Chapitre 12

Produit scalaire et orthogonalité

12.1 Norme et distance

Définition 12.1. Si $\mathbf{x} \in \mathbb{R}^n$, et si $x_1, \ldots, x_n$ sont ses composantes relativement à la base canonique, alors sa **norme** est définie par le réel

$$\|\mathbf{x}\| := \sqrt{x_1^2 + \cdots + x_n^2}\,.$$

Proposition 13. *(Propriétés de la norme)*

- ⋆ $\|\lambda \mathbf{x}\| = |\lambda|\|\mathbf{x}\|$ *pour tous* $\lambda \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n$
- ⋆ $\|\mathbf{x}\| \geqslant 0$ *pour tout* $\mathbf{x}$, *et* $\|\mathbf{x}\| = 0$ *si et seulement si* $\mathbf{x} = 0$.
- ⋆ *(Inégalité triangulaire)* $\|\mathbf{x}+\mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$ *pour tous* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Preuve: Pour la première propriété,

$$\begin{aligned}
\|\lambda \mathbf{x}\| &= \sqrt{(\lambda x_1)^2 + \cdots + (\lambda x_n)^2}\\
&= \sqrt{\lambda^2 x_1^2 + \cdots + \lambda^2 x_n^2}\\
&= \sqrt{\lambda^2 (x_1^2 + \cdots + x_n^2)}\\
&= |\lambda|\sqrt{x_1^2 + \cdots + x_n^2}\\
&= |\lambda|\|\mathbf{x}\|\,.
\end{aligned}$$

Ensuite, $\|\mathbf{x}\| \geqslant 0$ est évidente, et remarquons que $\|\mathbf{x}\| = 0$ si et seulement si $\|\mathbf{x}\|^2 = 0$, qui est équivalente à

$$x_1^2 + \cdots + x_n^2 = 0\,.$$

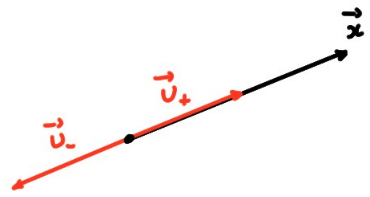
Or une somme de nombres non-négatifs est nulle si et seulement chacun de ces nombres est nul, $x_k^2 = 0$, et donc $x_k = 0$ pour chaque $k = 1, \ldots, n$.

On démontrera l'inégalité triangulaire dans la section suivante. □

Définition 12.2. $\mathbf{x} \in \mathbb{R}^n$ est **unitaire** (ou **normalisé**) si $\|\mathbf{x}\| = 1$.

Remarque 12.3. Pour tout vecteur non-nul $\mathbf{x}$, il existe exactement deux vecteurs unitaires qui sont colinéaires à $\mathbf{x}$, donnés par

$$\mathbf{u}_\pm := \pm\frac{\mathbf{x}}{\|\mathbf{x}\|}\,.$$

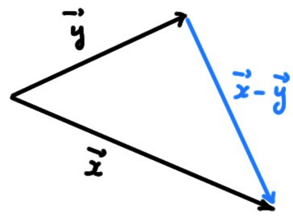


◇

La notion de norme permet de définir encore deux notions géométriques classiques :

Définition 12.4. La **distance** entre $\mathbf{x}$ et $\mathbf{y}$ est définie par

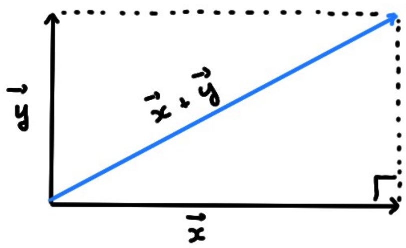
$$\operatorname{dist}(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| .$$



Définition 12.5. Deux vecteurs $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ sont **orthogonaux** (ou **perpendiculaires**) si

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 .$$

Si $\mathbf{x}$ et $\mathbf{y}$ sont orthogonaux, on écrit $\mathbf{x} \perp \mathbf{y}$.



12.2 Définition du produit scalaire

Définition 12.6. Soient $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Le **produit scalaire** de $\mathbf{x}$ et $\mathbf{y}$ est défini par

$$\mathbf{x} \cdot \mathbf{y} := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n .$$

Remarque 12.7. Il sera souvent utile de récrire le produit scalaire en le réinterprétant comme un produit matriciel un peu particulier :

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \cdots + x_n y_n = \underbrace{[x_1 \cdots x_n]}_{1\times n} \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{n\times 1} = \mathbf{x}^T \mathbf{y} .$$

◇

Proposition 14. *(Propriétés du produit scalaire)*

a) $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$

b) $(\lambda \mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (\lambda \mathbf{y}) = \lambda(\mathbf{x} \cdot \mathbf{y})$

c) $\mathbf{x} \cdot (\mathbf{y}_1 + \mathbf{y}_2) = \mathbf{x} \cdot \mathbf{y}_1 + \mathbf{x} \cdot \mathbf{y}_2$

d) $(\mathbf{x}_1 + \mathbf{x}_2) \cdot \mathbf{y} = \mathbf{x}_1 \cdot \mathbf{y} + \mathbf{x}_2 \cdot \mathbf{y}$

e) *(Lien avec la norme)* $\mathbf{x} \cdot \mathbf{x} = \|\mathbf{x}\|^2$

f) *(Inégalité de Cauchy-Schwartz)*

$$|\mathbf{x} \cdot \mathbf{y}| \leqslant \|\mathbf{x}\| \|\mathbf{y}\| .$$

Preuve: Les cinq premières propriétés suivent directement de la définition du produit scalaire. Démontrons l'inégalité de Cauchy-Schwartz.

Pour commencer, remarquons que l'inégalité est triviale dès que $\mathbf{y}$ (ou $\mathbf{x}$) est nul. On peut donc supposer que $\mathbf{y} \neq \mathbf{0}$. Ensuite, remarquons que pour tout $t \in \mathbb{R}$,

$$\begin{aligned} 0 \leqslant \|\mathbf{x} - t\mathbf{y}\|^2 &= (\mathbf{x} - t\mathbf{y}) \cdot (\mathbf{x} - t\mathbf{y}) \\ &= \mathbf{x} \cdot \mathbf{x} - 2t(\mathbf{x} \cdot \mathbf{y}) + t^2(\mathbf{y} \cdot \mathbf{y}) \\ &= \|\mathbf{x}\|^2 - 2t(\mathbf{x} \cdot \mathbf{y}) + t^2\|\mathbf{y}\|^2 \end{aligned}$$

Comme cette inégalité est vraie pour tout t, on peut l'utiliser pour la valeur t_* qui minimise le polynôme du deuxième degré défini par

$$t \mapsto t^2\|\mathbf{y}\|^2 - 2t(\mathbf{x} \cdot \mathbf{y}) + \|\mathbf{x}\|^2 .$$

Ce dernier est minimal lorsque

$$t_* = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{y}\|^2} .$$

Si on récrit l'inégalité du haut avec t_*, on obtient

$$0 \leqslant \|\mathbf{x}\|^2 - 2\frac{(\mathbf{x} \cdot \mathbf{y})^2}{\|\mathbf{y}\|^2} + \frac{(\mathbf{x} \cdot \mathbf{y})^2}{\|\mathbf{y}\|^2} = \|\mathbf{x}\|^2 - \frac{(\mathbf{x} \cdot \mathbf{y})^2}{\|\mathbf{y}\|^2} ,$$

qui entraîne

$$(\mathbf{x} \cdot \mathbf{y})^2 \leqslant \|\mathbf{x}\|^2\|\mathbf{y}\|^2 .$$

On obtient l'inégalité de Cauchy-Schwartz en prenant la racine carrée des deux côtés. □

Remarque 12.8. $\mathbb{R}^n$, muni du produit scalaire, est un cas particulier de ce que nous appellerons plus tard un **espace Euclidien**. ◇

Informel 12.9. En petites dimensions ($n = 2$ ou 3), le produit scalaire est relié à l'angle θ fait par $\mathbf{x}$ et $\mathbf{y}$, par la relation fondamentale suivante :

$$\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos(\theta) .$$

Nous ne ferons pas usage de cette relation, puisque nous utiliserons le produit scalaire surtout en grandes dimensions, et qu'on ne sait pas vraiment comment définir l'angle entre deux vecteurs.

On peut utiliser l'inégalité de Cauchy-Schwartz pour démontrer l'inégalité triangulaire de la section précédente :

$$\begin{aligned}\|\mathbf{x}+\mathbf{y}\|^2 &= (\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}+\mathbf{y}) \\ &= \|\mathbf{x}\|^2 + 2(\mathbf{x}\cdot\mathbf{y}) + \|\mathbf{y}\|^2 \\ &\leqslant \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2\,.\end{aligned}$$

Orthogonalité

Le produit scalaire est surtout utilisé, en algèbre linéaire, *pour résoudre des problèmes dans $\mathbb{R}^n$ à l'aide d'arguments géométriques* empruntés à la géométrie du plan et de l'espace. Et la première notion qui joue un rôle en géométrie est celle *d'orthogonalité*.

Commençons par écrire l'égalité du dessus sous une forme un peu différente :

$$(\mathbf{x}\cdot\mathbf{y}) = \frac{1}{2}\left(\|\mathbf{x}+\mathbf{y}\|^2 - \left(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2\right)\right)\,.$$

Ceci implique en particulier que $\mathbf{x}\perp\mathbf{y}$ si et seulement si $\mathbf{x}\cdot\mathbf{y} = 0$. On peut donc donner une définition alternative de l'orthogonalité, basée uniquement sur le produit scalaire :

Définition 12.10. Deux vecteurs $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$ sont **orthogonaux** (ou **perpendiculaires**) si $\mathbf{x}\cdot\mathbf{y} = 0$.

Exemple 12.11. Dans $\mathbb{R}^5$, les vecteurs

$$\mathbf{x} = \begin{pmatrix}3\\1\\-2\\0\\2\end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix}2\\2\\5\\3\\1\end{pmatrix}$$

sont orthogonaux, puisque $\mathbf{x}\cdot\mathbf{y} = 0$. ◇

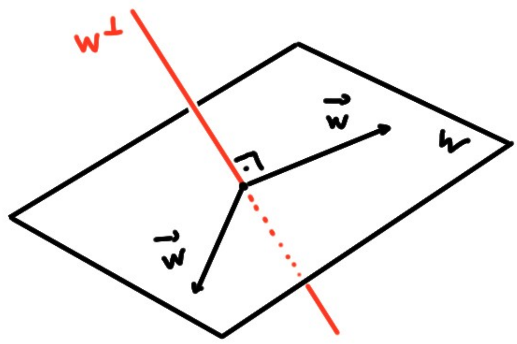
En géométrie, on considère souvent un objet géométrique, généralement une droite ou un plan, défini comme étant perpendiculaire à un autre. En algèbre linéaire, on définit un ensemble de vecteurs qui sont tous orthogonaux aux vecteurs d'un autre ensemble :

Définition 12.12. Soit W un sous-espace vectoriel de $\mathbb{R}^n$. Le **complément orthogonal de** W est l'ensemble

$$W^\perp := \left\{\mathbf{v}\in\mathbb{R}^n \,\middle|\, \mathbf{v}\perp\mathbf{w}\ \forall\mathbf{w}\in W\right\}.$$

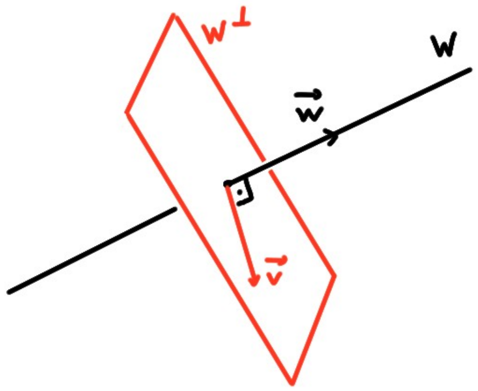
Commençons par comprendre intuitivement le sens de $W^\perp$, en petites dimensions :

Exemple 12.13. Si W est un plan (passant par l'origine) de $\mathbb{R}^3$, alors $W^\perp$ est la droite perpendiculaire à W, passant par l'origine :



(En effet, un vecteur $\mathbf{v}$ quelconque sur la droite est perpendiculaire à tous les vecteurs $\mathbf{w}$ du plan.) ◇

Exemple 12.14. Si W est une droite (passant par l'origine) de $\mathbb{R}^3$, alors $W^\perp$ est le plan perpendiculaire à W, passant par l'origine :



(En effet, un vecteur $\mathbf{v}$ quelconque sur le plan est perpendiculaire à tous les vecteurs $\mathbf{w}$ de la droite.) ◇

Ces deux derniers exemples illustrent bien les propriétés générales ci-dessous :

Proposition 15. *(Propriétés du complément orthogonal)*

a) $W^\perp$ *est un sous-espace vectoriel de* $\mathbb{R}^n$

b) $(W^\perp)^\perp = W$

c) $\dim(W) + \dim(W^\perp) = n$

Preuve: On vérifie que $W^\perp$ est un sous-espace vectoriel de $\mathbb{R}^n$:

- ⋆ Clairement, le vecteur nul appartient à W puisque $\mathbf{0} \cdot \mathbf{w} = 0$ pour tout $\mathbf{w} \in W$.
- ⋆ Si $\mathbf{v} \in W^\perp$, alors pour tout scalaire $\lambda \in \mathbb{R}$,

$$(\lambda v) \cdot \mathbf{w} = \lambda(\mathbf{v} \cdot \mathbf{w}) = 0\,,$$

donc $\lambda\mathbf{v} \in W^\perp$.

- ⋆ Si $\mathbf{v}_1, \mathbf{v}_2 \in W^\perp$, alors pour tout $\mathbf{w} \in W$,

$$(\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{w} = \mathbf{v}_1 \cdot \mathbf{w} + \mathbf{v}_2 \cdot \mathbf{w} = 0 + 0 = 0\,,$$

donc $\mathbf{v}_1 + \mathbf{v}_2 \in W^\perp$.

Les autes propriétés seront démontrées en exercice. □

Dans la définition, $W^\perp$ est défini comme l'ensemble des vecteurs qui sont orthogonaux à *tous* les vecteurs de W. Ceci implique que d'un point de vue calculatoire, on devrait a priori vérifier une infinité de conditions pour savoir si un vecteur appartient à $W^\perp$. Mais lorsqu'on possède une base les choses sont plus simples :

Lemme 16. *Soit W un sous-espace vectoriel de $\mathbb{R}^n$, et soit $\mathcal{B} = (\mathbf{w}_1, \ldots, \mathbf{w}_k)$ une base de W. Alors $\mathbf{v} \in W^\perp$ si et seulement si $\mathbf{v} \perp \mathbf{w}_j$ pour tout $j = 1, \ldots, k$.*

Preuve: Supposons que $W = \mathrm{Vect}\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$.

Si $\mathbf{v}$ est orthogonal à tous les vecteurs de W, il est en particulier orthogonal à chacun des éléments de la base.

Inversément, supposons que $\mathbf{v}$ est orthogonal à chacun des éléments de la base. Comme un élément quelconque $\mathbf{w} \in W$ peut se décomposer dans la base, $\mathbf{w} = \alpha_1 \mathbf{w}_1 + \cdots + \alpha_k \mathbf{w}_k$, la linéarité du produit scalaire implique que

$$\begin{aligned}\mathbf{v} \cdot \mathbf{w} &= \mathbf{v} \cdot (\alpha_1 \mathbf{w}_1 + \cdots + \alpha_k \mathbf{w}_k) \\ &= \alpha_1 \underbrace{(\mathbf{v} \cdot \mathbf{w}_1)}_{=0} + \cdots + \alpha_k \underbrace{(\mathbf{v} \cdot \mathbf{w}_k)}_{=0} \\ &= 0\,,\end{aligned}$$

et donc $\mathbf{v} \in W^\perp$. □

Exemple 12.15. Dans $\mathbb{R}^3$, considérons les vecteurs

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \qquad \mathbf{w}_2 = \begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix},$$

et considérons le plan

$$W = \mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_2\}\,.$$

Le lemme précédent dit que

$$W^\perp = \{\mathbf{v} \in \mathbb{R}^3 \,:\, \mathbf{v} \perp \mathbf{w}_1,\, \mathbf{v} \perp \mathbf{w}_2\}$$

Donc on cherche les vecteurs $\mathbf{v} \in \mathbb{R}^3$ tels que les deux conditions suivantes soient satisfaites simultanément :

$$\begin{cases} \mathbf{v} \cdot \mathbf{w}_1 = 0\,, \\ \mathbf{v} \cdot \mathbf{w}_2 = 0\,. \end{cases}$$

Si $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$, ceci est équivalent à

$$\left\{ \begin{array}{rcrcrcl} v_1 & + & 2v_2 & & & = & 0 \\ -v_1 & + & 3v_2 & + & v_3 & = & 0 \end{array} \right.$$

On peut prendre v_1 comme variable libre, et donc on voit que $W^\perp$ est une droite :

$$W^\perp = \left\{ \mathbf{v} = \begin{pmatrix} x_1 \\ -x_1/2 \\ 5x_1/2 \end{pmatrix} \,\middle|\, x_1 \in \mathbb{R} \right\} = \mathrm{Vect}\left\{ \begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix} \right\}.$$

On vérifie bien dans ce cas que

$$\dim(W) + \dim(W^\perp) = 2 + 1 = 3\,.$$

◇

Informel 12.16. Dans ce dernier exemple, l'intuition géométrique aurait peut-être suggéré de trouver un vecteur directeur de la droite $W^\perp$ en calculant le **produit vectoriel** de $\mathbf{w}_1$ et $\mathbf{w}_2$. Mais ce produit (que nous ne traiterons pas dans ce cours) n'existe que dans $\mathbb{R}^3$, alors que la méthode que nous avons utilisée fonctionne en toute dimension.

Exemple 12.17. Dans $\mathbb{R}^4$, considérons les vecteurs

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 2 \\ -1 \\ 0 \end{pmatrix}, \qquad \mathbf{w}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 2 \end{pmatrix},$$

et considérons le plan

$$W = \mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_2\}.$$

Puisque $\dim(W) = 2$ et que

$$\dim(W) + \dim(W^\perp) = 4,$$

on sait que $\dim(W^\perp) = 2$, et donc $W^\perp$ doit aussi être un plan. Et effectivement, un calcul semblable à celui de l'exemple précédent (voir exercices) montre que

$$W^\perp = \mathrm{Vect}\{\mathbf{p}_1, \mathbf{p}_2\}.$$

où

$$\mathbf{p}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \mathbf{p}_2 = \begin{pmatrix} 2 \\ -1 \\ 0 \\ 1 \end{pmatrix}.$$

◇

12.3 À propos de $\mathrm{Col}(A)$ et $\mathrm{Lgn}(A)$

Rappelons que si A est une matrice $m \times n$,

- ⋆ $\mathrm{Col}(A) \subset \mathbb{R}^m$ est le sous-espace engendré par ses colonnes, et
- ⋆ $\mathrm{Lgn}(A) \subset \mathbb{R}^n$ est le sous-espace engendré par ses lignes.

Théorème 12.18. *Si A est $m \times n$, alors*

a) $\mathrm{Lgn}(A)^\perp = \ker(A)$,

b) $\mathrm{Col}(A)^\perp = \ker(A^T)$.

Preuve: Nous avons déjà vu (**ici** (lien web)) que l'on peut toujours exprimer une matrice $m \times n$ à l'aide de ses lignes :

$$A = \begin{pmatrix} \ell_1^T \\ \vdots \\ \ell_m^T \end{pmatrix},$$

où $\ell_1, \dots, \ell_m \in \mathbb{R}^n$.

a) On a

$$\begin{aligned} \mathbf{v} \in \mathrm{Lgn}(A)^\perp &\iff \quad \mathbf{v} \cdot \ell_j = 0 \quad \forall j = 1, \dots, m \\ &\iff \quad \ell_j^T \mathbf{v} = 0 \quad \forall j = 1, \dots, m\,. \end{aligned}$$

On peut exprimer ces m conditions simultanément en écrivant

$$\begin{pmatrix} \ell_1^T \\ \vdots \\ \ell_m^T \end{pmatrix} \mathbf{v} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

qui n'est autre que $A\mathbf{v} = \mathbf{0}$.

b) Puisque $\mathrm{Col}(A) = \mathrm{Lgn}(A^T)$, l'affirmation suit de la première partie :

$$\mathrm{Col}(A)^\perp = (\mathrm{Lgn}(A^T))^\perp = \ker(A^T)\,.$$

□

12.4 Familles orthogonales

Définition 12.19. Une famille de vecteurs $\{\mathbf{w}_1, \dots, \mathbf{w}_k\} \subset \mathbb{R}^n$ est dite

- ⋆ **orthogonale** si ses vecteurs sont orthogonaux deux à deux ($\mathbf{w}_i \perp \mathbf{w}_j$ pour tout $i \neq j$),
- ⋆ **orthonormale** (ou **orthonormée**) si elle est orthogonale et si, de plus, tous ses vecteurs sont unitaires ($\|\mathbf{w}_j\| = 1$ pour tout j).

Exemple 12.20. La base canonique de $\mathbb{R}^n$, $\mathcal{B}_{\mathrm{can}} = (\mathbf{e}_1, \dots, \mathbf{e}_n)$, est une famille orthonormale, puisque

$$\mathbf{e}_i \cdot \mathbf{e}_j = \begin{cases} 1 & \text{si } i = j\,, \\ 0 & \text{si } i \neq j\,. \end{cases}$$

◇

Remarque 12.21. Si $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ est orthogonale, et si aucun de ses vecteurs n'est le vecteur nul, alors on la rend orthonormale en divisant chacun de ses vecteurs par sa norme :

$$\left\{ \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|}, \dots, \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|} \right\}$$

◇

Exemple 12.22. Dans $\mathbb{R}^3$, la famille

$$\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix} \right\}$$

est orthogonale, mais pas orthonormale. Comme aucun de ses vecteurs n'est nul, on peut le diviser par sa norme,

$$\left\{ \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \frac{1}{\sqrt{12}} \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix} \right\}$$

pour obtenir une famille orthonormale. ◇

Une propriété importante des familles orthogonales :

Lemme 17. *Si $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ est orthogonale, et si aucun de ses vecteurs n'est nul, alors elle est libre.*

Preuve: Considérons la relation
$$\alpha_1 \mathbf{w}_1 + \cdots + \alpha_k \mathbf{w}_k = \mathbf{0}\,.$$
Si on effectue le produit scalaire de cette relation avec $\mathbf{w}_j$,
$$\alpha_1 \underbrace{(\mathbf{w}_j \cdot \mathbf{w}_1)}_{=0} + \cdots + \alpha_j \underbrace{(\mathbf{w}_j \cdot \mathbf{w}_j)}_{\neq 0} + \cdots + \alpha_k \underbrace{(\mathbf{w}_j \cdot \mathbf{w}_k)}_{=0} = \mathbf{0}\,,$$
qui donne $\alpha_j \|\mathbf{w}_j\|^2 = 0$. Puisque par hypothèse $\mathbf{w}_j \neq \mathbf{0}$, ceci implique $\alpha_j = 0$. Comme ceci vaut pour tout $j = 1, \dots, k$, on a bien montré que la famille est libre. □

Le grand avantage de travailler avec des bases orthogonales :

Théorème 12.23. *Soit W un sous-espace vectoriel de $\mathbb{R}^n$, et soit $\mathcal{B} = (\mathbf{w}_1, \dots, \mathbf{w}_k)$ une base de W, orthogonale. Alors la décomposition d'un $\mathbf{w} \in W$ relativement à $\mathcal{B}$,*
$$\mathbf{w} = \gamma_1 \mathbf{w}_1 + \cdots + \gamma_k \mathbf{w}_k\,,$$
a ses coefficients γ_j donnés par
$$\gamma_j = \frac{\mathbf{w} \cdot \mathbf{w}_j}{\mathbf{w}_j \cdot \mathbf{w}_j} = \frac{\mathbf{w} \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2}\,.$$
En particulier, si $\mathcal{B}$ est orthonormale, alors $\gamma_j = \mathbf{w} \cdot \mathbf{w}_j$.

Preuve: Considérons la décomposition
$$\mathbf{w} = \gamma_1 \mathbf{w}_1 + \cdots + \gamma_k \mathbf{w}_k\,.$$
En prenant le produit scalaire de cette expression avec $\mathbf{w}_j$, l'orthogonalité de la base fait qu'il ne survit qu'un seul terme dans le membre de droite :
$$\mathbf{w}_j \cdot \mathbf{w} = \gamma_j (\mathbf{w}_j \cdot \mathbf{w}_j)\,.$$
Ceci démontre l'affirmation. □

Informel 12.24. Rappelons qu'en principe, trouver les composantes d'un vecteur relativement à une base se fait en résolvant un système. Ici, on voit le grand avantage de travailler avec des bases orthogonales : pour avoir une composante, il suffit de calculer un produit scalaire.

Exemple 12.25. Considérons
$$\mathcal{B} = \left(\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix} \right)\,.$$
On a vu plus haut que cette famille est orthogonale, et donc libre puisqu'aucun de ses vecteurs n'est nul, ce qui en fait une base de $\mathbb{R}^3$. Si on prend un vecteur quelconque de $\mathbb{R}^n$, par exemple
$$\mathbf{v} = \begin{pmatrix} 7 \\ -5 \\ 3 \end{pmatrix}\,,$$

on calcule ses composantes relativement à $\mathcal{B}$:

$$\begin{aligned}\gamma_1 &= \frac{\mathbf{v}\cdot\mathbf{b}_1}{\|\mathbf{b}_1\|^2} = \frac{0}{6} = 0\,,\\ \gamma_2 &= \frac{\mathbf{v}\cdot\mathbf{b}_2}{\|\mathbf{b}_2\|^2} = \frac{4}{2} = 2\,,\\ \gamma_3 &= \frac{\mathbf{v}\cdot\mathbf{b}_3}{\|\mathbf{b}_3\|^2} = \frac{-30}{12} = \frac{-5}{2}\,,\end{aligned}$$

ce qui donne

$$[\mathbf{v}]_{\mathcal{B}} = \begin{pmatrix}0\\2\\-5/2\end{pmatrix}\,,$$

c'est-à-dire

$$\mathbf{v} = 0\begin{pmatrix}1\\2\\1\end{pmatrix} + 2\begin{pmatrix}1\\0\\-1\end{pmatrix} - \frac{5}{2}\begin{pmatrix}-2\\2\\-2\end{pmatrix}\,.$$

Bien-sûr, on trouverait la même chose en cherchant les composantes comme on le faisait avant, en étudiant le système

$$\gamma_1\begin{pmatrix}1\\2\\1\end{pmatrix} + \gamma_2\begin{pmatrix}1\\0\\-1\end{pmatrix} + \gamma_3\begin{pmatrix}-2\\2\\-2\end{pmatrix} = \begin{pmatrix}7\\-5\\3\end{pmatrix}\,.$$

◇

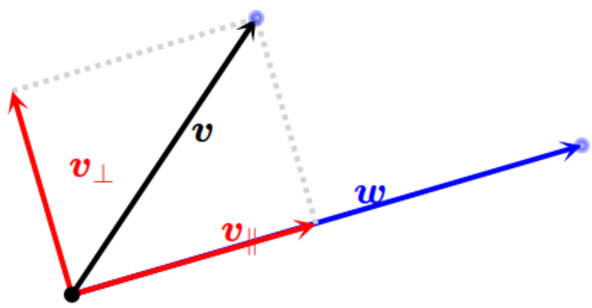
12.5 Projection sur un vecteur

La notion d'orthogonalité permet d'introduire en algèbre linéaire plusieurs notions géométriques très utiles. La première est celle de **projection**.

Comme motivation, fixons un vecteur $\mathbf{w} \in \mathbb{R}^n$, et posons la question suivante : Pour un deuxième $\mathbf{v} \in \mathbb{R}^n$ donné, comment définir la *projection orthogonale de* $\mathbf{v}$ *sur* $\mathbf{w}$?

Informel 12.26. On a déjà considéré, dans le plan, la projection d'un vecteur sur une droite. Mais ici, on est en dimension quelconque n ! Et nous allons commencer par projeter sur un vecteur, mais plus loin nous projetterons sur un sous-espace vectoriel quelconque de $\mathbb{R}^n$.

Un schéma peut aider à comprendre comment nous allons procéder (attention : cette image est représentée dans le plan, mais l'argument qui suit fonctionne en toute dimension !) :



La projection orthogonale de $\mathbf{v}$ sur $\mathbf{w}$, que nous noterons $\mathbf{v}_{\parallel}$ pour commencer, doit permettre de décomposer $\mathbf{v}$ en deux composantes vectorielles,

$$\mathbf{v} = \mathbf{v}_{\parallel} + \mathbf{v}_{\perp}\,,$$

où

a) $\mathbf{v}_{\parallel}$ est colinéaire (parallèle) à $\mathbf{w}$,
b) $\mathbf{v}_{\perp}$ est orthogonal à $\mathbf{w}$.

Il se trouve que ces deux conditions caractérisent entièrement $\mathbf{v}_{\parallel}$ et $\mathbf{v}_{\perp}$.
En effet, pour que $\mathbf{v}_{\parallel}$ soit colinéaire à $\mathbf{w}$, il doit exister un scalaire α tel que

$$\mathbf{v}_{\parallel} = \alpha\mathbf{w}\,.$$

Puis, pour que $\mathbf{v}_{\perp}$ soit orthogonal à $\mathbf{w}$, il faut que

$$0 = \mathbf{v}_{\perp}\cdot\mathbf{w} = (\mathbf{v}-\mathbf{v}_{\parallel})\cdot\mathbf{w} = (\mathbf{v}-\alpha\mathbf{w})\cdot\mathbf{w}\,.$$

De cette dernière relation, on tire que

$$\alpha = \frac{\mathbf{v}\cdot\mathbf{w}}{\mathbf{w}\cdot\mathbf{w}} = \frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{w}\|^2}\,.$$

En utilisant ce scalaire particulier dans $\mathbf{v}_{\parallel} = \alpha\mathbf{w}$, ceci motive la définition suivante :

Définition 12.27. Soit $\mathbf{w}\in\mathbb{R}^n$, $\mathbf{w}\neq\mathbf{0}$. La **projection orthogonale de** $\mathbf{v}\in\mathbb{R}^n$ **sur** $\mathbf{w}$ est définie par

$$\mathrm{proj}_{\mathbf{w}}(\mathbf{v}) := \frac{\mathbf{v}\cdot\mathbf{w}}{\mathbf{w}\cdot\mathbf{w}}\mathbf{w} = \frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{w}\|^2}\mathbf{w}\,.$$

Exemple 12.28. Dans $\mathbb{R}^5$, la projection orthogonale de

$$\mathbf{v} = \begin{pmatrix}0\\2\\-3\\1\\-1\end{pmatrix} \qquad \text{sur } \mathbf{w} = \begin{pmatrix}1\\0\\-1\\0\\1\end{pmatrix}$$

est donnée par

$$\mathrm{proj}_{\mathbf{w}}(\mathbf{v}) = \frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{w}\|^2}\mathbf{w} = \frac{2}{3}\begin{pmatrix}1\\0\\-1\\0\\1\end{pmatrix} = \begin{pmatrix}2/3\\0\\-2/3\\0\\2/3\end{pmatrix}$$

◇

Remarque 12.29. La définition de $\mathrm{proj}_{\mathbf{w}}(\mathbf{v})$ dépend uniquement de la *direction* de $\mathbf{w}$, pas de son sens ni de sa norme. En effet, la projection sur un vecteur colinéaire à $\mathbf{w}$, $\mathbf{w}' = \lambda\mathbf{w}$, donne le même résultat, puisque

$$\begin{aligned}\mathrm{proj}_{\mathbf{w}'}(\mathbf{v}) &= \frac{\mathbf{v}\cdot\mathbf{w}'}{\|\mathbf{w}'\|^2}\mathbf{w}'\\ &= \frac{\mathbf{v}\cdot(\lambda\mathbf{w})}{\|\lambda\mathbf{w}\|^2}(\lambda\mathbf{w})\\ &= \frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{w}\|^2}\mathbf{w}\\ &= \mathrm{proj}_{\mathbf{w}}(\mathbf{v})\,.\end{aligned}$$

Donc il est plus juste de penser à la projection sur un vecteur comme à la projection sur *la droite engendrée par* ce vecteur. ◇

La projection de $\mathbf{v}$ sur $\mathbf{w}$ est aussi optimale, dans le sens où c'est elle qui réalise la distance minimale entre $\mathbf{v}$ et un vecteur quelconque de la droite dirigée par $\mathbf{w}$:

Théorème 12.30. *Soit* $\mathbf{w} \in \mathbb{R}^n$ *non-nul, et soit* $W = \mathrm{Vect}\{\mathbf{w}\}$. *Alors*

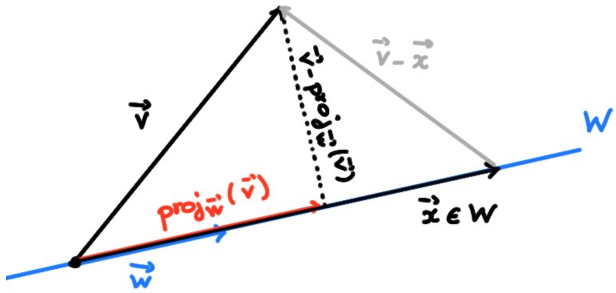
$$\|\mathbf{v} - \mathrm{proj}_{\mathbf{w}}(\mathbf{v})\| \leqslant \|\mathbf{v} - \mathbf{x}\| \qquad \forall \mathbf{x} \in W\,.$$

Puisque $\mathrm{proj}_{\mathbf{w}}(\mathbf{v}) \in W$, *ceci implique*

$$\|\mathbf{v} - \mathrm{proj}_{\mathbf{w}}(\mathbf{v})\| = \min_{\mathbf{x}\in W} \|\mathbf{v} - \mathbf{x}\|\,.$$

De plus, $\mathrm{proj}_W(\mathbf{v})$ *est l'unique vecteur de* W *qui réalise ce minimum.*

En d'autres termes, $\mathrm{proj}_{\mathbf{w}}(\mathbf{v})$ est *le* vecteur de W dont la distance à $\mathbf{v}$ est minimale :



Preuve: Pour tout $\mathbf{x} \in W$, on peut écrire

$$\mathbf{v} - \mathbf{x} = \underbrace{(\mathbf{v} - \mathrm{proj}_{\mathbf{w}}(\mathbf{v}))}_{\in W^\perp} + \underbrace{(\mathrm{proj}_{\mathbf{w}}(\mathbf{v}) - \mathbf{x})}_{\in W}$$

On a donc

$$\begin{aligned}\|\mathbf{v} - \mathbf{x}\|^2 &= \|\mathbf{v} - \mathrm{proj}_{\mathbf{w}}(\mathbf{v})\|^2 + \underbrace{\|\mathrm{proj}_{\mathbf{w}}(\mathbf{v}) - \mathbf{x}\|^2}_{\geqslant 0}\\ &\geqslant \|\mathbf{v} - \mathrm{proj}_{\mathbf{w}}(\mathbf{v})\|^2\end{aligned}$$

Supposons maintenant qu'il existe, en plus de $\mathbf{v}_\parallel = \mathrm{proj}_{\mathbf{w}}(\mathbf{v})$, un autre vecteur de W satisfaisant la même propriété ; notons-le $\mathbf{v}'_\parallel$. Alors par définition,

$$\|\mathbf{v} - \mathbf{v}_\parallel\| = \min_{\mathbf{x}\in W} \|\mathbf{v} - \mathbf{x}\| = \|\mathbf{v} - \mathbf{v}'_\parallel\|\,.$$

Aussi,

$$\begin{aligned}\|\mathbf{v} - \mathbf{v}'_\parallel\|^2 &= \|\underbrace{(\mathbf{v} - \mathbf{v}_\parallel)}_{\in W^\perp} + \underbrace{(\mathbf{v}_\parallel - \mathbf{v}'_\parallel)}_{\in W}\|^2\\ &= \|\mathbf{v} - \mathbf{v}_\parallel\|^2 + \|\mathbf{v}_\parallel - \mathbf{v}'_\parallel\|^2\end{aligned}$$

On a donc

$$\|\mathbf{v}_\parallel - \mathbf{v}'_\parallel\|^2 = 0\,,$$

qui implique $\mathbf{v}_\parallel = \mathbf{v}'_\parallel$. □

Le fait que $\mathrm{proj}_{\mathbf{w}}(\mathbf{v})$ réalise un minimum indique que certains *problèmes d'optimisation* pourront trouver une solution par l'utilisation de projections. (Voir plus loin, *Méthode des moindres carrés*.)

12.6 Projection sur un sous-espace vectoriel

Motivation : projection sur un plan de $\mathbb{R}^3$

Pour motiver la définition générale de projection sur un sous-espace vectoriel W, nous commencerons par un cas légèrement plus compliqué que la projection sur une droite (section précédente), en considérant une projection sur un plan.

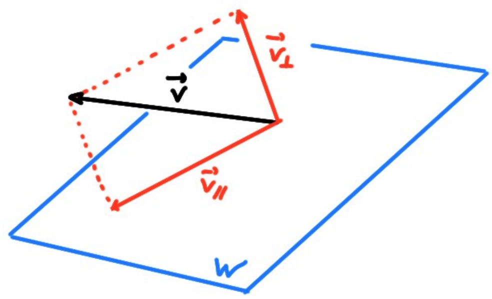
Exemple 12.31. Considérons les deux vecteurs non-colinéaires

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \qquad \mathbf{w}_2 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix},$$

ainsi que le plan contenant l'origine, engendré par ces deux vecteurs :

$$W = \mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_2\}.$$

Si $\mathbf{v} = \begin{pmatrix} 5 \\ 2 \\ -3 \end{pmatrix}$, comment calculer sa **projection orthogonale sur** W ?



Comme dans la section précédente, nous commencerons par représenter la projection de $\mathbf{v}$ sur W à l'aide du symbole $\mathbf{v}_\parallel$. Cette projection doit permettre de décomposer $\mathbf{v}$ en deux composantes vectorielles,

$$\mathbf{v} = \mathbf{v}_\parallel + \mathbf{v}_\perp,$$

où

a) $\mathbf{v}_\parallel \in W$

b) $\mathbf{v}_\perp \in W^\perp$

La première condition impose que $\mathbf{v}_\parallel$ soit une combinaison linéaire de $\mathbf{w}_1$ et $\mathbf{w}_2$:

$$\mathbf{v}_\parallel = \alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2,$$

et la deuxième impose que

$$\begin{cases} 0 = \mathbf{v}_\perp \cdot \mathbf{w}_1 = (\mathbf{v} - \alpha_1\mathbf{w}_1 - \alpha_2\mathbf{w}_2) \cdot \mathbf{w}_1, \\ 0 = \mathbf{v}_\perp \cdot \mathbf{w}_2 = (\mathbf{v} - \alpha_1\mathbf{w}_1 - \alpha_2\mathbf{w}_2) \cdot \mathbf{w}_2. \end{cases}$$

Comme $\|\mathbf{w}_1\|^2 = 3$, $\|\mathbf{w}_2\|^2 = 6$, $\mathbf{w}_1 \cdot \mathbf{w}_2 = -2$, $\mathbf{v} \cdot \mathbf{w}_1 = 10$, $\mathbf{v} \cdot \mathbf{w}_2 = -11$, on en déduit que les coefficients α_1, α_2 sont solutions du système

$$(*) \left\{ \begin{array}{rcrcr} 3\alpha_1 & - & 2\alpha_2 & = & 10 \\ -2\alpha_1 & + & 6\alpha_2 & = & -11. \end{array} \right.$$

On a donc $\alpha_1 = \frac{19}{7}$, $\alpha_2 = -\frac{13}{14}$. Ainsi, la projection de $\mathbf{v}$ sur le plan W est donnée par

$$\mathbf{v}_{\parallel} = \frac{19}{7}\mathbf{w}_1 - \frac{13}{14}\mathbf{w}_2 = \begin{pmatrix} 32/7 \\ 25/14 \\ -51/14 \end{pmatrix} .$$

◇

Cas général

Dans le cas général, énonçons d'abord le résultat général qui garantit que la projection orthogonale sur un sous-espace vectoriel existe toujours :

Théorème 12.32. *Soit W un sous-espace vectoriel de $\mathbb{R}^n$, et soit $\mathbf{v} \in \mathbb{R}^n$. Alors il existe une unique paire de vecteurs, $\mathbf{v}_{\parallel}$ et $\mathbf{v}_{\perp}$, telle que*

$$\mathbf{v} = \mathbf{v}_{\parallel} + \mathbf{v}_{\perp} ,$$

et telle que

a) $\mathbf{v}_{\parallel} \in W$

b) $\mathbf{v}_{\perp} \in W^{\perp}$

*Le vecteur $\mathbf{v}_{\parallel}$ est appelé **projection orthogonale de $\mathbf{v}$ sur W**, et sera noté*

$$\mathbf{v}_{\parallel} \equiv \mathrm{proj}_W(\mathbf{v}) .$$

De plus, $\mathrm{proj}_W(\mathbf{v})$ est l'unique vecteur de W qui minimise la distance à $\mathbf{v}$:

$$\|\mathbf{v} - \mathrm{proj}_W(\mathbf{v})\| = \min_{\mathbf{x} \in W} \|\mathbf{v} - \mathbf{x}\| .$$

Dans le cas où on connaît une famille génératrice pour W,

$$W = \mathrm{Vect}\{\mathbf{w}_1, \dots, \mathbf{w}_k\} ,$$

on peut calculer $\mathrm{proj}_W(\mathbf{v})$ comme on l'a fait dans la section précédente, en commençant par l'écrire comme une combinaison linéaire

$$\mathbf{v}_{\parallel} = \alpha_1 \mathbf{w}_1 + \cdots \alpha_k \mathbf{w}_k ,$$

où les coefficients doivent satisfaire simultanément aux k conditions suivantes :

$$\begin{cases} 0 &= (\mathbf{v} - \alpha_1 \mathbf{w}_1 \cdots - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_1 , \\ 0 &= (\mathbf{v} - \alpha_1 \mathbf{w}_1 \cdots - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_2 , \\ &\vdots \\ 0 &= (\mathbf{v} - \alpha_1 \mathbf{w}_1 \cdots - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_k . \end{cases}$$

Sans présenter de difficulté particulière, cette approche requiert malgré tout la résolution d'un système $n \times k$.

Cas où W est décrit par une base orthogonale

Lorsque W est décrit à l'aide d'une base orthogonale, la projection sur W prend une forme plus explicite :

Théorème 12.33. *Soit W un sous-espace vectoriel de $\mathbb{R}^n$, et soit $\mathcal{B} = (\mathbf{w}_1, \dots, \mathbf{w}_k)$ une base **orthogonale** de W. Alors la projection orthogonale d'un vecteur $\mathbf{v} \in \mathbb{R}^n$ sur W est donnée par*

$$\mathrm{proj}_W(\mathbf{v}) = \sum_{j=1}^{k} \frac{\mathbf{v} \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2} \mathbf{w}_j \,.$$

En particulier, l'application $\mathbf{v} \mapsto \mathrm{proj}_W(\mathbf{v})$ est linéaire.

Preuve: Comme décrit plus haut, la projection est de la forme

$$\mathrm{proj}_W(\mathbf{v}) = \alpha_1 \mathbf{w}_1 + \cdots \alpha_k \mathbf{w}_k \,,$$

où les α_j doivent satisfaire

$$\begin{cases} 0 &= (\mathbf{v} - \alpha_1 \mathbf{w}_1 \cdots - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_1 \,, \\ 0 &= (\mathbf{v} - \alpha_1 \mathbf{w}_1 \cdots - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_2 \,, \\ &\vdots \\ 0 &= (\mathbf{v} - \alpha_1 \mathbf{w}_1 \cdots - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_k \,. \end{cases}$$

Mais puisque la base $\mathcal{B}$ est orthogonale, $\mathbf{w}_i \cdot \mathbf{w}_j = 0$ si $i \neq j$. Il reste donc

$$\begin{cases} 0 &= (\mathbf{v} - \alpha_1 \mathbf{w}_1) \cdot \mathbf{w}_1 \,, \\ &\vdots \\ 0 &= (\mathbf{v} - \alpha_k \mathbf{w}_k) \cdot \mathbf{w}_k \,, \end{cases}$$

qui donne bien $\alpha_j = \frac{\mathbf{v} \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2}$ pour tout $j = 1, \dots, k$.

Vérifions la linéarité :

$$\begin{aligned} \mathrm{proj}_W(\beta_1 \mathbf{v}_1 + \beta_2 \mathbf{v}_2) &= \sum_{j=1}^{k} \frac{(\beta_1 \mathbf{v}_1 + \beta_2 \mathbf{v}_2) \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2} \mathbf{w}_j \\ &= \sum_{j=1}^{k} \left(\beta_1 \frac{\mathbf{v}_1 \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2} \mathbf{w}_j + \beta_2 \frac{\mathbf{v}_2 \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2} \mathbf{w}_j \right) \\ &= \beta_1 \sum_{j=1}^{k} \frac{\mathbf{v}_1 \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2} \mathbf{w}_j + \beta_2 \sum_{j=1}^{k} \frac{\mathbf{v}_2 \cdot \mathbf{w}_j}{\|\mathbf{w}_j\|^2} \mathbf{w}_j \\ &= \beta_1 \mathrm{proj}_W(\mathbf{v}_1) + \beta_2 \mathrm{proj}_W(\mathbf{v}_2) \,. \end{aligned}$$

□

La linéarité de la projection fait qu'on peut chercher sa matrice relativement à une base.

Exemple 12.34. Considérons les deux vecteurs non-colinéaires

$$\mathbf{w}_1 = \begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix} , \qquad \mathbf{w}_2 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} ,$$

ainsi que le plan contenant l'origine, engendré par ces deux vecteurs :

$$W = \mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_2\} \,.$$

Commençons par prendre un vecteur, par exemple

$$\mathbf{v} = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} ,$$

et calculons sa projection sur W. On pourrait procéder comme on l'a fait plus haut, mais on remarque tout de suite que cette fois, $\{\mathbf{w}_1, \mathbf{w}_2\}$ est orthogonale car $\mathbf{w}_1 \cdot \mathbf{w}_2 = 0$. On peut donc écrire la projection directement à l'aide de la formule du théorème :

$$\begin{aligned} \operatorname{proj}_W(\mathbf{v}) &= \frac{\mathbf{v} \cdot \mathbf{w}_1}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 + \frac{\mathbf{v} \cdot \mathbf{w}_2}{\|\mathbf{w}_2\|^2}\mathbf{w}_1 \\ &= \frac{-11}{30}\begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix} + \frac{-1}{6}\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2/5 \\ -2 \\ 1/5 \end{pmatrix} . \end{aligned}$$

Considérons ensuite la matrice de la projection, relativement à la base canonique :

$$\begin{aligned} [\operatorname{proj}_W]_{\mathcal{B}_{\text{can}}} &= \left[[\operatorname{proj}_W(\mathbf{e}_1)]\,[\operatorname{proj}_W(\mathbf{e}_2)]\,[\operatorname{proj}_W(\mathbf{e}_3)]\right] \\ &= \begin{pmatrix} 4/5 & 0 & -2/5 \\ 0 & 1 & 0 \\ -2/5 & 0 & 1/5 \end{pmatrix} \end{aligned}$$

◇

Cas où W est décrit par une base orthonomale

Si on exige en plus que la famille qui engendre W soit formée de vecteurs unitaires, alors la projection est encore plus simple à décrire :

Théorème 12.35. *Soit W un sous-espace vectoriel de $\mathbb{R}^n$, et soit $\mathcal{B} = (\mathbf{u}_1, \dots, \mathbf{u}_k)$ une base* ***orthonormée*** *de W. Alors la projection orthogonale d'un vecteur $\mathbf{v} \in \mathbb{R}^n$ sur W est donnée par*

$$\operatorname{proj}_W(\mathbf{v}) = \sum_{j=1}^{k} (\mathbf{v} \cdot \mathbf{u}_j)\mathbf{u}_j .$$

De plus, la matrice de proj_W, relativement à la base canonique, est donnée par

$$[\operatorname{proj}_W]_{\mathcal{B}_{\text{can}}} = UU^T ,$$

où U est la matrice $n \times k$ dont les colonnes sont les vecteurs de la base $\mathcal{B}$ exprimés dans la base canonique :

$$U = \left[[\mathbf{u}_1]_{\mathcal{B}_{\text{can}}} \cdots [\mathbf{u}_k]_{\mathcal{B}_{\text{can}}}\right] .$$

Preuve: Par le théorème précédent, et puisque $\|\mathbf{u}_j\| = 1$ pour tout j, la projection est bien donnée par

$$\operatorname{proj}_W(\mathbf{v}) = (\mathbf{v} \cdot \mathbf{u}_1)\mathbf{u}_1 + \cdots + (\mathbf{v} \cdot \mathbf{u}_k)\mathbf{u}_k .$$

Pensons maintenant en composantes relativement à la base canonique, et profitons de la structure de cette expression pour la récrire sous forme d'un produit matriciel :

$$\begin{aligned}
[\mathrm{proj}_W(\mathbf{v})]_{\mathcal{B}_{\mathrm{can}}} &= (\mathbf{v}\cdot\mathbf{u}_1)[\mathbf{u}_1]_{\mathcal{B}_{\mathrm{can}}} + \cdots + (\mathbf{v}\cdot\mathbf{u}_k)[\mathbf{u}_k]_{\mathcal{B}_{\mathrm{can}}} \\
&= \big[[\mathbf{u}_1]_{\mathcal{B}_{\mathrm{can}}} \cdots [\mathbf{u}_k]_{\mathcal{B}_{\mathrm{can}}}\big] \begin{pmatrix} \mathbf{v}\cdot\mathbf{u}_1 \\ \vdots \\ \mathbf{v}\cdot\mathbf{u}_k \end{pmatrix} \\
&= \big[[\mathbf{u}_1]_{\mathcal{B}_{\mathrm{can}}} \cdots [\mathbf{u}_k]_{\mathcal{B}_{\mathrm{can}}}\big] \begin{pmatrix} [\mathbf{u}_1]^T_{\mathcal{B}_{\mathrm{can}}}[\mathbf{v}]_{\mathcal{B}_{\mathrm{can}}} \\ \vdots \\ [\mathbf{u}_k]^T_{\mathcal{B}_{\mathrm{can}}}[\mathbf{v}]_{\mathcal{B}_{\mathrm{can}}} \end{pmatrix} \\
&= \underbrace{\big[[\mathbf{u}_1]_{\mathcal{B}_{\mathrm{can}}} \cdots [\mathbf{u}_k]_{\mathcal{B}_{\mathrm{can}}}\big]}_{=:U} \underbrace{\begin{bmatrix} [\mathbf{u}_1]^T_{\mathcal{B}_{\mathrm{can}}} \\ \vdots \\ [\mathbf{u}_k]^T_{\mathcal{B}_{\mathrm{can}}} \end{bmatrix}}_{=U^T} [\mathbf{v}]_{\mathcal{B}_{\mathrm{can}}}
\end{aligned}$$

□

Remarque 12.36. La projection est une application de $\mathbb{R}^n$ dans $\mathbb{R}^n$, donc la matrice qui la représente est $n \times n$. C'est bien le cas ici puisque

$$[\mathrm{proj}_W]_{\mathcal{B}_{\mathrm{can}}} = \underbrace{\underbrace{U}_{n\times k}\;\underbrace{U^T}_{k\times n}}_{n\times n}$$

◇

Exemple 12.37. Considérons la même projection orthogonale que celle vue plus haut, sur le plan W engendré par

$$\mathbf{w}_1 = \begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix}, \qquad \mathbf{w}_2 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix},$$

On sait que la base

$$\mathcal{B} = (\mathbf{w}_1, \mathbf{w}_2)$$

est orthogonale, et on peut la rendre orthonormée en divisant chaque vecteur par sa norme :

$$\mathcal{B}' = \left(\frac{\mathbf{w}_1}{\|\mathbf{w}_1\|}, \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|}\right).$$

On peut maintenant utiliser le théorème pour obtenir la matrice de la projection sur W relativement à la base canonique :

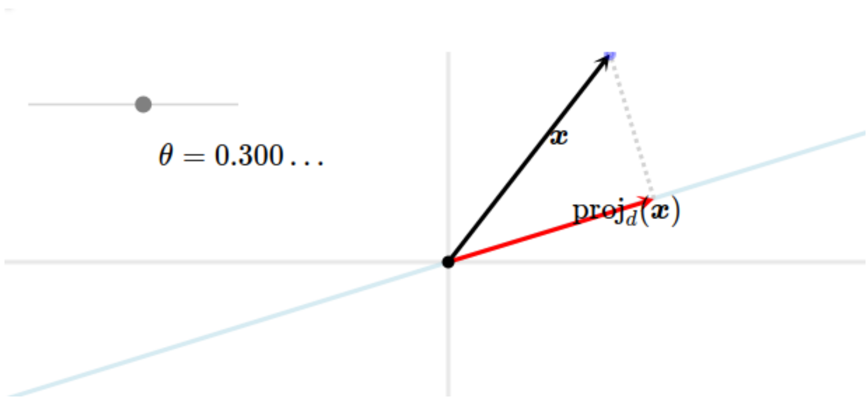
$$\begin{aligned}
[\mathrm{proj}_W]_{\mathcal{B}_{\mathrm{can}}} &= \underbrace{U}_{3\times 2}\;\underbrace{U^T}_{2\times 3} \\
&= \begin{bmatrix} \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} & \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} \end{bmatrix} \begin{bmatrix} (\frac{\mathbf{w}_1}{\|\mathbf{w}_1\|})^T \\ (\frac{\mathbf{w}_2}{\|\mathbf{w}_2\|})^T \end{bmatrix} \\
&= \begin{pmatrix} 2/\sqrt{30} & -2/\sqrt{6} \\ 5/\sqrt{30} & 1/\sqrt{6} \\ -1/\sqrt{30} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} 2/\sqrt{30} & 5/\sqrt{30} & -1/\sqrt{30} \\ -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{pmatrix} \\
&= \begin{pmatrix} 4/5 & 0 & -2/5 \\ 0 & 1 & 0 \\ -2/5 & 0 & 1/5 \end{pmatrix},
\end{aligned}$$

qui est bien la même que nous avions trouvé plus haut. On peut maintenant utiliser cette matrice pour projeter n'importe quel vecteur sur W. Par exemple,

$$\mathrm{proj}_W\left(\begin{pmatrix}1\\-2\\3\end{pmatrix}\right) = \begin{pmatrix}4/5 & 0 & -2/5\\0 & 1 & 0\\-2/5 & 0 & 1/5\end{pmatrix}\begin{pmatrix}1\\-2\\3\end{pmatrix} = \begin{pmatrix}-2/5\\-2\\1/5\end{pmatrix}$$

◇

Exemple 12.38. Considérons la projection proj_d sur une droite d passant par l'origine et faisant un angle de θ avec $\mathbf{e}_1$:



Cette droite d est un sous-espace de $\mathbb{R}^2$, engendrée par le vecteur unitaire

$$\mathbf{u}_1 = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}.$$

Par la formule du théorème ci-dessus, sa matrice relativement à la base canonique est donc donnée par

$$\begin{aligned}[\mathrm{proj}_d]_{\mathcal{B}_{can}} &= UU^T\\ &= \mathbf{u}_1\mathbf{u}_1^T\\ &= \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}\begin{pmatrix}\cos\theta & \sin\theta\end{pmatrix}\\ &= \begin{pmatrix}\cos^2\theta & \cos\theta\sin\theta\\ \cos\theta\sin\theta & \sin^2\theta\end{pmatrix}.\end{aligned}$$

◇

Remarque 12.39. Il est important d'apprécier l'ordre des matrices apparaissant dans la formule

$$[\mathrm{proj}_W]_{\mathcal{B}_{\mathrm{can}}} = UU^T$$

Le fait que les matrices soient dans cet ordre ("U fois U^T") font de leur produit une application linéaire non-triviale, qui projette sur l'espace engendré par les colonnes de U. Car si on multiplie ces matrices dans l'ordre inverse, on obtient une matrice $k \times k$ contenant les produits scalaires

$$\mathbf{u}_i^T\mathbf{u}_j = \mathbf{u}_i \cdot \mathbf{u}_j = \begin{cases}\|\mathbf{u}_i\| = 1 & \text{si } i = j\,,\\ 0 & \text{sinon,}\end{cases}$$

et donc

$$
\begin{aligned}
U^T U &= \begin{pmatrix} \mathbf{u}_1 \cdot \mathbf{u}_1 & \mathbf{u}_1 \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_1 \cdot \mathbf{u}_k \\ \mathbf{u}_2 \cdot \mathbf{u}_1 & \cdots & & \mathbf{u}_2 \cdot \mathbf{u}_k \\ \vdots & & & \vdots \\ \mathbf{u}_k \cdot \mathbf{u}_1 & \cdots & & \mathbf{u}_k \cdot \mathbf{u}_k \end{pmatrix} \\
&= \begin{pmatrix} \|\mathbf{u}_1\|^2 & 0 & \cdots & 0 \\ 0 & \cdots & & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & & \|\mathbf{u}_k\|^2 \end{pmatrix} \\
&= I_k \,.
\end{aligned}
$$

◇

12.7 Le procédé d'orthogonalisation de Gram-Schmidt

Les sections précédentes ont montré tout l'avantage de travailler avec une base orthogonale (ou orthonormale) pour un sous-espace W, puisque cela permet par exemple d'accéder directement aux composantes d'un vecteur relativement à cette base, ou de calculer plus facilement des projections orthogonales sur W.

Mais il se peut que le sous-espace W soit défini dès le départ par une base $\mathcal{B}$ qui n'est *pas* orthogonale. Pour profiter des avantages décrits ci-dessus, il est donc naturel de chercher une autre base de W, $\mathcal{B}'$, qui soit elle orthogonale.

Nous allons voir qu'une telle base existe toujours, et nous verrons comment la construire en *modifiant* la base de départ, par un algorithme appelé le *procédé d'orthogonalisation de Gram-Schmidt*.

Informel 12.40. L'idée est de "tordre" un à un les vecteurs de $\mathcal{B}$, de façon à les rendre progressivement orthogonaux deux-à-deux, et en garantissant *qu'ils engendrent toujours* W.

Voyons comment faire sur un exemple très simple d'une base ne contenant que deux vecteurs.

L'idée, sur un exemple où $\dim(W) = 2$

Considérons le plan de $\mathbb{R}^3$, $W = \text{Vect}\{\mathbf{w}_1, \mathbf{w}_2\}$, où

$$
\mathbf{w}_1 = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}, \qquad \mathbf{w}_2 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}.
$$

La paire $\mathcal{B} = (\mathbf{w}_1, \mathbf{w}_2)$ est une base de W, mais elle n'est pas orthogonale car

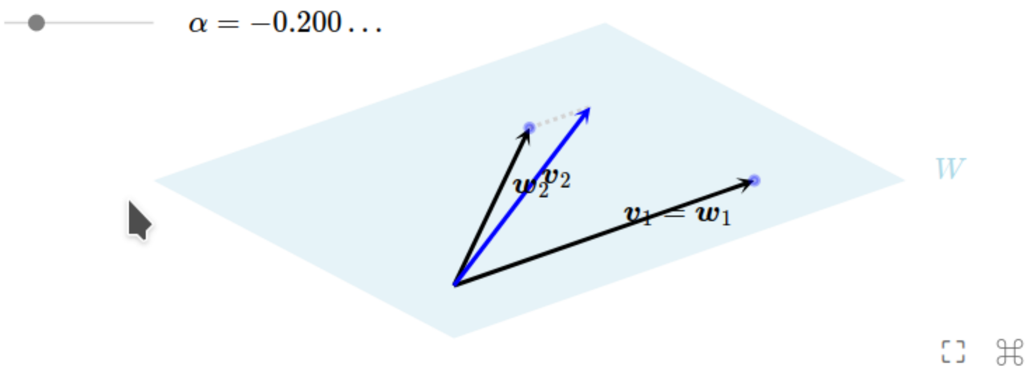
$$
\mathbf{w}_1 \cdot \mathbf{w}_2 = 3 \neq 0\,.
$$

Voyons comment modifier $\mathcal{B}$ de façon à la transformer en une autre base pour W, orthogonale cette fois.

La nouvelle base sera $\mathcal{B}' = (\mathbf{v}_1, \mathbf{v}_2)$, avec

$$\begin{aligned}\mathbf{v}_1 &:= \mathbf{w}_1\,,\\ \mathbf{v}_2 &:= \mathbf{w}_2 - \alpha\mathbf{w}_1\,.\end{aligned}$$

Voyons ce qui se passe lorsque α varie :



Informel 12.41. Remarquons que l'on "tord" $\mathbf{w}_2$ en lui rajoutant un multiple de $\mathbf{w}_1$, ce qui fait que $\mathbf{v}_2$ *reste dans le plan* W !

C'est évident sur l'animation ci-dessus, mais écrivons-le explicitement :

Lemme 18. *Peu importe la valeur du scalaire α, $\mathcal{B}' = (\mathbf{v}_1, \mathbf{v}_2)$ est toujours une base de W.*

Preuve: (en exercice) □

Il s'agit ensuite de choisir α de façon à ce que $\mathcal{B}'$ soit orthogonale. Or la seule condition à satisfaire est que

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = 0\,,$$

qui se traduit par

$$\mathbf{w}_1 \cdot (\mathbf{w}_2 - \alpha\mathbf{w}_1) = 0\,,$$

et qui implique

$$\alpha = \frac{\mathbf{w}_1 \cdot \mathbf{w}_2}{\|\mathbf{w}_1\|^2}$$

Ainsi, $\alpha\mathbf{w}_1 = \frac{\mathbf{w}_1 \cdot \mathbf{w}_2}{\|\mathbf{w}_1\|^2}\mathbf{w}_1$, qui n'est autre que la projection de $\mathbf{w}_2$ sur $\mathbf{w}_1$ (c'est-à-dire sur $\mathbf{v}_1$). En résumé, on a pris $\mathcal{B}' = (\mathbf{v}_1, \mathbf{v}_2)$, où

$$\begin{aligned}\mathbf{v}_1 &:= \mathbf{w}_1\,,\\ \mathbf{v}_2 &:= \mathbf{w}_2 - \mathrm{proj}_{\mathbf{v}_1}(\mathbf{w}_2)\,,\end{aligned}$$

qui donne

$$\mathbf{v}_1 = \begin{pmatrix}3\\-1\\2\end{pmatrix}\,,\qquad \mathbf{v}_2 = \frac{1}{14}\begin{pmatrix}5\\31\\8\end{pmatrix}\,.$$

Maintenant, $\mathcal{B}' = (\mathbf{v}_1, \mathbf{v}_2)$ est orthogonale puisque $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

La construction décrite dans l'exemple ci-dessus n'a rien de particulier à $\mathbb{R}^3$, et peut s'utiliser pour orthogonaliser la base de n'importe quel sous-espace de dimension 2 :

Exemple 12.42. Considérons le plan de $\mathbb{R}^5$ engendré par

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \qquad \mathbf{w}_2 = \begin{pmatrix} 0 \\ -2 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

La base $\mathcal{B} = (\mathbf{w}_1, \mathbf{w}_2)$ de ce plan n'est pas orthogonale, mais en prenant $\mathbf{v}_1 := \mathbf{w}_1$, et

$$\mathbf{v}_2 := \mathbf{w}_2 - \operatorname{proj}_{\mathbf{w}_1}(\mathbf{w}_2) = \begin{pmatrix} 0 \\ -2 \\ 0 \\ 1 \\ 1 \end{pmatrix} - \frac{-1}{3} \begin{pmatrix} 1 \\ 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/3 \\ -2 \\ 1/3 \\ 2/3 \\ 1 \end{pmatrix},$$

on obtient une base $\mathcal{B}' = (\mathbf{v}_1, \mathbf{v}_2)$ orthogonale. ◇

Cas général

Dans le cas général, considérons un sous-espace W de $\mathbb{R}^3$, de dimension $k \leqslant n$, muni d'une base (a priori pas orthogonale)

$$\mathcal{B} = (\mathbf{w}_1, \dots, \mathbf{w}_k),$$

et voyons comment l'utiliser pour construire une nouvelle base de W,

$$\mathcal{B}' = (\mathbf{v}_1, \dots, \mathbf{v}_k),$$

qui soit orthogonale. Cette construction est le **procédé d'orthogonalisation de Gram-Schmidt**.

L'idée est de procéder de manière **inductive**, le j-ème vecteur $\mathbf{v}_j$ de $\mathcal{B}'$ étant construit à partir des j premiers vecteurs de $\mathcal{B}$, de façon à ce que pour tout $j = 1, \dots, k$, deux conditions soient satisfaites :

- ⋆ $\{\mathbf{v}_1, \dots, \mathbf{v}_j\}$ est orthogonale (et donc libre),
- ⋆ $\operatorname{Vect}\{\mathbf{v}_1, \dots, \mathbf{v}_j\} = \operatorname{Vect}\{\mathbf{w}_1, \dots, \mathbf{w}_j\}$.

La vérification de ces conditions implique qu'à la fin, lorsque $j = k$, on a bien construit une famille $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ qui est orthogonale (et donc libre), et qui engendre W.

L'exemple précédent a suggéré de commencer par modifier $\mathbf{w}_2$ en lui soustrayant sa projection sur $\mathbf{w}_1$. Pour les suivants, on peut continuer à soustraire à chaque vecteur sa projection sur l'espace engendré par les précédents :

$$\begin{aligned}
\mathbf{v}_1 &:= \mathbf{w}_1, \\
\mathbf{v}_2 &:= \mathbf{w}_2 - \operatorname{proj}_{\mathrm{Vect}\{\mathbf{w}_1\}}(\mathbf{w}_2), \\
\mathbf{v}_3 &:= \mathbf{w}_3 - \operatorname{proj}_{\mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_2\}}(\mathbf{w}_3), \\
&\vdots \\
\mathbf{v}_j &:= \mathbf{w}_j - \operatorname{proj}_{\mathrm{Vect}\{\mathbf{w}_1, \dots, \mathbf{w}_{j-1}\}}(\mathbf{w}_j) \\
&\vdots \\
\mathbf{v}_k &:= \mathbf{w}_k - \operatorname{proj}_{\mathrm{Vect}\{\mathbf{w}_1, \dots\dots, \mathbf{w}_{k-1}\}}(\mathbf{w}_k).
\end{aligned}$$

Remarquons que le procédé nécessite, à l'étape j, de calculer la projection sur le sous-espace $\mathrm{Vect}\{\mathbf{w}_1, \ldots, \mathbf{w}_{j-1}\}$. Or, puisque

$$\mathrm{Vect}\{\mathbf{w}_1, \ldots, \mathbf{w}_{j-1}\} = \mathrm{Vect}\{\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}\}\,,$$

on a, pour tout j,

$$\mathrm{proj}_{\mathrm{Vect}\{\mathbf{w}_1,\ldots,\mathbf{w}_{j-1}\}}(\mathbf{w}_j) = \mathrm{proj}_{\mathrm{Vect}\{\mathbf{v}_1,\ldots,\mathbf{v}_{j-1}\}}(\mathbf{w}_j)\,.$$

Maintenant, comme $\{\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}\}$ est orthogonale, la formule de la section précédente permet d'écrire cette dernière projection comme

$$\mathrm{proj}_{\mathrm{Vect}\{\mathbf{v}_1,\ldots,\mathbf{v}_{j-1}\}}(\mathbf{w}_j) = \sum_{i=1}^{j-1} \frac{\mathbf{w}_j \cdot \mathbf{v}_i}{\|\mathbf{v}_i\|^2}\mathbf{v}_i\,.$$

Donc on peut écrire le procédé comme suit :

$$\begin{aligned}
\mathbf{v}_1 &:= \mathbf{w}_1\,,\\
\mathbf{v}_2 &:= \mathbf{w}_2 - \frac{\mathbf{w}_2 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1\\
\mathbf{v}_3 &:= \mathbf{w}_3 - \sum_{i=1}^{2} \frac{\mathbf{w}_3 \cdot \mathbf{v}_i}{\|\mathbf{v}_i\|^2}\mathbf{v}_i\,,\\
&\vdots\\
\mathbf{v}_k &:= \mathbf{w}_k - \sum_{i=1}^{k-1} \frac{\mathbf{w}_k \cdot \mathbf{v}_i}{\|\mathbf{v}_i\|^2}\mathbf{v}_i\,,
\end{aligned}$$

Remarque 12.43.

- ⋆ Une fois le procédé terminé, on peut toujours normaliser les vecteurs de $\mathcal{B}'$ pour en faire une base *orthonormée* de W :

$$\mathcal{B}'' = \left(\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \ldots, \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}\right)\,.$$

- ⋆ La convention est que l'algorithme du procédé de Gram-Schmidt se fait en respectant l'ordre qui est est fixé dans la base de départ.

◇

Exemple 12.44. Considérons, dans $\mathbb{R}^4$, le sous-espace W défini par

$$W = \mathrm{Vect}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}\,,$$

où

$$\mathbf{w}_1 := \begin{pmatrix}1\\0\\1\\0\end{pmatrix}\,,\quad \mathbf{w}_2 := \begin{pmatrix}-1\\0\\0\\1\end{pmatrix}\,,\quad \mathbf{w}_3 := \begin{pmatrix}0\\1\\-1\\0\end{pmatrix}\,.$$

Appliquons le procédé de Gram-Schmidt. D'abord, $\mathbf{v}_1 := \mathbf{w}_1$, puis

$$\begin{aligned}
\mathbf{v}_2 &:= \mathbf{w}_2 - \frac{\mathbf{w}_2 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1\\
&= \begin{pmatrix}-1\\0\\0\\1\end{pmatrix} - \frac{(-1)}{2}\begin{pmatrix}1\\0\\1\\0\end{pmatrix} = \begin{pmatrix}-1/2\\0\\1/2\\1\end{pmatrix}\,,
\end{aligned}$$

et pour finir

$$\begin{aligned}\mathbf{v}_3 &:= \mathbf{w}_3 - \frac{\mathbf{w}_3\cdot\mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\mathbf{w}_3\cdot\mathbf{v}_2}{\|\mathbf{v}_2\|^2}\mathbf{v}_2\\ &= \begin{pmatrix}0\\1\\-1\\0\end{pmatrix} - \frac{-1}{2}\begin{pmatrix}1\\0\\1\\0\end{pmatrix} - \frac{-1/2}{3/2}\begin{pmatrix}-1/2\\0\\1/2\\1\end{pmatrix}\\ &= \begin{pmatrix}1/3\\1\\-1/3\\1/3\end{pmatrix}.\end{aligned}$$

Remarquons que $\mathcal{B}' = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ est bien orthogonale puisque, par construction,

$$\mathbf{v}_1\cdot\mathbf{v}_2 = \mathbf{v}_1\cdot\mathbf{v}_3 = \mathbf{v}_2\cdot\mathbf{v}_3 = 0\,.$$

◇

Dans ce dernier exemple, on aurait pu remarquer dès le début que $\mathbf{w}_2 \perp \mathbf{w}_3$, et donc obtenir une base orthogonale $(\mathbf{v}_1', \mathbf{v}_2', \mathbf{v}_3')$, en gardant deux vecteurs inchangés, et en ne modifiant que $\mathbf{w}_1$:

$$\begin{aligned}\mathbf{v}_2' &:= \mathbf{w}_2\,,\\ \mathbf{v}_3' &:= \mathbf{w}_3\,,\\ \mathbf{v}_1' &:= \mathbf{w}_1 - \frac{\mathbf{w}_1\cdot\mathbf{v}_2'}{\|\mathbf{v}_2'\|^2}\mathbf{v}_2' - \frac{\mathbf{w}_1\cdot\mathbf{v}_3'}{\|\mathbf{v}_3'\|^2}\mathbf{v}_3'\,.\end{aligned}$$

Donc en général, il y a plusieurs façons d'orthogonaliser une base, mais en général, lorsqu'on implémente le procédé de Gram-Schmidt, la convention est *de modifier les vecteurs dans l'ordre donné par la base de départ*.

12.8 La décomposition QR

La décomposition QR est une méthode qui permet de *factoriser* une matrice, c'est-à-dire de l'écrire comme un produit de deux autres matrices particulières (un peu comme une matrice carrée inversible peut être factorisée en un produit de matrices élémentaires).

On le verra, pouvoir écrire une matrice comme un produit de matrices plus simples possèdera de nombreux avantages.

Lorsque les colonnes de A sont indépendantes

Soit A une matrice $m \times n$, que l'on écrit à l'aide de ses colonnes :

$$A = [\mathbf{a}_1 \cdots \mathbf{a}_n]\,,$$

où chaque $\mathbf{a}_j \in \mathbb{R}^m$.

Si les colonnes de A sont linéairement indépendantes, elles forment une base de $\mathrm{Im}(A)$. Cette base n'est a priori pas orthogonale, on peut donc lui appliquer le procédé de Gram-Schmidt :

$$\begin{aligned}
\mathbf{v}_1 &:= \mathbf{a}_1\,,\\
\mathbf{v}_2 &:= \mathbf{a}_2 - \frac{\mathbf{a}_2\cdot\mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1\\
\mathbf{v}_3 &:= \mathbf{a}_3 - \sum_{i=1}^{2}\frac{\mathbf{a}_3\cdot\mathbf{v}_i}{\|\mathbf{v}_i\|^2}\mathbf{v}_i\,,\\
&\vdots\\
\mathbf{v}_n &:= \mathbf{a}_n - \sum_{i=1}^{n-1}\frac{\mathbf{a}_n\cdot\mathbf{v}_i}{\|\mathbf{v}_i\|^2}\mathbf{v}_i\,,
\end{aligned}$$

Normalisons chacun des $\mathbf{v}_j$, en définissant

$$\mathbf{u}_1 := \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}\,,\quad \dots \quad, \mathbf{u}_k := \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}\,.$$

Les n relations ci-dessus permettent maintenant d'écrire

$$\begin{aligned}
\mathbf{a}_1 &= \|\mathbf{v}_1\|\mathbf{u}_1\,,\\
\mathbf{a}_2 &= (\mathbf{a}_2\cdot\mathbf{u}_1)\mathbf{u}_1 + \|\mathbf{v}_2\|\mathbf{u}_2\\
\mathbf{a}_3 &= \left(\sum_{i=1}^{2}(\mathbf{a}_3\cdot\mathbf{u}_i)\mathbf{u}_i\right) + \|\mathbf{v}_3\|\mathbf{u}_3\,,\\
&\vdots\\
\mathbf{a}_n &= \left(\sum_{i=1}^{n-1}(\mathbf{a}_n\cdot\mathbf{u}_i)\mathbf{u}_i\right) + \|\mathbf{v}_n\|\mathbf{u}_n\,,
\end{aligned}$$

Cette décomposition des colonnes de A en combinaisons linéaires des vecteurs $\mathbf{u}_1,\dots,\mathbf{u}_n$ peut s'exprimer à l'aide de deux matrices.

- $\star$ D'abord, considérons la matrice Q, $m\times n$, définie par

  $$Q := [\mathbf{u}_1 \cdots \mathbf{u}_n]\,.$$

  Les colonnes de cette matrice étant unitaires et orthogonales deux à deux, on a (lire la remarque de la fin de la section précédente) :

  $$Q^TQ = I_n\,.$$

- $\star$ Ensuite, considérons la matrice R, triangulaire supérieure $n\times n$, formée à partir des coefficients apparaissant dans les combinaisons linéaires ci-dessus :

$$R := \begin{pmatrix}
\|\mathbf{v}_1\| & \mathbf{a}_2\cdot\mathbf{u}_1 & \mathbf{a}_3\cdot\mathbf{u}_1 & \cdots & \mathbf{a}_n\cdot\mathbf{u}_1\\
0 & \|\mathbf{v}_2\| & \mathbf{a}_3\cdot\mathbf{u}_2 & \cdots & \mathbf{a}_n\cdot\mathbf{u}_2\\
\vdots & \ddots & \ddots & \ddots & \vdots\\
\vdots & & 0 & \|\mathbf{v}_{n-1}\| & \mathbf{a}_n\cdot\mathbf{u}_{n-1}\\
0 & \cdots & & 0 & \|\mathbf{v}_n\|
\end{pmatrix}.$$

En d'autres termes, les coefficients apparaissant dans la k-ème colonne de R sont les coefficients de la combinaison linéaire donnant $\mathbf{a}_k$.

Par définition, $QR = A$. Pour le voir explicitement, on peut écrire

$$R = [\mathbf{r}_1 \cdots \mathbf{r}_n]\,,$$

où

$$\begin{aligned}
\mathbf{r}_1 &= \|\mathbf{v}_1\|\mathbf{e}_1\\
\mathbf{r}_2 &= (\mathbf{a}_1 \cdot \mathbf{u}_1)\mathbf{e}_1 + \|\mathbf{v}_2\|\mathbf{e}_2\\
&\vdots\\
\mathbf{r}_n &= \left(\sum_{i=1}^{n-1}(\mathbf{a}_i \cdot \mathbf{u}_i)\mathbf{e}_i\right) + \|\mathbf{v}_n\|\mathbf{e}_n\,.
\end{aligned}$$

Ainsi, la k-ème colonne de $QR = [Q\mathbf{r}_1 \cdots Q\mathbf{r}_n]$ est donnée par

$$\begin{aligned}
Q\mathbf{r}_k &= Q\left(\sum_{i=1}^{k-1}(\mathbf{a}_i \cdot \mathbf{u}_i)\mathbf{e}_i + \|\mathbf{v}_k\|\mathbf{e}_k\right)\\
&= \sum_{i=1}^{k-1}(\mathbf{a}_i \cdot \mathbf{u}_i)(Q\mathbf{e}_i) + \|\mathbf{v}_k\|Q\mathbf{e}_k\\
&= \sum_{i=1}^{k-1}(\mathbf{a}_i \cdot \mathbf{u}_i)\mathbf{u}_i + \|\mathbf{v}_k\|\mathbf{u}_k\\
&= \mathbf{a}_k\,.
\end{aligned}$$

On a donc démontré :

Théorème 12.45. *Soit A une matrice $m \times n$ de rang $\operatorname{rang}(A) = n$. Alors il existe*

a) une matrice Q ($m \times n$) telle que $Q^TQ = I_n$ (c'est-à-dire dont les colonnes forment une famille orthonormale), et

b) une matrice R ($n \times n$) triangulaire supérieure,

telles que

$$A = QR\,.$$

*Cette factorisation est une **factorisation** QR **de** A.*

- ⋆ Une décomposition QR n'est pas unique, puisque l'on peut toujours par exemple changer les signes des colonnes de Q, ce qui change les coefficients de R aussi mais qui ne change pas la validité de l'identité $A = QR$.
- ⋆ A priori, R s'obtient à l'aide de l'expression ci-dessus, pleine de produits scalaires. Mais on peut l'obtenir plus simplement. En effet, remarquons que si on multiplie (à gauche) les deux côtés de "$A = QR$" par Q^T,

$$Q^TA = Q^T(QR) = (Q^TQ)R = I_nR = R\,.$$

Donc la factorisation QR d'une matrice A dont les colonnes sont linéairement indépendantes peut s'obtenir comme suit :

a) Appliquer le procédé de Gram-Schmidt aux colonnes de $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$, et normaliser les vecteurs obtenus, pour obtenir $Q := [\mathbf{u}_1 \cdots \mathbf{u}_n]$.

b) Calculer $R = Q^T A$.

Exemple 12.46. Calculons une factorisation QR de

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 0 & 2 \end{pmatrix} .$$

Ici, $A = [\mathbf{a}_1\,\mathbf{a}_2]$, où $\mathbf{a}_1$ et $\mathbf{a}_2$ sont indépendants, donc le théorème s'applique. Le procédé de Gram-Schmidt donne

$$\mathbf{v}_1 := \mathbf{a}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} , \qquad \mathbf{v}_2 := \mathbf{a}_2 - \frac{\mathbf{a}_2 \cdot \mathbf{a}_1}{\|\mathbf{a}_1\|^2}\mathbf{a}_1 = \begin{pmatrix} -1/2 \\ 1/2 \\ 2 \end{pmatrix} .$$

On a donc, après normalisation,

$$Q = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/6 \\ \sqrt{2}/2 & \sqrt{2}/6 \\ 0 & 2\sqrt{2}/3 \end{pmatrix}$$

Ensuite,

$$\begin{aligned} R = Q^T A &= \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ -\sqrt{2}/6 & \sqrt{2}/6 & 2\sqrt{2}/3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 0 & 2 \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{2} & 3\sqrt{2}/2 \\ 0 & 3\sqrt{2}/2 \end{pmatrix} \end{aligned}$$

Remarquons que cette dernière est bien

$$R = \begin{pmatrix} \|\mathbf{v}_1\| & \mathbf{a}_2 \cdot \mathbf{u}_1 \\ 0 & \|\mathbf{v}_2\| \end{pmatrix} ,$$

où $\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$. ◇

Cas général

La décomposition QR existe même si les colonnes de la matrice ne sont pas linéairement indépendantes :

Théorème 12.47. *Soit A une matrice $m \times n$ de rang $\mathrm{rang}(A) = r$. Alors il existe*

a) une matrice Q ($m \times r$) telle que $Q^T Q = I_r$ (c'est-à-dire dont les colonnes forment une famille orthonormale),

b) une matrice R ($r \times n$) triangulaire supérieure,

telles que

$$A = QR .$$

La différence avec le cas précédent est que lorsqu'on calcule Q en appliquant le procédé de Gram-Schmidt, on tombe sur des colonnes nulles, que l'on peut éliminer pour ne garder qu'une base orthonormale de $\mathrm{Col}(A)$. Après, R s'obtient aussi par $R = Q^T A$.