HIGHER-ORDER CONE PROGRAMMING

LIJUN DING AND LEK-HENG LIM

Abstract. We introduce a conic embedding condition that gives a hierarchy of cones and cone programs. This condition is satisfied by a large number of convex cones including the cone of copositive matrices, the cone of completely positive matrices, and all symmetric cones. We discuss properties of the intermediate cones and conic programs in the hierarchy. In particular, we demonstrate how this embedding condition gives rise to a family of cone programs that interpolates between LP, SOCP, and SDP. This family of $k$th order cones may be realized either as cones of $n$-by-$n$ symmetric matrices or as cones of $n$-variate even degree polynomials. The cases $k = 1, 2, n$ then correspond to LP, SOCP, SDP; or, in the language of polynomial optimization, to DSOS, SDSOS, SOS.

1. Introduction

Given a convex proper cone we will show how to construct a hierarchy of cones with associated cone programs, provided that a certain embedding property (defined below) is satisfied. This generalizes the work of Ahmadi and Majumdar in [AM17] where they constructed a sequence of polynomial conic programs, particularly the DSOS and SDSOS conic programs, to approximate the SOS cone program. We will show how such a construction can be carried out for a large number of conic programming problems including:

(i) the nonnegative orthant;
(ii) the second-order cone;
(iii) the cone of symmetric positive semidefinite matrices;
(iv) the cone of copositive matrices;
(v) the cone of completely positive matrices;
(vi) all symmetric cones, i.e., any cone is constructed out of a direct sum of $\mathbb{H}$, $\mathbb{H}$, or the cones of Hermitian positive semidefinite matrices over $\mathbb{C}$, $\mathbb{H}$, and $\mathbb{O}$ (quaternions and octonions);
(vii) any norm cones where the norm satisfies a consistency condition, which includes $l^p$-norms, Schatten and Ky Fan norms, operator $(p, q)$-norms, etc.

For each of these cones, we can build a sequence of intermediate cones and conic programs in the hierarchy. In the case of $\mathbb{H}$, we obtain a family of cone programs that interpolates between LP, SOCP, and SDP. This family of $k$th order cones may be realized either as cones of $n$-by-$n$ symmetric matrices or as cones of $n$-variate even degree polynomials. The cases $k = 1, 2, n$ then correspond to LP, SOCP, SDP; or, in the language of polynomial optimization, to DSOS, SDSOS, SOS.

Notations. Throughout this article, we write $\mathbb{N} := \{1, 2, 3, \ldots \}$ for the set of positive integers. The skew field of quaternions will be denoted as $\mathbb{H}$ and the division ring of octonions as $\mathbb{O}$. We will slightly abuse terminologies and refer to $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, $\mathbb{O}$ as ‘fields.’ We will write $S^d_F$ for the $F$-vector space (or, strictly speaking, $F$-module when $F$ is not a field) of $d \times d$ Hermitian matrices over $F = \mathbb{R}, \mathbb{C}, \mathbb{H}, \mathbb{O}$. When the choice of $F$ is implicit or immaterial, we will just write $S^d$. For a vector $x \in \mathbb{F}^d$, the notation $x \geq 0$ means each component of $x$ is greater or equal to 0.

We write $[d] := \{1, \ldots, d\}$ for any $d \in \mathbb{N}$. We denote the set of all increasing sequences of length $k$ in $[d]$ as $\binom{[d]}{k} = \{(i_1, \ldots, i_k) \mid 1 \leq i_1 < \cdots < i_k \leq d\}$.

For a matrix $A = [a_{ij}]_{ij} \in S^d$, we write $\text{tr}(A) = \sum_{i=1}^d a_{ii}$. The inner product $\langle \cdot, \cdot \rangle : S^d \times S^d \to \mathbb{R}$ we use in this article is the standard trace inner product $\langle A, B \rangle = \text{tr}(AB)$. The topology is then
defined via the distance metric induced by the trace inner product. We write the interior of a set \( S \subset \mathbb{F}^d \) as \( \text{int}(S) \).

2. Conic embedding property

To standardize our terminologies, the cones in this article will all be represented as cones of symmetric matrices over some field \( \mathbb{F} \); although we will see that this is hardly a limitation — conic programs involving cones in other common \( \mathbb{F} \)-vector spaces, e.g., of vectors in \( \mathbb{F}^n \) or polynomials in \( \mathbb{F}[x] \) or \( \mathbb{F} \)-valued functions on some set, can often be transformed to a symmetric matrix setting.

We start by defining two linear maps. Let \( k \leq d \) be positive integers. For \( \{i_1, \ldots, i_k\} \in \binom{[d]}{k} \), i.e., \( 1 \leq i_1 < \cdots < i_k \leq d \), the truncation operator is the projection \( \tau_{i_1 \cdots i_k}^d : \mathbb{S}^d \to \mathbb{S}^k \) defined by

\[
\tau_{i_1 \cdots i_k}^d(Z) := \begin{bmatrix} z_{i_1 i_1} & \cdots & z_{i_1 i_k} \\
z_{i_2 i_1} & \cdots & z_{i_2 i_k} \\
& \vdots & \\
z_{i_k i_1} & \cdots & z_{i_k i_k} \end{bmatrix}
\]

for any \( Z \in \mathbb{S}^d \); the lift operator is the injection \( \varepsilon_{i_1 \cdots i_k}^d : \mathbb{S}^k \to \mathbb{S}^d \) defined by

\[
[e_{i_1 \cdots i_k}^d(X)]_{pq} = \begin{cases} x_{pq} & p, q \in \{1, \ldots, k\}, \\
0 & \text{otherwise.} \end{cases}
\]

for any \( X \in \mathbb{S}^k \). In other words, the truncation operator takes a \( d \times d \) matrix to its \( k \times k \) submatrix; whereas the lifting operator takes a \( k \times k \) matrix and embed it as a \( d \times d \) matrix by filling-in the extra entries as zeros. Clearly for a fixed index set \( \{i_1, \ldots, i_k\} \), \( \tau_{i_1 \cdots i_k}^d \) is a left inverse of \( \varepsilon_{i_1 \cdots i_k}^d \), i.e.,

\[
\tau_{i_1 \cdots i_k}^d \circ \varepsilon_{i_1 \cdots i_k}^d = \text{id}_{\mathbb{S}^k}.
\]

We now state our embedding property.

**Definition 2.1.** Let \( \mathbb{F} = \mathbb{R}, \mathbb{C}, \mathbb{H}, \text{ or } \mathbb{O} \) and \( \mathbb{S}^k = \mathbb{S}^k_{\mathbb{F}} \). Let \( k_0 \in \mathbb{N} \) and \( \{K^k : k \in \mathbb{N}, k \geq k_0\} \) be a sequence of convex proper cones where \( K^k \subseteq \mathbb{S}^k \) for each \( k \geq k_0 \). We say that the sequence \( \{K^k\}_{k=k_0}^{\infty} \) satisfies the *embedding property* with index map

\[
I : \{(d, k) \in \mathbb{N} \times \mathbb{N} \mid d \geq k\} \to \bigcup_{k_0 \leq k \leq d} \binom{[d]}{k}
\]

if for any \( d \geq k \geq k_0, (i_1, \ldots, i_k) \in I(d, k) \), we have

\[
\tau_{i_1 i_2 \cdots i_k}^d(Z) \in K^k \quad \text{and} \quad \varepsilon_{i_1 i_2 \cdots i_k}^d(X) \in K^d
\]

for all \( Z \in K^d \) and \( X \in K^k \).

We caution our reader that the “higher-order cones” in the title of this article do not refer to \( \{K^k\}_{k=k_0}^{\infty} \) but will be constructed out of these cones. In several instances, the index map is given simply by

\[
I(d, k) = \binom{[d]}{k},
\]

and in which case we will drop any reference to the index map and just say that \( \{K^k\}_{k=k_0}^{\infty} \) satisfies the embedding property. If in addition \( k_0 = 1 \), we will say that \( \{K^k\}_{k=1}^{\infty} \) satisfies the embedding property *thoroughly.*

The embedding property simply says for a \( d \times d \) matrix \( Z \in K^d \), its \( k \times k \) principle submatrix belongs to the lower dimension cone \( K^k \); conversely, for a \( k \times k \) matrix \( X \in K^k \), embedding as it a principle submatrix of a \( d \times d \) matrix with all other entries set to be zero gives a matrix in \( K^d \).
A simple example is the cone of symmetric diagonally dominant matrices with nonnegative diagonals,

\[ \text{DD}_d := \left\{ M \in \mathbb{S}^d : m_{ii} \geq \sum_{j \neq i} |m_{ij}|, \ i = 1, \ldots, d \right\}, \]

where it is easy to see that \( \{\text{DD}_k\}_{k=1}^\infty \) satisfies the embedding property thoroughly. We will see many more examples of cones satisfying the embedding property over the next few sections.

We may now define the higher order cones in the title of this article. They are obtained by lifting cones in lower dimension to higher dimension. The benefit is that though the cones defined are in high dimension, they are expressible by cones in lower dimension and property of cones in lower dimension might be utilized. These higher cone might be served as an inner approximation of cones in high dimension.

As usual, in the following we let \( F = \mathbb{R}, \mathbb{C}, \mathbb{H}, \) or \( \mathbb{O} \) and write \( \mathbb{S}^d = \mathbb{S}^d_F \).

**Definition 2.2.** Let \( \{K^k\}_{k=k_0}^\infty \) be a sequence of cones that satisfies the embedding property with index map \( I \). The \( k \)th order cone with index set \( J \subseteq I(d, k) \) induced by \( K^k \) is

\[ K^d_k(J) := \left\{ M \in \mathbb{S}^d : M = \sum_{(i_1, \ldots, i_k) \in J} \varepsilon_{i_1 \ldots i_k}^{d} (M_{i_1 \ldots i_k}), \ M_{i_1 \ldots i_k} \in K^k \right\}, \]

If \( J = I(d, k) \), we will just write \( K^d_k \) for \( K^d_k(I(d, k)) \).

We will establish some basic properties of higher order cones.

**Proposition 2.3.** Let \( \{K^d_k\}_{k=k_0}^\infty \) satisfy the embedding property with index mapping \( I \). Then the following properties hold:

(i) Nested cones: Suppose a sequence of index sets \( \{J_k\}_{k=k_0}^d \subseteq I(d, k) \) satisfies that for any \( k \geq k_0 \) and any \( s \in J_k \), there is an \( s' \in J_{k+1} \) such that all the components of \( s \) appears in \( s' \) (This property is satisfied by \( I(d, k) \)). Then we have

\[ K^d_{k_0}(J_{k_0}) \subseteq K^d_{k_0+1}(J_{k_0+1}) \subseteq \cdots \subseteq K^d_d(J_d). \]

In particular, if \( \{I(d, k)\}_{k=k_0}^d \) is such sequence of index sets, then for every \( d \geq k_0 \), we have

\[ K^d_{k_0} \subseteq K^d_{k_0+1} \subseteq \cdots \subseteq K^d_d. \]

(ii) Dual cones: the dual cone of \( K^d_k(J) \) is

\[ (K^d_k(J))^* = \left\{ A \in \mathbb{S}^d : A \text{ is Hermitian and for all } (i_1, \ldots, i_k) \in J, \tau_{i_1 \ldots i_k}(A) \in (K^k)^* \right\}. \]

(iii) Membership: If \( I(d, k) = \binom{d}{k} \) for every \( d \), we have \( X_1 \in K^k_1, X_2 \in K^s_k \iff \text{diag}(X_1, X_2) \in K^{k+s}_k \), ditto for the dual cones of \( K^k_d \).

(iv) Inheritance: If the embedding property is satisfied thoroughly by \( \{K^k\}_{k=1}^\infty \), then for each \( k \geq 1 \), the sequence of cones \( \{K^k_d\}_{k=k_0}^\infty \) satisfies the embedding property.

**Proof.** (i) Consider \( k_0 + i \) and \( k_0 + i + 1 \) where \( 0 \leq i \leq d - k_0 - 1 \). The cones \( K^d_{k_0+i}(J_{k_0+i}) \) and \( K^d_{k_0+i+1}(J_{k_0+i+1}) \) can be expressed as

\[(2.2) \quad K^d_{k_0+i}(J_{k_0+i}) = \sum_{(j_1, \ldots, j_{k_0+i}) \in J_{k_0+i}} \varepsilon_{j_1 \ldots j_{k_0+i}}^{d}(K^{k_0+i}) \]

and

\[(2.3) \quad K^d_{k_0+i+1}(J_{k_0+i+1}) = \sum_{(j_1, \ldots, j_{k_0+i+1}) \in J_{k_0+i+1}} \varepsilon_{j_1 \ldots j_{k_0+i+1}}^{d}(K^{k_0+i+1}) \]
By the assumption on \( \{J_k^d\}_{k=0}^d \), we know for each \((j_1, \ldots, j_{k_0+i}) \in J_{k_0+i}\), there is some \( j \) such that \( \{j_1, \ldots, j_{k_0+i}, j\} \) after ordering is in \( J_{k_0+i+1} \). Without loss of generality, we may assume \( j \) is the largest among \( \{j_1, \ldots, j_{k_0+i}, j\} \). Thus

\[
\varepsilon_{j_1 \ldots j_{k_0+i}}^d(K_{k_0+i}) = \varepsilon_{j_1 \ldots j_{k_0+i}+j}^d \left[ \begin{array}{c} K_{k_0+i}^d \\ 0 \end{array} \right],
\]

where (a) is because of \( \left[ K_{k_0+i}^d \\ 0 \right] \subseteq K_{k_0+i+1}^d \) using the embedding property. Thus we see each summand in the decomposition \( (2.2) \) is a subset of a summand in the decomposition of \( (2.3) \). Using the conic property that \( a, b \in K_{k_0+i+1}^d \Rightarrow a + b \in K_{k_0+i+1}^d \), we see indeed

\[
K_{k_0+i}^d \subseteq K_{k_0+i}^d.
\]

Since \( i \) is arbitrary, we see we have the cones are nested.

(ii) We use the following simple fact [Roc70, Corollary 16.3.2] that for convex cone \( K_1, K_2 \),

\[
(K_1 + K_2)^* = K_1^* \cap K_2^*.
\]

By definition, \( \text{K}^d_k(J) \) can be expressed as

\[
\text{K}^d_k(J) = \sum_{(i_1, \ldots, i_k) \in J} \varepsilon_{i_1 \ldots i_k}^d(K^k),
\]

where each \( \varepsilon_{i_1 \ldots i_k}^d(K^k) \) is a convex cone in \( S^d \). The dual cone of \( \varepsilon_{i_1 \ldots i_k}^d(K^k) \) is

\[
(\varepsilon_{i_1 \ldots i_k}^d(K^k))^* = \{ A \in S^d : \tau_{i_1 \ldots i_k}(A) \in (K^k)^* \}.
\]

Applying previous fact, we get the characterization of the dual cone.

(iii) We first show that \( X_1 \in K_{k_0}^d, X_2 \in K_{k_0}^d \Rightarrow \text{diag}(X_1, X_2) \in K_{k_0}^{s+t} \). We know there are \( M_{i_1 \ldots i_k}, Y_{j_1 \ldots j_k} \in K_{k_0}^d \) such that

\[
X_1 = \sum_{(i_1, \ldots, i_k) \in \binom{[d]}{k}} \varepsilon_{i_1 \ldots i_k}^s(M_{i_1 \ldots i_k}),
\]

and

\[
X_2 = \sum_{(j_1, \ldots, j_k) \in \binom{[d]}{k}} \varepsilon_{j_1 \ldots j_k}^t(Y_{j_1 \ldots j_k}).
\]

Thus

\[
\text{diag}(X_1, X_2) = \sum_{(i_1, \ldots, i_k) \in \binom{[d]}{k}} \text{diag}(\varepsilon_{i_1 \ldots i_k}^s(M_{i_1 \ldots i_k}), 0) + \sum_{(j_1, \ldots, j_k) \in \binom{[d]}{k}} \text{diag}(0, \varepsilon_{j_1 \ldots j_k}^t(Y_{j_1 \ldots j_k}))
\]

\[
= \sum_{(i_1, \ldots, i_k) \in \binom{[d]}{k}} \varepsilon_{i_1 \ldots i_k}^{s+t}(M_{i_1 \ldots i_k}) + \sum_{(j_1, \ldots, j_k) \in \binom{[d]}{k}} \varepsilon_{j_1 \ldots j_k}^{s+t}(Y_{j_1 \ldots j_k}).
\]
HIGHER-ORDER CONE PROGRAMMING

Since \( \binom{[s]}{k}, \binom{[t]}{k} \subseteq \binom{[s+t]}{k} \), we see the above indeed gives a valid decomposition of \( k \)th order cone induced by \( \{K^k\}_{k=1}^\infty \). Now suppose \( \text{diag}(X_1, X_2) \in K^{s+t}_k \). This gives
\[
\text{diag}(X_1, X_2) = \sum_{(i_1, \ldots, i_k) \in \binom{[s+t]}{k}} \varepsilon_{i_1}^{s+t} \cdots \varepsilon_{i_k}^{s+t} (Z_{i_1} \cdots i_k)
\]
where \( Z_{i_1} \cdots i_k \in K^k \). Apply \( \tau_{1,2,\ldots,s} \) and \( \tau_{s+1,s+2,\ldots,s+t} \) to both sides of the above equality gives valid decompositions of \( X_1 \in K^s_k \) and \( X_2 \in K^t_k \) due to the embedding property.

For the dual cones, note that if \( X \in (K^k)^* \), then for each \( l \leq k \), \( \tau_{i_1} \cdots i_l(X) \in (K^l)^* \) because of the embedding property. The rest of the proof is similar to previous one.

(iv) Fix \( k \leq l < m \). Consider an increasing sequence \( (i_1, \ldots, i_l) \in \binom{[m]}{l} \) and any \( X \in K^k_k, Z \in K^m_k \). Then there are some \( M_{j_1, \ldots, j_k} \in K^k_k, Y_{n_1, \ldots, n_k} \in K^k_k \) such that
\[
\varepsilon_{i_1}^m \cdots i_l(X) = \varepsilon_{i_1}^m \cdots i_l \left( \sum_{(j_1, \ldots, j_k) \in \binom{[l]}{k}} \varepsilon_{j_1}^{m-l} \cdots \varepsilon_{j_k}^{m-l} (M_{j_1} \cdots j_k) \right)
\]
and
\[
\tau_{i_1}^m \cdots i_l(Z) = \tau_{i_1}^m \cdots i_l \left( \sum_{(n_1, \ldots, n_k) \in \binom{[m]}{k}} \varepsilon_{n_1}^m \cdots n_k (Y_{n_1} \cdots n_k) \right)
\]
Since \( \varepsilon_{i_1}^m \cdots i_l \circ (\varepsilon_{j_1}^{m-l} \cdots j_k) = \varepsilon_{i_1}^m \cdots i_l \), we see \( \varepsilon_{i_1}^m \cdots i_l(X) \) is indeed a member of \( K^m_k \). Using the embedding property for each \( \tau_{i_1}^m \cdots i_l \left( \varepsilon_{n_1}^m \cdots n_k (Y_{n_1} \cdots n_k) \right) \), we see \( \tau_{i_1}^m \cdots i_l(Z) \in K^l_k \).

\[\square\]

Given the definition of higher order cone and dual cone, we can consider their corresponding conic programs. We assume the underlying filed is real for simplicity. More precisely, the \( k \)th order cone program (standard form) is
\[
\begin{align*}
\text{minimize} & \quad tr(A_0 X) \\
\text{subject to} & \quad tr(A_i X) = b_i, \quad i = 1, \ldots, p \\
& \quad X \in K^d_k(J)
\end{align*}
\]
where \( A_1, \ldots, A_p \in \mathbb{R}^{n \times n}, b_1, \ldots, b_p \in \mathbb{R} \).

Alternatively, \( k \)th order cone program (inequality form) is
\[
\begin{align*}
\text{minimize} & \quad q^T x \\
\text{subject to} & \quad x_1 P_1 + x_2 P_2 + \cdots + x_k P_k + P_0 \in K^d_k(J)
\end{align*}
\]
where \( P_0, \ldots, P_k \in S^d \) and \( q \in \mathbb{R}^n \). The constraint here is called linear matrix inequality (LMI).

We call these programs \( kOCP \) induced by \( K^k \) with set \( J \) and simply \( kOCP \) if the underlying cone \( K^k \) is clear from the context and \( J = I(k, d) \). We note that the ambient dimension \( d \) might change from problem to problem as the case of semidefinite programming where the ambient dimension \( d \) is not specified, i.e., we write \( X \succeq 0 \) meaning \( X \) is positive semidefinite but did not specify the size of \( X \).
If the nested cones property is satisfied by the underlying cone \( K^k \) (which is true when \( I(d, k) \) satisfies the condition of first item, Nested Cones, of Proposition \( 2.3 \)), the above program serves as inner approximation of \( K^d \) program.

We state an equivalence theorem of the two form when \( K^k \) satisfies the embedding property thoroughly.

**Theorem 2.4.** If \( \{K^k\}_{k=1}^{\infty} \) satisfies the embedding property thoroughly, then the inequality form and the standard form are equivalent.

**Proof.** Without loss of generality, we assume that \( K^1 = \mathbb{R}_+ \) (because one dimensional proper cone is either \( \mathbb{R}_+ \) or \( \mathbb{R}_- \)) and \( p \geq k \) (we can repeat a few constraints if \( p < k \).

By Lemma 9.1 proved in the Appendix, we find that for any \( x \in \mathbb{F}^d \),

\[
\text{diag}(x) \in K^d_k \iff x \geq 0,
\]

where \( x \geq 0 \) means each component of \( x \) is greater or equal to 0.

We first prove the direction from the standard form to inequality form, i.e., \( 2.4 \) to \( 2.5 \):

By treating \( X \) as a long vector, the objective and the conic constraint \( X \in K^d_k \) can be transformed in a standard way. Indeed, the objective is just \( \sum_{ij}(A_0)_{ij}x_{ij} \). For conic constraint, we have

\[
X \in K^d_k \iff \sum_{j>k} x_{jk}(E_{jk} + E_{kj}) + \sum_{i=1}^n x_{jj}E_{jj} \in K^d_k,
\]

where \( E_{jk} \) are the matrices with only non-zero entry 1 at \((j,k)\)th entry. The linear constraint \( \text{tr}(A_iX) = b_i \) can be encoded by

\[
\text{diag}(\text{tr}(A_iX) - [b_i]_{i=1}^p) \in K^d_k, \quad \text{diag}([b_i - \text{tr}(A_iX)]_{i=1}^p) \in K^d_k.
\]

Finally, using membership property in Lemma 2.3 the transformed linear constraints \( 2.7 \) and the transformed conic constraint \( 2.6 \) can be made into one big \( k \)th order cone linear matrix inequality.

We now prove the direction from inequality form to standard form, i.e., \( 2.5 \) to \( 2.4 \):

First we can write \( x = x^+ - x^- \) as two non-negative vectors (element wise non-negative). Let \( \bar{X} = x_1P_1 + x_2P_2 + \ldots + x_kP_k + P_0 \), then the inequality form \( 2.5 \) can be transformed to

\[
\begin{align*}
\text{minimize} & \quad q^T x^+ - q^T x^- \\
\text{subject to} & \quad \sum_{i=1}^k (x_i^+P_i - x_i^-P_i) - \bar{X} = -P_0 \\
& \quad \bar{X} \in K^d_k, \quad x^+ \geq 0, \quad x^- \geq 0.
\end{align*}
\]

It can then be transformed to \( 2.4 \). We may let the \( X \) in \( 2.4 \) be

\[
X = \begin{bmatrix}
\text{diag}(x^+) \\
\text{diag}(x^-) \\
\bar{X}
\end{bmatrix}.
\]

The objective in \( 2.4 \) then can be easily formulated as \( D_0 = \begin{bmatrix} \text{diag}(q) & \quad -\text{diag}(q) \quad 0 \end{bmatrix} \). The equality constraints are just a re-statement of the elementwise version of \( \sum_{i=1}^k (x_i^+P_i - x_i^-P_i) - \bar{X} = -P_0 \). So \( D_i, f_i \) are set so that \( \text{tr}(D_i X) = [\sum_{i=1}^k (x_i^+P_i - x_i^-P_i) - \bar{X}]_{jk} = -P_{jk} f_i \). A total of \( \frac{n(n+1)}{2} \) constraints can be obtained from this method. To enforce the 0 in \( X \), we can put more \( \text{tr}(E_{ij}X) = 0 \) constraints on \( X \) with position index \((i,j)\) of 0 in \( X \) where \( E_{ij} \) is defined as the previous part. These linear constraints implies that for \( n \geq k \), \( X \in K^{d+2n}_k \) if and only if
$x^-, x^+ \geq 0, \bar{X} \in \mathbb{K}_d^k$ because the membership property and Lemma 9.1. If $n < k$, we may simply repeat $x^-, x^+$ in $X$ and enforce the repetition by adding more linear constraints. □

The dual $k$OCP (standard form) is
\[
\begin{align*}
\text{minimize} & \quad \text{tr}(A_0 X) \\
\text{subject to} & \quad \text{tr}(A_i X) = b_i, \quad i = 1, \ldots, p \\
& \quad X \in (\mathbb{K}_d^k(J))^* 
\end{align*}
\]
where $A_1, \ldots, A_p \in \mathbb{R}^{n \times n}$, $b_1, \ldots, b_p \in \mathbb{R}$ and the dual $k$OCP (inequality form) is
\[
\begin{align*}
\text{minimize} & \quad q^\top x \\
\text{subject to} & \quad x_1 P_1 + x_2 P_2 + \cdots + x_k P_k + P_0 \in (\mathbb{K}_d^k(J))^* 
\end{align*}
\]
where $P_0, \ldots, P_k \in \mathbb{S}^n$.

3. Positive semidefinite cone

Our first example is the cone of positive semidefinite matrices with dimension $d$:
\[
\mathbb{S}_+^d := \{ A \in \mathbb{S}^d : A = FF^\top \text{ for some } F \in \mathbb{R}^{d \times r}, r \in \mathbb{N} \}.
\]

Clearly, the sequence of cones $\{ \mathbb{S}_+^k \}_{k=1}^d$ satisfy the embedding property thoroughly. The first two order cones are:

(i) $\mathbb{S}_+^d_1 = \text{diag}(\mathbb{R}_+^d) \cong \mathbb{R}_+^d$. Note that from the inheritance property, fourth item of Proposition 2.3, the nonnegative orthant $\text{diag}(\mathbb{R}_+^d) \cong \mathbb{R}_+^d$ satisfies the embedding property throughly as well, which can also be directly verified.

(ii) $\mathbb{S}_+^d_2 = \{ A : A = \sum_{i<j} \varepsilon_{ij}(M^{ij}), \quad M^{ij} \in \mathbb{S}_+^2 \}$. Note this series of cone also satisfied the embedding property throughly by attaching $\mathbb{R}_+$ to the series $\{(\mathbb{S}_+^d_2)_{d=2} \}_{d=2}$.

It turns out that the second order cone $\mathbb{S}_+^d_2$ actually is the same as the set of symmetric scaled diagonally dominant matrices with nonnegative diagonals (SDD), $\mathbb{S}^d$, $\mathbb{S}^d_{\text{SDD}} := \{ M \in \mathbb{S}^d : \text{there exists } d > 0, d_i a_{ii} \geq \sum_{j \neq i} d_j |a_{ij}|, \text{ for all } i = 1, \ldots, d \}$, as shown in the following lemma, which appeared in [BCPT05, Theorems 8 and 9] and [AM17, Lemma 9].

Lemma 3.1. $\mathbb{S}_+^d_2 = \mathbb{S}^d$. We provide a simple, different and self-contained proof of this lemma based on the following lemma which can be found in Appendix.

Lemma 3.2. Denote $M(A) = [\alpha_{ij}]$ where $\alpha_{ii} = a_{ii}$ for all $i$ and $\alpha_{ij} = -|a_{ij}|$ for all $i \neq j$ and $\rho(A) = \max \{|\lambda| : \lambda \text{ is an eigenvalue of } A \}$. The following are all equivalent when $A \in \mathbb{S}^n$.

(i) $A$ is SDD;
(ii) $M(A)$ is positive semi-definite.

Proof of Lemma 3.2. We let $M(A) = [\alpha_{ij}] \in \mathbb{S}^d$ where $\alpha_{ii} = a_{ii}$ for $i = 1, \ldots, d$, and $\alpha_{ij} = -|a_{ij}|$ for all $i \neq j$; this is often called the comparison matrix [BP94] of $A$. To show that $\mathbb{S}^d \supset (\mathbb{S}_+^d_2)$, suppose $A \in (\mathbb{S}_+^d_2)$. Then $A = \sum_{i<j} M^{ij}$. Since $M(A) = \sum_{i<j} M(M^{ij})$ with $M(M^{ij}) \in \mathbb{S}_+^2$, $M(A)$ belongs to both $(\mathbb{S}_+^d_2)$ and $\mathbb{S}^d$.

Now suppose $A \in \mathbb{S}^d$. There exists $d = (d_1, \ldots, d_n) > 0$ such that $d_i a_{ii} \geq \sum_{j \neq i} d_j |a_{ij}|$ for each $i$, which allows us to define $M^{ij}$ by
\[
m_{ij} = a_{ij}, \quad m_{ii} = \frac{d_j}{a_{ii}} a_{ij}, \quad m_{jj} = \frac{d_i}{d_j} a_{ij}.
\]
We may then increase the values of $m_{ij}^{ij}$ and $n_{ij}^{ij}$ appropriately so that they sum up to the respective diagonal entries of $A$. This shows that $\operatorname{SDD}^d \subset (S^d_+)_2$. □

For general $J \subseteq \binom{[d]}{2}$, the equality in Lemma 3.1 does not hold, i.e., $(S^d_+)_2(J) \neq \operatorname{SDD}^d$ for general $J$.

The $kOCP$ in this case is actually very interesting. The $1OCP$ is simply Linear Program (LP) since $(S^d_+)_1 = \text{diag}(\mathbb{R}_+) \cong \mathbb{R}_+$, the $2OCP$ in this case is SDD program. We show in the following theorem that SDD program is the same as Second Order Cone Program (SOCP):

\begin{align}
\text{minimize} & \quad a^T x \\
\text{subject to} & \quad \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, q, \\
& \quad B x = e.
\end{align}

**Theorem 3.3.** SDD program is equivalent to SOCP, i.e., SOCP can be casted into SDD Program and vice versa.

**Proof.** The fact that SDD program can be optimized using SOCP has been shown in [AM17, Theorem 10], which is just an easy consequence of Lemma 3.1. We are only left to show the other direction. We will show one can transform a SOCP to the inequality form of SDD program. The equivalence between inequality form and standard form of SDD program follows from Theorem 2.4.

Our only difficulty is to transform a SOC constraint,

$$\|A_i x + b_i\|_2 \leq c_i^T x + d_i,$$

to a SDD constraint. We know

$$\|A_i x + b_i\|_2 \leq c_i^T x + d_i \iff \begin{bmatrix} (c_i^T x + d) I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d \end{bmatrix} \in S^n_+ \iff \begin{bmatrix} (c_i^T x + d) I & -|A_i x + b_i| \\ -|A_i x + b_i|^T & c_i^T x + d \end{bmatrix} \in S^n_+$$

for appropriate $n$ by the Schur complement condition for positive semi-definiteness, i.e.,

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \in S^n_+ \iff A \in S^m_+, C - B^T A^{-1} B \in S^h_+,$$

where $m, h$ are number of rows of $A$ and $C$. Now using Lemma 3.2, we see

$$\begin{bmatrix} (c_i^T x + d) I & -|A_i x + b_i| \\ -|A_i x + b_i|^T & c_i^T x + d \end{bmatrix} \in S^n_+ \iff \begin{bmatrix} (c_i^T x + d) I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d \end{bmatrix} \in (S^d_+)_2.$$ 

The last equation is a linear $(S^d_+)_2$ constraint and we see SOCP can be transformed to SDD program and so the two are equivalent. □

Thus we have shown that the intermediate program between LP, SOCP and SDP are kOCP and

- $1OCP = LP$,
- $2OCP = SOCP$,
- $dOCP = SDP$,
- $kOCP$ for $k = 3, \ldots, d - 1$ are intermediate programs:

$$\text{minimize} \quad \text{tr}(A_0 X)$$
$$\text{subject to} \quad \text{tr}(A_i X) = b_i, \quad i = 1, \ldots, p$$
$$X \in (S^d_+)^k,$$

where $(S^d_+)^k = \{ M : M = \sum_{(i_1, \ldots, i_k) \in \binom{[d]}{k}} c_{i_1 \cdots i_k}^d (M_{i_1 \cdots i_k}), \quad M_{i_1 \cdots i_k} \in S^{k}_+ \}$.

The elements in higher order cone $(S^d_+)_k$ with $k \geq 3$ turns out to be known as factor-width $k$ matrices [BCPT05]. The corresponding program has being introduced in [PP14] before.

The dual cones are:

$$((S^d_+)_k)^* = \{ A \in S^d : \text{for all } (i_1, \ldots, i_k) \in J, \tau_{i_1 \cdots i_k}(A) \in S^k_+ \}.$$
In the case of semidefinite cone, the nested inclusion for higher order cones \( \{(S^+_d)_k\}_{k=1}^d \) and its dual cone series \( \{(S^+_d)^*_k\}_{k=1}^d \) are strict as shown in the following lemma.

**Lemma 3.4.** We have

\[
(S^+_d)_1 \subset (S^+_d)_2 \subset \cdots \subset (S^+_d) = S^d
\]

and

\[
((S^+_d)_1)^* \supset ((S^+_d)_2)^* \supset \cdots \supset ((S^+_d)) = S^d.
\]

**Proof.** Both inclusion are easy consequences of first and second item of Proposition 2.3. We now prove the inclusion is strict. We first prove that the strict inclusion for the dual cones. Denote \( \mathbf{1}_d = (1, \ldots, 1)^T \) and \( I_d \) be the identity matrix in \( S^d \). The matrix

\[
\begin{bmatrix}
\sqrt{k-1} & \mathbf{1}_{d-1}^T \\
\mathbf{1}_{d-1} & \sqrt{k-1}I_{d-1}
\end{bmatrix}
\]

is always in \( ((S^+_d)^*_k)^* \) but not in \( ((S^+_d)^*_k+1)^* \).

Since \( ((S^+_d)^*_k)^* = \bigcap_{(i_1, \ldots, i_k) \in [d]^k} K_{i_1 \cdots i_k} \) where

\[
K_{i_1 \cdots i_k} = \{ A \in S^d : \tau_{i_1 \cdots i_k}(A) \in S^k \},
\]

\( (K_{i_1 \cdots i_k})^* = \varepsilon_{i_1 \cdots i_k}(S^k) \) and the identity matrix \( I \in \text{int}(K_{i_1 \cdots i_k}) \) for all \( (i_1, \ldots, i_k) \in [d]^k \), the Krein-Rutman Theorem [BL10, Corollary 3.3.13] implies that

\[
((S^+_d)^*_k)^* = \sum_{i_1 \cdots i_k} \varepsilon_{i_1 \cdots i_k}(S^k) = (S^d)_k.
\]

Thus strict inclusion in the dual cones implies the strict inclusion in the cones \( (S^d)_k \). The equality \( (S^d)_d = S^d = ((S^d)_d)^* \) is because \( S^d \) is self-dual. \( \square \)

So far we have mostly dealing with index set \( J_k = [d]^k \). By changing the index \( J_k \) of the \( k \)th order cone, we obtain new cones and new conic program. In real problems, the choice of the subset \( J_k \) of \( [d]^k \) represents some prior knowledge of the problem. The corresponding higher order cone and dual higher order cone program can enojoy less computational budget because of the smaller size of \( J_k \). This has been explored in the literature of chordal structure of SDP [WKKM06, DK10].

### 4. Sum-of-squares cone

A real coefficient polynomial \( p(x) \) is a sum-of-square (SOS) if it can be written as \( p(x) = \sum_{i=1}^m q_i^2(x) \) for some polynomial \( q_i \). It is clear that the set of sum of square polynomials form a convex cone.

It is well-known that a polynomial with \( n \) variable and degree \( 2d \) is a sum of square if and only if there exists a positive semidefinite symmetric \( A \) such that

\[
p(x) = m(x)^T Am(x)
\]

where \( m(x) \) is the vector of all monomials (so in total \( \binom{n+d}{d} \) tuples) that have degree less than or equal to \( d \) [Par00]. Due to this equivalence and our previous discussion on \( k \)OCP induced by \( S^k_+ \), we define the following \( k \)DDSOS.

**Definition 4.1.** Let \( i_1, \ldots, i_k \in \{1, 2, \ldots, \binom{n+d}{d} \} \). A polynomial \( p \) is \( k \)-th-diagonally-dominant-sum-of-squares (\( k \)DDSOS) if it can be written as

\[
p = \sum_{i_1 \cdots i_k} \sum_j \left( \sum_{l=1}^k \alpha_{j_1 \cdots i_k} m_{i_l} \right)^2
\]
for some monomials $m_{ij}$ and some constants $\alpha_{ji}^{i_1 \cdots i_k} \in \mathbb{R}$.

It directly follows from the definition that a polynomial $p$ (with $n$ variable and $2d$ degree) is SOS if and only if it is $\binom{n+d}{d}$-DDSOS. The cases $k=1,2$ has been explored intensively in [AM17] under the name DSOS and SDSOS.

In the definition, we did not require $i_1 < \cdots < i_k$ as we did in defining $k$OC. We show in the following lemma that this requirement is not necessary.

**Lemma 4.2.** Suppose the monomials having $n$ variables with degree less than or equal to $d$ are indexed by $\{1,2,\ldots,\binom{n+d}{d}\}$ according to some order. A polynomial $p$ with degree $2d$, $n$ variables is $k$DDSOS if and only if it can be written as

$$p = \sum_{(i_1,\ldots,i_k) \in \binom{\{1,\ldots,n\}}{d}} \sum_{j=1}^k \left( \sum_{l=1}^k \alpha_{ji}^{i_1 \cdots i_k} m_{ij_l} \right)^2,$$

where $(m_{ij})_{i=1}^k$ are different for different $(i_l)_{l=1}^k$.

**Proof.** It is easy to see a polynomial can be written in the above form is a $k$DDSOS.

Now suppose $p$ is a $k$DDSOS, by rearrange the brackets and adding 0 terms if necessary, we could write $p$ in the form

$$p = \sum_{(i_1,\ldots,i_k) \in \binom{\{1,\ldots,n\}}{d}} \sum_{j=1}^k \left( \sum_{l=1}^k \alpha_{ji}^{i_1 \cdots i_k} m_{ij_l} \right)^2,$$

where $m_{ij}$s are different when $i$'s are not equal. So the thing left to do is to make sure there are $k$ brackets in the second sum, i.e., the sum over $j$. Since

$$\left( \sum_{l=1}^k \alpha_{ji}^{i_1 \cdots i_k} m_{ij_l} \right)^2 = m_{i_1 \cdots i_k}^T (\alpha_{ji}^{i_1 \cdots i_k})^T \alpha_{ji}^{i_1 \cdots i_k} m_{i_1 \cdots i_k},$$

where $m_{i_1 \cdots i_k} = (m_{i_1}, \ldots, m_{i_k}), \alpha_{ji}^{i_1 \cdots i_k} = (\alpha_{ji}^{i_1}, \ldots, \alpha_{ji}^{i_k})$. The sum $\sum_j (\alpha_{ji}^{i_1 \cdots i_k})^T \alpha_{ji}^{i_1 \cdots i_k}$ is still a non-negative definite matrix and thus has a Cholesky decomposition, i.e., $\sum_j (\alpha_{ji}^{i_1 \cdots i_k})^T \alpha_{ji}^{i_1 \cdots i_k} = D^T D$. This means

$$p = \sum_{i_1 \cdots i_k} (m_{i_1 \cdots i_k} D)^T D (m_{i_1 \cdots i_k}),$$

which shows there can be exactly $k$ brackets in the second sum. \hfill \Box

The following theorem connects our $k$DDSOS polynomial with our $k$th order cone induced by $S^k_+$.  

**Theorem 4.3.** A polynomial $p$ of degree $2d$ with $n$ variables is $k$DDSOS if and only if it admits a representation as $p(x) = m^T \tau^T(x) A m(x)$, where $m(x)$ is the standard monomial vector of degree $d$ (so in total $\binom{n+d}{d}$ tuples with different entries), and $A \in (S^h_+)_k$ for some $h \leq \binom{n+d}{d}$.

**Proof.** If $p$ admits a representation

$$p = m^T \tau^T A m,$$

where $A \in (S^h_+)_k$ for some $h$ and $m$ is the vector of all monomials with degree less than $d$. Since $A \in (S^h_+)_k$, $A$ has the decomposition $A = \sum_{(i_1,\ldots,i_k) \in \binom{\{1,\ldots,n\}}{d}} M_{i_1 \cdots i_k}$. $M_{i_1 \cdots i_k}$ are zero except for those $(i,j), i,j \in \{i_1,\ldots,i_k\}$ entries. $\tau_{i_1 \cdots i_k}^h(M_{i_1 \cdots i_k})$ are positive semi-definite and thus has the Cholesky
decomposition $\tau_{i_1\cdots i_k}^h (M^{i_1\cdots i_k}) = N_{i_1\cdots i_k}^T N_{i_1\cdots i_k}$. Thus, we have

$$p = m^T Am$$

$$= \sum_{i_1\cdots i_k} m^T M^{i_1\cdots i_k} m$$

$$= \sum_{i_1\cdots i_k} \begin{bmatrix} m_{i_1} & \cdots & m_{i_k} \end{bmatrix} \tau_{i_1\cdots i_k}^d (M^{i_1\cdots i_k}) \begin{bmatrix} m_{i_1} \\ \vdots \\ m_{i_k} \end{bmatrix}$$

$$= \sum_{i_1\cdots i_k} \left( \begin{bmatrix} m_{i_1} & \cdots & m_{i_k} \end{bmatrix} N_{i_1\cdots i_k}^T \right) \left( \begin{bmatrix} m_{i_1} \\ \vdots \\ m_{i_k} \end{bmatrix} \right).$$

The last expression shows that $p$ is a $k$DDSOS.

Now if $p$ is a $k$DDSOS, as shown in lemma 4.2, we could write

$$p = \sum_{i_1\cdots i_k} \sum_{j=1}^k \left( \sum_{i=1}^k \alpha_{i,j} m_{i,j} \right)^2,$$

where $m_{i,j}$ are different for different $i$. This gives our

$$N_{i_1\cdots i_k} = \begin{bmatrix} \alpha_{1,i_1} & \cdots & \alpha_{1,i_k} \\ \alpha_{2,i_1} & \cdots & \alpha_{2,i_k} \\ \vdots & \cdots & \vdots \\ \alpha_{k,i_1} & \cdots & \alpha_{k,i_k} \end{bmatrix}.$$

We then can construct $M^{i_1\cdots i_k}$ and $A$.

We define the corresponding $k$DDSOS program here.

**Definition 4.4.** Denote the cone of $k$DDSOS with degree $2d$ and $n$ variables as $kSOS_{n,d}$. We call the following optimization $k$DDSOS programming.

$$\begin{array}{ll}
\text{minimize} & r^T u \\
\text{subject to} & r_{0,t} + r_{1,t}(x) u_1 + \cdots + r_{s_t,t}(x) u_{s_t} \in kSOS_{n_t,d_t}, t = 1, 2, \ldots, N,
\end{array}$$

(4.1)

where $r(x)$s are given polynomials and $n_t, d_t$ depends on $r(x)$. $n_t$ is the total number of variables of $r(x)$ in the same inequality. $d_t$ is half the highest degree of $r(x)$ in the same inequality.

To link to our previous discussion of $kOCP$ induced by $S_+^k$, we show that these two programs are equivalent.

**Theorem 4.5.** $k$DDSOS programming is equivalent to $(S_+^d)^k$ cone programming ($kOCP$ induced by $S_+^k$).

**Proof.** We first show how to reduce $(S_+^d)^k$ cone program to $k$DDSOS program:

We may suppose $(S_+^d)^k$ program is in its standard form, i.e., the form in (2.4) (the equivalence between standard form and inequality form for $(S_+^d)^k$ can be proved via standard techniques). To avoid confusion, suppose $U$ is the variable matrix in $(S_+^d)^k$ cone program.

Then our $r$ in $k$DDSOS program (4.1) is just $\text{vec}(A_0)$. The linear equality can be incorporated into a $k$DDSOS inequality by let $rs$ in the inequality in (4.1) be constant and matches $(D_{i,j,k}, -(D_{i,j,k})$. 

...
as the following
\[ \text{tr}(D_i U) = f_i \iff -f_i + \sum_{jk} (D_i)_{jk} u_{ij} \in k\text{SOS}_{1,1} \text{ and } f_i + \sum_{jk} -(D_i)_{jk} u_{ij} \in k\text{SOS}_{1,1}. \]

The condition \( U \in (S^d_+)_k \) is the same as
\[ \sum_{1 \leq i,j \leq d} x_{ij} u_{ij} \in k\text{SOS}_{d,1} \]
by Theorem 4.3.

Next we show how to reduce \( k\text{DDSOS} \) program to \((S^d_+)_k\) cone program in its inequality form. The objective is the same for both programs.

The constraint \( p_l(x) = r_{0,l}(x) + r_{1,l}(x) u_1 + \cdots + r_{s_l,t}(x) u_k \in k\text{SOS}_{n_t,d_t} \) is the same as there is one \( A = [a_{ij}]_{ij} \in (S^h_+)_k \) for some \( h \) such that \( p_l(x) = m^T A m \). Thus
\[ p_l(x) = r_{0,l}(x) + r_{1,l}(x) u_1 + \cdots + r_{s_l,t}(x) u_t \in k\text{SOS}_{nt,d_t} \]
if and only if there exists
\[ A \in (S^h_+)_{nt+d_t}, \text{ and linear constraints on } a_{ij}, u_i, \]
where the linear constraints come from matching coefficients of
\[ p_l(x) = m^T A m = r_{0,l}(x) u_1 + \cdots + r_{s_l,t}(x) u_t. \]
The condition \( A \in (S^h_+)_{nt+d_t} \) is a \( k\text{OC} \) constraint and we could add variable \( a_{ij} \) to \( k\text{OCP} \). This shows the other direction.

\[ \square \]

5. Completely positive cone and copositive cone

Recall the following definition of completely positive matrices and copositive matrices:

- The set of copositive matrices with dimension \( d \), \( \text{COP}^d \):
  \[ \text{COP}^d := \{ M \in S^d : x^T M x \geq 0 \text{ for all } x \in \mathbb{R}^d \}. \]

- The set of complete positive matrices with dimension \( d \), \( \text{CP}^d \):
  \[ \text{CP}^d := \{ B^T B : B \in \mathbb{R}^{d \times m}, m \text{ is an integer} \}. \]

These two cones satisfy the embedding property throughly by verifying the definition directly. The corresponding copositive programming and copositive programming gives a lot modeling power in combinatorics and nonconvex problems [Dur10, Bur15]. However, these programs are NP-hard to solve in general.

Using the construction of \( k\text{OCP} \) induced by \( \text{COP}^k \) or \( \text{CP}^k \), for \( k = 1, 2, 3, 4 \), we are able to solve the the inner approximation of copositive programming and copositive programming. The case \( k = 2 \) of \( \text{CP}^k \) has been explored in [BGPI18].

**Theorem 5.1.** \( 2, 3, 4\text{-OCP} \) with index set \( J \) induced by \( \text{COP}^k \) or \( \text{CP}^k \) can be casted into \( 2, 3, 4\text{-OCP induced by } S^k_+ \).

The theorem is mainly due to the following lemma:

**Lemma 5.2.** [MM62] Denote \( N^k_+ = (\mathbb{R}_+^{k \times k}) \cap S^k_+ \), we have for \( k = 1, 2, 3, 4 \)
\[ \text{COP}^k = S^k_+ + N^k_+, \quad \text{CP}^k = S^k_+ \cap N^k_+. \]
Proof of Theorem 5.1. We may suppose $i$-OCP with index set $J$ induced by $\text{COP}^k$ or $\text{CP}^k$ are in its standard form \(^{(2.4)}\) where $i = 2, 3, 4$. The case of inequality form is similar.

The constraint $X \in \text{COP}_i^d$ is the same as

$$X \in \text{COP}_i^d \iff X = \sum_{\{j_1, \ldots, j_i\} \in J} \varepsilon_{j_1 \ldots j_i}(M_{j_1 \ldots j_i}) \quad \text{and} \quad M_{j_1 \ldots j_i} \in \text{COP}^i.$$ 

Since $M_{j_1 \ldots j_i} \in \text{COP}^i$ if and only if $M_{j_1 \ldots j_i} = S_{j_1 \ldots j_i} + N_{j_1 \ldots j_i}$ for some $S_{j_1 \ldots j_i} \in S^i_+, N_{j_1 \ldots j_i} \in N^i_+$ by Lemma 5.2. We see $i$-OCP with index set $J$ induced by $\text{COP}^k$ can be casted into $i$OCP induced by $S^i_+$ (the constraint $S_{j_1 \ldots j_i} \in S^i_+$ can be casted into $(S^i_+)_k$ constraint by setting $d = k$ and the nonnegative constraint can be handled via $\text{diag}(x) \in (S^i_+)_k \iff x \geq 0$).

For $i$-OCP with index set $J$ induced by $\text{CP}^k$. We note that

$$X \in \text{COP}_i^d \iff X = \sum_{\{j_1, \ldots, j_i\} \in J} \varepsilon_{j_1 \ldots j_i}(M_{j_1 \ldots j_i}) \quad \text{and} \quad M_{j_1 \ldots j_i} \in \text{CP}^i.$$ 

Since $M_{j_1 \ldots j_i} \in \text{CP}^i$ if and only if $M_{j_1 \ldots j_i} \in S^i_+$ and $M_{j_1 \ldots j_i} \in N^i_+$ by Lemma 5.2. We see $i$-OCP with index set $J$ induced by $\text{COP}^k$ can be casted into $i$OCP induced by $S^i_+$.

By adjusting the set $J$ and an $U \in \mathbb{R}^{d \times d}$, we may consider solving

\[
\begin{align*}
\text{minimize} & \quad \text{tr}(A_0X) \\
\text{subject to} & \quad \text{tr}(A_iX) = b_i, \quad i = 1, \ldots, p \\
& \quad X \in UC\text{OP}_k^d(J)U^T.
\end{align*}
\]

and

\[
\begin{align*}
\text{minimize} & \quad \text{tr}(A_0X) \\
\text{subject to} & \quad \text{tr}(A_iX) = b_i, \quad i = 1, \ldots, p \\
& \quad X \in UC\text{P}_k^d(J)U^T.
\end{align*}
\]

This formulation gives us more modeling power and can also be casted into $k$OCP induced by $S^i_+$ for $k = 2, 3, 4$.

6. Symmetric cones

6.1. Positive semidefinite matrices in $\mathbb{R}^{d \times d}$, $\mathbb{C}^{d \times d}$, $\mathbb{H}^{d \times d}$ and $\mathbb{O}^{d \times d}$. Let us first recall the five irreducible symmetric cones\(^1\).

(i) Symmetric real positive semidefinite matrices in $\mathbb{S}^d$

(ii) Hermitian complex positive semidefinite matrices in $\mathbb{C}^{d \times d}$

(iii) Hermitian quaternion positive semidefinite matrices in $\mathbb{H}^{d \times d}$

(iv) Hermitian octonian positive semidefinite $\mathbb{O}^{3 \times 3}$ matrices

(v) Second order cone in $\mathbb{R}^{d+1}$: $\text{SOC}^{d+1} = \{(t, x) \mid \|x\|_2 \leq t, x \in \mathbb{R}^d, t \in \mathbb{R}\}$.

For the first three cones, they satisfy the embedding property throughly as they are all of the form

$$\{A \in \mathbb{S}^d : x^*Ax \geq 0, \text{ for all } x \in \mathbb{R}^d\},$$

where $F = \mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$.

Let

$$\mathcal{HO}^d_+ = \{A \in \mathbb{O}^{d \times d} : A = A^*, x^*Ax \geq 0, \text{ for all } x \in \mathbb{O}^d\}.$$ 

We may consider the series $\{\mathcal{HO}^k\}_{k=1}^\infty$ so that the cone of Hermitian octonian positive semidefinite $\mathbb{O}^{3 \times 3}$ matrices is a member of it. The series satisfies the embedding property throughly.

\(^1\)A cone is symmetric if it is self-dual and its autonomous group acts transitively on it. A symmetric cone is irreducible means it cannot be written as a Cartesian product of other symmetric cones.
6.2. **Second Order Cone.** We need to first transform the second order cone into the space of symmetric matrices. This can be done through:

\[ A : A = \text{diag}(t, x), \] for some \((t, x) \in \text{SOC}^k \}

We abuse the notation and call the above set as \(\text{SOC}^k\) as well. Moreover, we define \(\text{SOC}^1 = \mathbb{R}_+\). The index map for \(\{\text{SOC}^k\}_{k=1}^\infty\) is then

\[ I\text{SOC}(d, k) = \{s : s = (1, i_1, \ldots, i_{k-1}), 2 \leq i_1 < \cdots < i_{k-1} \leq d\} \]

for \(k \geq 2\) and is simply \(\{1\}\) if \(k = 1\). It can be easily verified that \(\{\text{SOC}^k\}_{k=1}^\infty\) satisfies the index map with index map \(I(d, k)\).

We can avoid lifting the second order cone to \(d \times d\) matrices. First, we define \(\tau_{i_1 \cdots i_k} : \mathbb{R}^d \to \mathbb{R}^k\), \(\varepsilon_{i_1 \cdots i_k}^d : \mathbb{R}^k \to \mathbb{R}^d\) for every \((i_1, \ldots, i_k) \in \binom{d}{k}\) such that

\[ \tau_{i_1 \cdots i_k}(x) = (x_{i_1}, \ldots, x_{i_k}), \quad [\varepsilon_{i_1 \cdots i_k}^d(y)]_i = \begin{cases} y_j & i = j \text{ for some } i_j \\ 0 & \text{otherwise} \end{cases} \]

The \(k\)th higher order cone of \(\text{SOC}^d\) is then

\[ \text{SOC}^d_k = \{x \in \mathbb{R}^d : x = \sum_{(1, i_1, \ldots, i_{k-1}) \in I\text{SOC}(d, k)} \varepsilon_{1i_1 \cdots i_{k-1}}(x_{1i_1 \cdots i_{k-1}}), x_{1i_1 \cdots i_{k-1}} \in \text{SOC}^k\}. \]

The following Lemma shows the strict inclusion of \(\{\text{SOC}^d_k\}_{k=1}^\infty\) is strict.

**Lemma 6.1.** We have

\[ \text{SOC}^d_1 \subset \text{SOC}^d_2 \subset \text{SOC}^d_3 \subset \cdots \subset \text{SOC}^d_d = \text{SOC}^d. \]

**Proof.** The inclusion follows easily from the Nested Cone property in Proposition 2.3. We now prove the inclusion is actually strict. First, we consider the dual cones

\[ (\text{SOC}^d_k)^* = \{x \in \mathbb{R}^d : \tau_{i_1 \cdots i_k}^d(x) \in \text{SOC}^k \text{ for all } (1, i_1, \ldots, i_{k-1}) \in I\text{SOC}(d, k)\} . \]

An application of first and second item of 2.3 tells us that

\[ (\text{SOC}^d_k)^* \supseteq (\text{SOC}^d_2)^* \supseteq (\text{SOC}^d_3)^* \supseteq \cdots \supseteq (\text{SOC}^d_d)^* = (\text{SOC}^d)^* . \]

Consider \((\sqrt{k-1}, 1_{d-1})\), where \(1\) is a all one vector with length \(d - 1\). This vector belongs \((\text{SOC}^d_k)^*\) but not \((\text{SOC}^d_{k+1})^*\). Thus the inclusion in the dual cones is strict.

Since \((\text{SOC}^d_k)^* = \bigcap_{(1, i_1, \ldots, i_{k-1}) \in I\text{SOC}(d, k)} K_{i_1 \cdots i_{k-1}}\) where \(K_{i_1 \cdots i_{k-1}} = \{x \in \mathbb{R}^d : \tau_{i_1 \cdots i_{k-1}}^d(x) \in \text{SOC}^k\}\), and \((\sqrt{d+1}, 1_{d-1}) \in \text{int}(K_{i_1 \cdots i_{k-1}})\), by the Krein-Rutman Theorem [BL10, Corollary 3.3.13], we have

\[ (\text{SOC}^d_k)^{**} = \sum_{(1, i_1, \ldots, i_{k-1}) \in I\text{SOC}(d, k)} \varepsilon_{1i_1 \cdots i_{k-1}}^d(\text{SOC}^k) = \text{SOC}^d_k. \]

Thus the strict inclusion in the dual cone implies that strict inclusion in \(\{\text{SOC}^d_k\}_{k=1}^d\). \(\square\)

7. **Norm Cones**

The embedding property property is also satisfied by a large class of norm cones. Specifically, the property we need is the following.

**Definition 7.1.** Suppose a norm \(\|\cdot\|\) is defined on \(\mathbb{S}^d\) (or \(\text{diag}(\mathbb{R}^d)\)) for all \(d\). For any \(1 \leq k \leq d\), \(X \in \mathbb{S}^d\) (or \(\text{diag}(\mathbb{R}^d)\)) and any \((i_1, \ldots, i_k) \in \binom{d}{k}\), it is

(i) **consistent** if \(\|\tau_{i_1 \cdots i_k}(X)\| = \|\varepsilon_{i_1 \cdots i_k}(\tau_{i_1 \cdots i_k}(X))\|\);

(ii) **monotonic** if \(\|\tau_{i_1 \cdots i_k}(X)\| \leq \|X\|\).

Norms satisfied the consistency and monotonicity are abundant, for example,
(a) All $\ell_p$ norms on $\mathbb{R}^d$: $\|x\|_p = (\sum_{i=1}^{d}|x_i|^p)^{\frac{1}{p}}$ for any $x \in \mathbb{R}^d, p \geq 1$.

(b) All Schatten norm on $\mathcal{S}^d$ with underlying field being $\mathbb{R}$ or $\mathbb{C}$: $\|X\|_p = (\sum_{i=1}^{d} |\sigma_i(X)|^p)^{\frac{1}{p}}$ for all $X \in \mathcal{S}^d$ where $\sigma_i(X)$ is the $i$th largest singular value of $X$. The monotonicity is due to Cauchy’s interlace theorem.

(c) All Ky-Fan $k$ norm on $\mathcal{S}^d$ with underlying field being $\mathbb{R}$ or $\mathbb{C}$: $\|X\|_{KF_k} = \sum_{i=1}^{k} \sigma_i(X)$ for all $X \in \mathcal{S}^d$ and $\sigma_i = 0$ for $i > d$. The monotonicity is also due to Cauchy’s interlace theorem.

(d) The operator norm of matrix induced by $\ell_p, \ell_q$ vector norms: $\|A\|_{p,q} = \sup_{\|x\|_p = 1} \|Ax\|_q$ for any $1 \leq p, q \leq \infty$.

In fact, these two properties turn out to be the characterization of norms so that its corresponding norm cones having embedding property as $\text{SOC}$. This fact is shown by the following theorem.

**Theorem 7.2 (Characterization of Norm Cones satisfying embedding property as SOC).** For a norm $\|\cdot\|$ defined on $\mathcal{S}^d$ (or diag($\mathbb{R}^d$)) for all $d \geq 1$, let the norm cone in $\mathcal{S}^{d+1}$ (or diag($\mathbb{R}^{d+1}$) be $N^{d+1}_{\|\cdot\|} = \{\text{diag}(t, X) \mid \|X\| \leq t, X \in \mathcal{L}^d, t \in \mathbb{R}\}$, and $N^1_{\|\cdot\|} = \mathbb{R}_+$ where $\mathcal{L}^d = \mathcal{S}^d$ or diag($\mathbb{R}^d$). If the norm is consistent and monotonic, then the series of norm cones $\{N^k_{\|\cdot\|}\}_{k=1}^{\infty}$ satisfies the embedding property with index map $I_{\text{SOC}}(d, k)$. The converse is also true.

**Proof.** We prove the case of $\mathcal{S}^d$. The case of diag($\mathbb{R}^d$) follows exactly the same line.

We first show that consistency with monotonicity implies that $N^d_{\|\cdot\|}$ satisfies the embedding property with index map $I_{\text{SOC}}$. For any $1 \leq k \leq d$, $(1, i_1, \ldots, i_{k-1}) \in I_{\text{SOC}}(d, k)$, diag$(t, X) \in N^d_{\|\cdot\|}$, and diag$(s, Z) \in N^k_{\|\cdot\|}$, the consistency implies that

$$\varepsilon_{1i_1 \ldots i_k}(\text{diag}(s, Z)) \in N^d_{\|\cdot\|},$$

since

$$\varepsilon_{1i_1 \ldots i_k}(\text{diag}(s, Z)) = \text{diag}(s, \varepsilon_{i_1 \ldots i_k}(Z)), \quad \text{and} \quad \|\varepsilon_{1i_1 \ldots i_k}(Z)\| = \|Z\| \leq s.$$  

The monotonicity implies that

$$\tau_{1i_1 \ldots i_k}(\text{diag}(t, X)) \in N^{k+1}_{\|\cdot\|},$$

since

$$\tau_{1i_1 \ldots i_k}(\text{diag}(t, X)) = \text{diag}(t, \tau_{i_1 \ldots i_k}(X)), \quad \text{and} \quad \|\tau_{i_1 \ldots i_k}(X)\| \leq \|X\| \leq t.$$  

The case $k = 0$ is trivial.

Next we show the embedding property of $\{N^k_{\|\cdot\|}\}_{k=1}^{\infty}$ implies its consistency and monotonicity. Due to the embedding property of $\{N^k_{\|\cdot\|}\}_{k=2}^{d}$, we have for any $1 \leq k \leq d$, $(1, i_1, \ldots, i_{k-1}) \in I_{\text{SOC}}(d, k)$, $Z \in \mathcal{S}^k$,

$$\text{diag}(\|Z\|, Z) \in N^{k+1}_{\|\cdot\|} \implies \varepsilon_{1i_1 \ldots i_k}(\text{diag}(\|Z\|, Z)) \in N^{d+1}_{\|\cdot\|} \implies \|\varepsilon_{1i_1 \ldots i_k}(Z)\| \leq \|Z\|.$$  

Now consider

$$\text{diag}(\|\varepsilon_{1i_1 \ldots i_k}(Z)\|, \varepsilon_{1i_1 \ldots i_k}(Z)) \in N^{d+1}_{\|\cdot\|} \implies \tau_{1i_1 \ldots i_k}(\text{diag}(\|\varepsilon_{1i_1 \ldots i_k}(Z)\|, \varepsilon_{1i_1 \ldots i_k}(Z))) \in N^{k+1}_{\|\cdot\|}.$$  

But $\tau_{1i_1 \ldots i_k}(\text{diag}(\|\varepsilon_{1i_1 \ldots i_k}(Z)\|, \varepsilon_{1i_1 \ldots i_k}(Z)))) = \text{diag}(\|\varepsilon_{1i_1 \ldots i_k}(Z)\|, Z)$ and we have

$$\|\varepsilon_{1i_1 \ldots i_k}(Z)\| \geq \|Z\|.$$  

This shows the consistency by taking $Z = \tau_{i_1 \ldots i_k}(X)$.

To prove monotonicity, we have for any $X \in \mathcal{S}^d$,

$$\tau_{1i_1 \ldots i_k}(\text{diag}(t, X)) \in N^{k+1}_{\|\cdot\|} \implies \|\tau_{i_1 \ldots i_k}(X)\| \leq t$$  

and taking $t = \|X\|$ shows the monotonicity.  

$\square$
Thus the norm cones of our previous mentioned four kinds of norm (1) $\ell_p$ norm on $\mathbb{R}^d$, (2) Schatten norm on $\mathbb{S}^d$ with underlying field being $\mathbb{R}$ or $\mathbb{C}$, (3) Ky-Fan $k$ norm with underlying field being $\mathbb{R}$ or $\mathbb{C}$, and (4) operator norm induced by $p,q$ norms all satisfies the embedding property with index map $I_{\text{SOC}}$. This means our previous discussion on $\text{SOC}$ is just a special case of norm cones with embedding property.

Here we give two more concrete examples of norm with consistency and monotonicity and studies its $k$th order cone. Let us first consider the $\ell_1$ norm in $\mathbb{R}^d$. As in the case of the second order cone, we don’t need to lift the space to matrices. The second order cone induced by $N_{\ell_1}^{d+1}$ is

$$(N_{\ell_1}^{d+1})_2 = \{ (t, x) \mid (t, x) = (\sum_{i=1}^{d} t_i + \cdots + t_d, x_1, \ldots, x_d), \text{ for all } i, t_i \geq |x_i|, t_i, x_i \in \mathbb{R} \}.$$ 

which is simply $N_{\ell_1}^{d+1}$! Thus by first item of Proposition 2.3 we know the $k$th order cone induced by $N_{\ell_1}^{d+1}$ is just itself $N_{\ell_1}^{d+1}$ for $k \geq 2$. We don’t gain new cones from this construction except the trivial cone $(N_{\ell_1})^{d+1} = \mathbb{R}_+ \times \{ 0_d \}$ where $0_d$ is the zero vector of length $d$. Note that this is not the case for the second order cone.

Next we consider the nuclear norm $\| \cdot \|_*$:

$$\|X\|_* = \sum_{i=1}^{d} \sigma_i(X), \text{ for all } X \in \mathbb{S}^d$$

with underlying field being real or complex. The $(k+1)$th order cone induced by $N_{\|\cdot\|_*}^{d+1}$ is

$$(N_{\|\cdot\|_*}^{d+1})_{k+1} = \big\{ \sum_{(1, i_1, \ldots, i_k)} \varepsilon_{i_1 \cdots i_k} (\text{diag}(t_{i_1 \cdots i_k}, X_{i_1 \cdots i_k})) \mid \text{diag}(t_{i_1 \cdots i_k}, X_{i_1 \cdots i_k}) \in N_{\|\cdot\|_*}^{k+1} \big\}.$$ 

Since $(N_{\|\cdot\|_*}^{d+1})^* = N_{\|\cdot\|_2}^{d+1}$ where $\| \cdot \|_2$ is the operator two norm, we know from second item of Proposition 2.3 the dual cone of $(N_{\|\cdot\|_*}^{d+1})_{k+1}$ is

$$((N_{\|\cdot\|_*}^{d+1})_{k+1})^* = \{ \text{diag}(t, X) \mid \tau_{1 i_1 \cdots i_k} (\text{diag}(t, X)) \in N_{\|\cdot\|_*}^{k+1}, \text{ for all } (1, i_1, \ldots, i_{k-1}) \in I_{\text{SOC}}(d, k) \}.$$ 

Moreover, by an application of first and second items of Proposition 2.3 we have

$$((N_{\|\cdot\|_*}^{d+1})_{d+1})^* \subseteq ((N_{\|\cdot\|_*}^{d+1})_d^*)^* \subseteq \cdots \subseteq ((N_{\|\cdot\|_*}^{d+1})_2^*)^* \subseteq ((N_{\|\cdot\|_*}^{d+1})_1^*)^*.$$ 

By considering $\text{diag}(k, I_d + 11^\top)$ with $k = 1, 2, \ldots, d$ where $I_d$ is the identity matrix in $\mathbb{S}^d$ and $1$ is the all one vector, we find that

$$((N_{\|\cdot\|_*}^{d+1})_{d+1})^* \subseteq ((N_{\|\cdot\|_*}^{d+1})_d^*)^* \subseteq \cdots \subseteq ((N_{\|\cdot\|_*}^{d+1})_2^*)^* \subseteq ((N_{\|\cdot\|_*}^{d+1})_1^*)^*.$$ 

Since $\text{diag}(d+1, I_d) \in \text{int}(K_{1 \cdots i_k}) = \text{int}\{ (\text{diag}(t, X) : \tau_{1 i_1 \cdots i_k} (\text{diag}(t, X)) \in N_{\|\cdot\|_*}^{k+1}) \}$ and $((N_{\|\cdot\|_*}^{d+1})_{k+1})^* = \cap_{(1, i_1, \ldots, i_{k-1}) \in I_{\text{SOC}}(d, k)} K_{1 \cdots i_k}$, by the Krein-Rutman Theorem [BL10 Corollary 3.3.13], we find that

$$(N_{\|\cdot\|_*}^{d+1})_{d+1} \supseteq (N_{\|\cdot\|_*}^{d+1})_d \supseteq \cdots \supseteq (N_{\|\cdot\|_*}^{d+1})_2 \supseteq (N_{\|\cdot\|_*}^{d+1})_1.$$ 

Thus, unlike the case of $\ell_1$ norm cone, we indeed obtain new cones here.

Finally, the $k$th order cone program induced by $(N_{\|\cdot\|_*}^{d+1})_k$ for monotonic and consistent norm $\| \cdot \|$ is

minimize \quad $\text{tr}(A_0 X) + a_0 t$
subject to \quad $\text{tr}(A_i X) + a_i t = b_i, \quad i = 1, \ldots, p$
$\text{diag}(t, X) \in (N_{\|\cdot\|_*}^{d+1})_k$,

where $A_i \in \mathbb{S}^d$, $a_i$, $b_i \in \mathbb{R}$, for all $i = 0, \ldots, p$. 
8. KKT Condition and Self-Concordance

8.1. KKT Condition. Here we list the KKT condition for our higher order cone program. The primal form of our $K^d_k(J)$ program is

$$\begin{align*}
\text{minimize} & \quad \text{tr}(A_0 X) \\
\text{subject to} & \quad \text{tr}(A_i X) = b_i, \quad i = 1, \ldots, p \\
& \quad X \in K^d_k(J)
\end{align*}$$

(8.1)

The dual of the above program is

$$\begin{align*}
\text{minimize} & \quad b^T y \\
\text{subject to} & \quad A_0 - \sum_{i=1}^p A_i y_i^* \in (K^d_k(J))^*.
\end{align*}$$

(8.2)

Let $X^*, y^*$ be a primal and dual solution pair of the above programs. Also let $Z^* = A_0 - \sum_{i=1}^p A_i y_i^*$. If strong duality holds:

$$\text{tr}(A_0 X^*) = b^T y^*,$$

we have the KKT condition as

$$\begin{align*}
X^* & \in K^d_k(J) \\
\text{tr}(A_i X^*) & = b_i, \quad i = 1, \ldots, p \\
Z^* & \in (K^d_k(J))^*, \\
\text{tr}(Z^* X^*) & = 0, \\
A_0 - \sum_{i=1}^p A_i y_i^* & = Z^*.
\end{align*}$$

(8.3)

8.2. Self-Concordance. We assume the original cone $K^k$ and its $k$th order cone $K^d_k$ are proper and the underlying field is $\mathbb{R}$. The index set is $I(d, k)$. The assumption of properness on $k$-th order is true for all previous mentioned examples in $\mathbb{S}^d$.

Recall the definition of self-concordance and a few propositions of it.

**Definition 8.1.** Let $K$ be a convex closed cone. A continuous function $f : K \to \mathbb{R} \cup \{+\infty\}$ is called a barrier function of $K$ if it satisfies

$$f(x) < \infty \quad \text{for every } x \in \text{int}(K) \quad f(x) = +\infty \quad \text{for every } x \in \partial K,$$

where $K^\infty$ means taking the interior of $K$ and $\partial K$ means the boundary of $K$ induced by the usual topology in $\mathbb{R}^n$.

A convex third order differentiable function $f(x)$ on $K$ is self-concordant if for every $x \in K^\infty$ and $h \in \mathbb{R}^n$ the univariate function $\phi(\alpha) = f(x + \alpha h)$ satisfies the property

$$|\phi'''(0)| \leq 2|\phi''(0)|^{\frac{3}{2}}.$$ 

A barrier $f(x)$ of $K$ is logarithmically homogeneous of degree $\theta$ if

$$f(tx) = f(x) - \theta \log(t).$$

The following property is an easy consequence of the definition of self-concordance and can be found in section 9.6 in [BV04].

**Proposition 8.2.** If $f_1, f_2$ are self-concordant functions on $K \subseteq \mathbb{R}^n$, then the following functions are also self-concordant.

(i) $af$, for all $a \geq 1$
(ii) \( f_1 + f_2 \)
(iii) \( f(Ax + b) \) for all \( A \in \mathbb{R}^{n \times m}, b \in \mathbb{R}^n \).

The following theorem is adapted from Theorem 2.4.2, Theorem 2.4.4, and Proposition 2.4.1 in [NN94]. One can also find this in section 11.6 in [BV04].

**Theorem 8.3.** Let \( K \) be a proper cone, i.e., \( K \) is solid, convex, pointed and closed, in \( \mathbb{R}^n \) and let \( f \) be a \( \theta \)-logarithmically homogeneous self-concordant barrier for \( K \). Then the Fenchel conjugate \( f^* \) of \( f \) is a \( \theta \)-logarithmically homogeneous self-concordant barrier for \(-K^*\), i.e, the polar dual of \( K \).

Moreover, we have the interior of \(-K^*\) to be

\[
\text{int}(-K^*) = \{ \nabla f(x) : x \in \text{int}(K) \},
\]

and

\[
f^*(x) + f(y) + \theta \log(-x^Ty) \geq \theta \log(\theta) - \theta
\]

where the equality holds for if and only if \( x = t\nabla f(y) \) for some \( t > 0 \).

We prove the following theorem when a self-concordance function of the dual cones \((K^k)^*\) is available.

**Theorem 8.4.** Let \( g \) be a \( \theta \)-logarithmically self-concordant barrier of the dual cone \((K^k)^*\). Also let \( f(Y) = \sum_{(i_1, \ldots, i_k) \in I(d,k)} g(\tau^d_{i_1 \ldots i_k}(Y)) \), \( Y \in \text{int}(\text{int}(K^k)^*) \). Assuming \( \nabla f := x \mapsto \nabla f(x) \) is invertible, and \((K^d)^*\) is a proper cone, the function \( F(X) = -\text{tr}(X(\nabla f)^{-1}(-X)) - f((\nabla f)^{-1}(-X)) \) is a \( \theta \) \( \text{card}(I(d,k)) \)-logarithmically self-concordant barrier for \( K^d_k \).

**Proof.** We first show \( f \) is a \( \theta \text{card}(I(d,k)) \)-logarithmically self-concordant barrier of the dual cone \((K^d)^*\) where \( \text{card}(I(d,k)) \) is the cardinality of \( I(d,k) \).

The barrier property follows from the fact that the boundary of \((K^d_k)^*\) are those \( Y \)s such that some of \( \tau^d_{i_1 \ldots i_k}(Y) \) are on the boundary of \((K^d)^*\).

To verify that \( f \) is self-concordant, we only need to show that for all \( X \) in the interior of \((K^d_k)^*\), \( V \in S^n \), \( \phi(t) = f(X + tV) = \sum_{(i_1, \ldots, i_k) \in I(d,k)} g(\tau^d_{i_1 \ldots i_k}(X + tV)) \) is self-concordant.

By Proposition 8.2, it is enough to show \( g(\tau^d_{i_1 \ldots i_k}(X + tV)) \) is self-concordant. Since \( X \) is in the interior, we know \( \tau^d_{i_1 \ldots i_k}(X + tV) \) is indeed in \( \text{int}(\text{int}(K^k)^*) \) for all small \( t \) and so \( g(\tau^d_{i_1 \ldots i_k}(X + tV)) \) is self-concordant as \( g \) is. This proves \( f \) is self-concordant on \((K^d_k)^*\).

From the following computation,

\[
f(tY) = \sum_{(i_1, \ldots, i_k) \in I(d,k)} g(\tau^d_{i_1 \ldots i_k}(tY)) \]
\[
\overset{(a)}{=} \sum_{(i_1, \ldots, i_k) \in I(d,k)} g(\tau^d_{i_1 \ldots i_k}(Y)) - \theta \log t \]
\[
= \sum_{i(i_1, \ldots, i_k) \in I(d,k)} g(\tau^d_{i_1 \ldots i_k}(Y)) - \text{card}(I(d,k))\theta \log t,
\]

where (a) is because \( g \) is \( \theta \)-logarithmically homogeneous. We see \( f \) is indeed logarithmically homogeneous of degree \( \text{card}(I(d,k))\theta \).

By Theorem 8.3 we know that \( f^*(X) \) is indeed \( \theta \)-logarithmically homogeneous self-concordant barrier for \(-K^d_k\). The Fenchel-Young’s inequality asserts that

\[
f^*(X) + f(Y) \geq \text{tr}(XY)
\]

and this becomes equality if \( X = \nabla f(Y) \).
Since $\nabla f$ is invertible from $\text{int}((K^d_k)^*)$ to its image which from Theorem 8.3 is just $\text{int}(-K^d_k)$, $\nabla f$ is bijective from the interior of the dual cone to the interior of $-K^d_k$. Thus the notation $(\nabla f)^{-1}$ always makes sense. We have
\[
 f^*(X) = \text{tr}(X(\nabla f)^{-1}(X)) - f((\nabla f)^{-1}(X))
\]
and so $F(X)$ is indeed a $\theta \text{card}(I(d,k))$-logarithmically self-concordant barrier of $K^d_k$.

The condition $\nabla f := x \mapsto \nabla f(x)$ is invertible is satisfied when $f$ has positive definite Hessian (see Lemma 9.3 in Appendix). This is the case for $(S^d_k)^*).

**Lemma 8.5.** The function $f(x) = \sum_{(i_1,\ldots,i_k)\in I(d,k)} -\log(\det(\tau^d_{i_1\ldots i_k}(Y)))$ is a $k \binom{n}{k}$-logarithmically homogeneous self-concordant, strictly convex barrier on $(S^d_k)^*$ and has positive definite Hessian on the interior of $(S^d_k)^*$.

**Proof.** The cone $(S^d_k)^*$ can be easily verified to be proper. We only need to show the Hessian is positive definite as other parts are due to $-\log(\det(S))$ is self-concordant for $S \in \text{int}(S^d_k)$.

Since first order approximation of $f$ is

\[
 f(Y + \alpha H) = \sum_{i_1\ldots i_k} -\log(\det(\tau^d_{i_1\ldots i_k}(Y + \alpha H)))
\]

\[
 = \sum_{(i_1,\ldots,i_k)\in I(d,k)} -\log(\det(\tau^d_{i_1\ldots i_k}(Y))) - \alpha \text{tr}(\tau^d_{i_1\ldots i_k}(Y)^{-1}\tau^d_{i_1\ldots i_k}(H)) + O(\alpha^2)
\]

The first order derivative is

\[
 f'(Y) = -\sum_{(i_1,\ldots,i_k)\in I(d,k)} \varepsilon^d_{i_1\ldots i_k}(\tau^d_{i_1\ldots i_k}(Y)^{-1}).
\]

Now if we approximate the derivative up to the first order, we have

\[
 f'(Y + \alpha H) = \sum_{(i_1,\ldots,i_k)\in I(d,k)} -\varepsilon^d_{i_1\ldots i_k}(\tau^d_{i_1\ldots i_k}(Y)^{-1}) + \alpha \varepsilon^d_{i_1\ldots i_k}(\tau^d_{i_1\ldots i_k}(Y)^{-1}\tau^d_{i_1\ldots i_k}(H)) + O(\alpha^2)
\]

Thus we see

\[
 D^2 f(X)[H,H] = \text{tr}(\sum_{(i_1,\ldots,i_k)\in I(d,k)} \varepsilon^d_{i_1\ldots i_k}(\tau^d_{i_1\ldots i_k}(Y)^{-1}\tau^d_{i_1\ldots i_k}(H)\tau^d_{i_1\ldots i_k}(Y)^{-1}H))
\]

\[
 = \sum_{(i_1,\ldots,i_k)\in I(d,k)} \text{tr}((\tau^d_{i_1\ldots i_k}(Y)^{-1}\tau^d_{i_1\ldots i_k}(H)\tau^d_{i_1\ldots i_k}(Y)^{-1}\tau^d_{i_1\ldots i_k}(H))
\]

\[
 = \sum_{(i_1,\ldots,i_k)\in I(d,k)} \text{tr}((\tau^d_{i_1\ldots i_k}(Y)^{-1}\frac{1}{2}\tau^d_{i_1\ldots i_k}(H)\tau^d_{i_1\ldots i_k}(Y)^{-1}\frac{1}{2})^2)
\]

where $D^2 f[H,H]$ denotes the value of second differential of $f$ taken at $x$ along the direction $H,H$. The last term is greater than zero for non-zero $H$. This means that $f$ is strictly convex and its Hessian is positive definite.

9. **Appendix**

Here we prove a few results in the main text. We first prove a simple Lemma used in proving Theorem 2.4 which states the equivalence between standard form and inequality form of $kOCP,$
Lemma 9.1. Suppose \( \{K^k\}_{k=1}^\infty \) satisfies the embedding property thoroughly. If \( x \in \mathbb{F}^d \) and \( d \geq k \) and \( K^1 = \mathbb{R}_+ \), then
\[
\text{diag}(x) \in K_k^d \iff x \geq 0,
\]
where \( x \geq 0 \) means each component of \( x \) is greater or equal to 0.

Proof. If \( \text{diag}(x) \in K_k^d \), then
\[
\text{diag}(x) = \sum_{(i_1, \ldots, i_k) \in \binom{d}{k}} \varepsilon_{i_1 \cdots i_k}^d (M_{i_1 \cdots i_k}^k), \quad \text{and } M_{i_1 \cdots i_k}^k \in K^k.
\]
The embedding property and our assumption on \( K^1 \) implies that the diagonal of \( M_{i_1 \cdots i_k}^k \) are non-negative. Thus we have \( x \geq 0 \).
Conversely, if \( x \geq 0 \), we can write
\[
\text{diag}(x) = \sum_{i=1}^d \text{diag}(x_i e_i)
\]
where \( e_i \) is the \( i \)-th standard vector in \( \mathbb{F}^d \). Because of our assumption on \( K^1 \) and the \( \{K^k\}_{k=1}^\infty \) satisfies the embedding property thoroughly, each \( \text{diag}(x_i e_i) \in K_k^d \) and this is a valid decomposition in \( K_k^d \). Thus \( \text{diag}(x) \in K_k^d \). \hfill \Box

We note the assumption \( K^1 = \mathbb{R}_+ \) has no loss of generality since for nonempty one dimensional cone in \( \mathbb{S}^1 \), it is either \( \mathbb{R}_- \) or \( \mathbb{R}_+ \).

The following Theorem includes Lemma 3.2 as a special case. See item (i) and (vi) of the theorem. The same result can also be found in [BCPT05 Theorem 8.9] but we give a different proof.

Theorem 9.2. For a matrix \( A = [a_{ij}]_{ij} \), denote \( M(A) = [\alpha_{ij}] \) where \( \alpha_{ii} = a_{ii} \) for all \( i \) and \( \alpha_{ij} = -|a_{ij}| \) for all \( i \neq j \) and \( \rho(A) = \max\{\lambda : \lambda \text{ is an eigenvalue of } A\} \). The following are all equivalent when \( A \in \mathbb{S}^n \).

(i) \( A \in \text{SDD}^d \);
(ii) \( M(A) \in \text{SDD}^d \);
(iii) there exists \( D = \text{diag}(d), d > 0 \), i.e., elementwise positive, such that \( D^T AD = DD_d \);
(iv) there exists a permutation matrix \( P \) such that \( P^T AP \in \text{SDD}^d \);
(v) \( M(A) = sI - B \) for some \( s \) and \( B \) where \( B \) is a non-negative matrix and \( s \) is greater or equal to the largest absolute value of eigenvalue of \( B \), i.e., \( s \geq \rho(B) \);
(vi) \( M(A) \) is positive semi-definite.

Proof. We begin with the equivalence between (i)–(iv). It directly follows from the definition that (i) and (ii) are equivalent. By multiplying out \( D^T AD \) and examining row by row, one finds the condition \( D^T AD \in \mathbb{D}_d \) is the same as \( A \in \text{SDD}^d \). Thus (iii) is equivalent to (i). The equivalence between (i) and (iv) can also be easily verified from the definition.

Next we show that (v) and (vi) are equivalent. First (v) implies (vi) since for symmetric matrix, \( \|B\|_2 = \rho(B) \) and \( \max_{\|v\|_2 = 1} v^T B v = \rho(B) \), \( \min_{\|v\|_2 = 1} v^T M(A) v = \min_{\|v\|_2 = 1} s - \max_{\|v\|_2 = 1} v^T B v = s - \rho(B) \geq 0 \). Also, (vi) implies (v): If \( M(A) \) is positive semi-definite, then we know all its eigenvalues are non-negative and the largest eigenvalue is positive (the case \( M(A) \) is a zero matrix is trivially true for the implication). Denote the eigenvalue of \( M(A) \) to be \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0 \) (counting multiplicity), then \( \lambda_1 I - M(A) \in \mathbb{S}^d_+ \). Furthermore, since \( B = \lambda_1 I - M(A) \) is positive semi-definite, the diagonal element is non-negative and so \( B \) is non-negative. We also have \( \rho(B) = \lambda_1 - \lambda_n \leq \lambda_1 \). This shows (vi) implies (v).

Lastly we deduce that (v)–(vi) and (i)–(iv) are equivalent. Suppose \( A \in \text{SDD}^d \) and so is \( M(A) \), then by characterization (iii) and the fact that diagonally dominant matrix are positive semi-definite which follows from Gerschgorin circle theorem, we see \( M(A) \) is positive semi-definite. This shows
(i)–(iv) implies (v)–(vi). Conversely, suppose $M(A) = sI - B$ where $B$ is non-negative and $s \geq \rho(B)$. Since $B$ is symmetric, there always exists a permutation matrix $P$ such that

$$P^TBP = \begin{bmatrix} B_1 & & \\ & B_2 & \\ & & \ddots \\ & & & B_k \end{bmatrix},$$

and $B_i$ are all irreducible and square matrices and for all $i$, $\rho(B_i) \leq s$. Now by the the Perron–Frobenius theorem, we know for each $B_i$, there is an elementwise positive vector $v_i$ such that $B_i v_i = \rho(B_i) v_i$. Then if we multiply the vector $v = (v_1, \ldots, v_k)$ on the right to $P^T M(A) P$, we have $P^T M(A) P v = sv - (\rho(B_1) v_1, \ldots, \rho(B_k) v_k) \geq 0$. This shows that $P^T M(A) P \in \text{SDD}^d$ is and so are $M(A)$ and $A$. \hfill \Box

**Lemma 9.3.** Suppose $f$ is a real valued second order differentiable function defined on an open convex cone $K \subseteq \mathbb{R}^n$. If $f$ has positive definite Hessian, then $\nabla f$ is an injection.

**Proof.** For every $x \in K$ and $x + h \in K, h \neq 0$, we have

$$h^T (\nabla f(x + h) - \nabla f(x)) = \int_0^1 h^T \nabla^2 f(x + th)h dt > 0$$

as $\nabla^2 f$, the Hessian, is positive definite. This means $f(x + h) \neq f(x)$ and $f$ is injective. \hfill \Box

**References**

[AM17] Amir Ali Ahmadi and Anirudha Majumdar. DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization. arXiv preprint arXiv:1706.02586, 2017.

[BCPT05] Erik G Boman, Doron Chen, Ojas Parekh, and Sivan Toledo. On factor width and symmetric H-matrices. Linear algebra and its applications, 405:239–248, 2005.

[BGP18] Mina Saee Bostanabad, João Gouveia, and Ting Kei Pong. Inner approximating the completely positive cone via the cone of scaled diagonally dominant matrices. arXiv preprint arXiv:1807.00379, 2018.

[BL10] Jonathan Borwein and Adrian S Lewis. Convex analysis and nonlinear optimization: theory and examples. Springer Science & Business Media, 2010.

[BP94] Abraham Berman and Robert J Plemmons. Nonnegative matrices in the mathematical sciences, volume 9. Siam, 1994.

[Bur15] Samuel Burer. A gentle, geometric introduction to copositive optimization. Mathematical Programming, 151(1):89–116, 2015.

[BV04] Stephen Boyd and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.

[DK10] Etienne De Klerk. Exploiting special structure in semidefinite programming: A survey of theory and applications. European Journal of Operational Research, 201(1):1–10, 2010.

[Dü10] Mirjam Dür. Copositive programming–a survey. In Recent advances in optimization and its applications in engineering, pages 3–20. Springer, 2010.

[MM62] John E Maxfield and Henryk Minc. On the matrix equation $XX' = A$. Proceedings of the Edinburgh Mathematical Society, 13(2):125–129, 1962.

[NN94] Yurii Nesterov and Arkadi Nemirovskii. Interior-point polynomial algorithms in convex programming, volume 13. Siam, 1994.

[Par00] Pablo A Parrilo. Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. PhD thesis, California Institute of Technology, 2000.

[PP14] Frank Permenter and Pablo Parrilo. Partial facial reduction: simplified, equivalent sdps via approximations of the psl cone. Mathematical Programming, pages 1–54, 2014.

[Roc70] RT Rockafellar. Convex analysis. 1970.

[WKKM06] Hayato Waki, Sunyoung Kim, Masakazu Kojima, and Masakazu Muramatsu. Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. SIAM Journal on Optimization, 17(1):218–242, 2006.
DEPARTMENT OF STATISTICS, UNIVERSITY OF CHICAGO, CHICAGO, IL 60637-1514.

E-mail address: ld446@cornell.edu

COMPUTATIONAL AND APPLIED MATHEMATICS INITIATIVE, DEPARTMENT OF STATISTICS, UNIVERSITY OF CHICAGO, CHICAGO, IL 60637-1514.

E-mail address: lekheng@galton.uchicago.edu