Linear models: Algebra

Linear models are a pillar of modern data analysis. Many phenomena can be modeled as linear, at least approximately. In addition, linear models tend to be easy to interpret (*quantity A is proportional to quantity B* is simpler to understand than *quantity A is proportional to e^B if $B < 1$ and to $1/B$ otherwise*). Finally, linear models are often convenient from a pragmatic point of view: fitting them to the data tends to be easy and computationally tractable. We will begin with an overview of linear algebra; being comfortable with these concepts is essential to understand linear-modeling methods.

1 Vector spaces

You are no doubt familiar with **vectors** in $\mathbb{R}^2$ or $\mathbb{R}^3$, i.e.

$$x = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad y = \begin{bmatrix} -1.1 \\ 0 \\ 5 \end{bmatrix}. \tag{1}$$

From the point of view of algebra, vectors are much more general objects. They are elements of sets called **vector spaces** that satisfy the following definition.

Definition 1.1 (Vector space)**.** *A vector space consists of a set $\mathcal{V}$ and two operations $+$ and $\cdot$ satisfying the following conditions.*

1. *For any pair of elements $x, y \in \mathcal{V}$ the* ***vector sum*** *$x + y$ belongs to $\mathcal{V}$.*
2. *For any $x \in \mathcal{V}$ and any scalar $\alpha \in \mathbb{R}$ the* ***scalar multiple*** *$\alpha \cdot x \in \mathcal{V}$.*
3. *There exists a* ***zero vector*** *or* ***origin*** *0 such that $x + 0 = x$ for any $x \in \mathcal{V}$.*
4. *For any $x \in \mathcal{V}$ there exists an additive inverse y such that $x + y = 0$, usually denoted as $-x$.*
5. *The vector sum is commutative and associative, i.e. for all $x, y \in \mathcal{V}$*
$$x + y = y + x, \quad (x + y) + z = x + (y + z). \tag{2}$$
6. *Scalar multiplication is associative, for any $\alpha, \beta \in \mathbb{R}$ and $x \in \mathcal{V}$*
$$\alpha\,(\beta \cdot x) = (\alpha\,\beta) \cdot x. \tag{3}$$

7. Scalar and vector sums are both distributive, i.e. for all $\alpha, \beta \in \mathbb{R}$ and $x, y \in \mathcal{V}$

$$(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x, \quad \alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y. \tag{4}$$

*A **subspace** of a vector space $\mathcal{V}$ is a subset of $\mathcal{V}$ that is also itself a vector space.*

From now on, for ease of notation we will ignore the symbol for the scalar product $\cdot$, writing $\alpha \cdot x$ as $\alpha\, x$.

Remark 1.2 (More general definition)**.** *We can define vector spaces over an arbitrary field, instead of $\mathbb{R}$, such as the complex numbers $\mathbb{C}$. We refer to any linear algebra text for more details.*

We can easily check that $\mathbb{R}^n$ is a valid vector space together with the usual vector addition and vector-scalar product. In this case the zero vector is the all-zero vector $\begin{bmatrix} 0 & 0 & 0 & \dots \end{bmatrix}^T$. When thinking about vector spaces it is a good idea to have $\mathbb{R}^2$ or $\mathbb{R}^3$ in mind to gain intuition, but it is also important to bear in mind that we can define vector sets over many other objects, such as infinite sequences, polynomials, functions and even random variables as in the following example.

Example 1.3 (The vector space of zero-mean random variables)**.** Zero-mean random variables belonging to the same probability space form a vector space together with the usual operations for adding random variables together and for multiplying random variables and scalars. This follows almost automatically from the fact that linear combinations of random variables are also random variables and from linearity of expectation. You can check for instance that if X and Y are zero-mean random variables, for any scalars α and β the random variable $\alpha X + \beta Y$ is also a zero-mean random variable. The zero vector of this vector space is the random variable equal to 0 with probability one.

The definition of vector space guarantees that any **linear combination** of vectors in a vector space $\mathcal{V}$, obtained by adding the vectors after multiplying by scalar coefficients, belongs to $\mathcal{V}$. Given a set of vectors, a natural question to ask is whether they can be expressed as linear combinations of each other, i.e. if they are **linearly dependent** or **independent**.

Definition 1.4 (Linear dependence/independence)**.** *A set of m vectors $x_1, x_2, \dots, x_m$ is linearly dependent if there exist m scalar coefficients $\alpha_1, \alpha_2, \dots, \alpha_m$ which are **not** all equal to zero and such that*

$$\sum_{i=1}^{m} \alpha_i\, x_i = 0. \tag{5}$$

*Otherwise, the vectors are **linearly independent**.*

Equivalently, at least one vector in a linearly dependent set can be expressed as the linear combination of the rest, whereas this is not the case for linearly independent sets.

Let us check the equivalence. Equation (5) holds with $\alpha_j \neq 0$ for some j if and only if

$$x_j = \frac{1}{\alpha_j} \sum_{i \in \{1,\ldots,m\}/\{j\}} \alpha_i \, x_i. \tag{6}$$

We define the **span** of a set of vectors $\{x_1, \ldots, x_m\}$ as the set of all possible linear combinations of the vectors:

$$\operatorname{span}(x_1, \ldots, x_m) := \left\{ y \mid y = \sum_{i=1}^{m} \alpha_i \, x_i \quad \text{for some } \alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R} \right\}. \tag{7}$$

This turns out to be a vector space.

Lemma 1.5. *The span of any set of vectors $x_1, \ldots, x_m$ belonging to a vector space $\mathcal{V}$ is a subspace of $\mathcal{V}$.*

Proof. The span is a subset of $\mathcal{V}$ due to Conditions 1 and 2 in Definition 1.1. We now show that it is a vector space. Conditions 5, 6 and 7 in Definition 1.1 hold because $\mathcal{V}$ is a vector space. We check Conditions 1, 2, 3 and 4 by proving that for two arbitrary elements of the span

$$y_1 = \sum_{i=1}^{m} \alpha_i \, x_i, \quad y_2 = \sum_{i=1}^{m} \beta_i \, x_i, \quad \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m \in \mathbb{R}, \tag{8}$$

$\gamma_1 \, y_1 + \gamma_2 \, y_2$ also belongs to the span. This holds because

$$\gamma_1 \, y_1 + \gamma_2 \, y_2 = \sum_{i=1}^{m} (\gamma_1 \, \alpha_i + \gamma_2 \, \beta_i) \, x_i, \tag{9}$$

so $\gamma_1 \, y_1 + \gamma_2 \, y_2$ is in $\operatorname{span}(x_1, \ldots, x_m)$. Now to prove Condition 1 we set $\gamma_1 = \gamma_2 = 1$, for Condition 2 $\gamma_2 = 0$, for Condition 3 $\gamma_1 = \gamma_2 = 0$ and for Condition 4 $\gamma_1 = -1, \gamma_2 = 0$. $\square$

When working with a vector space, it is useful to consider the set of vectors with the smallest cardinality that spans the space. This is called a **basis** of the vector space.

Definition 1.6 (Basis)**.** *A basis of a vector space $\mathcal{V}$ is a set of independent vectors $\{x_1, \ldots, x_m\}$ such that*

$$\mathcal{V} = \mathit{span}\,(x_1, \ldots, x_m)\,. \tag{10}$$

An important property of all bases in a vector space is that they have the same cardinality.

Theorem 1.7. *If a vector space $\mathcal{V}$ has a basis with finite cardinality then every basis of $\mathcal{V}$ contains the same number of vectors.*

This theorem, which is proven in Section A of the appendix, allows us to define the **dimension** of a vector space.

Definition 1.8 (Dimension)**.** *The dimension $\dim(\mathcal{V})$ of a vector space $\mathcal{V}$ is the cardinality of any of its bases, or equivalently the smallest number of linearly independent vectors that span $\mathcal{V}$.*

This definition coincides with the usual geometric notion of dimension in $\mathbb{R}^2$ and $\mathbb{R}^3$: a line has dimension 1, whereas a plane has dimension 2 (as long as they contain the origin). Note that there exist infinite-dimensional vector spaces, such as the continuous real-valued functions defined on $[0,1]$ or an iid sequence $X_1, X_2, \ldots$.

The vector space that we use to model a certain problem is usually called the **ambient space** and its dimension the **ambient dimension**. In the case of $\mathbb{R}^n$ the ambient dimension is n.

Lemma 1.9 (Dimension of $\mathbb{R}^n$)**.** *The dimension of $\mathbb{R}^n$ is n.*

Proof. Consider the set of vectors $e_1, \ldots, e_n \subseteq \mathbb{R}^n$ defined by

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ \cdots \\ 0 \end{bmatrix}, \quad \ldots, \quad e_n = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 1 \end{bmatrix}. \tag{11}$$

One can easily check that this set is a basis. It is in fact the **standard basis** of $\mathbb{R}^n$. □

2 Inner product product and norm

Up to now, the only operations that we know how to apply to the vectors in a vector space are addition and multiplication by a scalar. We now introduce a third operation, the **inner product** between two vectors.

Definition 2.1 (Inner product)**.** *An inner product on a vector space $\mathcal{V}$ is an operation $\langle \cdot, \cdot \rangle$ that maps pairs of vectors to $\mathbb{R}$ and satisfies the following conditions.*

- *It is symmetric, for any $x, y \in \mathcal{V}$*

$$\langle x, y \rangle = \langle y, x \rangle . \tag{12}$$

- *It is linear, i.e. for any $\alpha \in \mathbb{R}$ and any $x, y, z \in \mathcal{V}$*

$$\langle \alpha\, x, y \rangle = \alpha \langle y, x \rangle, \quad \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle . \tag{13}$$

- *It is positive semidefinite: for any $x \in \mathcal{V}$ $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ implies $x = 0$.*

A vector space endowed with an inner product is called an **inner-product space**.

Example 2.2 (Dot product)**.** We define the **dot product** between two vectors $x, y \in \mathbb{R}^n$ as

$$x \cdot y := \sum_i x[i]\ y[i], \tag{14}$$

where $x[i]$ is the ith entry of x. It is easy to check that the dot product is a valid inner product. $\mathbb{R}^n$ endowed with the dot product is usually called an Euclidean space of dimension n.

Example 2.3 (Covariance as an inner product)**.** The covariance of two zero-mean random variables X and Y is equal to $\mathrm{E}(XY)$. It is a valid inner product in the vector space of zero-mean random variables. It is obviously symmetric and linearity follows from linearity of expectation. Finally, $\mathrm{E}(X^2) \geq 0$ because it is the sum or integral of a nonnegative quantity and by Chebyshev's inequality $\mathrm{E}(X^2) = 0$ implies that $X = 0$ with probability one.

The **norm** of a vector is a generalization of the concept of *length*.

Definition 2.4 (Norm)**.** *Let $\mathcal{V}$ be a vector space, a norm is a function $||\cdot||$ from $\mathcal{V}$ to $\mathbb{R}$ that satisfies the following conditions.*

- *It is homogeneous. For all $\alpha \in \mathbb{R}$ and $x \in \mathcal{V}$*

$$||\alpha\, x|| = |\alpha|\, ||x|| . \tag{15}$$

- *It satisfies the **triangle inequality***

$$||x + y|| \leq ||x|| + ||y|| . \tag{16}$$

In particular, it is nonnegative (set $y = -x$).

- $||x|| = 0$ *implies that* x *is the zero vector* 0.

A vector space equipped with a norm is called a normed space. Distances in a normed space can be measured using the norm of the difference between vectors.

Definition 2.5 (Distance)**.** *The distance between two vectors in a normed space with norm* $||\cdot||$ *is*

$$d(x, y) := ||x - y|| . \tag{17}$$

Inner-product spaces are normed spaces because we can define a valid norm using the inner product. The norm **induced** by an inner product is obtained by taking the square root of the inner product of the vector with itself,

$$||x||_{\langle\cdot,\cdot\rangle} := \sqrt{\langle x, x\rangle}. \tag{18}$$

The norm induced by an inner product is clearly homogeneous by linearity and symmetry of the inner product. $||x||_{\langle\cdot,\cdot\rangle} = 0$ implies $x = 0$ because the inner product is positive semidefinite. We only need to establish that the triangle inequality holds to ensure that the inner-product is a valid norm.

Theorem 2.6 (Cauchy-Schwarz inequality)**.** *For any two vectors* x *and* y *in an inner-product space*

$$|\langle x, y\rangle| \leq ||x||_{\langle\cdot,\cdot\rangle} \, ||y||_{\langle\cdot,\cdot\rangle} . \tag{19}$$

Assume $||x||_{\langle\cdot,\cdot\rangle} \neq 0$,

$$\langle x, y\rangle = - ||x||_{\langle\cdot,\cdot\rangle} \, ||y||_{\langle\cdot,\cdot\rangle} \iff y = - \frac{||y||_{\langle\cdot,\cdot\rangle}}{||x||_{\langle\cdot,\cdot\rangle}} x, \tag{20}$$

$$\langle x, y\rangle = ||x||_{\langle\cdot,\cdot\rangle} \, ||y||_{\langle\cdot,\cdot\rangle} \iff y = \frac{||y||_{\langle\cdot,\cdot\rangle}}{||x||_{\langle\cdot,\cdot\rangle}} x. \tag{21}$$

Proof. If $||x||_{\langle\cdot,\cdot\rangle} = 0$ then $x = 0$ because the inner product is positive semidefinite, which implies $\langle x, y\rangle = 0$ and consequently that (19) holds with equality. The same is true if $||y||_{\langle\cdot,\cdot\rangle} = 0$.

Now assume that $||x||_{\langle\cdot,\cdot\rangle} \neq 0$ and $||y||_{\langle\cdot,\cdot\rangle} \neq 0$. By semidefiniteness of the inner product,

$$0 \leq \left|\left| \, ||y||_{\langle\cdot,\cdot\rangle} \, x + ||x||_{\langle\cdot,\cdot\rangle} \, y \right|\right|^2 = 2 \, ||x||^2_{\langle\cdot,\cdot\rangle} \, ||y||^2_{\langle\cdot,\cdot\rangle} + 2 \, ||x||_{\langle\cdot,\cdot\rangle} \, ||y||_{\langle\cdot,\cdot\rangle} \, \langle x, y\rangle , \tag{22}$$

$$0 \leq \left|\left| \, ||y||_{\langle\cdot,\cdot\rangle} \, x - ||x||_{\langle\cdot,\cdot\rangle} \, y \right|\right|^2 = 2 \, ||x||^2_{\langle\cdot,\cdot\rangle} \, ||y||^2_{\langle\cdot,\cdot\rangle} - 2 \, ||x||_{\langle\cdot,\cdot\rangle} \, ||y||_{\langle\cdot,\cdot\rangle} \, \langle x, y\rangle . \tag{23}$$

These inequalities establish (19).

Let us prove (20) by proving both implications.

($\Rightarrow$) Assume $\langle x, y\rangle = -\left|\left|x\right|\right|_{\langle\cdot,\cdot\rangle} \left|\left|y\right|\right|_{\langle\cdot,\cdot\rangle}$. Then (22) equals zero, so $\left|\left|y\right|\right|_{\langle\cdot,\cdot\rangle} x = -\left|\left|x\right|\right|_{\langle\cdot,\cdot\rangle} y$ because the inner product is positive semidefinite.

($\Leftarrow$) Assume $\left|\left|y\right|\right|_{\langle\cdot,\cdot\rangle} x = -\left|\left|x\right|\right|_{\langle\cdot,\cdot\rangle} y$. Then one can easily check that (22) equals zero, which implies $\langle x, y\rangle = -\left|\left|x\right|\right|_{\langle\cdot,\cdot\rangle} \left|\left|y\right|\right|_{\langle\cdot,\cdot\rangle}$.

The proof of (21) is identical (using (23) instead of (22)). □

Corollary 2.7. *The norm induced by an inner product satisfies the triangle inequality.*

Proof.

$$\begin{aligned}
\left|\left|x+y\right|\right|_{\langle\cdot,\cdot\rangle}^2 &= \left|\left|x\right|\right|_{\langle\cdot,\cdot\rangle}^2 + \left|\left|y\right|\right|_{\langle\cdot,\cdot\rangle}^2 + 2\langle x, y\rangle && (24)\\
&\leq \left|\left|x\right|\right|_{\langle\cdot,\cdot\rangle}^2 + \left|\left|y\right|\right|_{\langle\cdot,\cdot\rangle}^2 + 2\left|\left|x\right|\right|_{\langle\cdot,\cdot\rangle}\left|\left|y\right|\right|_{\langle\cdot,\cdot\rangle} \quad \text{by the Cauchy-Schwarz inequality} && (25)\\
&= \left(\left|\left|x\right|\right|_{\langle\cdot,\cdot\rangle} + \left|\left|y\right|\right|_{\langle\cdot,\cdot\rangle}\right)^2. && (26)
\end{aligned}$$

□

Example 2.8 (Euclidean norm)**.** The Euclidean or ℓ_2 norm is the norm induced by the dot product in $\mathbb{R}^n$,

$$\left|\left|x\right|\right|_2 := \sqrt{x \cdot x} = \sqrt{\sum_{i=1}^{n} x_i^2}. \tag{27}$$

In the case of $\mathbb{R}^2$ or $\mathbb{R}^3$ it is what we usually think of as the length of the vector.

Example 2.9 (The standard deviation as a norm)**.** The standard deviation or root mean square

$$\sigma_X = \sqrt{\mathrm{E}\left(X^2\right)} \tag{28}$$

is the norm induced by the covariance inner product in the vector space of zero-mean random variables.

---

3 Orthogonality

An important concept in linear algebra is orthogonality.

Definition 3.1 (Orthogonality)**.** *Two vectors x and y are orthogonal if*

$$\langle x, y \rangle = 0. \tag{29}$$

A vector x is orthogonal to a set $\mathcal{S}$, if

$$\langle x, s \rangle = 0, \quad \textit{for all } s \in \mathcal{S}. \tag{30}$$

Two sets of $\mathcal{S}_1, \mathcal{S}_2$ are orthogonal if for any $x \in \mathcal{S}_1, y \in \mathcal{S}_2$

$$\langle x, y \rangle = 0. \tag{31}$$

The orthogonal complement of a subspace $\mathcal{S}$ is

$$\mathcal{S}^\perp := \{x \mid \langle x, y \rangle = 0 \quad \textit{for all } y \in \mathcal{S}\}. \tag{32}$$

Distances between orthogonal vectors measured in terms of the norm induced by the inner product are easy to compute.

Theorem 3.2 (Pythagorean theorem)**.** *If x and y are orthogonal vectors*

$$||x + y||^2_{\langle \cdot, \cdot \rangle} = ||x||^2_{\langle \cdot, \cdot \rangle} + ||y||^2_{\langle \cdot, \cdot \rangle}. \tag{33}$$

Proof. By linearity of the inner product

$$\begin{aligned} ||x + y||^2_{\langle \cdot, \cdot \rangle} &= ||x||^2_{\langle \cdot, \cdot \rangle} + ||y||^2_{\langle \cdot, \cdot \rangle} + 2 \langle x, y \rangle && (34) \\ &= ||x||^2_{\langle \cdot, \cdot \rangle} + ||y||^2_{\langle \cdot, \cdot \rangle}. && (35) \end{aligned}$$

□

If we want to show that a vector is orthogonal to a certain subspace, it is enough to show that it is orthogonal to every vector in a basis of the subspace.

Lemma 3.3. *Let x be a vector and $\mathcal{S}$ a subspace of dimension n. If for any basis $b_1, b_2, \ldots, b_n$ of $\mathcal{S}$,*

$$\langle x, b_i \rangle = 0, \quad 1 \leq i \leq n, \tag{36}$$

then x is orthogonal to $\mathcal{S}$.

Proof. Any vector $v \in \mathcal{S}$ can be represented as $v = \sum_i \alpha_{i=1}^n b_i$ for $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, from (36)

$$\langle x, v \rangle = \left\langle x, \sum_i \alpha_{i=1}^n b_i \right\rangle = \sum_i \alpha_{i=1}^n \langle x, b_i \rangle = 0. \tag{37}$$

□

We now introduce orthonormal bases.

Definition 3.4 (Orthonormal basis)**.** *A basis of mutually orthogonal vectors with norm equal to one is called an* ***orthonormal*** *basis.*

It is very easy to find the coefficients of a vector in an orthonormal basis: we just need to compute the dot products with the basis vectors.

Lemma 3.5 (Coefficients in an orthonormal basis)**.** *If $\{u_1, \ldots, u_n\}$ is an orthonormal basis of a vector space $\mathcal{V}$, for any vector $x \in \mathcal{V}$*

$$x = \sum_{i=1}^n \langle u_i, x \rangle u_i. \tag{38}$$

Proof. Since $\{u_1, \ldots, u_n\}$ is a basis,

$$x = \sum_{i=1}^m \alpha_i u_i \quad \text{for some } \alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}. \tag{39}$$

Immediately,

$$\langle u_i, x \rangle = \left\langle u_i, \sum_{i=1}^m \alpha_i u_i \right\rangle = \sum_{i=1}^m \alpha_i \langle u_i, u_i \rangle = \alpha_i \tag{40}$$

because $\langle u_i, u_i \rangle = 1$ and $\langle u_i, u_j \rangle = 0$ for $i \neq j$. □

For *any* subspace of $\mathbb{R}^n$ we can obtain an orthonormal basis by applying the Gram-Schmidt method to a set of linearly independent vectors spanning the subspace.

Algorithm 3.6 (Gram-Schmidt)**.**

Input*: A set of linearly independent vectors $\{x_1, \ldots, x_m\} \subseteq \mathbb{R}^n$.*
Output*: An orthonormal basis $\{u_1, \ldots, u_m\}$ for span $(x_1, \ldots, x_m)$.*

Initialization*: Set $u_1 := x_1 / ||x_1||_2$.*
For $i = 1, \ldots, m$

1. Compute

$$v_i := x_i - \sum_{j=1}^{i-1} \langle u_j, x_i \rangle \, u_j. \tag{41}$$

2. Set $u_i := v_i / \left|\left|v_i\right|\right|_2$.

This implies in particular that we can always assume that a subspace has an orthonormal basis.

Theorem 3.7. *Every finite-dimensional vector space has an orthonormal basis.*

Proof. To see that the Gram-Schmidt method produces an orthonormal basis for the span of the input vectors we can check that $\operatorname{span}(x_1, \ldots, x_i) = \operatorname{span}(u_1, \ldots, u_i)$ and that $u_1, \ldots, u_i$ is set of orthonormal vectors. $\square$

A Proof of Theorem 1.7

We prove the claim by contradiction. Assume that we have two bases $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_n\}$ such that $m < n$ (or the second set has infinite cardinality). The proof follows from applying the following lemma m times (setting $r = 0, 1, \ldots, m-1$) to show that $\{y_1, \ldots, y_m\}$ spans $\mathcal{V}$ and hence $\{y_1, \ldots, y_n\}$ must be linearly dependent.

Lemma A.1. *Under the assumptions of the theorem, if $\{y_1, y_2, \ldots, y_r, x_{r+1}, \ldots, x_m\}$ spans $\mathcal{V}$ then $\{y_1, \ldots, y_{r+1}, x_{r+2}, \ldots, x_m\}$ also spans $\mathcal{V}$ (possibly after rearranging the indices $r+1, \ldots, m$) for $r = 0, 1, \ldots, m-1$.*

Proof. Since $\{y_1, y_2, \ldots, y_r, x_{r+1}, \ldots, x_m\}$ spans $\mathcal{V}$

$$y_{r+1} = \sum_{i=1}^{r} \beta_i \, y_i + \sum_{i=r+1}^{m} \gamma_i \, x_i, \quad \beta_1, \ldots, \beta_r, \gamma_{r+1}, \ldots, \gamma_m \in \mathbb{R}, \tag{42}$$

where at least one of the γ_j is non zero, as $\{y_1, \ldots, y_n\}$ is linearly independent by assumption. Without loss of generality (here is where we might need to rearrange the indices) we assume that $\gamma_{r+1} \neq 0$, so that

$$x_{r+1} = \frac{1}{\gamma_{r+1}} \left(\sum_{i=1}^{r} \beta_i \, y_i - \sum_{i=r+2}^{m} \gamma_i x_i \right). \tag{43}$$

This implies that any vector in the span of $\{y_1, y_2, \ldots, y_r, x_{r+1}, \ldots, x_m\}$, i.e. in $\mathcal{V}$, can be represented as a linear combination of vectors in $\{y_1, \ldots, y_{r+1}, x_{r+2}, \ldots, x_m\}$, which completes the proof. $\square$