Lecture 15

Semidefinite programming — Part 1/2

Positive semi-definite matrices

- A matrix $X \in \mathbb{R}^{n \times n}$ is **symmetric** if $X_{ij} = X_{ji}$ for all $i, j \in [n]$

Positive semi-definite matrices

- A matrix $X \in \mathbb{R}^{n \times n}$ is **symmetric** if $X_{ij} = X_{ji}$ for all $i, j \in [n]$
- **Fact.** For a symmetric matrix, all Eigenvalues are real.

Positive semi-definite matrices

- A matrix $X \in \mathbb{R}^{n \times n}$ is **symmetric** if $X_{ij} = X_{ji}$ for all $i, j \in [n]$
- **Fact.** For a symmetric matrix, all Eigenvalues are real.

Definition

A symmetric matrix $X \in \mathbb{R}^{n \times n}$ is **positive semidefinite** if all its Eigenvalues are non-negative.

- We write $X \succeq 0 \Leftrightarrow X$ is PSD.

Positive semi-definite matrices

- A matrix $X \in \mathbb{R}^{n \times n}$ is **symmetric** if $X_{ij} = X_{ji}$ for all $i, j \in [n]$
- **Fact.** For a symmetric matrix, all Eigenvalues are real.

Definition

A symmetric matrix $X \in \mathbb{R}^{n \times n}$ is **positive semidefinite** if all its Eigenvalues are non-negative.

- We write $X \succeq 0 \Leftrightarrow X$ is PSD.
- For $A, B \in \mathbb{R}^{n \times n}$ we write

$$\langle A, B \rangle := \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} \cdot B_{ij}$$

as the **Frobenius inner product**.

Positive semi-definite matrices (2)

Lemma

For a symmetric matrix $X \in \mathbb{R}^{n \times n}$, the following is equivalent

a) $a^T X a \geq 0 \; \forall a \in \mathbb{R}^n$.

b) *X is positive semidefinite.*

c) *There exists a matrix U so that $X = UU^T$.*

d) *There are $u_1, \ldots, u_n \in \mathbb{R}^r$ with $X_{ij} = \langle u_i, u_j \rangle$ for $i, j \in [n]$.*

Positive semi-definite matrices (2)

Lemma

For a symmetric matrix $X \in \mathbb{R}^{n \times n}$, the following is equivalent

a) $a^T X a \geq 0 \; \forall a \in \mathbb{R}^n$.

b) *X is positive semidefinite.*

c) *There exists a matrix U so that $X = UU^T$.*

d) *There are $u_1, \dots, u_n \in \mathbb{R}^r$ with $X_{ij} = \langle u_i, u_j \rangle$ for $i, j \in [n]$.*

Proof:

- Any symmetric real matrix is **diagonalizable**, that means $X = WDW^T = \sum_{i=1}^n \lambda_i v_i v_i^T$ for diagonal D, orth. W.

Then

- $a) \Rightarrow b)$. $0 \leq v_i^T X v_i = \lambda_i \|v_i\|_2^2 = \lambda_i$
- $b) \Rightarrow c)$. $X = WDW^T = UU^T$ for $U := W\sqrt{D}$.
- $c) \Leftrightarrow d)$. Choose u_i as ith row of U.
- $c) \Rightarrow a)$. For any $a \in \mathbb{R}^n$, $a^T X a = \|Ua\|_2^2 \geq 0$.

Positive semi-definite matrices 3

Definition

The **cone of PSD matrices** is

$$\begin{aligned}\mathbb{S}^n_{\geq 0} &:= \{X \in \mathbb{R}^{n \times n} \mid X \text{ symmetric}, X \succeq 0\} \\ &= \{X \in \mathbb{R}^{n \times n} \mid X \text{ symmetric}, \ \langle X, aa^T \rangle \geq 0 \ \forall a \in \mathbb{R}^n\}\end{aligned}$$

Positive semi-definite matrices 3

Definition

The **cone of PSD matrices** is

$$\begin{aligned}\mathbb{S}^n_{\geq 0} &:= \{X \in \mathbb{R}^{n\times n} \mid X \text{ symmetric}, X \succeq 0\} \\ &= \{X \in \mathbb{R}^{n\times n} \mid X \text{ symmetric}, \langle X, aa^T\rangle \geq 0 \; \forall a \in \mathbb{R}^n\}\end{aligned}$$

- **Fact:** $\mathbb{S}^n_{\geq 0}$ is convex.

Positive semi-definite matrices 3

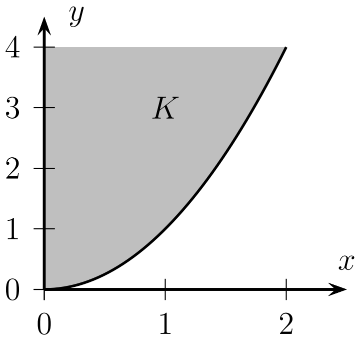
Definition

The **cone of PSD matrices** is

$$\begin{aligned}\mathbb{S}^n_{\geq 0} &:= \{X \in \mathbb{R}^{n\times n} \mid X \text{ symmetric}, X \succeq 0\} \\ &= \{X \in \mathbb{R}^{n\times n} \mid X \text{ symmetric}, \langle X, aa^T\rangle \geq 0 \; \forall a \in \mathbb{R}^n\}\end{aligned}$$

- **Fact:** $\mathbb{S}^n_{\geq 0}$ is convex.

$$K := \left\{(x,y) \mid \begin{pmatrix} 1 & x \\ x & y \end{pmatrix} \succeq 0\right\}$$



A semidefinite program

- A **semidefinite program** is of the form:

$$\begin{aligned} \max \langle C, X \rangle & & \\ \langle A_k, X \rangle &\leq b_k \quad \forall k = 1, \ldots, m \\ X & \quad \text{symmetric} \\ X &\succeq 0 \end{aligned}$$

where $C, A_1, \ldots, A_m \in \mathbb{R}^{n \times n}$.

A semidefinite program

- A **semidefinite program** is of the form:

$$\begin{aligned} \max \langle C, X\rangle & & \\ \langle A_k, X\rangle & \leq & b_k \quad \forall k = 1, \ldots, m \\ X & & \text{symmetric} \\ X & \succeq & 0 \end{aligned}$$

where $C, A_1, \ldots, A_m \in \mathbb{R}^{n \times n}$.

Less well behaved than LPs:

- **Issue 1:** Strong duality might fail.
- **Issue 2:** Possibly all solutions are irrational
- **Issue 3:** Possibly exact solutions have exponential encoding length

Solvability of Semidefinite Programs

Theorem

Given rational input $A_1, \dots, A_m, b_1, \dots, b_m, C, R$ *and* $\varepsilon > 0$*, suppose*

$$SDP = \max\{\langle C, X\rangle \mid \langle A_k, X\rangle \leq b_k \ \forall k; \ X \, symmetric; \ X \succeq 0\}$$

is feasible and all feasible points are contained in $B(\mathbf{0}, R)$*. Then one can find a* X^* *with*

$$\langle A_k, X^*\rangle \leq b_k + \varepsilon, \ X^* \ symmetric, \ X^* \succeq 0$$

such that $\langle C, X^*\rangle \geq SDP - \varepsilon$*. The running time is polynomial in the input length,* $\log(R)$ *and* $\log(1/\varepsilon)$ *(in the Turing machine model).*

Vector programs

Idea:

- $Y \succeq 0$ holds iff $Y_{ij} = \langle v_i, v_j \rangle$ for some vectors v_i

Vector programs

Idea:

- $Y \succeq 0$ holds iff $Y_{ij} = \langle v_i, v_j \rangle$ for some vectors v_i

SDP:

$$\max \sum_{i,j} C_{ij} Y_{ij}$$

$$\begin{array}{rcll} \sum_{i,j} A_{ij}^k \cdot Y_{ij} & \leq & b_k & \forall k \\ Y & & \text{sym.} & \\ Y & \succeq & 0 & \end{array}$$

Vector program

$$\max \sum_{i,j} C_{ij} \langle v_i, v_j \rangle$$

$$\begin{array}{rcll} \sum_{i,j} A_{ij}^k \cdot \langle v_i, v_j \rangle & \leq & b_k & \forall k \\ v_i & \in & \mathbb{R}^r & \forall i \end{array}$$

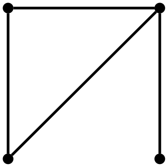
Observation

The SDP and the vector program are equivalent.

MaxCut

MaxCut
Input: An undirected graph $G = (V, E)$
Goal: Find the cut $S \subseteq V$ that maximizes the number $|\delta(S)|$ of cut edges.

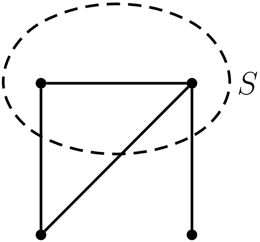
Example:



MaxCut

MaxCut
Input: An undirected graph $G = (V, E)$
Goal: Find the cut $S \subseteq V$ that maximizes the number $|\delta(S)|$ of cut edges.

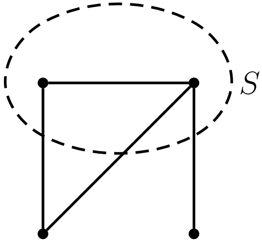
Example:



MaxCut

> MaxCut
> **Input:** An undirected graph $G = (V, E)$
> **Goal:** Find the cut $S \subseteq V$ that maximizes the number $|\delta(S)|$ of cut edges.

Example:



- **NP**-hard to find a solution that cuts even 94% of what the optimum cuts [Hastad 1997]
- Simple greedy algorithm cuts at least $|E|/2$ edges.

MaxCut SDP

SDP:

$$\begin{aligned} \max \quad & \frac{1}{2} \sum_{\{i,j\} \in E} (1 - X_{ij}) \\ X &\succeq 0 \\ X_{ii} &= 1 \quad \forall i \in V \\ X &\in \mathbb{R}^{n \times n} \end{aligned}$$

Vector program

$$\begin{aligned} \max \quad & \frac{1}{2} \sum_{\{i,j\} \in E} (1 - \langle u_i, u_j \rangle) \\ \|u_i\|_2 &= 1 \quad \forall i \in V \\ u_i &\in \mathbb{R}^r \end{aligned}$$

MaxCut SDP

SDP:

$$\begin{aligned} \max \quad & \frac{1}{2} \sum_{\{i,j\} \in E} (1 - X_{ij}) \\ X & \succeq 0 \\ X_{ii} & = 1 \quad \forall i \in V \\ X & \in \mathbb{R}^{n \times n} \end{aligned}$$

Vector program

$$\begin{aligned} \max \quad & \frac{1}{2} \sum_{\{i,j\} \in E} (1 - \langle u_i, u_j \rangle) \\ \|u_i\|_2 & = 1 \quad \forall i \in V \\ u_i & \in \mathbb{R}^r \end{aligned}$$

Lemma

If $S^* \subseteq V$ is opt. solution for MaxCut, then $SDP \geq |\delta(S^*)|$.

MaxCut SDP

SDP:

$$\begin{aligned} \max \quad & \frac{1}{2}\sum_{\{i,j\}\in E}(1 - X_{ij}) \\ X &\succeq 0 \\ X_{ii} &= 1 \quad \forall i \in V \\ X &\in \mathbb{R}^{n\times n} \end{aligned}$$

Vector program

$$\begin{aligned} \max \quad & \frac{1}{2}\sum_{\{i,j\}\in E}(1 - \langle u_i, u_j\rangle) \\ \|u_i\|_2 &= 1 \quad \forall i \in V \\ u_i &\in \mathbb{R}^r \end{aligned}$$

Lemma

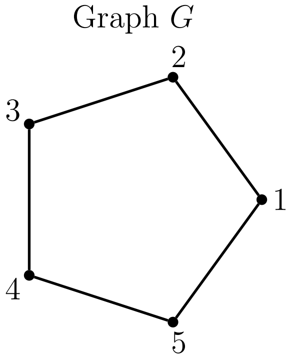
If $S^* \subseteq V$ is opt. solution for MaxCut, then $SDP \geq |\delta(S^*)|$.

Proof:

- We set $r := 1$ and define $u_i \in \mathbb{R}^1$ by

$$u_i := \begin{cases} 1 & \text{if } i \in S^* \\ -1 & \text{if } i \in V \setminus S^* \end{cases}$$

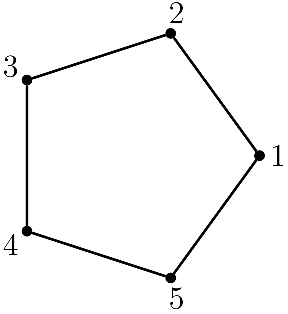
Example MaxCut SDP



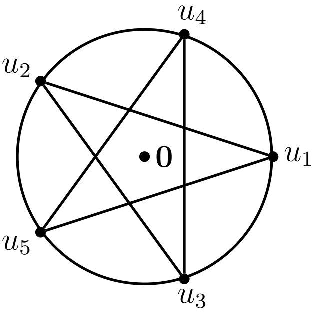
- Optimum MaxCut = 4

Example MaxCut SDP

Graph G



SDP solution:



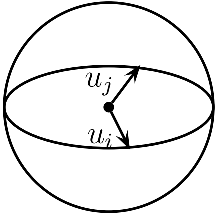
- Optimum MaxCut $= 4$
- Choose $u_i \in \mathbb{R}^2$ with $u_i := (\cos(\frac{4i\pi}{4}), \sin(\frac{4\pi}{5}))$ and we get vector program solution of value $5 \cdot \frac{1}{2}(1 - \cos(\frac{4}{5}\pi) \approx 4.522$

The Hyperplane Rounding algorithm

(1) Solve the SDP

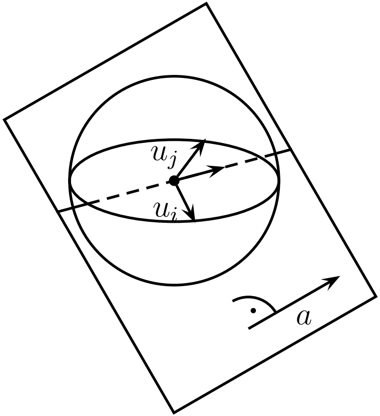
(2) Take a random standard Gaussian $a \in \mathbb{R}^r$

(3) Define $S := \{i \in V \mid \langle a, u_i \rangle \geq 0\}$



The Hyperplane Rounding algorithm

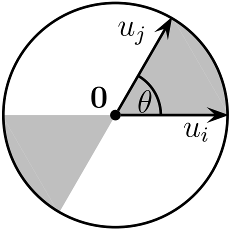
(1) Solve the SDP
(2) Take a random standard Gaussian $a \in \mathbb{R}^r$
(3) Define $S := \{i \in V \mid \langle a, u_i \rangle \geq 0\}$



The Hyperplane Rounding algorithm (2)

Lemma

For $\{i, j\} \in E$ *one has* $\Pr[\{i, j\} \in \delta(S)] = \frac{1}{\pi} arccos(\langle u_i, u_j \rangle)$.



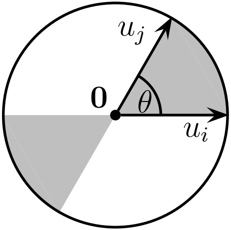
The Hyperplane Rounding algorithm (2)

Lemma

For $\{i,j\} \in E$ *one has* $\Pr[\{i,j\} \in \delta(S)] = \frac{1}{\pi} arccos(\langle u_i, u_j \rangle)$.

Proof.

- The angle between vectors is exactly $\theta := \arccos(\langle u_i, u_j \rangle)$ (as $\langle u_i, u_j \rangle = \cos(\theta)$).
- Only projection of a into $U := \text{span}\{u_i, u_j\}$ matters.



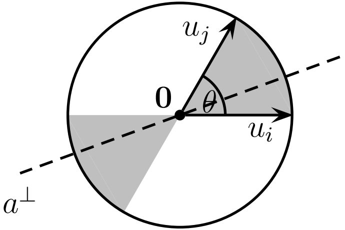
The Hyperplane Rounding algorithm (2)

Lemma

For $\{i,j\} \in E$ *one has* $\Pr[\{i,j\} \in \delta(S)] = \frac{1}{\pi} arccos(\langle u_i, u_j \rangle)$.

Proof.

- The angle between vectors is exactly $\theta := \arccos(\langle u_i, u_j \rangle)$ (as $\langle u_i, u_j \rangle = \cos(\theta)$).
- Only projection of a into $U := \text{span}\{u_i, u_j\}$ matters.
- Then $\Pr[u_i, u_j \text{ separated}] = \frac{2\theta}{2\pi}$.



The Hyperplane Rounding algorithm (3)

Theorem

One has $\mathbb{E}[|\delta(S)|] \geq 0.878 \cdot SDP \geq 0.878 \cdot |\delta(S^*)|$.

The Hyperplane Rounding algorithm (3)

Theorem

One has $\mathbb{E}[|\delta(S)|] \geq 0.878 \cdot SDP \geq 0.878 \cdot |\delta(S^*)|$.

- By **linearity of expectation** it suffices to show that every edge $\{i,j\} \in E$ one has

  $\Pr[\{i,j\} \in \delta(S)] \geq \frac{1}{2}(1 - \langle u_i, u_j \rangle) =$ contribution to SDP obj.fct

The Hyperplane Rounding algorithm (3)

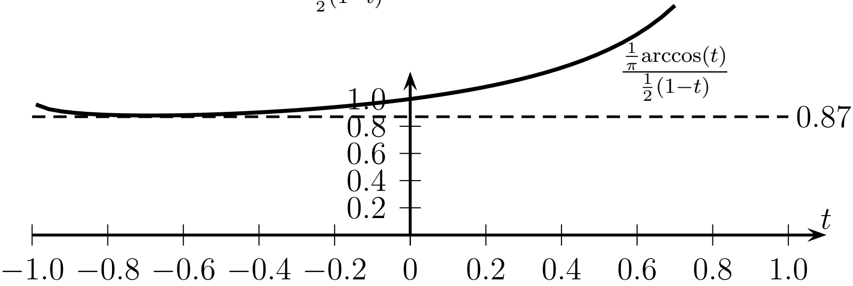
Theorem

One has $\mathbb{E}[|\delta(S)|] \geq 0.878 \cdot SDP \geq 0.878 \cdot |\delta(S^*)|$.

- By **linearity of expectation** it suffices to show that every edge $\{i,j\} \in E$ one has

  $\Pr[\{i,j\} \in \delta(S)] \geq \frac{1}{2}(1 - \langle u_i, u_j \rangle) =$ contribution to SDP obj.fct

- Set $t := \langle u_i, u_j \rangle$ and $\frac{\frac{1}{\pi}\arccos(t)}{\frac{1}{2}(1-t)} \geq 0.878 \quad \forall t \in [-1,1]$



Lecture 16

Semidefinite programming — Part 2/2

Grothendieck's Inequality

For a matrix $A \in \mathbb{R}^{m \times n}$ define

$$
\begin{aligned}
INT(A) &:= \max\left\{ \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} x_i y_j \mid x \in \{-1,1\}^m, y \in \{-1,1\}^n \right\} \\
SDP(A) &:= \max\left\{ \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} \langle u_i, v_j \rangle \mid \|u_i\|_2 = \|v_j\|_2 = 1 \right\}
\end{aligned}
$$

Grothendieck's Inequality

For a matrix $A \in \mathbb{R}^{m \times n}$ define

$$INT(A) := \max \left\{ \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} x_i y_j \mid x \in \{-1,1\}^m, y \in \{-1,1\}^n \right\}$$

$$SDP(A) := \max \left\{ \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} \langle u_i, v_j \rangle \mid \|u_i\|_2 = \|v_j\|_2 = 1 \right\}$$

Theorem (Grothendieck's Inequality)

For any matrix $A \in \mathbb{R}^{m \times n}$ one has

$$INT(A) \leq SDP(A) \leq C_G \cdot INT(A)$$

where $C_G \leq 1.783$.

- Grothendieck proved that C_G is indeed a constant
- [Krivine 1979] proved that $C_G \leq 1.783$

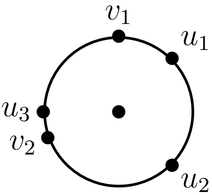
Hyperplane rounding

Random experiment:

(1) Given vectors $u_i, v_j \in \mathbb{R}^r$.

(2) Sample a **Gaussian** g in $\mathbb{R}^r$ and set

$$x_i := \text{sign}(\langle u_i, g\rangle) \quad \text{and} \quad y_j := \text{sign}(\langle v_j, g\rangle)$$



- Recall that

$$\text{sign}(z) := \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{if } z < 0 \end{cases}$$

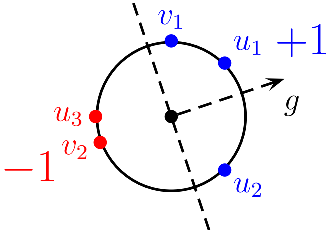
Hyperplane rounding

Random experiment:

(1) Given vectors $u_i, v_j \in \mathbb{R}^r$.

(2) Sample a **Gaussian** g in $\mathbb{R}^r$ and set

$$x_i := \text{sign}(\langle u_i, g\rangle) \quad \text{and} \quad y_j := \text{sign}(\langle v_j, g\rangle)$$



- Recall that

$$\text{sign}(z) := \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{if } z < 0 \end{cases}$$

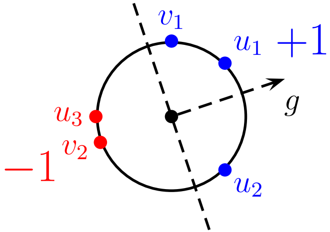
Hyperplane rounding

Random experiment:

(1) Given vectors $u_i, v_j \in \mathbb{R}^r$.

(2) Sample a **Gaussian** g in $\mathbb{R}^r$ and set

$$x_i := \text{sign}(\langle u_i, g \rangle) \quad \text{and} \quad y_j := \text{sign}(\langle v_j, g \rangle)$$



- Recall that

$$\text{sign}(z) := \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{if } z < 0 \end{cases}$$

- **Question:** How does $\mathbb{E}[A_{ij} x_i y_j]$ relate to $A_{ij} \langle u_i, v_j \rangle$?

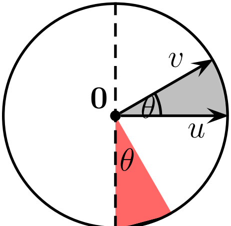
Hyperplane rounding (2)

Lemma

Let $u, v \in \mathbb{R}^r$ with $\|u\|_2 = \|v\|_2 = 1$. Then

$$\mathop{\mathbb{E}}_{g\ Gaussian}\left[sign(\langle g, u\rangle) \cdot sign(\langle g, v\rangle)\right] = \frac{2}{\pi} arcsin(\langle u, v\rangle)$$

- In words: Probability that u, v end up on the same side of a hyperplane is exactly $\frac{2}{\pi}\arcsin(\langle u, v\rangle)$



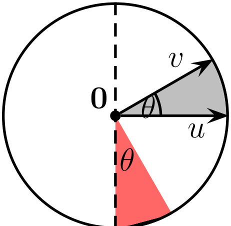
Hyperplane rounding (2)

Lemma

Let $u, v \in \mathbb{R}^r$ with $\|u\|_2 = \|v\|_2 = 1$. Then

$$\mathop{\mathbb{E}}_{g\ Gaussian}\left[sign(\langle g, u\rangle) \cdot sign(\langle g, v\rangle)\right] = \frac{2}{\pi} arcsin(\langle u, v\rangle)$$

- In words: Probability that u, v end up on the same side of a hyperplane is exactly $\frac{2}{\pi}\arcsin(\langle u, v\rangle)$
- Set $\cos(\theta) = \langle u, v\rangle$. Then $\Pr[u, v \text{ separated}] = \frac{\theta}{\pi}$



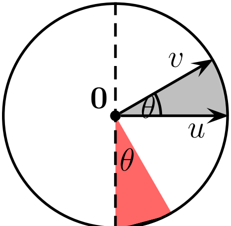
Hyperplane rounding (2)

Lemma

Let $u, v \in \mathbb{R}^r$ with $\|u\|_2 = \|v\|_2 = 1$. Then

$$\mathop{\mathbb{E}}_{g\ Gaussian}\left[sign(\langle g,u\rangle)\cdot sign(\langle g,v\rangle)\right] = \frac{2}{\pi} arcsin(\langle u,v\rangle)$$

- In words: Probability that u, v end up on the same side of a hyperplane is exactly $\frac{2}{\pi}\text{arcsin}(\langle u,v\rangle)$
- Set $\cos(\theta) = \langle u,v\rangle$. Then $\Pr[u,v \text{ separated}] = \frac{\theta}{\pi}$
- $\mathbb{E}[\cdots] = 1 - 2\Pr[u,v \text{ separated}] = 1 - \frac{2\theta}{\pi} = \frac{2}{\pi}\text{arcsin}(\langle u,v\rangle)$
- Recall: $\arccos(t) = \frac{\pi}{2} - \arcsin(t)$

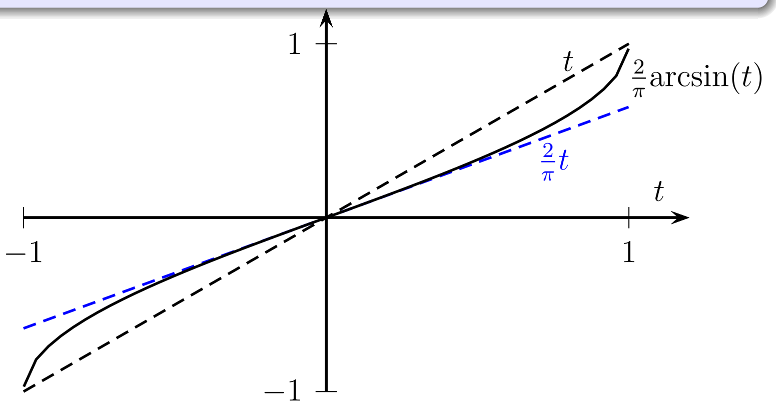


Hyperplane rounding (3)

Lemma

Let $u, v \in \mathbb{R}^r$ with $\|u\|_2 = \|v\|_2 = 1$. Then

$$\mathop{\mathbb{E}}_{g \text{ Gaussian}} \left[sign(\langle g, u \rangle) \cdot sign(\langle g, v \rangle) \right] = \frac{2}{\pi} arcsin(\langle u, v \rangle)$$



For $t \geq 0$, $\frac{2}{\pi}t \leq \frac{2}{\pi}\arcsin(t) \leq t$

Preliminary conclusion

We can conclude that:

- For $A_{ij} \geq 0$ and $\langle u_i, u_j \rangle \geq 0$ one has $\mathbb{E}[A_{ij} x_i y_j] \geq \frac{2}{\pi} \cdot A_{ij} \langle u_i, v_j \rangle$
- For $A_{ij} < 0$ and $\langle u_i, u_j \rangle \geq 0$ one has $\mathbb{E}[A_{ij} x_i y_j] \geq A_{ij} \langle u_i, v_j \rangle$

Preliminary conclusion

We can conclude that:

- For $A_{ij} \geq 0$ and $\langle u_i, u_j \rangle \geq 0$ one has $\mathbb{E}[A_{ij} x_i y_j] \geq \frac{2}{\pi} \cdot A_{ij} \langle u_i, v_j \rangle$
- For $A_{ij} < 0$ and $\langle u_i, u_j \rangle \geq 0$ one has $\mathbb{E}[A_{ij} x_i y_j] \geq A_{ij} \langle u_i, v_j \rangle$

Problem: Due to the non-linearity, this does bound $INT(A)$ in terms of $SDP(A)$!!

Tensors

Definition

A k**th order tensor** $A \in \mathbb{R}^{n_1 \times \ldots \times n_k}$ is a k-dimensional array of numbers; we write $A = (A_{i_1,\ldots,i_k})_{i_1 \in [n_1],\ldots,i_k \in [n_k]}$.

- A vector $a \in \mathbb{R}^n$ is a 1-order tensor
- A matrix $A \in \mathbb{R}^{n_1 \times n_2}$ is a 2-order tensor

Tensors

Definition

A **kth order tensor** $A \in \mathbb{R}^{n_1 \times \ldots \times n_k}$ is a k-dimensional array of numbers; we write $A = (A_{i_1,\ldots,i_k})_{i_1 \in [n_1],\ldots,i_k \in [n_k]}$.

- A vector $a \in \mathbb{R}^n$ is a 1-order tensor
- A matrix $A \in \mathbb{R}^{n_1 \times n_2}$ is a 2-order tensor
- For two tensors $A, B \in \mathbb{R}^{n_1 \times \ldots \times n_k}$ we can define an **inner product**

$$\langle A, B \rangle := \sum_{i_1,\ldots,i_k} A_{i_1,\ldots,i_k} \cdot B_{i_1,\ldots,i_k}$$

Tensors

Definition

A **kth order tensor** $A \in \mathbb{R}^{n_1 \times \ldots \times n_k}$ is a k-dimensional array of numbers; we write $A = (A_{i_1,\ldots,i_k})_{i_1 \in [n_1],\ldots,i_k \in [n_k]}$.

- A vector $a \in \mathbb{R}^n$ is a 1-order tensor
- A matrix $A \in \mathbb{R}^{n_1 \times n_2}$ is a 2-order tensor
- For two tensors $A, B \in \mathbb{R}^{n_1 \times \ldots \times n_k}$ we can define an **inner product**

$$\langle A, B \rangle := \sum_{i_1,\ldots,i_k} A_{i_1,\ldots,i_k} \cdot B_{i_1,\ldots,i_k}$$

- For vector $u \in \mathbb{R}^n$ and $k \in \mathbb{N}$, define the **tensor product**

$$u \otimes \ldots \otimes u := u^{\otimes k} := (u_{i_1} \cdot \ldots \cdot u_{i_k})_{i_1 \in [n],\ldots,i_k \in [n]}$$

Tensors

Definition

A **kth order tensor** $A \in \mathbb{R}^{n_1 \times \ldots \times n_k}$ is a k-dimensional array of numbers; we write $A = (A_{i_1,\ldots,i_k})_{i_1 \in [n_1],\ldots,i_k \in [n_k]}$.

- A vector $a \in \mathbb{R}^n$ is a 1-order tensor
- A matrix $A \in \mathbb{R}^{n_1 \times n_2}$ is a 2-order tensor
- For two tensors $A, B \in \mathbb{R}^{n_1 \times \ldots \times n_k}$ we can define an **inner product**

$$\langle A, B \rangle := \sum_{i_1,\ldots,i_k} A_{i_1,\ldots,i_k} \cdot B_{i_1,\ldots,i_k}$$

- For vector $u \in \mathbb{R}^n$ and $k \in \mathbb{N}$, define the **tensor product**

$$u \otimes \ldots \otimes u := u^{\otimes k} := (u_{i_1} \cdot \ldots \cdot u_{i_k})_{i_1 \in [n],\ldots,i_k \in [n]}$$

- **Fact:** For vectors $u, v \in \mathbb{R}^n$ one has $\langle u^{\otimes k}, v^{\otimes k} \rangle = \langle u, v \rangle^k$.

Tensors

Definition

We call a function $f : \mathbb{R} \to \mathbb{R}$ **(real) analytic** if it can be written as a convergent power series $f(x) = \sum_{k=0}^{\infty} a_k x^k$ for all $x \in \mathbb{R}$.

Tensors

Definition

We call a function $f : \mathbb{R} \to \mathbb{R}$ **(real) analytic** if it can be written as a convergent power series $f(x) = \sum_{k=0}^{\infty} a_k x^k$ for all $x \in \mathbb{R}$.

- For fixed r, we can define a **Hilbert space / infinite-dimensional vector space** of the form

$$H = \{(U^0, U^1, U^2, U^3, \dots) \mid U^k \text{ is a } k\text{-tensor of size } r^k\}$$

using the natural inner product.

A vector transformation

Now we can "bend" any vectors to give any analytic function that we like:

Lemma

Let $f(x) = \sum_{k=0}^{\infty} a_k x^k$ and fix a dimension $r \in \mathbb{N}$. Then there is a Hilbert space H and maps $F, G : \mathbb{R}^r \to H$ so that

$$\langle F(u), G(v) \rangle = f(\langle u, v \rangle) \quad \forall u, v \in \mathbb{R}^r$$

Moreover the length of the mapped vectors satisfies

$$\|F(u)\|_2^2 = \|G(u)\|_2^2 = \sum_{k=0}^{\infty} |a_k| \cdot \|u\|_2^{2k}$$

A vector transformation (2)

Proof:

- The maps are

$$F(u) := (\sqrt{|a_k|}\cdot u^{\otimes k})_{k\in\mathbb{Z}_{\geq 0}}, \quad G(u) := (\mathrm{sign}(a_k)\cdot\sqrt{|a_k|}\cdot u^{\otimes k})_{k\in\mathbb{Z}_{\geq 0}}$$

A vector transformation (2)

Proof:

- The maps are

$$F(u) := (\sqrt{|a_k|}\cdot u^{\otimes k})_{k\in\mathbb{Z}_{\geq 0}}, \quad G(u) := (\operatorname{sign}(a_k)\cdot\sqrt{|a_k|}\cdot u^{\otimes k})_{k\in\mathbb{Z}_{\geq 0}}$$

- Then for vectors $u, v \in \mathbb{R}^r$ one has

$$\begin{aligned} \langle F(u), G(v)\rangle &= \sum_{k\geq 0} \operatorname{sign}(a_k)\cdot(\sqrt{|a_k|})^2 \cdot \langle u^{\otimes k}, v^{\otimes k}\rangle \\ &= \sum_{k\geq 0} a_k \cdot \langle u, v\rangle^k = f(\langle u, v\rangle). \end{aligned}$$

A vector transformation (2)

Proof:

- The maps are

$$F(u) := (\sqrt{|a_k|}\cdot u^{\otimes k})_{k\in\mathbb{Z}_{\geq 0}}, \quad G(u) := (\text{sign}(a_k)\cdot\sqrt{|a_k|}\cdot u^{\otimes k})_{k\in\mathbb{Z}_{\geq 0}}$$

- Then for vectors $u, v \in \mathbb{R}^r$ one has

$$\begin{aligned}\langle F(u), G(v)\rangle &= \sum_{k\geq 0} \text{sign}(a_k)\cdot(\sqrt{|a_k|})^2\cdot\langle u^{\otimes k}, v^{\otimes k}\rangle \\ &= \sum_{k\geq 0} a_k\cdot\langle u, v\rangle^k = f(\langle u, v\rangle).\end{aligned}$$

- We can verify that the lengths are

$$\|F(u)\|_2^2 = \|G(u)\|_2^2 = \sum_{k\geq 0}(\sqrt{|a_k|})^2\cdot\|u^{\otimes k}\|_2^2 = \sum_{k\geq 0}|a_k|\cdot\|u\|_2^{2k}$$

as claimed. □

Applying the vector transformation

Lemma

Let $r \in \mathbb{N}$. Then there are maps $F, G : \mathbb{R}^r \to H$ so that

$$\langle F(u), G(v) \rangle = \sin\left(\beta \frac{\pi}{2} \langle u, v \rangle \right)$$

where $\beta = \frac{2}{\pi} \ln(1 + \sqrt{2}) \approx \frac{1}{1.783}$. Moreover $\|F(u)\|_2^2 = \|G(u)\|_2^2 = 1$ for all $u \in \mathbb{R}^r$ with $\|u\|_2^2 = 1$.

Note that this is equivalent to

$$\frac{2}{\pi} \arcsin(\langle F(u), G(v) \rangle) = \beta \cdot \langle u, v \rangle$$

Applying the vector transformation (2)

Proof.

- Consider $f(x) = \sin(\beta \frac{\pi}{2} x)$.

Applying the vector transformation (2)

Proof.

- Consider $f(x) = \sin(\beta \frac{\pi}{2} x)$.
- Recall that
  $\sin(x) = \sum_{k \geq 0} \frac{(-1)^k}{(2k+1)!} x^{2k+1} = x - \frac{1}{3!} x^3 + \frac{1}{5!} x^5 - \dots$
  $\sinh(x) = \sum_{k \geq 0} \frac{1}{(2k+1)!} x^{2k+1}$

Applying the vector transformation (2)

Proof.

- Consider $f(x) = \sin(\beta \frac{\pi}{2} x)$.
- Recall that
  $\sin(x) = \sum_{k \geq 0} \frac{(-1)^k}{(2k+1)!} x^{2k+1} = x - \frac{1}{3!} x^3 + \frac{1}{5!} x^5 - \dots$
  $\sinh(x) = \sum_{k \geq 0} \frac{1}{(2k+1)!} x^{2k+1}$
- Then for $\|u\|_2 = 1$,

$$\|F(u)\|_2^2 = \sum_{k \geq 0} \left| \frac{(-1)^k}{(2k+1)!} \cdot \left(\beta \frac{\pi}{2}\right)^{2k+1} \right| = \sinh\left(\beta \frac{\pi}{2}\right) \stackrel{\beta := \frac{2}{\pi} \operatorname{arcsinh}(1)}{=} 1$$

Applying the vector transformation (2)

Proof.

- Consider $f(x) = \sin(\beta \frac{\pi}{2} x)$.
- Recall that
  $\sin(x) = \sum_{k \geq 0} \frac{(-1)^k}{(2k+1)!} x^{2k+1} = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots$
  $\sinh(x) = \sum_{k \geq 0} \frac{1}{(2k+1)!} x^{2k+1}$
- Then for $\|u\|_2 = 1$,

$$\|F(u)\|_2^2 = \sum_{k \geq 0} \left| \frac{(-1)^k}{(2k+1)!} \cdot \left(\beta \frac{\pi}{2}\right)^{2k+1} \right| = \sinh\left(\beta \frac{\pi}{2}\right) \overset{\beta := \frac{2}{\pi}\operatorname{arcsinh}(1)}{=} 1$$

- One can check that

$$\beta = \frac{2}{\pi}\operatorname{arcsinh}(1) = \frac{2}{\pi}\ln(1 + \sqrt{2}) \approx \frac{1}{1.783}.$$

Applying the vector transformation (3)

- Consider $A \in \mathbb{R}^{m \times n}$ and $u_i, v_j \in \mathbb{R}^r$ with $\|u_i\|_2 = 1 = \|v_j\|_2$.
- Sample a **Gaussian** g in H and set

$$x_i := \operatorname{sign}(\langle g, F(u_i) \rangle) \quad \text{and} \quad y_j := \operatorname{sign}(\langle g, G(v_j) \rangle)$$

Applying the vector transformation (3)

- Consider $A \in \mathbb{R}^{m \times n}$ and $u_i, v_j \in \mathbb{R}^r$ with $\|u_i\|_2 = 1 = \|v_j\|_2$.
- Sample a **Gaussian** g in H and set

$$x_i := \text{sign}(\langle g, F(u_i) \rangle) \quad \text{and} \quad y_j := \text{sign}(\langle g, G(v_j) \rangle)$$

- Then

$$\mathbb{E}[x_i y_j] = \frac{2}{\pi} \arcsin(\langle F(u_i), G(v_i) \rangle) = \beta \cdot \langle u_i, v_i \rangle$$

Applying the vector transformation (3)

- Consider $A \in \mathbb{R}^{m \times n}$ and $u_i, v_j \in \mathbb{R}^r$ with $\|u_i\|_2 = 1 = \|v_j\|_2$.
- Sample a **Gaussian** g in H and set

$$x_i := \text{sign}(\langle g, F(u_i) \rangle) \quad \text{and} \quad y_j := \text{sign}(\langle g, G(v_j) \rangle)$$

- Then

$$\mathbb{E}[x_i y_j] = \frac{2}{\pi} \arcsin(\langle F(u_i), G(v_i) \rangle) = \beta \cdot \langle u_i, v_i \rangle$$

- By linearity of expectation

$$\mathbb{E}\Big[\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} x_i y_j\Big] = \underbrace{\beta}_{\approx \frac{1}{1.783}} \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} \langle u_i, v_j \rangle \quad \square$$