

# HYPERBOLIC RELAXATIONS FOR SEMIDEFINITE PROGRAMS WITH SPARSITY CONSTRAINTS

KEVIN SHU

**ABSTRACT.** Semidefinite programs are valuable computational methods for approximating quadratic programs with quadratic constraints. In addition to continuous constraints, it is often of interest to introduce constraints requiring that the final solution have only a certain number of nonzero entries. For instance, both linear regression and the maximum singular value problem can be formulated as quadratic programs, whose semidefinite programming relaxations are tight. The sparse versions of these problems, where the final output vector is required to have at most  $k$  nonzero entries are known to be hard to approximate. We introduce a family of tractible convex problems which approximate such sparse semidefinite programs, which arise naturally from the study of hyperbolic polynomials, which have been of recent interest. Our main contributions are formulations of these relaxations and computational methods solving these relaxations. We will give particular attention to the sparse linear regression problem and the sparse principal components analysis problem, obtaining particularly simple and efficient methods for approximating these expressions. We will also give computational evidence that these methods can be effective.

## 1. INTRODUCTION

The main object of interest in this paper optimization over linear slices of the *factor width  $k$*  cone, which is the convex cone defined as

$$\mathcal{FW}_k^n = \text{conv}\{xx^\top : x \in \mathbb{R}^n, \|x\|_0 \leq k\}$$

Here,  $\|x\|_0$  denotes the number of nonzero entries of  $x$ . Optimizing over such cones (when  $k$  is allowed to vary) can capture problems such as sparse linear regression [?] and sparse PCA [?], and as such it is NP-hard.

We will actually be interested in the more general *structured sparse semidefinite cones*, which are parameterized by a collection of subsets  $\Delta \subseteq 2^{[n]}$ , and are defined as

$$\mathcal{M}(\Delta) = \text{conv}\{xx^\top : x \in \mathbb{R}^n, \text{supp}(x) \in \Delta\}.$$

Here,  $\text{supp}(x) = \{i \in [n] : x_i \neq 0\}$ .

We may equivalently write the factor-width  $k$  cone in these terms:

$$\mathcal{FW}_k^n = \mathcal{M}\left(\left(\begin{bmatrix} n \\ k \end{bmatrix}\right)\right),$$

where  $\begin{bmatrix} n \\ k \end{bmatrix} = \{S \subseteq [n] : |S| = k\}$ . As can be seen in this example, we will often want to consider  $\Delta$  where the size of  $\Delta$  grows quickly with  $n$  and  $k$ .

The dual to  $\mathcal{M}(\Delta)$  can be seen to be

$$\mathcal{P}(\Delta) = \{X \in \mathbb{R}_{\text{sym}}^{n \times n} : \forall S \in \Delta, X|_S \succeq 0\}.$$

Here,  $X|_S$  denotes the principal submatrix of  $X$  obtained by restricting to the rows and columns of  $X$  indexed by elements of  $S$ .

In this paper, we will be primarily concerned with how these cones relate to a family of polynomials which are *linear combinations of principal minors* (LPM). Precisely, we say that a polynomial  $p$  with inputs in  $\mathbb{R}_{\text{sym}}^{n \times n}$  is LPM with support  $\Delta$  if it is of the form

$$p(X) = \sum_{S \in \Delta} a_S \det(X|_S)$$

for some coefficients  $a_S$ .

---

*Date:* May 27, 2022.

We would like to thank Greg Blekherman, Santanu Dey, Mohit Singh, Shengding Sun, Santosh Vempala for many insightful conversations.

We say that an LPM polynomial  $p(X)$  is homogeneous if every set in its support has the same size. A homogeneous LPM polynomial  $p(X)$  is said to be *hyperbolic* with respect to the identity if  $p(I) > 0$ , and for any  $X \in \mathbb{R}_{\text{sym}}^{n \times n}$ , all of the complex roots of the univariate polynomial

$$p_X(t) = p(X + tI)$$

are real. The study of hyperbolic LPM polynomials was begun in [?], which characterized the hyperbolic LPM polynomials.

These LPM polynomials are of interest because they define convex cones, known as *hyperbolicity cones*. The hyperbolicity cone of  $p$  can be defined as

$$\Lambda_I(p) = \{X \in \mathbb{R}_{\text{sym}}^{n \times n} : \forall t \geq 0, p(X + tI) \geq 0\}.$$

The fact that these cones are convex was first shown by Gårding in [?] when studying differential equations.

For our purposes, the main fact that is of interest is that if  $p$  is a hyperbolic LPM polynomial with support  $\Delta$ , then  $\mathcal{P}(\Delta)$  is contained in  $\Lambda_I(p)$ . Moreover, as long as we can efficiently compute the value of  $p$  at a matrix  $X$ , the cone  $\Lambda_I(p)$  is tractible, in the sense that we can apply interior point methods to solve convex optimization problems over  $\Lambda_I(p)$ . In this way, we can obtain tractible relaxations of  $\mathcal{P}(\Delta)$  as long as we can find an efficiently computable hyperbolic LPM polynomial with support  $\Delta$ .

This fact was used in [?] in the case when  $\Delta = \binom{[n]}{k}$ , and when

$$p(X) = \sum_{S \subseteq [n]: |S|=k} \det(X|_S).$$

Using this polynomial, the authors were able to obtain tight lower bounds on the minimum eigenvalue of a matrix in  $\mathcal{P}(\binom{[n]}{k})$ .

The main object of this paper is to apply this more general fact about LPM polynomials towards finding lower bounds on sparse semidefinite programs. We next describe the practical problems to which these ideas can be applied.

**1.1. Sparse Semidefinite Programs and Quadratic Programs.** A sparse QCQP is an optimization problem of the following form:[?]

$$\begin{aligned} (1) \quad & \max \quad q_0(x) \\ & \text{s.t.} \quad q_i(x) = a_i \text{ for } i \in \{1, \dots, k\} \\ & \quad x \in \mathbb{R}^n \\ & \quad \text{supp}(x) \in \Delta \end{aligned}$$

Here, each  $q_i$  denotes a quadratic polynomial in  $x$  of the form  $x^\top A_i x + b_i^\top x$ .

Associated to such a sparse QCQP is its *Schur relaxation*[?]:

$$\begin{aligned} (2) \quad & \max \quad \text{tr}(A_0 X) + b_0^\top x \\ & \text{s.t.} \quad \text{tr}(A_i X) + b_i^\top x = a_i \text{ for } i \in [k] \\ & \quad \begin{pmatrix} X & x \\ x^\top & 1 \end{pmatrix} \in \mathcal{M}(\Delta') \end{aligned}$$

Here,  $\Delta'$  consists of those  $S \subseteq [n+1]$  so that  $n+1 \in S$ , and  $S \cap [n] \in \Delta$ .

It can be seen that if  $x$  is a feasible solution to Equation (1), then letting  $X = xx^\top$  yields a feasible solution to Equation (2) with the same value. Thus, this is a relaxation in the sense that the optimal value of Equation (2) is an upper bound for the value of Equation (1).

The question of how closely the relaxation approximates the true solution to the QCQP is a difficult and fascinating one, but we will only be concerned with Equation (2) in this paper. In particular, our techniques will only produce lower bounds on Equation (2), which may not apply for ???. However, for the particular problems of interest which we describe below, the Schur relaxation is equivalent to the QCQP, and so, our upper bounds also apply to the QCQP.

Our technique is essentially to first take the dual of Equation (2), which is

$$(3) \quad \begin{aligned} \min \quad & \sum_{i=1}^k a_i y_i + z \\ \text{s.t.} \quad & \sum_{i=1}^k \begin{pmatrix} A_i & b_i \\ b_i^\top & 0 \end{pmatrix} y_i + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} z - \begin{pmatrix} A_0 & b_0 \\ b_0 & 0 \end{pmatrix} \in \mathcal{P}(\Delta') \end{aligned}$$

We will need to assume that strong duality holds in this case. We then take a hyperbolic LPM polynomial  $p$  with support  $\Delta$  and consider the relaxation of Equation (4)

$$(4) \quad \begin{aligned} \min \quad & \sum_{i=1}^k a_i y_i + z \\ \text{s.t.} \quad & \sum_{i=1}^k \begin{pmatrix} A_i & b_i \\ b_i^\top & 0 \end{pmatrix} y_i + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} z - \begin{pmatrix} A_0 & b_0 \\ b_0 & 0 \end{pmatrix} \in \Lambda_I(p) \end{aligned}$$

**1.2. Sparse Regression and Sparse PCA.** There are in fact a number of cases when Equation (2) in fact gives the exact same value as Equation (1), for instance, when there is only one constraint.

This is the case for the sparse principal components analysis (PCA) problem, which is defined for a symmetric matrix  $A \in \mathbb{R}^{n \times n}$  by

$$\max\{x^\top A x : \|x\|_2^2 = 1, \|x\|_0 \leq k\}.$$

PCA is often used to reduce the dimensionality of a data set by only considering those dimensions in which the data has the highest variance. The dimensions returned by PCA are often linear combinations of all of the input dimensions, which can make it difficult to interpret semantically what the PCA directions mean. Sparse PCA ensures that the dimensions output in the results are linear combinations of only a small number of input features. This technique is of interest in data science and other fields [?].

This also occurs in the sparse linear regression problem, which is defined for  $A \in \mathbb{R}^{m \times n}$  and  $b \in \mathbb{R}^m$  by

$$\min\{\|Ax - b\|_2^2 : x \in \mathbb{R}^n, \|x\|_0 \leq k\}.$$

This is not quite in the form of a sparse QCQP defined above, but we will see below that it can be brought into the standard form, and that the Schur relaxation is tight in that case. This problem is of fundamental interest in data science as well, because it can be thought of as selecting the  $k$  best features of the data for performing linear regression. In this sense, sparse linear regression is often used as a preprocessing step to reduce the number of features in a dataset before performing more complicated analyses on the data [?].

We may also consider the more general problem of finding a sparse regressor whose support is constrained to be in some fixed set, for example

$$\min\{\|Ax - b\|_2^2 : x \in \mathbb{R}^n, \text{supp}(x) \in \Delta\}.$$

This may be of interest if experimenters have some prior knowledge about what a good set of features will be before applying these

**1.3. Paper Outline.** In this paper, we will describe general procedures for optimizing

(Kevin Shu) SCHOOL OF MATHEMATICS, GEORGIA INSTITUTE OF TECHNOLOGY, 686 CHERRY STREET, ATLANTA, GA 30332, USA

*Email address:* kshu8@gatech.edu