SIAM J. APPL. ALGEBRA GEOMETRY
Vol. 2, No. 4, pp. 535–596

A General Theory of Singular Values with Applications to Signal Denoising*

Harm Derksen†

Abstract. We study the Pareto frontier for two competing norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ on a vector space. For a given vector c, the Pareto frontier describes the possible values of $(\|a\|_X, \|b\|_Y)$ for a decomposition $c = a + b$. The singular value decomposition of a matrix is closely related to the Pareto frontier for the spectral and nuclear norm. We will develop a general theory that extends the notion of singular values of a matrix to arbitrary finite dimensional Euclidean vector spaces equipped with dual norms. This also generalizes the *diagonal singular value decompositions* (DSVDs) for tensors introduced by the author in previous work. We can apply the results to denoising, where c is a noisy signal, a is a sparse signal, and b is noise. Applications include 1D total variation denoising, 2D total variation Rudin–Osher–Fatemi image denoising, LASSO, basis pursuit denoising, and tensor decompositions.

Key words. singular values, tensor decomposition, signal processing, image processing, taut string, denoising

AMS subject classifications. 15A18, 15A69, 90C25

DOI. 10.1137/17M1156149

1. Introduction.

Sound, images, and videos can be corrupted by noise. Noise removal is a fundamental problem in signal and image processing. In the additive noise model, we have an original signal a, additive noise b, and a corrupted signal $c = a + b$. We will work with discrete signals and view a, b, and c as vectors or arrays. One of our goals is to formulate the problem of noise removal in a general framework of competing norms on a vector space using the Pareto frontier. The Pareto frontier defines the optimal trade-off between the two norms. The Pareto frontier was used in the L-curve method in Tikhonov regularization (see [44, 50, 27, 28] and [36, Chapter 26]), and in basis pursuit denoising (see [6, 48, 26]) to find optimal regularization parameters. The Pareto frontier is a continuous convex curve and has a continuous derivative if one of the norms is the Euclidean norm (see [6], Lemma 3.2, and Proposition 4.5).

For noise removal to be possible, the original signal and the additive noise should be qualitatively different. We will assume that the original signal a is *sparse*. A vector is sparse when it has few nonzero values. We will also consider other notions of sparseness. For example, a piecewise constant function on an interval can be considered sparse because its derivative has few nonzero values (or values where it is not defined), and a piecewise linear function can be considered sparse because the second derivative has few nonzero values. A sound signal from music is sparse because it contains only a few frequencies. A typical image is sparse because it has large connected areas of the same color; i.e., the image is piecewise constant.

*Received by the editors November 9, 2017; accepted for publication (in revised form) August 10, 2018; published electronically October 23, 2018.
http://www.siam.org/journals/siaga/2-4/M115614.html
Funding: The author was partially supported by the National Science Foundation, grant DMS 1601229, and the U.S. Department of Defense, grant W81XWH-17-2-0012.
†Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 USA (hderksen@umich.edu).

This is exploited in the total variation image denoising method of Rudin, Osher, and Fatemi [46]. A matrix of low rank can also be viewed as sparse because it is the sum of a few rank 1 matrices. In this context, principal component analysis can be viewed as a method for recovering a sparse signal. The noise signal b, on the other hand, is not sparse. For example, white noise contains all frequencies in the sound spectrum. Gaussian additive noise in an image will be completely discontinuous and not locally constant at all.

There are many ways to measure sparseness. Examples of sparseness measures are the ℓ_0 "norm" (which actually is not a norm), the rank of a matrix, and the number of different frequencies in a sound signal. It is difficult to use these measures because they are not convex. We deal with this using *convex relaxation*, i.e., we replace the nonconvex sparseness measure by a convex one. In this paper, we will measure sparseness using a norm $\|\cdot\|_X$ on the vector space of all signals. These norms coming from convex relaxation are typically ℓ_1-type norms. For example, we may replace the ℓ_0 "norm" by the ℓ_1 norm, or replace the rank of a matrix by the nuclear norm. Noise will be measured using a different norm $\|\cdot\|_Y$. This typically will be a Euclidean ℓ_2 norm or perhaps an ℓ_∞-type norm. The quotient $\|a\|_Y/\|a\|_X$ is large for the sparse signal a, and $\|b\|_Y/\|b\|_X$ is small for the noise signal b. To denoise a signal c we search for a decomposition $c = a + b$ where $\|a\|_X$ and $\|b\|_Y$ are small. Minimizing $\|a\|_X$ and minimizing $\|b\|_Y$ are two competing objectives. The trade-off between these two objectives is governed by the Pareto frontier. The concept of Pareto efficiency was used by Vilfredo Pareto (1848–1923) to describe economic efficiency. A point $(x, y) \in \mathbb{R}^2$ is called Pareto efficient if there exists a decomposition $c = a + b$ with $\|a\|_X = x$ and $\|b\|_Y = y$ such that for every decomposition $c = a' + b'$ we have $\|a'\|_X > x$, $\|b'\|_Y > y$, or $(\|a'\|_X, \|b'\|_Y) = (x, y)$. If (x, y) is Pareto efficient, then we will call the decomposition $c = a + b$ an XY-decomposition. Many methods such as LASSO, basis pursuit denoising, the Dantzig selector, total variation denoising, and principal component analysis can be formulated as finding an XY-decomposition for certain norms $\|\cdot\|_X$ and $\|\cdot\|_Y$.

Most of the theory developed in this paper is concerned with the particularly interesting situation where the space of signals has a positive definite inner product and the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ are dual to each other. In this context, we introduce new notions such as the *Pareto subfrontier*, the *slope decomposition*, and the *singular value region*. The inner product gives a Euclidean norm defined by $\|c\|_2 = \sqrt{\langle c, c\rangle}$. We now have three distinct norms. Using duality, we will show that $X2$-decompositions and $2Y$-decompositions are the same. We define the *Pareto subfrontier* as the set of all points $(\|a\|_X, \|b\|_Y)$ where $c = a+b$ is an $X2$-decomposition (or, equivalently, a $2Y$-decomposition). The Pareto subfrontier lies on or above the Pareto frontier (by the definition of the Pareto frontier). A vector c is called *tight* if its Pareto frontier and Pareto subfrontier coincide. If every vector is tight, then the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ are called *tight*. We show that for tight vectors, the Pareto (sub)frontier is piecewise linear. For a tight vector c, we will define the slope decomposition of c which can be thought of as a generalization of the singular value decomposition.

The singular value decomposition of a matrix will be a guiding example. Our objective is to generalize the singular value decomposition of matrices to any vector space with competing dual norms on a vector space. Let us briefly explain how the singular value decomposition relates to matrix norms. The vector space $V = \mathbb{C}^{m\times n}$ of complex $m \times n$-matrices can be identified with $\mathbb{R}^{2mn}$. This space has an inner product $\langle\cdot,\cdot\rangle$ defined by

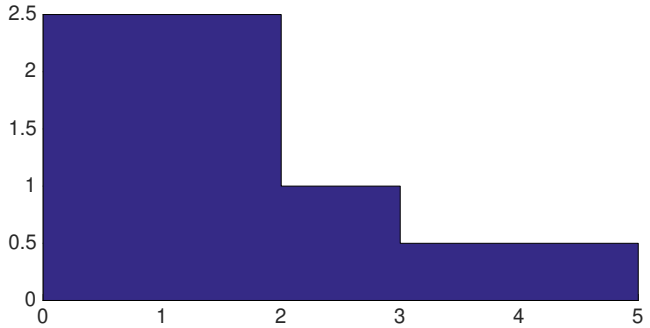

Figure 1.

$\langle A, B\rangle = \Re(\text{trace}(A^\star B)) = \Re(\text{trace}(B^\star A))$, where $A^\star$ denotes the conjugate transpose of A and $\Re(\cdot)$ denotes the real part. The Euclidean norm (or Frobenius norm) on V is given by $\|A\|_2 = \sqrt{\langle A, A\rangle} = \sqrt{\text{trace}(A^\star A)}$ (also often denoted by $\|A\|_F$). We have two more norms on V, namely the nuclear norm $\|\cdot\|_\star$ and the spectral norm (or operator norm) $\|\cdot\|_\sigma$. The matrix $A^\star A$ is nonnegative definite Hermitian and has a unique positive nonnegative square root $\sqrt{A^\star A}$, and the nuclear norm is defined by $\|A\|_\star = \text{trace}(\sqrt{A^\star A})$. The spectral norm is the operator norm and is given by $\|A\|_\sigma = \max\{\|Av\|_2 \mid v \in \mathbb{C}^n \text{ and } \|v\|_2 = 1\}$. A matrix $A \in V$ has a singular value decomposition $A = U_1 D U_2^\star$ where U_1 and U_2 are unitary matrices, and D is a diagonal matrix with nonnegative real entries. If a real number λ appears m times on the diagonal of D, then we say that λ is a singular value of A with multiplicity m. If the singular values of a matrix A are $\lambda_1 > \lambda_2 > \cdots > \lambda_r$ with multiplicities $m_1, m_2, \ldots, m_r$, respectively, we have the following well-known formulas for the spectral, nuclear, and Euclidean norms:

$$\|A\|_\sigma = \lambda_1, \tag{1}$$

$$\|A\|_\star = m_1\lambda_1 + m_2\lambda_2 + \cdots + m_r\lambda_r, \tag{2}$$

$$\|A\|_2 = \sqrt{m_1\lambda_1^2 + m_2\lambda_2^2 + \cdots + m_r\lambda_r^2}. \tag{3}$$

The nuclear norm and spectral norm of a matrix are dual to each other (see, for example, [12, 45] and Lemma 7.1), and we will show that these norms are tight in section 7. This implies that we can define the Pareto frontier and Pareto subfrontier of a matrix, and that these are the same.

The singular values of a matrix can be graphically represented by the *singular value region*. The singular value region of A is a bar of height λ_1 and width m_1, followed by a bar of height λ_2 and width m_2, etc. For example, if a matrix A has eigenvalues 2.5 with multiplicity 2, 1 with multiplicity 1, and 0.5 with multiplicity 2, then the singular value region is shown in Figure 1. The height of the singular value region is the spectral norm, the width is the rank, the area is the nuclear norm, and if we integrate $2y$ over the region, we obtain the square of the Frobenius norm $\|A\|_2^2$. The Pareto frontier of a matrix (which is also its Pareto subfrontier) encodes the singular values of the matrix, and the slope decomposition is closely related to the singular value decomposition. The singular value region can be computed from the Pareto (sub)frontier. This allows us to generalize the singular value region to any vector space with

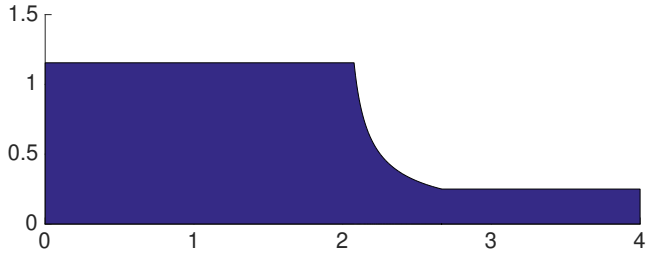

Figure 2.

two competing norms.

We will study the application of our theory to tensors in more detail. Using the general setup, we can define the singular value region for arbitrary tensors, and for some tensors we can also define a slope decomposition. The slope decomposition is a generalization of the singular value decomposition and is related to tensor decompositions. Let $V = \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2} \otimes \cdots \otimes \mathbb{R}^{n_d}$ be the tensor product of d vector spaces. Elements of V are called d-way tensors and can be viewed as d-dimensional arrays of size $n_1 \times n_2 \times \cdots \times n_d$. Let e_j be the jth basis vector of $\mathbb{R}^{n_k}$. The tensor $e_{j_1} \otimes e_{j_2} \otimes \cdots \otimes e_{j_d}$ is the array with a 1 in position $(j_1, j_2, \ldots, j_d)$ and 0 everywhere else. Such tensors form a basis of the tensor product space V. For $d \geq 3$, one can generalize the rank of a matrix to tensor rank [31]. The tensor rank is closely related to the canonical polyadic decomposition of a tensor (CP-decomposition). This decomposition is also known as the PARAFAC [29] or the CANDECOMP model [8]. The nuclear norm of a matrix can be generalized to the nuclear norm of a tensor [25, 47], and this can be viewed as a convex relaxation of the tensor rank. The spectral norm of a matrix can be generalized to a spectral norm of a tensor, and this norm is dual to the nuclear norm of a tensor [19]. Not every tensor is tight. A tensor that is tight will have a slope decomposition which generalizes the diagonal singular value decomposition introduced by the author in [19]. Every tensor that has a diagonal singular value has a slope decomposition, but the converse is not true. We will define singular values and multiplicities for tight tensors such that the formulas (1), (2), (3) are satisfied. The multiplicities of the singular values of tensors are nonnegative, but are not always integers. For example, in section 13 we will show that the tensor

$$e_1 \otimes e_2 \otimes e_3 + e_1 \otimes e_3 \otimes e_2 + e_2 \otimes e_1 \otimes e_3 + e_2 \otimes e_3 \otimes e_1 + e_3 \otimes e_1 \otimes e_2 + e_3 \otimes e_2 \otimes e_1 \in \mathbb{R}^{3\times 3\times 3}$$

is tight and has the singular value $\frac{2}{\sqrt{3}}$ with multiplicity $\frac{9}{2}$ (and not singular value 1 with multiplicity 6 as one might expect). For tensors that are not tight we can still define the singular value region, but the singular value interpretation may be more esoteric. For example, we will show that the tensor

$$e_2 \otimes e_1 \otimes e_1 + e_1 \otimes e_2 \otimes e_1 + e_1 \otimes e_1 \otimes e_2 \in \mathbb{R}^{2\times 2\times 2}$$

has the singular value region shown in Figure 2.

We will define the singular value region in a very general context. Whenever V is a finite dimensional Euclidean vector space, and $\|\cdot\|_X$ and $\|\cdot\|_Y$ are norms that are dual to each other, we can define the singular value region for any $c \in V$.

Notation.

$\langle \cdot, \cdot \rangle$ positive definite bilinear form, page 540
$\alpha^c_{YX}(x)$ solution curve to $\mathbf{M}^c_{YX}(x)$, page 550
$\|A\|_\star$ nuclear norm of a matrix or tensor, page 537
$\|A\|_\sigma$ spectral norm of a matrix or tensor, page 537
$B_X(x)$ ball of radius x for norm $\|\cdot\|_X$, page 544
$\det_n$ determinant tensor, page 589
$\mathcal{F}_X(x)$ smallest face of B_X containing $\alpha_{2X}(x)/x$, page 562
f^c_{YX} Pareto frontier, page 539
$\mathrm{gsparse}_X(\cdot)$ geometric sparseness, page 545
h^c_{YX} Pareto sub-frontier, page 541
$\mu_{XY}(c)$ slope $\|c\|_Y/\|c\|_X$, page 543
$\mathbf{M}^c_{YX}(x)$ minimize $\|c-a\|_Y$ subject to $\|a\|_X \le x$, page 548
$\|\cdot\|_p$ ℓ_p norm, page 540
perm_n permanent tensor, page 589
$\mathrm{proj}_X(c,x)$ projection of c onto the ball $B_X(x)$, page 544
$\Re(\cdot)$ real part, page 537
$\mathrm{shrink}_X(c,x)$ shrinkage of c by x, page 545
$\mathrm{sparse}_X(\cdot)$ sparseness measure related to $\|\cdot\|_X$, page 545
$\mathbf{TS}^c(\varepsilon)$ Taut String problem, page 579
$u \star v$ concatenation of two functions, page 551
$\|\cdot\|_X, \|\cdot\|_Y$ norms, page 535

2. Main results.

2.1. The Pareto frontier. Let us consider a finite dimensional $\mathbb{R}$-vector space V equipped with two norms, $\|\cdot\|_X$ and $\|\cdot\|_Y$. Suppose that $c \in V$. We are looking for decompositions $c = a + b$ that are optimal in the sense that we cannot reduce $\|a\|_X$ without increasing $\|b\|_Y$ and we cannot reduce $\|b\|_Y$ without increasing $\|a\|_X$. We recall the definition from the introduction.

Definition 2.1. *A pair $(x,y) \in \mathbb{R}^2_{\ge 0}$ is called* Pareto efficient *if there exists a decomposition $c = a + b$ with $\|a\|_X = x$, $\|b\|_Y = y$ such that for every decomposition $c = a' + b'$ we have $\|a'\|_X > x$, $\|b'\|_Y > y$, or $(\|a'\|_X, \|b'\|_Y) = (x,y)$. If (x,y) is a Pareto efficient pair, then we call $c = a + b$ an XY-*decomposition.

By symmetry, $c = a + b$ is an XY-decomposition if and only if $c = b + a$ is a YX-decomposition. The *Pareto frontier* consists of all Pareto efficient pairs (see [6]). The Pareto frontier is the graph of a strictly decreasing, continuous convex function

$$f^c_{YX} : [0, \|c\|_X] \to [0, \|c\|_Y]$$

(see [6] and Lemmas 3.2 and 3.3). If we change the role of X and Y, we get the graph of f^c_{XY}, so f^c_{XY} and f^c_{YX} are inverse functions of each other.

Example 2.2. Consider the vector space $V = \mathbb{R}^n$. Sparseness of vectors in $\mathbb{R}^n$ can be

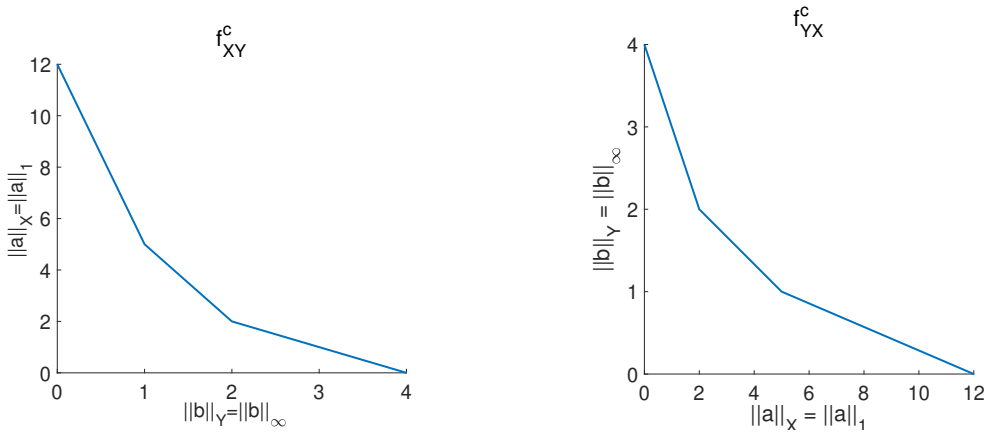

Figure 3.

measured by the number of nonzero entries. For $c \in V$ we define

$$\|c\|_0 = |\{i \mid 1 \leq i \leq n,\ c_i \neq 0\}| = \lim_{p \to 0} \|c\|_p^p.$$

Note that $\|\cdot\|_0$ is not a norm on V because it does not satisfy $\|\lambda c\|_0 = |\lambda|\|c\|_0$ for $\lambda \in \mathbb{R}$. Convex relaxation of $\|\cdot\|_0$ gives us the ℓ_1 norm $\|\cdot\|_1$. This means that the unit ball for the norm $\|\cdot\|_1$ is the convex hull of all vectors c with $\|c\|_0 = \|c\|_2 = 1$. Let us take $\|\cdot\|_X = \|\cdot\|_1$ and $\|\cdot\|_Y = \|\cdot\|_\infty$ and describe the Pareto frontier. Suppose that $c = (c_1 \cdots c_n)^t \in \mathbb{R}^n$ and $0 \leq y \leq \|c\|_\infty$. If $\|b\|_\infty = y$, then we have

$$\|c - b\|_1 = \textstyle\sum_{i=1}^n |c_i - b_i| \geq \sum_{i=1}^n \max\{|c_i| - y, 0\}.$$

If we take $b_i = \operatorname{sgn}(c_i)\min\{|c_i|, y\}$ for all i, then we have equality. So $c = a + b$ is an XY-decomposition where $a := c - b$. This shows that

$$f^c_{XY}(y) = \|a\|_1 = \textstyle\sum_{i=1}^n \max\{0, |c_i| - y\}.$$

For the vector $c = (-1, 2, 4, 1, -2, 1, -1)^t$ we plotted $f^c_{XY} = f^c_{1\infty}$ and $f^c_{YX} = f^c_{\infty 1}$ (see also Example 2.11) in Figure 3.

For example, if we take $y = \frac{3}{2}$, then we get the decomposition with $a = (0, \frac{1}{2}, \frac{5}{2}, 0, -\frac{3}{2}, 0, 0)^t$ and $b = (-1, \frac{3}{2}, \frac{3}{2}, 1, -\frac{3}{2}, 1, -1)^t$ and we have $\|a\|_1 = \frac{7}{2}$ and $\|a\|_0 = 3$. The vector a is sparser than c. This procedure of noise reduction is *soft thresholding* in its simplest form.

In section 3, we will study the Pareto frontier and XY-decompositions in more detail.

2.2. Dual norms and the Pareto subfrontier. We now assume that we have a positive definite bilinear form $\langle \cdot, \cdot \rangle$ on the finite dimensional vector space V. The Euclidean ℓ_2 norm on V is defined by $\|v\|_2 := \sqrt{\langle v, v \rangle}$. Suppose that $\|\cdot\|_X$ is another norm on V. We may think of $\|\cdot\|_X$ as a norm that measures sparseness. For denoising, we compare the norms $\|\cdot\|_X$ and $\|\cdot\|_2$. We also consider the *dual norm* on V defined by

$$\|v\|_Y = \max\{\langle v, w \rangle \mid w \in V,\ \|w\|_X = 1\}.$$

The dual norm of $\|\cdot\|_Y$ is $\|\cdot\|_X$ again. There is an interesting interplay between the three norms, and the $X2$-decompositions, $2Y$-decompositions, and XY-decompositions are closely connected. The following proposition and other results in this section will be proved in section 4.

Proposition 2.3. *For a vector $c \in V$, the following three statements are equivalent:*
(1) *$c = a + b$ is an $X2$-decomposition;*
(2) *$c = a + b$ is a $2Y$-decomposition;*
(3) *$c = a + b$ and $\langle a, b\rangle = \|a\|_X \|b\|_Y$.*

Definition 2.4. *If the statements* (1)–(3) *in Proposition 2.3 imply*
(4) *$c = a + b$ is an XY-decomposition,*
then c is called tight. *If all vectors in V are tight, then the norm $\|\cdot\|_X$ is called* tight.

Definition 2.5. *The* Pareto subfrontier *is the set of all pairs $(x, y) \in \mathbb{R}^2_{\geq 0}$ such that there exists an $X2$-decomposition $c = a + b$ with $\|a\|_X = x$ and $\|b\|_Y = y$.*

We will show in section 4 that the Pareto subfrontier is the graph of a decreasing Lipschitz continuous function

$$h^c_{YX} : [0, \|c\|_X] \to [0, \|c\|_Y].$$

By symmetry, h^c_{XY} is the inverse function of h^c_{YX}. From the definitions it is clear that $f^c_{YX}(x) \leq h^c_{YX}(x)$. If $f^c_{XY} = h^c_{XY}$, then every $X2$-decomposition is automatically an XY-decomposition, and c is tight.

Corollary 2.6. *A vector $c \in V$ is tight if and only if $f^c_{XY} = h^c_{XY}$.*

If $f^c_{XY} = h^c_{XY}$, then there is no space between the two graphs, and they fit together *tightly*, which explains the name of this property. Let us work out an example where the norms are not tight.

Example 2.7. Define a norm $\|\cdot\|_X$ on $\mathbb{R}^2$ by

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

Its dual norm is given by

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

For $t \in [\frac{1}{2}, 2]$, the decomposition

$$c = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = a(t) + b(t)$$

is an $X2$-decomposition, where

$$a(t) = \begin{pmatrix} 4 - 2t \\ 2t^{-1} - 1 \end{pmatrix} \text{ and } b(t) = \begin{pmatrix} 2t - 1 \\ 4 - 2t^{-1} \end{pmatrix}.$$

To verify this, we compute

$$\|a(t)\|_X = \sqrt{2}(2t^{-1} - 1)\sqrt{t^2 + 1}, \quad \|b(t)\|_Y = \sqrt{2}(2 - t^{-1})\sqrt{t^2 + 1},$$

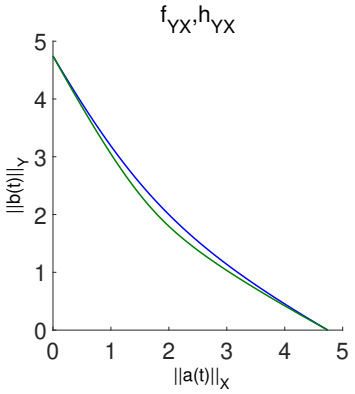

Figure 4.

and

$$\langle a(t), b(t)\rangle = (4-2t)(2t-1)+(2t^{-1}-1)(4-2t^{-1}) = 2(2t^{-1}-1)(2-t^{-1})(t^2+1) = \|a(t)\|_X\|b(t)\|_Y.$$

The Pareto subfrontier is parameterized by

$$(\sqrt{2}(2t^{-1}-1)\sqrt{t^2+1}, \sqrt{2}(2-t^{-1})\sqrt{t^2+1})), \quad t \in [\tfrac{1}{2}, 2].$$

The XY-decompositions of c are $c = \widetilde{a}(t) + \widetilde{b}(t)$ for $t \in [\frac{1}{4}, 4]$, where

$$\widetilde{a}(t) = \frac{1}{5}\begin{pmatrix} 16-4t \\ 4t^{-1}-1 \end{pmatrix} \quad \text{and} \quad \widetilde{b}(t) = \frac{1}{5}\begin{pmatrix} 4t-1 \\ 16-4t^{-1} \end{pmatrix}.$$

The Pareto frontier is parameterized by

$$(\tfrac{\sqrt{2}}{5}(4t^{-1}-1)\sqrt{4t^2+1}, \tfrac{\sqrt{2}}{5}(4-t^{-1})\sqrt{t^2+4})).$$

In Figure 4, we plotted the Pareto frontier in green and the Pareto subfrontier in blue. The Pareto frontier and subfrontier are not the same, so the vector c is not tight.

The Pareto subfrontier encodes crucial information about the $X2$-decompositions. If (x_0, y_0) is a point on the Pareto subfrontier and $c = a + b$ is an $X2$-decomposition with $\|a\|_X = x_0$ and $\|b\|_Y = y_0$, then we can read off $\|c\|_X$, $\|c\|_Y$, $\|c\|_2$, $\|a\|_X$, $\|a\|_2$, $\|b\|_Y$, $\|b\|_2$, and $\langle a, b\rangle$ from the Pareto subfrontier using the following proposition.

Proposition 2.8. *Suppose that (x_0, y_0) is a point on the XY-Pareto subfrontier of $c \in V$, and let $c = a + b$ be an $X2$-decomposition with $\|a\|_X = x_0$ and $\|b\|_Y = y_0$. (See Figure 5.)*

(1) *The area below the subfrontier is equal to $\frac{1}{2}\|c\|_2^2$.*
(2) *The area below the subfrontier and to the right of $x = x_0$ is equal to $\frac{1}{2}\|b\|_2^2$.*
(3) *The area below the subfrontier and above $y = y_0$ is equal to $\frac{1}{2}\|a\|_2^2$.*
(4) *The area of the rectangle $[0, x_0] \times [0, y_0]$ is $\langle a, b\rangle$.*

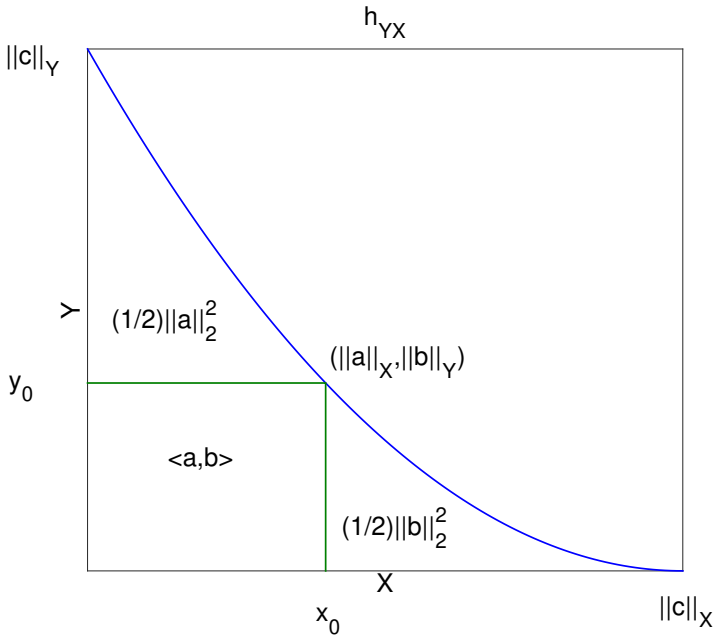

Figure 5.

2.3. The slope decomposition. Proofs of results in this subsection will be given in section 6. We define the slope of a nonzero vector $c \in V$ as the ratio

$$\mu_{XY}(c) := \frac{\|c\|_Y}{\|c\|_X}.$$

If $\|\cdot\|_X$ is a norm that is small for sparse signals, then μ_{XY} is large for sparse signals and small for Gaussian noise. Note that $\mu_{YX}(c) = (\mu_{XY}(c))^{-1}$. Using the slope function, we define the slope decomposition. If $\|\cdot\|_X$ is the nuclear norm for matrices, then the slope decomposition is closely related to the singular value decomposition. So we can think of the slope decomposition as a generalization of the singular value decomposition.

Definition 2.9. *An expression $c = c_1 + c_2 + \cdots + c_r$ is called an XY-*slope decomposition *if $c_1, \ldots, c_r$ are nonzero, $\langle c_i, c_j \rangle = \|c_i\|_X \|c_j\|_Y$ for all $i \le j$, and $\mu_{XY}(c_1) > \mu_{XY}(c_2) > \cdots > \mu_{XY}(c_r)$.*

Note that because of symmetry, $c = c_1 + c_2 + \cdots + c_r$ is an XY-slope decomposition if and only if $c = c_r + c_{r-1} + \cdots + c_1$ is a YX-slope decomposition. There may be vectors that do not have a slope decomposition. We will prove the following result.

Theorem 2.10.

(1) *A vector $c \in V$ is tight if and only if c has a slope decomposition.*
(2) *Suppose that $c = c_1 + c_2 + \cdots + c_r$ is a slope decomposition, and let $x_i = \|c_i\|_X$ and $y_i = \|c_i\|_Y$ for all i. Then $\|c\|_X = \sum_{i=1}^r x_i$, $\|c\|_Y = \sum_{i=1}^r y_i$, and the Pareto frontier (which is the same as the Pareto subfrontier) is the piecewise linear curve through the*

points

$$(x_1 + \cdots + x_i, y_{i+1} + \cdots + y_r), \quad i = 0, 1, 2, \ldots, r.$$

Example 2.11. Let us go back to Example 2.2. The norms $\|\cdot\|_X = \|\cdot\|_1$ and $\|\cdot\|_Y = \|\cdot\|_\infty$ are dual to each other. We will show that these norms are tight. If we integrate the function

$$f^c_{XY}(y) = \sum_{i=1}^{n} \max\{0, |c_i| - y\}$$

from 0 to $\|c\|_Y = \max\{|c_1|, |c_2|, \ldots, |c_n|\}$, we get $\frac{1}{2}\sum_{i=1}^n c_i^2 = \frac{1}{2}\|c\|_2^2$, which is the area under the graph of h^c_{XY}. So the areas under the graphs of f^c_{XY} and h^c_{XY} are the same, and we deduce that $h^c_{XY} = f^c_{XY}$. This shows that c is tight. Since c is arbitrary, the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ are tight By Theorem 2.10 above, every vector has a slope decomposition. For example,

$$c = \begin{pmatrix} -1 \\ 2 \\ 4 \\ 1 \\ -2 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 2 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} = c_1 + c_2 + c_3$$

is a slope decomposition. Let $x_i = \|c_i\|_1$ and $y_i = \|c_i\|_Y$. Then we have $x_1 = 2$, $x_2 = 3$, $x_3 = 7$, $y_1 = 2$, $y_2 = 1$, $y_3 = 1$. The Pareto curve f^c_{YX} is the piecewise linear function going through

$$(0, 2+1+1) = (0, 4), (2, 1+1) = (2, 2), (2+3, 1) = (5, 1), (2+3+7, 0) = (12, 0).$$

2.4. Geometry of the unit ball. For $x \geq 0$ we define the X-ball of radius x by

$$B_X(x) = \{v \in V \mid \|v\|_X \leq x\}.$$

We explain denoising in terms of the geometry of the X-balls. Suppose we want to denoise a signal c, such that the denoised signal a is sparse. We impose the constraint $\|a\|_X \leq x$. Under this constraint, we minimize the amount of noise by minimizing the ℓ_2 norm of $b := c - a$. This means that the a is the vector inside the ball $B_X(x)$ that is closest to the vector c. We call a the projection of c onto the ball $B_X(x)$ and write $a = \mathrm{proj}_X(c, x)$. The function $\mathrm{proj}_X(\cdot, x)$ is a retraction of $\mathbb{R}^n$ onto the ball $B_X(x)$. If $x_1 \leq x_2$, then it is clear that

$$\mathrm{proj}_X(\mathrm{proj}_X(c, x_1), x_2) = \mathrm{proj}_X(c, x_1).$$

For $x_1 > x_2$ one might expect that $\mathrm{proj}_X(\mathrm{proj}_X(c, x_1), x_2) = \mathrm{proj}_X(c, x_2)$. This is not always true, but it is true in the case where c is tight by Proposition 2.12 below.

We also define a shrinkage operator by

$$\mathrm{shrink}_X(c, x) = c - \mathrm{proj}_X(c, x).$$

If $c = a + b$ is an $X2$-decomposition, $\|a\|_X = x$, and $\|b\|_Y = y$, then we have

$$a = \operatorname{proj}_X(c, x) = \operatorname{shrink}_Y(c, y)$$

and

$$b = \operatorname{proj}_Y(c, y) = \operatorname{shrink}_X(c, x).$$

The function $\operatorname{shrink}_Y(\cdot, y)$ can be seen as a denoising function where y is the noise level.

A nice property of tight vectors is the transitivity of denoising.

Proposition 2.12. *If c is tight, then we have*
(1) $\operatorname{proj}_X(\operatorname{proj}_X(c, x_1), x_2) = \operatorname{proj}_X(c, \min\{x_1, x_2\})$ *and*
(2) $\operatorname{shrink}_X(\operatorname{shrink}_X(c, x_1), x_2) = \operatorname{shrink}_X(c, x_1 + x_2)$.

The unit ball $B_X = B_X(1)$ is a closed convex set but not always a polytope. We recall the definition of a face of a closed convex set.

Definition 2.13. *A face of a closed convex set $B \subseteq V$ is a convex closed subset $F \subseteq B$ with the following property: if $a, b \in B$ and t is any real number with $0 < t < 1$ such that $0 < t < 1$ and $ta + (1 - t)b \in F$, then we must have $a, b \in F$.*

More generally, if F is a face of B and a point of F is a convex combination of finitely many points from B, then all those points must lie in F (see Lemma 6.3). We will study faces of the unit ball $B_X = B_X(1)$ and the cones associated to it.

Definition 2.14. *A* facial X-cone *is a cone of the form $\mathbb{R}_{\geq 0}F$ where $F \subset B_X$ is a (proper) face of the unit ball B_X. The set $\{0\}$ is considered a facial X-cone as well.*

If C is a nonzero facial X-cone, then $C = \mathbb{R}_{\geq 0}F$ for some proper face F of the unit ball B_X. In that case, we have $F = C \cap \partial B_X$, where $\partial B_X = \{c \in V \mid \|c\|_X = 1\}$ is the unit sphere. We now will discuss two notions of sparseness related to a norm $\|\cdot\|_X$.

Definition 2.15. *For a nonzero vector c we define its* X-sparseness $\operatorname{sparse}_X(c)$ *as the smallest nonnegative integer r such that we can write*

$$c = c_1 + c_2 + \cdots + c_r,$$

where $c_i/\|c_i\|_X$ is an extremal point of the unit ball B_X for $i = 1, 2, \ldots, r$. The geometric X-sparseness $\operatorname{gsparse}_X(c)$ *is $\dim C$, where C is the smallest facial X-cone containing c.*

Each notion of sparseness has its merits. We have

$$\operatorname{sparse}_X(a + b) \leq \operatorname{sparse}_X(a) + \operatorname{sparse}_X(b),$$

but a similar inequality does not always hold for geometric sparseness. On the other hand, the set

$$\{c \mid \operatorname{gsparse}_X(c) \leq k\}$$

of k-geometric X-sparse vectors is closed, but the set

$$\{c \mid \operatorname{sparse}_X(c) \leq k\}$$

of k-X-sparse vectors is not always closed.

We have

$$\operatorname{sparse}_X(c) \leq \operatorname{gsparse}_X(c)$$

by the Carathéodory theorem (see [4, Theorem 2.3]).

Example 2.16. Consider $\mathbb{R}^n$, and let $\|\cdot\|_X = \|\cdot\|_1$ and $\|\cdot\|_Y = \|\cdot\|_\infty$. We have

$$\operatorname{sparse}_X(c) = \operatorname{gsparse}_X(c) = \|c\|_0,$$

where

$$\|c\|_0 = \#\{i \mid c_i \neq 0\} = \lim_{p\to 0} \|c\|_p^p.$$

The function $\|\cdot\|_0$ is sometimes referred to as the ℓ_0 norm but is strictly speaking not a norm.

$$\operatorname{gsparse}_Y(c) = 1 + \#\big\{i \,\big|\, |c_i| \neq \max\{|c_1|, \ldots, |c_n|\}\big\}.$$

We have

$$\begin{pmatrix}2\\0\\0\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix} + \begin{pmatrix}1\\-1\\-1\end{pmatrix}$$

but

$$\operatorname{gsparse}_Y \begin{pmatrix}2\\0\\0\end{pmatrix} = 3 > 1+1 = \operatorname{gsparse}_Y \begin{pmatrix}1\\1\\1\end{pmatrix} + \operatorname{gsparse}_Y \begin{pmatrix}1\\-1\\-1\end{pmatrix}.$$

Example 2.17. If $\|\cdot\|_X = \|\cdot\|_\star$ is the nuclear norm on a space $V = \mathbb{R}^{m\times n}$ of $m \times n$ matrices, and $c \in V$ is a matrix, then $\operatorname{sparse}_X(c) = \operatorname{gsparse}_X(c) = \operatorname{rank}(c)$ is the rank of c.

Example 2.18. Let $\|\cdot\|_X$ be the nuclear norm on the tensor product space $V = \mathbb{R}^{2\times 2\times 2} = \mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2$. The value $\operatorname{sparse}_X(c)$ is the tensor rank of $c \in V$. It is known that the set of all tensors of rank ≤ 2 is not closed. If $\{e_1, e_2\}$ is the standard orthogonal basis of $\mathbb{R}^2$, then the tensor

$$c = e_2 \otimes e_1 \otimes e_1 + e_1 \otimes e_2 \otimes e_1 + e_1 \otimes e_1 \otimes e_2$$

has rank 3 but is a limit of tensors of rank 2. In this case, the geometric sparseness and sparseness are not the same. For example, if

$$d = (e_1 + (0.1)e_2) \otimes (e_1 + (0.1)e_2) \otimes (e_1 + (0.1)e_2) - e_1 \otimes e_1 \otimes e_1,$$

then $\operatorname{sparse}_X(d) = 2$ is the tensor rank, but $\operatorname{gsparse}_X(d) > 2$.

Theorem 2.19. *Suppose that $c = a + b$ is an $X2$-decomposition. Then we have*

$$\operatorname{gsparse}_X(a) + \operatorname{gsparse}_Y(b) \leq n + 1,$$

where $n = \dim V$. Moreover, if c is tight, then we have

$$\operatorname{gsparse}_X(a) \leq \operatorname{gsparse}_X(c) \text{ and } \operatorname{gsparse}_Y(b) \leq \operatorname{gsparse}_Y(c).$$

Since $\operatorname{gsparse}_X \geq \operatorname{sparse}_X$, the theorem also implies that

$$\operatorname{sparse}_X(a) + \operatorname{sparse}_Y(b) \leq n + 1.$$

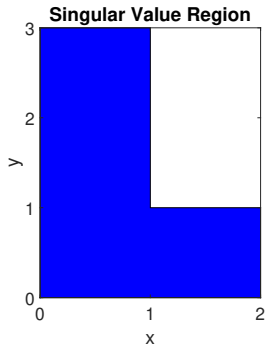

Figure 6.

2.5. The singular value region. We can generalize the notion of the singular value region to an arbitrary finite dimensional vector space V with dual norms $\|\cdot\|_X$ and $\|\cdot\|_Y$. Let $y = h^c_{YX}(x)$ be the Pareto subfrontier of $c \in V$. For the definition of the singular value region, we view x as a function of y which gives $x = h^c_{XY}(y)$. The function $h^c_{XY}(y)$ is Lipschitz and decreasing and is differentiable almost everywhere.

Definition 2.20. *The* singular value region *is the region in the first quadrant to the left of the graph of* $x = -\frac{d}{dy}h^c_{XY}(y)$.

If the graph of $-\frac{d}{dy}h^c_{XY}(y)$ is a step function, then we can interpret the region as singular values with multiplicities similarly as in the case of matrices.

Example 2.21. Suppose that the Pareto subfrontier for a vector $c \in V$ is given by

$$h^c_{YX}(x) = \begin{cases} 3 - x & \text{if } 0 \le x \le 2, \\ 1 - \frac{1}{2} & \text{if } 2 \le x \le 4. \end{cases}$$

The inverse is h^c_{XY} and is given by

$$h^c_{YX}(y) = \begin{cases} 4 - 2y & \text{if } 0 \le y \le 1, \\ 3 - y & \text{if } 1 \le y \le 3. \end{cases}$$

The function $x = -\frac{d}{dy}h^c_{XY}(y)$ is equal to 2 for $0 \le y \le 1$ and 1 for $1 \le y \le 3$. The singular value region is shown in Figure 6. We see that c has singular value 3 with multiplicity 1 and singular value 1 with multiplicity 1.

If a vector $c \in V$ has a slope decomposition, then we can easily find the singular values and multiplicities from Theorem 2.10. Recall that $\mu_{XY}(c) = \|c\|_Y/\|c\|_X$ and $\mu_{YX}(c) = \|c\|_X/\|c\|_Y$.

Corollary 2.22. *If*

$$c = c_1 + c_2 + \cdots + c_r$$

is an XY*-slope decomposition, then the singular values are*

$$\|c_1\|_Y, \|c_2\|_Y, \ldots, \|c_r\|_Y$$

with multiplicities

$$\mu_{YX}(c_1), \mu_{YX}(c_2) - \mu_{YX}(c_1), \ldots, \mu_{YX}(c_r) - \mu_{YX}(c_{r-1}),$$

respectively.

As we will see in section 7, this notion of the singular value region is indeed a generalization of the singular value region defined for matrices in the introduction.

3. The Pareto curve.

3.1. Optimization problems. We will formulate signal denoising as an optimization problem. Suppose that V is an n-dimensional $\mathbb{R}$-vector space equipped with two norms, $\|\cdot\|_X$ and $\|\cdot\|_Y$. Let $c \in V$ be a fixed vector. Given a noisy signal c, we would like to find a decomposition $c = a + b$ where both $\|a\|_X$ and $\|b\|_Y$ are small. We can do this by minimizing $\|a\|_X$ under the constraint $\|b\|_Y \leq y$ for some fixed $y \geq 0$, or by minimizing $\|b\|_Y$ under the constraint $\|a\|_X \leq x$ for some x with $0 \leq x \leq \|c\|_X$. We formally define the following optimization problem.

Problem $\mathbf{M}^c_{YX}(x)$. *Minimize $\|c - a\|_Y$ for $a \in V$ under the constraint $\|a\|_X \leq x$.*

This has a solution because the function $a \mapsto \|c - a\|_Y$ is continuous, and the domain $\{a \in \mathbb{R}^n \mid \|a\|_X \leq x\}$ is compact. Let $f^c_{YX}(x)$ be the smallest value of $\|c - a\|_Y$. As we will see later in Lemma 3.3, the graph of $f^c_{YX}(x)$ consists of all Pareto efficient pairs (x, y), and we call f^c_{YX} the *Pareto curve* or *Pareto frontier*. Since c will be fixed most of the time, we may just write $\mathbf{M}_{YX}(x)$ and $f_{YX}(x)$ instead of $\mathbf{M}^c_{YX}(x)$ and $f^c_{YX}(x)$. We will prove some properties of the Pareto curve.

Lemma 3.1. *If $0 \leq x \leq \|c\|_X$ and a is a solution to $\mathbf{M}^c_{YX}(x)$, then $\|a\|_X = x$.*

Proof. Suppose that a is a solution to $\mathbf{M}_{YX}(x)$ and $\|a\|_X < x$. Note that $\|a\|_X < x \leq \|c\|_X$, so $c \neq a$ and $\|c - a\|_Y > 0$. Choose $\varepsilon > 0$ such that $\|a\|_X + \varepsilon\|c\|_X \leq x$, and define $a' = (1 - \varepsilon)a + \varepsilon c$. Then we have $\|a'\|_X \leq (1 - \varepsilon)\|a\|_X + \varepsilon\|c\|_X \leq x$ and

$$\|c - a'\|_Y = \|(1 - \varepsilon)(c - a)\|_Y = (1 - \varepsilon)\|c - a\|_Y < \|c - a\|_Y.$$

Contradiction. ∎

3.2. Properties of the Pareto curve. The following lemma gives some basic properties of the Pareto curve. In various contexts, these properties are already known. We formulate the properties for the problem of two arbitrary competing norms on a finite dimensional vector space.

Lemma 3.2. *The function f_{YX} is a convex, strictly decreasing, continuous function on the interval $[0, \|c\|_X]$.*

Proof. Suppose that $0 \leq x_1 < x_2 \leq \|c\|_X$. There exist $a_1, a_2 \in V$ with $\|a_i\|_X = x_i$ and $\|c - a_i\|_Y = f_{YX}(x_i)$ for $i = 1, 2$.

Let $a = (1 - t)a_1 + ta_2$ for some $t \in [0, 1]$. Then we have $\|a\|_X = \|(1 - t)a_1 + ta_2\|_X \leq (1 - t)x_1 + tx_2$, so by definition we have

$$\begin{aligned}(1 - t)f_{YX}(x_1) + tf_{YX}(x_2) &= (1 - t)\|c - a_1\|_Y + t\|c - a_2\|_Y \\ &\geq \|(1 - t)(c - a_1) + t(c - a_2)\|_Y = \|c - a\|_Y \geq f_{YX}((1 - t)x_1 + tx_2).\end{aligned}$$

This proves that f is convex.

For some $t \in (0,1]$ we can write $x_2 = (1-t)x_1 + t\|c\|_X$. We have

$$f_{YX}(x_2) \le (1-t)f_{YX}(x_1) + tf_{YX}(\|c\|_X) = (1-t)f_{YX}(x_1) < f_{YX}(x_1).$$

This shows that f is strictly decreasing.

Since f_{YX} is convex, it is continuous on $(0, \|c\|_X)$. Because f is decreasing, it is continuous at $x = \|c\|_X$. We will show that f is also continuous at $x = 0$. Suppose that $x_1, x_2, \dots$ is a sequence in $[0, \|c\|_X]$ with $\lim_{n\to\infty} x_n = 0$. Choose $a_n \in V$ such that $\|a_n\|_X = x_n$ and $\|c - a_n\|_Y = f_{YX}(x_n)$. Since $\lim_{n\to\infty} \|a_n\|_X = \lim_{n\to\infty} x_n = 0$, we have that $\lim_{n\to\infty} a_n = 0$. It follows that $\lim_{n\to\infty} f_{YX}(x_n) = \lim_{n\to\infty} \|c - a_n\|_Y = \|c\|_Y = f_{YX}(0)$ because $\|\cdot\|_Y$ is a continuous function. ∎

We show now that the graph of f_{YX} consists of all Pareto efficient pairs.

Lemma 3.3. *Suppose that $c \in V$.*

(1) *$c = a + b$ is an XY-decomposition if and only if a is a solution to $\mathbf{M}^c_{YX}(\|a\|_X)$.*

(2) *The pair (x, y) is Pareto efficient if and only if $0 \le x \le \|c\|_X$ and $y = f^c_{YX}(x)$.*

Proof. (1) Suppose that $c = a + b$. If $c = a + b$ is an XY-decomposition, then it is clear from the definitions that a is a solution to $\mathbf{M}^c_{YX}(\|a\|_X)$.

On the other hand, suppose that a is a solution to $\mathbf{M}^c_{YX}(\|a\|_X)$ and $c = a' + b'$. Let $x = \|a\|_X$ and $y = \|b\|_Y$. Assume that $\|a'\|_X \le \|a\|_X = x$. Then we have $\|b'\|_Y = \|c - a'\|_Y \ge \|c - a\|_Y = \|b\|_Y = y$. If $\|b'\|_Y = y$, then a' is also a solution to $\mathbf{M}^c_{YX}(x)$ and $\|a'\|_X = x$ by Lemma 3.1. This shows that $c = a + b$ is an XY-decomposition.

(2) Suppose that (x, y) is Pareto efficient. Assume that $\|c\|_X < x$. From the decomposition $c = c + 0$ follows that $0 = \|0\|_Y > y$. Contradiction. This shows that $0 \le x \le \|c\|_X$. There exists a decomposition $c = a + b$ with $\|a\|_X = x$ and $\|c - a\|_Y = \|b\|_Y = y$. Because $\|a\|_X \le x$, we have $y = \|c - a\|_Y \ge f^c_{YX}(x)$. Suppose that a' is a solution of $\mathbf{M}^c_{XY}(x)$. Then $\|a'\|_X = x$ by Lemma 3.1. Because (x, y) is Pareto efficient, we have $f^c_{YX}(x) = \|c - a'\|_Y \ge y$. We conclude that $f^c_{YX}(x) = y$.

Conversely, suppose that $0 \le x \le \|c\|_X$ and $y = f^c_{YX}(x)$. Let a be a solution of $\mathbf{M}^c_{YX}(x)$ and $b := c - a$. Suppose that $c = a' + b'$ is another decomposition. If $\|a'\|_X < x$, then we have $\|b'\|_Y = \|c - a'\|_Y \ge f^c_{YX}(\|a'\|_X) > f^c_{YX}(x) = y$. If $\|a'\|_X = x$, then we have $\|b'\|_Y = \|c - a'\|_Y \ge f^c_{YX}(\|a'\|_X) = f^c_{YX}(x) = y$. We conclude that (x, y) is Pareto efficient. ∎

Corollary 3.4.

(1) *The function $f_{YX} : [0, \|c\|_X] \to [0, \|c\|_Y]$ is a homeomorphism and its inverse is f_{XY}.*

(2) *A vector a is a solution to $\mathbf{M}_{YX}(x)$ if and only if $c - a$ is a solution to $\mathbf{M}_{XY}(f_{YX}(x))$.*

Proof. (1) If $0 \le x \le \|c\|_X$ and $0 \le y \le \|c\|_Y$, then we have

$$f_{YX}(x) = y \Leftrightarrow (x, y) \text{ is Pareto efficient} \Leftrightarrow f_{XY}(y) = x.$$

So f_{XY} and f_{YX} are inverse of each other. Since both functions are continuous, the functions are homeomorphisms.

If $y = f_{XY}(x)$ and $b = c - a$, then we have

$$a \text{ is a solution to } \mathbf{M}_{YX}(x) \Leftrightarrow c = a + b \text{ is an } XY\text{-decomposition} \Leftrightarrow$$
$$\Leftrightarrow c - a = b \text{ is a solution to } \mathbf{M}_{XY}(y) = \mathbf{M}_{XY}(f_{YX}(x)). \quad \blacksquare$$

3.3. Rigid norms. The problem $\mathbf{M}^c_{YX}(x)$ does not always have a unique solution.

Definition 3.5. *We say that $c \in V$ is* rigid *if $\mathbf{M}^c_{YX}(x)$ has a* unique *solution for all $x \in [0, \|c\|_X]$.*

Let us give an example of a vector that is not rigid.

Example 3.6. Suppose $V = \mathbb{R}^2$ and $\|c\|_X = \|c\|_Y = \|c\|_\infty$ for all $c \in V$. Then $(1,1)^t$ is rigid because for $x \in [0,1]$ the vector $(x,x)^t$ is the unique solution to $\mathbf{M}^c_{YX}(x)$. The vector $c = (1,0)^t$ is not rigid: If $0 < x < 1$, then $\mathbf{M}^c_{YX}(x)$ has infinitely many solutions, namely $(x,s)^t$, $|s| \leq \min\{x, 1-x\}$.

If $c \in V$ is rigid, then we can study how the unique solution of $\mathbf{M}^c_{YX}(x)$ varies as we change the value of x. The lemma below shows that the solution varies continuously. In various contexts, this property is well known and is used in homotopy continuation methods (see, for example, [42, 43, 21, 6]) for some optimization problems.

Lemma 3.7. *Suppose that $c \in V$ is rigid, and let $\alpha_{YX}(x) = \alpha^c_{YX}(x)$ be the unique solution to $\mathbf{M}^c_{YX}(x)$ for $x \in [0, \|c\|_X]$. Then $\alpha_{YX} : [0, \|c\|_X] \to V$ is continuous.*

Proof. Suppose that $x_1, x_2, \ldots \in [0, \|c\|_X]$ is a sequence for which $x = \lim_{n\to\infty} x_n$ exists. We assume that $\lim_{n\to\infty} \alpha_{YX}(x_n)$ does not exist, or that it is not equal to $\alpha_{YX}(x)$. By replacing $x_1, x_2, \ldots$ by a subsequence, we may assume that $a = \lim_{n\to\infty} \alpha_{YX}(x_n)$ exists but that it is not equal to $\alpha_{YX}(x)$. We have $\|a\|_X = \lim_{n\to\infty} \|\alpha_{YX}(x_n)\|_X = \lim_{n\to\infty} x_n = x$. Also, we get $\|c - a\|_Y = \lim_{n\to\infty} \|c - \alpha_{YX}(x_n)\|_Y = \lim_{n\to\infty} f_{YX}(x_n) = f_{YX}(x)$ because f is continuous. Because $\mathbf{M}_{YX}(x)$ has a unique solution, we conclude that $a = \alpha_{YX}(x)$. Contradiction. We conclude that $\lim_{n\to\infty} \alpha_{YX}(x_n) = \alpha_{YX}(x)$. This proves that α_{YX} is continuous. $\blacksquare$

Definition 3.8. *The norm $\|\cdot\|_Y$ is called* strictly convex *if $\|a+b\|_Y = \|a\|_Y + \|b\|_Y$ implies that a and b are linearly dependent.*

The ℓ_2 norm on $\mathbb{R}^n$ is strictly convex, for example.

Lemma 3.9. *If $\|\cdot\|_Y$ is strictly convex, then every vector is rigid.*

Proof. Suppose that $\|\cdot\|_Y$ is strictly convex and that $c \in V$. We will prove that c is rigid. Suppose $\|a\|_X = \|a'\|_X = x$ and $\|b\|_Y = \|b'\|_Y = f_{YX}(x)$, where $b = c - a$ and $b' = c - a'$. Let $\overline{a} = (a + a')/2$. Then we have $\|\overline{a}\|_X \leq \frac{1}{2}(\|a\|_X + \|a'\|_X) = x$. By definition, we have $\|c - \overline{a}\|_Y \geq f_{YX}(x)$. It follows that

$$\|b + b'\|_Y = \|2(c - \overline{a})\|_Y = 2\|c - \overline{a}\|_Y \geq 2f_{YX}(x) = \|b\|_Y + \|b'\|_Y.$$

Since $\|\cdot\|_Y$ is strictly convex, b and b' are linearly dependent. Because $\|b\|_Y = \|b'\|_Y$, it follows that $b = \pm b'$. If $b = -b'$, then we have $b + b' = 0$. From $\|b\|_Y + \|b'\|_Y \leq \|b + b'\|_Y = 0$ follows that $b = b' = 0$ and $a = a' = c$, and the uniqueness is established. If $b = b'$, then $a = a'$, and we have again uniqueness. $\blacksquare$

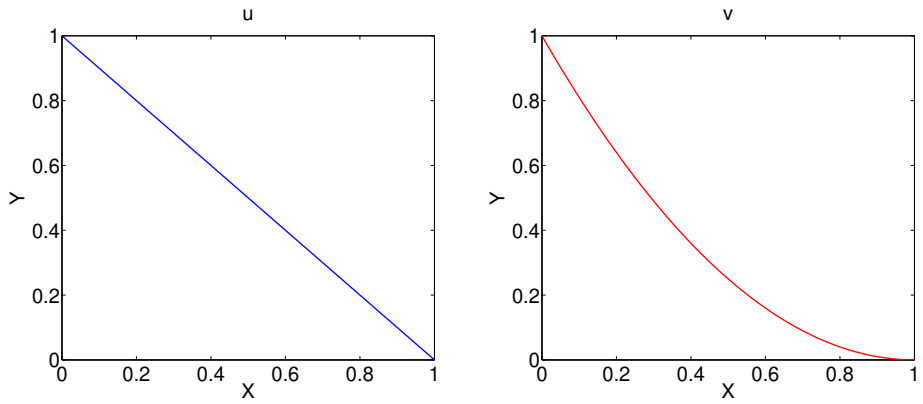

Figure 7.

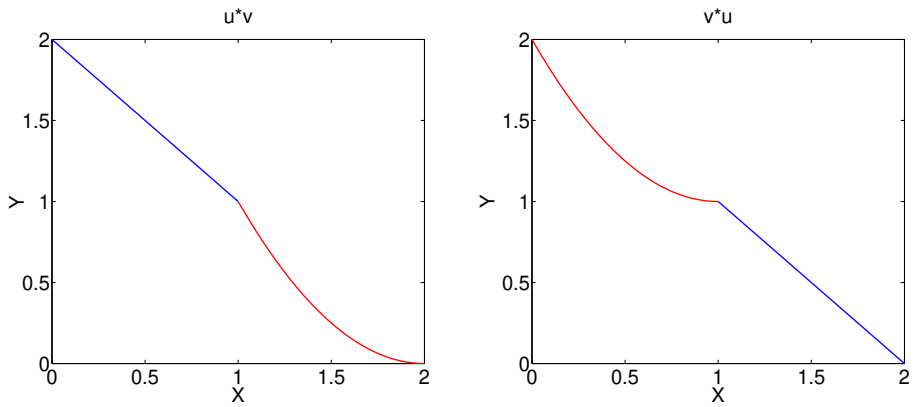

Figure 8.

3.4. The Pareto curve of sums. Next we will compare the Pareto curve of f^{a+b}_{YX} with the Pareto curves f^a_{YX} and f^b_{YX}. For this purpose, we introduce the concatenation of two functions. Suppose that $u : [0,t] \to \mathbb{R}$ and $v : [0,s] \to \mathbb{R}$ are two functions with $u(t) = v(s) = 0$. The *concatenation* $u \star v : [0, s+t] \to \mathbb{R}$ is defined by

$$u \star v(x) = \begin{cases} u(x) + v(0) & \text{if } 0 \le x \le t, \\ v(x-t) & \text{if } t \le x \le s+t. \end{cases}$$

Note that $(u \star v)(0) = u(0) + v(0)$ and $(u \star v)(s+t) = v(s) = 0$. If u and v are decreasing, then so is $u \star v$. If u and v are continuous, then so is $u \star v$. Note that concatenation is associative: $(u \star v) \star w = u \star (v \star w)$. However, it is not commutative.

Example 3.10. Suppose that $u, v : [0,1] \to \mathbb{R}$ are defined by $u(x) = 1 - x$ and $v = (1-x)^2$ (see Figure 7). The graphs of $u \star v$ and $v \star u$ are shown in Figure 8.

Lemma 3.11. *Suppose that $a, b \in V$. If $0 \le x \le \|a+b\|_X$, then we have*

$$f^{a+b}_{YX}(x) \le (f^a_{YX} \star f^b_{YX})(x).$$

Proof. Suppose that $0 \leq x \leq \|a\|_X$. Choose $d \in V$ such that $\|d\|_X = x$ and $\|a-d\|_Y = f^a_{YX}(x)$. Then we have

$$f^{a+b}_{YX}(x) \leq \|a+b-d\|_Y \leq \|a-d\|_Y + \|b\|_Y = f^a_{YX}(x) + f^b_{YX}(0) = (f^a_{YX} \star f^b_{YX})(x).$$

Reversing the roles of X and Y and of a and b gives

$$f^{a+b}_{XY}(y) \leq f^b_{XY}(y) + f^a_{XY}(0) = f^b_{XY}(y) + \|a\|_X$$

if $0 \leq y \leq \|b\|_Y$. Substituting $y = f^{a+b}_{YX}(x)$ where $\|a\|_X \leq x \leq \|a+b\|_X$ yields

$$x - \|a\|_X \leq f^b_{XY}(f^{a+b}_{YX}(x)).$$

Applying the decreasing function f^b_{YX} gives

$$(f^a_{YX} \star f^b_{YX})(x) = f^b_{YX}(x - \|a\|_X) \geq f^{a+b}_{YX}(x). \qquad \blacksquare$$

4. Duality.

4.1. $X2$- and $2Y$-decompositions. Throughout this section V is a finite dimensional vector space equipped with a positive definite bilinear form $\langle \cdot, \cdot \rangle$ and a norm $\|\cdot\|_X$. The bilinear form gives an ℓ_2 norm $\|\cdot\|_2$ and $\|\cdot\|_Y$ will always be the dual norm of $\|\cdot\|_X$. In this section we will study $X2$-decompositions, which turn out to be the same as $2Y$-decompositions. We start with the following characterization of an $X2$-decomposition.

Proposition 4.1. *The expression $c = a + b$ is an $X2$-decomposition if and only if $\langle a, b \rangle = \|a\|_X \|b\|_Y$, where $\|\cdot\|_Y$ is the dual norm of $\|\cdot\|_X$.*

Proof. Suppose that $c = a + b$ is an $X2$-decomposition. Choose a vector d such that $\langle b, d \rangle = \|b\|_Y$ and $\|d\|_X = 1$. Let $\varepsilon > 0$. Define $a' = (1-\varepsilon)a + \varepsilon \|a\|_X d$ and $b' = c - a'$. We have $\|a'\|_X \leq (1-\varepsilon)\|a\|_X + \varepsilon\|a\|_X = \|a\|_X$. Therefore, $\|b'\|_2 \geq \|b\|_2$. It follows that

$$\begin{aligned} 0 \leq \|b'\|_2^2 - \|b\|_2^2 = \|b + (a-a')\|_2^2 - \|b\|_2^2 = 2\langle a - a', b\rangle + \|a-a'\|_2^2 \\ = 2\varepsilon \langle a - \|a\|_X d, b\rangle + \varepsilon^2 \|a - \|a\|_X d\|_2^2. \end{aligned}$$

Taking the limit $\varepsilon \downarrow 0$ yields the inequality

$$0 \leq \langle a - \|a\|_X d, b\rangle = \langle a, b\rangle - \|a\|_X \|b\|_Y,$$

so $\langle a, b\rangle \geq \|a\|_X \|b\|_Y$. The opposite inequality $\langle a, b\rangle \leq \|a\|_X \|b\|_Y$ holds because the norms are dual to each other. We conclude that $\langle a, b\rangle = \|a\|_X \|b\|_Y$.

Conversely, suppose that $\langle a, b\rangle = \|a\|_X \|b\|_Y$. Let $x = \|a\|_X$, let a' be the solution to $\mathbf{M}^c_{2X}(x)$, and define $b' = c - a'$. Then $c = a' + b'$ is an $X2$-decomposition.

$$\begin{aligned} \|a' - a\|_2^2 = \langle a' - a, b - b'\rangle = \langle a', b\rangle + \langle a, b'\rangle - \langle a', b'\rangle - \langle a, b\rangle \\ = \langle a', b\rangle + \langle a, b'\rangle - x\|b\|_Y - x\|b'\|_Y \leq x\|b'\|_Y + x\|b\|_Y - x\|b'\|_Y - x\|b\|_Y = 0. \end{aligned}$$

So we conclude that $a = a'$, and $c = a + b$ is an $X2$-decomposition. $\blacksquare$

The equivalence between the $X2$-decomposition and $2Y$-decomposition (Proposition 2.3) now easily follows.

Proof of Proposition 2.3. In Proposition 2.3, (1) and (3) are equivalent because of Proposition 4.1. Dually, (2) and (3) are equivalent. $\blacksquare$

4.2. The Pareto subfrontier. We define $h_{YX} = h^c_{YX} : [0, \|c\|_X] \to [0, \|c\|_Y]$ by $h_{YX}(x) = \|c - \alpha_{2X}(x)\|_Y$. The graph of h^c_{YX} is the Pareto subfrontier. Indeed, if $c = a + b$ is an $X2$-decomposition with $\|a\|_X = x$ and $\|b\|_Y = y$, then we have $\alpha_{2X}(x) = a$ and $h_{YX}(x) = \|c - \alpha_{2X}(x)\|_Y = \|c - a\|_Y = \|b\|_Y = y$. We now prove some properties of the Pareto subfrontier.

Lemma 4.2. *We have* $\alpha_{2Y}(h_{YX}(x)) = c - \alpha_{2X}(x)$.

Proof. Let $a = \alpha_{2X}(x)$ and $b = c - a$. Then $c = a + b$ is an $X2$-decomposition and therefore also a $2Y$-decomposition. So $c = b + a$ is a $Y2$-decomposition. Let $y = \|b\|_Y = h_{YX}(x)$. Then we have $b = \alpha_{2Y}(y) = \alpha_{2Y}(h_{YX}(x))$. ∎

Lemma 4.3. *The function* $h_{YX}(x)$ *is a strictly decreasing homeomorphism, and its inverse is* $h_{XY}(x)$.

Proof. We have

$$\alpha_{2X}(h_{XY}(h_{YX}(x))) = c - \alpha_{2Y}(h_{YX}(x)) = \alpha_{2X}(x).$$

The function α_{2X} is injective, because the function $f_{2X}(x) = \|c - \alpha_{2X}(x)\|_2$ is injective. It follows that $h_{XY}(h_{YX}(x)) = x$. By symmetry, we also have $h_{YX}(h_{XY}(y)) = y$, so h_{XY} is the inverse of h_{YX}.

The function $h_{YX}(x) = \|c - \alpha_{2X}(x)\|_Y$ is continuous, because α_{2X} and $\|\cdot\|_Y$ are continuous. This proves that h_{YX} is a homeomorphism. By the intermediate value theorem, it has to be strictly increasing or strictly decreasing. Since $h_{YX}(\|c\|_X) = 0 \le h_{YX}(0) = \|c\|_Y$, the function h_{YX} must be strictly decreasing. ∎

Proposition 4.4. *The function* h_{YX} *is Lipschitz continuous.*

Proof. Since $\|\cdot\|_2$ and $\|\cdot\|_X$ are norms on a finite dimensional vector space, there exists positive constant $K > 0$ such that $\|c\|_Y \le K\|c\|_2$. Suppose that $c = a_1 + b_1$ and $c = a_2 + b_2$ are XY-decompositions, with $x_2 := \|a_2\|_X > x_1 := \|a_1\|_X$. It follows that $y_2 := \|b_2\|_Y < y_1 := \|b_1\|_Y$. We have

$$\begin{aligned} K^{-2}(y_1 - y_2)^2 \le K^{-2}\|b_1 - b_2\|_X^2 \le \|b_1 - b_2\|_2^2 &= \langle a_2 - a_1, b_1 - b_2\rangle \\ \le \langle a_2, b_1\rangle + \langle a_1, b_2\rangle - \langle a_1, b_1\rangle - \langle a_2, b_2\rangle &= \langle a_2, b_1\rangle + \langle a_1, b_2\rangle - x_1y_1 - x_2y_2 \\ \le x_2y_1 + x_1y_2 - x_1y_1 - x_2y_2 &= (x_2 - x_1)(y_1 - y_2). \end{aligned}$$

We conclude that

$$\frac{|y_2 - y_1|}{|x_2 - x_1|} = \frac{y_1 - y_2}{x_2 - x_1} \le K^2.$$ ∎

4.3. Differentiating the Pareto curve. The function $f_{2X}(x)$ is differentiable. A special case (but with a similar proof) was treated in [6, sect. 2].

Proposition 4.5. *The function* $f_{2X}(x)$ *is differentiable on* $[0, \|c\|_X]$, *and*

$$f'_{2X}(x) f_{2X}(x) = -h_{YX}(x).$$

Proof. If $0 \le x_1, x_2 \le \|c\|_X$, then we have

$$\begin{aligned}(f_{2X}(x_2))^2 - (f_{2X}(x_1))^2 &\ge \|c - \alpha_{2X}(x_2)\|_2^2 - \|c - \alpha_{2X}(x_1)\|_2^2 - \|\alpha_{2X}(x_2) - \alpha_{2X}(x_1)\|_2^2 \\ = 2\langle c - \alpha_{2X}(x_1), \alpha_{2X}(x_1) - \alpha_{2X}(x_2)\rangle &= 2x_1 h_{YX}(x_1) - 2\langle c - \alpha_{2X}(x_1), \alpha_{2X}(x_2)\rangle \\ &\ge 2(x_1 - x_2)h_{YX}(x_1) = -2(x_2 - x_1)h_{YX}(x_1).\end{aligned}$$

Reversing the roles of x_1, x_2 gives us

$$(f_{2X}(x_2))^2 - (f_{2X}(x_1))^2 \le -2(x_2 - x_1)h_{YX}(x_2).$$

If $x_2 > x_1$, then we obtain

$$-2h_{YX}(x_1) \le \frac{(f_{2X}(x_2))^2 - (f_{2X}(x_1))^2}{x_2 - x_1} \le -2h_{YX}(x_2),$$

and if $x_1 > x_2$, then we have

$$-2h_{YX}(x_2) \le \frac{(f_{2X}(x_2))^2 - (f_{2X}(x_1))^2}{x_2 - x_1} \le -2h_{YX}(x_1).$$

Since h_{YX} is continuous, it follows that $(f_{2X})^2$ is differentiable on $[0, \|c\|_X]$ with derivative $-2h_{YX}$. Since $f_{2X}(x)$ is positive on $[0, \|c\|_X)$, it is differentiable on $[0, \|c\|_X)$. We have

$$2f'_{2X}(x)f_{2X}(x) = \frac{d}{dx}\Big(f_{2X}(x)\Big)^2 = -2h_{YX}(x).$$ ∎

Proof of Proposition 2.8. From Proposition 4.5 follows that

$$\tfrac{d}{dx}\big(-\tfrac{1}{2}f_{2X}(x)^2\big) = -f'_{2X}(x)f_{2X}(x) = h_{YX}(x).$$

So the area to the right of $x = \|a\|_X$ is

$$\int_{\|a\|_X}^{\|c\|_X} h_{YX}(x)\,dx = \big[-\tfrac{1}{2}f_{2X}(x)^2\big]_{\|a\|_X}^{\|c\|_X} = -\tfrac{1}{2}(0^2 - \|b\|_2^2) = \tfrac{1}{2}\|b\|_2^2.$$

Similarly, the area below the graph and above the line $y = \|b\|_Y$ is equal to $\|a\|_2^2$. The area below the graph of h_{YX} is equal to $\frac{1}{2}\|c\|_2^2 = \frac{1}{2}\|a\|_2^2 + \frac{1}{2}\|b\|_2^2 + \langle a, b\rangle$. The area of the $\|a\|_X \times \|b\|_Y$ rectangle is $\|a\|_X\|b\|_Y = \langle a, b\rangle$. ∎

The solution for $\mathbf{M}_{2X}^c(\lambda)$ can be obtained from a regularized quadratic minimization problem.

Proposition 4.6. *The vector a is a solution to $\mathbf{M}_{2X}^c(\lambda)$ if and only if*

$$\|c - a\|_Y + \tfrac{1}{2\lambda}\|a\|_2^2$$

is minimal, where $\|\cdot\|_Y$ is the dual norm of $\|\cdot\|_X$.

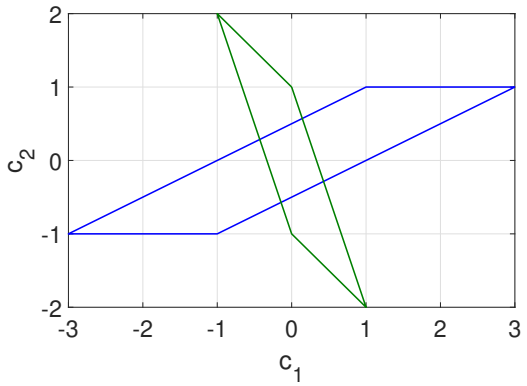

Figure 9.

Proof. We can choose a such that $\|c-a\|_Y + \frac{1}{2\lambda}\|a\|_2^2$ is minimal. Let $b = c - a$. Then $c = a + b$ is an $X2$-decomposition, so $\langle a, b\rangle = \|a\|_X\|b\|_Y$. The function

$$f(t) = \|tb\|_Y + \tfrac{1}{2\lambda}\|c - tb\|_2^2 = t\|b\|_Y + \tfrac{1}{\lambda}\langle c - bt, c - bt\rangle$$

has a minimum at $t = 1$. So we have

$$0 = f'(1) = \|b\|_Y - \tfrac{1}{\lambda}\langle a, b\rangle = \|b\|_Y - \tfrac{1}{\lambda}\|a\|_X\|b\|_Y$$

and $\|a\|_X = \lambda$. This shows that a is a solution to $\mathbf{M}^c_{2X}(\lambda)$. Since $M^c_{2X}(\lambda)$ has a unique solution, this unique solution a must minimize $\|c-a\|_Y + \frac{1}{2\lambda}\|a\|_2^2$. ∎

A similar argument shows that a is a solution to $\mathbf{M}^c_{2X}(\lambda)$ if and only if $\|c-a\|_Y + \frac{1}{\lambda}\|a\|_2$ is minimal.

5. Tight vectors. Throughout this section, V is an n-dimensional vector space with a positive definite bilinear form, and $\|\cdot\|_X$ and $\|\cdot\|_Y$ are norms which are dual to each other. From the definitions it is clear that $f^c_{YX}(x) \le h^c_{YX}(x)$. Recall that c is *tight* if we have equality for all $x \in [0, \|c\|_X]$. If $c \in V$ is tight and rigid, then $\alpha_{2X}(x) = \alpha_{YX}(x)$ for $x \in [0, \|c\|_X]$. If every vector in V is tight, then $\|\cdot\|_X$ and $\|\cdot\|_Y$ are called *tight norms*. In this section we study properties of tight vectors and tight norms.

5.1. An example of a norm that is not tight. Consider the norm on $\mathbb{R}^2$ defined by $\|c\|_X = \max\{|c_2|, |2c_2 - c_1|\}$ for $c = (c_1, c_2)^t$. Its dual norm is given by $\|c\|_Y = \max\{|c_1 + c_2|, |3c_1 + c_2|\}$. The unit balls are polar duals of each other (see Figure 9).

Consider the vector $c = (3, 12)^t$. Figure 10 shows the functions f^c_{YX} and h^c_{YX}. We see that f^c_{YX} and h^c_{YX} are not the same, so c is not tight. The example shows that h^c_{YX} is not always convex.

The trajectories of α^c_{YX} (green) and α^c_{2X} (blue) are sketched in Figure 11.

For every positive value of x, $\alpha^c_{2X}(x)/x$ lies on the unit ball B_X. In fact, $\alpha^c_{2X}(x)/x$ is the vector in B_X that is closest to c/x with respect to the Euclidean distance. In Figure 12, c/x and $\alpha^c_{2X}(x)/x$ are plotted for various values of x. Note that $\alpha^c_{2X}(x)/x$ is constant on the

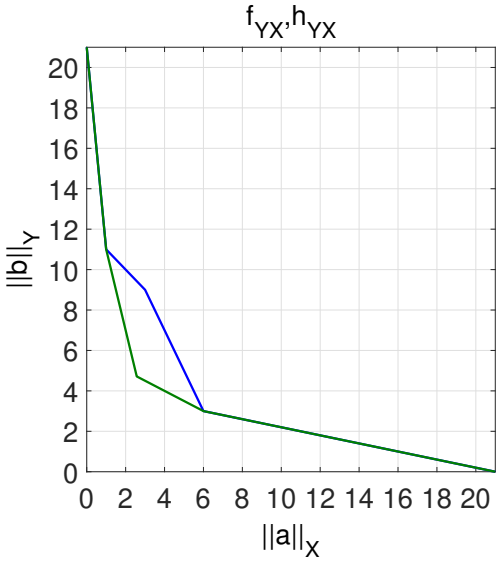

Figure 10.

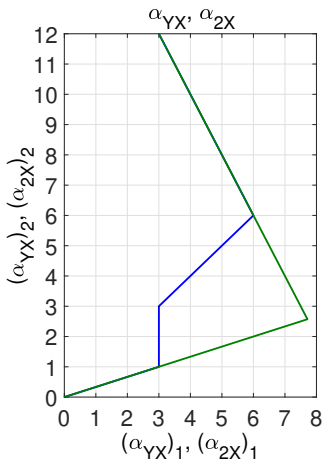

Figure 11.

intervals $(0,1]$ and $[3,6]$. On these intervals $\alpha^c_{2X}(x)$ moves on a line through the origin. On the other intervals $[1,3]$ and $[6,21]$, $\alpha^c_{2X}(x)$ moves on a line through c.

The singular value region for the vector c is shown in Figure 13.

If we want to interpret the singular value region in terms of singular values, we must allow negative multiplicities. The singular values of c are 21 with multiplicity 0.1, 11 with multiplicity 0.9, 9 with multiplicity -0.5, and 3 with multiplicity 4.5.

5.2. The Pareto subfrontier of sums. We have defined the concatenation of two graphs, and now we define the concatenation of two paths in V in a similar manner. If $\beta : [0,t] \to V$

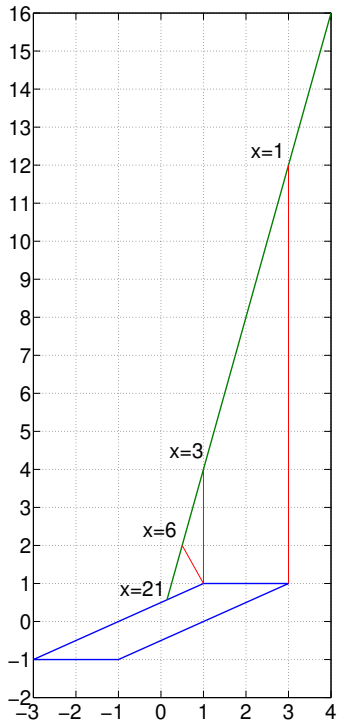

Figure 12.

and $\gamma : [0, s] \to V$ are curves with $\beta(0) = \gamma(0) = 0$, then we define the concatenation $\beta \star \gamma : [0, s + t] \to V$ by

$$(\beta \star \gamma)(x) = \begin{cases} \beta(x) & \text{if } 0 \le x \le t, \\ \beta(t) + \gamma(s - t) & \text{if } t \le x \le s + t. \end{cases}$$

Theorem 5.1. *Suppose that $\|\cdot\|_X$ and $\|\cdot\|_Y$ are dual norms, $c \in V$ is tight, and $c = a + b$ is an XY-decomposition.*

(1) *a and b are tight as well;*
(2) *$\|c\|_X = \|a\|_X + \|b\|_X$ and $\|c\|_Y = \|a\|_Y + \|b\|_Y$;*
(3) *$f^c_{YX} = f^a_{YX} \star f^b_{YX}$;*
(4) *$\alpha^c_{YX} = \alpha^a_{YX} \star \alpha^b_{YX}$.*

Proof. Suppose that $0 \le x \le \|a\|_X$. Choose a vector d such that $\|d\|_X = x$ and $\|a-d\|_Y = f^a_{YX}(x)$. We have

$$h^c_{YX}(x) - \|b\|_Y = f^c_{YX}(x) - \|b\|_Y \le \|c - d\|_Y - \|c - a\|_Y \le \|a - d\|_Y = f^a_{YX}(x) \le h^a_{YX}(x).$$

Suppose that one of the inequalities above is strict for some $x \in [0, \|a\|_X]$. Integrating t from

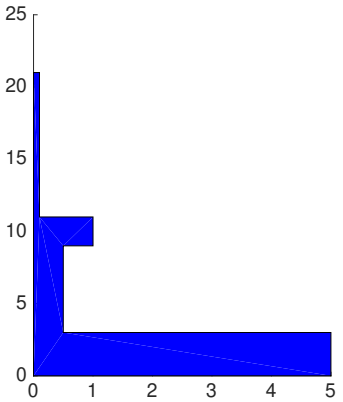

Figure 13.

0 to x yields

$$\tfrac{1}{2}\|a\|_2^2 = \int_0^{\|a\|_X} h^a_{YX}(t)\,dt > \int_0^{\|a\|_X} (h^c_{YX}(x) - \|b\|_Y)\,dt$$
$$= \tfrac{1}{2}(\|c\|_2^2 - \|a\|_2^2) - \|a\|_X\|b\|_Y = \tfrac{1}{2}(\|a+b\|_2^2 - \|b\|_2^2 - 2\langle a,b\rangle) = \tfrac{1}{2}\|a\|_2^2.$$

Contradiction. We conclude that $h^a_{YX}(x) = f^a_{YX}(x) = h^c_{YX}(x) - \|b\|_Y$ for all $x \in [0, \|a\|_X]$. In particular, a is tight. By symmetry, b is tight as well. This proves (1). For $x = 0$ we get $\|a\|_Y = h^a_{YX}(0) = h^c_{YX}(0) - \|b\|_Y = \|c\|_Y - \|b\|_Y$. So we have $\|c\|_Y = \|a\|_Y + \|b\|_Y$, and by symmetry we also have $\|c\|_X = \|a\|_X + \|b\|_X$. This proves (2).

If $0 \le x \le \|a\|_X$, then we have

$$f^c_{YX}(x) = h^c_{YX}(x) = h^a_{YX}(x) + \|b\|_Y = f^a_{YX}(x) + f^b_{YX}(0) = (f^a_{YX} \star f^b_{YX})(x).$$

By symmetry, if $0 \le y \le \|b\|_Y$, then we have

$$f^c_{XY}(y) = f^b_{XY}(y) + \|a\|_X.$$

It follows that

$$f^b_{YX}(f^c_{XY}(y) - \|a\|_X) = f^b_{YX}(f^b_{XY}(y)) = y.$$

Substituting $y = f^c_{YX}(x)$ yields

$$(f^a_{YX} \star f^b_{YX})(x) = f^b_{YX}(x - \|a\|_X) = f^c_{YX}(x).$$

This proves (3).

If $x \in [0, \|a\|_X]$, then we have

$$\|(\alpha^a_{YX} \star \alpha^b_{YX})(x)\|_X = \|\alpha^a_{YX}(x)\|_X = x.$$

So we get

$$\|c - (\alpha^a_{YX} \star \alpha^b_{YX})(x)\|_Y = \|c - \alpha^a_{YX}(x)\|_Y \le \|b\|_Y + \|a - \alpha^a_{YX}(x)\|_Y = \|b\|_Y + h^a_{YX}(x) = f^c_{YX}(x).$$

If $x \in [\|a\|_X, \|c\|_X]$, then we have

$$\|(\alpha^a_{YX} \star \alpha^b_{YX})(x)\|_X = \|a + \alpha^b_{YX}(x - \|a\|_X)\|_X \\ \leq \|a\|_X + \|\alpha^b_{YX}(x - \|a\|_X)\|_X = \|a\|_X + (x - \|a\|_X) = x$$

and

$$\|c - (\alpha^a_{YX} \star \alpha^b_{YX})(x)\|_Y = \|c - a - \alpha^b_{YX}(x - \|a\|_X)\|_Y \\ = \|b - \alpha^b_{YX}(x - \|a\|_X)\|_Y = f^b_{YX}(x - \|a\|_X) = f^c_{YX}(x).$$

So for all $x \in [0, \|c\|_X]$ we have

$$\|(\alpha^a_{YX} \star \alpha^b_{YX})(x)\|_X \leq x$$

and

$$\|c - (\alpha^a_{YX} \star \alpha^b_{YX})(x)\|_Y \leq f^c_{YX}(x).$$

So $(\alpha^a_{YX} \star \alpha^b_{YX})(x)$ is the unique solution to $\mathbf{M}^c_{YX}(x)$ and therefore equal to $\alpha^c_{YX}(x)$. This proves (4). ∎

Proof of Proposition 2.12. We already know part (1) in the case $x_1 \leq x_2$. Assume that $x_1 > x_2$. Let $c = a + b$ be the XY-decomposition with $\|a\|_X = x_1$. This is also an $X2$-decomposition and $a = \mathrm{proj}_X(c, x_1)$. Now a and b are tight by Theorem 5.1(1). Let $a = a_1 + a_2$ be an XY-decomposition with $\|a_1\|_X = x_2$. Now $a = a_1 + a_2$ is also an $X2$-decomposition and $a_1 = \mathrm{proj}_X(a, x_2) = \mathrm{proj}_X(\mathrm{proj}_X(c, x_1), x_2)$. We have $a_1 = \alpha^a_{YX}(x_1) = (\alpha^a_{YX} \star \alpha^b_{YX})(x_1) = \alpha^c_{YX}(x_1)$. So $c = a_1 + (c - a_1)$ is an XY-decomposition (and an $X2$-decomposition) and $a_1 = \mathrm{proj}_X(c, x_2)$. This proves that $\mathrm{proj}_X(c, x_2) = \mathrm{proj}_X(\mathrm{proj}_X(c, x_1), x_2)$, and part (1) has been proved.

Suppose that $c = a + b$ is an $X2$-decomposition with $\|a\|_X = x_1$. Then $b = \mathrm{shrink}_X(c, x_1)$. Let $b = b_1 + b_2$ be an $X2$-decomposition with $\|b_1\|_X = x_2$. Then $b_2 = \mathrm{shrink}_X(b, x_2) = \mathrm{shrink}_X(\mathrm{shrink}_X(c, x_1), x_2)$. Similar reasoning as before shows that $c = (a + b_1) + b_2$ is an $X2$-decomposition. Also $(a + b_1)$ is tight and $(a + b_1) = a + b_1$ is an XY-decomposition, so $\|a + b_1\|_X = \|a\|_X + \|b_1\|_X = x_1 + x_2$. Therefore, $b_2 = \mathrm{shrink}_X(c, x_1 + x_2)$ and we are done. ∎

Lemma 5.2. *If $c = a + b$ is an $X2$-decomposition and a and b are tight, then we have $h^c_{YX} = h^a_{YX} \star h^b_{YX}$.*

Proof. Suppose that $0 \leq x \leq \|a\|_X$, and let $d = \alpha^a_{YX}(x) = \alpha^a_{2X}(x)$. We get

$$\|d\|_X\|c - d\|_Y \geq \langle d, c - d\rangle = \langle d, a - d\rangle + \langle d, b\rangle = \|d\|_X\|a - d\|_Y + \langle a, b\rangle - \langle a - d, b\rangle \\ \geq \|d\|_X\|a-d\|_Y + \|a\|_X\|b\|_Y - \|a-d\|_X\|b\|_Y = \|d\|_X\|a-d\|_Y + \|d\|_X\|b\|_Y \geq \|d\|_X\|c-d\|_Y.$$

It follows that

$$\|d\|_X\|c - d\|_Y = \langle d, c - d\rangle = \|d\|_X(\|a - d\|_Y + \|b\|_Y).$$

So $c = d + (c - d)$ is an $X2$-decomposition and $\|c - d\|_Y = \|a - d\|_Y + \|b\|_Y$. We get

$$h^c_{YX}(x) = \|c - d\|_Y = \|b\|_Y + \|a - d\|_Y = h^a_{YX}(x) + \|b\|_Y = (h^a_{YX} \star h^b_{YX})(x).$$

Suppose that $\|a\|_X \leq x \leq \|c\|_X$, and define $y = h^c_{YX}(x)$. Then we have $0 \leq y \leq \|b\|_Y$, and by symmetry we get

$$x = h^c_{XY}(y) = (h^b_{XY} \star h^a_{XY})(y) = h^b_{XY}(y) + \|b\|_X = h^b_{XY}(h^c_{YX}(x)) + \|b\|_X$$

and

$$(h^a_{YX} \star h^b_{YX})(x) = h^b_{YX}(x - \|b\|_X) = h^b_{YX}(h^b_{XY}(h^c_{YX}(x))) = h^c_{YX}(x).$$

We conclude that $h^c_{YX} = h^a_{YX} \star h^b_{YX}$. ∎

We will show later in Proposition 6.11 that under the assumptions of Lemma 5.2 the vector c is tight.

6. The slope decomposition.

6.1. Unitangent vectors. Throughout this section $\|\cdot\|_X$ and $\|\cdot\|_Y$ will again be dual norms in an n-dimensional Euclidean vector space V. In this section we study the slope decomposition. We show that a vector c is tight if and only if it has a slope decomposition. We also will show that the Pareto frontier is always piecewise linear for a tight vector. In that case, the different slopes in the Pareto frontier correspond to different summands in the slope decomposition of c.

For a vector $c \in V$ we have the inequality $\|c\|_2^2 = \langle c, c\rangle \leq \|c\|_X\|c\|_Y$.

Definition 6.1. *We call a vector $c \in V$* unitangent *if $\|c\|_2^2 = \|c\|_X\|c\|_Y$.*

Unitangent vectors are the simplest kind of tight vectors. As we will see, their Pareto frontier is linear, i.e., has only one slope. Recall that $\|c\|_Y$ is the maximal value of the functional $\langle c, \cdot\rangle$ on the unit ball B_X. Now c is unitangent if and only if the maximum of $\langle c, \cdot\rangle$ is attained at $c/\|c\|_X \in B_X$.

Proposition 6.2. *If $c \in V$ is unitangent, then it is tight, and we have $\alpha^c_{YX}(x) = (x/\|c\|_X)c$ and $f^c_{YX}(x) = \|c\|_Y(1 - x/\|c\|_X)$ for $x \in [0, \|c\|_X]$.*

Proof. Suppose that $\|a\|_X \leq x$. Then we have

$$\begin{aligned}\|c - a\|_Y\|c\|_X \geq \langle c - a, c\rangle = \|c\|_2^2 - \langle a, c\rangle \geq \|c\|_2^2 - \|a\|_X\|c\|_Y \\ = \|c\|_Y(\|c\|_X - \|a\|_X) \geq \|c\|_Y(\|c\|_X - x).\end{aligned}$$

It follows that

$$\|c - a\|_Y \geq \|c\|_Y(1 - x/\|c\|_X).$$

Since a was arbitrary, we conclude that $f^c_{YX}(x) \geq \|c\|_Y(1 - x/\|c\|_X)$.

If we take $a = (x/\|c\|_X)c$, then we have $\|a\|_X = x$ and $\|c - a\|_Y = \|c\|_Y(1 - x/\|c\|_X) \leq f^c_{YX}(x)$. We conclude that $f^c_{YX}(x) = \|c\|_Y(1 - x/\|c\|_X)$ and $\alpha^c_{YX}(x) = a = (x/\|c\|_X)c$. ∎

6.2. Faces of the unit ball. Suppose that B is a compact convex subset of a finite dimensional $\mathbb{R}$-vector space V. Recall that a convex closed subset F of B is a face if $v, w \in B$, $0 < t < 1$, and $tv + (1-t)w \in F$ implies that $v, w \in F$.

Lemma 6.3. *If F is a face of B, $v_1, v_2, \ldots, v_r \in B$, $t_1, t_2, \ldots, t_r > 0$ such that $t_1 + t_2 + \cdots + t_r = 1$ and $t_1v_1 + \cdots + t_rv_r \in F$, then $v_1, v_2, \ldots, v_r \in F$.*

Proof. We prove the statement by induction on r. This is clear from the definition of a face for $r \leq 2$. Suppose $r > 2$. Let $t'_i = t_i/(1-t_r)$ and $w = t'_1 v_1 + \cdots + t'_{r-1} v_{r-1}$. We have $t_r v_r + (1-t_r)w = t_1 v_1 + \cdots + t_r v_r \in F$, so $v_r, w \in F$. Since $t'_1 + \cdots + t'_{r-1} = 1$ and $w = t'_1 v_1 + \cdots + t'_{r-1} v_{r-1} \in F$, we get $v_1, \ldots, v_{r-1}$ by induction. ∎

Lemma 6.4. *If $B \subseteq V$ is a compact convex set, then the smallest face containing $v \in B$ is*

$$F = \{w \in B \mid (1+t)v - tw \in B \text{ for some } t > 0\}.$$

Proof. Suppose that $w_1, w_2 \in F$ and $sw_1 + (1-s)w_2 \in B$ for some $s > 0$. There exists $t > 0$ such that $(1+t)v - tw_1, (1+t)v - tw_2 \in B$, so we have

$$s\big((1+t)v - tw_1\big) + (1-s)\big((1+t)v - tw_2\big) = (1+t)v - t\big(sw_1 + (1-s)w_2\big) \in B,$$

and therefore $sw_1 + (1-s)w_2 \in F$. This proves that F is a face of B.

Suppose that F' is any face of B containing v. If $w \in F$, then there exists $t > 0$ such that $(1+t)v - tw \in B$. Since $v \in F'$ is a convex combination of $(1+t)v - tw, w \in B$, we have $w, (1+t)v - tw \in F'$. So F' contains F. ∎

Let B_X be the unit ball for the norm $\|\cdot\|_X$.

Lemma 6.5. *A convex cone C in V is a facial X-cone if and only if it has the following norm-sum property: If $a, b \in V$, $a+b \in C$ and $\|a+b\|_X = \|a\|_X + \|b\|_X$ then we have $a, b \in C$.*

Proof. Suppose that C is a cone in V. If $C = \{0\}$ then C is a facial X-cone and has the norm-sum property. Assume now that $C \neq \{0\}$. Let $\partial B_X = \{c \in V \mid \|c\|_X = 1\}$ be the unit sphere. Take $F = C \cap \partial B_X$ so that $C = \mathbb{R}_{\geq 0} F$. We have to show that F is a face of B_X if and only if C has the norm-sum property.

Suppose that F is a face. If $a + b \in C$ and $\|a+b\|_X = \|a\|_X + \|b\|_X$, then we have

$$\frac{a+b}{\|a+b\|_X} = t\frac{a}{\|a\|_X} + (1-t)\frac{b}{\|b\|_X},$$

where $t = \|a\|_X/\|a+b\|_X$ and $a/\|a\|_X, b/\|b\|_X, (a+b)/\|a+b\|_X \in F$. If $t = 0$ or $t = 1$, then $a = 0$ or $b = 0$ and $a, b \in C$. Otherwise, $0 < t < 1$ and $a/\|a\|_X, b/\|b\|_X \in F$ because F is a face. We conclude that $a, b \in C$. So C has the norm-sum property.

Conversely, suppose that C has the norm-sum property, $a, b \in B_X$, and $0 < t < 1$, such that $ta + (1-t)b \in F$; then we have

$$\|ta + (1-t)b\|_X = 1 = t + (1-t) \geq \|ta\|_X + \|(1-t)b\|_X \geq \|ta + (1-t)b\|_X$$

and the inequalities are equalities. It follows that $\|a\|_X = \|b\|_X = 1$. The norm-sum property gives $ta, (1-t)b \in C$, so $a, b \in C$. We conclude that $a, b \in C \cap \partial B_X = F$. ∎

Lemma 6.6. *Suppose that $a \in V$ is nonzero, and C is the set of all $b \in V$ for which there exists $\varepsilon > 0$ such that $\|a - \varepsilon b\|_X + \|\varepsilon b\|_X = \|a\|_X$. Then C is the smallest facial X-cone containing a.*

Proof. If $a \in V$, then we have $\|a - \varepsilon b\|_X + \|\varepsilon b\|_X = \|a\|_X$, and every facial X cone containing a must also contain εb and b by Lemma 6.5. Now C itself is a facial X cone: if

$b_1, b_2 \in C$, then there exists $\varepsilon_1, \varepsilon_2 > 0$ such that $\|a - \varepsilon_i b_i\|_X + \|\varepsilon_i b_i\|_X = \|a\|_X$ for $i = 1, 2$. We can replace ε_1 and ε_2 by the minimum of the two and assume that $\varepsilon_1 = \varepsilon_2 = \varepsilon$. We have

$$\begin{aligned}\|a\|_X &\le \tfrac{1}{2}(\|a - \varepsilon b_1\|_X + \|\varepsilon b_1\|_X + \|a - \varepsilon b_2\|_X + \|\varepsilon b_2\|_X) \\ &\le \|a - \tfrac{1}{2}\varepsilon(b_1 + b_2)\|_X + \|\tfrac{1}{2}\varepsilon(b_1 + b_2)\|_X \le \|a\|_X.\end{aligned}$$

We must have equalities everywhere, so $\frac{1}{2}(b_1 + b_2)$ and $b_1 + b_2$ lie in C by Lemma 6.5. ∎

Definition 6.7. *For $c \in V$ and $0 < x < \|c\|_X$ we define $\mathcal{F}_X(x)$ as the smallest face of B_X containing $\alpha_{2X}(x)/x$.*

Lemma 6.8. *Suppose that c is a tight vector. If $0 < x_1 < x_2 < \|c\|_X$, then we have $\mathcal{F}_X(x_1) \subseteq \mathcal{F}_X(x_2)$.*

Proof. Suppose that c is tight and $0 < x_1 < x_2 < \|c\|_X$. For $t = x_2/(x_2 - x_1)$ we have

$$(1 + t)\alpha_{YX}(x_2)/x_2 - t\alpha_{YX}(x_1)/x_1 = (\alpha_{YX}(x_2) - \alpha_{YX}(x_1))/(x_2 - x_1)$$

which lies in the unit ball B_X. This proves that $\alpha_{YX}(x_1)/x_1$ lies in $\mathcal{F}_X(x_2)$. We conclude that $\mathcal{F}_X(x_1) \subseteq \mathcal{F}_X(x_2)$. ∎

Proposition 6.9. *If $c \in V$ is tight, then $\alpha_{YX}(x)$ and $f_{YX}(x)$ are piecewise linear.*

Proof. We can divide the interval $[0, \|c\|_X]$ into finitely many intervals such that on each interval $\mathcal{F}_X$ is constant. Suppose that (x_1, x_2) is an open interval on which $\mathcal{F}_X$ is equal to F. The affine hull of F is of the form $d + W$ where $W \subset V$ is a subspace and $d \in W^\perp$. Now $\alpha_{YX}(x)$ is the vector in $xd + W$ closest to c (in the Euclidean norm). If we define $a(x) = \alpha_{YX}(x) - dx$, then $a(x) \in W$ is the vector closest to $c - dx$. So $a(x)$ is the orthogonal projection of $c - dx$ onto W. Since $d \in W^\perp$, $a(x)$ is the orthogonal projection of c onto W, so $a(x) = a$ is constant. This proves that $\alpha_{YX}(x) = a + dx$ is linear.

Because $\langle c - a, a\rangle = 0$, we have

$$xh_{YX}(x) = \langle c - \alpha_{XY}(x), \alpha_{XY}(x)\rangle = \langle c - a - dx, a + dx\rangle = \langle c, d\rangle x + \langle d, d\rangle x^2.$$

So $h_{YX}(x)$ is linear for $x \in (x_1, x_2)$. ∎

Proof of Theorem 2.19. Suppose that $c = a + b$ is an $X2$-decomposition, and let $x = \|a\|_X$ and $y = \|b\|_Y$. The smallest face of B_X containing $x^{-1}a$ is $\mathcal{F}_X(x)$, and the smallest face containing $y^{-1}b$ is $\mathcal{F}_Y(y)$. For every $a' \in \mathcal{F}_X(x)$ and every $b' \in \mathcal{F}_Y(y)$ we have $\langle a', b'\rangle \le 1$. We have $\langle x^{-1}a, y^{-1}b\rangle = 1$. Since $y^{-1}b$ lies in the relative interior of $\mathcal{F}_Y(y)$, we have $\langle x^{-1}a, b'\rangle = 1$ for all $b' \in \mathcal{F}_Y(y)$. Since $x^{-1}a$ lies in the relative interior of $\mathcal{F}_X(x)$, we have $\langle a', b'\rangle = 1$ for all $a' \in \mathcal{F}_X(x)$ and all $b' \in \mathcal{F}_Y(y)$. It follows that

$$\operatorname{gsparse}_X(a) + \operatorname{gsparse}_Y(b) = \dim \mathcal{F}_X(x) + 1 + \dim \mathcal{F}_Y(y) + 1 \le n + 1.$$

If c is tight, then we have $\mathcal{F}_X(x) \subseteq \mathcal{F}_X(\|c\|_X)$ and

$$\operatorname{gsparse}_X(a) = \dim \mathcal{F}_X(x) + 1 \le \dim \mathcal{F}_X(\|c\|_X) + 1 = \operatorname{gsparse}_X(c).$$ ∎

6.3. Proof of Theorem 2.10. Suppose that c is tight. Then $h^c_{YX} = f^c_{YX}$ is piecewise linear by Proposition 6.9. Suppose that $z_0 < z_1 < \cdots < z_r = \|c\|_X$ such that h^c_{YX} is linear on each interval $[z_{i-1}, z_i]$, and that h^c_{YX} is not differentiable at $z_1, \dots, z_{r-1}$. Let $a_i = \alpha^c_{2X}(z_i)$, and define $c_i = a_i - a_{i-1}$ for $i = 1, 2, \dots, r$. We have $a_{i-1} = \alpha^c_{2X}(z_{i-1}) = \alpha^{a_i}_{2X}(z_{i-1})$, so $a_i = a_{i-1} + c_i$ is an $X2$-decomposition for all i. By induction we get

$$f^c_{YX} = f^{c_1}_{YX} \star \cdots \star f^{c_r}_{YX}.$$

The area under the graph of $f^{c_i}_{YX} = h^{c_i}_{YX}$ is $\frac{1}{2}\|c_i\|_2^2$. The area under the graph of $f^{c_1}_{YX} \star \cdots \star f^{c_r}_{YX}$ is

$$\sum_{i<j} \|c_i\|_X \|c_j\|_Y + \tfrac{1}{2} \sum_i \|c_i\|_X \|c_i\|_Y.$$

So we have

$$\sum_{i<j} \langle c_i, c_j \rangle + \tfrac{1}{2} \sum_i \langle c_i, c_i \rangle = \tfrac{1}{2}\|c\|_2^2 = \sum_{i<j} \|c_i\|_X \|c_j\|_Y + \tfrac{1}{2}\|c_i\|_X \|c_i\|_Y \geq \sum_{i<j} \langle c_i, c_j \rangle + \tfrac{1}{2} \sum_i \langle c_i, c_i \rangle.$$

This proves that $\langle c_i, c_j \rangle = \|c_i\|_X \|c_j\|_Y$ for all $i \leq j$. This shows that $c = c_1 + c_2 + \cdots + c_r$ is a slope decomposition.

Conversely, suppose that $c = c_1 + c_2 + \cdots + c_r$ is a slope decomposition. We will show that c is tight. Since $c_1 + \cdots + c_{r-1}$ is also a slope decomposition, by induction we may assume that $c_1 + \cdots + c_{r-1}$ is tight, and

$$h^{c_1+\cdots+c_{r-1}}_{YX} = f^{c_1+\cdots+c_{r-1}}_{YX} = f^{c_1}_{YX} \star \cdots \star f^{c_{r-1}}_{YX} = h^{c_1}_{YX} \star \cdots \star h^{c_{r-1}}_{YX}.$$

Since $c = (c_1 + \cdots + c_{r-1}) + c_r$ is an $X2$-decomposition, it follows from Lemma 5.2 that

$$h^c_{YX} = h^{c_1+\cdots+c_{r-1}}_{YX} \star h^{c_r}_{YX} = h^{c_1}_{YX} \star \cdots \star h^{c_r}_{YX}.$$

Suppose that

$$\|c_1\|_X + \cdots + \|c_{i-1}\|_X \leq x \leq \|c_1\|_X + \cdots + \|c_i\|_X.$$

We have

$$\begin{aligned}\sum_{j<i} \|c_j\|_X \|c_i\|_Y + \sum_{j\geq i} \|c_j\|_Y \|c_i\|_X &= \sum_{j<i} \langle c_j, c_i \rangle + \sum_{j\geq i} \langle c_j, c_i \rangle \\ &= \langle c - a, c_i \rangle + \langle a, c_i \rangle \leq \|c - a\|_Y \|c_i\|_X + \|a\|_X \|c_i\|_Y = f_{YX}(x)\|c_i\|_X + x\|c_i\|_Y.\end{aligned}$$

So we have

$$f^c_{YX}(x) \geq \sum_{j=i}^{r} \|c_j\|_Y + \frac{\|c_i\|_Y}{\|c_i\|_X}\Big(\sum_{j=1}^{i-1} \|c_j\|_X - x\Big) = (f^{c_1}_{YX} \star f^{c_2}_{YX} \star \cdots \star f^{c_r}_{YX})(x).$$

It follows that

$$f^c_{YX} \leq h^c_{YX} = h^{c_1}_{YX} \star \cdots \star h^{c_r}_{YX} = f^{c_1}_{YX} \star \cdots \star f^{c_r}_{YX} \leq f^c_{YX}.$$

We get $f^c_{YX} = h^c_{YX}$, so c is tight. We have proven part (1).

(2) Suppose that c is tight and $c = c_1 + c_2 + \cdots + c_r$ is a slope decomposition. Then the graph $f^{c_i}_{YX}$ is a straight line segment from $(0, y_i)$ to $(x_i, 0)$, where $y_i = \|c_i\|_Y$ and $x_i = \|c_i\|_X$. Since $f^c_{YX} = f^{c_1}_{YX} \star f^{c_2}_{YX} \star \cdots \star f^{c_r}_{YX}$, we have that f^c_{YX} is the graph through the points $(x_1 + \cdots + x_i, y_{i+1} + \cdots + y_r)$ for $i = 0, 1, 2, \dots, r$.

6.4. Properties of the slope decomposition.

Lemma 6.10. *Suppose that $\|\cdot\|_X$ and $\|\cdot\|_Y$ are dual norms. If $c = c_1 + \cdots + c_m$ is an XY-slope decomposition, then $c_1, \ldots, c_m$ are linearly independent.*

Proof. Suppose that we have

$$c_m = \sum_{i=1}^{m-1} \lambda_i c_i.$$

Because $\mu_{XY}(c_i) = \|c_i\|_Y/\|c_i\|_X > \mu_{XY}(c_m) = \|c_m\|_Y/\|c_m\|_X$ we get

$$\begin{aligned}\|c_m\|_X\|c_m\|_Y = \langle c_m, c_m\rangle &= \sum_{i=1}^{m-1} \lambda_i \langle c_i, c_m\rangle \\ &\le \sum_{i=1}^{m-1} |\lambda_i| \|c_i\|_X\|c_m\|_Y < \sum_{i=1}^{m-1} |\lambda_i| \|c_i\|_Y\|c_m\|_X \le \|c_m\|_X\|c_m\|_Y.\end{aligned}$$

Contradiction. This proves that $c_1, \ldots, c_r$ are linearly independent. ∎

Proposition 6.11. *Suppose that $c = a + b$ is an $X2$-decomposition and a and b are tight. Then c is tight.*

Proof. Since a and b are tight, they have slope decompositions, say $a = a_1 + \cdots + a_r$ and $b = b_1 + \cdots + b_s$. We have

$$\begin{aligned}\|a\|_X\|b\|_Y = \langle a, b\rangle = \sum_{i=1}^{r}\sum_{j=1}^{s} \langle a_i, b_j\rangle &\le \sum_{i=1}^{r}\sum_{j=1}^{s} \|a_i\|_X\|b_j\|_Y \\ &= \left(\sum_{i=1}^{r} \|a_i\|_X\right)\left(\sum_{j=1}^{s} \|b_j\|_Y\right) \le \|a\|_X\|b\|_Y.\end{aligned}$$

It follows that $\langle a_i, b_j\rangle = \|a_i\|_X\|b_j\|_Y$ for all $i < j$. If $\mu_{XY}(a_r) > \mu_{XY}(b_1)$, then

$$c = a_1 + \cdots + a_r + b_1 + \cdots + b_s$$

is a slope decomposition.

Suppose that $\mu_{XY}(a_r) \le \mu_{XY}(b_1)$. We have

$$\begin{aligned}\|a_r\|_Y\|a_r + b_1\|_X &\ge \langle a_r, a_r + b_1\rangle \\ &= \langle a_r, a_r\rangle + \langle a_r, b_1\rangle = \|a_r\|_X\|a_r\|_Y + \|a_r\|_X\|b_1\|_Y \ge \|a_r\|_X\|a_r + b_1\|_Y,\end{aligned}$$

so $\mu_{XY}(a_r) \ge \mu_{XY}(a_r + b_1)$. Similarly, we have

$$\begin{aligned}\|a_r + b_1\|_Y\|b_1\|_X &\ge \langle a_r + b_1, b_1\rangle \\ &= \langle a_r, b_1\rangle + \langle b_1, b_1\rangle = \|a_r\|_X\|b_1\|_Y + \|b_1\|_X\|b_1\|_Y \ge \|a_r + b_1\|_X\|b_1\|_Y,\end{aligned}$$

so $\mu_{XY}(a_r + b_1) \geq \mu_{XY}(b_1) \geq \mu_{XY}(a_r) \geq \mu_{XY}(a_r + b_1)$. We conclude that $\mu_{XY}(a_r) = \mu_{XY}(b_1) = \mu_{XY}(a_r + b_1)$ and

$$c = a_1 + \cdots + a_{r-1} + (a_r + b_1) + b_2 + \cdots + b_s$$

is a slope decomposition.

Since c has a slope decomposition, it is tight. ∎

6.5. The unit ball of a tight norm.

Proposition 6.12. *A norm $\|\cdot\|_X$ is tight if and only if every face F of the unit ball B_X contains a unitangent vector that is perpendicular to F.*

Proof. Suppose that $\|\cdot\|_X$ is tight and that F is a face of B_X (other than B_X itself). Choose $c \in F$ in the relative interior. Then we have a slope decomposition

$$c = c_1 + c_2 + \cdots + c_r.$$

If $t_i = \|c_i\|_X / \|c\|_X$, then we have $t_1 + t_2 + \cdots + t_r = 1$ and

$$\frac{c}{\|c\|_X} = t_1 \frac{c_1}{\|c_1\|_X} + t_2 \frac{c_2}{\|c_2\|_X} + \cdots + \cdots + t_r \frac{c_r}{\|c_r\|_X}.$$

Now $c_1, \ldots, c_r \in F$ by Lemma 6.3. We have $\langle c_r, c\rangle = \|c\|_X \|c_r\|_Y = \|c_r\|_Y$. For any other vector $b \in B_X$ we have $\langle c_r, b\rangle \leq \|b\|_X \|c_r\|_Y \leq \|c_r\|_Y$. So the functional $\langle c_r, \cdot\rangle$ on F is maximal at c and therefore maximal and constant on the face F. It follows that c_r is perpendicular to F.

Now we show the converse. Suppose that every face F of the unit ball B_X contains a unitangent vector that is perpendicular to F. Suppose that $c \in V$ is a vector with $\|c\|_X = 1$. Let F be the smallest face of B_X that contains $c/\|c\|_X$. By induction on $\dim F$ we show that c is tight. The case $\dim F = 0$ is clear. Suppose that $\dim F > 0$. There exists a vector $b \in F$ that is unitangent and orthogonal to c. Choose t maximal such that $v = c + t(c - b) = (1+t)c - tb \in F$. Let $a' = \frac{1}{1+t} v$ and $b' = \frac{t}{t+1} b$. Then we have $c = a' + b'$. Since v lies in a face of smaller dimension, we know by induction that v and a' are tight. Because b' is unitangent, it is also tight. We have

$$\|a'\|_X = \frac{\|v\|_X}{t+1} = \frac{1}{t+1} = \frac{\|b\|_X}{t+1},$$
$$\|b'\|_Y = \frac{t\|b\|_Y}{t+1},$$
$$\langle a', b'\rangle = \frac{t\langle v, b\rangle}{(t+1)^2} = \frac{t\langle b, b\rangle}{(t+1)^2}.$$

So $\langle a', b'\rangle = \|a'\|_X \|b'\|_Y$. It follows that $c = a' + b'$ is an $X2$-decomposition. By Lemma 6.10, c is tight. ∎

Example 6.13. Consider again Examples 2.11 and 2.2. Suppose that $c = (c_1, \ldots, c_n)^t \in \mathbb{R}^n$, and define $\lambda_1 > \lambda_2 > \cdots > \lambda_r > 0$ by

$$\{|c_1|, |c_2|, \ldots, |c_n|\} \setminus \{0\} = \{\lambda_1, \ldots, \lambda_r\}.$$

Define m_i as the multiplicity of λ_i; i.e., m_i is the number of values of j for which $|c_j| = \lambda_i$. We define vectors $c^{(1)}, \dots, c^{(r)} \in \mathbb{R}^n$ as follows:

$$c^{(i)} = (c_1^{(i)}, \dots, c_n^{(i)})$$

and

$$c_j^{(i)} = \begin{cases} \operatorname{sgn}(c_j)(\lambda_i - \lambda_{i+1}) & \text{if } |c_j| \geq \lambda_i \text{ and} \\ 0 & \text{if } |c_j| < \lambda_i. \end{cases}$$

We use the convention $\lambda_{r+1} = 0$. We have

$$c = c^{(1)} + \cdots + c^{(r)}, \tag{4}$$

where $\|c^{(i)}\|_\infty = \lambda_i - \lambda_{i+1}$ and

$$\|c^{(i)}\|_1 = (m_1 + m_2 + \cdots + m_i)(\lambda_i - \lambda_{i+1}).$$

If $i < j$, then we have

$$\langle c^{(i)}, c^{(j)} \rangle = (m_1 + m_2 + \cdots + m_j)(\lambda_i - \lambda_{i+1})(\lambda_j - \lambda_{j+1}) = \|c^{(i)}\|_\infty \|c^{(i)}\|_1.$$

We have

$$\mu_{1\infty}(c^{(i)}) = \frac{\|c^{(i)}\|_\infty}{\|c^{(i)}\|_1} = \frac{1}{m_1 + m_2 + \cdots + m_i}.$$

This shows that (4) is a slope decomposition. So the norms $\|\cdot\|_\infty$ and $\|\cdot\|_1$ are tight.

7. The singular value decomposition of a matrix.

7.1. Matrix norms. In this section, we will study the singular value decomposition of a matrix using our terminology and the results we have obtained. Let $V = \mathbb{C}^{m \times n}$ be the space of $m \times n$-matrices with an inner product defined by $\langle A, B \rangle = \Re(\operatorname{trace}(B^\star A))$. The proof of the following well-known result will be useful for the discussion that follows.

Lemma 7.1. *The norms $\|\cdot\|_\sigma$ and $\|\cdot\|_\star$ are dual to each other.*

Proof. Let $A \in V$. We use the notation as before, where $\sqrt{A^\star A} = UDU^\star$, and D is the diagonal matrix whose diagonal entries are the singular values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ (where each singular value is listed as many times as its multiplicity). Let $u_1, u_2, \dots, u_n$ be the columns of U. These vectors form an orthonormal basis, and for any $B \in V$ we have

$$\begin{aligned} \langle A, B \rangle = \Re(\operatorname{trace}(B^\star A)) = \Re(\operatorname{trace}(U^\star B^\star A U)) &= \Re\Big(\sum_{i=1}^n \langle Au_i, Bu_i \rangle \Big) \\ &\leq \sum_{i=1}^n \|Au_i\|_2 \|Bu_i\|_2 = \sum_{i=1}^n \lambda_i \|B\|_\sigma = \|A\|_\star \|B\|_\sigma. \end{aligned} \tag{5}$$

Because $(AU)^\star(AU) = D^2$, the columns of AU are orthogonal. The matrices $W = AUD^{-1}$ are unitary. So $A = WDU^\star$. If $w_1, \dots, w_n$ are the columns of W, then the singular value decomposition of A is

$$\sum_{i=1}^n \lambda_i w_i u_i^\star.$$

Let r be maximal such that $\lambda_r \neq 0$, and define the block matrix

$$E = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

For $B = WEU^\star$ we have

$$\begin{aligned} (6)\quad \langle A, B\rangle &= \Re(\operatorname{trace}(B^\star A)) = \Re(\operatorname{trace}(UEW^\star A)) = \Re(\operatorname{trace}(EW^\star AU)) \\ &= \Re(\operatorname{trace}(ED)) = \Re(\operatorname{trace}(D)) = \lambda_1 + \cdots + \lambda_r = \|A\|_\star. \end{aligned}$$

From (5) and (6) follows that $\|\cdot\|_\star$ is the dual norm of $\|\cdot\|_\sigma$. ∎

7.2. Slope decomposition for matrices. Suppose that C is a complex $m \times n$ matrix, and let $\lambda_1 > \lambda_2 > \cdots > \lambda_r > 0$ be the nonsingular values of C with multiplicities $m_1, m_2, \ldots, m_r$, respectively. We can write $C = WDU^\star$, where U, W are unitary and D is of the form

$$\left(\begin{array}{cccc|c} \lambda_1 I_{m_1} & & & & 0 \\ & \lambda_2 I_{m_2} & & & 0 \\ & & \ddots & & \vdots \\ & & & \lambda_r I_{m_r} & 0 \\ \hline 0 & 0 & \cdots & 0 & 0 \end{array}\right)$$

where I_d is the $d \times d$ identity matrix and the zeros represent possible empty zero blocks. Recall that the norms can be expressed by $\|C\|_\star = \sum_{i=1}^r m_i\lambda_i$, $\|C\|_\sigma = \lambda_1$, and $\|C\|_2 = \sqrt{\sum_{i=1}^r m_i\lambda_i^2}$. Define

$$C^{(j)} = (\lambda_j - \lambda_{j+1}) W \begin{pmatrix} I_{k_j} & 0 \\ 0 & 0 \end{pmatrix} U^\star$$

for $j = 1, 2, \ldots, r$, where $k_j = m_1 + m_2 + \cdots + m_j$ and $\lambda_{r+1} = 0$. We have

$$(7)\qquad C = C^{(1)} + \cdots + C^{(r)}.$$

Proposition 7.2. *The expression (7) is a slope decomposition. In particular, the spectral and the nuclear norms are tight.*

Proof. We have

$$\begin{aligned} \|C^{(j)}\|_Y &= \|A^{(j)}\|_\sigma = \lambda_j - \lambda_{j+1}, \\ \|C^{(j)}\|_X &= \|C^{(j)}\|_\star = k_j(\lambda_j - \lambda_{j+1}), \\ \mu_{XY}(C^{(j)}) &= \frac{1}{k_j}. \end{aligned}$$

In particular, $\mu_{XY}(C^{(j)})$ is strictly decreasing as j increases.

If $i \leq j$, then we have

$$\begin{aligned}\langle C^{(i)}, C^{(j)}\rangle &= \Re\Big(\operatorname{trace}\left((C^{(i)})^\star C^{(j)}\right)\Big)\\ &= (\lambda_i-\lambda_{i+1})(\lambda_j-\lambda_{j+1})\Re\Big(\operatorname{trace}\left(U\begin{pmatrix} I_{k_i} & 0\\ 0 & 0\end{pmatrix}W^\star W\begin{pmatrix} I_{k_j} & 0\\ 0 & 0\end{pmatrix}U^\star\right)\Big)\\ &= (\lambda_i-\lambda_{i+1})(\lambda_j-\lambda_{j+1})\Re\Big(\operatorname{trace}\left(U\begin{pmatrix} I_{k_i} & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix} I_{k_j} & 0\\ 0 & 0\end{pmatrix}U^\star\right)\Big)\\ &= (\lambda_i-\lambda_{i+1})(\lambda_j-\lambda_{j+1})\Re\Big(\operatorname{trace}\left(U\begin{pmatrix} I_{k_i} & 0\\ 0 & 0\end{pmatrix}U^\star\right)\Big)\\ &= k_i(\lambda_i-\lambda_{i+1})(\lambda_j-\lambda_{j+1}) = \|C^{(i)}\|_\star\|C^{(j)}\|_\sigma.\end{aligned}$$

This proves that (7) is the slope decomposition. ∎

7.3. Principal component analysis. In principal component analysis, one finds a low rank matrix that approximates the matrix A by truncating the singular value decomposition. For a given threshold y, let s be maximal such that $\lambda_s > y$. Then

$$C' = W\left(\begin{array}{cccc|c} \lambda_1 I_{m_1} & & & & 0\\ & \lambda_2 I_{m_2} & & & 0\\ & & \ddots & & \vdots\\ & & & \lambda_r I_{m_r} & 0\\ \hline 0 & 0 & \cdots & 0 & 0\end{array}\right)U^\star$$

is a low rank approximation of C. This method is called *hard thresholding*. Replacing C by the approximation C' is an effective way to reduce the dimension of a large scale problem. Let us compare this to the XY-decomposition (or equivalently the $X2$- or $2Y$-decomposition) of C. Let us define

$$A = W\left(\begin{array}{cccc|c} (\lambda_1-y) I_{m_1} & & & & 0\\ & (\lambda_2-y) I_{m_2} & & & 0\\ & & \ddots & & \vdots\\ & & & (\lambda_s-y) I_{m_s} & 0\\ \hline 0 & 0 & \cdots & 0 & 0\end{array}\right)U^\star$$

and

$$B = C - A = W\left(\begin{array}{ccccccc|c} yI_{m_1} & & & & & & & 0\\ & \ddots & & & & & & \vdots\\ & & yI_{m_s} & & & & & 0\\ & & & \lambda_{s+1}I_{m_{s+1}} & & & & 0\\ & & & & \ddots & & & \vdots\\ & & & & & & \lambda_r I_{m_r} & 0\\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0\end{array}\right)U^\star.$$

Lemma 7.3. *The expression $C = A + B$ is an XY-decomposition.*

Proof. We have $\|A\|_\star = \|A\|_X = \sum_{i=1}^s m_i(\lambda_i - y)$ and $\|B\|_\sigma = \|B\|_Y = y$. Now the lemma follows from

$$\langle A, B\rangle = \sum_{i=1}^{s} m_i(\lambda_i - y)y = \|A\|_\star \|B\|_\sigma = \|A\|_X \|B\|_Y. \qquad \blacksquare$$

In particular, $A = \operatorname{shrink}_Y(C, y)$. The operator $\operatorname{shrink}_Y(\cdot, y)$ is soft thresholding with threshold level y (see [20]). Unlike hard thresholding, soft thresholding is continuous. The Pareto frontier (which is also the subfrontier) is given by

$$f^C_{XY}(y) = h^C_{XY}(y) = m_i \sum_{i=1}^{r} m_i \max\{\lambda_i - y, 0\}.$$

7.4. The singular value region for matrices. The Pareto frontier of C encodes the singular values. We will describe in detail how to obtain the singular values from the Pareto frontier. Note that we consider $x = h_{XY}(y)$ as a function of y, rather then considering its inverse function $y = h_{YX}(x)$. If we differentiate $h^C_{XY}(y)$ with respect to y, we get

$$(h^C_{XY}(y))' = -\sum_{i=1}^{s} m_i$$

if $\lambda_s > y > \lambda_{s+1}$ and $1 \le s \le r$ (with the convention that $\lambda_{r+1} = 0$).

If we plot y against $x = -(h^C_{XY}(y))'$, then we get the singular value region. From the descriptions above, it is now clear that this region can be described as an $m_1 \times \lambda_1$ bar, followed by an $m_2 \times \lambda_2$ bar, etc. So the singular value region from Definition 2.20 is the same as the singular value region for matrices as described in the introduction.

7.5. Rank minimization and low rank matrix completion. Let $V = \mathbb{C}^{m\times n}$ be the set of $m \times n$ matrices. We will study the low rank matrix completion from the viewpoint of competing dual norms.

Problem 7.4 (low rank matrix completion (LRMC)). *Given an $m \times n$-matrix C where the entries $(i_1, j_1), \ldots, (i_s, j_s)$ are missing, fill in the missing entries such that the resulting matrix has minimal rank.*

The LRMC problem has applications in collaborative filtering and recommender systems such as the Netflix problem. In the Netflix problem, we are given a partially filled matrix. The rows of the matrix correspond to the Netflix user, the columns of the matrix correspond to movies, and the entry at position (i, j) is the rating of the jth movie by the ith user. If the (i, j) entry is missing, then the ith user did not rate the jth movie. In the Netflix problem one would like to estimate this entry, meaning that we would like to know whether this user would like the movie. The assumption is that there are only a few factors that determine whether a person likes a certain movie (for example, genre, or the actors that are in the movie), and this means that the completed matrix has low rank.

The LRMC problem is a special case of the rank minimization (RM) problem.

Problem 7.5 (rank minimization (RM)). *Suppose that W is a subspace of V, and let $A \in V$. Find a matrix $B \in A + W$ of minimal rank.*

The LRMC problem can be formulated as an RM problem as follows. Complete C to a matrix A in some way (for example, set all the missing entries equal to 0). Then, let $W \subseteq V$ be the subspace spanned by all matrices e_{i_k,j_k}, $k = 1, 2, \ldots, s$. Here $e_{p,q}$ is the matrix with all 0's except for a 1 in position (p, q). Find $B \in A + W$ with minimal rank using RM. Then B is also the solution to the LRMC problem.

The LRMC problem and the RM problem are difficult to solve. However, there are heuristic approaches using optimization that work well in practice. For this we use the philosophy of convex relaxation. The RM problem is very noncovex, and we modify the problem by a similar convex problem that can more easily be solved. Instead of RM, we consider the following problem (see [12, 15, 45]).

Problem 7.6. *Find a matrix $D \in C + W$ with $\|D\|_\star$ minimal.*

Let Z be the orthogonal complement of W, and let π_Z be the orthogonal projection onto Z. The problem does not change when we replace C by $\pi_Z(C)$, so we may assume that $C \in Z$ without loss of generality. We define a norm $\|\cdot\|_X$ on Z by

$$\|C\|_X = \min\{\|D\|_\star \mid D \in C + W\}.$$

So Problem 7.6 is essentially the problem of determining the value of $\|C\|_X$.

In the presence of noise, we would like to find a matrix A such that $\|A\|_X$ and $\|C - A\|_2$ are small. This problem can be formulated in the context of the Pareto subfrontier. Namely, this is exactly the optimization problem $\mathbf{M}^C_{X2}$ (or $\mathbf{M}^C_{2X}$).

The dual norm to $\|\cdot\|_X$ on Z is given by $\|B\|_Y = \|B\|_\sigma$. So by dualizing the problem $\mathbf{M}^C_{X2}$ we get the optimization problem $\mathbf{M}^C_{2Y}$ (or $\mathbf{M}^C_{Y2}$). This shows that the convex relaxation of the rank minimization with noise can be formulated in our general framework of denoising. Optimal solutions correspond to $2Y$ decompositions, where $\|\cdot\|_Y = \|\cdot\|_\sigma$ is the restriction of the spectral norm to the subspace $Z \subseteq \mathbb{C}^{m\times n}$.

8. Restricting norms. Suppose that V is a finite dimensional $\mathbb{R}$-vector space with a positive definite bilinear form $\langle\cdot,\cdot\rangle$, and $\|\cdot\|_X$ is a norm on V. For a subspace W of V, it is natural to ask whether the $X2$-decompositions of vectors in W are always within the space W. In this section we will give a sufficient criterion for W to have this property.

Definition 8.1. *A subspace $W \subseteq V$ is called a* nice slice *if we have $\|\pi_W(c)\|_X \leq \|c\|_X$ for all $c \in V$, where $\pi_W : V \to W$ is the orthogonal projection.*

Lemma 8.2. *If W is a nice slice, then we also have $\|\pi_W(c)\|_Y \leq \|c\|_Y$ for all $c \in V$, where $\|\cdot\|_Y$ is the norm dual to $\|\cdot\|_X$.*

Proof. Choose a vector d with $\|d\|_X = 1$ and $\langle d, \pi_W(c)\rangle = \|\pi_W(c)\|_Y$. We have

$$\|\pi_W(c)\|_Y = \langle d, \pi_W(c)\rangle = \langle \pi_W(d), c\rangle \leq \|\pi_W(d)\|_X\|c\|_Y \leq \|d\|_X\|c\|_Y = \|c\|_Y. \qquad \blacksquare$$

Let $\mathrm{O}(V)$ be the orthogonal group consisting of all $g \in \mathrm{GL}(V)$ with the property

$$\langle g \cdot v, g \cdot w\rangle = \langle v, w\rangle$$

for all $v, w \in V$.

Lemma 8.3. *Suppose that $G \subseteq \mathrm{O}(V)$ is a subgroup with the properties $\|g \cdot v\|_X = \|v\|_X$ for all $v, w \in V$ and all $g \in G$. Then the space $V^G = \{v \in V \mid \forall g \in G\ g \cdot v = v\}$ of G-invariant vectors is a nice slice.*

Proof. We can replace G with its closure, so without loss of generality we may assume that G is a compact Lie group. Let π_W be the projection onto $W := V^G$. We have

$$\pi_W(v) = \int_{g \in G} g \cdot v \, d\mu,$$

where $d\mu$ is the normalized Haar measure. This shows that $\pi_W(v)$ lies in the convex hull of all $g \cdot v$, $g \in G$. Since $\|g \cdot v\|_X \le \|v\|_X$ for all $v \in V$, we also have $\|\pi_W(v)\|_X \le \|v\|_X$ for all $g \in G$. ∎

For the remainder of this section, $W \subseteq V$ is a nice slice, and $\|\cdot\|_{X'}$ and $\|\cdot\|_{Y'}$ are the restrictions of the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ to W.

Lemma 8.4. *The norms $\|\cdot\|_{X'}$ and $\|\cdot\|_{Y'}$ are also dual to each other.*

Proof. If $v, u \in W$ and $\|u\|_Y = 1$, then we have

$$\langle v, u \rangle \le \|v\|_X \|u\|_Y = \|v\|_{X'}.$$

For a given $v \in W$, there exists $w \in V$ with $\|w\|_Y = 1$ and

$$\langle v, w \rangle = \|v\|_X = \|v\|_{X'}.$$

We get

$$\langle v, \pi_W(w) \rangle = \langle \pi_W(v), w \rangle = \langle v, w \rangle = \|v\|_{X'}.$$

Define $u = \pi_W(w)/\|\pi_W(w)\|_{Y'}$. We have $\|u\|_{Y'} = 1$. It follows that

$$\|v\|_{X'} \ge \langle v, u \rangle = \frac{\langle v, \pi_W(w) \rangle}{\|\pi_W(w)\|_{Y'}} = \frac{\|v\|_{X'}}{\|\pi_W(w)\|_{Y'}} \ge \|v\|_{X'},$$

because $\|\pi_W(w)\|_{Y'} = \|\pi_W(w)\|_Y \le \|w\|_Y = 1$. So $\langle v, u \rangle = \|v\|_{X'}$. ∎

Lemma 8.5. *Suppose that $c \in W$.*

(1) *If $c = a + b$ is an $X2$-decomposition, then $a, b \in W$ and $c = a + b$ is an $X'2$-decomposition.*
(2) *If $c = a + b$ is an XY-decomposition, then $c = \pi_W(a) + \pi_W(b)$ is an $X'Y'$-decomposition, $\|\pi_W(a)\|_{X'} = \|a\|_X$, and $\|\pi_W(b)\|_{Y'} = \|b\|_Y$.*

Proof. (1) If $c = a + b$ is an $X2$-decomposition, then we have

$$c = \pi_W(c) = \pi_W(a) + \pi_W(b),$$

$\|\pi_W(a)\|_X \le \|a\|_X$, and $\|\pi_W(b)\|_2 \le \|b\|_2$. It follows that $c = \pi_W(a) + \pi_W(b)$ is also an $X2$-decomposition and by uniqueness we have $\pi_W(a) = a$ and $\pi_W(b) = b$. Suppose that $c = a' + b'$ with $a', b' \in W$ and $\|b'\|_2 = \|b\|_2$. We get

$$\|a'\|_{X'} = \|a'\|_X \ge \|a\|_X = \|a\|_{X'}$$

because $c = a+b$ is an $X2$-decomposition. This shows that $c = a+b$ is an $X'2$-decomposition.

(2) Suppose that $c = a + b$ is an XY-decomposition. We get

$$c = \pi_W(c) = \pi_W(a) + \pi_W(b),$$

$\|\pi_W(a)\|_X \leq \|a\|_X$, and $\|\pi_W(b)\|_Y \leq \|b\|_Y$. It follows that $c = \pi_W(a) + \pi_W(b)$ is also an XY-decomposition. An argument similar to that in (1) shows that this is also an $X'Y'$-decomposition. Since $\|\pi_W(a)\|_X \leq \|a\|_X$, we must have $\|b\|_Y \geq \|\pi_W(b)\|_Y \geq \|b\|_Y$, because $c = a + b$ is an XY-decomposition. It follows that

$$\|\pi_W(b)\|_{Y'} = \|\pi_W(b)\|_Y = \|b\|_Y.$$

Similarly, we get $\|\pi_W(a)\|_{X'} = \|a\|_X$. ∎

In particular, we have $f^c_{YX} = f^c_{Y'X'}$ and $h^c_{YX} = h^c_{Y'X'}$.

Example 8.6. Suppose that $V = \mathbb{C}^{m\times n}$ and $W = \mathbb{R}^{m\times n} \subseteq V$. The orthogonal projection $\pi_W : V \to W$ is given by $\pi_W(A) = \Re(A)$. Suppose that $A = A_1 + A_2 i \in \mathbb{C}^{m\times n}$ where $A_1, A_2 \in \mathbb{R}^{m\times n}$. Choose a unit vector $v \in \mathbb{R}^n$ such that $\|\pi_W(A)v\|_2 = \|\pi_W(A)\|_\sigma$. Then we have

$$\|\pi_W(A)\|_\sigma = \|\pi_W(A)v\|_2 = \|A_1 v\|_2 \leq \sqrt{\|A_1 v\|_2^2 + \|A_2 v\|_2^2} = \|Av\|_2 \leq \|A\|_\sigma \|v\|_2 = \|A\|_\sigma.$$

This shows that W is a nice slice.

9. 1D total variation denoising.

9.1. The total variation norm and its dual. In this section we discuss the application to 1D total variation denoising. This example is particularly interesting because the corresponding norms are tight.

Define the difference map $D : \mathbb{R}^{n+1} \to \mathbb{R}^n$ by

$$D(b) = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n+1} \end{pmatrix} = \begin{pmatrix} b_2 - b_1 \\ b_3 - b_2 \\ \vdots \\ b_{n+1} - b_n \end{pmatrix}.$$

The map D is surjective, and the kernel is spanned by the vector $\mathbf{1} = (1\ 1 \cdots 1)^t$. The dual map $D^\star : \mathbb{R}^n \to \mathbb{R}^{n+1}$ is given by

$$D^\star \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} -a_1 \\ a_1 - a_2 \\ \vdots \\ a_{n-1} - a_n \\ a_n \end{pmatrix}.$$

If we compose the two maps, we get

$$DD^\star(a) = \begin{pmatrix} 2a_1 - a_2 \\ 2a_2 - a_1 - a_3 \\ 2a_3 - a_2 - a_4 \\ \vdots \\ 2a_{n-1} - a_{n-2} - a_n \\ 2a_n - a_{n-1} \end{pmatrix}.$$

The linear map $DD^\star$ is invertible. Let $V \subset \mathbb{R}^{n+1}$ be the subspace defined by

$$V = \{a \in \mathbb{R}^{n+1} \mid a_1 + \cdots + a_{n+1} = 0\}.$$

The image of $D^\star$ is V. If $b = D^\star(DD^\star)^{-1}a$, then b is the vector of minimal length with the property $Db = a$.

We define a norm $\|\cdot\|_X$ on $\mathbb{R}^n$ by

$$\|a\|_X = \|D^\star a\|_1.$$

Another norm $\|\cdot\|_Y$ on $\mathbb{R}^n$ is given by

$$\|a\|_Y = \min\{\|b\|_\infty \mid b \in \mathbb{R}^{n+1} \text{ and } Db = a\}.$$

Lemma 9.1. *Suppose that b is a vector with $Db = a$. Let $m_+(b) = \max\{b_1, \ldots, b_{n+1}\}$ and $m_-(b) = \min\{b_1, \ldots, b_{n+1}\}$. Then we have $\|a\|_Y = \frac{1}{2}(m_+(b) - m_-(b))$.*

Proof. The vectors that map to a under D are of the form $b - \lambda\mathbf{1}$. We have $\|b - \lambda\mathbf{1}\|_\infty = \max\{m_+ - \lambda, m_- - \lambda\}$. This quantity is minimal if $\lambda = \frac{1}{2}(m_+(b) + m_-(b))$. In that case we have $\|b - \lambda\mathbf{1}\|_\infty = \frac{1}{2}(m_+(b) - m_-(b))$. ∎

Lemma 9.2. *The norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ are dual to each other.*

Proof. Suppose that $a, b \in \mathbb{R}^n$. Choose $e \in \mathbb{R}^{n+1}$ such that $De = b$ and $\|e\|_\infty = \|b\|_Y$. Then we have

$$\langle a, b\rangle = \langle a, De\rangle = \langle D^\star a, e\rangle \le \|D^\star a\|_1 \|e\|_\infty = \|a\|_X \|b\|_Y.$$

Suppose that a is nonzero, and let $c = D^\star a \neq 0$. Define

$$e = \begin{pmatrix} \operatorname{sgn}(c_1) \\ \operatorname{sgn}(c_2) \\ \vdots \\ \operatorname{sgn}(c_{n+1}) \end{pmatrix}.$$

Because $c_1 + \cdots + c_{n+1} = 0$, the set $\{c_1, \ldots, c_{n+1}\}$ has positive and negative elements. This implies that $m_+(e) = 1$ and $m_-(e) = -1$. If $b = De$, then we have $\|b\|_Y = \frac{1}{2}(m_+(e) - m_-(e)) = 1$.

$$\langle a, b\rangle = \langle a, De\rangle = \langle D^\star a, e\rangle = \langle c, e\rangle = \|c\|_1 = \|a\|_X.$$

This shows that the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ are dual. ∎

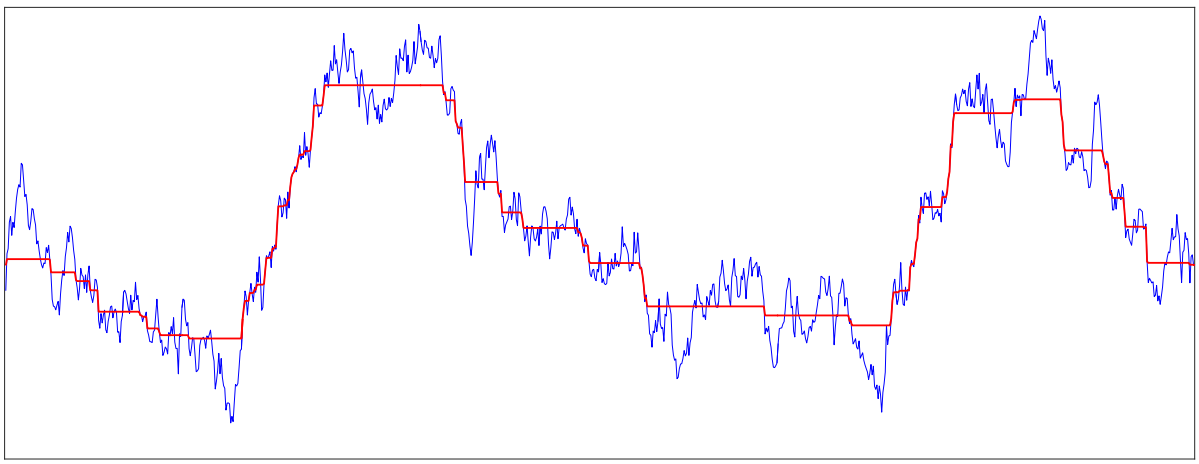

Figure 14.

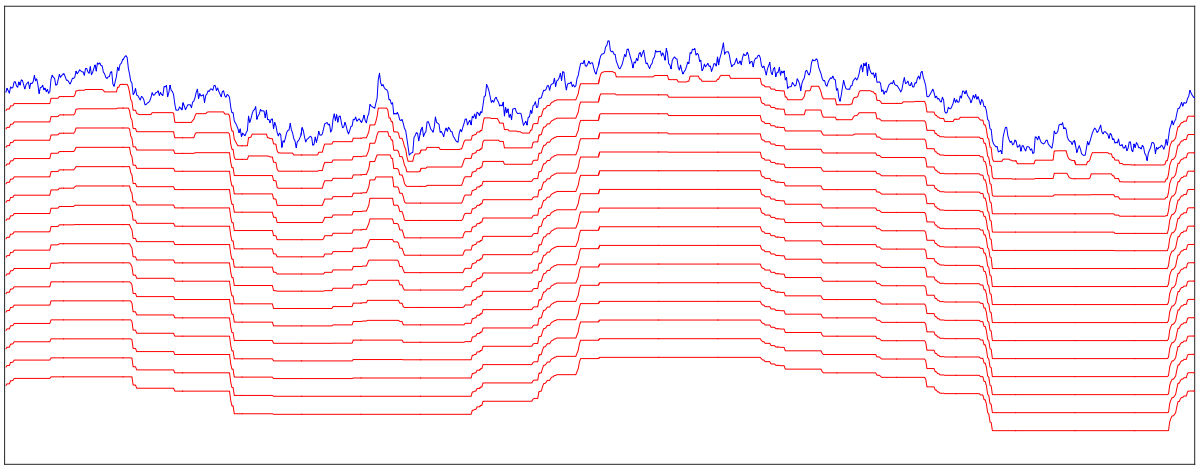

Figure 15.

For a vector $a \in \mathbb{R}^n$, $\|D^\star a\|_1$ is its total variation. Given a signal c and an $\varepsilon > 0$, a solution a to the problem $\mathbf{M}_{X2}(\varepsilon)$ minimizes the total variation $\|D^\star a\|_1$ under the constraint $\|c - a\|_2 \leq \varepsilon$. The function a is typically a piecewise constant function. Figure 14 shows an example where the blue function is c and the red function is a.

As we increase the value of ε, the sparsity decreases. In Figure 15 we draw the signal c in blue and (a vertical translation of) the denoised signal $a = \text{shrink}_Y(\varepsilon)$ for various values of ε in red.

9.2. Description of the unit ball. We now will describe the unit balls B_X and B_Y.

Definition 9.3. *A vector* $(s_1\ s_2\ \cdots\ s_{n+1})^t$ *is called a* signature sequence *if* $s_1, s_2, \ldots, s_{n+1} \in \{-1, 0, 1\}$ *and* $\{1, -1\} \subseteq \{s_1, \ldots, s_{n+1}\}$. *For a signature sequence* s, *we define* $\widetilde{F}_s$ *as the set of all vectors* $x \in \mathbb{R}^{n+1}$ *such that* $x_i = s_i$ *when* $s_i = \pm 1$, *and* $|x_i| \leq 1$ *if* $s_i = 0$. *We define* $F_s = D(\widetilde{F}_s)$.

Lemma 9.4. *The set* F_s *is a face of the unit ball* B_Y.

Proof. The set $\widetilde{F}_s$ is a face of the unit ball $B_1 \subset \mathbb{R}^{n+1}$ for the ℓ_1 norm. Suppose that $b = tb_1 + (1-t)b_2 \in F_s$ with $b_1, b_2 \in B_Y$ and $0 < t < 1$. Then we can write $b_i = D(\widetilde{b}_i)$ with $\|\widetilde{b}_i\|_\infty \leq 1$ for $i = 1, 2$. We have

$$D(t\widetilde{b}_1 + (1-t)\widetilde{b}_2) = tb_1 + (1-t)b_2 = b = D\widetilde{b}$$

for some $\widetilde{b} \in \widetilde{F}_s$. We get that $\widetilde{b}$ and $t\widetilde{b}_1 + (1-t)\widetilde{b}_2$ differ by a multiple of $\mathbf{1}$. Since the maximum and minimum entries of $\widetilde{b}$ are 1 and -1, respectively, and $\|t\widetilde{b}_1 + (1-t)\widetilde{b}_2\|_\infty \leq 1$, we deduce that $\widetilde{b} = t\widetilde{b}_1 + (1-t)\widetilde{b}_2$. Now $\widetilde{b}_1, \widetilde{b}_2$ lie in the unit ball B_1, and $\widetilde{b}$ lies in the face $\widetilde{F}_s$. It follows that $\widetilde{b}_1, \widetilde{b}_2 \in \widetilde{F}_s$ and $b_1 = D(\widetilde{b}_1), b_2 = D(\widetilde{b}_2) \in F_s$. ∎

The dimension of F_s is equal to the number of 0's in the signature sequence s. The restriction of D to $\widetilde{F}_s$ is injective, so F_s and $\widetilde{F}_s$ have the same dimension.

Proposition 9.5. *The faces of* B_Y *of dimension* $< n$ *are exactly all* F_s, *where* s *is a signature sequence.*

Proof. Suppose that F is a proper face of the polytope B_Y. Then there exists a vector $a \in \mathbb{R}^n$ with $\|a\|_X = 1$ and $F = \{b \in B_Y \mid \langle a, b\rangle = 1\}$. Let $\widetilde{a} = D^\star a$ and $s_i = \operatorname{sgn}(\widetilde{a}_i)$ for $i = 1, 2, \ldots, n+1$. Since $\widetilde{a} \neq 0$, and $\widetilde{a}_1 + \cdots + \widetilde{a}_{n+1} = 0$, the vector $\widetilde{a}$ must have positive and negative coordinates. In the sequence $s_1, \ldots, s_{n+1}$ the elements 1 and -1 both must appear. For a vector $b \in \mathbb{R}^n$ with $\|b\|_Y = 1$, let $\widetilde{b} \in \mathbb{R}^{n+1}$ be the unique vector with $D\widetilde{b} = b$ and $\|\widetilde{b}\|_\infty = 1$. Then we have

$$\langle a, b\rangle = \langle a, D\widetilde{b}\rangle = \langle D^\star a, \widetilde{b}\rangle = \langle \widetilde{a}, \widetilde{b}\rangle$$

with $\|\widetilde{a}\|_1 = \|\widetilde{b}\|_\infty = 1$. Now $\langle \widetilde{a}, \widetilde{b}\rangle = \sum_{i=1}^{n+1} \widetilde{a}_i \widetilde{b}_i = 1$ if and only if for every i, $\widetilde{a}_i = s_i = 0$ or $\widetilde{b}_i = s_i$. In other words, $\langle \widetilde{a}, \widetilde{b}\rangle = 1$ if and only if $\widetilde{b} \in \widetilde{F}_s$. So $\langle a, b\rangle = \langle \widetilde{a}, \widetilde{b}\rangle = 1$ if and only if $b = D\widetilde{b} \in F_s$. ∎

Theorem 9.6. *The norms* $\|\cdot\|_X$ *and* $\|\cdot\|_Y$ *are tight.*

Proof. Suppose that s is a signature vector. It suffices to construct a unitangent vector $u \in F_s$ by Proposition 6.12. We define a vector $v \in \mathbb{R}^{n+1}$ as follows. If $s_i = \pm 1$, then $v_i = s_i$. The other coordinates of v are obtained by linear interpolation. If $i < j$, $s_i, s_j \neq 0$, and $s_{i+1} = \cdots = s_{j-1} = 0$, then we define

$$v_k = \left(\frac{k-i}{j-i}\right) s_i + \left(\frac{j-k}{j-i}\right) s_j$$

whenever $i < k < j$. If $s_1 = s_2 = \cdots = s_{i-1} = 0$ and $s_i \neq 0$, then we define $v_k = s_i$ for $k < i$. If $s_{j+1} = s_{j+2} = \cdots = s_{n+1} = 0$ and $s_j \neq 0$, then we define $v_k = s_j$ for $k > j$. For example, if $s = (0, 0, 1, 0, 0, 0, -1, 0, -1, 0, 0, 1)$, then $v = (1, 1, 1, \frac{1}{2}, 0, -\frac{1}{2}, -1, -1, -1, -\frac{1}{3}, \frac{1}{3}, 1)$.

From the construction follows that for every i we have $(D^\star Dv)_i s_i = |(D^\star Dv)_i|$. We define $u = Dv$. We have $\|u\|_Y = \|v\|_\infty = 1$. From the construction it is clear that $v \in \widetilde{F}_s$ and $u \in F_s$. We have

$$\langle u, u\rangle = \langle Dv, Dv\rangle = \langle D^\star Dv, v\rangle = \sum_{i=1}^{n+1} (D^\star Dv)_i v_i = \sum_{i=1}^{n+1} |(D^\star u)_i| = \|u\|_X = \|u\|_X \|u\|_Y,$$

so u is unitangent. ∎

We briefly discuss the combinatorics of the polytope B_Y. Let h_i be the number of faces of B_Y of dimension i. For $i < n$, h_i is the number of signature sequences with i zeros. We also have $h_n = 1$. The generating function for $h_0, h_1, \ldots, h_n$ is $H(t) = h_0 + h_1 t + \cdots + h_n t^n$. The generating function for the set $\{1, 0, -1\}^{n+1}$ is $(2+t)^{n+1}$. The generating function for $\{1, 0\}^{n+1}$ and for $\{-1, 0\}^{n+1}$ is $(1+t)^{n+1}$. The generating function for $\{(0, 0, \ldots, 0)\}$ is t^{n+1}. So the generating function for the set of signature sequences, using inclusion-exclusion, is $(2+t)^{n+1} - 2(1+t)^{n+1} + t^{n+1}$. There is one face of dimension n that does not correspond to a signature sequences, so we have

$$H(t) = (2+t)^{n+1} - 2(1+t)^{n+1} + t^{n+1} + t^n.$$

In particular, $h_0 = 2^{n+1} - 2$ is the number of vertices of the ball B_Y, and $h_{n-1} = 4\binom{n+1}{2} - 2\binom{n+1}{2} = n^2 + n$ is the number of facets of B_Y, which is the number of vertices of B_X. The total number of faces of B_Y (and B_X) is $H(1) = 3^{n+1} - 2^{n+2} + 2$.

Example 9.7. Let $n = 3$. We have the following signature sequences and corresponding

unitangent vectors (written as row vectors):

$(1,1,1,-1),(1,0,1,-1),(0,1,1,-1),(0,0,1,-1)$	$(0,0,-2)^*$
$(1,1,0,-1),(0,1,0,-1)$	$(0,-1,-1)$
$(1,1,-1,1),(0,1,-1,1)$	$(0,-2,2)^*$
$(1,1,-1,-1),(1,1,-1,0),(0,1,-1,-1),(0,1,-1,0)$	$(0,-2,0)^*$
$(1,0,0,-1)$	$(-\frac{2}{3},-\frac{2}{3},-\frac{2}{3})$
$(1,0,-1,1)$	$(-1,-1,2)$
$(1,0,-1,-1),(1,0,-1,0)$	$(-1,-1,0)$
$(1,-1,1,1),(1,-1,1,0)$	$(-2,2,0)^*$
$(1,-1,1,-1)$	$(-2,2,-2)^*$
$(1,-1,0,1)$	$(-1,-1,1)$
$(1,-1,-1,1)$	$(-2,0,2)^*$
$(1,-1,-1,-1),(1,-1,-1,0),(1,-1,0,-1),(1,-1,0,0)$	$(-2,0,0)^*$
$(-1,1,1,1),(-1,1,0),(-1,1,0,1),(-1,1,0,0)$	$(2,0,0)^*$
$(-1,1,1,-1)$	$(2,0,-2)^*$
$(-1,1,0,-1)$	$(1,1,-1)$
$(-1,1,-1,1)$	$(2,-2,2)^*$
$(-1,1,-1,-1),(-1,1,-1,0)$	$(2,-2,0)^*$
$(-1,0,1,1),(-1,0,1,0)$	$(1,1,0)$
$(-1,0,1,-1)$	$(1,1,-2)$
$(-1,0,0,1)$	$(\frac{2}{3},\frac{2}{3},\frac{2}{3})$
$(-1,-1,1,1),(-1,-1,1,0),(0,-1,1,1),(0,-1,1,0)$	$(0,2,0)^*$
$(-1,-1,1,-1),(0,-1,1,-1)$	$(0,2,-2)^*$
$(-1,-1,0,1),(0,-1,0,1)$	$(0,1,1)$
$(-1,-1,-1,1)(-1,0,-1,1),(0,-1,-1,1),(0,0,-1,1)$	$(0,0,2)^*$

Every vertex of the unit ball B_Y is unitangent. These vectors are marked with $\star$, and they correspond to signature sequences that have no zeros. Figure 16 shows a 2D projection of the unit ball B_Y.

The unit ball B_X is dual to the polytope B_Y and is shown in Figure 17.

An example of an XY-slope decomposition is

$$\begin{pmatrix} 5 \\ -2 \\ 3 \end{pmatrix} = \frac{3}{4}\begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} + \frac{5}{4}\begin{pmatrix} 2 \\ -2 \\ 2 \end{pmatrix}.$$

If $a \in \mathbb{R}^n$, then we have

$$\operatorname{gsparse}_Y(a) = |\{i \mid 1 \le i \le n-1,\ a_i \ne a_{i+1}\}| + 1.$$

9.3. The taut string method. The restriction of $D : \mathbb{R}^{n+1} \to \mathbb{R}^n$ to V gives an isomorphism between V and $\mathbb{R}^n$. Now we can view the bilinear form $\langle \cdot, \cdot \rangle$ as a bilinear form on V, and for $a, b \in V$ we have

$$\langle a, b \rangle = \langle Da, Db \rangle = \sum_{i=1}^{n} (a_{i+1} - a_i)(b_{i+1} - b_i).$$

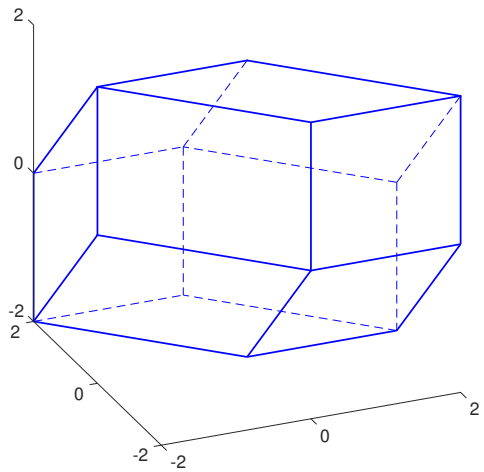

Figure 16.

In particular, we have

$$\|a\|_2 = \sqrt{\langle a, a\rangle} = \sqrt{\sum_{i=1}^{n}(a_{i+1} - a_i)^2}.$$

We can also view $\|\cdot\|_X$ and $\|\cdot\|_Y$ as norms on V, and for $c = (c_1, \ldots, c_{n+1}) \in V$ we have

$$\|c\|_X = \|D^\star Dc\|_1$$

and

$$\|c\|_Y = \tfrac{1}{2}(\max_i\{c_i\} - \min_i\{c_i\}).$$

For $c \in \mathbb{R}^{n+1}$, we consider the following optimization problem.

TS$^c(\varepsilon)$. *Find a vector $a \in \mathbb{R}^{n+1}$ with $\|c - a\|_\infty \leq \varepsilon$ such that $\|D^\star Da\|_1$ is minimal.*

We call this the Taut String problem. This problem, and some generalizations to higher order, were studied in [39]. If the vectors a, b, c are discretized functions, then the graph a lies between $c - \varepsilon$ and $c + \varepsilon$. Now $D^\star Da$ is a discrete version of the second derivative. The value $|(D^\star Da)_i|$ is a measure of how much the graph bends at vertex i. So $\|D^\star Da\|_1$ is the total amount of bending, and we try to minimize this. Visually we can see a as a string between $c - \varepsilon$ and $c + \varepsilon$, and we pull a on both ends so that the string is taut. The *Taut String Algorithm* described in [18] computes a in time $O(n)$ (see also [11]). The function a is piecewise linear, and $D(a)$ is piecewise constant. Total variation denoising of time signals has applications in statistics to estimate a density function from a collection of measurements (see [3]). It was also used in [5, 40] for analyzing heart rate variability signals to predict

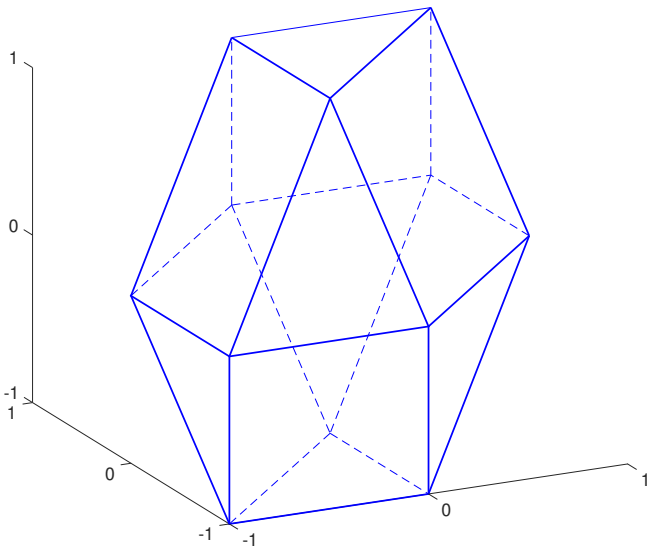

Figure 17.

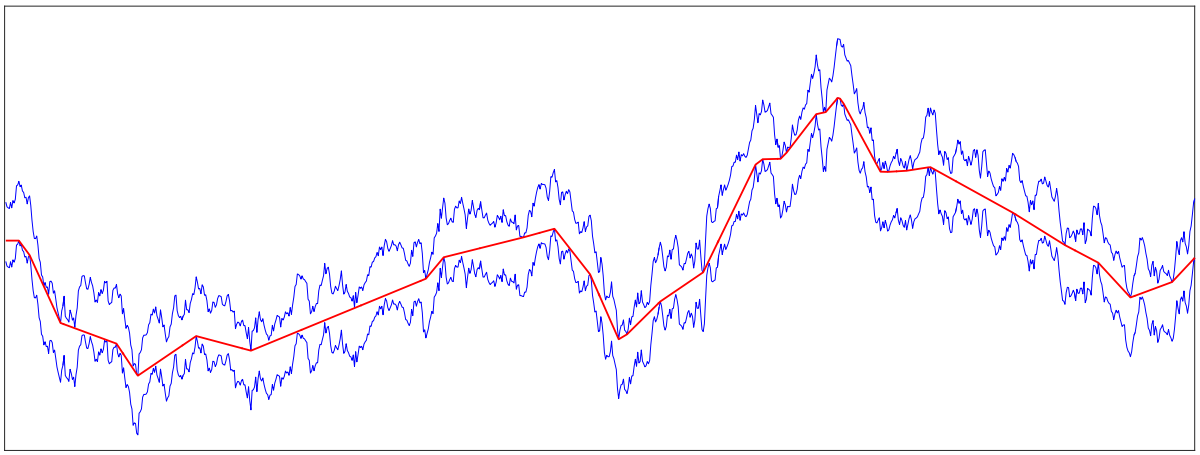

Figure 18.

hemodynamic decompensation. In Figure 18, we have drawn $c-\varepsilon$ and $c+\varepsilon$ for a function c, as well as the solution a to the Taut String problem $\mathbf{TS}^c(\varepsilon)$. It appears as a tight string that is in between the graphs of $c-\varepsilon$ and $c+\varepsilon$ and is a piecewise linear approximation of c.

As the value of ε increases, $\mathrm{gsparse}_Y(a)$ decreases (see Figure 19).

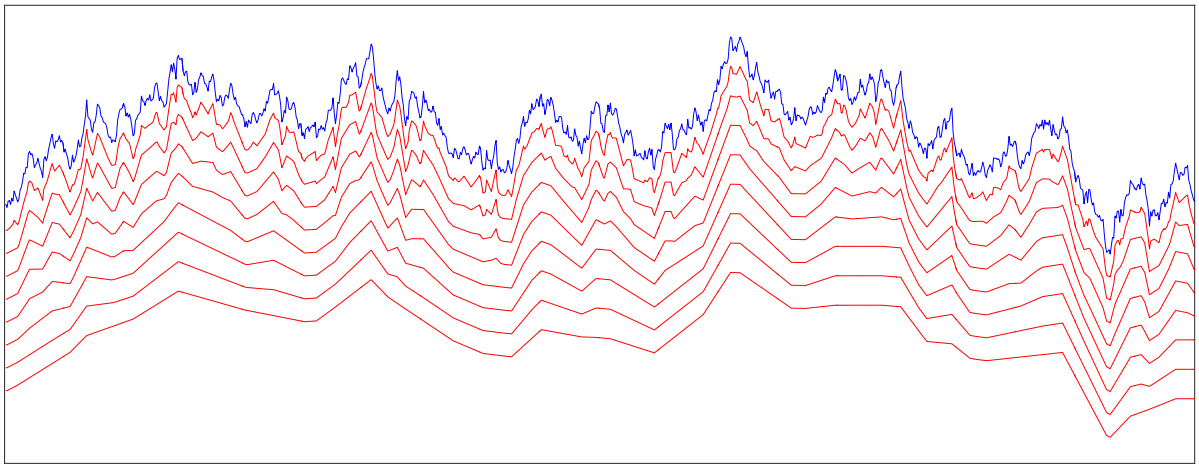

Figure 19.

Let $\pi : \mathbb{R}^{n+1} \to V$ be the projection onto V defined by

$$\pi(a_1, \ldots, a_{n+1}) = (a_1, \ldots, a_{n+1}) - \frac{\sum_{i=1}^{n+1} a_i}{n+1}(1, 1, \ldots, 1).$$

Lemma 9.8. *Suppose that $c \in V$ and $a \in \mathbb{R}^{n+1}$ minimizes $\|D^\star Da\|_1$ under the constraint $\|c-a\|_\infty \leq \varepsilon$. Then $\pi(a)$ is a solution to $\mathbf{M}^c_{XY}(\varepsilon)$ and $c = \pi(a)+\pi(b)$ is an XY-decomposition, where $b = c - a$.*

Proof. We have $\|\pi(b)\|_Y \leq \|b\|_\infty \leq \varepsilon$. Suppose that $c = a' + b'$ with $a', b' \in V$ and $\|b'\|_Y \leq \varepsilon$. Then there is a vector $\overline{b} \in \mathbb{R}^{n+1}$ with $\pi(\overline{b}) = b$ and $\|\overline{b}\|_\infty = \|b'\|_X \leq \varepsilon$. If we define $\overline{a} = c - \overline{b}$, then $\pi(\overline{a}) = a'$, and we have

$$\|a'\|_X = \|D^\star Da'\|_1 = \|D^\star D\overline{a}\|_1 \geq \|D^\star Da\|_1 = \|D^\star D\pi(a)\|_1 = \|\pi(a)\|_X.$$

This shows that $\pi(a)$ is a solution to $\mathbf{M}^c_{XY}(\varepsilon)$. ∎

9.4. An ECG example. For a 500 Hz noisy electrocardiogram (ECG) signal $c \in \mathbb{R}^{2000}$ of 4 seconds, we graph the Pareto frontier (and subfrontier) of c (see Figure 20).

(If $c \notin V$, then we can remove the baseline by replacing c with $\pi(c) \in V$.) The graph is L-shaped, where the vertical leg corresponds to the sparse signal, and the horizontal leg corresponds to noise. The vertical leg starts near the point $(300, 8)$. This means that there exists a decomposition $c = a + b$ with $\|b\|_\infty = 8$ and $\|D^\star Da\|_1 \approx 300$. The signal a is the denoised signal. The singular value region for c is shown in Figure 21.

The x-axis is cut off here in order to better visualize the graph. The graph approaches the x-axis slowly and meets the x-axis near $x = 1000$. The horizontal leg corresponds to noise. To estimate the noise level, we look at where the horizontal leg starts. To find a cutoff for the singular values, one proceeds as in principal component analysis. A reasonable cutoff is

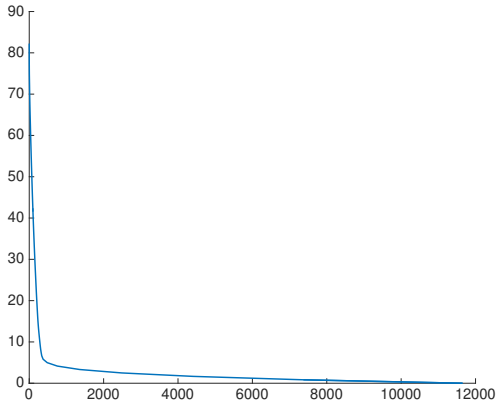

Figure 20.

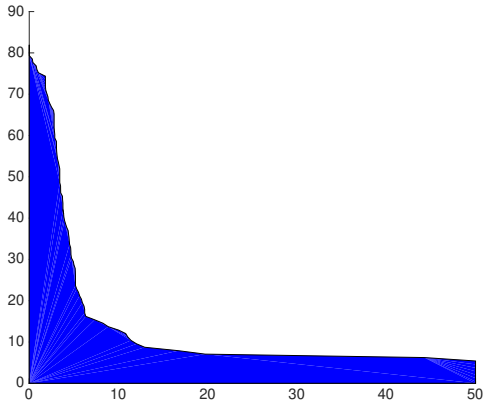

Figure 21.

again $y = 8$. The value 8 is an estimation of the maximum amplitude of noise, which is more than the standard deviation (which is closer to 4 in this case). In Figure 22 we graph the noisy signal and the denoised signals for $y = 2, 4, 6, 8, 10, 12$. The denoised signal for $y = 8$ is colored green.

9.5. Higher order total variation denoising. Suppose that $c : [0, 1] \to \mathbb{R}$ is a function. One can also denoise by using a higher derivative to regularize the ℓ_2 norm by minimizing

$$\tfrac{1}{2}\|c - a\|_2^2 + \lambda\|a^{(k)}\|_1 = \frac{1}{2}\int_0^1 (c(x) - a(x))^2\,dx + \lambda\int_0^1 |a^{(k)}(x)|\,dx.$$

For $k = 1$ we just get 1D total variation denoising. We consider a discrete version.

Define $D^{(k)} : \mathbb{R}^{n+k} \to \mathbb{R}^n$ as the composition $\underbrace{D \circ D \circ \cdots \circ D}_{k}$. We can view $D^{(k)}$ as the

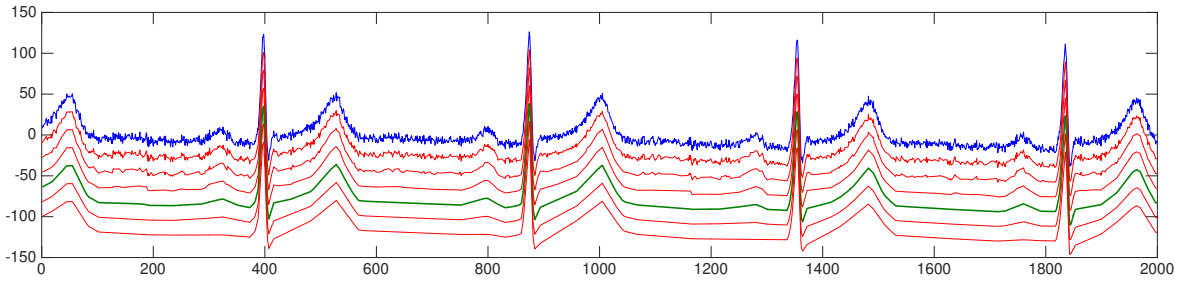

Figure 22.

kth discrete derivative. For example, for $k = 2$ we have

$$D^{(2)}(v_1, v_2, \ldots, v_{n+2}) = (v_1 - 2v_2 + v_3, v_2 - 2v_3 + v_4, \ldots, v_n - 2v_{n+1} + v_{n+2}).$$

Define a seminorm on $\mathbb{R}^{n+k}$ by $\|v\|_X = \|D^{(k)}v\|_1$. Restricting the norm to

$$V = \{(v_1, \ldots, v_{n+k}) \in \mathbb{R}^{n+k} \mid \textstyle\sum_{i=1}^{n+k} v_i i^d = 0 \text{ for } d = 0, 1, \ldots, k-1\}$$

gives a norm; let $\|\cdot\|_Y$ be the dual to this norm.

Problem 9.9. *Given $c \in \mathbb{R}^{n+k}$, minimize $\frac{1}{2}\|c - a\|_2^2 + \lambda\|a\|_X$.*

For $k = 2$, this problem is called ℓ_1-trend filtering. An overview of ℓ_1-trend filtering is given in [33]. Some applications are in financial time series [53], macroeconomics [52], automatic control [41], oceanography [51], and geophysics [34]. In ℓ_1-trend filtering, the function a is piecewise linear. More generally, for $k > 2$, the $(k-1)$th derivative of a will be piecewise constant. This case has been studied in [39]. For $k \geq 2$, the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ are not tight.

10. The ISTA algorithm for XY-decompositions. A general formulation of the Iterative Shrinkage-Thresholding Algorithm (ISTA) was given in [17]. We give here a different general version of an ISTA algorithm that is formulated in terms of norms. A map $f : \mathbb{R}^n \to \mathbb{R}^n$ is called *nonexpansive* if it is Lipschitz with Lipschitz constant 1, i.e., $\|f(v) - f(w)\|_2 \leq \|v - w\|_2$ for all $v, w \in \mathbb{R}^n$.

Lemma 10.1. *If $\|\cdot\|_X$ is a norm on $\mathbb{R}^n$, then the functions $\operatorname{proj}_X(\cdot, x)$ and $\operatorname{shrink}_X(\cdot, x)$ are nonexpansive.*

Proof. Suppose that $v, w \in \mathbb{R}^n$, and let $v' = \operatorname{proj}_X(v, x)$ and $w' = \operatorname{proj}_X(w, x)$. For $t \in [0, 1]$ we have $(1-t)v' + tw' \in B_X(x)$. By definition of $v' = \operatorname{proj}_X(v, x)$ we have

$$\|(v - v') + t(v' - w')\|_2 = \big\|v - \big((1-t)v' + tw'\big)\big\|_2 \geq \|v - v'\|_2.$$

Squaring both sides yields

$$2t\langle v - v', v' - w'\rangle + t^2\|w' - v'\|_2^2 \geq 0.$$

If we take the limit $t \to 0$, we get

$$\langle v - v', v' - w' \rangle \geq 0.$$

By symmetry, we also get

$$\langle w' - w, v' - w' \rangle = \langle w - w', w' - v' \rangle \geq 0.$$

Adding both equations yields

$$\|v - w\|_2^2 - \|v' - w'\|_2^2 - \|(v - w) - (v' - w')\|_2^2 = 2\langle (v - w) - (v' - w'), v' - w' \rangle \geq 0.$$

It follows that

$$\|v' - w'\|_2 \leq \|v - w\|_2 \quad \text{and} \quad \|(v - w) - (v' - w')\|_2 \leq \|v - w\|_2.$$

This shows that $\text{proj}_X(\cdot, x)$ and $\text{shrink}_X(\cdot, x)$ are nonexpansive. ∎

Suppose that $V = \mathbb{R}^n$. Let $D : \mathbb{R}^m \to \mathbb{R}^n$ be a surjective linear map, and suppose that $\|\cdot\|_{\overline{X}}$ is a norm on $\mathbb{R}^m$. We define a norm $\|\cdot\|_X$ on $\mathbb{R}^n$ by

$$\|c\|_X = \min\{\|\overline{c}\|_{\overline{X}} \mid \overline{c} \in \mathbb{R}^m,\ D\overline{c} = c\}.$$

The dual norm $\|\cdot\|_Y$ of $\|\cdot\|_X$ is defined by

$$\|c\|_Y := \|D^\star c\|_{\overline{Y}},$$

where $\|\cdot\|_{\overline{Y}}$ is the dual norm to $\|\cdot\|_{\overline{X}}$. Assume that we can easily compute the norms $\|\cdot\|_{\overline{X}}$, $\|\cdot\|_{\overline{Y}}$, and the projection function $\text{proj}_{\overline{X}}$. We will also assume that the singular values of D all lie in $[0, 1]$. We have the following algorithm for computing $\text{proj}_X(c, x)$ for $c \in V$ and x with $0 \leq x < \|c\|_X$. There is one more parameter, δ, which specifies the accuracy of the output. We assume $0 < \delta < 1$ and the closer δ is to 1, the more accurate the output will be.

function $\text{proj}_X(c, x, \delta)$
  $e \leftarrow 0$
  while $\langle De, c - De \rangle \leq \delta \cdot \|e\|_{\overline{X}} \cdot \|D^\star(c - De)\|_{\overline{Y}}$ **do**
    $e \leftarrow \text{proj}_{\overline{X}}(e + D^\star(c - De), x)$
  end while
  return De
end function

Proposition 10.2. *The algorithm terminates for every δ with $0 < \delta < 1$. Let $a = \text{proj}_X(c, x)$ and $a(\delta) = \text{proj}_X(c, x, \delta)$ (the output of the algorithm). Then we have $\lim_{\delta \to 1} a(\delta) = a$.*

Proof. Define $a = \text{proj}_X(c, x)$ and $b = c - a$ so that $c = a + b$ is an $X2$-decomposition. Because of the definition of the norm $\|\cdot\|_X$, there exists a vector e with $De = a$ and $\|e\|_{\overline{X}} = \|a\|_X$. We set $f = D^\star b + e$. We have

$$\langle e, D^\star b \rangle = \langle De, b \rangle = \langle a, b \rangle = \|a\|_X \|b\|_Y = \|e\|_{\overline{X}} \|D^\star b\|_{\overline{Y}}, \tag{8}$$

so $f = e + D^\star b$ is an $\overline{X}2$-decomposition. It follows that $e = \mathrm{proj}_{\overline{X}}(f, x)$.

Let e_n be the value of e after the nth iteration of the while-loop. We define $a_n = De_n$, $b_n = c - a_n$, and $f_{n+1} = e_n + D^\star b_n$. Then we have $e_{n+1} = \mathrm{proj}_{\overline{X}}(f_{n+1}, x)$.

Because $\mathrm{proj}_{\overline{X}}$ is nonexpansive and the singular values of D are ≤ 1, we have

$$\begin{aligned}(9)\quad \|e_{n+1} - e\|_2^2 \leq \|f_{n+1} - f\|_2^2 &= \|(I - D^\star D)(e_n - e)\|_2^2 \\ &= \|e_n - e\|_2^2 - 2\langle e_n - e, D^\star D(e_n - e)\rangle + \|D^\star D(e_n - e)\|_2^2 \\ &= \|e_n - e\|_2^2 - 2\|D(e_n - e)\|_2^2 + \|D^\star D(e_n - e)\|_2^2 \leq \|e_n - e\|_2^2 - \|D(e_n - e)\|_2^2.\end{aligned}$$

So the sequence $n \mapsto \|e_n - e\|_2$ is nonincreasing and bounded below by 0. So it must converge. Taking the limit $n \to \infty$ on both sides shows that $\|D(e_n - e)\|_2$ and $D(e_n - e) = a_n - a$ tend to 0. So we have $\lim_{n\to\infty} a_n = a$. In particular, $\langle a_n, b_n\rangle = \langle a_n, c - a_n\rangle$ converges to $\langle a, c - a\rangle = \langle a, b\rangle$, and $\|D^\star b_n\|_{\overline{Y}}$ converges to $\|D^\star b\|_{\overline{Y}}$. It follows that

$$\frac{\langle a_n, b_n\rangle}{x\|D^\star b_n\|_{\overline{Y}}} = \frac{\langle a_n, b_n\rangle}{x\|b_n\|_Y} \text{ tends to } \frac{\langle a, b\rangle}{x\|b\|_Y} = \frac{\langle a, b\rangle}{\|a\|_X\|b\|_Y} = 1.$$

So for some n we have

$$\langle De_n, c - De_n\rangle = \langle a_n, b_n\rangle > \delta x\|b_n\|_Y = \delta x\|D^\star(c - De_n)\|_{\overline{Y}},$$

and the algorithm terminates.

Let $n(\delta)$ be the number of times the algorithm runs through the loop with input (c, x, δ). Then, as $\delta \to 1$, we have $n(\delta) \to \infty$. Because $\lim_{n\to\infty} a_n = a$, we get $a(\delta) = a_{n(\delta)} \to a$ as $\delta \to 1$. ∎

11. Basis pursuit denoising, LASSO, and the Dantzig selector. In basis pursuit (BP) one tries to solve the equation $Av = c$, where A is a given $n \times m$ matrix and v is a sparse vector (i.e., there are few nonzero entries). We will assume that A has rank n and that the system is underdetermined, (i.e., $m > n$). It was shown in [13] that v often can be found by minimizing $\|v\|_1$ under the constraint $Av = c$. This can be done efficiently using linear programming. We define a norm $\|\cdot\|_X$ on $\mathbb{R}^n$ by

$$\|c\|_X = \min\{\|v\|_1 \mid Av = c\}.$$

Note that evaluating the norm $\|c\|_X$ of some vector c is a BP problem. We also have

$$\mathrm{sparse}_X(c) = \min\{\|v\|_0 \mid Av = c\}.$$

In the presence of noise, one minimizes $\|v\|_1$ under the constraint $\|Av - c\|_2 \leq y$. This is called basis pursuit denoising (BPDN); see [9, 16]. If we set $a = Av$, then we minimize $\|a\|_X$ under the the constraint $\|c - a\|_2 \leq y$. If we set $b = c - a$, then we minimize $\|c - b\|_X$ under the constraint $\|b\|_2 \leq y$. This is the optimization problem $\mathbf{M}^c_{X2}(y)$.

Sometimes BPDN is formulated as the problem of minimizing ℓ_1-regularized function

$$\tfrac{1}{2}\|Av - c\|_2^2 + \lambda\|v\|_1.$$

This is the same problem as minimizing

$$\tfrac{1}{2}\|c-a\|_2^2+\lambda\|a\|_X.$$

The equivalence between the two formulations of BPDN is well known (see also Proposition 4.6).

The LASSO problem (see [49]) asks to minimize $\|c-Av\|_2$ under the constraint $\|v\|_1 \le x$. This is equivalent to minimizing $\|c-a\|_2$ under the constraint $\|a\|_X \le x$. This is the optimization problem $\mathbf{M}^c_{2X}(x)$.

The dual norm of $\|\cdot\|_X$ is defined by

$$\|c\|_Y := \|A^\star c\|_\infty.$$

Since the $X2$-decompositions and the $2Y$-decompositions are the same, we have two more approaches for finding the $X2$-decompositions (dual LASSO/BPDN):

(1) $\mathbf{M}^c_{2Y}(y)$: Under the constraint $\|A^\star(c-a)\|_\infty \le y$ we minimize $\|a\|_2$ [43, 6].

(2) $\mathbf{M}^c_{Y2}(x)$: Under the constraint $\|a\|_2 \le x$, we minimize $\|A^\star(c-a)\|_\infty$.

To find an XY-decomposition we can minimize $\|a\|_X$ under the constraint $\|c-a\|_Y \le y$ ($\mathbf{M}^c_{XY}(y)$). If we set $a=Av$, then this is equivalent to minimizing $\|v\|_1$ under the constraint $\|A^\star(Av-c)\|_\infty \le y$. This optimization problem is called the *Dantzig selector* [14].

The Dantzig selector does not always have the same solution as LASSO (or the other equivalent problems). Some conditions were given in [1] when the Dantzig selector and LASSO have the same solutions. If $a_1, a_2, \dots, a_m$ are the columns of A, then the unit ball B_X is the convex hull of $a_1, \dots, a_m, -a_1, \dots, -a_m$. Theorem 6.12 gives a necessary and sufficient condition for this polytope so that the Dantzig selector and LASSO have the same solutions for every $c \in V$.

12. Total variation denoising in imaging.

12.1. The 2D total variation norm. We can view a grayscale image as a function $c : [0,1]\times[0,1] \to \mathbb{R}$. In the anisotropic Rudin–Osher–Fatemi [46] total variation model, we seek a decomposition $c = a + b$ such that

$$\frac{1}{2}\int_0^1\int_0^1 (c(x,y)-a(x,y))^2\,dx\,dy + \lambda\int_0^1\int_0^1 \|\nabla a(x,y)\|_1\,dx\,dy$$

is small. (In the isotropic model we replace $\|\nabla a(x,y)\|_1$ by $\|\nabla a(x,y)\|_2$.)

A discrete formulation of the model is as follows. We can view a grayscale image of $m\times n$ pixels as a matrix $c \in \mathbb{R}^{m\times n}$. We define a map $F : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times(n-1)} \times \mathbb{R}^{(m-1)\times n}$ by

$$F(c) = (d, e),$$

where $d_{i,j} = c_{i,j} - c_{i,j+1}$ and $d_{i,j} = e_{i,j} - e_{i,j+1}$ for all i, j. We define a total variation seminorm by

$$\|c\|_X = \|Fc\|_1 = \sum_{i=1}^{m}\sum_{j=1}^{n-1}|c_{i,j}-c_{i,j+1}| + \sum_{i=1}^{m-1}\sum_{j=1}^{n}|c_{i,j}-c_{i+1,j}|.$$

The restriction of $\|\cdot\|_X$ to the set

$$V = \{c \in \mathbb{R}^{m\times n} \mid \textstyle\sum_{i,j} c_{i,j} = 0\}$$

is a norm. We can always normalize an image by subtracting the average value to obtain an element of V. Let $F^\star : \mathbb{R}^{m\times(n-1)} \times \mathbb{R}^{(m-1)\times n} \to \mathbb{R}^{m\times n}$ be the dual of F. We have

$$F^\star(d,e)_{i,j} = d_{i,j} - d_{i,j-1} + e_{i,j} - e_{i-1,j}, \quad 1 \le i \le m,\ 1 \le j \le n,$$

with the conventions that $d_{i,0} = d_{i,n} = e_{0,j} = e_{m,j} = 0$ for all i, j.

The dual norm of $\|\cdot\|_X$ is

$$\|c\|_Y = \min\{\|(d,e)\|_\infty \mid F^\star(d,e) = c\}.$$

Total variation denoising is usually formulated as follows.

Problem 12.1. *Minimize*

$$\tfrac{1}{2}\|c-a\|_2^2 + \lambda\|Fa\|_1.$$

Since the paper of Rudin, Osher, and Fatemi [46], several other algorithms have been proposed—for example, Chambolle's algorithm [10], the split Bregman method [24], and the efficient primal-dual hybrid gradient algorithm [54].

12.2. Sparseness and total variation. We will study the notion of X-sparsity in this context. Suppose that $F \in V$ is an image. An F-region is a maximal connected subset of $\{1,2,\dots,m\} \times \{1,2,\dots,n\}$ on which F is constant.

Proposition 12.2. *For an image F we have* $\operatorname{gsparse}_X(F) = d-1$, *where d is the number of connected regions on which F is constant.*

Proof. Let C be the smallest facial X-cone containing F. By Lemma 6.6, F lies in C if and only if $\|F - \varepsilon G\|_X + \|\varepsilon G\|_X = \|F\|_X$ for some $\varepsilon > 0$. So G lies in C if and only if the following properties are satisfied for all i, j:

(1) $G(i,j) > G(i,j+1)$ implies $F(i,j) > F(i,j+1)$;
(2) $G(i,j) < G(i,j+1)$ implies $F(i,j) < F(i,j+1)$;
(3) $G(i,j) > G(i+1,j)$ implies $F(i,j) > F(i+1,j)$;
(4) $G(i,j) < G(i+1,j)$ implies $F(i,j) < F(i+1,j)$.

Taking the contrapositive in each statement (and changing the indexing), we see that for all i, j we have

(1) $F(i,j) \ge F(i,j+1)$ implies $G(i,j) \ge G(i,j+1)$;
(2) $F(i,j) \le F(i,j+1)$ implies $G(i,j) \le G(i,j+1)$;
(3) $F(i,j) \ge F(i+1,j)$ implies $G(i,j) \ge G(i+1,j)$;
(4) $F(i,j) \le F(i+1,j)$ implies $G(i,j) \le G(i+1,j)$.

It is clear that for all $G \in C$, we have that G is constant on the connected regions on which F is constant. The function G can have arbitrary values on these d connected regions as long as the four inequalities above and the linear constraint $\sum_{i,j} G(i,j) = 0$ are satisfied. It follows that $\operatorname{gsparse}_X(F) = \dim C = d-1$. ∎

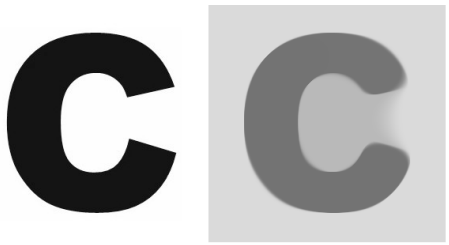

Figure 23.

12.3. The total variation norm is not tight. In Figure 23 we denoised the image of the letter C using Rudin–Osher–Fatemi total variation denoising. The original image has only two colors, black and white, and has geometric X-sparsity 1. In the denoised images, there are various shades of gray, and the geometric X-sparsity is more than 1. So Rudin–Osher–Fatemi denoising may increase the geometric sparsity, so the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ are not tight.

13. Tensor decompositions.

13.1. CP decompositions. One of the motivations for writing this paper is the study of tensor decompositions. Suppose that $\mathbb{F}$ is the field $\mathbb{C}$ or $\mathbb{R}$, and that $V^{(1)}, \dots, V^{(d)}$ are finite dimensional $\mathbb{F}$-vector spaces. Define

$$V = V^{(1)} \otimes_{\mathbb{F}} V^{(2)} \otimes_{\mathbb{F}} \cdots \otimes_{\mathbb{F}} V^{(d)}.$$

By a tensor we will mean an element of V. Elements of V can be thought of as multiway arrays of size $n_1 \times n_2 \times \cdots \times n_d$, where $n_i = \dim V^{(i)}$. A *simple tensor* (also called a *rank one tensor*) is a tensor of the form

$$v^{(1)} \otimes v^{(2)} \otimes \cdots \otimes v^{(d)},$$

where $v^{(i)} \in V^{(i)}$ for all i. Not every tensor is simple, but every tensor can be written as a sum of simple tensors.

Problem 13.1 (tensor decomposition). *Given a tensor T, find a decomposition $T = v_1 + v_2 + \cdots + v_r$ where $v_1, \dots, v_r$ are simple tensors and r is minimal.*

Hitchcock defined the rank of the tensor T in [31] as the smallest r for which such a decomposition exists, and this minimal rank decomposition is called the *canonical polyadic decomposition*. Problem 13.1 is also known as the PARAFAC [29] or CANDECOMP [8] model. Finding the rank of a tensor is an NP-hard problem. Over $\mathbb{Q}$ this was shown in [30], and in our case, $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, this was proved in [32].

13.2. The CoDe model and the nuclear norm. Even in relatively small dimensions, there are examples of tensors for which the rank is unknown. Using the heuristic of convex relaxation, we consider the following problem.

Problem 13.2 (CoDe model). *Given a tensor T, find a decomposition $T = v_1 + v_2 + \cdots + v_r$ where $v_1, \dots, v_r$ are simple tensors and $\sum_{i=1}^{r} \|v_i\|_2$ is minimal.*

The nuclear norm for tensors was explicitly given [37, 38], but the ideas go back to [25] and [47].

Definition 13.3. *The nuclear norm $\|T\|_\star$ of the tensor is the smallest possible value of $\sum_{i=1}^r \|v_i\|_2$ such that $v_1, \dots, v_r$ are simple tensors and $T = \sum_{i=1}^r v_i$.*

A matrix can be viewed as a 2-way tensor, and in this case the nuclear norm for tensors coincides with the nuclear norm of the matrix, which is defined as

$$\|A\|_\star = \operatorname{trace}(\sqrt{A^\star A}) = \lambda_1 + \lambda_2 + \cdots + \lambda_r,$$

where $A^\star$ is the complex conjugate transpose of A, $\sqrt{A^\star A}$ is the unique nonnegative definite Hermitian matrix whose square is $A^\star A$, and $\lambda_1, \lambda_2, \dots, \lambda_r$ are the singular values of A. Although finding the nuclear norm of a higher order tensor is also NP-complete (see [23]), it is often easier than determining its rank. In [19] some examples of tensors are given for which the nuclear norm and the optimal decomposition can be computed, but where the rank of the tensors are unknown.

Let $\|\cdot\|_X = \|\cdot\|_\star$ be the nuclear norm. The dual norm, $\|\cdot\|_Y$, is equal to the spectral norm.

Definition 13.4. *The spectral norm of a tensor T is defined by*

$$\|T\|_\sigma = \max\{|\langle T, v\rangle| \mid v \text{ is a simple tensor with } \|v\|_2 = 1\}.$$

Finding the spectral norm of a higher-order tensor is also an NP-complete problem (see [32]).

From now on we consider the case $\mathbb{F} = \mathbb{C}$, where V is the tensor product (over $\mathbb{C}$) of several finite dimensional Hilbert spaces. We have a positive definite Hermitian inner product $\langle\cdot,\cdot\rangle_{\mathbb{C}}$ on V. A real inner product is given by $\langle\cdot,\cdot\rangle = \Re\langle\cdot,\cdot\rangle_{\mathbb{C}}$.

13.3. Examples of unitangent tensors.

The following examples come from [19].

Example 13.5. The space $\mathbb{C}^{p\times q}$ of complex $p \times q$ matrices has the usual basis $e_{i,j}$, where $1 \le i \le p$ and $1 \le j \le q$. Define $V = \mathbb{C}^{p\times q} \otimes \mathbb{C}^{q\times r} \otimes \mathbb{C}^{r\times p}$, and define the tensor

$$T_{p,q,r} = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} e_{i,j} \otimes e_{j,k} \otimes e_{k,i}. \tag{10}$$

This tensor is related to matrix multiplication. It is known that if $\operatorname{rank}(T_{p,q,r}) \le d$, then two $n \times n$-matrices can be multiplied using $O(n^{3\log(d)/\log(pqr)})$ arithmetic operations in $\mathbb{C}$. For most p, q, r, the rank of $T_{p,q,r}$ is unknown. For example, the best known lower bound for $\operatorname{rank}(T_{3,3,3})$ is 19 and follows from [7]. The best known upper bound is 23 and comes from [35]. It was shown in [19] that $\|T_{p,q,r}\|_\sigma = 1$ and $\|T_{p,q,r}\|_\star = pqr$. Now (10) is a convex decomposition, because

$$pqr = \|T_{p,q,r}\|_\star = \sum_{i=1}^{p}\sum_{i=1}^{q}\sum_{i=1}^{r} \|e_{i,j} \otimes e_{j,k} \otimes e_{k,i}\|_2.$$

Also, we have

$$\|T_{p,q,r}\|_2^2 = pqr = \|T_{p,q,r}\|_\star \|T_{p,q,r}\|_\sigma,$$

so $T_{p,q,r}$ is unitangent. This means that

$$T_{p,q,r} = \Big(\lambda T_{p,q,r}\Big) + \Big((1-\lambda)T_{p,q,r}\Big)$$

is an $X2$-decomposition, a $2Y$-decomposition, and an XY-decomposition (as well as $2X$, $Y2$, and YX) if $0 \le \lambda \le 1$. If we take $\lambda = \|S\|_\star / \|T_{p,q,r}\|_\star = \|S\|_\star/(pqr)$, then we have $\|S\|_\star \le \|\lambda T_{p,q,r}\|_\star$. It follows that we get

$$\|T_{p,q,r} - S\|_2 \ge \|(1-\lambda)T_{p,q,r}\|_2 = \left(1 - \frac{\|S\|_\star}{pqr}\right)\sqrt{pqr} = \sqrt{pqr} - \frac{\|S\|_\star}{\sqrt{pqr}}.$$

Similarly, we get inequalities

$$\|T_{p,q,r} - S\|_2 \ge \sqrt{pqr} - \sqrt{pqr}\|S\|_\sigma$$

and

$$\|T_{p,q,r} - S\|_\sigma \ge 1 - \frac{\|S\|_\star}{pqr}.$$

Example 13.6. Let Σ_n be the set of permutations of $\{1, 2, \dots, n\}$, and for a permutation τ denote its sign by $\operatorname{sgn}(\tau)$. The determinant tensor is defined by

$$\det_n = \sum_{\tau \in \Sigma_n} \operatorname{sgn}(\tau) e_{\tau(1)} \otimes e_{\tau(2)} \otimes \cdots \otimes e_{\tau(n)} \in \mathbb{C}^n \otimes \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n.$$

It was shown in [19] that $\|\det_n\|_\sigma = 1$, $\|\det_n\|_\star = n!$, and $\|\det_n\|_2 = \sqrt{n!}$. In particular, $\det_n$ is unitangent. Let

$$\operatorname{perm}_n = \sum_{\tau \in \Sigma_n} e_{\tau(1)} \otimes e_{\tau(2)} \otimes \cdots \otimes e_{\tau(n)} \in \mathbb{C}^n \otimes \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n$$

be the permanent tensor. In [19] it was calculated that $\|\operatorname{perm}_n\|_\sigma = n!/n^{n/2}$, $\|\operatorname{perm}_n\|_\star = n^{n/2}$, and $\|\operatorname{perm}_n\|_2 = \sqrt{n!}$. The tensor perm_n is also unitangent.

13.4. The diagonal singular value decomposition and the slope decomposition. Following [19], we give the following definitions.

Definition 13.7. *Suppose that $v_1, v_2, \dots, v_r$ are simple tensors with $\|v_i\|_2 = 1$ for all i. For a real number $t \ge 1$ we say that $v_1, \dots, v_r$ are t-orthogonal if*

$$\sum_{j=1}^{r} |\langle v_i, w\rangle_{\mathbb{C}}|^{2/t} \le 1$$

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

Note that t-orthogonality implies orthogonality because we can take $w = v_j$ so that

$$\sum_{j=1}^{r} |\langle v_i, w\rangle_{\mathbb{C}}|^{2/t} = 1 + \sum_{j \ne i} |\langle v_i, v_j\rangle_{\mathbb{C}}|^{2/t} \le 1$$

implies that v_j is orthogonal to all v_i with $i \ne j$. By Pythagoras's theorem, orthogonality is equivalent to 1-orthogonality.

Definition 13.8. *The expression*

$$T = \lambda_1 v_1 + \cdots + \lambda_r v_r \tag{11}$$

is called a diagonal singular value decomposition (DSVD) if $v_1, \ldots, v_r$ *are* 2*-orthogonal simple tensors of length* 1 *and* $\lambda_1 \geq \cdots \geq \lambda_r > 0$.

If (11) is a DSVD, then we have

$$\begin{aligned}\|T\|_\star &= \lambda_1 + \lambda_2 + \cdots + \lambda_r,\\ \|T\|_\sigma &= \max\{\lambda_1, \lambda_2, \ldots, \lambda_r\},\\ \|T\|_2 &= \sqrt{\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_r^2}.\end{aligned}$$

Theorem 13.9. *Suppose that a tensor* T *has a DSVD with singular values* $\lambda_1 > \lambda_2 > \cdots > \lambda_r > 0$ *and multiplicities* $m_1, m_2, \ldots, m_r$*, respectively. Then we can write*

$$T = \lambda_1 w_1 + \lambda_2 w_2 + \cdots + \lambda_r w_r$$

such that

$$w_i = v_{i,1} + v_{i,2} + \cdots + v_{i,m_i}$$

for all i *and*

$$v_{1,1}, \ldots, v_{1,m_1}, v_{2,1}, \ldots, v_{2,m_2}, \ldots, v_{r,1}, \ldots, v_{r,m_r}$$

is a sequence of 2*-orthogonal simple unit tensors. Then the slope decomposition of* T *is given by*

$$T = u_1 + u_2 + \cdots + u_r,$$

where

$$u_i = (\lambda_i - \lambda_{i+1})(w_1 + w_2 + \cdots + w_i).$$

Proof. We have

$$\begin{aligned}\|u_i\|_\sigma &= (\lambda_i - \lambda_{i+1})\|w_1 + \cdots + w_i\|_\sigma = (\lambda_i - \lambda_{i+1}),\\ \|u_i\|_\star &= (\lambda_i - \lambda_{i+1})\|w_1 + \cdots + w_i\|_\star = (\lambda_i - \lambda_{i+1})(m_1 + m_2 + \cdots + m_i),\\ \mu_{\star\sigma}(u_i) &= \frac{\|u_i\|_\sigma}{\|u_i\|_\star} = (m_1 + m_2 + \cdots + m_i)^{-1},\end{aligned}$$

so we have

$$\mu_{\star\sigma}(u_1) > \mu_{\star\sigma}(u_2) > \cdots > \mu_{\star\sigma}(u_s) > 0.$$

For $i < j$ we have

$$\begin{aligned}\langle u_i, u_j\rangle_{\mathbb{C}} &= (\lambda_i - \lambda_{i+1})(\lambda_j - \lambda_{j+1})\langle w_1 + \cdots + w_i, w_1 + \cdots + w_j\rangle_{\mathbb{C}}\\ &= (\lambda_i - \lambda_{i+1})(\lambda_j - \lambda_{j+1})(m_1 + m_2 + \cdots + m_i) = \|u_i\|_\star \|u_j\|_\sigma.\end{aligned}$$

So we also have

$$\langle u_i, u_j\rangle = \Re\langle u_i, u_j\rangle_{\mathbb{C}} = \|u_i\|_\star \|u_j\|_\sigma.$$

This proves that $T = u_1 + u_2 + \cdots + u_s$ is the slope decomposition. ∎

It was shown in [19] that the tensor $T_{p,q,r}$ has a DSVD, but perm_n and $\det_n$ do not for $n \geq 3$.

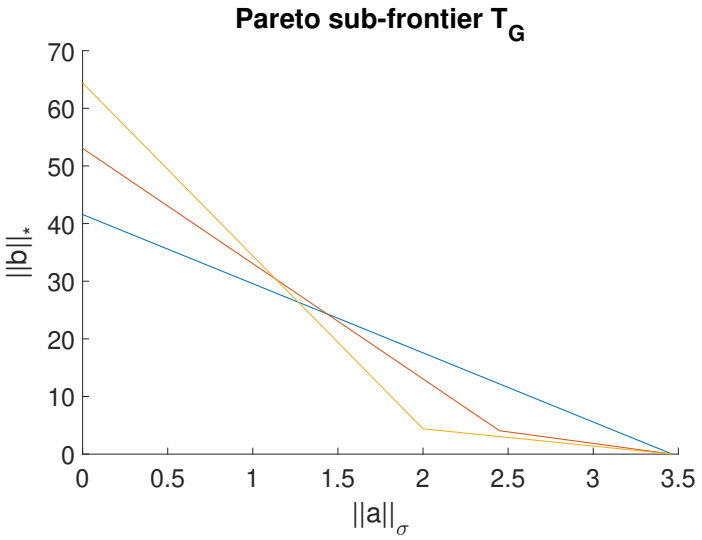

Figure 24.

13.5. Group algebra tensors. Suppose that G is a finite group of order n. Suppose that there are m_d irreducible representations of dimension d. Then we have $\sum_d m_d d^2 = n$. Let e_g, $g \in G$, be an orthonormal basis of $\mathbb{C}^n$, and consider the tensor

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

which is related to the multiplication in the group algebra. Then T_G has singular value $\sqrt{n/d}$ with multiplicity $m_d d^3$ for all d (see [19]). We have

$$f_{*\sigma}(\lambda) = h_{*\sigma}(\lambda) = \sum_d m_d d^3 \max\left\{\lambda - \sqrt{\frac{n}{d}}, 0\right\}.$$

In Figure 24 we draw the Pareto subfrontier of T_G for all groups G of order G. The blue graph represents the abelian groups $\mathbb{Z}/12$ and $\mathbb{Z}/6 \times \mathbb{Z}/2$ with only 1D representations, the red graph represents the dihedral group D_6 and and the semidirect product $\mathbb{Z}/4 \ltimes \mathbb{Z}/3$ with representations of dimension $1, 1, 1, 1, 2, 2$, and the yellow graph represents the alternating group $\mathbf{A}_4$ with representations of dimension $1, 1, 2, 3, 3$.

13.6. Symmetric tensors in $\mathbb{R}^{2\times2\times2}$. For the remainder of this section, let us consider the tensor product space $\mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2$. In particular, for $t \in \mathbb{R}$, we will study the symmetric tensor

$$U_t = e_2 \otimes e_1 \otimes e_1 + e_1 \otimes e_2 \otimes e_1 + e_1 \otimes e_1 \otimes e_2 + te_2 \otimes e_2 \otimes e_2.$$

Proposition 13.10.

(1) *We have*

$$\|U_t\|_\sigma = \begin{cases} |t| & \text{if } t \geq 2 \text{ or } t \leq -1, \\ \frac{2}{\sqrt{3-t}} & \text{if } -1 \leq t \leq 2. \end{cases}$$

(2) *We have*

$$\|U_t\|_\star = \begin{cases} 3-t & \text{if } t \le \frac{1}{3}, \\ \frac{(1+t)^{3/2}}{\sqrt{t}} & \text{if } t \ge \frac{1}{3}. \end{cases}$$

Proof.
(1) By the definition of the spectral norm, we have

$$\|U_t\|_\sigma = \max\{\langle U_t, v_1 \otimes v_2 \otimes v_3\rangle \mid \|v_1\| = \|v_2\| = \|v_3\| = 1\}.$$

By Banach's theorem (see [2, 22]), we may take $v_1 = v_2 = v_3 = v$. If we write $v = xe_1 + ye_2$ with $x^2 + y^2 = 1$, then we have

$$\begin{aligned}\|U_t\|_\sigma &= \max_{x^2+y^2=1} \langle U_t, (xe_1 + ye_2) \otimes (xe_1 + ye_2) \otimes (xe_1 + ye_2)\rangle\} \\ &= \max_{x^2+y^2=1} 3x^2y + ty^3 = \max_{|y|\le 1} 3y + (t-3)y^3.\end{aligned}$$

Let $g_t(y) = 3y + (t-3)y^3$ for $y \in [-1, 1]$. We get $g_t'(y) = 3 + 3(t-3)y^2$. If $t > 2$, then $g_t'(y)$ has no roots in $[-1, 1]$ and

$$\|U_t\|_\sigma = \max\{g(1), g(-1)\} = t.$$

If $t < 2$, then the roots of $g_t'(y)$ are

$$y = \pm\frac{1}{\sqrt{3-t}}.$$

We have

$$\begin{aligned}\|U_t\|_\sigma &= \max\Big\{g(1), g(-1), g\Big(\frac{1}{\sqrt{3-t}}\Big), g\Big(-\frac{1}{\sqrt{3-t}}\Big)\Big\} \\ &= \max\Big\{|t|, \frac{2}{\sqrt{3-t}}\Big\} = \begin{cases} |t| & \text{if } t \le -1, \\ \frac{2}{\sqrt{3-t}} & \text{if } -1 \le t \le 2. \end{cases}\end{aligned}$$

(2) Note that

$$U_t = \frac{1}{2\sqrt{t}}\Big((e_1 + \sqrt{t}e_2) \otimes (e_1 + \sqrt{t}e_2) \otimes (e_1 + \sqrt{t}e_2) - (e_1 - \sqrt{t}e_2) \otimes (e_1 - \sqrt{t}e_2) \otimes (e_1 - \sqrt{t}e_2)\Big).$$

This implies that

$$\|U_t\|_\star \le \frac{(1+t)^{3/2}}{\sqrt{t}}.$$

If $t \ge \frac{1}{3}$, then let $s = 2 - t^{-1} \in [-1, 2]$. We get

$$2 + 2t = 3 + st = \langle U_t, U_s\rangle \le \|U_t\|_\star\|U_s\|_\sigma = \|U_t\|_\star\frac{2}{\sqrt{3-s}} = \|U_t\|_\star\frac{2\sqrt{t}}{\sqrt{1+t}}.$$

So it follows that

$$\|U_t\|_\star \ge \frac{(1+t)^{3/2}}{t},$$

so we must have equality.

For $t = \frac{1}{3}$ we get $\|U_{1/3}\|_\star = \frac{8}{3}$. For $t < \frac{1}{3}$ we get

$$\|U_t\|_\star \leq \|U_{1/3}\|_\star + \|U_t - U_{1/3}\|_\star = \tfrac{8}{3} + \|(t - \tfrac{1}{3})e_2 \otimes e_2 \otimes e_2\|_\star = \tfrac{8}{3} + \tfrac{1}{3} - t = 3 - t.$$

We have

$$3 - t = \langle U_t, U_{-1} \rangle \leq \|U_t\|_\star \|U_{-1}\|_\sigma = \|U_t\|_\star,$$

so this shows that $\|U_t\|_\star = 3 - t$. ∎

Corollary 13.11. *For* $0 \leq t \leq \frac{1}{3}$

$$U_0 = \Big(\frac{1}{t+1}\Big)U_t + \Big(\frac{t}{1+t}\Big)U_{-1}$$

is a $\star\sigma$*-decomposition, and for* $\frac{1}{3} \leq t \leq \frac{1}{2}$

$$U_0 = \left(\frac{1-2t}{(1-t)^2}\right) U_t + \Big(\frac{t^2}{(1-t)^2}\Big)U_{2-t^{-1}}$$

is a $\star\sigma$*-decomposition. A parameterization of the Pareto subfrontier is given by*

$$\left(\frac{3-t}{t+1}, \frac{t}{1+t}\right)$$

if $0 \leq t \leq \frac{1}{3}$ *and*

$$\left(\frac{(1-2t)(1+t)^{3/2}}{(1-t)^2\sqrt{t}}, \frac{2t^{5/2}}{(1-t)^2\sqrt{1+t}}\right)$$

if $\frac{1}{3} \leq t \leq \frac{1}{2}$.

We plot the Pareto subfrontier in Figure 25. The blue part of the graph is linear, but the red part is nonlinear. The graph is not piecewise linear, so U_0 does not have a slope decomposition. This shows that the nuclear norm and the spectral norm on $\mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \mathbb{R}^2$ are not tight.

In Figure 26 we have plotted the singular value region. The singular value $\frac{2}{\sqrt{3}}$ appears with multiplicity $\frac{27}{13}$, and the singular value $\frac{1}{4}$ appears with multiplicity $\frac{4}{3}$. The singular values between $\frac{1}{4}$ and $\frac{2}{\sqrt{3}}$ appear with infinitesimal multiplicities. The height of the region is the spectral norm $\|U_0\|_\sigma = \frac{2}{\sqrt{3}}$, the area of the region is the nuclear norm $\|U_0\|_\star = 3$, and if we integrate $2y^2$ over the region, we get the square of the Euclidean norm, which is $\|U_0\|_2^2 = 3$.

14. Conclusion.

We developed a general theory of denoising by studying two competing norms that are dual to each other. Examples that fit in this framework include total variation denoising for time signals or images, principal component analysis, LASSO and basis pursuit denoising and the convex decomposition model for tensor decompositions. We have developed useful notions such as the Pareto subfrontier, the singular value region, tight vectors and norms, and the slope decomposition that are applicable to a wide range of applications.

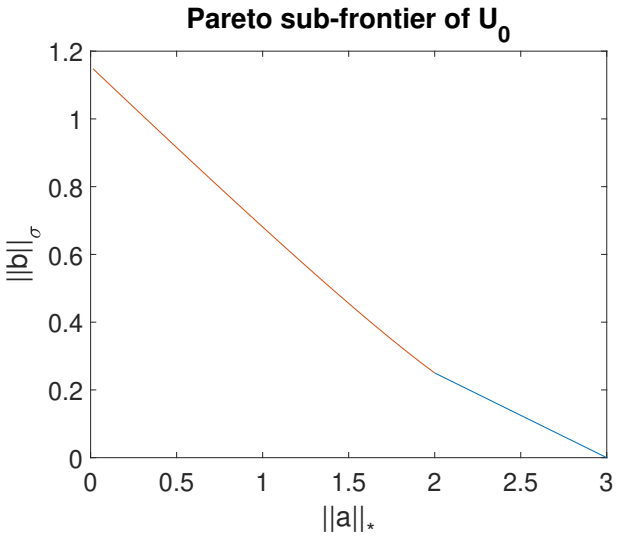

Figure 25.

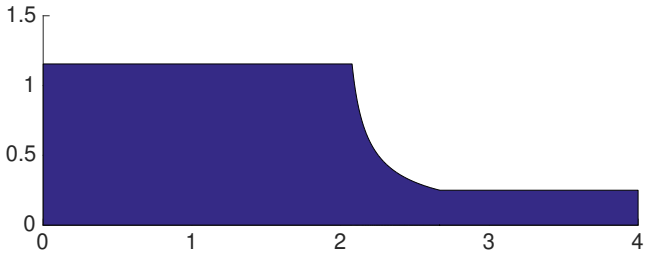

Figure 26.

In some applications, such as 1D total variation denoising, the competing norms are tight. In these cases one can always define singular values and their multiplicities, and the slope decomposition which is similar to singular value decomposition. Just as in principal component analysis, small singular values correspond to noise. In the general case where the norms may not be tight, one can still define the singular value region which is similar to the singular spectrum of a matrix. Like in principal component analysis, the singular value region gives us information about the noise.

Of particular interest is the example of tensor product spaces. In [19], the author introduced the diagonal singular value decomposition for tensors, which is a generalization of the singular value decomposition of a matrix. This notion is closely related to canonical polyadic decompositions of tensors, but a DSVD does not exist for all tensors. The slope decomposition generalizes the notion of the DSVD and can be defined for more tensors. Moreover, the singular value region can be defined for all tensors.

Acknowledgments. I would like to thank Lek-Heng Lim for discussions that inspired me to write this paper, and Neriman Tokcan for helpful comments.

REFERENCES

[1] M. S. Asif and J. Romberg, *On the LASSO and Dantzig selector equivalence*, in the 44th Conference on Information Sciences and Systems (CISS) (Princeton, NJ, 2010), IEEE, Washington, DC, 2010.

[2] S. Banach, *Über homogene Polynome in (L^2)*, Studia Math., 7 (1938), pp. 36–44.

[3] R. Barlow, D. Bartholomew, J. Bremner, and H. Brunk, *Statistical Inference under Order Restrictions*, Wiley, New York, 1972.

[4] A. Barvinok, *A Course in Convexity*, Grad. Stud. Math. 54, AMS, Providence, RI, 2002.

[5] A. Belle, S. Asgari, M. Spadafore, V. A. Convertino, K. R. Ward, H. Derksen, and K. Najarian, *A signal processing approach for detection of hemodynamic instability before decompensation*, PLoS ONE, 11 (2016), e0148544.

[6] E. van den Berg and M. P. Friedlander, *Probing the Pareto frontier for basic pursuit solutions*, SIAM J. Sci. Comput., 31 (2008), pp. 890–912, https://doi.org/10.1137/080714488.

[7] M. Bläser, *On the complexity of the multiplication of matrices of small formats*, J. Complexity, 19 (2003), pp. 43–60.

[8] J. D. Carroll and J. Chang, *Analysis of individual differences in multidimensional scaling via an N-way generalization of an Eckart-Young decomposition*, Psychometrika, 35 (1970), pp. 283–319.

[9] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Rev., 43 (2001), pp. 129–159, https://doi.org/10.1137/S003614450037906X.

[10] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imag. Vis., 20 (2004), pp. 89–97.

[11] L. Condat, *A direct algorithm for 1D total variation denoising*, IEEE Signal Process. Lett., 20 (2013), pp. 1054–1057.

[12] E. Candès and B. Recht, *Exact matrix completion via convex optimization*, Found. Comput. Math., 9 (2009), pp. 717–772.

[13] E. J. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4215.

[14] E. J. Candès and T. Tao, *The Dantzig selector: Statistical estimation when p is much larger than n*, Ann. Statist., 35 (2007), pp. 2313–2351.

[15] E. J. Candès and T. Tao, *The power of convex relaxation: Near-optimal matrix completion*, IEEE Trans. Inform. Theory, 56 (2010), pp. 2053–2080.

[16] E. J. Candès, J. K. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.

[17] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457.

[18] P. L. Davies and A. Kovac, *Local extremes, runs, strings and multiresolution*, Ann. Statist., 29 (2001), pp. 1–65.

[19] H. Derksen, *On the nuclear norm and the singular value decomposition of tensors*, Found. Comput. Math., 16 (2016), pp. 779–811.

[20] D. L. Donoho, *De-noising by soft thresholding*, IEEE Trans. Inform. Theory, 41 (1995), pp. 613–627.

[21] B. Efron, I. Johnstone, T. Hastie, and R. Tibshirani, *Least angle regression*, Ann. Statist., 32 (2004), pp. 407–499.

[22] S. Friedland, *Best rank-one approximation of real symmetric tensors can be chosen symmetric*, Front. Math. China, 8 (2013), pp. 19–40.

[23] S. Friedland and L.-H. Lim, *Nuclear norm of higher-order tensors*, Math. Comp., 87 (2018), pp. 1255–1281.

[24] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343, https://doi.org/10.1137/080725891.

[25] A. Grothendieck, *Produits tensoriels topologiques et espaces nucléaires*, Mem. Amer. Math. Soc., 1955 (1955), 16.

[26] M. Gu, L.-H. Lim, and C. J. Wu, *ParNes: A rapidly convergent algorithm for accurate recovery of sparse and approximately sparse signals*, Numer. Algorithms, 64 (2013), pp. 321–347.

[27] P. C. Hansen, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., 34 (1992), pp. 561–580, https://doi.org/10.1137/1034115.

[28] P. C. Hansen and D. P. O'Leary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput., 14 (1993), pp. 1487–1503, https://doi.org/10.1137/0914086.
[29] R. A. Harshman, *Foundations of the PARAFAC procedure: Model and conditions for an explanatory multi-mode factor analysis*, UCLA Working Papers in Phonetics, 16 (1970), 1.
[30] J. Håstad, *Tensor rank is NP-complete*, J. Algorithms, 11 (1990), pp. 644–654.
[31] F. L. Hitchcock, *The expression of a tensor or a polyadic as a sum of products*, J. Math. Phys., 6 (1927), pp. 164–189.
[32] C. J. Hillar and L.-H. Lim, *Most tensor problems are NP-hard*, J. ACM, 60 (2013), 45.
[33] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky, ℓ_1 *trend filtering*, SIAM Rev., 51 (2009), pp. 339–360, https://doi.org/10.1137/070690274.
[34] Y. Klinger, *Relation between continental strike-slip earthquake segmentation and thickness of the crust*, J. Geophys. Res., 115 (2010), B07306.
[35] J. Laderman, *A noncommutative algorithm for multiplying* 3×3 *matrices using* 23 *multiplication*, Bull. Amer. Math. Soc., 82 (1976), pp. 180–182.
[36] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Classics Appl. Math. 15, SIAM, Philadelphia, 1995, https://doi.org/10.1137/1.9781611971217; first published by Prentice–Hall, 1974.
[37] L.-H. Lim and P. Comon, *Multiarray signal processing: Tensor decomposition meets compressed sensing*, C. R. Acad. Sci. Paris Ser. IIB Mech., 338 (2010), pp. 311–320.
[38] L. H. Lim and P. Comon, *Blind multilinear identification*, IEEE Trans. Inform. Theory, 60 (2014), pp. 1260–1280.
[39] E. Mammen and S. van de Geer, *Locally adaptive regression spines*, Ann. Statist., 25 (1997), pp. 387–413.
[40] K. Najarian, A. Belle, K. Ward, and H. Derksen, *Early Detection of Hemodynamic Decompensation Using Taut-String Transformation*, U.S. Provisional Patent Ser. 62/018,336; filed June 26, 2015.
[41] H. Ohlsson, F. Gustafsson, L. Ljung, and S. Boyd, *Smoothed state estimates and abrupt changes using sum of norms regularization*, Automatica J. IFAC, 48 (2012), pp. 595–605.
[42] M. R. Osborne, B. Presnell, and B. A. Turlach, *A new approach to variable selection in least square problems*, IMA J. Numer. Anal., 20 (2000), pp. 389–404.
[43] M. R. Osborne, B. Presnell, and B. A. Turlach, *On the LASSO and its dual*, J. Comput. Graph. Statist., 9 (2000), pp. 319–337.
[44] D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, J. ACM, 9 (1962), pp. 84–97.
[45] B. Recht, M. Fazel, and P. A. Parillo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501, https://doi.org/10.1137/070697835.
[46] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[47] R. Schatten, *A Theory of Cross-Spaces*, Princeton University Press, Princeton, NJ, 1950.
[48] H. N. Tehrani, A. McEwan, C. Jin, and A. van Schaik, *L1 regularization method in electrical impedance tomography by using the L1-curve (Pareto frontier curve)*, Appl. Math. Model., 36 (2012), pp. 1095–1105.
[49] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B, 58 (1996), pp. 267–288.
[50] A. N. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Soviet Math. Dokl., 4 (1963), pp. 1035–1038; English translation of Dokl. Akad. Nauk. SSSR, 151 (1963), pp. 501–504.
[51] C. Wunsch, *Toward a midlatitude ocean frequency—wavenumber spectral density and trend determination*, J. Physical Oceanography, 40 (2010).
[52] H. Yamada and L. Jin, *Japan's output gap estimation and* ℓ_1 *trend filtering*, Empirical Econom., 45 (2013), pp. 81–88.
[53] H. Yamada and G. Yoon, *When Grilli and Yang meet Prebisch and Singer: Piecewise linear trends in primary commodity prices*, J. Internat. Money Finance, 42 (2014), pp. 193–207.
[54] M. Zhu and T. Chan, *An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration*, UCLA CAM Report 08-34, UCLA, Los Angeles, CA, 2008.