

# Tensors Decompositions and the Nuclear Norm

Harm Derksen

University of Michigan

TDA 2016

January 19, 2016

# Tensor Rank

$\mathbb{F}$  a field,  $\mathbb{F} = \mathbb{R}$  or  $\mathbb{F} = \mathbb{C}$

$V^{(i)} \cong \mathbb{F}^{n_i}$  for  $i = 1, 2, \dots, d$

$V = V^{(1)} \otimes V^{(2)} \otimes \dots \otimes V^{(d)} \cong \mathbb{F}^{n_1 \times \dots \times n_d}$  tensor product space

## Definition

A *pure* tensor is a tensor of the form  $v^{(1)} \otimes v^{(2)} \otimes \dots \otimes v^{(d)}$   
( $v^{(i)} \in V^{(i)}$ ).

# Tensor Rank

$\mathbb{F}$  a field,  $\mathbb{F} = \mathbb{R}$  or  $\mathbb{F} = \mathbb{C}$

$V^{(i)} \cong \mathbb{F}^{n_i}$  for  $i = 1, 2, \dots, d$

$V = V^{(1)} \otimes V^{(2)} \otimes \dots \otimes V^{(d)} \cong \mathbb{F}^{n_1 \times \dots \times n_d}$  tensor product space

## Definition

A *pure* tensor is a tensor of the form  $v^{(1)} \otimes v^{(2)} \otimes \dots \otimes v^{(d)}$   
( $v^{(i)} \in V^{(i)}$ ).

## Definition (tensor rank)

The *rank* of  $T$  is the smallest positive integer  $r$  such that  $T$  is the sum of  $r$  pure tensors.

# The Canonical Polyadic (CP) Model

aka PARAFAC, CANDECOMP

## Problem

*Given a tensor  $T \in V$ , write  $T = v_1 + v_2 + \dots + v_r$  where  $v_1, v_2, \dots, v_r$  are pure tensors and  $r$  is minimal.*

# The Canonical Polyadic (CP) Model

aka PARAFAC, CANDECOMP

## Problem

*Given a tensor  $T \in V$ , write  $T = v_1 + v_2 + \dots + v_r$  where  $v_1, v_2, \dots, v_r$  are pure tensors and  $r$  is minimal.*

- ▶ the CP-decomposition may not be unique
- ▶ not numerically stable: there is no upper bound for  $\max\{\|v_1\|, \dots, \|v_r\|\}$  as a function of  $\|T\|$ .
- ▶ difficult to compute: for many tensors of interest the rank is unknown

To make it more numerically stable, we could allow an error term.

# The Canonical Polyadic (CP) Model

## Problem

*Given a tensor  $T \in V$  and fixed  $r$ , write  $T = v_1 + v_2 + \cdots + v_r + E$  where  $v_1, v_2, \dots, v_r$  are pure tensors and  $\|E\|$  is minimal.*

# The Canonical Polyadic (CP) Model

## Problem

Given a tensor  $T \in V$  and fixed  $r$ , write  $T = v_1 + v_2 + \dots + v_r + E$  where  $v_1, v_2, \dots, v_r$  are pure tensors and  $\|E\|$  is minimal.

This optimization problem might be ill-posed. Define

$$T = e_2 \otimes e_1 \otimes e_1 + e_1 \otimes e_2 \otimes e_1 + e_1 \otimes e_1 \otimes e_2,$$

$$v_1(t) = t^{-1}(e_1 + te_2) \otimes (e_1 + te_2) \otimes (e_1 + te_2), \quad v_2(t) = -t^{-1}e_1 \otimes e_1 \otimes e_1$$

Then  $T = v_1(t) + v_2(t) + E(t)$  with  $\|E(t)\| = t\sqrt{3+t}$ .

So we can write  $T = v_1 + v_2 + E$  with  $\|E\| = \varepsilon$  for every  $\varepsilon > 0$ , but we cannot write  $T = v_1 + v_2 + E$  with  $\|E\| = 0$ .

# Motivation: Compressed Sensing and Convex Relaxation

For  $x \in \mathbb{R}^n$ , its sparsity is measured by  $\|x\|_0 = \#\{i \mid x_i \neq 0\}$ .

## Problem

*Given  $A \in \text{Mat}_{m,n}(\mathbb{R})$ ,  $b \in \mathbb{R}^m$ , find a solution  $x \in \mathbb{R}^n$  for  $Ax = b$  with  $\|x\|_0$  minimal (a sparsest solution).*

But,  $\|\cdot\|_0$  is not convex and this optimization problem is difficult,



# Motivation: Compressed Sensing and Convex Relaxation

For  $x \in \mathbb{R}^n$ , its sparsity is measured by  $\|x\|_0 = \#\{i \mid x_i \neq 0\}$ .

## Problem

*Given  $A \in \text{Mat}_{m,n}(\mathbb{R})$ ,  $b \in \mathbb{R}^m$ , find a solution  $x \in \mathbb{R}^n$  for  $Ax = b$  with  $\|x\|_0$  minimal (a sparsest solution).*

But,  $\|\cdot\|_0$  is not convex and this optimization problem is difficult, so instead we consider:

## Problem (Basis Pursuit)

*Given  $A \in \text{Mat}_{m,n}(\mathbb{R})$ ,  $b \in \mathbb{R}^m$ , find a solution  $x \in \mathbb{R}^n$  for  $Ax = b$  with  $\|x\|_1$  minimal.*

Basis Pursuit can be solved by linear programming and is generally fast. Under reasonable assumptions, Basis Pursuit also gives the sparsest solutions (Candès-Tao, Donoho).

# The Nuclear and Spectral Norms

Low rank tensors are sparse in some sense, and by convex relaxation:

## Definition

The nuclear norm  $\|T\|_*$  is the smallest value of  $\sum_{i=1}^r \|v_i\|$  where  $T = \sum_{i=1}^r v_i$  and  $v_1, \dots, v_r$  are pure tensors. (well-defined)

# The Nuclear and Spectral Norms

Low rank tensors are sparse in some sense, and by convex relaxation:

## Definition

The nuclear norm  $\|T\|_*$  is the smallest value of  $\sum_{i=1}^r \|v_i\|$  where  $T = \sum_{i=1}^r v_i$  and  $v_1, \dots, v_r$  are pure tensors. (well-defined)

$V^{(1)}, \dots, V^{(d)}, V$  a spaces with a positive definite bilinear/hermitian form  $\langle \cdot, \cdot \rangle$

# The Nuclear and Spectral Norms

Low rank tensors are sparse in some sense, and by convex relaxation:

## Definition

The nuclear norm  $\|T\|_*$  is the smallest value of  $\sum_{i=1}^r \|v_i\|$  where  $T = \sum_{i=1}^r v_i$  and  $v_1, \dots, v_r$  are pure tensors. (well-defined)

$V^{(1)}, \dots, V^{(d)}, V$  a spaces with a positive definite bilinear/hermitian form  $\langle \cdot, \cdot \rangle$

## Definition

The spectral norm is defined by

$$\|T\|_\sigma = \max\{|\langle T, v \rangle| \mid v \text{ pure tensor with } \|v\| = 1\}.$$

The spectral norm is *dual* to the nuclear norm, in particular

$$|\langle T, S \rangle| \leq \|T\|_* \|S\|_\sigma \quad \text{for all tensors } S, T.$$

# Example: Permanent Tensor

$$P_n = \sum_{\sigma \in \mathcal{S}_n} e_{\sigma(1)} \otimes e_{\sigma(2)} \otimes \cdots \otimes e_{\sigma(n)} \in \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n.$$

$$\|P_n\|_{\sigma} = \max\{|\text{perm}(v_1 v_2 \cdots v_n)| \mid \|v_1\| = \cdots = \|v_n\| = 1\}$$

# Example: Permanent Tensor

$$P_n = \sum_{\sigma \in S_n} e_{\sigma(1)} \otimes e_{\sigma(2)} \otimes \cdots \otimes e_{\sigma(n)} \in \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n.$$
$$\|P_n\|_{\sigma} = \max\{|\text{perm}(v_1 v_2 \cdots v_n)| \mid \|v_1\| = \cdots = \|v_n\| = 1\}$$

Theorem (Carlen, Lieb and Moss, 2006)

$$\max\{\text{perm}(v_1 v_2 \cdots v_n) \mid \|v_1\| = \cdots = \|v_n\| = 1\} = n!/n^{n/2}$$

$$\|P_n\|_{\star} = \frac{n^{n/2}}{n!} \|P_n\|_{\star} \|P_n\|_{\sigma} \geq \frac{n^{n/2}}{n!} \langle P_n, P_n \rangle = n^{n/2}.$$

# Example: Permanent Tensor

Theorem (Glynn 2010)

$$P_n = \frac{1}{2^{n-1}} \sum_{\delta} \left( \prod_{i=1}^n \delta_i \right) (\sum_{i=1}^n \delta_i \mathbf{e}_i) \otimes \cdots \otimes (\sum_{i=1}^n \delta_i \mathbf{e}_i)$$

where  $\delta$  runs over  $\{1\} \times \{-1, 1\}^{n-1}$ .

# Example: Permanent Tensor

Theorem (Glynn 2010)

$$P_n = \frac{1}{2^{n-1}} \sum_{\delta} \left( \prod_{i=1}^n \delta_i \right) (\sum_{i=1}^n \delta_i \mathbf{e}_i) \otimes \cdots \otimes (\sum_{i=1}^n \delta_i \mathbf{e}_i)$$

where  $\delta$  runs over  $\{1\} \times \{-1, 1\}^{n-1}$ .

In particular,  $\binom{n}{\lfloor n/2 \rfloor} \leq \text{rank}(P_n) \leq 2^{n-1}$  and  $\|P_n\|_{\star} \leq n^{n/2}$ , so

Theorem (D.)

$$\|P_n\|_{\star} = n^{n/2}$$



# The Convex Decomposition (CoDe) Model

## Problem

*Given a tensor  $T \in V$ , write  $T = v_1 + v_2 + \cdots + v_r$  where  $v_1, v_2, \dots, v_r$  are pure unit tensors and  $\|v_1\| + \|v_2\| + \cdots + \|v_r\|$  is minimal.*

Well-defined but the decomposition may not be unique.

# The Convex Decomposition (CoDe) Model

## Problem

*Given a tensor  $T \in V$ , write  $T = v_1 + v_2 + \cdots + v_r$  where  $v_1, v_2, \dots, v_r$  are pure unit tensors and  $\|v_1\| + \|v_2\| + \cdots + \|v_r\|$  is minimal.*

Well-defined but the decomposition may not be unique.

We can also allow an error term ...

# The Convex Decomposition (CoDe) Model

## Problem

*Given a tensor  $T \in V$  and fixed  $x \geq 0$ , write  $T = v_1 + v_2 + \cdots + v_r + E$  where  $r$  is a nonnegative integer,  $v_1, v_2, \dots, v_r$  are pure tensors,  $\|v_1\| + \|v_2\| + \cdots + \|v_r\| \leq x$  and  $\|E\|$  is minimal.*

The decomposition is not unique.

If  $A = v_1 + v_2 + \cdots + v_r$  then  $A$  is the unique tensor with  $\|A\|_* \leq x$  for which  $\|T - A\|$  is minimal.

# The Convex Decomposition (CoDe) Model

## Problem

Given a tensor  $T \in V$  and fixed  $x \geq 0$ , write  $T = v_1 + v_2 + \cdots + v_r + E$  where  $r$  is a nonnegative integer,  $v_1, v_2, \dots, v_r$  are pure tensors,  $\|v_1\| + \|v_2\| + \cdots + \|v_r\| \leq x$  and  $\|E\|$  is minimal.

The decomposition is not unique.

If  $A = v_1 + v_2 + \cdots + v_r$  then  $A$  is the unique tensor with  $\|A\|_* \leq x$  for which  $\|T - A\|$  is minimal.

## Problem

Given a tensor  $T \in V$  and fixed  $y \geq 0$ , write  $T = v_1 + v_2 + \cdots + v_r + E$  where  $r$  is a nonnegative integer,  $v_1, v_2, \dots, v_r$  are pure tensors,  $\|E\| \leq y$  and  $\|v_1\| + \|v_2\| + \cdots + \|v_r\|$  is minimal.

## Definition

Pure tensors  $v_1, v_2, \dots, v_r$  are  $t$ -orthogonal if

$$\sum_{i=1}^r |\langle v_i, w \rangle|^{2/t} \leq 1$$

for every pure tensor  $w$  with  $\|w\|_2 = 1$ .

## Definition

Pure tensors  $v_1, v_2, \dots, v_r$  are  $t$ -orthogonal if

$$\sum_{i=1}^r |\langle v_i, w \rangle|^{2/t} \leq 1$$

for every pure tensor  $w$  with  $\|w\|_2 = 1$ .

1-orthogonal  $\Leftrightarrow$  orthogonal

If  $t > s$  then  $t$ -orthogonal  $\Rightarrow$   $s$ -orthogonal.

## Definition

Pure tensors  $v_1, v_2, \dots, v_r$  are  $t$ -orthogonal if

$$\sum_{i=1}^r |\langle v_i, w \rangle|^{2/t} \leq 1$$

for every pure tensor  $w$  with  $\|w\|_2 = 1$ .

1-orthogonal  $\Leftrightarrow$  orthogonal

If  $t > s$  then  $t$ -orthogonal  $\Rightarrow$   $s$ -orthogonal.

## Theorem (D.)

If  $v_1, \dots, v_r \in V$  are  $t$ -orthogonal, then  $r \leq \dim(V)^{1/t}$ .

# Horizontal and Vertical Tensor Product

Theorem (“horizontal tensor product”, D.)

*If  $v_1, \dots, v_r$  are  $t$ -orthogonal, and  $w_1, \dots, w_r$  are  $s$ -orthogonal, then  $v_1 \otimes w_1, \dots, v_r \otimes w_r$  are  $(s + t)$ -orthogonal.*



# Horizontal and Vertical Tensor Product

Theorem (“horizontal tensor product”, D.)

*If  $v_1, \dots, v_r$  are  $t$ -orthogonal, and  $w_1, \dots, w_r$  are  $s$ -orthogonal, then  $v_1 \otimes w_1, \dots, v_r \otimes w_r$  are  $(s + t)$ -orthogonal.*

If  $V = V^{(1)} \otimes \dots \otimes V^{(d)}$  and  $W = W^{(1)} \otimes \dots \otimes W^{(d)}$ , then

$$V \boxtimes W := (V^{(1)} \otimes W^{(1)}) \otimes \dots \otimes (V^{(d)} \otimes W^{(d)}).$$

$$(v_1 \otimes \dots \otimes v_d) \boxtimes (w_1 \otimes \dots \otimes w_d) = (v_1 \otimes w_1) \otimes (v_2 \otimes w_2) \otimes \dots \otimes (v_d \otimes w_d)$$

Theorem (“vertical tensor product”, D.)

*If  $v_1, v_2, \dots, v_r \in V$  and  $w_1, \dots, w_s \in W$  are  $t$ -orthogonal, then  $\{v_i \boxtimes w_j \mid 1 \leq i \leq r, 1 \leq j \leq s\}$  are  $t$ -orthogonal.*

# The Diagonal Singular Value Decomposition

## Definition

Suppose that  $(\star) : T = \sum_{i=1}^r \lambda_i v_i$  such that  $\lambda_1 \geq \dots \geq \lambda_r > 0$  and  $v_1, \dots, v_r$  are 2-orthogonal pure tensors of unit length, then  $(\star)$  is called a *diagonal singular value decomposition* of  $T$  (DSVD).

If  $d = 2$  (tensor product of 2 spaces) then the DSVD is the usual singular value decomposition. For  $d > 2$ , the DSVD is different from the *Higher Order Singular Value Decomposition* defined by De Lathauer, De Moor, and Vandewalle. Not every tensor has a DSVD.

# The Diagonal Singular Value Decomposition

## Theorem (D.)

If  $T$  has a DSVD then

$$\|T\|_{\star} = \sum_i \lambda_i, \quad \|T\|_2 = \sqrt{\sum_i \lambda_i^2}, \quad \|T\|_{\sigma} = \lambda_1$$

# The Diagonal Singular Value Decomposition

## Theorem (D.)

If  $T$  has a DSVD then

$$\|T\|_{\star} = \sum_i \lambda_i, \quad \|T\|_2 = \sqrt{\sum_i \lambda_i^2}, \quad \|T\|_{\sigma} = \lambda_1$$

## Theorem (D.)

If  $\lambda_1 > \lambda_2 > \dots > \lambda_r$  then the DSVD is unique.

## Theorem (D.)

If  $v_1, \dots, v_r$  are  $t$ -orthogonal with  $t > 2$ , then the DSVD is unique.

# Example: Matrix Multiplication Tensor

$e_1, \dots, e_n \in \mathbb{C}^n$  are orthogonal

$e_1 \otimes e_1, \dots, e_n \otimes e_n \in \mathbb{C}^n \otimes \mathbb{C}^n$  are 2-orthogonal

# Example: Matrix Multiplication Tensor

$e_1, \dots, e_n \in \mathbb{C}^n$  are orthogonal

$e_1 \otimes e_1, \dots, e_n \otimes e_n \in \mathbb{C}^n \otimes \mathbb{C}^n$  are 2-orthogonal

$e_1 \otimes e_1 \otimes 1, \dots, e_n \otimes e_n \otimes e_1 \in \mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}$  are 2-orthogonal

$e_1 \otimes 1 \otimes e_1, \dots, e_n \otimes 1 \otimes e_n \in \mathbb{C}^n \otimes \mathbb{C} \otimes \mathbb{C}^n$  are 2-orthogonal

$1 \otimes e_1 \otimes e_1, \dots, 1 \otimes e_n \otimes e_n \in \mathbb{C} \otimes \mathbb{C}^n \otimes \mathbb{C}^n$  are 2-orthogonal

## Example: Matrix Multiplication Tensor

$e_1, \dots, e_n \in \mathbb{C}^n$  are orthogonal

$e_1 \otimes e_1, \dots, e_n \otimes e_n \in \mathbb{C}^n \otimes \mathbb{C}^n$  are 2-orthogonal

$e_1 \otimes e_1 \otimes 1, \dots, e_n \otimes e_n \otimes e_1 \in \mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}$  are 2-orthogonal

$e_1 \otimes 1 \otimes e_1, \dots, e_n \otimes 1 \otimes e_n \in \mathbb{C}^n \otimes \mathbb{C} \otimes \mathbb{C}^n$  are 2-orthogonal

$1 \otimes e_1 \otimes e_1, \dots, 1 \otimes e_n \otimes e_n \in \mathbb{C} \otimes \mathbb{C}^n \otimes \mathbb{C}^n$  are 2-orthogonal

Using vertical tensor product, we get

$$\{(e_i \otimes e_j) \otimes (e_j \otimes e_k) \otimes (e_k \otimes e_i) \mid 1 \leq i, j, k \leq n\}$$

are 2-orthogonal.

# Example: Matrix Multiplication Tensor

## Theorem (D.)

*The matrix multiplication tensor*

$$T_n = \sum_{i,j,k=1}^n e_{i,j} \otimes e_{j,k} \otimes e_{k,i}$$

*is a DSVD.*

The singular values of  $T_n$  are

$$\underbrace{1, 1, \dots, 1}_{n^3}$$

In particular,

$$\|T_n\|_* = \sum_{i=1}^{n^3} 1 = n^3.$$



# Example: Group Algebra Multiplication Tensor

$G$  is a group with  $n$  elements and  $\mathbb{C}G \cong \mathbb{C}^n$  is the group algebra

$$T_G = \sum_{g,h \in G} g \otimes h \otimes h^{-1}g^{-1}.$$

DFT case corresponds to  $G = \mathbb{Z}/n\mathbb{Z}$ .

# Example: Group Algebra Multiplication Tensor

$G$  is a group with  $n$  elements and  $\mathbb{C}G \cong \mathbb{C}^n$  is the group algebra

$$T_G = \sum_{g, h \in G} g \otimes h \otimes h^{-1}g^{-1}.$$

DFT case corresponds to  $G = \mathbb{Z}/n\mathbb{Z}$ .

## Theorem (D.)

$T_G$  has a DSVD and its singular values are

$$\underbrace{\sqrt{\frac{n}{d_1}}, \dots, \sqrt{\frac{n}{d_1}}}_{d_1^3}, \dots, \underbrace{\sqrt{\frac{n}{d_s}}, \dots, \sqrt{\frac{n}{d_s}}}_{d_s^3}$$

where  $d_1, d_2, \dots, d_s$  are the dimension of the irreducible representations of  $G$ .

Suppose that  $P_n$  has a DSVD with singular values  $\lambda_1, \dots, \lambda_r$ .

$$\|P_n\|_{\star} = n^{n/2} = \sum_{i=1}^r \lambda_i$$

$$\|P_n\|_{\sigma} = \frac{n!}{n^{n/2}} = \lambda_1$$

$$\|P_n\|^2 = n! = \sum_{i=1}^r \lambda_i^2$$

$$\lambda_1 \sum_{i=1}^r \lambda_i = n! = \sum_{i=1}^r \lambda_i^2$$

so  $\lambda_1 = \dots = \lambda_r$ , and  $r = \frac{r\lambda_1}{\lambda_1} = \frac{\|D_n\|_{\star}}{\|D_n\|_{\sigma}} = \frac{n^n}{n!}$ .

If  $n > 2$  then  $r$  is not an integer!

# Noisy Signal Model

Suppose that  $V$  is a finite dimensional  $\mathbb{R}$ -vector space of signals. If  $c \in V$  is a noisy signal, then there is a decomposition

$$c = a + b$$

where  $a \in V$  is a sparse original signal, and  $b \in V$  is additive noise.

Using convex relaxation, sparsity is measured by some  $\ell_1$ -type norm  $\|\cdot\|_X$ . The noise is measured by some  $\ell_2$ -type norm  $\|\cdot\|_Y$ .

# Noisy Signal Model

Suppose that  $V$  is a finite dimensional  $\mathbb{R}$ -vector space of signals. If  $c \in V$  is a noisy signal, then there is a decomposition

$$c = a + b$$

where  $a \in V$  is a sparse original signal, and  $b \in V$  is additive noise.

Using convex relaxation, sparsity is measured by some  $\ell_1$ -type norm  $\|\cdot\|_X$ . The noise is measured by some  $\ell_2$ -type norm  $\|\cdot\|_Y$ .

For example,  $c$  is a tensor in  $V = V^{(1)} \otimes \dots \otimes V^{(d)}$ ,  $\|\cdot\|_X = \|\cdot\|_*$  is the nuclear norm, and  $\|\cdot\|_Y = \|\cdot\|_2 = \|\cdot\|$  is the usual  $\ell_2$ -norm. (see the CoDe Model)

# The Pareto Frontier

$V$  finite dimensional  $\mathbb{R}$ -vector space

$\|\cdot\|_X, \|\cdot\|_Y$  norms on  $V$

$c \in V$

## Definition

A pair  $(x, y) \in \mathbb{R}^2$  is Pareto-efficient if there exists a decomposition  $c = a + b$  with  $\|a\| = x, \|b\| = y$  and for every decomposition  $c = a' + b'$  we have  $\|a'\|_X > x, \|b'\|_Y > y$  or  $(\|a'\|_X, \|b'\|_Y) = (x, y)$ . We call  $c = a + b$  an  $XY$ -decomposition. The Pareto-frontier consists of all Pareto-efficient pairs.

# The Pareto Frontier

$V$  finite dimensional  $\mathbb{R}$ -vector space

$\|\cdot\|_X, \|\cdot\|_Y$  norms on  $V$

$c \in V$

## Definition

A pair  $(x, y) \in \mathbb{R}^2$  is Pareto-efficient if there exists a decomposition  $c = a + b$  with  $\|a\| = x, \|b\| = y$  and for every decomposition  $c = a' + b'$  we have  $\|a'\|_X > x, \|b'\|_Y > y$  or  $(\|a'\|_X, \|b'\|_Y) = (x, y)$ . We call  $c = a + b$  an  $XY$ -decomposition. The Pareto-frontier consists of all Pareto-efficient pairs.

## Lemma

*The Pareto-frontier is the graph of a strictly decreasing, convex, homeomorphism  $f_{YX}^c : [0, \|c\|_X] \rightarrow [0, \|c\|_Y]$ .*

# The Pareto Sub-Frontier

$\langle \cdot, \cdot \rangle$  positive definite bilinear form

$\|v\|_2 := \|v\| = \sqrt{\langle v, v \rangle}$   $\ell_2$ -norm

$\|\cdot\|_X, \|\cdot\|_Y$  dual to each others



# The Pareto Sub-Frontier

$\langle \cdot, \cdot \rangle$  positive definite bilinear form

$\|v\|_2 := \|v\| = \sqrt{\langle v, v \rangle}$   $\ell_2$ -norm

$\|\cdot\|_X, \|\cdot\|_Y$  dual to each others

## Theorem

*Equivalent:*

1.  $c = a + b$  is an  $X^2$ -decomposition
2.  $c = a + b$  is an  $2Y$ -decomposition
3.  $\langle a, b \rangle = \|a\|_X \|b\|_Y$ .

# The Pareto Sub-Frontier

$\langle \cdot, \cdot \rangle$  positive definite bilinear form

$\|v\|_2 := \|v\| = \sqrt{\langle v, v \rangle}$   $\ell_2$ -norm

$\|\cdot\|_X, \|\cdot\|_Y$  dual to each others

## Theorem

*Equivalent:*

1.  $c = a + b$  is an  $X^2$ -decomposition
2.  $c = a + b$  is an  $2Y$ -decomposition
3.  $\langle a, b \rangle = \|a\|_X \|b\|_Y$ .

## Definition

The Pareto sub-frontier consists of all pairs  $(x, y)$  such that there exists an  $X^2$ -decomposition  $c = a + b$  with  $\|a\|_X = x$  and  $\|b\|_Y = y$ .

# The Pareto Sub-Frontier

## Lemma

*The Pareto sub-frontier is the graph of a strictly decreasing homeomorphism  $h_{YX}^c : [0, \|c\|_X] \rightarrow [0, \|c\|_Y]$ .*

Clearly,  $f_{YX}^c \leq h_{YX}^c$ .

## Definition

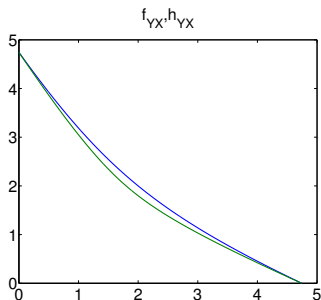
A vector  $c \in V$  is called tight if  $f_{YX}^c = h_{YX}^c$ . If all vectors in  $V$  are tight then the norms  $\|\cdot\|_X$  and  $\|\cdot\|_Y$  are called tight.

# (Sub)-Pareto Example

$$c = (3 \ 3)^t \in V = \mathbb{R}^2$$

$$\left\| \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right\|_X = \sqrt{\frac{1}{2}z_1^2 + 2z_2^2}, \quad \left\| \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right\|_Y = \sqrt{2z_1^2 + \frac{1}{2}z_2^2}.$$

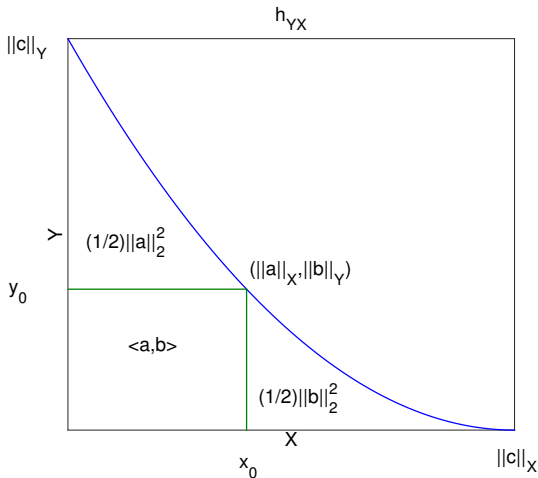
dual norms



So  $c$  is not tight.

# The Area Below the Pareto Sub-Frontier

Suppose  $c = a + b$  is an  $X^2$ -decomposition with  $\|a\|_X = x_0$  and  $\|b\|_Y = y_0$ . The area under the Pareto sub-frontier is  $\frac{1}{2}\|c\|^2 = \frac{1}{2}\|a\|^2 + \langle a, b \rangle + \frac{1}{2}\|b\|^2$ .



# The Area Below the Pareto Sub-Frontier

The sub-Pareto frontier applies to:

- ▶ denoising a sparse 1D signal:  $\|\cdot\|_1$  vs.  $\|\cdot\|_\infty$  (tight!)
- ▶ denoising of a matrix by soft thresholding:  $\|\cdot\|_\star$  vs.  $\|\cdot\|_\sigma$  (tight!)
- ▶ denoising tensors:  $\|\cdot\|_\star$  vs.  $\|\cdot\|_\sigma$
- ▶ compressive sensing (Basis Pursuit Denoising, LASSO, Dantzig Selector)
- ▶ Fatemi-Osher-Rudin image denoising: total variation norm vs. its dual
- ▶ 1D total variation denoising: Taut String Method

# Example: Singular Value Decomposition

$$\mathbb{R}^{n \times m} = \mathbb{R}^n \otimes \mathbb{R}^m, \quad r = \min\{n, m\}$$

$A \in \mathbb{R}^{n \times m}$  with singular values  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq 0$

$$\|A\|_X = \|A\|_* = \lambda_1 + \lambda_2 + \dots + \lambda_r$$

$$\|A\|_Y = \|A\|_\sigma = \lambda_1$$

$$\|A\| = \sqrt{\lambda_1^2 + \lambda_2^2 + \dots + \lambda_r^2}$$

# Example: Singular Value Decomposition

$$\mathbb{R}^{n \times m} = \mathbb{R}^n \otimes \mathbb{R}^m, \quad r = \min\{n, m\}$$

$A \in \mathbb{R}^{n \times m}$  with singular values  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq 0$

$$\|A\|_X = \|A\|_* = \lambda_1 + \lambda_2 + \dots + \lambda_r$$

$$\|A\|_Y = \|A\|_\sigma = \lambda_1$$

$$\|A\| = \sqrt{\lambda_1^2 + \lambda_2^2 + \dots + \lambda_r^2}$$

## Lemma

*The norms  $\|\cdot\|_*$  and  $\|\cdot\|_\sigma$  are tight.*

$XY$ -decompositions are related to soft thresholding



# Example: Singular Value Decomposition

(note that the Pareto and sub-Pareto frontiers are the same here)

## Lemma

*If  $A$  has singular values  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq 0$  then the Pareto (sub-)frontier is the piecewise linear function through the points:*

$$(\lambda_1 + \lambda_2 + \dots + \lambda_k - k\lambda_{k+1}, \lambda_{k+1}), \quad k = 0, 1, 2, \dots, r$$

*(with the convention  $\lambda_{r+1} = 0$ ).*

# Example: Singular Value Decomposition

(note that the Pareto and sub-Pareto frontiers are the same here)

## Lemma

*If  $A$  has singular values  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq 0$  then the Pareto (sub-)frontier is the piecewise linear function through the points:*

$$(\lambda_1 + \lambda_2 + \dots + \lambda_k - k\lambda_{k+1}, \lambda_{k+1}), \quad k = 0, 1, 2, \dots, r$$

*(with the convention  $\lambda_{r+1} = 0$ ).*

## Corollary

*The singular values are uniquely determined by the (sub-)Pareto curve, and vice versa.*

## Definition

$c = c_1 + c_2 + \dots + c_s$  is called a slope decomposition if  $\langle c_i, c_j \rangle = \|c_i\|_X \|c_j\|_Y$  for all  $i \leq j$  and

$$\frac{\|c_1\|_X}{\|c_1\|_Y} < \frac{\|c_2\|_X}{\|c_2\|_Y} < \dots < \frac{\|c_r\|_X}{\|c_r\|_Y}.$$

# Slope decomposition

## Definition

$c = c_1 + c_2 + \dots + c_s$  is called a slope decomposition if  $\langle c_i, c_j \rangle = \|c_i\|_X \|c_j\|_Y$  for all  $i \leq j$  and

$$\frac{\|c_1\|_X}{\|c_1\|_Y} < \frac{\|c_2\|_X}{\|c_2\|_Y} < \dots < \frac{\|c_r\|_X}{\|c_r\|_Y}.$$

## Theorem

$c \in V$  has a slope decomposition if and only if  $c$  is tight.

# Slope decomposition

## Definition

$c = c_1 + c_2 + \dots + c_s$  is called a slope decomposition if  $\langle c_i, c_j \rangle = \|c_i\|_X \|c_j\|_Y$  for all  $i \leq j$  and

$$\frac{\|c_1\|_X}{\|c_1\|_Y} < \frac{\|c_2\|_X}{\|c_2\|_Y} < \dots < \frac{\|c_r\|_X}{\|c_r\|_Y}.$$

## Theorem

$c \in V$  has a slope decomposition if and only if  $c$  is tight.

## Theorem

The slope decomposition is unique when it exists.

# Singular values for tight vectors

Suppose  $c \in V$  is tight, with slope decomposition  
 $c = c_1 + c_2 + \cdots + c_s$ .

## Definition

Let  $\mu_i = \|c_i\|_Y$ , and  $\lambda_i = \mu_i + \mu_{i+1} + \cdots + \mu_s$  for all  $i$ . Then  $c$  has singular value  $\lambda_i$  with multiplicity  $\frac{\|c_i\|_X}{\|c_i\|_Y} - \frac{\|c_{i-1}\|_X}{\|c_{i-1}\|_Y}$ .

Multiplicities and singular values are positive, but not necessarily integers.

# Singular values for tight vectors

Suppose  $c \in V$  is tight, with slope decomposition  
 $c = c_1 + c_2 + \cdots + c_s$ .

## Definition

Let  $\mu_i = \|c_i\|_Y$ , and  $\lambda_i = \mu_i + \mu_{i+1} + \cdots + \mu_s$  for all  $i$ . Then  $c$  has singular value  $\lambda_i$  with multiplicity  $\frac{\|c_i\|_X}{\|c_i\|_Y} - \frac{\|c_{i-1}\|_X}{\|c_{i-1}\|_Y}$ .

Multiplicities and singular values are positive, but not necessarily integers.

## Theorem

*If  $c$  has singular values  $\lambda_1, \dots, \lambda_s$  with multiplicities  $m_1, \dots, m_s$  respectively, then  $\|c\|_X = \sum_i m_i \lambda_i$ ,  $\|c\|_Y = \lambda_1$  and  $\|T\| = \sqrt{\sum_i m_i \lambda_i^2}$ .*

# Singular Values for Tensors

Norms  $\|\cdot\|_X = \|\cdot\|_*$ ,  $\|\cdot\|_Y = \|\cdot\|_\sigma$

## Theorem

*If  $T = \sum_{i=1}^r \lambda_i v_i$  is a diagonal singular value decomposition, then the singular values of  $T$  are  $\lambda_1, \dots, \lambda_r$ .*



# Singular Values for Tensors

Norms  $\|\cdot\|_X = \|\cdot\|_*$ ,  $\|\cdot\|_Y = \|\cdot\|_\sigma$

## Theorem

*If  $T = \sum_{i=1}^r \lambda_i v_i$  is a diagonal singular value decomposition, then the singular values of  $T$  are  $\lambda_1, \dots, \lambda_r$ .*

The permanent tensor  $P_n$  is tight, but has no DSVD.  
It has singular value  $\frac{n!}{n^{n/2}}$  with multiplicity  $\frac{n^n}{n!}$ .

Thanks

Thank You!