geometric programming. *Optimization and Engineering*, 8:67–127, 2007.
- [12] S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [13] M. Chiang. Geometric programming for communication systems. *Foundations and Trends in Communications and Information Theory*, 2:1–154, 2005.
- [14] M. Chiang and S. Boyd. Geometric programming duals of channel capacity and rate distortion. *IEEE Transactions on Information Theory*, 50:245–258, 2004.
- [15] J. J. Dinkel, G. A. Kochenberger, and S. N. Wong. Entropy maximization and geometric programming. *Environment and Planning A*, 9:419–427, 1977.
- [16] J. H. Drew and C. R. Johnson. The maximum permanent of a 3-by-3 positive semidefinite matrix, given the eigenvalues. *Linear and Multilinear Algebra*, 25:243–251, 1989.
- [17] R. J. Duffin, E. L. Peterson, and C. M. Zener. *Geometric Programming: Theory and Application*. John Wiley and Sons, 1967.
- [18] G. P. Egorychev. Proof of the Van der Waerden conjecture for permanents (english translation; original in russian). *Siberian Mathematical Journal*, 22:854–859, 1981.

- [19] D. I. Falikman. Proof of the Van der Waerden conjecture regarding the permanent of a doubly stochastic matrix (english translation; original in russian). *Mathematical Notes*, 29:475–479, 1981.
- [20] F. Glineur. An extended conic formulation for geometric optimization. *Foundations of Computing and Decision Sciences*, 25:161–174, 2000.
- [21] M. Goemans and D. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42:1115–1145, 1995.
- [22] D. S. Gonalves, C. Lavor, M. A. Gomes-Ruggiero, A. T. Cesrio, R. O. Vianna, and T. O. Maciel. Quantum state tomography with incomplete data: Maximum entropy and variational quantum tomography. *Physical Review A*, 87, 2013.
- [23] J. Gouveia, P. Parrilo, and R. Thomas. Lifts of convex sets and cone factorizations. *Mathematics of Operations Research*, 38:248–264, 2013.
- [24] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008.
- [25] R. Grone, C. R. Johnson, S. A. Eduardo, and H. Wolkowicz. A note on maximizing the permanent of a positive definite hermitian matrix, given the eigenvalues. *Linear and Multilinear Algebra*, 19:389–393, 1986.
- [26] L. Gurvits. Van der Waerden/Schrijver-Valiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: one theorem for all. *Electronic Journal of Combinatorics*, 15, 2008.
- [27] L. Gurvits and A. Samorodnitsky. A deterministic algorithm for approximating the mixed discriminant and mixed volume, and a combinatorial corollary. *Discrete and Computational Geometry*, 27:531–550, 2002.
- [28] K. Hellwig and K. Krauss. Operations and measurements II. *Communications of Mathematical Physics*, 16:142–147, 1970.
- [29] J. W. Helton and J. Nie. Sufficient and necessary conditions for semidefinite representability of convex hulls and sets. *SIAM Journal on Optimization*, 20:759–791, 2009.
- [30] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44:269–273, 1998.
- [31] K. L. Hsiung, S. J. Kim, and S. Boyd. Tractable approximate robust geometric programming. *Optimization and Engineering*, 9:95–118, 2008.
- [32] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review Series II*, 106:620–630, 1957.
- [33] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries. *Journal of the ACM*, 51:671–697, 2004.
- [34] D. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *Journal of the ACM*, 45:246–265, 1998.

- [35] B. Kulis, M. Sustik, and I. Dhillon. Low-rank kernel learning with bregman matrix divergences. *Journal of Machine Learning Research*, 10:341–376, 2009.
- [36] A. S. Lewis. The convex analysis of unitarily invariant matrix functions. *Journal of Convex Analysis*, 2:173–183, 1995.
- [37] N. Linial, A. Samorodnitsky, and A. Wigderson. A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents. *Combinatorica*, 20, 2000.
- [38] H. Minc. *Permanents*. Cambridge University Press, 1984.
- [39] Y. Nesterov and A. Nemirovski. *Interior-Point Polynomial Algorithms in Convex Programming*. Society of Industrial and Applied Mathematics, 1994.
- [40] Y. Nesterov and M. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM Journal on Optimization*, 8:324–364, 1998.
- [41] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2011.
- [42] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, 1999.
- [43] S. Prajna and A. Jadbabaie. Safety verification of hybrid systems using barrier certificates. *Hybrid Systems: Computation and Control, Springer Lecture Notes in Computer Science*, 2004.
- [44] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization. *SIAM Review*, 52:471–501, 2010.
- [45] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization*. Society for Industrial and Applied Mathematics, 1997.
- [46] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [47] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56:131–138, 1997.
- [48] C. H. Scott and T. R. Jefferson. Trace optimization problems and generalized geometric programming. *Journal of Mathematical Analysis and Applications*, 58:373–377, 1977.
- [49] A. Shapiro. Weighted minimum trace factor analysis. *Psychometrika*, 47:243–264, 1982.
- [50] P. W. Shor. Capacities of quantum channels and how to find them. *Mathematical Programming, B*, 97:311–335, 2003.
- [51] R. H. Tutuncu, K. C. Toh, and M. J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Mathematical Programming, Series B*, 95:189–217, 2003.
- [52] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38:49–95, 1994.
- [53] L. Vandenberghe, S. Boyd, and S. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19:499–533, 1998.
- [54] V. Vazirani. *Approximation Algorithms*. Springer, 2004.

- [55] J. Von Neumann. Some matrix-inequalities and metrization of matrix space. *Tomsk University Review*, 1:286–300, 1937.
- [56] T. Wall, D. Greening, and R. Woolsey. Solving complex chemical equilibria using a geometric programming based technique. *Operations Research*, 34:345–355, 1986.
- [57] H. Yazarel and G. Pappas. Geometric programming relaxations for linear system reachability. In *Proceedings of the 2004 American Control Conference*, 2004.