Inclusion of spectrahedra, free spectrahedra and coin tossing

(joint work with Bill Helton, Igor Klep and Scott McCullough)

Markus Schweighofer

Universität Konstanz

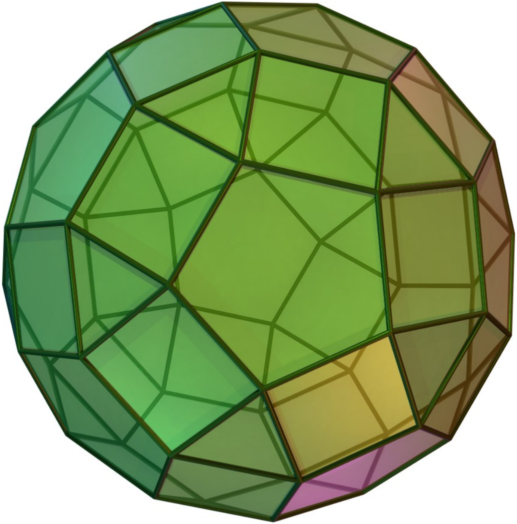
Monday Lecture

Graduiertenkolleg "Methods for Discrete Structures"

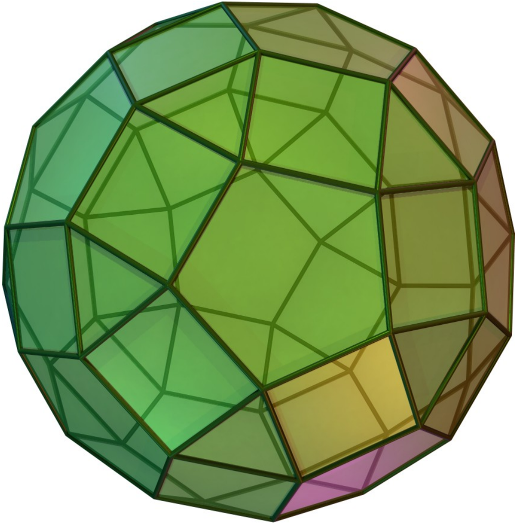
TU Berlin

July 13, 2015

A (closed convex) polyhedron

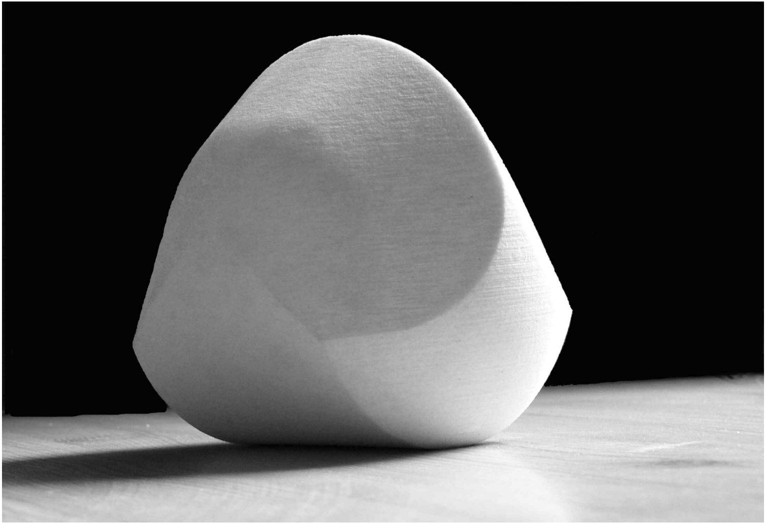


A (closed convex) polyhedron



...called rhombicosidodecahedron.

A spectrahedron



Spectrahedra

A pencil (of size d in n variables) is a monic linear symmetric real matrix polynomial

$$\begin{aligned} A &= I_d + A_1\mathtt{x}_1 + \dots A_n\mathtt{x}_n \\ &= \begin{pmatrix} 1 + a_{11}^{(1)}\mathtt{x}_1 + \cdots + a_{11}^{(n)}\mathtt{x}_n & a_{12}^{(1)}\mathtt{x}_1 + \cdots + a_{12}^{(n)}\mathtt{x}_n & \dots \\ a_{21}^{(1)}\mathtt{x}_1 + \cdots + a_{21}^{(n)}\mathtt{x}_n & 1 + a_{22}^{(1)}\mathtt{x}_1 + \cdots + a_{22}^{(n)}\mathtt{x}_n & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \\ &\in \mathbb{R}[\mathtt{x}_1, \dots, \mathtt{x}_n]^{d\times d} = \mathbb{R}[\mathtt{x}]^{d\times d} \end{aligned}$$

where $A_i = (a_{k\ell}^{(i)})_{1\leq k,\ell\leq d} \in S\mathbb{R}^{d\times d}$.

Spectrahedra

For $x \in \mathbb{R}^n$

$$
\begin{aligned}
A(x) &= I_d + x_1 A_1 + \dots x_n A_n \\
&= \begin{pmatrix} 1 + a_{11}^{(1)} x_1 + \cdots + a_{11}^{(n)} x_n & a_{12}^{(1)} x_1 + \cdots + a_{12}^{(n)} x_n & \dots \\ a_{21}^{(1)} x_1 + \cdots + a_{21}^{(n)} x_n & 1 + a_{22}^{(1)} x_1 + \cdots + a_{22}^{(n)} x_n & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \\
&\in \mathbb{R}^{d \times d}
\end{aligned}
$$

where $A_i = (a_{k\ell}^{(i)})_{1 \le k,\ell \le d} \in S\mathbb{R}^{d \times d}$.

Spectrahedra

A pencil (of size d in n variables) is a monic linear symmetric real matrix polynomial

$$\begin{aligned} A &= I_d + A_1\mathrm{x}_1 + \dots A_n\mathrm{x}_n \\ &= \begin{pmatrix} 1 + a_{11}^{(1)}\mathrm{x}_1 + \dots + a_{11}^{(n)}\mathrm{x}_n & a_{12}^{(1)}\mathrm{x}_1 + \dots + a_{12}^{(n)}\mathrm{x}_n & \dots \\ a_{21}^{(1)}\mathrm{x}_1 + \dots + a_{21}^{(n)}\mathrm{x}_n & 1 + a_{22}^{(1)}\mathrm{x}_1 + \dots + a_{22}^{(n)}\mathrm{x}_n & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \\ &\in \mathbb{R}[\mathrm{x}_1, \dots, \mathrm{x}_n]^{d\times d} = \mathbb{R}[\mathrm{x}]^{d\times d} \end{aligned}$$

where $A_i = (a_{k\ell}^{(i)})_{1\le k,\ell\le d} \in S\mathbb{R}^{d\times d}$.

$S_A(1) := \{x \in \mathbb{R}^n \mid A(x) \succeq 0\}$ is the spectrahedron defined by A.

Spectrahedra

A pencil (of size d in n variables) is a monic linear symmetric real matrix polynomial

$$\begin{aligned} A &= I_d + A_1\mathrm{x}_1 + \dots A_n\mathrm{x}_n \\ &= \begin{pmatrix} 1 + a_{11}^{(1)}\mathrm{x}_1 + \cdots + a_{11}^{(n)}\mathrm{x}_n & a_{12}^{(1)}\mathrm{x}_1 + \cdots + a_{12}^{(n)}\mathrm{x}_n & \dots \\ a_{21}^{(1)}\mathrm{x}_1 + \cdots + a_{21}^{(n)}\mathrm{x}_n & 1 + a_{22}^{(1)}\mathrm{x}_1 + \cdots + a_{22}^{(n)}\mathrm{x}_n & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \\ &\in \mathbb{R}[\mathrm{x}_1, \dots, \mathrm{x}_n]^{d \times d} = \mathbb{R}[\mathrm{x}]^{d \times d} \end{aligned}$$

where $A_i = (a_{k\ell}^{(i)})_{1 \le k,\ell \le d} \in S\mathbb{R}^{d \times d}$.

$S_A(1) := \{x \in \mathbb{R}^n \mid A(x) \succeq 0\}$ is the spectrahedron defined by A.

The $S_A(1)$ with A a pencil are exactly the spectrahedra with 0 in their interior.

Spectrahedra

A pencil (of size d in n variables) is a monic linear symmetric real matrix polynomial

$$\begin{aligned} A &= I_d + A_1\mathrm{x}_1 + \ldots A_n\mathrm{x}_n \\ &= \begin{pmatrix} 1 + a_{11}^{(1)}\mathrm{x}_1 + \cdots + a_{11}^{(n)}\mathrm{x}_n & a_{12}^{(1)}\mathrm{x}_1 + \cdots + a_{12}^{(n)}\mathrm{x}_n & \ldots \\ a_{21}^{(1)}\mathrm{x}_1 + \cdots + a_{21}^{(n)}\mathrm{x}_n & 1 + a_{22}^{(1)}\mathrm{x}_1 + \cdots + a_{22}^{(n)}\mathrm{x}_n & \ldots \\ \vdots & \vdots & \ddots \end{pmatrix} \\ &\in \mathbb{R}[\mathrm{x}_1, \ldots, \mathrm{x}_n]^{d\times d} = \mathbb{R}[\mathrm{x}]^{d\times d} \end{aligned}$$

where $A_i = (a_{k\ell}^{(i)})_{1\leq k,\ell\leq d} \in S\mathbb{R}^{d\times d}$.

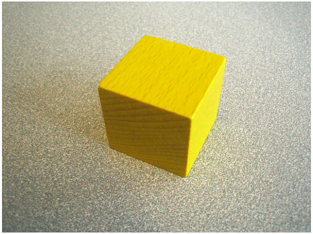
$S_A(1) := \{x \in \mathbb{R}^n \mid A(x) \succeq 0\}$ is the spectrahedron defined by A.

The $S_A(1)$ with A a diagonal pencil are exactly the polyhedra with 0 in their interior.

The cube

$$C_n := \begin{pmatrix} 1+\mathtt{x}_1 & & & & & & \\ & 1-\mathtt{x}_1 & & & & & \\ & & 1+\mathtt{x}_2 & & & & \\ & & & 1-\mathtt{x}_2 & & & \\ & & & & \ddots & & \\ & & & & & 1+\mathtt{x}_n & \\ & & & & & & 1-\mathtt{x}_n \end{pmatrix}$$

defines the cube $S_{C_n}(1) = [-1,1]^n$.



The disk

$$A := \begin{pmatrix} 1 + x_1 & x_2 \\ x_2 & 1 - x_1 \end{pmatrix} \qquad \text{and} \qquad B := \begin{pmatrix} 1 & x_1 & x_2 \\ x_1 & 1 & 0 \\ x_2 & 0 & 1 \end{pmatrix}$$

define both the disk

$$S_A(1) = \{x \in \mathbb{R}^2 \mid \|x\| \leq 1\} = S_B(1)$$

since $\det A = 1 - x_1^2 - x_2^2 = \det B$.

What is this talk (not) about?

It is about detecting inclusion (containment) of two spectrahedra whose interiors contain both 0 (or another known point).

What is this talk (not) about?

It is about detecting inclusion (containment) of two spectrahedra whose interiors contain both 0 (or another known point).

Mainly, it is about detecting inclusion of a cube in a spectrahedron.

What is this talk (not) about?

It is about detecting inclusion (containment) of two spectrahedra whose interiors contain both 0 (or another known point).

Mainly, it is about detecting inclusion of a cube in a spectrahedron.

It is not about testing emptiness or low-dimensionality of spectrahedra.

Certifying inclusion of spectrahedra

Observation. Let $A \in \mathbb{R}[\mathrm{x}]^{m\times m}$ and $B \in \mathbb{R}[\mathrm{x}]^{d\times d}$ be pencils. If there exist $P \in \mathbb{R}^{d\times d}$ and $Q_i \in \mathbb{R}^{m\times d}$ such that

$$(*) \qquad B = P^*P + \sum_i Q_i^* A Q_i,$$

then $S_A(1) \subseteq S_B(1)$.

Certifying inclusion of spectrahedra

Observation. Let $A \in \mathbb{R}[\mathrm{x}]^{m\times m}$ and $B \in \mathbb{R}[\mathrm{x}]^{d\times d}$ be pencils. If there exist $P \in \mathbb{R}^{d\times d}$ and $Q_i \in \mathbb{R}^{m\times d}$ such that

$$(*) \qquad B = P^*P + \sum_i Q_i^* A Q_i,$$

then $S_A(1) \subseteq S_B(1)$.

The search for certificates $(*)$ can be done with semidefinite programming and is therefore tractable.

Certifying inclusion of spectrahedra

Observation. Let $A \in \mathbb{R}[\mathrm{x}]^{m\times m}$ and $B \in \mathbb{R}[\mathrm{x}]^{d\times d}$ be pencils. If there exist $P \in \mathbb{R}^{d\times d}$ and $Q_i \in \mathbb{R}^{m\times d}$ such that

$$(*) \qquad B = P^*P + \sum_i Q_i^* A Q_i,$$

then $S_A(1) \subseteq S_B(1)$.

The search for certificates $(*)$ can be done with semidefinite programming and is therefore tractable.

Example. With $A := \begin{pmatrix} 1+\mathrm{x}_1 & \mathrm{x}_2 \\ \mathrm{x}_2 & 1-\mathrm{x}_1 \end{pmatrix}$ and $B := \begin{pmatrix} 1 & \mathrm{x}_1 & \mathrm{x}_2 \\ \mathrm{x}_1 & 1 & 0 \\ \mathrm{x}_2 & 0 & 1 \end{pmatrix}$ from above, we have

$$2B = \begin{pmatrix} 0 & 1 \\ 0 & -1 \\ 1 & 0 \end{pmatrix} A \begin{pmatrix} 0 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} A \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Certifying inclusion of spectrahedra

Observation. Let $A \in \mathbb{R}[\mathrm{x}]^{m\times m}$ and $B \in \mathbb{R}[\mathrm{x}]^{d\times d}$ be pencils. If there exist $P \in \mathbb{R}^{d\times d}$ and $Q_i \in \mathbb{R}^{m\times d}$ such that

$$(*) \qquad B = P^*P + \sum_i Q_i^* A Q_i,$$

then $S_A(1) \subseteq S_B(1)$. We will see that the converse fails in general.

The search for certificates $(*)$ can be done with semidefinite programming and is therefore tractable.

Example. With $A := \begin{pmatrix} 1+\mathrm{x}_1 & \mathrm{x}_2 \\ \mathrm{x}_2 & 1-\mathrm{x}_1 \end{pmatrix}$ and $B := \begin{pmatrix} 1 & \mathrm{x}_1 & \mathrm{x}_2 \\ \mathrm{x}_1 & 1 & 0 \\ \mathrm{x}_2 & 0 & 1 \end{pmatrix}$ from above, we have

$$2B = \begin{pmatrix} 0 & 1 \\ 0 & -1 \\ 1 & 0 \end{pmatrix} A \begin{pmatrix} 0 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} A \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Free spectrahedra

Consider again a pencil

$$
\begin{aligned}
A &= I_d + A_1 \mathtt{x}_1 + \ldots A_n \mathtt{x}_n \\
&= \begin{pmatrix} 1 + a_{11}^{(1)} \mathtt{x}_1 + \cdots + a_{11}^{(n)} \mathtt{x}_n & a_{12}^{(1)} \mathtt{x}_1 + \cdots + a_{12}^{(n)} \mathtt{x}_n & \ldots \\ a_{21}^{(1)} \mathtt{x}_1 + \cdots + a_{21}^{(n)} \mathtt{x}_n & 1 + a_{22}^{(1)} \mathtt{x}_1 + \cdots + a_{22}^{(n)} \mathtt{x}_n & \ldots \\ \vdots & \vdots & \ddots \end{pmatrix} \\
&\in \mathbb{R}[\mathtt{x}]^{d \times d}
\end{aligned}
$$

where $A_i = (a_{k\ell}^{(i)})_{1 \leq k, \ell \leq d} \in S\mathbb{R}^{d \times d}$.

Free spectrahedra

For $X \in (S\mathbb{R}^{m\times m})^n$

$$
\begin{aligned}
A(X) &= I_d \otimes I_m + A_1 \otimes X_1 + \dots A_n \otimes X_n \\
&= \begin{pmatrix} I_m + a_{11}^{(1)} X_1 + \cdots + a_{11}^{(n)} X_n & a_{12}^{(1)} X_1 + \cdots + a_{12}^{(n)} X_n & \cdots \\ a_{21}^{(1)} X_1 + \cdots + a_{21}^{(n)} X_n & I_m + a_{22}^{(1)} X_1 + \cdots + a_{22}^{(n)} X_n & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \\
&\in \mathbb{R}^{dm\times dm}
\end{aligned}
$$

where $A_i = (a_{k\ell}^{(i)})_{1\leq k,\ell\leq d} \in S\mathbb{R}^{d\times d}$.

Free spectrahedra

Consider again a pencil

$$
\begin{aligned}
A &= I_d + A_1\mathtt{x}_1 + \dots A_n\mathtt{x}_n \\
&= \begin{pmatrix} 1 + a_{11}^{(1)}\mathtt{x}_1 + \cdots + a_{11}^{(n)}\mathtt{x}_n & a_{12}^{(1)}\mathtt{x}_1 + \cdots + a_{12}^{(n)}\mathtt{x}_n & \dots \\ a_{21}^{(1)}\mathtt{x}_1 + \cdots + a_{21}^{(n)}\mathtt{x}_n & 1 + a_{22}^{(1)}\mathtt{x}_1 + \cdots + a_{22}^{(n)}\mathtt{x}_n & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \\
&\in \mathbb{R}[\mathtt{x}]^{d\times d}
\end{aligned}
$$

where $A_i = (a_{k\ell}^{(i)})_{1\le k,\ell\le d} \in S\mathbb{R}^{d\times d}$.

$S_A(m) := \{X \in (S\mathbb{R}^{m\times m})^n \mid A(X) \succeq 0\}$

$S_A := \bigcup_{m\in\mathbb{N}} S_A(m)$ is the free spectrahedron defined by A.

Free spectrahedra

Consider again a pencil

$$
\begin{aligned}
A &= I_d + A_1\mathtt{x}_1 + \dots A_n\mathtt{x}_n \\
&= \begin{pmatrix} 1 + a_{11}^{(1)}\mathtt{x}_1 + \cdots + a_{11}^{(n)}\mathtt{x}_n & a_{12}^{(1)}\mathtt{x}_1 + \cdots + a_{12}^{(n)}\mathtt{x}_n & \dots \\ a_{21}^{(1)}\mathtt{x}_1 + \cdots + a_{21}^{(n)}\mathtt{x}_n & 1 + a_{22}^{(1)}\mathtt{x}_1 + \cdots + a_{22}^{(n)}\mathtt{x}_n & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \\
&\in \mathbb{R}[\mathtt{x}]^{d\times d}
\end{aligned}
$$

where $A_i = (a_{k\ell}^{(i)})_{1\leq k,\ell\leq d} \in S\mathbb{R}^{d\times d}$.

$S_A(m) := \{X \in (S\mathbb{R}^{m\times m})^n \mid A(X) \succeq 0\}$

$S_A := \bigcup_{m\in\mathbb{N}} S_A(m)$ is the free spectrahedron defined by A.

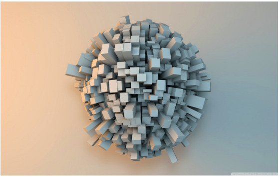
Condition $(*)$ certifies not only $S_A(1) \subseteq S_B(1)$ but even $S_A \subseteq S_B$.

The free cube

$$C_n = \begin{pmatrix} 1+\mathrm{x}_1 & & & & & & \\ & 1-\mathrm{x}_1 & & & & & \\ & & 1+\mathrm{x}_2 & & & & \\ & & & 1-\mathrm{x}_2 & & & \\ & & & & \ddots & & \\ & & & & & 1+\mathrm{x}_n & \\ & & & & & & 1-\mathrm{x}_n \end{pmatrix}$$

defines the free cube

$$\mathscr{C}_n := S_{C_n} = \bigcup_{m \in \mathbb{N}} \left\{ X \in (S\mathbb{R}^{m\times m})^n \mid \|X_i\| \leq 1 \right\}.$$



The free disk

With $A := \begin{pmatrix} 1+\mathtt{x}_1 & \mathtt{x}_2 \\ \mathtt{x}_2 & 1-\mathtt{x}_1 \end{pmatrix}$ and $B := \begin{pmatrix} 1 & \mathtt{x}_1 & \mathtt{x}_2 \\ \mathtt{x}_1 & 1 & 0 \\ \mathtt{x}_2 & 0 & 1 \end{pmatrix}$ from above,

$$S_B = \bigcup_{m\in\mathbb{N}} \{X \in (S\mathbb{R}^{m\times m})^2 \mid X_1^2 + X_2^2 \preceq I_m\}$$

is the free disk but $S_A \neq S_B$ since

$$\left(\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \frac{3}{4} \\ \frac{3}{4} & 0 \end{pmatrix}\right) \in S_B \setminus S_A.$$

Although we have $S_A(1) = S_B(1)$, we have $S_B \not\subseteq S_A$.

Certifying inclusion of free spectrahedra

Theorem (Helton, Klep, McCullough 2012).
Let $A \in \mathbb{R}[\mathrm{x}]^{m\times m}$ and $B \in \mathbb{R}[\mathrm{x}]^{d\times d}$ be pencils.
Then there exist $P \in \mathbb{R}^{d\times d}$ and $Q_i \in \mathbb{R}^{m\times d}$ such that

$$(*) \qquad B = P^*P + \sum_i Q_i^* A Q_i,$$

if and only if $S_A \subseteq S_B$.

Certifying inclusion of free spectrahedra

Theorem (Helton, Klep, McCullough 2012).
Let $A \in \mathbb{R}[\mathrm{x}]^{m\times m}$ and $B \in \mathbb{R}[\mathrm{x}]^{d\times d}$ be pencils.
Then there exist $P \in \mathbb{R}^{d\times d}$ and $Q_i \in \mathbb{R}^{m\times d}$ such that

$$(*) \qquad B = P^*P + \sum_i Q_i^* A Q_i,$$

if and only if $S_A \subseteq S_B$.

The proof uses Arveson's extension theorem and Stinespring's dilation theorem.

Certifying inclusion of free spectrahedra

Theorem (Helton, Klep, McCullough 2012).
Let $A \in \mathbb{R}[\mathrm{x}]^{m\times m}$ and $B \in \mathbb{R}[\mathrm{x}]^{d\times d}$ be pencils.
Then there exist $P \in \mathbb{R}^{d\times d}$ and $Q_i \in \mathbb{R}^{m\times d}$ such that

$$(*) \qquad B = P^*P + \sum_i Q_i^* A Q_i,$$

if and only if $S_A \subseteq S_B$.

The proof uses Arveson's extension theorem and Stinespring's dilation theorem.

Helton, Klep, McCullough: The matricial relaxation of a linear matrix inequality, Math. Program. 138 (2013), no. 1-2, Ser. A, 401–445 (was first but appeared later)
http://arxiv.org/abs/1003.0908.pdf

Helton, Klep, McCullough: The convex Positivstellensatz in a free algebra, Adv. Math. 231 (2012), no. 1, 516–534
http://arxiv.org/abs/1102.4859.pdf

Certifying inclusion of free spectrahedra

Theorem (Helton, Klep, McCullough 2012).
Let $A \in \mathbb{R}[\mathrm{x}]^{m\times m}$ and $B \in \mathbb{R}[\mathrm{x}]^{d\times d}$ be pencils.
Then there exist $P \in \mathbb{R}^{d\times d}$ and $Q_i \in \mathbb{R}^{m\times d}$ such that

$$(*) \qquad B = P^*P + \sum_i Q_i^* A Q_i,$$

if and only if $S_A \subseteq S_B$.

The proof uses Arveson's extension theorem and Stinespring's dilation theorem.

Kellner, Theobald, Trabandt: Containment problems for polytopes and spectrahedra, SIAM J. Optim. 23 (2013), no. 2, 1000–1020
`http://arxiv.org/abs/1204.4313`

Kellner, Theobald, Trabandt: A Semidefinite Hierarchy for Containment of Spectrahedra
`http://arxiv.org/abs/1308.5076`

Inclusion of free spectrahedra

Theorem. Let $A \in \mathbb{R}[\mathrm{x}]^{m\times m}$ and $B \in \mathbb{R}[\mathrm{x}]^{d\times d}$ be pencils with $S_A = -S_A$ and $S_A(1) \subseteq S_B(1)$. Then $S_A \subseteq dS_B$.

Inclusion of free spectrahedra

Theorem. Let $A \in \mathbb{R}[\mathrm{x}]^{m \times m}$ and $B \in \mathbb{R}[\mathrm{x}]^{d \times d}$ be pencils with $S_A = -S_A$ and $S_A(1) \subseteq S_B(1)$. Then $S_A \subseteq dS_B$.

Example. With $A := \begin{pmatrix} 1 + \mathrm{x}_1 & \mathrm{x}_2 \\ \mathrm{x}_2 & 1 - \mathrm{x}_1 \end{pmatrix}$ and $B := \begin{pmatrix} 1 & \mathrm{x}_1 & \mathrm{x}_2 \\ \mathrm{x}_1 & 1 & 0 \\ \mathrm{x}_2 & 0 & 1 \end{pmatrix}$ from above,

$$S_B \subseteq S_A \subseteq 3S_B.$$

The matrix cube problem

Theorem (Ben Tal, Nemirovski 2002). For $d \in \mathbb{N}$, define $\vartheta(d) \in [1, \infty)$ by

$$\frac{1}{\vartheta(d)} = \min_{\substack{a \in \mathbb{R}^d \\ |a_1| + \cdots + |a_d| = d}} \int_{S^{d-1}} \left| \sum_{i=1}^{d} a_i \xi_i^2 \right| d\xi.$$

Then $\vartheta(1) = 1$, $\vartheta(2) = \frac{\pi}{2}$,
$\vartheta(d) \leq \frac{\pi}{2}\sqrt{d} \leq \sqrt{3d}$ ($\leq \sqrt{d^2} = d$ for $d \geq 3$) and if $A = I + A_1 \mathrm{x}_1 + \cdots + A_n \mathrm{x}_n$ is a pencil with real matrices A_i of rank at most d such that $[-1, 1]^n \subseteq S_A(1)$, then

$$\mathscr{C}_n \subseteq \vartheta(d) S_A.$$

The matrix cube problem

Theorem (Ben Tal, Nemirovski 2002). For $d \in \mathbb{N}$, define $\vartheta(d) \in [1, \infty)$ by

$$\frac{1}{\vartheta(d)} = \min_{\substack{a \in \mathbb{R}^d \\ |a_1| + \cdots + |a_d| = d}} \int_{S^{d-1}} \left| \sum_{i=1}^{d} a_i \xi_i^2 \right| d\xi.$$

Then $\vartheta(1) = 1$, $\vartheta(2) = \frac{\pi}{2}$,
$\vartheta(d) \leq \frac{\pi}{2}\sqrt{d} \leq \sqrt{3d}$ ($\leq \sqrt{d^2} = d$ for $d \geq 3$) and if $A = I + A_1 \mathrm{x}_1 + \cdots + A_n \mathrm{x}_n$ is a pencil with real matrices A_i of rank at most d such that $[-1, 1]^n \subseteq S_A(1)$, then

$$\mathscr{C}_n \subseteq \vartheta(d) S_A.$$

Ben-Tal, Nemirovski: On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty, SIAM J. Optim. 12 (2002), no. 3, 811–833

The matrix cube problem

Theorem (Ben Tal, Nemirovski 2002). For $d \in \mathbb{N}$, define $\vartheta(d) \in [1, \infty)$ by

$$\frac{1}{\vartheta(d)} = \min_{\substack{a \in \mathbb{R}^d \\ |a_1| + \cdots + |a_d| = d}} \int_{S^{d-1}} \left| \sum_{i=1}^{d} a_i \xi_i^2 \right| d\xi.$$

Then $\vartheta(1) = 1$, $\vartheta(2) = \frac{\pi}{2}$,
$\vartheta(d) \leq \frac{\pi}{2}\sqrt{d} \leq \sqrt{3d}$ ($\leq \sqrt{d^2} = d$ for $d \geq 3$) and if $A = I + A_1 \mathrm{x}_1 + \cdots + A_n \mathrm{x}_n$ is a pencil with real matrices A_i of rank at most d such that $[-1, 1]^n \subseteq S_A(1)$, then

$$\mathscr{C}_n \subseteq \vartheta(d) S_A.$$

Our contributions to this theorem:

- The theorem follows naturally from a new dilation theorem.
- Analytic expression for $\vartheta(d)$ for even d and implicit characterization of $\vartheta(d)$ for odd d.
- The scaling factor $\vartheta(d)$ is sharp.

Dilation theorem

We give here only a version of our dilation theorem from which the preceding theorem can be deduced in the case where each A_i is of size d (instead of rank at most d):

Dilation theorem

We give here only a version of our dilation theorem from which the preceding theorem can be deduced in the case where each A_i is of size d (instead of rank at most d):

Theorem. Let $d \in \mathbb{N}$. There is a Hilbert space H, an isometry $V: \mathbb{R}^d \to H$ and a set $\mathscr{T}$ of commuting self-adjoint contractions on H such that for each $X \in \mathscr{C}_n(d)$ there exists a $T \in \mathscr{T}$ with $X = \vartheta(d) V^* T V$.

Dilation theorem

We give here only a version of our dilation theorem from which the preceding theorem can be deduced in the case where each A_i is of size d (instead of rank at most d):

Theorem. Let $d \in \mathbb{N}$. There is a Hilbert space H, an isometry $V\colon \mathbb{R}^d \to H$ and a set $\mathscr{T}$ of commuting self-adjoint contractions on H such that for each $X \in \mathscr{C}_n(d)$ there exists a $T \in \mathscr{T}$ with $X = \vartheta(d) V^* T V$.

In the Ben-Tal & Nemirovski theorem, let A be of size d.

Dilation theorem

We give here only a version of our dilation theorem from which the preceding theorem can be deduced in the case where each A_i is of size d (instead of rank at most d):

Theorem. Let $d \in \mathbb{N}$. There is a Hilbert space H, an isometry $V\colon \mathbb{R}^d \to H$ and a set $\mathscr{T}$ of commuting self-adjoint contractions on H such that for each $X \in \mathscr{C}_n(d)$ there exists a $T \in \mathscr{T}$ with $X = \vartheta(d)V^*TV$.

In the Ben-Tal & Nemirovski theorem, let A be of size d. It was already known that to show $\mathscr{C}_n \subseteq \vartheta(d)S_A$ it suffices to prove $\mathscr{C}_n(d) \subseteq \vartheta(d)S_A(d)$.

Dilation theorem

We give here only a version of our dilation theorem from which the preceding theorem can be deduced in the case where each A_i is of size d (instead of rank at most d):

Theorem. Let $d \in \mathbb{N}$. There is a Hilbert space H, an isometry $V : \mathbb{R}^d \to H$ and a set $\mathscr{T}$ of commuting self-adjoint contractions on H such that for each $X \in \mathscr{C}_n(d)$ there exists a $T \in \mathscr{T}$ with $X = \vartheta(d) V^* T V$.

In the Ben-Tal & Nemirovski theorem, let A be of size d. It was already known that to show $\mathscr{C}_n \subseteq \vartheta(d) S_A$ it suffices to prove $\mathscr{C}_n(d) \subseteq \vartheta(d) S_A(d)$. With the above theorem, this reduces to $(V^* T_1 V, \dots, V^* T_n V) \in S_A(d)$ for all $T_1, \dots, T_n \in \mathscr{T}$.

Dilation theorem

We give here only a version of our dilation theorem from which the preceding theorem can be deduced in the case where each A_i is of size d (instead of rank at most d):

Theorem. Let $d \in \mathbb{N}$. There is a Hilbert space H, an isometry $V : \mathbb{R}^d \to H$ and a set $\mathscr{T}$ of commuting self-adjoint contractions on H such that for each $X \in \mathscr{C}_n(d)$ there exists a $T \in \mathscr{T}$ with $X = \vartheta(d) V^* T V$.

In the Ben-Tal & Nemirovski theorem, let A be of size d. It was already known that to show $\mathscr{C}_n \subseteq \vartheta(d) S_A$ it suffices to prove $\mathscr{C}_n(d) \subseteq \vartheta(d) S_A(d)$. With the above theorem, this reduces to $(V^* T_1 V, \ldots, V^* T_n V) \in S_A(d)$ for all $T_1, \ldots, T_n \in \mathscr{T}$. Assume H is finite-dimensional (which it is not but this can be repaired with the spectral theorem),

Dilation theorem

We give here only a version of our dilation theorem from which the preceding theorem can be deduced in the case where each A_i is of size d (instead of rank at most d):

Theorem. Let $d \in \mathbb{N}$. There is a Hilbert space H, an isometry $V : \mathbb{R}^d \to H$ and a set $\mathscr{T}$ of commuting self-adjoint contractions on H such that for each $X \in \mathscr{C}_n(d)$ there exists a $T \in \mathscr{T}$ with $X = \vartheta(d) V^* T V$.

In the Ben-Tal & Nemirovski theorem, let A be of size d. It was already known that to show $\mathscr{C}_n \subseteq \vartheta(d) S_A$ it suffices to prove $\mathscr{C}_n(d) \subseteq \vartheta(d) S_A(d)$. With the above theorem, this reduces to $(V^* T_1 V, \dots, V^* T_n V) \in S_A(d)$ for all $T_1, \dots, T_n \in \mathscr{T}$. Assume H is finite-dimensional (which it is not but this can be repaired with the spectral theorem), then WLOG $H = \mathbb{R}^m$ for some $m \in \mathbb{N}$

Dilation theorem

We give here only a version of our dilation theorem from which the preceding theorem can be deduced in the case where each A_i is of size d (instead of rank at most d):

Theorem. Let $d \in \mathbb{N}$. There is a Hilbert space H, an isometry $V\colon \mathbb{R}^d \to H$ and a set $\mathscr{T}$ of commuting self-adjoint contractions on H such that for each $X \in \mathscr{C}_n(d)$ there exists a $T \in \mathscr{T}$ with $X = \vartheta(d)V^*TV$.

In the Ben-Tal & Nemirovski theorem, let A be of size d. It was already known that to show $\mathscr{C}_n \subseteq \vartheta(d)S_A$ it suffices to prove $\mathscr{C}_n(d) \subseteq \vartheta(d)S_A(d)$. With the above theorem, this reduces to $(V^*T_1V,\dots,V^*T_nV) \in S_A(d)$ for all $T_1,\dots,T_n \in \mathscr{T}$. Assume H is finite-dimensional (which it is not but this can be repaired with the spectral theorem), then WLOG $H = \mathbb{R}^m$ for some $m \in \mathbb{N}$ and, since the T_i are commuting self-adjoints, WLOG $T_i \in \mathbb{R}^{m\times m}$ diagonal.

Dilation theorem

We give here only a version of our dilation theorem from which the preceding theorem can be deduced in the case where each A_i is of size d (instead of rank at most d):

Theorem. Let $d \in \mathbb{N}$. There is a Hilbert space H, an isometry $V\colon \mathbb{R}^d \to H$ and a set $\mathscr{T}$ of commuting self-adjoint contractions on H such that for each $X \in \mathscr{C}_n(d)$ there exists a $T \in \mathscr{T}$ with $X = \vartheta(d)V^*TV$.

In the Ben-Tal & Nemirovski theorem, let A be of size d. It was already known that to show $\mathscr{C}_n \subseteq \vartheta(d)S_A$ it suffices to prove $\mathscr{C}_n(d) \subseteq \vartheta(d)S_A(d)$. With the above theorem, this reduces to $(V^*T_1V, \dots, V^*T_nV) \in S_A(d)$ for all $T_1, \dots, T_n \in \mathscr{T}$. Assume H is finite-dimensional (which it is not but this can be repaired with the spectral theorem), then WLOG $H = \mathbb{R}^m$ for some $m \in \mathbb{N}$ and, since the T_i are commuting self-adjoints, WLOG $T_i \in \mathbb{R}^{m\times m}$ diagonal. Since the T_i are contractions and $[-1,1]^n \subseteq S_A(1)$, $A(T) \succeq 0$. Hence $A(V^*TV) = V^*A(T)V \succeq 0$.

Dilation theorem

We give here only a version of our dilation theorem from which the preceding theorem can be deduced in the case where each A_i is of size d (instead of rank at most d):

Theorem. Let $d \in \mathbb{N}$. There is a Hilbert space H, an isometry $V\colon \mathbb{R}^d \to H$ and a set $\mathscr{T}$ of commuting self-adjoint contractions on H such that for each $X \in \mathscr{C}_n(d)$ there exists a $T \in \mathscr{T}$ with $X = \vartheta(d)V^*TV$.

Proof idea: $H := L^2(O(d), \mathbb{R}^n)$, $V\colon \mathbb{R}^d \to H$, $v \mapsto (U \mapsto v)$, $V^*\colon H \to \mathbb{R}^d$, $f \mapsto \int_{O(d)} f(U)\, dU$, $\mathscr{T}$ consists of all operators $T_D\colon H \to H$, $f \mapsto (U \mapsto UD(U)U^* f(U))$ where $D\colon O(d) \to \mathbb{R}^{d \times d}$ is any measurable function taking diagonal contractive values.

Dilation theorem

We give here only a version of our dilation theorem from which the preceding theorem can be deduced in the case where each A_i is of size d (instead of rank at most d):

Theorem. Let $d \in \mathbb{N}$. There is a Hilbert space H, an isometry $V\colon \mathbb{R}^d \to H$ and a set $\mathscr{T}$ of commuting self-adjoint contractions on H such that for each $X \in \mathscr{C}_n(d)$ there exists a $T \in \mathscr{T}$ with $X = \vartheta(d)V^*TV$.

Proof idea: $H := L^2(O(d), \mathbb{R}^n)$, $V\colon \mathbb{R}^d \to H$, $v \mapsto (U \mapsto v)$, $V^*\colon H \to \mathbb{R}^d$, $f \mapsto \int_{O(d)} f(U)\, dU$, $\mathscr{T}$ consists of all operators $T_D\colon H \to H$, $f \mapsto (U \mapsto UD(U)U^*f(U))$ where $D\colon O(d) \to \mathbb{R}^{d\times d}$ is any measurable function taking diagonal contractive values. Note $V^*T_DV = \int_{O(d)} UD(U)U^*\, dU$.

Dilation theorem

We give here only a version of our dilation theorem from which the preceding theorem can be deduced in the case where each A_i is of size d (instead of rank at most d):

Theorem. Let $d \in \mathbb{N}$. There is a Hilbert space H, an isometry $V\colon \mathbb{R}^d \to H$ and a set $\mathscr{T}$ of commuting self-adjoint contractions on H such that for each $X \in \mathscr{C}_n(d)$ there exists a $T \in \mathscr{T}$ with $X = \vartheta(d) V^* T V$.

Proof idea: $H := L^2(O(d), \mathbb{R}^n)$, $V\colon \mathbb{R}^d \to H,\ v \mapsto (U \mapsto v)$, $V^*\colon H \to \mathbb{R}^d,\ f \mapsto \int_{O(d)} f(U)\, dU$, $\mathscr{T}$ consists of all operators $T_D\colon H \to H,\ f \mapsto (U \mapsto UD(U)U^* f(U))$ where $D\colon O(d) \to \mathbb{R}^{d\times d}$ is any measurable function taking diagonal contractive values. Note $V^* T_D V = \int_{O(d)} UD(U)U^* dU$. Since $\mathscr{T}$ is convex, consider an extreme point X of the compact convex set $\mathscr{C}_n(d)$.

Dilation theorem

We give here only a version of our dilation theorem from which the preceding theorem can be deduced in the case where each A_i is of size d (instead of rank at most d):

Theorem. Let $d \in \mathbb{N}$. There is a Hilbert space H, an isometry $V\colon \mathbb{R}^d \to H$ and a set $\mathscr{T}$ of commuting self-adjoint contractions on H such that for each $X \in \mathscr{C}_n(d)$ there exists a $T \in \mathscr{T}$ with $X = \vartheta(d)V^*TV$.

Proof idea: $H := L^2(O(d), \mathbb{R}^n)$, $V\colon \mathbb{R}^d \to H,\ v \mapsto (U \mapsto v)$, $V^*\colon H \to \mathbb{R}^d,\ f \mapsto \int_{O(d)} f(U)\, dU$, $\mathscr{T}$ consists of all operators $T_D\colon H \to H,\ f \mapsto (U \mapsto UD(U)U^*f(U))$ where $D\colon O(d) \to \mathbb{R}^{d\times d}$ is any measurable function taking diagonal contractive values. Note $V^*T_DV = \int_{O(d)} UD(U)U^*dU$. Since $\mathscr{T}$ is convex, consider an extreme point X of the compact convex set $\mathscr{C}_n(d)$.
Take $D\colon O(d) \to \mathbb{R}^{d\times d},\ U \mapsto \sum_{i=1}^d \operatorname{sgn}(e_i^*U^*(\lambda + \mu X)Ue_i)e_ie_i^*$ for certain carefully chosen $\lambda, \mu \in \mathbb{R}$. Then $X = \vartheta(d)V^*T_DV$.

Better bounds for $\vartheta(d)$

We considerably improve the upper bound on $\vartheta(d)$ given by Ben Tal and Nemirovski and prove also a lower bound.

Theorem. Let $d \in \mathbb{N}$. If d is even, then

$$\frac{\sqrt{\pi}}{2}\sqrt{d+1} \leq \vartheta(d) \leq \frac{\sqrt{\pi}}{2} \cdot \frac{d}{\sqrt{d-1}}.$$

If $d \neq 1$ is odd, then

$$\sqrt[4]{\left(1-\frac{1}{d+1}\right)^{d+1}\left(1+\frac{1}{d-1}\right)^{d-1}} \cdot \frac{\sqrt{\pi}}{2}\sqrt{d+\frac{3}{2}} \leq \vartheta(d) \leq \frac{\sqrt{\pi}}{2} \cdot \frac{d+2}{\sqrt{d+\frac{5}{2}}}.$$

We have $\displaystyle\lim_{d\to\infty} \frac{\vartheta(d)}{\sqrt{d}} = \frac{\sqrt{\pi}}{2}$.

Computing $\vartheta(d)$

Reminder. For $a > 0$: $\Gamma(x) = \int_0^x t^{a-1} e^{-t} dt$ ("gamma function")

Computing $\vartheta(d)$

Reminder. For $a > 0$: $\Gamma(x) = \int_0^x t^{a-1} e^{-t} dt$ ("gamma function")

For $a, b > 0$ and $0 \leq x \leq 1$:

$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$ ("beta function")

Computing $\vartheta(d)$

Reminder. For $a > 0$: $\Gamma(x) = \int_0^x t^{a-1} e^{-t} dt$ ("gamma function")

For $a, b > 0$ and $0 \leq x \leq 1$:

$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$ ("beta function")

$B_x(a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt$ ("incomplete beta function")

Computing $\vartheta(d)$

Reminder. For $a > 0$: $\Gamma(x) = \int_0^x t^{a-1} e^{-t} dt$ ("gamma function")

For $a, b > 0$ and $0 \leq x \leq 1$:

$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$ ("beta function")

$B_x(a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt$ ("incomplete beta function")

$I_x(a, b) = \frac{B_x(a,b)}{B(a,b)}$ ("regularized incomplete beta function")

Computing $\vartheta(d)$

Reminder. For $a > 0$: $\Gamma(x) = \int_0^x t^{a-1} e^{-t} dt$ ("gamma function")

For $a, b > 0$ and $0 \leq x \leq 1$:

$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$ ("beta function")

$B_x(a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt$ ("incomplete beta function")

$I_x(a, b) = \frac{B_x(a,b)}{B(a,b)}$ ("regularized incomplete beta function")

Theorem. Let $d \in \mathbb{N}$. If d is even, then $\vartheta(d) = \sqrt{\pi}\frac{\Gamma\left(1+\frac{d}{4}\right)}{\Gamma\left(\frac{1}{2}+\frac{d}{4}\right)}$.

Computing $\vartheta(d)$

Reminder. For $a > 0$: $\Gamma(x) = \int_0^x t^{a-1} e^{-t} dt$ ("gamma function")

For $a, b > 0$ and $0 \leq x \leq 1$:

$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$ ("beta function")

$B_x(a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt$ ("incomplete beta function")

$I_x(a, b) = \frac{B_x(a,b)}{B(a,b)}$ ("regularized incomplete beta function")

Theorem. Let $d \in \mathbb{N}$. If d is even, then $\vartheta(d) = \sqrt{\pi}\frac{\Gamma\left(1+\frac{d}{4}\right)}{\Gamma\left(\frac{1}{2}+\frac{d}{4}\right)}$.

Suppose $d \geq 3$ is odd. Then there is a unique $p \in [0, 1]$ satisfying $I_p\left(\frac{d+1}{4}, \frac{d+3}{4}\right) = I_{1-p}\left(\frac{d-1}{4}, \frac{d+5}{4}\right)$.

Computing $\vartheta(d)$

Reminder. For $a > 0$: $\Gamma(x) = \int_0^x t^{a-1} e^{-t} dt$ ("gamma function")

For $a, b > 0$ and $0 \leq x \leq 1$:

$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$ ("beta function")

$B_x(a,b) = \int_0^x t^{a-1}(1-t)^{b-1} dt$ ("incomplete beta function")

$I_x(a,b) = \frac{B_x(a,b)}{B(a,b)}$ ("regularized incomplete beta function")

Theorem. Let $d \in \mathbb{N}$. If d is even, then $\vartheta(d) = \sqrt{\pi}\frac{\Gamma\left(1+\frac{d}{4}\right)}{\Gamma\left(\frac{1}{2}+\frac{d}{4}\right)}$.

Suppose $d \geq 3$ is odd. Then there is a unique $p \in [0,1]$ satisfying $I_p\left(\frac{d+1}{4}, \frac{d+3}{4}\right) = I_{1-p}\left(\frac{d-1}{4}, \frac{d+5}{4}\right)$. For this p, we have $p \in [\frac{1}{2}, \frac{d+1}{2d}]$,

$$\vartheta_-(d) \leq \vartheta(d) = \frac{\Gamma\left(\frac{d+3}{4}\right)\Gamma\left(\frac{d+5}{4}\right)}{p^{\frac{d-1}{4}}(1-p)^{\frac{d+1}{4}}\Gamma\left(\frac{d}{2}+1\right)} \leq \min\{\vartheta_+(d), \vartheta_{++}(d)\}$$

where $\vartheta_-(d)$, $\vartheta_+(d)$ and $\vartheta_{++}(d)$ are given by

$\vartheta_-(d) = \sqrt[4]{\frac{d^{2d}}{(d+1)^{d+1}(d-1)^{d-1}}}\ \vartheta_{++}(d)$,

$\frac{1}{\vartheta_+(d)} = \frac{d-1}{d} I_{\frac{d+1}{2d}}\left(\frac{d+1}{4}, \frac{d+3}{4}\right) + \frac{d+1}{d} I_{\frac{d-1}{2d}}\left(\frac{d-1}{4}, \frac{d+5}{4}\right) - 1$ and

$\vartheta_{++}(d) = \sqrt{\frac{\pi}{2}}\frac{\Gamma\left(\frac{d+3}{2}\right)}{\Gamma\left(\frac{d}{2}+1\right)}$.

d	$\vartheta_{-}(d)$	$\vartheta(d)$	$\vartheta_{+}(d)$	$\vartheta_{++}(d)$
1	–	1	–	–
2	–	1.5708	–	–
3	1.73205	1.73482	1.77064	1.88562
4	–	2	–	–
5	2.15166	2.1527	2.17266	2.26274
6	–	2.35619	–	–
7	2.49496	2.49548	2.50851	2.58599
8	–	2.66667	–	–
9	2.79445	2.79475	2.80409	2.87332
10	–	2.94524	–	–
11	3.064	3.06419	3.07131	3.13453
12	–	3.2	–	–
13	3.31129	3.31142	3.31707	3.37565
14	–	3.43612	–	–
15	3.54114	3.54123	3.54585	3.6007
16	–	3.65714	–	–
17	3.75681	3.75688	3.76076	3.8125
18	–	3.86563	–	–

Computing $\vartheta(d)$

Let $d \in \mathbb{N}$ with $d \geq 2$. We have simplified the formula of Ben Tal and Nemirovski

$$\frac{1}{\vartheta(d)} = \min_{\substack{a \in \mathbb{R}^d \\ |a_1| + \cdots + |a_d| = d}} \int_{S^{d-1}} \left| \sum_{i=1}^{d} a_i \xi_i^2 \right| d\xi$$

Computing $\vartheta(d)$

Let $d \in \mathbb{N}$ with $d \geq 2$. We have simplified the formula of Ben Tal and Nemirovski

$$\frac{1}{\vartheta(d)} = \min_{\substack{a \in \mathbb{R}^d \\ |a_1|+\cdots+|a_d|=d}} \int_{S^{d-1}} \left| \sum_{i=1}^{d} a_i \xi_i^2 \right| d\xi$$

to

$$\frac{1}{\vartheta(d)} = \min_{\substack{s,t \in \mathbb{N} \\ s+t=d}} \min_{\substack{a,b \in \mathbb{R}_{\geq 0} \\ as+bt=d}} \int_{S^{d-1}} \left| a \sum_{i=1}^{s} \xi_i^2 - b \sum_{i=s+1}^{d} \xi_i^2 \right| d\xi.$$

Computing $\vartheta(d)$

Let $d \in \mathbb{N}$ with $d \geq 2$. We have simplified the formula of Ben Tal and Nemirovski

$$\frac{1}{\vartheta(d)} = \min_{\substack{a \in \mathbb{R}^d \\ |a_1| + \cdots + |a_d| = d}} \int_{S^{d-1}} \left| \sum_{i=1}^{d} a_i \xi_i^2 \right| d\xi$$

to

$$\frac{1}{\vartheta(d)} = \min_{\substack{s,t \in \mathbb{N} \\ s+t=d}} \min_{\substack{a,b \in \mathbb{R}_{\geq 0} \\ as+bt=d}} \int_{S^{d-1}} \left| a \sum_{i=1}^{s} \xi_i^2 - b \sum_{i=s+1}^{d} \xi_i^2 \right| d\xi.$$

We manage to compute the integral and reparameterize it to get

$$\frac{1}{\vartheta(d)} = \min_{\substack{s,t \in \mathbb{N} \\ s+t=d}} \min_{p \in [0,1]} \left(\frac{2(1-p)sI_{1-p}\left(\frac{t}{2}, 1+\frac{s}{2}\right) + 2ptI_p\left(\frac{s}{2}, 1+\frac{t}{2}\right)}{(1-p)s + pt} - 1 \right)$$

Computing $\vartheta(d)$

Let $d \in \mathbb{N}$ with $d \geq 2$. We have simplified the formula of Ben Tal and Nemirovski

$$\frac{1}{\vartheta(d)} = \min_{\substack{a \in \mathbb{R}^d \\ |a_1|+\cdots+|a_d|=d}} \int_{S^{d-1}} \left| \sum_{i=1}^{d} a_i \xi_i^2 \right| d\xi$$

to

$$\frac{1}{\vartheta(d)} = \min_{\substack{s,t \in \mathbb{N} \\ s+t=d}} \min_{\substack{a,b \in \mathbb{R}_{\geq 0} \\ as+bt=d}} \int_{S^{d-1}} \left| a \sum_{i=1}^{s} \xi_i^2 - b \sum_{i=s+1}^{d} \xi_i^2 \right| d\xi.$$

We manage to compute the integral and reparameterize it to get

$$\frac{1}{\vartheta(d)} = \min_{\substack{s,t \in \mathbb{N} \\ s+t=d}} \min_{p \in [0,1]} \left(\frac{2(1-p)sI_{1-p}\left(\frac{t}{2}, 1+\frac{s}{2}\right) + 2ptI_p\left(\frac{s}{2}, 1+\frac{t}{2}\right)}{(1-p)s+pt} - 1 \right)$$

and we prove that the inner minimum is assumed at the unique $p_{s,t} \in (0,1)$ satisfying

$$I_{p_{s,t}}\left(\frac{s}{2}, 1+\frac{t}{2}\right) = I_{1-p_{s,t}}\left(\frac{t}{2}, 1+\frac{s}{2}\right).$$

Quiz

Let $s, t \in \mathbb{N}$ such that $s \geq t$ and set $d := s + t$.

Suppose you toss a biased coin d times with probability for heads $\frac{s}{d}$ and probability for tails $\frac{t}{d}$.

What is more likely –
that you observe at least s heads or at least t tails?

Quiz

Let $s, t \in \mathbb{N}$ such that $s \geq t$ and set $d := s + t$.

Suppose you toss a biased coin d times with probability for heads $\frac{s}{d}$ and probability for tails $\frac{t}{d}$.

What is more likely –
that you observe at least s heads or at least t tails?

The expected number of heads is s but that seems only loosely related.

Quiz

Let $s, t \in \mathbb{N}$ such that $s \geq t$ and set $d := s + t$.

Suppose you toss a biased coin d times with probability for heads $\frac{s}{d}$ and probability for tails $\frac{t}{d}$.

What is more likely –
that you observe at least s heads or at least t tails?

The expected number of heads is s but that seems only loosely related.

The median number of heads could perhaps help. But it can be shown to be s also. That does not help!

Quiz

Let $s, t \in \mathbb{N}$ such that $s \geq t$ and set $d := s + t$.

Suppose you toss a biased coin d times with probability for heads $\frac{s}{d}$ and probability for tails $\frac{t}{d}$.

What is more likely –
that you observe at least s heads or at least t tails?

The expected number of heads is s but that seems only loosely related.

The median number of heads could perhaps help. But it can be shown to be s also. That does not help!

A theorem of Simmons from 1895 says:
s heads or more is as least as probable than t tails or more.

Quiz

Let $s, t \in \mathbb{N}$ such that $s \geq t$ and set $d := s + t$.

Suppose you toss a biased coin d times with probability for heads $\frac{s}{d}$ and probability for tails $\frac{t}{d}$.

What is more likely –
that you observe at least s heads or at least t tails?

The expected number of heads is s but that seems only loosely related.

The median number of heads could perhaps help. But it can be shown to be s also. That does not help!

A theorem of Simmons from 1895 says:
s heads or more is as least as probable than t tails or more.

A paper by Perrin and Redside from 2007 says something even more subtle: The difference grows when $s \notin \{0, d\}$ grows.

Computing $\vartheta(d)$

Let $d \in \mathbb{N}$ with $d \geq 2$. Breaking the symmetry in s and t,

$$\frac{1}{\vartheta(d)} = \min_{\substack{s,t\in\mathbb{N} \\ s+t=d \\ s \geq t}} \min_{p\in[0,1]} \left(\frac{2(1-p)sI_{1-p}\left(\frac{t}{2}, 1+\frac{s}{2}\right) + 2ptI_p\left(\frac{s}{2}, 1+\frac{t}{2}\right)}{(1-p)s+pt} - 1 \right)$$

Computing $\vartheta(d)$

Let $d \in \mathbb{N}$ with $d \geq 2$. Breaking the symmetry in s and t,

$$\frac{1}{\vartheta(d)} = \min_{\substack{s,t \in \mathbb{N} \\ s+t=d \\ s \geq t}} \min_{p \in [0,1]} \left(\frac{2(1-p)sI_{1-p}\left(\frac{t}{2}, 1+\frac{s}{2}\right) + 2ptI_p\left(\frac{s}{2}, 1+\frac{t}{2}\right)}{(1-p)s + pt} - 1 \right)$$

where the inner minimum is assumed at $p_{s,t} \in (0,1)$ defined by $I_{p_{s,t}}\left(\frac{s}{2}, 1+\frac{t}{2}\right) = I_{1-p_{s,t}}\left(\frac{t}{2}, 1+\frac{s}{2}\right)$.

Computing $\vartheta(d)$

Let $d \in \mathbb{N}$ with $d \geq 2$. Breaking the symmetry in s and t,

$$\frac{1}{\vartheta(d)} = \min_{\substack{s,t \in \mathbb{N} \\ s+t=d \\ s \geq t}} \min_{p \in [0,1]} \left(\frac{2(1-p)sI_{1-p}\left(\frac{t}{2}, 1+\frac{s}{2}\right) + 2ptI_p\left(\frac{s}{2}, 1+\frac{t}{2}\right)}{(1-p)s+pt} - 1 \right)$$

where the inner minimum is assumed at $p_{s,t} \in (0,1)$ defined by $I_{p_{s,t}}\left(\frac{s}{2}, 1+\frac{t}{2}\right) = I_{1-p_{s,t}}\left(\frac{t}{2}, 1+\frac{s}{2}\right)$. The outer minimum is assumed at $(s,t) = \left(\frac{d}{2}, \frac{d}{2}\right)$ for even d and at $(s,t) = \left(\frac{d+1}{2}, \frac{d-1}{2}\right)$ for odd d but this seems extremely hard to prove.

Computing $\vartheta(d)$

Let $d \in \mathbb{N}$ with $d \geq 2$. Breaking the symmetry in s and t,

$$\frac{1}{\vartheta(d)} = \min_{\substack{s,t\in\mathbb{N}\\ s+t=d\\ s\geq t}} \min_{p\in[0,1]} \left(\frac{2(1-p)sI_{1-p}\left(\frac{t}{2}, 1+\frac{s}{2}\right) + 2ptI_p\left(\frac{s}{2}, 1+\frac{t}{2}\right)}{(1-p)s+pt} - 1 \right)$$

where the inner minimum is assumed at $p_{s,t} \in (0,1)$ defined by $I_{p_{s,t}}\left(\frac{s}{2}, 1+\frac{t}{2}\right) = I_{1-p_{s,t}}\left(\frac{t}{2}, 1+\frac{s}{2}\right)$. The outer minimum is assumed at $(s,t) = \left(\frac{d}{2}, \frac{d}{2}\right)$ for even d and at $(s,t) = \left(\frac{d+1}{2}, \frac{d-1}{2}\right)$ for odd d but this seems extremely hard to prove.

For example, one ingredient in the proof is that $p_{s,t} \leq \frac{s}{d}$ (assuming $s, t \in \mathbb{N}$, $s+t=d$ and $s \geq t$) which is equivalent to

$$I_{\frac{s}{d}}\left(\frac{s}{2}, 1+\frac{t}{2}\right) \geq I_{\frac{t}{d}}\left(\frac{t}{2}, 1+\frac{s}{2}\right).$$

Simmons' theorem for half integers

Let $s, t \in \mathbb{N}$ such that $s \geq t$ and set $d := s + t$.

It turns out that for even s and t, the inequality

$$I_{\frac{s}{d}}\left(\frac{s}{2}, 1 + \frac{t}{2}\right) \geq I_{\frac{t}{d}}\left(\frac{t}{2}, 1 + \frac{s}{2}\right)$$

can be interpreted exactly as the statement of Simmons' theorem for $\left(\frac{d}{2}, \frac{s}{2}, \frac{t}{2}\right)$ instead of (d, s, t).

Simmons' theorem for half integers

Let $s, t \in \mathbb{N}$ such that $s \geq t$ and set $d := s + t$.

It turns out that for even s and t, the inequality

$$I_{\frac{s}{d}}\left(\frac{s}{2}, 1 + \frac{t}{2}\right) \geq I_{\frac{t}{d}}\left(\frac{t}{2}, 1 + \frac{s}{2}\right)$$

can be interpreted exactly as the statement of Simmons' theorem for $\left(\frac{d}{2}, \frac{s}{2}, \frac{t}{2}\right)$ instead of (d, s, t).

But what if s or t is odd?

Simmons' theorem for half integers

Let $s, t \in \mathbb{N}$ such that $s \geq t$ and set $d := s + t$.

It turns out that for even s and t, the inequality

$$I_{\frac{s}{d}}\left(\frac{s}{2}, 1 + \frac{t}{2}\right) \geq I_{\frac{t}{d}}\left(\frac{t}{2}, 1 + \frac{s}{2}\right)$$

can be interpreted exactly as the statement of Simmons' theorem for $\left(\frac{d}{2}, \frac{s}{2}, \frac{t}{2}\right)$ instead of (d, s, t).

But what if s or t is odd?

The only proof of Simmons' theorem that somewhat showed potential for generalization to half integers was the one of Perrin and Redside. With a lot of effort we could adapt their idea to find a proof for the half integer case.

Simmons' theorem for reals

Conjecture. For all $s, t \in \mathbb{R}$ such that $s \geq t > 0$, setting $d := s + t$, we have

$$I_{\frac{s}{d}}\left(s, 1+t\right) \geq I_{\frac{t}{d}}\left(t, 1+s\right).$$

Simmons' theorem for reals

Conjecture. For all $s, t \in \mathbb{R}$ such that $s \geq t > 0$, setting $d := s + t$, we have

$$I_{\frac{s}{d}}\left(s, 1+t\right) \geq I_{\frac{t}{d}}\left(t, 1+s\right).$$

It turns out that the above inequality is equivalent to

$$2I_{\frac{s}{d}}(s,t) + (s-t)\frac{s^{s-1}t^{t-1}}{d^d B(s,t)} \geq 1.$$

Simmons' theorem for reals

Conjecture. For all $s, t \in \mathbb{R}$ such that $s \geq t > 0$, setting $d := s + t$, we have

$$I_{\frac{s}{d}}\left(s, 1+t\right) \geq I_{\frac{t}{d}}\left(t, 1+s\right).$$

It turns out that the above inequality is equivalent to

$$2I_{\frac{s}{d}}(s,t) + (s-t)\frac{s^{s-1}t^{t-1}}{d^d B(s,t)} \geq 1.$$

With a completely different method, we show the following weakening of Simmons for reals:

Theorem. For all $s, t \in \mathbb{R}$ such that $s \geq t \geq 1$ and $s + t \geq 3$, setting $d := s + t$, we have

$$2I_{\frac{s}{d}}(s,t) + 2(s-t)\frac{s^{s-1}t^{t-1}}{d^d B(s,t)} \geq 1.$$

The median of the Beta distribution

Reminder. For $s, t \in \mathbb{R}_{>0}$, the beta distribution $\mathrm{Beta}(s, t)$ is the probability distribution on $[0, 1]$ with density $x \mapsto \frac{x^{s-1}x^{t-1}}{B(s,t)}$ and cumulative density $x \mapsto I_x(s, t)$.

The median of the Beta distribution

Reminder. For $s, t \in \mathbb{R}_{>0}$, the beta distribution $\mathrm{Beta}(s,t)$ is the probability distribution on $[0,1]$ with density $x \mapsto \frac{x^{s-1}x^{t-1}}{B(s,t)}$ and cumulative density $x \mapsto I_x(s,t)$.

From the weakening of Simmons' for reals, we deduce:

Theorem. For $s, t \in \mathbb{R}$ with $s \geq t \geq 1$ and $s + t \geq 3$, setting $d := s + t$, the median of $\mathrm{Beta}(s,t)$ lies between $\frac{s}{d}$ and $\frac{s}{d} + \frac{s-t}{d^2}$.

The median of the Beta distribution

Reminder. For $s, t \in \mathbb{R}_{>0}$, the beta distribution $\text{Beta}(s, t)$ is the probability distribution on $[0, 1]$ with density $x \mapsto \frac{x^{s-1}x^{t-1}}{B(s,t)}$ and cumulative density $x \mapsto I_x(s, t)$.

From the weakening of Simmons' for reals, we deduce:

Theorem. For $s, t \in \mathbb{R}$ with $s \geq t \geq 1$ and $s + t \geq 3$, setting $d := s + t$, the median of $\text{Beta}(s, t)$ lies between $\frac{s}{d}$ and $\frac{s}{d} + \frac{s-t}{d^2}$.

The previously known upper bound was only $\frac{s-1}{s-t-2}$ for $s \geq t > 1$.

The median of the Beta distribution

Reminder. For $s, t \in \mathbb{R}_{>0}$, the beta distribution $\mathrm{Beta}(s, t)$ is the probability distribution on $[0, 1]$ with density $x \mapsto \frac{x^{s-1}x^{t-1}}{B(s,t)}$ and cumulative density $x \mapsto I_x(s, t)$.

From the weakening of Simmons' for reals, we deduce:

Theorem. For $s, t \in \mathbb{R}$ with $s \geq t \geq 1$ and $s + t \geq 3$, setting $d := s + t$, the median of $\mathrm{Beta}(s, t)$ lies between $\frac{s}{d}$ and $\frac{s}{d} + \frac{s-t}{d^2}$.

The previously known upper bound was only $\frac{s-1}{s-t-2}$ for $s \geq t > 1$.

If Simmons for reals holds, then we can improve the upper bound further to $\frac{s}{d} + \frac{s-t}{2d^2}$ for $s \geq t \geq 1$.

The median of the Beta distribution

Reminder. For $s, t \in \mathbb{R}_{>0}$, the beta distribution $\mathrm{Beta}(s,t)$ is the probability distribution on $[0,1]$ with density $x \mapsto \frac{x^{s-1}x^{t-1}}{B(s,t)}$ and cumulative density $x \mapsto I_x(s,t)$.

From the weakening of Simmons' for reals, we deduce:

Theorem. For $s, t \in \mathbb{R}$ with $s \geq t \geq 1$ and $s + t \geq 3$, setting $d := s + t$, the median of $\mathrm{Beta}(s,t)$ lies between $\frac{s}{d}$ and $\frac{s}{d} + \frac{s-t}{d^2}$.

The previously known upper bound was only $\frac{s-1}{s-t-2}$ for $s \geq t > 1$.

If Simmons for reals holds, then we can improve the upper bound further to $\frac{s}{d} + \frac{s-t}{2d^2}$ for $s \geq t \geq 1$.

s	t	$\frac{s}{d}$	median	$\frac{s}{d} + \frac{s-t}{2d^2}$	$\frac{s}{d} + \frac{s-t}{d^2}$	$\frac{s-1}{s-t-2}$
2.5	1	0.714286	0.757858	0.77551	0.836735	1
3	1	0.75	0.793701	0.8125	0.875	1
3	2	0.6	0.614272	0.62	0.64	0.666667
4	2	0.666667	0.68619	0.694444	0.722222	0.75

Quiz

Let $d \in \mathbb{N}$ be given.

Quiz

Let $d \in \mathbb{N}$ be given.

Suppose you can choose $s, t \in \mathbb{N}_0$ such that $d = s + t$.
Then you are given a possibly biased coin with probabilities for head and tail $\frac{s}{d}$ and $\frac{t}{d}$, respectively.

Quiz

Let $d \in \mathbb{N}$ be given.

Suppose you can choose $s, t \in \mathbb{N}_0$ such that $d = s + t$.
Then you are given a possibly biased coin with probabilities for head and tail $\frac{s}{d}$ and $\frac{t}{d}$, respectively.

Now you can toss the coin d times.

Quiz

Let $d \in \mathbb{N}$ be given.

Suppose you can choose $s, t \in \mathbb{N}_0$ such that $d = s + t$.
Then you are given a possibly biased coin with probabilities for head and tail $\frac{s}{d}$ and $\frac{t}{d}$, respectively.

Now you can toss the coin d times.

If you obtain at least s times head, you pay me t dollars.
If you obtain at least t times tail, you pay me s dollars.
(Consequently, if you obtain exactly s times head, then you pay d dollars in total.)

Quiz

Let $d \in \mathbb{N}$ be given.

Suppose you can choose $s, t \in \mathbb{N}_0$ such that $d = s + t$.
Then you are given a possibly biased coin with probabilities for head and tail $\frac{s}{d}$ and $\frac{t}{d}$, respectively.

Now you can toss the coin d times.

If you obtain at least s times head, you pay me t dollars.
If you obtain at least t times tail, you pay me s dollars.
(Consequently, if you obtain exactly s times head, then you pay d dollars in total.)

Which coin should you choose to minimize the expected loss?