Note II — Linear Spaces*

Bruno Salcedo†

Fall 2016

1. Vectors

Let $\mathcal{F} = (F, +, \times, 1, 0)$ be a field. A *vector space* or *linear space* over $\mathcal{F}$ is a structure $\mathcal{V} = (V, +, \cdot, \mathbf{0})$ such that

(V1) $+ : V \times V \to V$ is an associative and commutative binary operation called *vector addition*, with $\mathbf{0} \in V$ as an identity elements, and such that every $\mathbf{v} \in V$ has an inverse $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.

(V2) $\cdot : F \times V \to V$ is a binary operation called *scalar product* with $1 \in F$ is the (left) identity element, i.e., $1 \cdot \mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$, and such that $(a \times b) \cdot \mathbf{v} = a \cdot (b \cdot \mathbf{v})$ for all $a, b \in F$ and $\mathbf{v} \in V$.

(V3) The scalar product is distributive respect to both scalar addition and vector addition i.e., $(a + b) \cdot \mathbf{v} = (a \cdot \mathbf{v}) + (b \cdot \mathbf{v})$ and $a \cdot (\mathbf{v} + \mathbf{u}) = (a \cdot \mathbf{v}) + (a \cdot \mathbf{u})$ for all $\mathbf{u}, \mathbf{v} \in V$ and $a, b \in F$

Elements of V are called *vectors* and I will denote them with bold-font letters. Elements of F are called *scalars* in relation to V. Typically, there is no risk of ambiguity, so we simply say that V is a vector field over F. Moreover, every vector

*Lecture notes for Econ 6170, Intermediate Mathematical Economics I, Cornell University.

†Department of Economics, Cornell University · brunosalcedo.com · salcedo@cornell.edu

space we consider in this class is a vector space over $\mathbb{R}$. With this understanding in mind, we can omit the reference to $\mathbb{R}$, and simply say that V is a vector field.

The most common vector space we work with is the set of n-tuples of real numbers $\mathbb{R}^n = \{\mathbf{x} = (x_1, \ldots, x_n)^\top \mid \forall i = 1, \ldots, n, x_i \in \mathbb{R}\}$ endowed with vector addition and scalar product defined by

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

and

$$a \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a \cdot x_1 \\ a \cdot x_2 \\ \vdots \\ a \cdot x_n \end{pmatrix}$$

The identity element of the sum is $\mathbf{0} = (0, \ldots, 0)^\top$, and the inverse of $\mathbf{x}$ is $-\mathbf{x} = (-x_1, \ldots, -x_n)^\top$. It is straightforward to verify that $\mathbb{R}^n$ is a linear space over the field of real numbers, it is called the n-dimensional *Euclidean space*. In the case of $\mathbb{R}^2$ and $\mathbb{R}^3$ we can think of vectors as arrows, and addition and scalar product can be drawn. See Figure 1.

Example 1.1 For any non-empty set X, let $\mathcal{C}(X)$ denote the set of all continuous functions $\mathbf{f} : X \to \mathbb{R}$. $\mathcal{C}(X)$ endowed with the usual addition $(\mathbf{f} + \mathbf{g})(x) = \mathbf{f}(x) + \mathbf{g}(x)$ and scalar product $(a \cdot \mathbf{f})(x) = a \times \mathbf{f}(x)$ is a vector space over $\mathbb{R}$.

Example 1.2 Let $(\Omega, \mathscr{F}, \Pr)$ be a probability space. A random variable is a $\mathscr{F}$-measurable function $\mathbf{x} : \Omega \to \mathbb{R}$. Since addition and scalar product of functions preserves measurability, the set of random variables over $(\Omega, \mathscr{F}, \Pr)$ with the usual addition and scalar product is a vector space over $\mathbb{R}$.

Given a vector space $(V, +, \cdot, \mathbf{0})$, a set $W \subseteq V$ is a *subspace* of V if it is a vector space on its own when endowed with the restrictions of $+$ and $\cdot$ to W. It is easy to show that $0 \cdot \mathbf{v} = \mathbf{0}$ and $-1 \cdot \mathbf{v} = -\mathbf{v}$ for every $\mathbf{v} \in V$. Hence, $W \subseteq V$ is a subspace if and only if $\mathbf{u} + \mathbf{v} \in W$ and $a \cdot \mathbf{v} \in W$ for all $\mathbf{u}, \mathbf{v} \in W$ and $a \in \mathbb{R}$ (why?). In

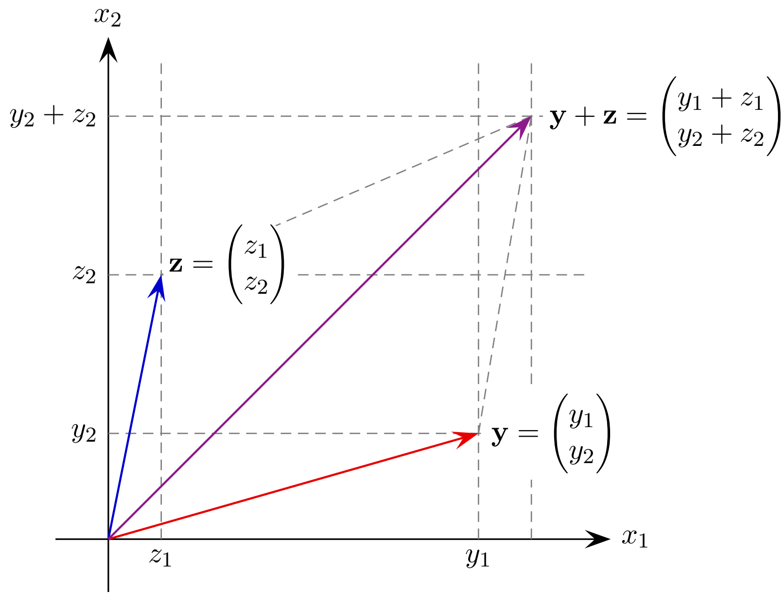

Figure 1 – Vector addition in $\mathbb{R}^2$.

particular, $\{\mathbf{0}\}$ is always a subspace. Moreover, every subspace contains $\mathbf{0}$.

Example 1.3 Let $X \subseteq \mathbb{R}$ and let $\mathcal{P}_n(X) \subseteq \mathcal{C}(X)$ be the set of all polynomials on X of order at most n. For instance $\mathbf{p}(x) = 3x^3 + 2x - 1$ is an element of $\mathcal{P}_n(X)$ for $n \geq 3$. Since the sum of polynomials of order at most n is a polynomial of order at most n, and the scalar product of a number times a polynomial of order at most n is also a polynomial of order at most n, it follows that $\mathcal{P}_n(X)$ is a subspace of $\mathcal{C}(X)$.

Example 1.4 For $M > 0$, let $\mathcal{B}_M(X) \subseteq \mathcal{C}(X)$ be the set of continuous functions on X that are uniformly bounded by M, i.e., such that $|\mathbf{f}(x)| \leq M$ for all $x \in X$. In particular, the constant function $\mathbf{f}_M(x) \equiv M$ belongs to $\mathcal{B}_M(X)$. The set $\mathcal{B}_M(X)$ is *not* a subspace of $\mathcal{C}(X)$, because $2 \cdot \mathbf{f}_M \notin \mathcal{B}_M(X)$.

Example 1.5 Fix an arbitrary vector $\mathbf{x}^* \in \mathbb{R}^n$ and a scalar $c \in \mathbb{R}$. The set $H = \{\mathbf{y} \in \mathbb{R}^n \mid \sum_i x_i^* y_i = c\}$ is a subspace of $\mathbb{R}^n$ if and only if $c = 0$. Indeed, it is easy to verify that, if $c = 0$, then the vector addition and scalar product are closed in H. On the other hand, if $c \neq 0$, then $\mathbf{0} \notin H$, and thus H is not a linear space.

2. Norm

A norm is a measure of the size of an object. In $\mathbb{R}^2$, where vectors can be thought of as arrows, we can thin of the size of the vector as the length of the corresponding arrow. By Pythagoras theorem, the length of $\mathbf{x} = (x_1, x_2)$ is given by $\sqrt{x_1^2 + x_2^2}$. See Figure 2. This notion of size is called the 2 and it can be generalized to $\mathbb{R}^n$. The Euclidean norm of $\mathbf{x} \in \mathbb{R}^n$ is defined to be

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

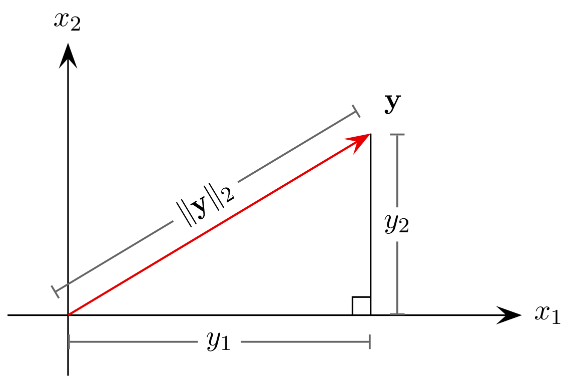

Figure 2 – The Euclidean norm of $\mathbf{y} \in \mathbb{R}^2$ satisfies $\|\mathbf{y}\|_2^2 = y_1^2 + y_2^2$.

The definition of Euclidean norm if fit only for finite tuples of real numbers, and cannot be directly used to measure the size of vectors in general spaces. However, it served to inspire a definition of the properties that a measure of vector sizes should satisfy. A *norm* over a vector space X is a function $\|\cdot\| : V \to \mathbb{R}$ such that

(N1) $\forall \mathbf{x} \in X, \quad \|\mathbf{x}\| \geq 0$ with equality if and only if $\mathbf{x} = \mathbf{0}$

(N2) $\forall \mathbf{x} \in X, \forall a \in \mathbb{R}, \quad \|a\mathbf{x}\| = |a|\|\mathbf{x}\|$

(N3) $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in X, \quad \|x + y\| \leq \|x\| + \|y\|$

The last property is called the *triangle inequality*. This is because, in $\mathbb{R}^2$, it is equivalent to saying that the closest distance between two points is a straight line. See Figure 3.

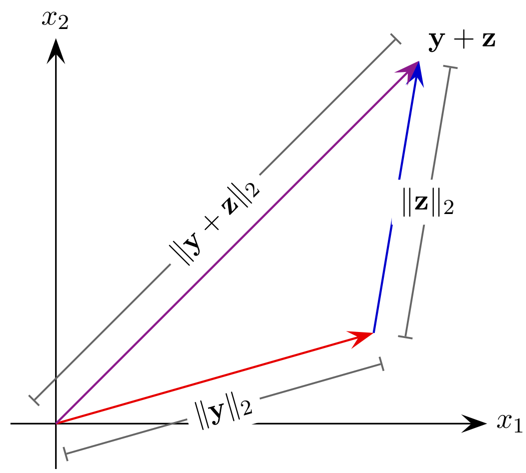

Figure 3 – Triangle inequality for the Euclidean norm in $\mathbb{R}^2$.

Proposition 2.1 *The Euclidean norm $\|\cdot\|_2$ is a norm in $\mathbb{R}^n$.*

Proving that the Euclidean norm satisfies (N1) and (N2) is straightforward. Establishing the triangle inequality requires some additional work. I will omit a formal proof, because it is a consequence of propositions 4.1 and 4.3, which are proved further below. The general definition of a norm allows us to define different ways of measuring the size of a vector which are useful in different applications, and it allows to measure vectors in general spaces other than $\mathbb{R}^n$.

Example 2.1 p-norm. For $p \geq 1$, the p-norm on $\mathbb{R}^n$ is defined by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

As with the euclidean norm, showing that $\|\cdot\|_p$ satisfies (N1) and (N2) is straightforward. The triangle inequality for p-norms is called Hölder inequality and requires a non-trivial proof (which you can easily find online if you are interested).

Example 2.2 Sup-norm. For functions $\mathbf{f} \in \mathcal{C}(X)$, let $\|\mathbf{f}\|_\infty = \sup\{|f(x)| \mid x \in X\}$. If X is infinite, then $\|\cdot\|_p$ is *not* a norm in $\mathcal{C}(X)$ because there exist functions for which $\|\mathbf{f}\|_\infty = \infty \notin \mathbb{R}$. However, $\|\cdot\|_\infty$ is a norm on the subspace $\mathcal{B}(X) \subseteq \mathcal{C}(X)$ consisting of all bounded and continuous functions on X. It is called the *sup-norm* and it plays an important role in many applications.

Example 2.3 Variance. The variance defined by $\mathbb{V}[\,\mathbf{x}\,] = \mathbb{E}[\,(\mathbf{x} - \mu_x)^2\,]$ is a norm on the space of random $\mathbf{x}$ with $\mathbb{E}[\,x\,] = 0$ and $\mathbb{E}[\,x^2\,] < \infty$.

A vector space endowed with a norm is called a *normed vector space*. A normed vector space satisfying a property called completeness, which we will define later on is called a *Banach space*. Banach spaces play an important role in many applications, for instance, to derive the contraction mapping theorem that you should be studying in the Macro sequence around this time.

3. Angles in $\mathbb{R}^2$

The *dot product* of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ is defined to be $\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2$. Note that the Euclidean norm in $\mathbb{R}^2$ can be expressed as $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x} \cdot \mathbf{x}}$. Moreover, being that the cosine function in injective in $[0, \pi)$, the following proposition implies that we can define the angle between $\mathbb{R}^2$ vectors in terms of the dot product.

Proposition 3.1 *The angle $\theta \in [0, \pi)$ between any two non-null vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2 \setminus \{\mathbf{0}\}$ satisfies $\cos(\theta) = \mathbf{x} \cdot \mathbf{y} / \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$.*

Proof. The proof is geometric. I only considers the case $0 < \theta < \pi/2$, so that $\mathbf{x}$ and $\mathbf{y}$ form an acute angle as in Figure 4. The case for obtuse angles follows from $\cos(\pi - \theta) = -\cos(\theta)$, and the cases $\theta = 0$ and $\theta = 1$ can be established by continuity of cos and $\cdot$.

The first step is to find the point on $\mathrm{span}(\mathbf{x})$ which is closest to $\mathbf{y}$. This point can be written as $\alpha \cdot \mathbf{x}$ for some $\alpha > 0$ (why?). Also, let $\mathbf{z} = \mathbf{y} - \alpha \cdot \mathbf{x}$. Note that $\mathbf{y}$, $\alpha \cdot \mathbf{x}$, and $\mathbf{z}$ have the same length as the sides of a right triangle with θ as one of the angles. Once again, see the figure.

On one hand, since the cosine of an angle equals the ratio of the adjacent side and the hypotenuse, we have that

$$\cos \theta = \frac{\alpha \|\mathbf{x}\|_2}{\|\mathbf{y}\|_2}. \tag{1}$$

On the other hand, from Pythagoras theorem, it follows that

$$\begin{aligned}\|\mathbf{y}\|_2^2 &= \|\alpha \cdot \mathbf{x}\|_2^2 + \|\mathbf{z}\|_2^2 \\ &= \alpha^2 \|\mathbf{x}\|_2^2 + (\mathbf{y} - \alpha \cdot \mathbf{x}) \cdot (\mathbf{y} - \alpha \cdot \mathbf{x}) \\ &= \alpha^2 \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 - 2\alpha \mathbf{x} \cdot \mathbf{y} + \alpha^2 \|\mathbf{x}\|_2^2\end{aligned}$$

Solving for α yields $\alpha = \mathbf{x} \cdot \mathbf{y} / \|\mathbf{x}\|_2^2$. And substituting this value into (1) yields the result. ■

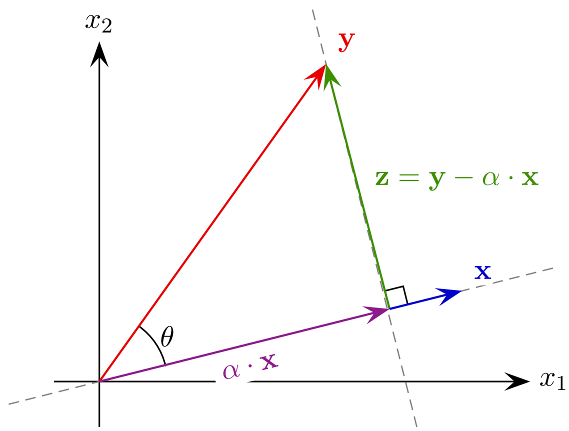

Figure 4 – Angle between vectors in $\mathbb{R}^2$.

The vector $\alpha \cdot \mathbf{x}$ in the proof of Proposition 3.1 is called the *projection* of $\mathbf{y}$ over $\mathbf{x}$. Having solved for the value of α, it is given by

$$p_{\mathbf{x}}(\mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{x} \cdot \mathbf{x}} \mathbf{x}. \tag{2}$$

Proposition 3.1 can be taken to be a definition of the angle between non-null vectors in $\mathbb{R}^2$ is terms of the dot product. If we could generalize the notion of dot product to general linear spaces, we could also generalize a notion of angle between vectors. The dot product in $\mathbb{R}^n$ is the binary operation $\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$x \cdot y = \sum_{i=1}^{n} x_i y_i.$$

For general linear spaces we use the notion of inner product defined in the next section.

4. Inner product

Given a linear space V, and *inner product* is a function $\langle \cdot, \cdot \rangle : V \times V \to F$ satisfying

(IP1) Commutativity — $\forall \mathbf{x}, \mathbf{y} \in X, \quad \langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$

(IP2) Linearity — $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in X, \forall a \in F \quad \langle a(\mathbf{x}+\mathbf{z}), \mathbf{z} \rangle = a(\langle \mathbf{x}+\mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle)$

(IP3) Positive definiteness — $\forall \mathbf{x} \in X, \quad \langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ with equality only if $\mathbf{x} = \mathbf{0}$

Proposition 4.1 *The dot product on $\mathbb{R}^n$ is an inner product.*

Proof. The proof follows almost directly from the algebraic properties of $\mathbb{R}$. Consider any $\mathbf{x}, \mathbf{y}, \mathbf{x} \in \mathbb{R}^n$ and $a \in A$. (*Commutativity*) $\mathbf{x} \cdot \mathbf{y} = \sum_i x_i y_i = \sum_i y_i x_i = \mathbf{y} \cdot \mathbf{x}$. (*Linearity*) $(a \cdot (\mathbf{x}+\mathbf{y})) \cdot \mathbf{y} = \sum_i a(x_i + y_i) z_i = a(\sum_i x_i z_i + \sum_i y_i z_i) = a((\mathbf{x} \cdot \mathbf{y}) + (\mathbf{y} \cdot \mathbf{z}))$. (*Positive definiteness*) $\mathbf{x} \cdot \mathbf{x} = \sum_i x_i^2 \geq 0$ with equality if and only if $x_i = 0$ for all $i = 1, \dots, n$. ■

Just like the Euclidean norm can be defined in terms of the dot product, every inner product on a general space induces a norm. Given any linear space V and an inner product $\langle \cdot, \cdot \rangle$, the norm induced by $\langle \cdot, \cdot \rangle$ is the function $\| \cdot \| : V \to \mathbb{R}$ defined by $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Normed spaces whose norm can be induced by an inner product are called *inner product spaces*, and if they are complete, they are called *Hilbert spaces*. Before proving that the norm indiced by an inner product is indeed a norm, we need to establish an important inequality.

Proposition 4.2 (Cauchy-Schwarz inequality) *Given any linear space V. If $\langle \cdot, \cdot \rangle$ is an inner product on V then $\langle \mathbf{x}, \mathbf{y} \rangle \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in V$, where $\| \cdot \|$ is the norm induced by $\langle \cdot, \cdot \rangle$.*

Proof. If $\mathbf{x} = \mathbf{0}$ then the result is trivial. The proof mimics the construction from the proof of Proposition 3.1. Let $\mathbf{z}$ be the vector defined by $\mathbf{z} = \mathbf{y} - \alpha \mathbf{x}$, where $\alpha = \langle \mathbf{x}, \mathbf{y} \rangle / \langle \mathbf{x}, \mathbf{x} \rangle$ (notice the similarity with 2). Now, note that

$$\begin{aligned} 0 \leq \langle \mathbf{z}, \mathbf{z} \rangle &= \langle \mathbf{y} - \alpha \cdot \mathbf{x}, \mathbf{y} - \alpha \cdot \mathbf{x} \rangle \\ &= \langle \mathbf{y}, \mathbf{y} \rangle - 2\alpha \langle \mathbf{y}, \mathbf{x} \rangle + \alpha^2 \langle \mathbf{x}, \mathbf{x} \rangle \end{aligned}$$

$$
\begin{aligned}
&= \langle \mathbf{y}, \mathbf{y} \rangle - 2 \left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} \right) \langle \mathbf{y}, \mathbf{x} \rangle + \left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} \right)^2 \langle \mathbf{x}, \mathbf{x} \rangle \\
&= \langle \mathbf{y}, \mathbf{y} \rangle - \frac{\langle \mathbf{x}, \mathbf{y} \rangle^2}{\langle \mathbf{x}, \mathbf{x} \rangle}
\end{aligned}
$$

Rearranging terms, this yields

$$
\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \langle \mathbf{y}, \mathbf{y} \rangle \langle \mathbf{x}, \mathbf{x} \rangle.
$$

Taking square root yields the desired inequality. ■

Proposition 4.3 *The norm induced by an inner product is a norm.*

Proof. Let V be a linear space with an inner product $\langle \cdot, \cdot \rangle$, and let $\|\cdot\|$ be the norm induced by it. The positive definiteness and linearity of the inner product trivially imply that $\|\cdot\|$ satisfies (N1) and (N2). It only remains to establish the triangle inequality. For that purpose, let $\mathbf{x}$ and $\mathbf{y}$ be arbitrary vectors in V. Note that

$$
\begin{aligned}
(\|\mathbf{x}\| + \|\mathbf{y}\|)^2 &= \langle \mathbf{x}, \mathbf{x} \rangle + 2 \|\mathbf{x}\| \, \|\mathbf{y}\| + \langle \mathbf{y}, \mathbf{y} \rangle \\
&\geq \langle \mathbf{x}, \mathbf{x} \rangle + 2\langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle = \|\mathbf{x} + \mathbf{y}\|^2,
\end{aligned}
$$

where the inequality follows from the Cauchy-Schwatz inequality, and the equalities follow from the linearity of the inner product. ■

Given an inner product space X and two non-null vectors $\mathbf{x}, \mathbf{y} \in X$, the angle between $\mathbf{x}$ and $\mathbf{y}$ is defined to be

$$
\angle(\mathbf{x}, \mathbf{y}) := \cos^{-1} \left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \right), \tag{3}
$$

where $\|\cdot\|$ denotes the norm induced by the inner product. As for null-vectors, for all $\mathbf{x} \in V$ we set $\angle(\mathbf{x}, \mathbf{0}) = 0$ by convention.

Vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be *orthogonal* if $\angle(\mathbf{x}, \mathbf{y}) \in \{\pi/2, 3\pi/2\}$, and are said to be *parallel* if $\angle(\mathbf{x}, \mathbf{y}) \in \{0, \pi, \}$. It is easy to show that $\mathbf{x}$ and $\mathbf{y}$ are parallel if and only if there exists some $a \in \mathbb{R}$ such that $\mathbf{x} = a\mathbf{y}$, and they are orthogonal if and only if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

Example 4.1 OLS. Given a probability space, let X be the space of random variables with mean 0. It is easy to show that the covariance $\mathbb{C}[\mathbf{x},\mathbf{y}] = \mathbb{E}[\mathbf{xy}]$ is an inner product. The norm induced by the covariance is the standard deviation defined by $\|\mathbf{x}\| = \sqrt{\mathbb{E}[\mathbf{x}^2]}$. Defining projections the same way we did for $\mathbb{R}^2$ the projection of $\mathbf{y}$ over $\mathbf{x}$ is then given by $p_{\mathbf{x}}(\mathbf{y}) = (\mathbb{E}[\mathbf{x}^2])^{-1}\mathbb{E}[\mathbf{xy}]\,\mathbf{x}$, which corresponds to the OLS predictions for $\mathbf{y}$ using $\mathbf{x}$ as a regressor. Moreover, the angle between $\mathbf{x}$ and $\mathbf{y}$ is nothing more than the cosine of their correlation.

5. Basis and dimension

Fix a linear space V. We say that $\mathbf{x} \in V$ is a linear combination of $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\} \subseteq V$, if there exists scalars $c_1, \ldots, c_n$ such that $\mathbf{x} = \sum_{i=1}^n c_i \cdot \mathbf{y}_i$. Given a set of vectors Y, the *span* of Y, denoted by $\operatorname{span}(Y)$ is the set of all linear combinations of elements of Y. Y is said to span V if $V = \operatorname{span}(Y)$.

Proposition 5.1 *For every non-empty set of vectors Y, $\operatorname{span}(Y)$ is a subspace.*

Proof. Let $\mathbf{x}, \mathbf{y} \in \operatorname{span}(Y)$ and $\alpha \in \mathbb{R}$. Then, there exist finite sets $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subseteq Y$, $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\} \subseteq Y$, and $\{c_1, \ldots, c_n\}, \{d_1, \ldots, d_m\} \subseteq \mathbb{R}$ such that $\mathbf{x} = \sum_{i=1}^n c_i\mathbf{x}_i$ and $\mathbf{y} = \sum_{i=1}^m d_i\mathbf{y}_i$. Then, $\mathbf{x} + \mathbf{y} \sum_{i=1}^n c_i\mathbf{x}_i + \sum_{i=1}^m d_i\mathbf{y}_i \in \operatorname{span}(Y)$, and $\alpha \cdot \mathbf{x} = \sum_{i=1}^n (\alpha c_i)\mathbf{x}_i \in \operatorname{span}(Y)$. ■

Example 5.1 The span of a vector $\mathbf{x}$ is the line $\operatorname{span}(\{\mathbf{x}\}) = \{\alpha \cdot \mathbf{x} \mid \alpha \in \mathbb{R}\}$.

Example 5.2 For $n \in \mathbb{N}$, $\mathbf{f}_n(x) = x^n$. The set $\{\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_n\}$ spans the set $P_n(\mathbb{R})$ consisting of all polynomials of order at most n. Indeed, a function $\mathbf{f} : \mathbb{R} \to \mathbb{R}$ is a polynomial of order at most n if and only if it can be written as $\mathbf{f}(x) = \sum_{i=0}^n a_i x_i^n$, that is, if and only if $\mathbf{f} = \sum_{i=0}^n a_i \cdot \mathbf{f}_i$ for some $a_0, \ldots, a_n \in \mathbb{R}$.

A set of vectors Y is said to be *linearly independent* if $\mathbf{y} \notin \operatorname{span}(Y \setminus \{\mathbf{y}\})$ for all $\mathbf{y} \in Y$. Equivalently, Y is linearly dependent if and only if there do *not* exist $\mathbf{y}_1, \ldots, \mathbf{y}_n \in Y$ and $c_1, \ldots, c_n \in \mathbb{R}$ not all zeros such that $\mathbf{0} = \sum_{i=1}^n c_i\mathbf{y}_i$. Otherwise, we say that the are linearly dependent.

A set $Y \subseteq V$ is a *basis* if it is linearly dependent and it spans V. Using the definition of minimum and maximal with respect to $\subseteq$, it is straightforward to see that Y is a basis if and only if is a maximal element of the set of linearly independent subsets on V. Also, Y is a basis if and only if is a minimum element of the set of subsets of V that span V. See the following example.

Example 5.3 A basis of $\mathbb{R}^n$ is $\{\mathbf{e}^i \mid i = 1, \ldots, n\}$. It is a minimum spanning set because, if we remove any element from it, the resulting set will not span $\mathbb{R}^n$. It is a maximal spanning set because, if we add any other vector, the resulting set is not linearly independent.

Example 5.4 The set $\{(1,1)^\mathsf{T}, (-1,1)^\mathsf{T}\}$ is also a basis for $\mathbb{R}^2$. Indeed, for any $\mathbf{x} \in \mathbb{R}^2$ we have that

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{x_2 + x_1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{x_2 - x_1}{2} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

There is nothing particular about the vectors in the previous example, except that they are not parallel. In fact, as we will show below, a subset of $\mathbb{R}^2$ is a basis if and only if it consists of exactly two linearly independent vectors. Since all basis of $\mathbb{R}^2$ have exactly two elements, we say that $\mathbb{R}^2$ is two-dimensional. Analogous results hold for general linear spaces.

Theorem 5.2 *All basis of the same space have the same cardinality.*

The proof of this theorem for general vector spaces require tools that are far beyond the scope of this course. So, we will only prove it for vectors that have a finite basis. The key step of the proof is to establish the following lemma.

Lemma 5.3 *Let V be a vector space. If $X \subseteq V$ is linearly independent and $Y \subseteq V$ is finite and spans V, then $\|X\| \leq \|Y\|$.*

Proof. Let $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$ and X be as in the statement of the lemma. First, we consider the case that X is finite, i.e., $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ for some $n \in \mathbb{N}$. We want to show that $m \geq n$. The case with $n = 1$ is trivial. So, for the rest of the proof, assume $n \geq 2$. Since X is linearly independent, we know that $\mathbf{x}_i \neq \mathbf{0}$ for all $i = 1, \ldots, n$ (why?). We will use this fact a couple of times.

Since $\operatorname{span}(Y) = V$, there exist scalars $c_{11}, \ldots, c_{1m}$ such that $\mathbf{x}_1 = \sum_{i=1}^m c_{1i}\mathbf{y}_i$.

Since $\mathbf{x}_1 \neq \mathbf{0}$, at least one of this scalars bust be different from 0, Assume without loss of generality that $c_{11} \neq 0$. Then, we can write

$$\mathbf{y}_1 = \frac{1}{c_{11}}\mathbf{x}_1 - \sum_{i=2}^{m} \frac{c_{1i}}{c_{11}}\mathbf{y}_i.$$

This means that $\mathbf{y}_1 \in \operatorname{span}(Y_1)$, where $Y_1 = (Y \setminus \{\mathbf{y}_1\}) \cup \{\mathbf{x}_1\}$. Therefore, $Y \subseteq \operatorname{span}(Y_1)$ and, consequently, $V = \operatorname{span}(Y) \subseteq \operatorname{span}(Y_1)$.

We can do a similar procedure for $\mathbf{x}_2$. Since $\mathbf{0} \neq \mathbf{x}_2 \in \operatorname{span}(Y_1)$, there exist scalars $c_{21}, \ldots, c_{2m}$ such that $\mathbf{x}_2 = c_{21}\mathbf{x}_1 + \sum_{i=2}^{m} c_{2i}\mathbf{y}_i$. Since X is linearly independent, it must be the case that $c_{2i} \neq 0$ for some $i > 1$ (otherwise $\mathbf{x}_2$ would be a linear combination of $\mathbf{x}_1$). Assume without loss of generality that $c_{22} \neq 0$. Then we can write

$$\mathbf{y}_2 = \frac{1}{c_{22}}\mathbf{x}_2 - \frac{c_{21}}{c_{22}}\mathbf{x}_2 - \sum_{i=3}^{m} \frac{c_{2i}}{c_{22}}\mathbf{y}_i.$$

As before, this implies that $\mathbf{y}_2 \in \operatorname{span}(Y_2)$, where $Y_2 = (Y \setminus \{\mathbf{y}_1, \mathbf{y}_2\}) \cup \{\mathbf{x}_1, \mathbf{x}_2\}$. And thus, $\operatorname{span}(Y_2) = V$.

If $n = 2$, then the proof would be complete because we already showed that Y contains two distinct elements $\mathbf{y}_1$ and $\mathbf{y}_2$. Otherwise, we can repeat the last step for $\mathbf{x}_3, \ldots, \mathbf{x}_n$ until we arrive to a set $Y_n = (Y \setminus \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}) \cup \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. Since each of the elements of $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\|$ is distinct and $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\} \subseteq Y$, we know that $m = \|Y\| \geq \|\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}\| = n$.

We have shown that if a finite set Y spans V, then any finite linearly independent set has at most $\|Y\|$ elements. Being that subsets of linear independent sets are also linearly independent, this implies that there cannot be any infinite linearly independent sets (because they would have finite subsets with more elements than Y). ■

Proof of Theorem 5.2 for finite-dimensional spaces. Suppose that V has a finite basis, and let $X, Y \subseteq V$ be bases. Since X and Y are linearly independent, Lemma 5.3 implies that they are finite. Since X is a finite spanning set and Y is linearly independent, Lemma 5.3 implies that $\|Y\| \leq \|X\|$. By an analogous argument, $\|X\| \leq \|Y\|$. ■

Theorem 5.2 asserts that the cardinality of a basis is a fundamental property of the linear space, that does not depend on which specific basis we consider. The *dimension* of V, denoted by $\dim(V)$ is the cardinality of its bases. A space is *finite-dimensional* if $\dim(V) \in \mathbb{N}$ and otherwise it is said to be *infinite-dimensional.*

Corollary 5.4 *For any $n \in \mathbb{N}$, the dimension of $\mathbb{R}^n$ is n.*

Proof. $\{\mathbf{e}^i \mid i = 1, \dots, n\}$ is a basis for n. ■

Proposition 5.5 *If V is a finitely-dimensional vector space, then $Y \subseteq V$ is a basis for V if and only if it is linearly independent and $\|Y\| = \dim(V)$.*

Proof. Necessity follows directly from the definitions of dimension and basis. For sufficiency, let $Y = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ be a linearly independent set with $\|Y\| = \dim(V)$, and consider an arbitrary vector $\mathbf{x} \notin Y$. Since $\|Y \cup \{\mathbf{x}\}\| > \dim(V)$, Lemma 5.3 implies that $Y \cup \{\mathbf{x}\}$ is linearly dependent. Therefore, there exist constants $\{c_0, c_1, \dots, c_n\}$, not all zeros, such that $c_0\mathbf{x} + \sum_{i=1}^n c_i\mathbf{y}_i = \mathbf{0}$. Since Y is linearly independent, it cannot be the case that $c_0 = 0$ (why?). Hence, we can write $\mathbf{x} = \sum_{i=1}^n (-c_i/c_0)\mathbf{y}_i$, which implies that $\mathbf{x} \in \operatorname{span}(Y)$. Since $\mathbf{x}$ was arbitrary, we can conclude that $\operatorname{span}(Y) = V$. ■

6. Convexity

Fix a linear space V. A vector $\mathbf{z} \in V$ is a *convex combination* of $\mathbf{x}$ and $\mathbf{y}$ if there exists some $\mu \in [0, 1]$ such that $\mathbf{z} = \mu\mathbf{x} + (1 - \mu)\mathbf{y}$. A set $X \subseteq V$ is *convex* if it contains all the convex combinations of its elements. Otherwise, X is said to be non-convex (there is no such a thing as concave sets).

Example 6.1 In $\mathbb{R}^2$ and $\mathbb{R}^3$, the set of convex combinations of $\mathbf{x}$ and $\mathbf{y}$ corresponds to the line segment between them. Hence, a set is convex if and only every line segment joining points in the set is completely included in the set. See the left panel of Figure 5.

Example 6.2 Let $\mathbf{f}_0, \mathbf{f}_1 : [-2, 2] \to \mathbb{R}$ be given by $\mathbf{f}_0(x) = 1$ and $\mathbf{f}_1(x) = x$. A function $\mathbf{f} : [-2, 2] \to \mathbb{R}$ is a convex combination of $\mathbf{f}_0$ and $\mathbf{f}_1$ if and only if it can be written as $\mathbf{f}(x) = \theta x + (1 - \theta)$ for some $\theta \in [0, 1]$. The graph of each such function is an upward sloping straight line within the shaded region in the right panel of Figure 5.

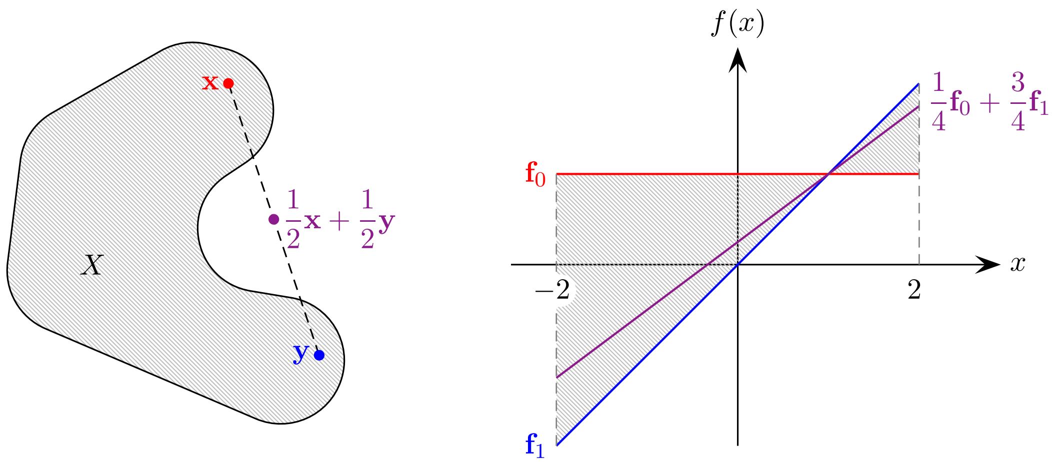

Figure 5 – Left panel: Non-convex set $X \subseteq \mathbb{R}^2$. Right panel: Convex combinations of $\mathbf{f}_0(x) = 1$ and $\mathbf{f}_1(x) = x$ in $\mathcal{C}([-2, 2])$.

The definition of convex combinations can be extended from pairs of vectors to finite sets. The n-dimensional *simplex* is the set $\Delta^n \subseteq \mathbb{R}^n$ consisting of vectors $\lambda \in \mathbb{R}^n$ such that $\sum_i \lambda_i = 1$, and $0 \leq \lambda_i \leq 1$ for all $i = 1, \ldots, n$. We say that $\mathbf{u} \in V$ is a convex combination of $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subseteq V$ if there exist a vector of weights $\lambda \in \Delta^n$ such that $\mathbf{u} = \sum_i \lambda_i \mathbf{v}_i$. The *convex hull* of a set X is the set denoted by $\mathrm{co}(X)$ consisting of all convex combinations of finite subsets of X.

Proposition 6.1 *A set is convex if and only if it coincides with its convex hull.*

Proof. The proof is assigned as a homework exercise in Problem Set 4. ■

Proposition 6.2 *The convex hull of a set $X \subseteq V$ is the smallest convex set containing X.*

Proof. Take any set $X \subseteq V$. Since any vector is a convex combination of itself ($\mathbf{x} = 1 \cdot \mathbf{x}$), it follows that $X \subseteq \mathrm{co}(X)$. Now take any $\mathbf{x}, \mathbf{y} \in \mathrm{co}(X)$ and $\theta \in [0, 1]$. It follows from the definition of convex hull that there exist finite collections $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}, \{\mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+m}\} \subseteq X$ and vectors of weights $(\lambda_1, \ldots, \lambda_n) \in \Delta^n$ and $(\lambda_{n_1}, \ldots, \lambda_{n+m}) \in \Delta^m$, such that $\mathbf{x} = \sum_{i=1}^{n} \lambda_i \mathbf{x}_i$ and $\mathbf{y} = \sum_{i=1}^{m} \lambda_{n+i} \mathbf{x}_{n+i}$.

Let $\mu \in \mathbb{R}^{n+m}$ be given by $\mu_i = \theta\lambda_i$ for $i = 1 \ldots, n$, and $\mu_i = (1-\theta)\lambda_i$ for $i = n+1, \ldots, n+m$. Note that

$$\theta\mathbf{x} + (1-\theta)\mathbf{y} = \sum_{i=1}^{n}(\theta\lambda_i)\mathbf{x}_i + \sum_{i=n+1}^{n+m}((1-\theta)\lambda_i)\mathbf{x}_i = \sum_{i=1}^{n+m}\mu_i\mathbf{x}_i.$$

Moreover it is straightforward to see that $0 \leq \mu_i \leq 1$ for all $i = 1, \ldots, n+m$, and

$$\sum_{i=1}^{n+m}\mu_i = \theta\sum_{i=1}^{n}\lambda_i + (1-\theta)\sum_{i=n+1}^{n+m}\lambda_i = \theta + (1-\theta) = 1.$$

Hence, $\mu \in \Delta^{n+m}$ and, consequently, $\theta\mathbf{x} + (1-\theta)\mathbf{y} \in \mathrm{co}(X)$. Since $\mathbf{x}$, $\mathbf{y}$ and θ where arbitrary, this implies that $\mathrm{co}(X)$ is a convex set containing X. It remains to show that it is the smallest of such sets.

For that purpose, we will first show that the convex hull operator is $\subseteq$-monotone. Let $U \subseteq W \subseteq V$. We will show that $\mathrm{co}(U) \subseteq \mathrm{co}(W)$. For any $\mathbf{x} \in \mathrm{co}(U)$, there exist $\{\mathbf{x}_1, \ldots \mathbf{x}_n\} \subseteq U$ and $\lambda \in \Delta^n$ such that $\mathbf{x} = \sum_i \lambda_i\mathbf{x}_i$. Since $U \subseteq V$ it follows that $\{\mathbf{x}_1, \ldots \mathbf{x}_n\} \subseteq V$ and, consequently, $\mathbf{x} \in \mathrm{co}(V)$.

Returning to the problem at hand, let $Y \subseteq V$ be a convex set containing X. Since $\mathrm{co}(\,\cdot\,)$ is $\subseteq$-monotone and $X \subseteq Y$, it follows that $\mathrm{co}(X) \subseteq \mathrm{co}(Y)$. Proposition 6.1 implies that $Y = \mathrm{co}(Y)$. Therefore, $\mathrm{co}(X) \subseteq Y$, thus completing the proof. ■

Geometrically, the convex hull is a very simple object. If we think of $\mathbb{R}^2$ as a board and each vector in X as a nail on it, then the convex hull of X is the set delimited by a elastic band wrapped around X. See Figure 6. However, the convex hull can be hard to characterize because it could potentially involve convex combinations of arbitrarily large sets. Thankfully, the following result guarantees that it suffices to consider convex combinations of at most $\dim(V) + 1$ vectors.

Theorem 6.3 (Caratheodory) *Let X be a subset of an n-dimensional space. For every point $\mathbf{x} \in \mathrm{co}(X)$, there exist $Y \subseteq X$ with $\|Y\| \leq n+1$, such that $\mathbf{x} \in \mathrm{co}(Y)$.*

Proof. There exist $Y = \{\mathbf{x}_1, \ldots \mathbf{x}_m\} \subseteq X$ and $\lambda \in \Delta^m$ such that

$$\mathbf{x} = \sum_{i=1}^{m}\lambda_i\mathbf{x}_i. \tag{4}$$

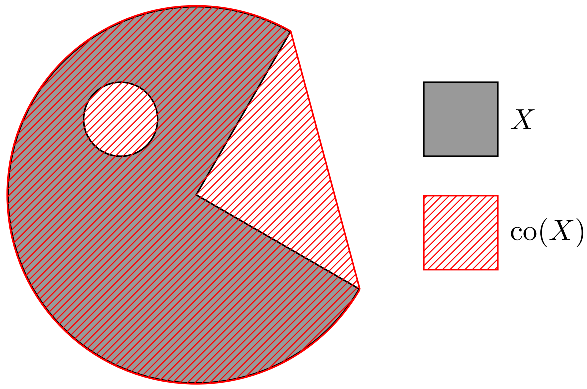

Figure 6 – Convex hull of a non-convex set X in $\mathbb{R}^2$.

If $m \leq n+1$, there is nothing to prove. Otherwise, we will show that there exists a *strict* subset of Y such that $\mathbf{x} \in \text{co}(Y)$. If were is some $i_0 \in \{1, \ldots, m\}$ such that $\lambda_i = 0$, then $\mathbf{x} \in \text{co}(Y \setminus \{\mathbf{x}_{i_0}\})$, and we would be done. For the rest of the proof, suppose that $\lambda_i > 0$ for all i.

Let $Z = \{\mathbf{x}_i - \mathbf{x}_m \mid i = 1, \ldots, m-1\}$. Since $\|Z\| = m - 1 \geq n$, Z is linearly dependent. Hence, there exist $\mu_1, \ldots, \mu_m \in \mathbb{R}$, not all zeros, such that $\sum_{i=1}^{m-1} \mu_i(\mathbf{x}_i - \mathbf{x}_m) = \mathbf{0}$. Define $\mu_m = -\sum_{i=1}^{m-1} \mu_i$ so that we can write

$$\sum_{i=1}^{m} \mu_i \mathbf{x}_1 = \mathbf{0} \tag{5}$$

and

$$\sum_{i=1}^{m} \mu_i = 0. \tag{6}$$

Condition (6) implies that there exists some i such that $\mu_i > 0$. Hence we can define $\alpha = \min\{\lambda_i/\mu_i \mid \mu_i > 0, i = 1, \ldots, m\}$. Let i_0 be an index such that $\alpha = \lambda_{i_0}/\mu_{i_0} > 0$. Let $\gamma \in \mathbb{R}^n$ be the vector given by $\gamma_i = \lambda_i - \alpha\mu_i$, for $i = 1, \ldots, m$. From (4) and (5), it follows that

$$\mathbf{x} = \sum_{i=1}^{m} \lambda_i \mathbf{x}_i - \alpha_0 \sum_{i=1}^{m} \mu_i \mathbf{x}_i = \sum_{i=1}^{m} \gamma_i \mathbf{x}_i.$$

Now, we will show that $\gamma \in \Delta^m$. Take an arbitrary index i. If $\mu_i \leq 0$, then $\gamma_i = \lambda_i - \alpha\mu_i \geq \lambda_i \geq 0$ because $\alpha > 0$. If $\mu_i > 0$, then $\lambda_i/\mu_i \geq \lambda_{i_0}/\mu_{i_0}$ from the definition of α. Therefor,e in this case,

$$\gamma_i = \lambda_i - \frac{\lambda_{i_0}}{\mu_{i_0}}\mu_i \geq \lambda_i - \frac{\lambda_i}{\mu_i}\mu_i = 0.$$

From (6), it follows that $\sum_i \gamma_i = \sum_i \lambda_i = 1$. Therefore $\gamma \in \Delta^m$. Moreover, note that $\gamma_{i_0} = 0$, which implies that $\mathbf{x}$ is a convex combination of $Y \setminus \{\mathbf{x}_{i_0}\}$. ∎