

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE
UNIVERSITE AMAR TELIDJI LAGHOUAT
FACULTE DES SCIENCES
DEPARTEMENT DE MATHEMATIQUES



*Mémoire de MASTER
(Master)*

Domaine : Mathématiques et Informatique

Filière : Mathématiques

Option : Analyse mathématique

Par : Smail Sara

THEME :

Séparation à vaste marge des convexes.

Soutenu publiquement devant le jury composé de :

Mr. Rahmoune Abdelaziz	M.C.B	Président
Mr. Bentobache Mohand	M.C.A	Examineur
Mr. Mokhtari Abdelkader	Professeur	Encadreur

Année Universitaire : 2019-2020

Remerciements

Tout d'abord je tiens à remercier Dieu, le tout puissant et miséricordieux, qui m'a donné la santé, la patience et la volonté d'accomplir ce modeste travail.

*En premier lieu, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de **Pr. Mokhtari Abdelkader**, on le remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur scientifique et sa disponibilité durant notre préparation de ce mémoire.*

*Mes vifs remerciements vont également aux membres du jury **Dr. Rahmoune Abdelaziz** et **Dr. Bentobache Mohand** pour l'intérêt qu'ils ont porté à notre mémoire en acceptant d'examiner notre travail et de l'enrichir par leurs propositions.*

*J'exprime toute ma reconnaissance et gratitude à l'administration et à l'ensemble du corps enseignant en particulier à **Dr. Rahmoune Abdelaziz** le chef de département de mathématiques de l'université de (Amar Telidji) pour leurs efforts à nous garantir la continuité et l'aboutissement de ce programme de Master.*

Enfin, je tiens également à remercier toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail, à mes amis, ma famille, merci.

Dédicaces

Je dédie ce modeste travail

À l'exemple de dévouement et de la source de tendresse mes chers parents, qui ont toujours fait leur maximum pour m'encourager, m'aider et me donner le support moral tout au long de mes études. Aucun mot sur cette page ne saurait exprimer ce que je vous dois, ni combien je vous aime. Qu'Allah vous bénisse, vous assiste, vous vienne en aide.

À ceux que j'aime beaucoup mon cher frère Islam Mohammed et ma chère sœur Asma,

A toute ma famille, et mes amis,

Et à tous ceux qui ont contribué de près ou de loin pour que ce travail soit possible, je vous dis merci.

Résumé :

Les séparateurs à vastes marges des convexes est une méthode de classification qui tente de séparer deux convexes non vides et disjoints d'un espace de dimension finie. La méthode cherche alors l'hyperplan qui les sépare, en garantissant que la marge entre ces deux convexes non vides et disjoints soit maximale.

Mots clés : séparation, convexe, hyperplan, Hahn-Banach, séparation à vaste marge (SVM), problème dual.

Abstract :

Support vector machine of convex sets is a classification method that attempts to separate two disjoint and nonempty convex sets from a finite dimensional space. The method then searches for the hyperplane that separates them, ensuring that the margin between these two disjoint and nonempty convex sets is maximum.

Mots clés :separation, convex, hyperplane, Hahn-Banach, support vector machine (SVM), dual problem.

اللمص

إن ناقلات أشعة السند للمجموعات المحدبة هي طريقة تصنيف تحاول الفصل بين مجموعتين محدبتين منفصلتين و غير خاليتين من فضاء ذو بعد منته. الطريقة تبحث بعد لك عن المستوي الذي يفصل بينهما، بضمان أن الهامش بين هاتين المجموعتين المنفصلتين و غير الخاليتين هو الحد الأقصى.

الكلمات المفتاحية فصل، محدب، مستو فائق، هان بناخ، ناقلات أشعة السند، مسألة ثنوية.

Table des matières

Introduction	3
1 Topologie des ensembles convexes	4
1.1 Rappel de topologie des espaces vectoriels normés	4
1.2 Ensembles convexes, exemples et premières propriétés	5
1.3 Topologie des ensembles convexes	8
1.4 Point extrémaux et faces des convexes	10
1.5 Fonctions convexes	10
1.6 Fonctions semi-continue inférieurement sur un evn	12
1.7 Fonctions convexes semi-continue inférieurement sur un evn	14
1.8 Théorème de projection sur un convexe fermé d'un espace euclidien de dimension finie	15
1.9 Intérieur relatif d'un convexe de dimension finie	16
2 Séparation des convexes	17
2.1 Fonctionnelles convexes	17
2.2 Théorème de Hahn-Banach	17
2.3 Théorèmes de séparation en dimension finie	20
2.4 Structure des convexes compacts de dimension finie	22
3 Continuité et calcul différentiel de champs scalaires et vectoriels	24
3.1 Fonctions de \mathbb{R}^n vers \mathbb{R}^m	24
3.2 Notion de continuité	25
3.2.1 Boules ouvertes et ensembles ouverts	25
3.2.2 Limite et continuité de champs scalaires et vectoriels	25
3.3 Divers notions de dérivations	28
3.3.1 La dérivée d'un champ scalaire par rapport à un vecteur	28
3.3.2 Dérivées directionnelles, dérivées partielles et dérivée de Gâteaux	29
3.3.3 Dérivées partielles d'ordre supérieur	29
3.3.4 Dérivées directionnelles et continuité	30
3.3.5 La dérivée totale	31
3.3.6 Le gradient d'un champ scalaire	32
3.3.7 Une condition suffisante de différentiabilité	32
3.4 Quelques règles et résultats utiles	33
3.4.1 Une règle de dérivation en chaîne pour les champs scalaires	33
3.4.2 Dérivée d'un champ vectoriel	33
3.4.3 La règle de dérivation en chaîne pour les champs de vecteurs	35

3.4.4	Conditions suffisantes pour avoir l'égalité des dérivées partielles mixtes	36
4	Séparation à vaste marge	37
4.1	Introduction	37
4.2	Historique	37
4.3	Principe général	38
4.4	Séparateur linéaire	38
4.4.1	Hyperplan	38
4.4.2	Les exemples	38
4.4.3	Le séparateur linéaire	39
4.5	Séparabilité	39
4.5.1	Vecteurs supports	40
4.5.2	Marge	40
4.5.3	Hyperplans canoniques	41
4.6	Le problème d'optimisation à résoudre par SVM pour le cas linéairement séparable	41
4.7	Les avantages et les inconvénients des SVM	45
5	Algorithme d'optimisation du problème	46
5.1	L'algorithme proposé	46
5.2	Applications	52
	Conclusion	59
	Bibliographie	60

Introduction

Les méthodes de classification ont pour but d'identifier les classes auxquelles appartiennent des objets à partir de certains paramètres descriptifs. Elles s'appliquent à un grand nombre d'activités humaines. Les méthodes utilisées pour la classification sont nombreuses, citons : la méthode des séparateurs à vastes marges (SVM), les Réseaux de Neurones, etc. Nous présentons dans la suite une étude de la technique SVM des convexes. Cette méthode a montré son efficacité dans de nombreux domaines d'applications tels que le traitement d'image, la catégorisation de textes et le diagnostic médicale.

Les machines à vecteurs de support (Support Vector Machine, SVM) appelés aussi séparateurs à vaste marge sont des techniques d'apprentissage destinées à résoudre des problèmes de classification. Les machines à vecteurs supports exploitent les concepts relatifs à la théorie de l'apprentissage statistique et à la théorie des bornes de Vapnik et Chervonenkis. La justification intuitive de cette méthode d'apprentissage est la suivante : si l'échantillon d'apprentissage est linéairement séparable, il semble naturel de séparer parfaitement les éléments des deux classes de telle sorte qu'ils soient le plus loin possibles de la frontière choisie.

Cette technique est une méthode de classification à deux classes qui tente de séparer les exemples positifs des exemples négatifs dans l'ensemble des exemples. La méthode cherche alors l'hyperplan qui sépare les exemples positifs des exemples négatifs, en garantissant que la marge entre le plus proche des positifs et des négatifs soit maximale. L'intérêt de cette méthode est la sélection de vecteurs supports qui représentent les vecteurs discriminant grâce auxquels est déterminé l'hyperplan. Les exemples utilisés lors de la recherche de l'hyperplan ne sont alors plus utiles et seuls ces vecteurs supports sont utilisés pour classer un nouveau cas, ce qui peut être considéré comme un avantage pour cette méthode.

Notre travail est divisé en cinq chapitres, le premier est un rappel de topologie des ensembles convexes, le deuxième chapitre nous présentons la technique de séparation des convexes et que selon le théorème de Hahn-Banach on peut toujours séparer deux convexes non vides et disjoints d'un espace de dimension finie, dans le troisième chapitre on expose quelques définitions et théorèmes sur la continuité et le calcul différentiel de champs scalaires et vectoriels et dans les chapitres quatre et cinq on présente la séparation à vaste marge et l'algorithme d'optimisation du problème.

Chapitre 1

Topologie des ensembles convexes

1.1 Rappel de topologie des espaces vectoriels normés

Soit $(E, \|\cdot\|)$ un espace vectoriel normé sur \mathbb{R} ou \mathbb{C} en bref un evn. E est donc un cas particulier d'espace métrique où la distance est donnée par :

$$d(x, y) = \|x - y\|.$$

On notera $B(x_0, r) = \{x \in E / \|x - x_0\| < r\}$, la boule ouverte de centre x et de rayon r et $\overline{B}(x_0, r) = \{x \in E / \|x - x_0\| \leq r\}$, la boule fermée de centre x et de rayon r .

Définition 1.1.1. i) Une partie $A \subset E$ d'un evn est ouverte si $\forall x_0 \in A, \exists r > 0$ tel que $B(x_0, r) \subset A$.

ii) Une partie $A \subset E$ d'un evn est fermée si son complémentaire est ouvert.

Propriétés 1.1.1. Dans un evn E , une partie A est fermée si et seulement si pour toute suite $(x_n)_{n \in \mathbb{N}} \subset A$, convergeant vers $x \in E$, alors $x \in A$.

Propriétés 1.1.2. i) Toute réunion et toute intersection finie d'ouverts est ouverte.

ii) Toute intersection et toute réunion finie de fermés est fermée.

Propriétés 1.1.3. Un point x est adhérent à un sous-ensemble A de E s'il est limite d'une suite de point de A . L'adhérence (ou fermeture) \overline{A} de A est l'ensemble des points adhérents à A . L'ensemble A est fermé si et seulement s'il contient tous ses points adhérents ($A = \overline{A}$).

Propriétés 1.1.4. Un point x est intérieur à une partie A de E s'il existe une boule $B(x, r)$ de centre x de rayon $r > 0$ incluse dans A . Il revient au même de dire que x n'est pas adhérent au complémentaire A^c de A . L'intérieur A^0 de A est l'ensemble des ses points intérieurs. L'ensemble A est ouvert si et seulement s'il est égale à son intérieur.

Définition 1.1.2. Une partie K de E est dite compacte si de tout recouvrement ouvert de K , on peut extraire un sous-recouvrement fini.

Proposition 1.1.1. *Une partie K d'un evn E est compacte si et seulement si de toute suite de points de K on peut extraire une sous-suite convergeant vers un point de K . Un compact est toujours fermé.*

Un produit de compacts, une partie fermée d'un compact, l'image d'un compact par une application continue sont compacts.

Si E est de dimension finie, ses parties compactes sont exactement ses fermés bornés.

Définition 1.1.3. *Le dual (topologique) E' d'un espace vectoriel normé E est l'ensemble des applications linéaires continues de E dans \mathbb{R} , appelées formes linéaires continues.*

Propriétés 1.1.5. *Le dual E' d'un evn E est un evn, la norme de E' est donnée par :*

$$\forall f \in E', \|f\| = \sup_{x \in E, \|x\| \leq 1} |f(x)|.$$

Notations. *Les éléments du dual E' de E seront notés de deux façons différentes selon le contexte :*

- i) *Ou bien la notation fonctionnelle $f : E \rightarrow \mathbb{R}$.*
- ii) *Ou bien la notation vectorielle par analogie avec le cas hilbertien : $y \in E'$ agit sur les éléments de E par le crochet de dualité $\langle y, x \rangle$, pour $x \in E$.*

Définition 1.1.4. *On dit qu'un evn E est réflexif si on peut l'identifier à son bidual E'' .*

Propriétés 1.1.6. *Dans un espace vectoriel de dimension finie, toutes les formes linéaires sont continues. Le dual d'un evn de dimension finie est un evn de même dimension. Les evn de dimension finie sont réflexifs.*

1.2 Ensembles convexes, exemples et premières propriétés

Dans cette partie, E est un espace vectoriel réel.

Définition 1.2.1. *Une partie C de E est dite convexe si et seulement si :*

$$\forall x, y \in C, \forall \theta \in [0, 1] : \theta x + (1 - \theta)y \in C.$$

Géométriquement, $z(\theta) = \theta x + (1 - \theta)y$ est le barycentre des points x et y affectés des coefficients respectifs θ et $1 - \theta$. Lorsque θ décrit l'intervalle $[0, 1]$, $z(\theta)$ décrit le segment fermé d'extrémités x et y . Dire que C est convexe, c'est donc dire que C contient un segment dès qu'elle contient ses extrémités.

En fait un convexe contient toute combinaison convexe de ses points :

Définition 1.2.2. *Pour toute famille finie x_1, \dots, x_n de points de E et tout système $\lambda_1, \dots, \lambda_n$ de réels positifs ou nuls avec $\sum_{i=1}^n \lambda_i = 1$, le point $\sum_{i=1}^n \lambda_i x_i$ s'appelle une combinaison convexe des points x_1, \dots, x_n .*

Proposition 1.2.1. *Soit C un convexe de E . Pour toute famille finie x_1, \dots, x_n de points de C et tout système $\lambda_1, \dots, \lambda_n$ de réels positifs ou nuls avec $\sum_{i=1}^n \lambda_i = 1$, le point $\sum_{i=1}^n \lambda_i x_i$ appartient à C .*

Démonstration. Par récurrence sur n . C'est évident si $n = 1$. Si $\lambda_1 = 1$, les autres coefficients sont nuls et il n'y a rien à montrer. Si $\lambda_1 < 1$ alors $\sum_{i=2}^n \lambda_i = 1 - \lambda_1 > 0$. Posons $\mu_i = \frac{\lambda_i}{1 - \lambda_1}$, on a $\sum_{i=2}^n \mu_i = 1$, d'où $y := \sum_{i=2}^n \mu_i x_i \in C$ (hypothèse de récurrence) et donc $x = \lambda_1 x_1 + (1 - \lambda_1)y \in C$. \square

Exemple 1.2.1. *(Boules de normes)*

Soit $(E, \|\cdot\|)$ une norme sur E . Les boules de E , ouvertes ou fermées, $B(x_0, r)$, $\bar{B}(x_0, r)$ sont des convexes.

Les boules centrées à l'origine sont des convexes particuliers, en ce qu'ils sont symétriques, c'est à dire que si $x \in B$ alors $-x \in B$ (et en particulier $0 \in B$).

Plus généralement, considérons une application $p : E \rightarrow \bar{\mathbb{R}}_+$ homogène, sous-additive, nulle en 0, c'est à dire :

- i) $p(\lambda x) = \lambda p(x) (\forall x \in E, \forall \lambda > 0)$.
- ii) $p(x + y) = p(x) + p(y) (\forall x, y \in E)$.
- iii) $p(0) = 0$.

Alors les ensembles $\{x \in E / p(x) < 1\}$ et $\{x \in E / p(x) \leq 1\}$ sont des convexes contenant 0.

Inversement, à tout convexe contenant 0, on peut associer une fonction homogène, sous-additive, nulle en 0.

Définition 1.2.3. *Si C est un convexe contenant 0, Sa jauge est la fonction $j_c : E \rightarrow \bar{\mathbb{R}}_+$ définie par :*

$$j_c(x) = \inf\{\lambda > 0 / \frac{x}{\lambda} \in C\}.$$

Alors vérifie les propriétés i) à ii) ci-dessus, et de plus :

$$\{x \in E / j_c(x) < 1\} \subset C \subset \{x \in E / j_c(x) \leq 1\}.$$

On remarquera :

- 1) j_c est à valeurs finies si et seulement si l'intersection de C avec toute demi-droite issue de 0 n'est pas réduite à $\{0\}$. (C est dit absorbant)
- 2) La fonction j_c est une norme si et seulement si de plus C est symétrique et ne contient aucune demi-droite issue de 0.

Définition 1.2.4. *(Sous-espaces affines)*

Un sous-ensemble $M \subset E$ est un sous-espace affine (en bref sea) de E si :

$$\forall x, y \in M, \forall \theta \in \mathbb{R} : \theta x + (1 - \theta)y \in M.$$

Géométriquement cela signifie que M contient une droite D dès qu'il contient deux points (distincts) de D . Un sea de E est clairement un convexe.

Définition 1.2.5. Soient M un sea de E et x_0 un élément de M , le sev $M - x_0$ s'appelle la direction du sea M .

Définition 1.2.6. Si A est une partie de E , le sous-espace affine engendré par A est le plus petit sea contenant A (l'intersection des sea contenant A). On le notera $\text{Aff}(A)$.

Définition 1.2.7. Les points x_1, \dots, x_k de E sont dits affinement indépendants si aucun d'eux n'appartient au sous-espace affine engendré par les $(k - 1)$ autres.

Définition 1.2.8. Si $E = \text{Aff}(x_1, \dots, x_p)$, on dit que x_1, \dots, x_p est un repère affine de E . Une catégorie importante de sous-espaces affines est celles des hyperplans affines.

Définition 1.2.9. Un hyperplan affine est un sous-espace propre maximal. Autrement dit M est un hyperplan affine si et seulement si :

- i) $M \neq E$.
- ii) E est le seul sea contenant strictement M .

Lemme 1.2.1. Les hyperplans affines de E sont exactement les ensembles de la forme $\{x \in E / f(x) = a\}$, où f est une forme linéaire non nulle sur E et a un réel.

Démonstration. En effet si H est un hyperplan affine, V sa direction, x_0 un point de E n'appartenant pas à V , on pose $f(v + \lambda x_0) = \lambda$ pour tout $v \in V$ et $\lambda \in \mathbb{R}$. Alors f est une forme linéaire sur E , nulle sur V , donc constante sur H . Si a est cette valeur, l'ensemble $\{x \in E / f(x) = a\}$ est un hyperplan affine contenant H , donc égale à H . □

Définition 1.2.10. Soit H un hyperplan affine ; soient f une forme linéaire non nulle sur E et a un réel tels que $H = \{x \in E / f(x) = a\}$. On appelle demi-espaces associés à l'hyperplan H les ensembles $D_+ = \{x \in E / f(x) \leq a\}$ et $D_- = \{x \in E / f(x) \geq a\}$. Les demi-espaces stricts associés à l'hyperplan H sont les ensembles $D'_+ = \{x \in E / f(x) < a\}$ et $D'_- = \{x \in E / f(x) > a\}$.

Ces ensembles sont convexes. Les paires $\{D_+, D_-\}$ et $\{D'_+, D'_-\}$ ne dépendent pas du couple (f, a) tel que $H = \{x \in E / f(x) = a\}$ (à permutation près).

On a évidemment $D'_+ = D_+ \setminus H$ et $D'_- = D_- \setminus H$.

Proposition 1.2.2. Si $(C_i)_{i \in I}$ est une famille quelconque de convexes, leur intersection $\bigcap_{i \in I} C_i$ est convexe.

En particulier toute intersection de demi-espaces affines (respectivement de demi-espaces affines stricts) est un convexe (éventuellement vide).

Définition 1.2.11. Soit A une partie de E . L'enveloppe convexe de A (notée $\text{co}A$) est l'intersection de tous les convexes contenant A .

L'enveloppe convexe $\text{co}A$ est évidemment un convexe contenant A , et c'est le plus petit. Si C est convexe contenant A , il contient $\text{co}A$.

Proposition 1.2.3. L'enveloppe convexe de A est l'ensemble de toutes les combinaisons convexes finies d'éléments de A :

$$\text{co}A = \left\{ \sum_{i=1}^n \lambda_i x_i / n \geq 1, x_i \in A, \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1 \right\}.$$

Démonstration. Comme $\text{co}A$ est convexe, cet ensemble contient les combinaisons convexes finies de ses éléments (proposition 1.2.1), donc des éléments de A . Inversement il est facile de voir que l'ensemble $C = \{\sum_{i=1}^n \lambda_i x_i / n \geq 1, x_i \in A, \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1\}$ est convexe; il contient A , donc aussi $\text{co}A$. \square

Définition 1.2.12. Une partie Γ de E est un cône de sommet x_0 si elle est stable par les homothéties de centre x_0 , de rapport $\lambda > 0$.

Autrement dit $x \in \Gamma, \lambda > 0 \Rightarrow x_0 + \lambda(x - x_0) \in \Gamma$. Si on ne précise pas le sommet x_0 , ce sera 0. Alors Γ est un cône si et seulement si $\lambda\Gamma \subset \Gamma$, pour tout $\lambda > 0$.

Définition 1.2.13. Un cône est pointé s'il contient son sommet, époinché sinon.

Proposition 1.2.4. Une partie Γ de E est un cône convexe (de sommet 0) si et seulement si elle est stable par addition et par multiplication par les réels strictement positifs.

Autrement dit $\Gamma + \Gamma \subset \Gamma$ et $\lambda\Gamma \subset \Gamma$ pour tout $\lambda > 0$. (notation : $A + B = \{a + b / a \in A, b \in B\}$).

Définition 1.2.14. Soit $A \subset E$. Le plus petit cône convexe pointé de sommet 0 contenant A s'appelle le cône engendré par A . On le note $\text{cone}A$.

1.3 Topologie des ensembles convexes

Soit $(E, \|\cdot\|)$ est un espace vectoriel normé.

Proposition 1.3.1. L'adhérence d'un convexe (respectivement d'un sea, respectivement d'un cône convexe) est un convexe (respectivement d'un sea, respectivement d'un cône convexe).

Démonstration. Si C est un convexe, x et y sont adhérents à C , il existe deux suites $(x_n)_n$ et $(y_n)_n$ dans C convergeant respectivement vers x et y . Pour tout $\theta \in [0, 1]$, la suite $(\theta x_n + (1 - \theta)y_n)_n$ est dans C et converge vers $\theta x + (1 - \theta)y$, qui est donc adhérent à C . Donc \overline{C} est convexe. Même preuve pour un sea ou un cône convexe. \square

Remarque 1.3.1. Le fait que $F \subset E$ soit fermé n'entraîne pas que $\text{co}F$ soit fermé.

Définition 1.3.1. L'enveloppe convexe fermée d'une partie A de E est l'adhérence de son enveloppe convexe. On la note $\overline{\text{co}A}$.

Proposition 1.3.2. 1) Si A est un ensemble fini, $\text{co}A$ est compacte.

2) Plus généralement si K_1, \dots, K_p sont des convexes compacts, $\text{co}(K_1 \cup \dots \cup K_p)$ est compacte.

Démonstration. Posons : $T_p = \{(\lambda_1, \dots, \lambda_p) \in [0, 1]^p / \sum_{i=1}^p \lambda_i = 1\}$: T_p est un compact (car fermé dans $[0, 1]^p$).

1) Si $A = \{a_1, \dots, a_p\}$, on considère l'application continue $f : T_p \rightarrow E$, $(\lambda_1, \dots, \lambda_p) \mapsto \sum_{i=1}^p \lambda_i a_i$; on a $\text{co}A = f(T_p)$, donc est compact (comme image continue d'un compact).

- 2) Preuve analogue en considérant $g : T_p \times K_1 \times \dots \times K_p \longrightarrow E$,
 $(\lambda_1, \dots, \lambda_p, x_1, \dots, x_p) \longmapsto \sum_{i=1}^p \lambda_i x_i$.

□

Remarque 1.3.2. Si E est de dimension finie, l'enveloppe convexe de tout compact de E est encore compacte (voir le chapitre 2). Toutefois ceci n'est plus vrai en dimension infinie.

Lemme 1.3.1. Soit C un convexe ayant un point intérieur x_0 . Pour tout point x adhérent à C , tout point y du segment semi-ouvert $[x_0, x[$ est intérieur à C .

Proposition 1.3.3. Si C est un convexe d'intérieur non vide, alors son intérieur C^0 est convexe; on a de plus $\overline{C^0} = \overline{C}$ et $\overline{C^0} = C^0$.

Démonstration. i) Si $x, y \in C^0$, tout le segment $[x, y]$ est inclus dans C^0 d'après le lemme précédent. Donc C^0 est convexe.

ii) L'inclusion $\overline{A^0} \subset \overline{A}$ est vrai pour toute partie A de E ; inversement si $x \in \overline{C}$, alors x est extrémité d'un segment ouvert inclus dans C^0 , d'après le lemme précédent, donc $x \in \overline{C^0}$.

iii) L'inclusion $\overline{A^0} \supset A^0$ est vrai pour toute partie A de E . Inversement soit x intérieur à \overline{C} ; soit $r > 0$ tel que $B(x, r) \subset \overline{C}$. Comme $\overline{C} = \overline{C^0}$, d'après ii), il existe $y \in C^0 \cap B(x, r)$. Soit $z \in E$ tel que x soit milieu du segment $[z, y]$; on a encore $z \in B(x, r)$, donc $z \in \overline{C}$; d'après le lemme précédent, le segment $]z, y]$ est inclus dans C^0 ; en particulier x est intérieur à C .

□

Remarque 1.3.3. Si E est de dimension infinie, l'égalité $\overline{C^0} = C^0$ peut être fausse lorsque C^0 est vide : par exemple, prendre $E = c_0$, l'espace des suites de réels convergent vers 0, muni de la norme $\|(x_n)_n\| = \sup_{x \in \mathbb{N}} |x_n|$ et $C = c_{00}$, le sous-espace des suites nulles à partir d'un certain rang. Alors $\overline{C} = E$, mais $C^0 = \emptyset$.

Toute fois ce phénomène se produit pas en dimension finie.

Dans le cas de demi-espaces affines, on peut dire un peu plus :

Soit H un hyperplan affine, D_+ et D_- les deux demi-espaces affines définis par H , et $D'_+ = D_+ \setminus H$, $D'_- = D_- \setminus H$ les demi-espaces stricts associés.

Proposition 1.3.4. Les assertions suivantes sont équivalentes :

- 1) H est fermé;
- 2) D_+ respectivement D_- est fermé;
- 3) D_+ respectivement D_- est d'intérieur non vide;
- 4) D'_+ respectivement D'_- est ouvert.

Remarque 1.3.4. Si $H = \{x \in E / f(x) = a\}$ où f est une forme linéaire sur E et $a \in \mathbb{R}$, alors H est fermé si et seulement si f est continue, c'est à dire $f \in E'$.

Corollaire 1.3.1. Un hyperplan affine est soit fermé, soit partout dense dans E .

Démonstration. Si l'hyperplan H n'est pas partout dense, son complémentaire $D'_+ \cup D'_-$ contient une boule $B(x, r)$ de rayon non nul. Cette boule ne peut intersecter à la fois D'_+ et D'_- (elle intersecterait alors H), donc est incluse par exemple dans D'_+ . Donc D_+ est d'intérieur non vide, et H fermé d'après la proposition précédente. □

1.4 Point extrémaux et faces des convexes

Définition 1.4.1. Soit C un convexe d'un espace vectoriel réel E . On dit qu'un point x de C est un point extrémal de C si x n'appartient à aucun segment ouvert dont les extrémités sont dans C .

Remarque 1.4.1. Il revient au même de dire que :

$$\text{Il n'existe pas } y, z \in C \text{ tels que } y \neq z \text{ et } x = \frac{y+z}{2}.$$

ou encore que :

$$y, z \in C, x = \frac{y+z}{2} \Rightarrow x = y = z.$$

Notations. On désignera par $\text{Ext}(C)$ l'ensemble des points extrémaux de C .

Définition 1.4.2. Une face d'un convexe C est un sous-convexe non vide F tel que :

$$x \in F, y, z \in C, x = \frac{y+z}{2} \Rightarrow y \text{ et } z \in F.$$

Proposition 1.4.1. Toute intersection non vide de faces de C est une face de C .

Un exemple triviale de face de C est le convexe C lui-même.

Un point $x_0 \in C$ est extrémal pour C si et seulement si $\{x_0\}$ est une face de C .

Définition 1.4.3. On dit qu'un hyperplan H est un hyperplan d'appui du convexe C si C est inclus dans l'un des deux demi-espaces définis par H et rencontre H .

Lemme 1.4.1. Si H est un hyperplan d'appui du convexe C alors $F = H \cap C$ est une face de C .

Notons que si C est d'intérieur non vide, tout hyperplan d'appui de C est fermé. Si de plus C est fermé, on obtient donc une face fermée $F = H \cap C$.

Dans la pratique les hyperplans d'appui seront obtenus au moyen d'une forme linéaire f atteignant son maximum M (ou son minimum m) sur le convexe C : alors $H = \{x \in E / f(x) = M\}$ (respectivement $H = \{x \in E / f(x) = m\}$) est un hyperplan d'appui de C . C'est le cas en particulier si f est continue et C compact (alors H est fermé).

Le lemme suivant est une application immédiate des définitions :

Lemme 1.4.2. Si F est une face de C alors $\text{Ext}F \subset \text{Ext}C$; et plus généralement toute face de F est une face de C .

1.5 Fonctions convexes

Dans ce qui suit E est un espace vectoriel.

Définition 1.5.1. Une fonction est $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ dite convexe si pour tous x, y dans E avec $f(x) < +\infty$, $f(y) < +\infty$ et tout $\theta \in [0, 1]$, on a :

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

On remarque aisément que si f est convexe, tous les ensembles $A_f(a) = \{x \in E \mid f(x) \leq a\}$ et $B_f(a) = \{x \in E \mid f(x) < a\}$ sont convexes ; la réciproque est fautive : si $f : \mathbb{R} \rightarrow \mathbb{R}$ est seulement supposée monotone, les ensembles $A_f(a)$, $B_f(a)$ sont convexes.

Définition 1.5.2. Le domaine effectif d'une fonction $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ est l'ensemble convexe :

$$\text{dom } f = \{x \in E \mid f(x) < +\infty\}.$$

Remarque 1.5.1. La fonction $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ est convexe si et seulement si :

- 1) son domaine effectif est convexe.
- 2) sa restriction $f|_{\text{dom } f}$ à son domaine effectif est convexe.

Définition 1.5.3. La fonction f est dite strictement convexe (sur son domaine effectif) si pour tous $x \neq y$ dans $\text{dom } f$ et tout $\theta \in]0, 1[$, on a l'inégalité stricte :

$$f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y)$$

Lorsqu'on la restreint à son domaine effectif, une fonction convexe devient une fonction convexe usuelle définie sur un convexe et à valeurs réelles.

Proposition 1.5.1. Si $f : C \rightarrow \mathbb{R}$ est une fonction convexe définie sur le convexe C , on peut l'étendre en une fonction convexe $\tilde{f} : E \rightarrow \mathbb{R} \cup \{+\infty\}$ en posant :

$$\tilde{f}(x) = \begin{cases} f(x) & \text{si } x \in C. \\ +\infty & \text{sinon.} \end{cases}$$

Cet artifice permet de manier simultanément des fonctions convexes f, g, \dots définies sur des convexes différents, et de considérer $f + g$, $\max(f, g), \dots$

Définition 1.5.4. Si f est une fonction $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$, son épigraphe est le sous-ensemble de $E \times \mathbb{R}$ défini par :

$$\text{epi } f = \{(x, t) \in E \times \mathbb{R} \mid f(x) \leq t\}.$$

L'épigraphe stricte de f est :

$$\text{epist } f = \{(x, t) \in E \times \mathbb{R} \mid f(x) < t\}.$$

Remarque 1.5.2. L'épigraphe de f est en fait un sous-ensemble de $\text{dom } f \times \mathbb{R}$; en fait la projection naturelle $E \times \mathbb{R} \rightarrow E$, $(x, t) \mapsto x$ envoie exactement $\text{epi } f$ sur $\text{dom } f$.

Proposition 1.5.2. La fonction $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ est convexe si et seulement si son épigraphe est convexe.

Démonstration. Si f est convexe, alors pour tous (x, t) , (y, u) dans $\text{epi } f$, et tout $\theta \in [0, 1]$, on a $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y) \leq \theta t + (1 - \theta)u$, donc $(\theta x + (1 - \theta)y, \theta t + (1 - \theta)u)$ appartient à $\text{epi } f$, qui est donc convexe.

Réciproquement, si $\text{epi } f$ est convexe, alors pour tous $x, y \in \text{dom } f$, et $\theta \in [0, 1]$, on a $(x, f(x)) \in \text{epi } f$, donc $(\theta x + (1 - \theta)y, (1 - \theta)f(y)) \in \text{epi } f$, ce qui signifie que $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$, et donc f est convexe. \square

Proposition 1.5.3. 1) *L'ensemble des fonctions convexes $E \longrightarrow \mathbb{R} \cup \{+\infty\}$ est un cône de l'ensemble des fonctions de $E \longrightarrow \mathbb{R} \cup \{+\infty\}$ (f, g convexes, $\alpha, \beta \geq 0$ entraînent $\alpha f + \beta g$ convexe).*

2) *Si $(f_i)_{i \in I}$ est une famille quelconque de fonctions convexes, l'enveloppe supérieure des $(f_i)_{i \in I}$, défini par $f = \sup_{i \in I} f_i$ est convexe.*

3) *Si g est limite (pour la convergence simple des fonctions) d'une suite de fonctions convexes, g est convexe.*

Dans certaines questions on peut avoir besoin de considérer des fonctions convexes à valeurs dans $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. On prend alors la propriété de convexité de l'épigraphe comme définition :

Définition 1.5.5. *La fonction $f : E \longrightarrow \overline{\mathbb{R}}$ est dite convexe si son épigraphe $\text{epi } f = \{(x, t) \in E \times \mathbb{R} / f(x) \leq t\}$ est convexe.*

Remarque 1.5.3. *Il est équivalent de dire que l'épigraphe strict de f est convexe.*

La définition 1.5.5 coïncide avec la définition usuelle lorsque f est à valeurs supérieure à $-\infty$. Lorsque ce n'est pas le cas, on a la caractérisation suivante :

Lemme 1.5.1. *La fonction $f : E \longrightarrow \overline{\mathbb{R}}$ est convexe si et seulement si pour tous $x, y \in C$ tels que $f(x), f(y) < +\infty$ et pour tout $\lambda \in [0, 1]$, on a l'inégalité de convexité*

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

avec les conventions usuelles $-\infty + t = -\infty, \forall t < +\infty, \quad a \cdot (-\infty) = -\infty, \forall a > 0,$
et $0 \cdot (-\infty) = 0$.

Notons que dans cette caractérisation, les points où f prend la valeur $+\infty$ n'interviennent pas.

Remarque 1.5.4. 1) *Si la fonction convexe $f : E \longrightarrow \overline{\mathbb{R}}$ prend la valeur $-\infty$ en un point x_0 , elle prend la valeur $-\infty$ sur tout l'intérieur de $\text{dom } f$. Inversement si $f(x_0) < +\infty$ et $f(x) > -\infty$ au voisinage du point x_0 (en particulier si f est finie et continue ou sci en x_0) elle ne prend nulle part la valeur $-\infty$.*

2) *La proposition 1.5.3 reste vraie pour les fonctions à valeurs dans $\overline{\mathbb{R}}$.*

Proposition 1.5.4. *Soit C un convexe d'un espace vectoriel et $f : F \longrightarrow \mathbb{R}$ une fonction strictement convexe bornée supérieurement. Tout point de C , s'il en existe, où f atteint son maximum est un point extrémal de C .*

1.6 Fonctions semi-continue inférieurement sur un evn

Soit $(E, \|\cdot\|)$ un evn. On sait qu'une caractérisation classique de la continuité d'une fonction numérique définie sur E , est la suivante :

Une fonction $f : E \longrightarrow \mathbb{R}$ est continue si et seulement si l'image réciproque de tout intervalle ouvert de \mathbb{R} est un ouvert de E .

On passe à la notion plus faible de demi-continuité inférieure si l'on se borne à demander

que les images réciproques des intervalles ouverts du type $]a, +\infty[$, soient ouverts. Si f est autorisée à prendre les valeurs $\pm\infty$, ce sont les intervalles $]a, +\infty[$ que l'on considère.

On a donc :

Définition 1.6.1. *On dit que la fonction numérique $f : E \longrightarrow \overline{\mathbb{R}}$ est semi-continue inférieurement (sci) si pour tout $a \in \overline{\mathbb{R}}$, l'ensemble $\{x \in E / f(x) > a\}$ est un ouvert de E . Il revient au même de dire que pour tout $a \in \overline{\mathbb{R}}$, l'ensemble $\{x \in E / f(x) \leq a\}$ est un fermé de E .*

En effet les ensemble $\{x \in E / f(x) > a\}$ et $\{x \in E / f(x) \leq a\}$ sont complémentaires ; le premier est ouvert si et seulement si le second est fermé.

Exemple 1.6.1. 1) *Toute fonction continue est sci.*

2) *Une fonction $f : E \longrightarrow \overline{\mathbb{R}}$ est continue si et seulement si f et $-f$ sont sci.*

Proposition 1.6.1. *La fonction numérique $f : E \longrightarrow \overline{\mathbb{R}}$ est sci si et seulement si son épigraphe est fermé.*

On a une autre caractérisation des fonctions sci en termes de limites inférieures. Si $f : E \longrightarrow \overline{\mathbb{R}}$, $x_0 \in E$, et $U \subset E$, on pose :

$$\lim_{x \rightarrow x_0} \inf_{x \in U} f(x) = \sup_{r > 0} \inf_{\substack{d(x, x_0) < r \\ x \in U}} f(x) = \lim_{r \rightarrow 0^+} \inf_{\substack{d(x, x_0) < r \\ x \in U}} f(x).$$

Proposition 1.6.2. *La fonction numérique $f : X \longrightarrow \overline{\mathbb{R}}$ est sci si et seulement si pour tout $x_0 \in X$ on a $\lim_{x \rightarrow x_0} \inf f(x) = f(x_0)$.*

Démonstration. Si f est sci, et $f(x_0) > -\infty$, alors pour tout $t < f(x_0)$ l'ensemble $\{x \in E / f(x) > t\}$ est un ouvert contenant x_0 , donc aussi une boule ouverte $B(x_0, r)$, avec $r > 0$. Donc $\lim_{x \rightarrow x_0} \inf f(x) \geq t$, et en faisant $t \longrightarrow f(x_0)$ on obtient $\lim_{x \rightarrow x_0} \inf f(x) \geq f(x_0)$. Ceci reste évidemment vrai si $f(x_0) = -\infty$. L'inégalité inverse est toujours vraie, donc $\lim_{x \rightarrow x_0} \inf f(x) = f(x_0)$.

Réciproquement, supposons cette égalité est vraie en tout point $x_0 \in E$. Soit $t \in]-\infty, +\infty[$ et supposons que $f(x_0) > t$. Il existe alors $r > 0$ tel que $\inf_{x \in B(x_0, r)} f(x) > t$. Le point x_0 est donc intérieur à l'ensemble $U_t = \{x \in X / f(x) > t\}$. Comme ceci est vraie pour tout $x_0 \in U_t$, cet ensemble est donc ouvert. Le cas $t = +\infty$ est trivial ($U_{+\infty} = \emptyset$). \square

Proposition 1.6.3. 1) *Si $(f_i)_{i \in I}$ est une famille quelconque de fonction sci de $E \longrightarrow \overline{\mathbb{R}}$, l'enveloppe supérieure des $(f_i)_{i \in I}$, $f = \sup_i f_i$, est également sci.*

2) *Si f et g sont des fonctions de X dans $\mathbb{R} \cup \{+\infty\}$, et α, β des réels strictement positifs, la fonction $\alpha f + \beta g$ est sci.*

Démonstration. 1) On a $\text{epi } f = \bigcap_i \text{epi } f_i$. C'est donc un fermé, comme intersection de fermés.

2) Pour tout sous-ensemble A de E , on a :

$$\inf_{x \in A} (\alpha f(x) + \beta g(x)) \geq \alpha \inf_{x \in A} f(x) + \beta \inf_{x \in A} g(x)$$

Il en résulte que pour tout $x_0 \in X$, on a :

$$\liminf_{x \rightarrow x_0} (\alpha f(x) + \beta g(x)) \geq \alpha \liminf_{x \rightarrow x_0} f(x) + \beta \liminf_{x \rightarrow x_0} g(x)$$

Si f et g sont sci, on en déduit

$$\liminf_{x \rightarrow x_0} (\alpha f(x) + \beta g(x)) \geq \alpha f(x_0) + \beta g(x_0).$$

L'inégalité inverse est triviale. □

Considérons maintenant un sous-ensemble Y de E et une fonction $f : Y \rightarrow \overline{\mathbb{R}}$ définie sur Y .

Définition 1.6.2. On dit que la fonction numérique $f : Y \rightarrow \overline{\mathbb{R}}$ est semi-continue inférieurement si pour tout $a \in \overline{\mathbb{R}}$, l'ensemble $\{x \in Y \mid f(x) > a\}$ est un ouvert de Y , c'est à dire l'intersection d'un ouvert de E avec Y .

Proposition 1.6.4. Si $f : Y \rightarrow \overline{\mathbb{R}}$ est une fonction sci, la fonction $\tilde{f} : E \rightarrow \overline{\mathbb{R}}$ définie par :

$$\tilde{f}(x) = \begin{cases} \liminf_{y \rightarrow x} \inf_{y \in Y} f(y) & \text{si } x \in \overline{Y}. \\ +\infty & \text{sinon.} \end{cases}$$

est un prolongement sci de f , et on a $\text{epi } \tilde{f} = \overline{(\text{epi } f)}$.

Démonstration. Le fait que $\tilde{f}(x) = f(x)$ pour $x \in Y$ résulte de la semi-continuité inférieure de f . Montrons que $\text{epi } \tilde{f} = \overline{(\text{epi } f)}$, il en résultera que \tilde{f} est sci.

- a) Si $(x, t) \in \overline{(\text{epi } f)}$, il existe une suite (x_n, t_n) convergeant vers (x, t) . En particulier $x \in \overline{Y}$. Pour tout $\varepsilon > 0$, on a $t_n \leq t + \varepsilon$, pour $n > n_0(\varepsilon)$, d'où $\tilde{f}(x) \leq \liminf f(x_n) \leq t + \varepsilon$; en faisant $\varepsilon \rightarrow 0$, on voit que $\tilde{f}(x) \leq t$, i.e. $(x, t) \in \text{epi } \tilde{f}$.
- b) Réciproquement, si $(x, t) \in \text{epi } \tilde{f}$, on a $\tilde{f}(x) < +\infty$, d'où $x \in \overline{Y}$ et $\tilde{f}(x) = \liminf_{y \rightarrow x} f(y)$. Il existe alors $x_n \rightarrow x$, $x_n \in Y$ telle que $f(x_n) \rightarrow \tilde{f}(x)$. Soit $\varepsilon > 0$; comme $t + \varepsilon > \tilde{f}(x)$, on a $f(x_n) \leq t + \varepsilon$ pour $n \geq n_0(\varepsilon)$, c'est à dire $(x_n, t + \varepsilon) \in \text{epi } f$, et par conséquent $(x, t + \varepsilon) \in \overline{(\text{epi } f)}$. En faisant $\varepsilon \rightarrow 0$, on obtient $(x, t) \in \overline{(\text{epi } f)}$. □

En particulier si Y est un ensemble fermé, toute fonction $f : Y \rightarrow \overline{\mathbb{R}}$ continue se prolonge en une fonction sci sur E en posant :

$$\tilde{f}(x) = \begin{cases} f(x) & \text{si } x \in Y. \\ +\infty & \text{sinon.} \end{cases}$$

1.7 Fonctions convexes semi-continue inférieurement sur un evn

Soit $(E, \|\cdot\|)$ un evn. En réunissant les résultats sur les fonctions convexes et les fonctions sci des deux sections précédentes, on obtient immédiatement les résultats suivants :

1.8. THÉORÈME DE PROJECTION SUR UN CONVEXE FERMÉ D'UN ESPACE EUCLIDIEN DE DIMENSION FINIE

Proposition 1.7.1. *Une fonction $f : E \rightarrow \overline{\mathbb{R}}$ est convexe sci si et seulement si son épigraphe est convexe fermé.*

Proposition 1.7.2. 1) *L'enveloppe supérieure d'une famille quelconque de fonctions convexes sci est convexe sci.*

2) *Si f et g sont des fonctions convexes sci de E dans $\mathbb{R} \cup \{+\infty\}$, et α, β des réels strictement positifs, la fonction $\alpha f + \beta g$ est convexe sci.*

Soit C un convexe de E .

Proposition 1.7.3. *Si $f : C \rightarrow \overline{\mathbb{R}}$ est une fonction convexe sci, la fonction $\tilde{f} : E \rightarrow \overline{\mathbb{R}}$ définie par :*

$$\tilde{f}(x) = \begin{cases} \lim_{y \rightarrow x} \inf_{y \in C} f(y) & \text{si } x \in \overline{C}. \\ +\infty & \text{sinon.} \end{cases}$$

est un prolongement convexe sci de f à E , et on a $\text{epi } \tilde{f} = \overline{(\text{epi } f)}$.

En particulier si C est un convexe fermé, toute fonction $f : C \rightarrow \mathbb{R}$ convexe continue se prolonge en une fonction sci en posant :

$$\tilde{f}(x) = \begin{cases} f(x) & \text{si } x \in C. \\ +\infty & \text{sinon.} \end{cases}$$

1.8 Théorème de projection sur un convexe fermé d'un espace euclidien de dimension finie

Théorème 1.8.1. *Soit C un convexe fermé non vide d'un espace euclidien de dimension finie et $x_0 \notin C$. Il existe un point unique $y_0 \in C$, appelé projection de x_0 sur C , qui réalise la distance de x_0 à C . La projection de x_0 sur C est caractérisée par :*

$$\forall y \in C, \quad \langle x_0 - y_0, y - y_0 \rangle \leq 0.$$

Démonstration. Soit $r = d(x_0, C) > 0$ et $B = C \cap \overline{B}(x_0, 2r)$. On écrit :

$$d(x_0, C) = \inf_{y \in C} \|y - x_0\| = \inf_{y \in B} \|y - x_0\|$$

Or B est un fermé borné dans un espace de dimension finie, il est donc compact. La fonction continue $y \rightarrow \|y - x_0\|$ atteint donc son minimum sur B . Le point y_0 où cette fonction atteint son minimum réalise la distance de x_0 à C et ceci prouve l'existence de la projection $y_0 \in C$ de x_0 sur C .

La projection y_0 vérifie : $\forall y \in C, \|x_0 - y\|^2 \geq \|x_0 - y_0\|^2$, d'où en développant :

$$\langle x_0 - y_0, y - y_0 \rangle \leq 1/2 \|y - y_0\|^2.$$

Soit $0 < \theta < 1$. On applique cette inégalité à $y_\theta = \theta y + (1 - \theta)y_0 \in C$ et on obtient :

$$\theta \langle x_0 - y_0, y - y_0 \rangle \leq (1/2)\theta^2 \|y - y_0\|^2.$$

En divisant par θ et en faisant tendre θ vers 0, on obtient bien :

$$\forall y \in C, \quad \langle x_0 - y_0, y - y_0 \rangle \leq 0.$$

Réciproquement, si $y_0 \in C$ vérifie $\forall y \in C, \langle x_0 - y_0, y - y_0 \rangle \leq 0$, on a bien :

$$\begin{aligned} \|x_0 - y\|^2 &= \|x_0 - y_0\|^2 + 2 \langle x_0 - y_0, y_0 - y \rangle + \|y_0 - y\|^2 \\ &\geq \|x_0 - y_0\|^2 + \|y_0 - y\|^2 \geq \|x_0 - y_0\|^2. \end{aligned}$$

et y_0 réalise la distance de x_0 à C . En plus, la dernière inégalité est stricte pour $y \neq y_0$, d'où l'unicité de la projection de x_0 sur C . \square

1.9 Intérieur relatif d'un convexe de dimension finie

Proposition 1.9.1. *Soit C un convexe dans un espace vectoriel de dimension finie E . Les assertions suivantes sont équivalentes :*

- i) $C^0 \neq \emptyset$.
- ii) $\text{Aff}(C) = E$.
- iii) C contient un repère affine.

Définition 1.9.1. *On appelle dimension d'un convexe C la dimension de $\text{Aff}(C)$. C'est aussi le nombre p maximal tel que C contiennent $p+1$ points affinement indépendants.*

Définition 1.9.2. *Soit C un convexe non vide d'un espace vectoriel de dimension finie E . On appelle intérieur relatif de C , et on note $\text{ri}(C)$ l'intérieur de C relatif au sous-espace affine $\text{Aff}(C)$.*

Un point x de C est donc relativement intérieur à C s'il existe $r > 0$ tel que la boule $B(x, r)$ vérifie $(B(x, r) \cap \text{Aff}(C)) \subset C$.

Remarque 1.9.1. *Cette notion d'intérieur relatif ne dépend pas de la norme considérée sur l'espace de dimension finie E .*

Remarque 1.9.2. *Tout convexe de dimension finie a même adhérence que son intérieur relatif. En particulier si le convexe est non vide, il en est de même de son intérieur relatif.*

Chapitre 2

Séparation des convexes

2.1 Fonctionnelles convexes

Définition 2.1.1. Une fonctionnelle non négative p , définie sur un espace vectoriel réel L , est dite convexe, si :

- i) $p(x + y) \leq p(x) + p(y)$, $\forall x, y \in L$;
- ii) $p(\alpha x) = \alpha p(x)$, $\forall x \in L$ et $\alpha > 0$.

Nous n'exigeons pas que la valeur de $p(x)$ soit finie pour tous les $x \in L$; autrement dit, nous admettons que pour certains $x \in L$ on peut avoir $p(x) = +\infty$.

Exemple 2.1.1. 1) Soit M l'espace des fonctions bornées x sur un ensemble quelconque S et soit s_0 un point fixe de S . Alors :

$$p_{s_0}(x) = |x(s_0)|$$

est une fonctionnelle convexe.

2) Soit m l'espace des suites numériques bornées $x = (x_1, x_2, \dots, x_n, \dots)$. La fonctionnelle :

$$p(x) = \sup_n |x_n|$$

est convexe.

2.2 Théorème de Hahn-Banach

Soient L un espace vectoriel réel et L_0 un sous-espace quelconque de L . Soit, en outre, f_0 une fonctionnelle linéaire définie sur le sous-espace L_0 . Une fonctionnelle f , définie sur l'espace L tout entier, est appelée prolongement de la fonctionnelle f_0 , si :

$$f(x) = f_0(x) \text{ pour tous les } x \in L_0.$$

Le problème du prolongement d'une fonctionnelle se présente fréquemment en analyse. Le rôle fondamental dans toutes ces questions appartient au théorème suivant.

Théorème 2.2.1. (de Hahn-Banach) Soit p une fonctionnelle convexe finie, définie sur l'espace vectoriel réel L , et soit L_0 un sous-espace vectoriel de L . Si f_0 est une fonctionnelle linéaire sur L_0 , majorée (sur L_0) par la fonctionnelle $p(x)$, c'est à dire si pour tout $x \in L_0$ on a :

$$f_0(x) \leq p(x). \quad (2.1)$$

alors f_0 peut être prolongée de façon à obtenir une fonctionnelle linéaire f sur L , majorée par $p(x)$ partout sur L .

Démonstration. Montrons que si $L_0 \neq L$, la fonctionnelle f_0 peut être prolongée à un sous-espace L' , plus vaste que L_0 , tout en respectant la condition (2.1). En effet, soit z un élément quelconque de L n'appartenant pas à L_0 et soit L' le sous-espace engendré par L_0 et z . Chaque élément de L' est de la forme :

$$tz + x, \text{ où } x \in L_0$$

Si f' est le prolongement cherché de la fonctionnelle f_0 à L' , on a :

$$f'(tz + x) = tf'(z) + f_0(x).$$

ou, si l'on pose $f'(z) = c$,

$$f'(tz + x) = tc + f_0(x).$$

Choisissons maintenant c de façon que la condition de majoration (2.1) soit conservée sur L' , c'est à dire de façon que pour tous les $x \in L_0$ et tous les t réels on ait l'inégalité :

$$f_0(x) + tc \leq p(x + tz).$$

pour $t > 0$ cette inégalité équivaut à la condition :

$$f_0\left(\frac{x}{t}\right) + c \leq p\left(\frac{x}{t} + z\right) \text{ ou } c \leq p\left(\frac{x}{t} + z\right) - f_0\left(\frac{x}{t}\right), \quad (2.2)$$

pour $t < 0$ elle est équivalente à la condition :

$$f_0\left(\frac{x}{t}\right) + c \geq -p\left(-\frac{x}{t} - z\right) \text{ ou } c \geq -p\left(-\frac{x}{t} - z\right) - f_0\left(\frac{x}{t}\right). \quad (2.3)$$

Montrons qu'il existe toujours un nombre c satisfaisant à ces deux conditions. Soient y' et y'' deux éléments arbitraires de L_0 . Alors :

$$-f_0(y'') + p(y'' + z) \geq -f_0(y') - p(-y' - z). \quad (2.4)$$

ceci résulte de l'inégalité :

$$f_0(y'') - f_0(y') \leq p(y'' - y') = p((y'' + z) - (y' + z)) \leq p(y'' + z) + p(-y' - z).$$

posons :

$$c'' = \inf_{y''} (-f_0(y'') + p(y'' + z)),$$

$$c' = \inf_{y'} (-f_0(y') - p(-y' - z)),$$

2.2. THÉORÈME DE HAHN-BANACH

Comme les éléments y' et y'' sont arbitraires, de (2.4) il résulte que $c'' \geq c'$. En choisissons c de façon que :

$$c' \leq c \leq c'',$$

définissons la fonctionnelle f' sur L' par la formule :

$$f'(tz + x) = tc + f_0(x).$$

cette fonctionnelle satisfait à la condition de majoration (2.1).

Ainsi, nous avons montré que si une fonctionnelle f_0 est définie sur un sous-espace $L_0 \subset L$ est vérifiée sur L_0 la condition (2.1), elle peut être prolongée à un sous-espace plus grand L' de façon que la condition (2.1) soit conservée.

Si dans L on peut choisir un système dénombrable d'éléments $x_1, x_2, \dots, x_n, \dots$ engendrant l'espace L tout entier, on construit la fonctionnelle sur L par récurrence, en considérant la suite de sous-espaces :

$$L^{(1)} = \{L_0, x_1\}, \quad L^{(2)} = \{L^{(1)}, x_2\}, \dots$$

(ici $\{L^{(k)}, x_{k+1}\}$ désigne le plus petit sous-espace vectoriel de L contenant $L^{(k)}$ et x_{k+1}). Comme tout élément $x \in L$ se trouve alors dans un sous-espace $L^{(k)}$, la fonctionnelle f_0 sera prolongée à l'espace L tout entier.

Dans le cas général (c'est à dire lorsqu'il est impossible de trouver un ensemble dénombrable d'éléments engendrant l'espace L) la démonstration fait un appel au lemme de Zorn. L'ensemble \mathfrak{F} de tous les prolongements de la fonctionnelle f_0 , vérifiant la condition (2.1), est ordonné, et tout sous-ensemble totalement ordonné $\mathfrak{F}_0 \subset \mathfrak{F}$ possède une borne supérieure. Cette borne supérieure n'est autre chose que la fonctionnelle définie sur la réunion des domaines de définition des fonctionnelles $f' \in \mathfrak{F}_0$ et coïncidant avec chacune de ces fonctionnelles f' sur son domaine de définition. En vertu du lemme de Zorn, l'ensemble \mathfrak{F} possède un élément maximal f . Cet élément maximal f est précisément la fonctionnelle cherchée. En effet, elle est un prolongement de la fonctionnelle initiale f_0 , vérifie la condition (2.1) et est définie sur tout l'espace L , car sinon nous l'aurions prolongée de la manière indiquée ci-dessus à un sous-espace plus grand que le sous-espace propre sur lequel elle est définie, de sorte que f ne serait pas un élément maximal de \mathfrak{F} . \square

Théorème 2.2.2. (de Hahn-Banach dans un espace normé) Soient E un espace normé réel, L un sous-espace de E et f_0 une fonctionnelle linéaire bornée sur L . Cette fonctionnelle peut être prolongée de façon à obtenir une fonctionnelle linéaire f sur l'espace E tout entier sans augmenter la norme, c'est à dire de façon que :

$$\|f_0\|_L = \|f\|_E.$$

Démonstration. Soit :

$$\|f_0\|_L = k.$$

Il est clair que $k\|x\|$ est une fonctionnelle convexe. En prenant cette fonctionnelle pour p et en appliquant le théorème général de Hahn-Banach, on obtient le résultat voulu. \square

Le théorème de Hahn-Banach sous cette forme admet l'interprétation géométrique suivante. L'équation :

$$f_0(x) = 1.$$

définit dans le sous-espace L un hyperplan situé à la distance $\frac{1}{\|f_0\|}$ de zéro. En prolongeant la fonctionnelle f_0 à l'espace E tout entier, sans augmenter sa norme, nous traçons par cet hyperplan "partiel" un hyperplan "plus grand" dans l'espace E tout entier, sans lui "permettre" de s'approcher de zéro.

Corollaire 2.2.1. *Soit E un espace normé. Quels que soient $x_1, x_2 \in E$, $x_1 \neq x_2$, il existe une fonctionnelle linéaire continue f qui sépare ces points, c'est à dire telle que $f(x_1) \neq f(x_2)$.*

Démonstration. Ceci équivaut à l'existence, pour tout $x_0 \neq 0$, d'une fonctionnelle séparant x_0 de 0, c'est à dire telle que $f(x_0) \neq 0$. Pour construire une telle fonctionnelle, considérons d'abord les éléments de la forme λx_0 et définissons sur le sous-espace formé par ces éléments une fonctionnelle f_0 , en posant $f(\lambda x_0) = \lambda$; ensuite, prolongeons cette fonctionnelle (en utilisant le théorème de Hahn-Banach) à l'espace E tout entier. En définitive, nous obtiendrons une fonctionnelle linéaire continue f vérifiant la condition $f(x_0) = 1 \neq 0$. \square

2.3 Théorèmes de séparation en dimension finie

Théorème 2.3.1. *(Séparation d'un point et d'un convexe fermé dans un espace de dimension finie) Si C est un convexe fermé non vide d'un espace E de dimension finie et x_0 un point de E n'appartient pas à C il existe un hyperplan affine H séparant strictement x_0 de C ; c'est à dire que C et x_0 sont inclus chacun dans un demi-espace ouvert différent défini par H . En particulier C est inclus dans l'un des demi-espaces affines associés à H .*

Démonstration. Soient x_0 et C comme dans l'énoncé. La distance $d(x_0, C) = \inf\{\|y - x_0\| \mid y \in C\}$ de x_0 à C est strictement positive, puisque $x_0 \notin C$ et C est fermé. Le théorème de projection sur un convexe fermé d'un espace euclidien 1.8.1, dit qu'il existe un unique point y_0 de C réalisant cette distance (i.e. $\|y_0 - x_0\| = d(x_0, C)$), caractérisé par les inéquations :

$$\forall y \in C, \quad \langle y - y_0, x_0 - y_0 \rangle \leq 0.$$

Soit H l'hyperplan médiateur du segment $[x_0, y_0]$. Pour tout $y \in C$, on a :

$$\|y - x_0\|^2 = \|y - y_0\|^2 + 2 \langle y - y_0, y_0 - x_0 \rangle + \|y_0 - x_0\|^2 > \|y - y_0\|^2.$$

donc y appartient au demi-espace ouvert $\{x \in E \mid d(x, y_0) < d(x, x_0)\}$ défini par H auquel x_0 n'appartient pas. \square

Remarque 2.3.1. *Il revient au même de dire que si x_0 n'appartient pas au convexe fermé C , il existe une forme linéaire f sur E (de la forme $x \mapsto \langle u, x \rangle$) telle que $\sup_{x \in C} f(x) < f(x_0)$.*

Remarque 2.3.2. *Les résultats précédents 2.3.1, sont vrais dans un espace vectoriel normé de dimension infinie; la preuve donnée ici est spécifique à la dimension finie.*

Théorème 2.3.2. (*Séparation d'un point et d'un convexe ouvert dans un espace de dimension finie*) Soit C un convexe ouvert non vide d'un espace E de dimension finie et x_0 un point de E n'appartient pas à C . Alors il passe un hyperplan disjoint de C .

Théorème 2.3.3. Si C est un convexe non vide d'un espace E de dimension finie et x_0 un point de E n'appartient pas à $\text{ri}(C)$, il existe un hyperplan affine H de E , contenant x_0 et disjoint de $\text{ri}(C)$. De plus C est inclus dans l'un des demi-espaces affines associés à H .

Remarque 2.3.3. On voit donc qu'il existe une forme linéaire non nulle f sur E (de la forme $x \mapsto \langle u, x \rangle$) telle que $\sup_{x \in C} f(x) \leq f(x_0)$.

D'après les théorèmes 2.3.1 et 2.3.2 de séparation, on déduit ce qu'on appelle habituellement les théorèmes de Hahn-Banach :

Corollaire 2.3.1. (*Théorème de Hahn-Banach, première forme géométrique dans les espaces de dimension finie*) Soient C_1 et C_2 deux convexes non vides et disjoints de l'espace vectoriel normé E . On suppose que C_1 est ouvert. Il existe alors un hyperplan affine H séparant strictement C_1 de C_2 . Plus précisément il existe une forme linéaire f sur E et un réel a tels que $\forall x_1 \in C_1, \forall x_2 \in C_2, f(x_1) < a \leq f(x_2)$.

Démonstration. On applique le théorème 2.3.3 au point 0 et au convexe non vide $C_1 - C_2 = \{y - z \mid y \in C_1, z \in C_2\}$. D'où il existe f forme linéaire telle que $\forall y \in C_1, \forall z \in C_2, f(y - z) < f(0) = 0$; on en déduit, en notant $a = \inf_{z \in C_2} f(z)$ que $\forall x_1 \in C_1, \forall x_2 \in C_2, f(x_1) \leq a \leq f(x_2)$. La première inéquation signifie que C_1 est inclus dans l'intérieur du demi-espace $f(x) \leq a$; comme C_1 est ouvert, il est en fait inclus dans l'intérieur du demi-espace, c'est à dire le demi-espace ouvert $f(x) < a$. \square

Corollaire 2.3.2. (*Théorème de Hahn-Banach, deuxième forme géométrique dans les espaces de dimension finie*) Soit C un convexe non vide de l'espace vectoriel normé E . Tout sous-espace affine M disjoint de C est inclus dans un hyperplan affine H disjoint de C .

Démonstration. Le convexe $C - M = \{x - y \mid x \in C, y \in M\}$ est non vide et ne contient pas le point 0 . Il existe donc d'après le théorème 2.3.3 une forme linéaire f telle que $f(x - y) > f(0) = 0$ pour tout $x \in C, y \in M$. Alors la forme linéaire f est constante sur M (si elle prend des valeurs distinctes en deux points $y_1, y_2 \in M$, sa restriction à la droite $y_1 y_2$ prend toutes les valeurs réelles, donc ne peut être majorée par $f(x)$). Soit a cette valeur constante. On a donc $M \subset H = \{y \mid f(y) = a\}$ qui est l'hyperplan affine recherché. \square

Corollaire 2.3.3. Par tout point non intérieur d'un convexe de E il passe un hyperplan d'appui.

Démonstration. Si le convexe C est d'intérieur non vide c'est une conséquence directe du théorème 2.3.2 (puisque $C^0 = \text{ri}(C)$). Si C est d'intérieur vide, $\text{Aff}(C)$ est un sous-espace affine de E , différent de E . Il est donc inclus dans un hyperplan affine H ; celui-ci contient C , donc est bien un hyperplan d'appui de C . \square

Corollaire 2.3.4. Tout convexe fermé de dimension finie distinct de l'espace entier est intersection des demi-espaces fermés qui le contiennent et qui sont associés à des hyperplans d'appui.

2.4 Structure des convexes compacts de dimension finie

Proposition 2.4.1. *Soit K un convexe compact d'un espace vectoriel normé E . Pour tout hyperplan vectoriel fermé V de E il existe un hyperplan d'appui fermé de K parallèle à V . De plus K est l'intersection des demi-espaces le contenant déterminés par ses hyperplans d'appui.*

Démonstration. Tout hyperplan vectoriel fermé V a pour équation $f(x) = 0$, où f est une forme linéaire non nulle. Cette forme linéaire est évidemment continue, elle atteint donc son maximum M et son minimum m sur K . Les hyperplans d'équations $f(x) = m$ et $f(x) = M$ rencontrent donc K . Comme K est inclus dans les demi-espaces $f(x) \geq m$ et $f(x) \leq M$, ces hyperplans sont bien des hyperplans d'appui de K , de direction V .

Si $x_0 \in E$ est un point n'appartenant pas à K , il existe d'après la proposition 2.3.1 une forme linéaire continue f telle que $M := \sup_{x \in K} f(x) < f(x_0)$. Le demi-espace $f(x) \leq M$ contient K mais non x_0 . \square

Remarque 2.4.1. *En fait on voit que soit K est inclus dans un hyperplan parallèle à V , soit il est inclus dans une bande définie par deux hyperplans d'appui parallèles à V .*

Le théorème qui suit est la version finie-dimensionnelle d'un théorème de Krein et Milman.

Théorème 2.4.1. *(Krein-Milman) Tout convexe compact K de dimension finie est l'enveloppe convexe de ses points extrémaux.*

En vue de préciser le théorème de structure précédent, on rappelle le théorème de Carathéodory :

Proposition 2.4.2. *(Théorème de Carathéodory) Soit E un espace vectoriel de dimension finie p , et A un sous-ensemble de E .*

- i) *Tout point du cône convexe engendré par A est combinaison linéaire à coefficients positifs d'au plus p éléments de A .*
- ii) *Tout point de l'enveloppe convexe de A est combinaison convexe d'au plus $p + 1$ éléments de A .*

Corollaire 2.4.1. *Dans un espace de dimension finie p , tout point d'un convexe compact est combinaison convexe d'au plus $p + 1$ points extrémaux du convexe.*

Une autre conséquence du théorème de Carathéodory est relative à l'enveloppe convexe d'un ensemble compact, ce résultat a été annoncé au chapitre 1, remarque 1.3.2.

Proposition 2.4.3. *Dans un espace vectoriel de dimension finie, l'enveloppe convexe de tout compact est encore compacte.*

Démonstration. Soit p la dimension de E , K un compact de E ;
posons : $T_p = \{(\lambda_1, \dots, \lambda_p) \in [0, 1]^p / \sum_{i=1}^p \lambda_i = 1\}$: T_p est un compact (car fermé dans $[0, 1]^p$). Soit h l'application continue $T_{p+1} \times K^{p+1} \longrightarrow E$,
 $(\lambda_1, \dots, \lambda_{p+1}, x_1, \dots, x_{p+1}) \longrightarrow \sum_{i=1}^{p+1} \lambda_i x_i$. On a $\text{co}K = h(T_{p+1} \times K^{p+1})$ d'après le théorème de Carathéodory, donc $\text{co}K$ est compacte (image d'un compact par une application continue). \square

Chapitre 3

Continuité et calcul différentiel de champs scalaires et vectoriels

3.1 Fonctions de \mathbb{R}^n vers \mathbb{R}^m

Nous considérons ici des fonctions

$$f : V \longrightarrow W.$$

où V et W sont des espaces vectoriels de dimensions finies. Plus précisément, nous considérons le choix : $V = \mathbb{R}^n$ et $W = \mathbb{R}^m$. Lorsque $m = n = 1$, une telle fonction est appelée fonction d'une variable réelle à valeurs réelles. Lorsque $n = 1$ et $m > 1$, cette fonction est appelée une fonction vectorielle d'une variable réelle à valeurs réelles. Nous faisons l'hypothèse ici que $n > 1$ et $m \geq 1$. Lorsque $m = 1$, la fonction est appelée fonction à valeurs réelles d'une variable vectorielle réelle, ou plus brièvement, un champ scalaire. Lorsque $m > 1$, elle est appelée fonction à valeurs vectorielles réelles d'une variable vectorielle, ou tout simplement champ de vecteurs (réel).

Nous allons nous intéresser ici à étendre les concepts connus, de limite, continuité et dérivée à des champs scalaires et vectoriels. Nous utilisons, dans la suite de chapitre, les notations suivantes. Si f est un champ scalaire défini en un point $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, les notations $f(\mathbf{x})$ et $f(x_1, \dots, x_n)$ seront utilisées pour désigner la valeur de f en ce point particulier. Si \mathbf{f} est un champ de vecteurs, nous écrivons également $\mathbf{f}(\mathbf{x})$ ou $\mathbf{f}(x_1, \dots, x_n)$.

Définissons le produit scalaire usuel de deux vecteurs réels comme :

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{k=1}^n x_k y_k, \quad \forall (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n,$$

et la norme associé :

$$\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{\frac{1}{2}}.$$

Les points dans le plan sont généralement notés (x, y) à la place de (x_1, x_2) et (x, y, z) plutôt que (x_1, x_2, x_3) dans le cas tridimensionnel.

Les champs scalaires et vectoriels définis sur des sous-ensembles de \mathbb{R}^2 ou \mathbb{R}^3 (voir plus) apparaissent très souvent et de manière naturelle dans les sciences de l'ingénieur. En effet, dans de nombreux problèmes, on s'intéresse aux variations d'un champ. Dans

le cas unidimensionnel, c'est la dérivée qui traduit cette idée. La notion de dérivée s'applique aux fonctions définies sur des ouverts. Généralisons cette idée dans le cas de \mathbb{R}^n .

3.2 Notion de continuité

3.2.1 Boules ouvertes et ensembles ouverts

Soit \mathbf{x}_0 un point de \mathbb{R}^n et $r > 0$ un nombre réel donné, strictement positif. L'ensemble des points \mathbf{x} de \mathbb{R}^n tels que : $\|\mathbf{x} - \mathbf{x}_0\| < r$, est appelée une n -boule ouverte de rayon r et de centre \mathbf{x}_0 . On la note $B(\mathbf{x}_0, r)$. Un exemple est donné en dimension un par un intervalle ouvert de centre \mathbf{x}_0 . Dans \mathbb{R}^2 , nous trouvons le disque circulaire ouvert de centre \mathbf{x}_0 et de rayon r . Dans \mathbb{R}^3 , c'est la boule usuelle ouverte de centre \mathbf{x}_0 et de rayon r .

Définition 3.2.1. (d'un point intérieur de \mathcal{S}). Soit \mathcal{S} un sous-ensemble de \mathbb{R}^n et soit $\mathbf{x}_0 \in \mathcal{S}$. Alors, \mathbf{x}_0 est appelée un point intérieur de \mathcal{S} s'il existe une n -boule ouverte de centre \mathbf{x}_0 , tel que tous ses points appartenant à \mathcal{S} . L'ensemble de tous les points intérieurs de \mathcal{S} est appelée l'intérieur de \mathcal{S} et noté $\text{int}\mathcal{S}$.

Un ouvert contenant un point \mathbf{x}_0 est appelée un voisinage de \mathbf{x}_0 .

Définition 3.2.2. (d'un ouvert). Un ensemble \mathcal{S} de \mathbb{R}^n est appelée ouvert si tous ses points sont des points intérieurs. En d'autres termes, si et seulement si $\mathcal{S} = \text{int}\mathcal{S}$.

Exemple 3.2.1. En dimension un, nous pouvons donner l'exemple d'un intervalle ouvert, ou encore d'une réunion d'intervalles ouverts. Un contre-exemple est un intervalle fermé. En dimension deux, un disque ouvert est un exemple (sans compter le bord). Un autre exemple est un rectangle du type $]a, b[\times]c, d[$. Un contre-exemple est un ouvert de \mathbb{R} considéré dans \mathbb{R}^2 .

Introduisons maintenant la notion d'extérieur et de frontière.

Définition 3.2.3. (d'extérieur et de frontière). Un point \mathbf{x} est dit être extérieur à un ensemble \mathcal{S} dans \mathbb{R}^n s'il existe une n -boule $B(\mathbf{x})$ ne contenant aucun point de \mathcal{S} . L'ensemble de tous les points dans \mathbb{R}^n extérieurs à \mathcal{S} est appelée l'extérieur de \mathcal{S} et est noté $\text{ext}\mathcal{S}$. Un point qui n'est ni dans l'extérieur ni dans l'intérieur de \mathcal{S} est appelé un point frontière de \mathcal{S} et est noté $\partial\mathcal{S}$. Un ensemble \mathcal{S} de \mathbb{R}^n est dit fermé si son complémentaire dans \mathbb{R}^n (noté $-\mathcal{S}$ ou encore \mathcal{S}^c) est ouvert.

3.2.2 Limite et continuité de champs scalaires et vectoriels

Les concepts de limite et continuité sont facilement étendus à des champs scalaires et vectoriels. Nous allons reformuler ce concept pour des champs vectoriels, celui-ci étant directement applicable aux champs scalaires.

Avant cela, commençons par rappeler la définition de la limite puis continuité d'une fonction dans \mathbb{R}^n .

3.2. NOTION DE CONTINUITÉ

Définition 3.2.4. Soit f une fonction \mathbb{R} de dans \mathbb{R} . Nous dirons que la fonction f admet comme limite ℓ en un point x_0 si :

$$\forall \varepsilon > 0, \exists \eta > 0, |x - x_0| < \eta \Rightarrow |\ell - f(x)| < \varepsilon.$$

Nous le noterons :

$$\lim_{x \rightarrow x_0} f(x) = \ell.$$

Définition 3.2.5. Soit f une fonction de \mathbb{R} dans \mathbb{R} . La fonction f est dite continue en x_0 si :

$$\forall \varepsilon > 0, \exists \eta > 0, |x - x_0| < \eta \Rightarrow |f(x_0) - f(x)| < \varepsilon.$$

Considérons maintenant une fonction $f : \mathcal{S} \rightarrow \mathbb{R}^m$, où \mathcal{S} est un sous-ensemble de \mathbb{R}^n . Soient $\mathbf{x}_0 \in \mathbb{R}^n$ et $\mathbf{L} \in \mathbb{R}^m$. Nous écrivons :

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \mathbf{f}(\mathbf{x}) = \mathbf{L}, \quad (3.1)$$

Ce que signifie que :

$$\lim_{\|\mathbf{x} - \mathbf{x}_0\| \rightarrow 0} \|\mathbf{f}(\mathbf{x}) - \mathbf{L}\| = 0. \quad (3.2)$$

Le symbole limite dans l'équation (3.2) est la limite au sens usuel du calcul élémentaire. Dans cette définition, il n'est pas nécessaire que \mathbf{f} soit définie en \mathbf{x}_0 .

Ecrivons $\mathbf{h} = \mathbf{x} - \mathbf{x}_0$; alors, l'équation (3.2) devient :

$$\lim_{\|\mathbf{h}\| \rightarrow 0} \|\mathbf{f}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{L}\| = 0.$$

Pour des points dans \mathbb{R}^2 , nous écrivons (x, y) pour \mathbf{x} et (x_0, y_0) pour \mathbf{x}_0 . Ainsi, la relation (3.1) prend la forme :

$$\lim_{(x,y) \rightarrow (x_0,y_0)} \mathbf{f}(x, y) = \mathbf{L}.$$

Pour des points dans \mathbb{R}^3 , nous considérons la notation $\mathbf{x} = (x, y, z)$ et $\mathbf{x}_0 = (x_0, y_0, z_0)$. Par conséquent, nous avons :

$$\lim_{(x,y,z) \rightarrow (x_0,y_0,z_0)} \mathbf{f}(x, y, z) = \mathbf{L}.$$

Une fonction est dite continue en \mathbf{x}_0 si \mathbf{f} est définie en \mathbf{x}_0 et si :

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_0).$$

Définition 3.2.6. Nous dirons que \mathbf{f} est continue en \mathbf{x}_0 sur un ensemble \mathcal{S} si \mathbf{f} est continue en tout point de \mathcal{S} . On le note $\mathbf{f} \in \mathcal{C}^0(\mathcal{S})$.

Puisque ces définitions sont des extensions directes de celles établies dans le cas unidimensionnel, il n'est pas surprenant d'apprendre que beaucoup de propriétés familières de la limite et de la continuité peuvent aussi être étendues. Pour les champs scalaires, les théorèmes basiques concernant les limites et continuités de sommes, produits et quotients ne sont pas définis mais nous avons les théorèmes suivantes.

Théorème 3.2.1. Si $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \mathbf{f}(\mathbf{x}) = \mathbf{L}$ et $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \mathbf{g}(\mathbf{x}) = \mathbf{M}$, nous avons également :

- a) $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} (\mathbf{f} + \mathbf{g})(\mathbf{x}) = \mathbf{L} + \mathbf{M}$,
- b) $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \lambda \mathbf{f}(\mathbf{x}) = \lambda \mathbf{L}$, $\forall \lambda \in \mathbb{R}$,
- c) $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} (\mathbf{f} \cdot \mathbf{g})(\mathbf{x}) = \mathbf{L} \cdot \mathbf{M}$,
- d) $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \|\mathbf{f}(\mathbf{x})\| = \|\mathbf{L}\|$.

Exemple 3.2.2. (continuité des composantes d'un champ de vecteurs). Si un champ vectoriel \mathbf{f} a ses valeurs dans \mathbb{R}^m , chaque valeur $\mathbf{f}(\mathbf{x})$ de la fonction a m composantes et nous pouvons écrire :

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x})).$$

Les champs scalaires f_1, \dots, f_m sont appelés composantes du champ de vecteur \mathbf{f} . On peut montrer que \mathbf{f} est continue en un point si et seulement si chaque composante f_k est continue en ce point.

On peut construire d'autres exemples de fonctions grâce au théorème suivant sur les fonctions composées.

Théorème 3.2.2. Soient \mathbf{f} et \mathbf{g} des fonctions telles que la fonction composée $\mathbf{f} \circ \mathbf{g}$ soit définie au voisinage de \mathbf{x}_0 , où :

$$(\mathbf{f} \circ \mathbf{g})(\mathbf{x}) = \mathbf{f}(\mathbf{g}(\mathbf{x})).$$

Si \mathbf{g} est continue en \mathbf{x}_0 et si \mathbf{f} est continue en $\mathbf{g}(\mathbf{x}_0)$, alors la composée $\mathbf{f} \circ \mathbf{g}$ est continue en \mathbf{x}_0 .

Exemple 3.2.3. Le théorème précédent implique la continuité des champs scalaires \mathbf{h} , où $\mathbf{h}(x, y)$ est donnée par les formules telle que :

- i) $\sin(x^2 y)$,
- ii) $\log(x^2 + y^2)$,
- iii) $\frac{\exp(x+y)}{x+y}$,
- iv) $\log(\cos(x^2 + y^2))$.

Ces exemples conduisent à des fonctions continues en tout point où la fonction est définie. La première est continue en tous les points du plan, la seconde en tous les points en dehors de l'origine, la troisième en tous les points tels que $x + y \neq 0$, et enfin, la quatrième en tous les points tels que $x^2 + y^2$ n'est pas un multiple impair de $\frac{\pi}{2}$. Plus précisément, ce dernier ensemble correspond aux points (x, y) tels que :

$$x^2 + y^2 = \frac{n\pi}{2}, \quad n = 1, 3, 5, \dots$$

C'est une famille de cercles centrés à l'origine. Ces exemples montrent notamment que l'ensemble des discontinuités d'une fonction de deux variables peut être un ou des points isolés, des courbes entières ou des familles de courbes.

Exemple 3.2.4. Une fonction de deux variables peut être continue en chacune des variables séparément et être discontinue comme une fonction de deux variables. Vous pouvez considérer à titre d'exemple la fonction définie par :

$$f(x, y) = \frac{xy}{x^2 + y^2}, \quad \text{si } (x, y) \neq (0, 0),$$
$$f(0, 0) = 0.$$

3.3 Divers notions de dérivations

Nous introduirons dans cette section plusieurs notions de dérivée d'un champ scalaire : dérivée par rapport à un vecteur, directionnelle, partielle, de Gâteaux et totale. Nous verrons que ces notions sont distinctes. Nous établirons cependant un certain nombre de liens entre elle et avec le propriété de continuité.

3.3.1 La dérivée d'un champ scalaire par rapport à un vecteur

Soit f un champ scalaire défini sur un ensemble \mathcal{S} de \mathbb{R}^n , et soit \mathbf{x}_0 un point intérieur de \mathcal{S} . Nous souhaitons étudier la manière dont varie un champ scalaire lorsque nous bougeons de \mathbf{x}_0 vers un point proche. Par exemple, supposons que $f(\mathbf{x}_0)$ représente la température en un point \mathbf{x}_0 donné dans une salle chauffée dont une fenêtre est ouverte. Si nous nous rapprochons de la fenêtre, la température tend à décroître, si nous nous rapprochons du chauffage, la température augmente. En général, la manière dont le champ change dépend de la direction selon laquelle nous nous dirigeons à partir de \mathbf{x}_0 . Supposons que nous spécifions cette direction par un second vecteur \mathbf{y} . Plus précisément, supposons que nous allions de \mathbf{x}_0 vers $\mathbf{x}_0 + \mathbf{y}$ le long de la ligne joignant \mathbf{x}_0 et $\mathbf{x}_0 + \mathbf{y}$. Chaque point de ce segment est alors de la forme $\mathbf{x}_0 + h\mathbf{y}$, où $h \in \mathbb{R}$. La distance de \mathbf{x}_0 à $\mathbf{x}_0 + h\mathbf{y}$ est $\|h\mathbf{y}\| = |h|\|\mathbf{y}\|$. Puisque \mathbf{x}_0 est un point intérieur de \mathcal{S} , il existe une n -boule $B(\mathbf{x}_0, r)$ contenue entièrement dans \mathcal{S} . Si h est choisit tel que $|h|\|\mathbf{y}\| < r$, le segment \mathbf{x}_0 reliant à $\mathbf{x}_0 + \mathbf{y}$ se trouve dans \mathcal{S} . Supposons que $h \neq 0$ mais suffisamment petit pour garantir que $\mathbf{x}_0 + h\mathbf{y} \in \mathcal{S}$ et formons le quotient :

$$\frac{f(\mathbf{x}_0 + h\mathbf{y}) - f(\mathbf{x}_0)}{h}. \quad (3.3)$$

Le numérateur de ce quotient nous dit comment varie la fonction lorsque nous bougeons de \mathbf{x}_0 à $\mathbf{x}_0 + h\mathbf{y}$. Le quotient est appelé le taux de variation de f sur le segment de ligne joignant \mathbf{x}_0 à $\mathbf{x}_0 + h\mathbf{y}$. Intéressons nous au comportement de ce quotient lorsque $h \rightarrow 0$.

Définition 3.3.1. (de la dérivée d'un champ scalaire par rapport à un vecteur). Soit $f : \mathcal{S} \rightarrow \mathbb{R}$ un champ scalaire donné. Soit \mathbf{x}_0 un point intérieur de \mathcal{S} et soit \mathbf{y} un point arbitraire dans \mathbb{R}^n . La dérivée de f en \mathbf{x}_0 par rapport à \mathbf{y} , notée $f'(\mathbf{x}_0, \mathbf{y})$, est définie par l'équation :

$$f'(\mathbf{x}_0, \mathbf{y}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x}_0 + h\mathbf{y}) - f(\mathbf{x}_0)}{h},$$

lorsque la limite définie dans le membre de droite de l'équation ci-dessus existe.

Théorème 3.3.1. (de la valeur moyenne pour les dérivées de champs scalaires). Supposons que la dérivée $f'(\mathbf{x}_0 + h\mathbf{y}, \mathbf{y})$ existe pour tout h dans l'intervalle $0 \leq h \leq 1$. Alors, pour un certain θ réel dans l'intervalle ouvert $0 < \theta < 1$, nous avons :

$$f(\mathbf{x}_0 + \mathbf{y}) - f(\mathbf{x}_0) = f'(\mathbf{z}, \mathbf{y}), \text{ où } \mathbf{z} = \mathbf{x}_0 + \theta\mathbf{y}.$$

3.3.2 Dérivées directionnelles, dérivées partielles et dérivée de Gâteaux

Dans la cas particulier où \mathbf{y} est un vecteur unitaire (c'est à dire $\|\mathbf{y}\| = 1$), la distance entre \mathbf{x}_0 et $\mathbf{x}_0 + h\mathbf{y}$ est $|h|$. Dans ce cas, le quotient (3.3) représente le taux de f variations de par unité de distance le long du segment joignant \mathbf{x}_0 à $\mathbf{x}_0 + h\mathbf{y}$; la dérivée $f'(\mathbf{x}_0, \mathbf{y})$ est appelée dérivée directionnelle.

Définition 3.3.2. (des dérivées directionnelles, partielles et de la dérivée de Gâteaux). Si \mathbf{y} est un vecteur unitaire, la dérivée $f'(\mathbf{x}_0, \mathbf{y})$ est appelée dérivée directionnelle de f en \mathbf{x}_0 selon la direction \mathbf{y} . En particulier, si $\mathbf{y} = \mathbf{e}_k$ (le k -ème vecteur unitaire des coordonnées), la dérivée directionnelle $f'(\mathbf{x}_0, \mathbf{e}_k)$ est appelée la dérivée partielle de f par rapport à \mathbf{e}_k et est notée $D_k f(\mathbf{x}_0)$. Ainsi, nous avons :

$$D_k f(\mathbf{x}_0) = f'(\mathbf{x}_0, \mathbf{e}_k).$$

La fonction f est dite Gâteaux-dérivable en \mathbf{x}_0 si et seulement si f est dérivable en \mathbf{x}_0 par rapport à tous les vecteurs \mathbf{y} , et que l'application $\mathbf{y} \rightarrow f'(\mathbf{x}_0, \mathbf{y})$ est linéaire. Cette dernière application linéaire est alors appelée dérivée de Gâteaux de f en \mathbf{x}_0 .

Les notations suivantes sont également utilisées pour les dérivées partielles en un point \mathbf{a} .

$$D_k f(a_1, \dots, a_n), \frac{\partial f}{\partial x_k}(a_1, \dots, a_n), f'_{x_k}(a_1, \dots, a_n).$$

Quelques fois, la dérivée f'_{x_k} est écrite sans le prime comme f_{x_k} . Dans \mathbb{R}^2 , les vecteurs unitaires des coordonnées s'écrivent \mathbf{i} et \mathbf{j} . Si $\mathbf{x}_0 = (x_0, y_0)$, les dérivées partielles $f'(\mathbf{x}_0, \mathbf{i})$ et $f'(\mathbf{x}_0, \mathbf{j})$ s'écrivent aussi :

$$\frac{\partial f}{\partial x}(x_0, y_0) \text{ et } \frac{\partial f}{\partial y}(x_0, y_0),$$

respectivement. Dans \mathbb{R}^3 , si $\mathbf{x}_0 = (x_0, y_0, z_0)$, les dérivées partielles $D_1 f(\mathbf{x}_0)$, $D_2 f(\mathbf{x}_0)$, et $D_3 f(\mathbf{x}_0)$ sont aussi notées :

$$\frac{\partial f}{\partial x}(x_0, y_0, z_0), \frac{\partial f}{\partial y}(x_0, y_0, z_0) \text{ et } \frac{\partial f}{\partial z}(x_0, y_0, z_0).$$

3.3.3 Dérivées partielles d'ordre supérieur

Les dérivées partielles produisent de nouveaux champs scalaires $D_1 f, \dots, D_n f$ pour un champ f donné. Les dérivées partielles de $D_1 f, \dots, D_n f$ sont appelées dérivées secondes de f . Pour les fonctions de variables, il y a quatre dérivées partielles secondes que l'on écrit :

$$D_1(D_1 f) = \frac{\partial^2 f}{\partial x^2}, D_1(D_2 f) = \frac{\partial^2 f}{\partial x \partial y}, D_2(D_1 f) = \frac{\partial^2 f}{\partial y \partial x} \text{ et } D_2(D_2 f) = \frac{\partial^2 f}{\partial y^2}.$$

On utilise quelques fois la notation $D_{i,j} f$ pour la dérivée seconde $D_i(D_j f)$. Par exemple, $D_{1,2} f = D_1(D_2 f)$. Selon la notation ∂ , on précise l'ordre de dérivation en écrivant :

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right).$$

3.3.4 Dérivées directionnelles et continuité

Dans le cas unidimensionnel, l'existence de la dérivée d'une fonction f en un point implique la continuité en ce point. Ceci se montre facilement en prenant $h \neq 0$ et en écrivant :

$$f(x_0 + h) - f(x_0) = \left(\frac{f(x_0 + h) - f(x_0)}{h} \right) h.$$

Lorsque $h \rightarrow 0$, le membre de droite tend vers 0 ($= f'(x_0) \cdot 0$) et ainsi $f(x_0 + h) \xrightarrow{h \rightarrow 0} f(x_0)$.

Appliquons maintenant le même argument à un champ scalaire général. Supposons que la dérivée $f'(\mathbf{x}_0, \mathbf{y})$ existe pour un certain \mathbf{y} . Alors, si $h \neq 0$, on peut écrire :

$$f(\mathbf{x}_0 + h\mathbf{y}) - f(\mathbf{x}_0) = \left(\frac{f(\mathbf{x}_0 + h\mathbf{y}) - f(\mathbf{x}_0)}{h} \right) h.$$

Lorsque h tend vers 0, le membre de droite tend vers $f'(\mathbf{x}_0, \mathbf{y}) \cdot h = 0$; ainsi, l'existence de $f'(\mathbf{x}_0, \mathbf{y})$ pour un \mathbf{y} donné implique que :

$$\lim_{h \rightarrow 0} f(\mathbf{x}_0 + h\mathbf{y}) = f(\mathbf{x}_0).$$

pour le même \mathbf{y} . Ceci signifie que $f(\mathbf{x}) \rightarrow f(\mathbf{x}_0)$ lorsque $\mathbf{x} \rightarrow \mathbf{x}_0$ le long d'une ligne droite passant par \mathbf{x}_0 et ayant comme direction \mathbf{y} . Si $f'(\mathbf{x}_0, \mathbf{y})$ existe pour tout vecteur \mathbf{y} , alors $f(\mathbf{x}) \rightarrow f(\mathbf{x}_0)$ lorsque $\mathbf{x} \rightarrow \mathbf{x}_0$ le long d'une ligne droite passant par \mathbf{x}_0 . Ceci semble suggérer que f est continue en \mathbf{x}_0 . De manière assez surprenante, cette conclusion peut être fautive. Les exemples suivants décrivent des champs scalaires qui possèdent une dérivée directionnelle selon chaque direction partant de $\mathbf{0}$ mais qui ne sont pas continus en ce point.

Un premier exemple est donné par la fonction f définie de la manière suivante :

$$\begin{aligned} f(x, y) &= 1 \text{ si } x < 0 \text{ ou } y > x^2, \\ f(0, 0) &= 0 \text{ sinon.} \end{aligned}$$

Il est clair que ce champ admet une dérivée directionnelle (nulle) selon chaque direction. Toutefois, le champ n'est pas continu à l'origine.

Soit maintenant $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ telle que :

$$\begin{aligned} f(x, y) &= \frac{xy^2}{x^2 + y^4}, \text{ si } x \neq 0, \\ f(0, 0) &= 0, \text{ sinon.} \end{aligned}$$

Soit $\mathbf{x}_0 = (0, 0)$ et $\mathbf{y} = (a, b)$ un vecteur. Si $a \neq 0$ et si $h \neq 0$, nous avons :

$$\frac{f(\mathbf{0} + h\mathbf{y}) - f(\mathbf{0})}{h} = \frac{f(ha, hb)}{h} = \frac{ab^2}{a^2 + h^2b^4}.$$

Soit $h \rightarrow 0$. Nous trouvons $f'(\mathbf{0}, \mathbf{y}) = \frac{b^2}{a}$. Si $\mathbf{y} = (0, b)$, nous trouvons, de manière similaire que $f'(\mathbf{0}, \mathbf{y}) = 0$. Ainsi, $f'(\mathbf{0}, \mathbf{y})$ existe pour toute direction \mathbf{y} . De plus, $f(\mathbf{x}) \rightarrow 0$ lorsque $\mathbf{x} \rightarrow \mathbf{0}$ le long de toute ligne droite partant de l'origine. Toutefois, en chaque point de la parabole $x = y^2$ (excepté à l'origine) la fonction a comme valeur

1/2. Puisque de tels points existent arbitrairement proche de l'origine et que $f(\mathbf{0}) = 0$, la fonction f n'est pas continue en $\mathbf{0}$. Cet exemple montre que l'existence de toutes les dérivées directionnelles n'implique pas la continuité en ce point. Pour cette raison, les dérivées directionnelles sont une extension insatisfaisante du concept de dérivée. Une généralisation plus satisfaisante existe. Elle implique la continuité et simultanément, nous permet d'étendre les principaux résultats rencontrés dans le cas unidimensionnel à des dimension supérieures. C'est ce que l'on appelle la dérivée totale.

3.3.5 La dérivée totale

Rappelons que, dans le cas unidimensionnel, une fonction f qui admet une dérivée en un point x_0 peut être approchée près de ce point par une approximation linéaire de Taylor. Si $f'(x_0)$ existe, nous notons $E(x_0, h)$ la différence :

$$E(x_0, h) = \frac{f(x_0 + h) - f(x_0)}{h} - f'(x_0), \text{ si } h \neq 0. \quad (3.4)$$

Définissons $E(x_0, 0) = 0$. De (3.4), nous obtenons la formule :

$$f(x_0 + h) = f(x_0) + f'(x_0)h + hE(x_0, h),$$

qui a également lieu pour $h = 0$. C'est ce que l'on appelle le développement de Taylor du premier ordre pour approcher $f(x_0 + h) - f(x_0)$ par $f'(x_0)h$. L'erreur commise est $hE(x_0, h)$. De (3.4), nous voyons que $E(x_0, h) \rightarrow 0$ lorsque $h \rightarrow 0$. En cela, l'erreur commise sur $hE(x_0, h)$ est d'ordre inférieur à h , pour h petit. Ce point de vue qui consiste à approcher une fonction différentiable par une fonction linéaire suggère une manière d'étendre le concept de différentiabilité à des dimensions supérieures.

Soit $f : \mathcal{S} \rightarrow \mathbb{R}$ un champ scalaire défini sur un ensemble \mathcal{S} de \mathbb{R}^n . Soit \mathbf{x}_0 un point intérieur de \mathcal{S} et $B(\mathbf{x}_0, r)$ une n -boule se trouvant dans \mathcal{S} . Soit \mathbf{v} un vecteur tel que $\|\mathbf{v}\| < r$; ainsi, $\mathbf{x}_0 + \mathbf{v} \in B(\mathbf{x}_0, r)$.

Définition 3.3.3. *Nous dirons que f est différentiable en \mathbf{x}_0 s'il existe une transformation linéaire*

$$T_{\mathbf{x}_0} : \mathbb{R}^n \rightarrow \mathbb{R},$$

et une fonction scalaire $E(\mathbf{x}_0, \mathbf{v})$ telle que :

$$f(\mathbf{x}_0 + \mathbf{v}) = f(\mathbf{x}_0) + T_{\mathbf{x}_0}(\mathbf{v}) + \|\mathbf{v}\|E(\mathbf{x}_0, \mathbf{v}), \quad (3.5)$$

pour $\|\mathbf{v}\| < r$, où $\lim_{\|\mathbf{v}\| \rightarrow 0} E(\mathbf{x}_0, \mathbf{v}) = 0$. La transformation linéaire $T_{\mathbf{x}_0}$ est appelée la dérivée totale¹ de f en \mathbf{x}_0 .

L'équation (3.5), qui a lieu pour $\|\mathbf{v}\| < r$, est appelée formule de Taylor du premier ordre pour $f(\mathbf{x}_0 + \mathbf{v})$. Ceci donne une approximation linéaire, $T_{\mathbf{x}_0}(\mathbf{v})$, de la différence $f(\mathbf{x}_0 + \mathbf{v}) - f(\mathbf{x}_0)$. L'erreur dans l'approximation est $\|\mathbf{v}\|E(\mathbf{x}_0, \mathbf{v})$, un terme qui est d'ordre inférieur à $\|\mathbf{v}\|$ lorsque $\|\mathbf{v}\| \rightarrow 0$ (c'est à dire $E(\mathbf{x}_0, \mathbf{x}) = o(\|\mathbf{v}\|)$ lorsque $\|\mathbf{v}\| \rightarrow 0$). Le prochain théorème montre que si la dérivée totale existe, alors elle est unique. Il nous dit également comment calculer la dérivée totale $T_{\mathbf{x}_0}(\mathbf{v})$, $\forall \mathbf{v} \in \mathbb{R}^n$.

1. La dérivée totale n'est pas un nombre mais une application linéaire. La quantité $T_{\mathbf{x}_0}(\mathbf{v})$ est un nombre; il est défini pour point \mathbf{v} de \mathbb{R}^n . La dérivée totale a été introduite par Young en 1908 puis par Fréchet en 1911.

Théorème 3.3.2. *Supposons que f soit différentiable en \mathbf{x}_0 et de dérivée totale $T_{\mathbf{x}_0}$. Alors, la dérivée $f'(\mathbf{x}_0, \mathbf{y})$ existe pour tout $\mathbf{y} \in \mathbb{R}^n$ et nous avons :*

$$T_{\mathbf{x}_0}(\mathbf{y}) = f'(\mathbf{x}_0, \mathbf{y}).$$

De plus, $f'(\mathbf{x}_0, \mathbf{y})$ est une combinaison linéaire des composantes de \mathbf{y} . En fait, en posant $\mathbf{y} = (y_1, \dots, y_n)$, nous avons :

$$f'(\mathbf{x}_0, \mathbf{y}) = \sum_{k=1}^n D_k f(\mathbf{x}_0) y_k = \langle Df(\mathbf{x}_0), \mathbf{y} \rangle. \quad (3.6)$$

3.3.6 Le gradient d'un champ scalaire

Nous pouvons récrire (3.6) comme :

$$f'(\mathbf{x}_0, \mathbf{y}) = \langle \nabla f(\mathbf{x}_0), \mathbf{y} \rangle,$$

où $\nabla f(\mathbf{x}_0)$ est le vecteur dont les composantes sont les dérivées partielles de f en \mathbf{x}_0 .

$$\nabla f(\mathbf{x}_0) = (D_1 f(\mathbf{x}_0), \dots, D_n f(\mathbf{x}_0)).$$

C'est ce que l'on appelle le gradient de f . Le gradient ∇f est un champ vectoriel défini en chaque point \mathbf{x}_0 où les dérivées partielles $D_1 f(\mathbf{x}_0), \dots, D_n f(\mathbf{x}_0)$ existent. On trouve également quelques fois la notation $\text{grad} f$.

Nous pouvons, sous ces notations, récrire la formule de Taylor au premier ordre :

$$f(\mathbf{x}_0 + \mathbf{v}) = f(\mathbf{x}_0) + \langle \nabla f(\mathbf{x}_0), \mathbf{v} \rangle + \|\mathbf{v}\| E(\mathbf{x}_0, \mathbf{v}), \quad (3.7)$$

où $E(\mathbf{x}_0, \mathbf{v}) \xrightarrow{\|\mathbf{v}\| \rightarrow 0} 0$. Cette formule est alors similaire à celle obtenue dans le cas unidimensionnel où $\nabla f(\mathbf{x}_0)$ joue le rôle de $f'(\mathbf{x}_0)$. A partir de la formule de Taylor, on peut montrer que la différentiabilité implique la continuité.

Théorème 3.3.3. *Si un champ f est différentiable en \mathbf{x}_0 . Alors f est continue en \mathbf{x}_0 . Dans le cas bidimensionnel, le vecteur gradient s'écrit :*

$$\nabla f(x, y) = \frac{\partial f}{\partial x}(x, y) \mathbf{i} + \frac{\partial f}{\partial y}(x, y) \mathbf{j}.$$

En dimension trois, nous avons :

$$\nabla f(x, y, z) = \frac{\partial f}{\partial x}(x, y, z) \mathbf{i} + \frac{\partial f}{\partial y}(x, y, z) \mathbf{j} + \frac{\partial f}{\partial z}(x, y, z) \mathbf{k}.$$

3.3.7 Une condition suffisante de différentiabilité

Si f est différentiable en \mathbf{x}_0 , alors toutes les dérivées partielles $D_1 f(\mathbf{x}_0), \dots, D_n f(\mathbf{x}_0)$ existent. Toutefois, l'inverse est faux (le contre-exemple précédent).

Le prochain théorème montre que l'existence de dérivées partielles continues en un point implique la différentiabilité en ce point.

Théorème 3.3.4. *(Une condition suffisante de différentiabilité). Supposons que les dérivées partielles $D_1 f(\mathbf{x}_0), \dots, D_n f(\mathbf{x}_0)$ existent dans une n -boule $B(\mathbf{x}_0)$ et sont continues en \mathbf{x}_0 . Alors, f est différentiable en \mathbf{x}_0 .*

Remarque 3.3.1. *Un champ scalaire satisfaisant les hypothèses du théorème précédent est appelé continument différentiable en \mathbf{x}_0 .*

3.4 Quelques règles et résultats utiles

3.4.1 Une règle de dérivation en chaîne pour les champs scalaires

Dans le cas unidimensionnel, la règle de calcul en chaîne permet de calculer la dérivée d'une fonction composée $g(t) = f(r(t))$ par la formule :

$$g'(t) = f'(r(t))r'(t).$$

Nous allons étendre cette formule lorsque :

- f est remplacée par un champ scalaire défini sur un ensemble d'un espace n -dimensionnel de f .
- r est remplacée par une fonction d'une variable réelle à valeurs dans le domaine de f .

Nous verrons ensuite comment étendre la formule pour f et r des champs vectoriels.

Théorème 3.4.1. *Soit f un champ scalaire défini sur un ensemble ouvert \mathcal{S} dans \mathbb{R}^n et soit \mathbf{r} une fonction à valeurs vectorielles réelles qui transporte un intervalle J de \mathbb{R} dans \mathcal{S} . Définissons la fonction composée $g = f \circ \mathbf{r}$ sur J par la relation :*

$$g(t) = f(\mathbf{r}(t)), \text{ si } t \in J.$$

Soit t_0 un point de J où $\mathbf{r}'(t_0)$ existe et supposons que f est différentiable en $\mathbf{r}(t_0)$. Alors, $g'(t_0)$ existe et est égale au produit scalaire :

$$g'(t_0) = \langle \nabla f(x_0), \mathbf{r}'(t_0) \rangle, \text{ où } x_0 = \mathbf{r}(t_0).$$

3.4.2 Dérivée d'un champ vectoriel

La théorie de la dérivation pour les champs de vecteurs est une extension directe de celle pour les champs scalaires. Soit $\mathbf{f} : \mathcal{S} \rightarrow \mathbb{R}^m$ un champ de vecteurs défini sur un sous-ensemble \mathcal{S} de \mathbb{R}^n . Si \mathbf{x}_0 est un point intérieur de \mathcal{S} et si \mathbf{y} est un vecteur de \mathbb{R}^n , on définit la dérivée totale $\mathbf{f}'(\mathbf{x}_0, \mathbf{y})$ par la formule :

$$\mathbf{f}'(\mathbf{x}_0, \mathbf{y}) = \lim_{h \rightarrow 0} \frac{\mathbf{f}(\mathbf{x}_0 + h\mathbf{y}) - \mathbf{f}(\mathbf{x}_0)}{h}$$

dès que la limite existe. La dérivée $\mathbf{f}'(\mathbf{x}_0, \mathbf{y})$ est un ouvert de \mathbb{R}^m .

Soit f_k la k -ème composante de \mathbf{f} . Remarquons que la dérivée $\mathbf{f}'(\mathbf{x}_0, \mathbf{y})$ existe si et seulement si $f'_k(\mathbf{x}_0, \mathbf{y})$ existe pour chaque $1 \leq k \leq m$, auquel cas nous avons :

$$\mathbf{f}'(\mathbf{x}_0, \mathbf{y}) = (f'_1(\mathbf{x}_0, \mathbf{y}), \dots, f'_m(\mathbf{x}_0, \mathbf{y})) = \sum_{k=1}^m f'_k(\mathbf{x}_0, \mathbf{y}) \mathbf{e}_k,$$

où \mathbf{e}_k est le k -ème vecteur des coordonnées.

Nous dirons que \mathbf{f} est différentiable en un point \mathbf{x}_0 s'il existe une transformation linéaire :

$$T_{\mathbf{x}_0} : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

3.4. QUELQUES RÈGLES ET RÉSULTATS UTILES

telle que :

$$\mathbf{f}(\mathbf{x}_0 + \mathbf{v}) = \mathbf{f}(\mathbf{x}_0) + T_{\mathbf{x}_0}(\mathbf{v}) + \|\mathbf{v}\|\mathbf{E}(\mathbf{x}_0, \mathbf{v}), \quad (3.8)$$

où $\mathbf{E}(\mathbf{x}_0, \mathbf{v}) \rightarrow \mathbf{0}$ lorsque $\mathbf{v} \rightarrow \mathbf{0}$. La formule de Taylor au premier ordre (3.8) a lieu pour tout \mathbf{v} tel que $\|\mathbf{v}\| < r$ pour des $r > 0$. Le terme $\mathbf{E}(\mathbf{x}_0, \mathbf{v})$ est un vecteur de \mathbb{R}^m . La transformation linéaire $T_{\mathbf{x}_0}$ est appelée dérivée totale \mathbf{f} de en \mathbf{x}_0 . Pour des champs scalaires, nous avons prouvé que $T_{\mathbf{x}_0}(\mathbf{y})$ est le produit scalaire du gradient $\nabla f(\mathbf{x}_0)$ avec \mathbf{y} . Pour les champs de vecteurs, nous allons prouver que $T_{\mathbf{x}_0}(\mathbf{y})$ est un vecteur dont la k -ème composante est le produit scalaire $\langle \nabla f_k(\mathbf{x}_0), \mathbf{y} \rangle$.

Théorème 3.4.2. *Supposons que \mathbf{f} soit différentiable en \mathbf{x}_0 et de dérivée totale $T_{\mathbf{x}_0}$. Alors, la dérivée $\mathbf{f}'(\mathbf{x}_0, \mathbf{y})$ existe pour tout \mathbf{x}_0 dans \mathbb{R}^n et nous avons :*

$$T_{\mathbf{x}_0}(\mathbf{y}) = \mathbf{f}'(\mathbf{x}_0, \mathbf{y}).$$

De plus, si $\mathbf{f} = (f_1, \dots, f_m)$ et si $\mathbf{y} = (y_1, \dots, y_n)$, nous avons :

$$T_{\mathbf{x}_0}(\mathbf{y}) = \sum_{k=1}^m \langle \nabla f_k(\mathbf{x}_0), \mathbf{y} \mathbf{e}_k \rangle = (\langle \nabla f_1(\mathbf{x}_0), \mathbf{y} \rangle, \dots, \langle \nabla f_m(\mathbf{x}_0), \mathbf{y} \rangle). \quad (3.9)$$

On peut en fait récrire plus simplement sous forme matricielle cette dernière relation :

$$T_{\mathbf{x}_0}(\mathbf{y}) = Df(\mathbf{x}_0)\mathbf{y},$$

où $Df(\mathbf{x}_0)$ est la matrice $m \times n$ dont la k -ème ligne est $\nabla f_k(\mathbf{x}_0)$ et où \mathbf{y} est regardé comme un vecteur de longueur n . La matrice $Df(\mathbf{x}_0)$ est appelée matrice jacobienne de \mathbf{f} en \mathbf{x}_0 . Son j -ème élément est la dérivée partielle $D_j f_k(\mathbf{x}_0)$. Ainsi, nous avons :

$$Df(\mathbf{x}_0) = \begin{pmatrix} D_1 f_1(\mathbf{x}_0) & \dots & D_n f_1(\mathbf{x}_0) \\ \dots & \dots & \dots \\ D_1 f_m(\mathbf{x}_0) & \dots & D_n f_m(\mathbf{x}_0) \end{pmatrix}$$

La matrice jacobienne $Df(\mathbf{x}_0)$ est définie en chacun des points où les mn dérivées partielles $D_j f_k(\mathbf{x}_0)$ existent. La dérivée totale $T_{\mathbf{x}_0}$ s'écrit également $\mathbf{f}'(\mathbf{x}_0)$. La dérivée $\mathbf{f}'(\mathbf{x}_0)$ est une transformation linéaire ; le jacobien $Df(\mathbf{x}_0)$ est une représentation matricielle de cette transformation. La formule de Taylor au premier ordre s'écrit :

$$\mathbf{f}(\mathbf{x}_0 + \mathbf{v}) = \mathbf{f}(\mathbf{x}_0) + \mathbf{f}'(\mathbf{x}_0)(\mathbf{v}) + \|\mathbf{v}\|\mathbf{E}(\mathbf{x}_0, \mathbf{v}),$$

où $\mathbf{E}(\mathbf{x}_0, \mathbf{v}) \rightarrow \mathbf{0}$ lorsque $\mathbf{v} \rightarrow \mathbf{0}$. Pour calculer les composantes du vecteur $\mathbf{f}'(\mathbf{x}_0)(\mathbf{v})$, on peut utiliser le produit matriciel $Df(\mathbf{x}_0)\mathbf{v}$ ou encore la formule (3.9).

De manière similaire au cas scalaire, nous avons que la différentiabilité d'un champ vectoriel implique la continuité de ce champ.

Théorème 3.4.3. *Si un champ de vecteur \mathbf{f} est différentiable en \mathbf{x}_0 , alors \mathbf{f} est continue en \mathbf{x}_0 .*

3.4.3 La règle de dérivation en chaîne pour les champs de vecteurs

Théorème 3.4.4. (*Dérivation en chaîne*). Soient $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ et $\mathbf{g} : \mathbb{R}^p \rightarrow \mathbb{R}^n$ des champs de vecteurs tels que la composition $\mathbf{h} = \mathbf{f} \circ \mathbf{g}$ soit définie dans un voisinage d'un point \mathbf{x}_0 . Supposons que \mathbf{g} soit différentiable dans un voisinage d'un point \mathbf{x}_0 , de dérivée totale $\mathbf{g}'(\mathbf{x}_0)$. Soit $\mathbf{y}_0 = \mathbf{g}(\mathbf{x}_0)$ et supposons que \mathbf{f} soit différentiable en \mathbf{y}_0 , de dérivée totale $\mathbf{f}'(\mathbf{y}_0)$. Alors, \mathbf{h} est différentiable en \mathbf{x}_0 , et la dérivée totale $\mathbf{h}'(\mathbf{x}_0)$ est donnée par :

$$\mathbf{h}'(\mathbf{x}_0) = \mathbf{f}'(\mathbf{y}_0) \circ \mathbf{g}'(\mathbf{x}_0).$$

Soit $\mathbf{h} = \mathbf{f} \circ \mathbf{g}$, où \mathbf{g} est différentiable en \mathbf{x}_0 et \mathbf{f} est différentiable en $\mathbf{y}_0 = \mathbf{g}(\mathbf{x}_0)$. La dérivation en chaîne nous donne :

$$\mathbf{h}'(\mathbf{x}_0) = \mathbf{f}'(\mathbf{y}_0) \circ \mathbf{g}'(\mathbf{x}_0).$$

On peut en fait exprimer la règle de dérivation en chaîne grâce aux matrices jacobienes $Dh(\mathbf{x}_0)$, $Df(\mathbf{y}_0)$ et $Dg(\mathbf{x}_0)$ qui représentent les transformations linéaires $\mathbf{h}'(\mathbf{x}_0)$, $\mathbf{f}'(\mathbf{y}_0)$ et $\mathbf{g}'(\mathbf{x}_0)$, respectivement. Puisque la composition d'applications correspond à la multiplication de matrices, nous obtenons :

$$Dh(\mathbf{x}_0) = Df(\mathbf{y}_0)Dg(\mathbf{x}_0), \quad (3.10)$$

où $\mathbf{y}_0 = \mathbf{g}(\mathbf{x}_0)$. On peut également exprimer la règle sous la forme d'un ensemble d'équations scalaires.

Supposons que $\mathbf{x}_0 \in \mathbb{R}^p$, $\mathbf{y}_0 = \mathbf{g}(\mathbf{x}_0) \in \mathbb{R}^n$ et $\mathbf{f}(\mathbf{y}_0) \in \mathbb{R}^m$. Alors, $\mathbf{h}(\mathbf{x}_0) \in \mathbb{R}^m$ et nous pouvons écrire :

$$\mathbf{g} = (g_1, \dots, g_n)$$

$$\mathbf{f} = (f_1, \dots, f_m)$$

$$\mathbf{h} = (h_1, \dots, h_m)$$

Alors, $Dh(\mathbf{x}_0)$ est une matrice $m \times p$, $Dh(\mathbf{y}_0)$ est une matrice $m \times n$ et $Dg(\mathbf{x}_0)$ est une matrice $n \times p$, données respectivement par :

$$Dh(\mathbf{x}_0) = [D_j h_i(\mathbf{x}_0)]_{i,j=1}^{m,p}$$

$$Df(\mathbf{y}_0) = [D_j f_i(\mathbf{y}_0)]_{i,j=1}^{m,n}$$

$$Dg(\mathbf{x}_0) = [D_j g_k(\mathbf{x}_0)]_{k,j=1}^{n,p}$$

L'équation (3.10) est équivalente à équations scalaires :

$$D_j h_i(\mathbf{x}_0) = \sum_{k=1}^n D_k f_i(\mathbf{y}_0) D_j g_k(\mathbf{x}_0), \quad \forall 1 \leq i \leq m, 1 \leq j \leq p.$$

Exemple 3.4.1. (*Règle de dérivation en chaîne développée pour les champs scalaires*). Si $m = 1$, \mathbf{f} est un champ scalaire, \mathbf{h} aussi. Les p équations (une pour chaque dérivée partielle de \mathbf{h}) s'écrivent :

$$D_j h(\mathbf{x}_0) = \sum_{k=1}^n D_k f(\mathbf{y}_0) D_j g_k(\mathbf{x}_0), \quad \forall 1 \leq j \leq p.$$

3.4. QUELQUES RÈGLES ET RÉSULTATS UTILES

Le cas particulier $p = 1$ donne une équation :

$$h'(\mathbf{x}_0) = \sum_{k=1}^n D_k f(\mathbf{y}_0) g'_k(\mathbf{x}_0).$$

Exemple 3.4.2. *Considérons $p = 2$ et $n = 2$. Écrivons $\mathbf{x}_0 = (s, t)$ et $\mathbf{b} = (x, y)$. Alors les composantes de x et y sont reliées à s et t par les équations :*

$$x = g_1(s, t) \text{ et } y = g_2(s, t).$$

La règle de dérivation en chaîne donne un couple d'équations pour les dérivées partielles de \mathbf{h}

$$D_1 \mathbf{h}(s, t) = D_1(\mathbf{f} \circ \mathbf{g}) = D_1 \mathbf{f}(x, y) D_1 g_1(s, t) + D_2 \mathbf{f}(x, y) D_1 g_2(s, t).$$

De même, nous avons :

$$D_2 \mathbf{h}(s, t) = D_1 \mathbf{f}(x, y) D_2 g_1(s, t) + D_2 \mathbf{f}(x, y) D_2 g_2(s, t).$$

Selon la notation ∂ , ce couple d'équations s'écrit :

$$\frac{\partial h}{\partial s} = \frac{\partial f}{\partial x} \frac{\partial g}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial g}{\partial s},$$

$$\frac{\partial h}{\partial t} = \frac{\partial f}{\partial x} \frac{\partial g}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial g}{\partial t}.$$

3.4.4 Conditions suffisantes pour avoir l'égalité des dérivées partielles mixtes

Si f est une fonction à valeurs réelles de deux variables, les deux dérivées mixtes $D_{1,2}f$ et $D_{2,1}f$ ne sont pas nécessairement égales.

Théorème 3.4.5. *(Une conditions suffisantes pour l'égalité des dérivées partielles mixtes). Supposons que f soit un champ scalaire tel que les dérivées partielles $D_1 f$, $D_2 f$, $D_{1,2}f$ et $D_{2,1}f$ existent sur un ouvert \mathcal{S} . Si (x_0, y_0) est un point de \mathcal{S} où $D_{1,2}f$ et $D_{2,1}f$ sont continues, nous avons alors :*

$$D_{1,2}f(x_0, y_0) = D_{2,1}f(x_0, y_0).$$

Théorème 3.4.6. *Soit f un champ scalaire tel que les dérivées partielles $D_1 f$, $D_2 f$ et $D_{2,1}f$ existent sur un ouvert \mathcal{S} contenant (x_0, y_0) . Supposons de plus que $D_{2,1}f$ est continue sur \mathcal{S} . Alors, les dérivées partielles $D_{1,2}(x_0, y_0)$ existent et nous avons :*

$$D_{1,2}f(x_0, y_0) = D_{2,1}f(x_0, y_0).$$

Chapitre 4

Séparation à vaste marge

4.1 Introduction

Les machines à vecteurs de support (ou séparateurs à vaste marge SVM) sont un ensemble de techniques d'apprentissage destinées à résoudre des problèmes de discrimination, c'est à dire décider à quelle classe appartient un échantillon, ou de régression, c'est à dire prédire la valeur numérique d'une variable.

SVM est une méthode de classification particulièrement bien adaptée pour traiter des données de très hautes dimensions telles que les textes, les images et la voix...

4.2 Historique

Les séparateurs à vastes marges reposent sur deux idées clés : la notion de marge maximale et la notion de fonction noyau. Ces deux notions existaient depuis plusieurs années avant qu'elles ne soient mises en commun pour construire les SVM.

L'idée des hyperplans à marge maximale a été explorée dès 1963 par Vladimir Vapnik et A. Lerner, et en 1973 par Richard Duda et Peter Hart dans leur livre *Pattern Classification*. Les fondations théoriques des SVM ont été explorées par Vapnik et ses collègues dans les années 70 avec le développement de la Théorie de Vapnik-Chervonenkis et par Valiant.

L'idée des fonctions noyaux n'est pas non plus nouvelle : le théorème de Mercer date de 1909, et l'utilité des fonctions noyaux dans le contexte de l'apprentissage artificiel a été montrée dès 1964 par Aizermann, Bravermann et Rozoener.

Ce n'est toutefois qu'en 1992 que ces idées seront bien comprises et rassemblées par Boser, Guyon et Vapnik dans un article, qui est l'article fondateur des séparateurs à vaste marge. L'idée des variables ressorts, qui permet de résoudre certaines limitations pratiques importantes, ne sera introduite qu'en 1995. À partir de cette date, qui correspond à la publication du livre de Vapnik, les SVM gagnent en popularité et sont utilisés dans de nombreuses applications.

4.3 Principe général

Les SVM peuvent être utilisés pour résoudre des problèmes de discrimination, c'est à dire décider à quelle classe appartient un échantillon, ou de régression, c'est à dire prédire la valeur numérique d'une variable.

La résolution de ces problèmes passe par la construction d'une fonction f qui à un vecteur d'entrée x fait correspondre une sortie $y : y = f(x)$.

On se limite pour l'instant à un problème de discrimination à deux classes (discrimination binaire), c'est à dire $y \in \{-1, 1\}$, le vecteur d'entrée x étant dans un espace X muni d'un produit scalaire. On peut prendre par exemple $X = \mathbb{R}^n$.

4.4 Séparateur linéaire

4.4.1 Hyperplan

Plaçons-nous dans le cas d'une classification binaire (i.e. les exemples à classifier réparties en 2 classes). On appelle hyperplan séparateur un hyperplan qui sépare les deux classes figure 4.1, en particulier il sépare leurs points d'apprentissage. Comme il n'est en générale pas possible d'en trouver un, on se contentera donc de chercher un hyperplan discriminant qui est une approximation au sens d'un critère à fixer (maximiser la distance entre ces deux classes).

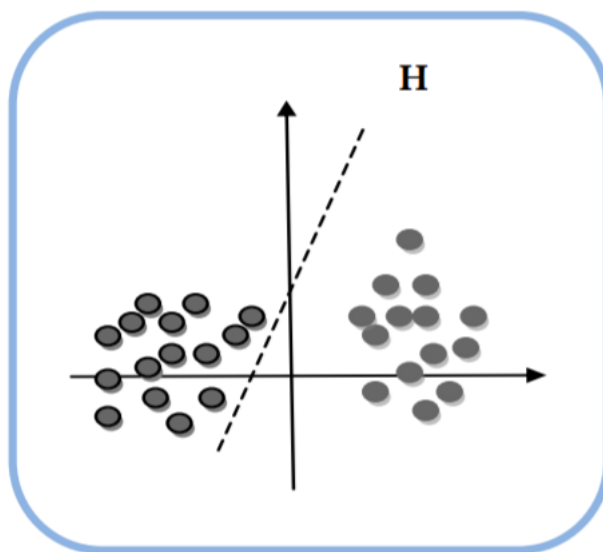


FIGURE 4.1 – L'hyperplan H qui sépare les deux ensembles de points.

4.4.2 Les exemples

Nous noterons x un vecteur de \mathbb{R}^n , dans notre cas, nous dirons que l'ensemble des vecteurs étiquetés que nous nous donnons est l'ensemble des exemples noté X , qui contient p éléments. $X = \{(x_i, y_i)\}_{1 \leq i \leq p}$ avec $\forall i, y_i \in \{-1, 1\}$.

L'appartenance d'un vecteur à une classe ou à l'autre est matérialisée ici par la valeur -1 ou 1 de l'étiquette y , ce qui nous arrangera pour les calculs.

4.4.3 Le séparateur linéaire

Nous noterons le produit scalaire de deux vecteurs $\langle w, x \rangle$. Cette notation posée, nous pouvons définir le séparateur linéaire $f_{w,b}$ par l'équation suivante :

$$f_{w,b}(x) = \langle w, x \rangle + b.$$

Ce séparateur ne fournit pas des valeurs valant exclusivement -1 ou 1 , mais nous dirons que quand le résultat $f_{w,b}(x)$ est positif, le vecteur x appartient à la même classe que les exemples d'étiquette 1 , et que quand ce résultat est négatif, le vecteur x appartient à la même classe que les exemples d'étiquette -1 .

Notons avant de discuter plus avant cette notation de séparateur linéaire que l'équation $f_{w,b}(x) = 0$ définit la frontière de séparation entre les deux classes, et que cette frontière est un hyperplan affine dans le cas du séparateur linéaire.

4.5 Séparabilité

Reprenons notre base d'exemples X , et décomposons-la en deux sous-ensembles selon la valeur de l'étiquette y . Définissons donc :

$$X^+ = \{x : (x, y) \in X \text{ et } y = 1\}.$$

$$X^- = \{x : (x, y) \in X \text{ et } y = -1\}.$$

Dire que X est linéairement séparable signifie qu'il existe w et b tel que :

$$f_{w,b}(x) > 0 \quad \forall x \in X^+.$$

$$f_{w,b}(x) < 0 \quad \forall x \in X^-.$$

Ce n'est pas toujours faisable, il peut y avoir des distributions d'étiquettes sur les vecteurs de X qui rendent X non linéairement séparable. Dans le cas d'exemples pris dans le plan, dire que la distribution d'exemples est linéairement séparable signifie qu'on peut tracer un trait (un hyperplan) tel que les exemples de la classe 1 et ceux de la classe -1 se retrouvent de part et d'autre de cette frontière.

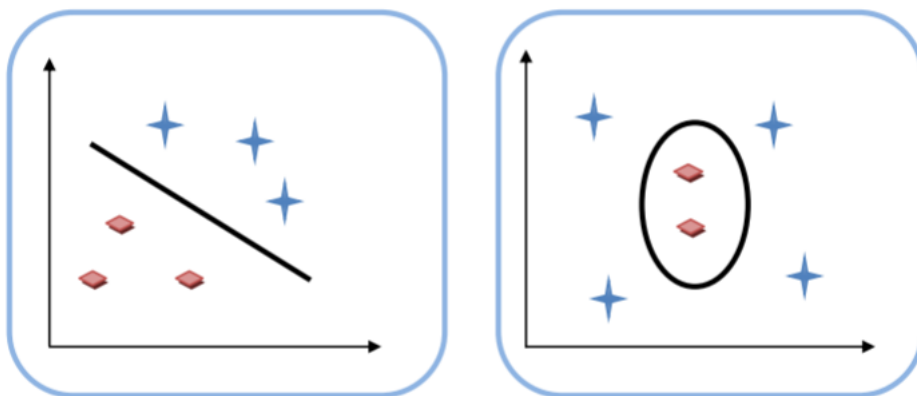


FIGURE 4.2 – À gauche cas linéairement séparable, à droite non linéairement séparable.

4.5.1 Vecteurs supports

Pour une tâche de détermination de l'hyperplan séparable des SVM est d'utiliser seulement les points les plus proches (i.e. les points de la frontière entre les deux classes des données) parmi l'ensemble total d'apprentissage, ces points sont appelés vecteurs de support figure 4.3.

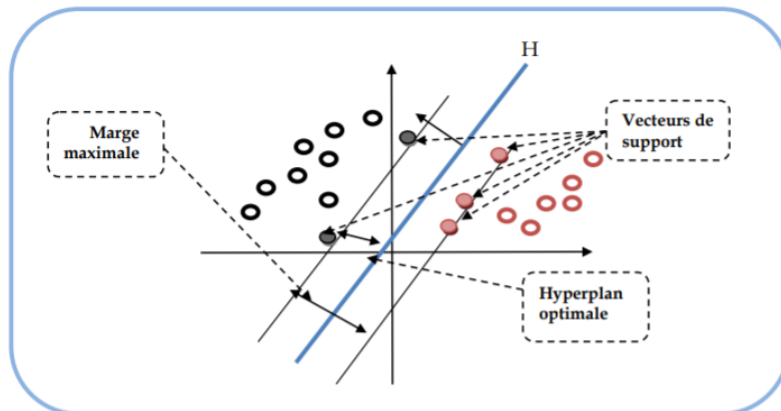


FIGURE 4.3 – Hyperplan optimal, vecteurs de support et marge maximale.

4.5.2 Marge

il existe une infinité d'hyperplans capable de séparer parfaitement les deux classes d'exemples. Le principe des SVM est de choisir celui qui va maximiser la distance minimale entre l'hyperplan et les exemples d'apprentissage (i.e. la distance entre l'hyperplan et les vecteurs de support), cette distance est appelée la marge (figure 4.3).

- Définition 4.5.1.**
- 1) On définit la marge d'un séparateur f pour un point (x, y) par $\gamma_{(x,y)}^f = y \cdot f(x)$ ce qui est proportionnel à la distance de (x, y) à la séparation.
 - 2) Pour la base d'exemples la marge est $\gamma_X^f = \min_{(x,y) \in X} \gamma_{(x,y)}^f$.

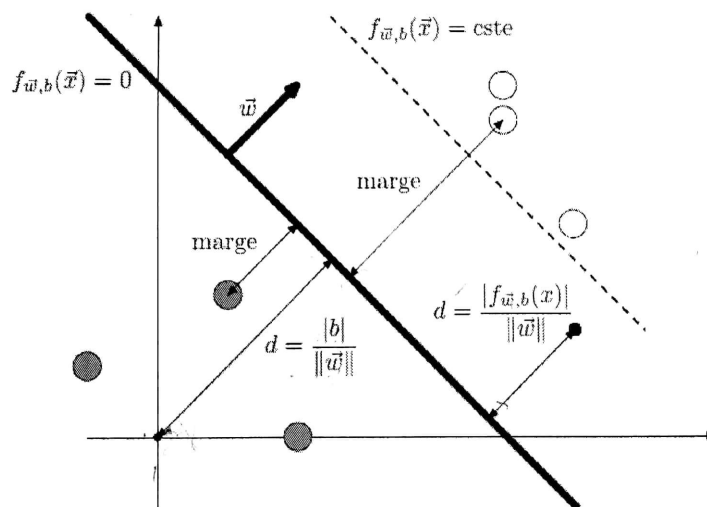


FIGURE 4.4 – Définition d'un séparateur $f_{w,b}$. Les grandeurs mentionnées sont les valeurs du séparateurs au niveau des points, non les distances euclidiennes.

4.5.3 Hyperplans canoniques

Dans le cadre de classificateurs à marge maximale, l'hyperplan séparateur correspond à la médiatrice du plus petit segment de droite reliant les enveloppes convexes des deux catégories.

Notons que l'on suppose aussi que l'ensemble d'apprentissage est linéairement séparable. Dès lors, on peut définir deux plans se trouvant de part et d'autre de l'hyperplan et parallèles à celui-ci, sur lesquels reposent les exemples les plus proches. La figure 4.5 illustre cette situation. Il est possible que différentes équations correspondent au même plan géométrique.

Il est donc possible de redimensionner w et b de telle sorte que les deux plans parallèles aient respectivement pour l'équation : $\langle w, x \rangle + b = 1$ et $\langle w, x \rangle + b = -1$.

Les deux hyperplans sont appelés hyperplans canoniques. Notons que la marge des hyperplans canoniques est $\frac{1}{\|w\|}$.

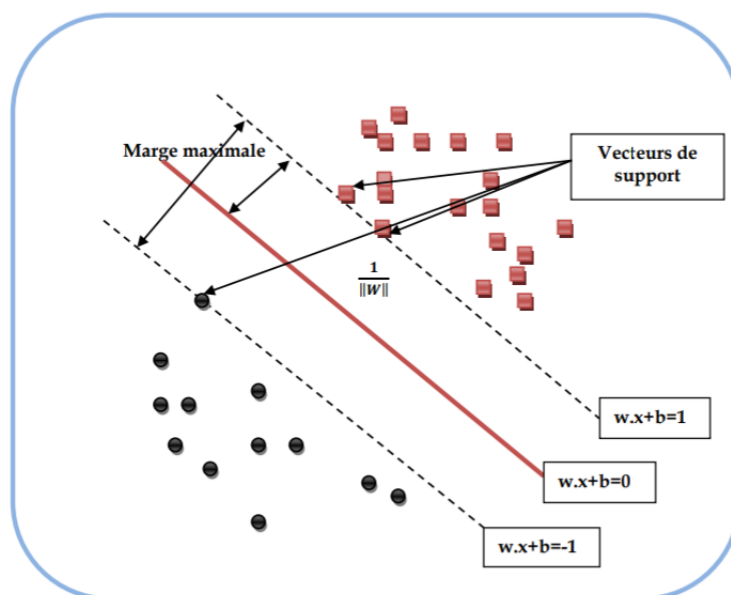


FIGURE 4.5 – Exemple graphique des données linéairement séparables.

4.6 Le problème d'optimisation à résoudre par SVM pour le cas linéairement séparable

Un classificateur est dit linéaire lorsqu'il est possible d'exprimer sa fonction de décision par une fonction linéaire en x . Dans la suite, nous supposons que les exemples nous sont fournis dans le format vectoriel. Notre espace d'entrée x correspond donc à \mathbb{R}^n où n est le nombre de composantes des vecteurs contenant les données :

$$f(x) = \langle w, x \rangle + b = \sum_{j=1}^n w_j x_j + b.$$

4.6. LE PROBLÈME D'OPTIMISATION À RÉSOUDRE PAR SVM POUR LE CAS LINÉAIREMENT SÉPARABLE

Si les données sont linéairement séparables, alors il existe un hyperplan d'équation :

$$\begin{aligned} &\langle w, x \rangle + b = 0 \text{ et tel que} \\ &\langle w, x \rangle + b \geq 1 \text{ si } y_i = 1. \\ &\langle w, x \rangle + b \leq -1 \text{ si } y_i = -1. \end{aligned}$$

On peut combiner ces deux inéquations en une seule $y_i(\langle w, x \rangle + b) \geq 1$. Nous n'allons plus nous contenter d'en trouver un hyperplan séparateur, mais nous allons en plus chercher parmi ceux-ci l'hyperplan optimal.

L'hyperplan optimal est celui qui maximise la marge γ_X^f .

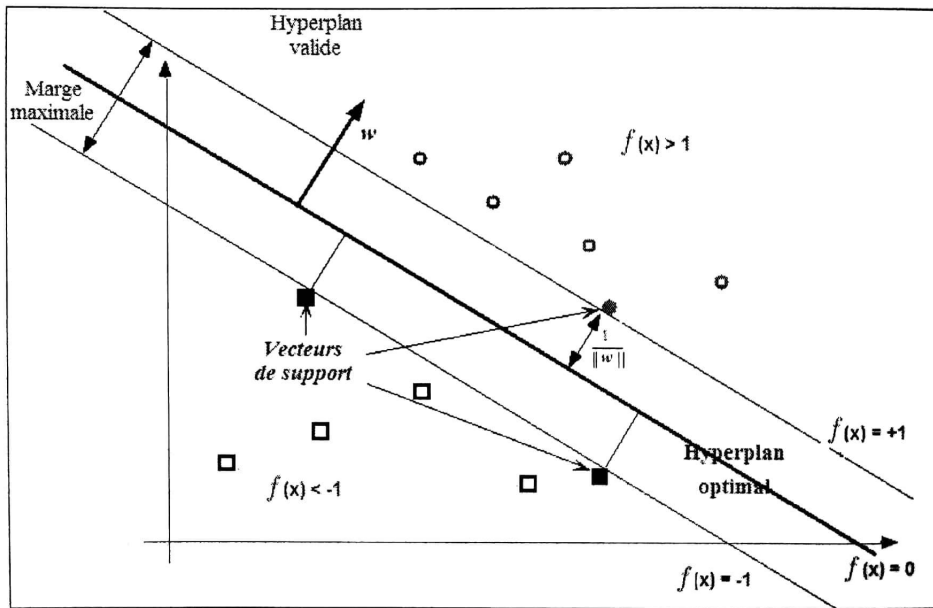


FIGURE 4.6 – L'hyperplan optimal séparant les points de deux classes est celui qui passe "au milieu" de ces classes.

Maximiser la marge γ_X^f est équivalent à maximiser la somme des distances euclidiennes des deux classes par rapport à l'hyperplan. Ainsi, la marge a l'expression mathématique suivante :

$$\begin{aligned} \gamma_X^f &= \min_{(x_i, y_i = -1)} \frac{|f_{(w,b)}(x_i)|}{\|w\|} + \min_{(x_i, y_i = 1)} \frac{|f_{(w,b)}(x_i)|}{\|w\|} \\ &= \frac{1}{\|w\|} \left[\min_{(x_i, y_i = -1)} |\langle w, x_i \rangle + b| + \min_{(x_i, y_i = 1)} |\langle w, x_i \rangle + b| \right] \\ &= \frac{1}{\|w\|} [1 + 1] \\ &= \frac{2}{\|w\|}. \end{aligned}$$

Maintenant que nous avons défini les notions de marge et d'hyperplans canoniques, nous pouvons formuler un problème d'optimisation mathématique tel que sa solution nous fournisse l'hyperplan optimal (maximisant la marge) :

$$(QP1) \begin{cases} \min \frac{1}{2} \|w\|^2 \\ y_i(\langle w, x_i \rangle + b) \geq 1 \quad \forall (x_i, y_i) \in X. \end{cases}$$

Il s'agit d'un problème quadratique dont la fonction objective est à minimiser. Cette fonction objective est le carré de l'inverse de la double marge. L'unique contrainte stipule que les exemples doivent être bien classés et qu'ils ne dépassent pas les hyperplans canoniques.

Dans cette formulation, les variables à fixer sont les composantes w_i et b . Le vecteur w possède un nombre de composantes égal à la dimension de l'espace d'entrée. En gardant cette formulation telle quelle, nous souffrons des mêmes problèmes que les méthodes classiques.

Pour éviter cela, il est nécessaire d'introduire une formulation dite duale du problème.

Un problème dual est un problème fournissant la même solution que le primal mais dont la formulation est différente.

Pour dualiser $QP1$, nous devons former ce que l'on appelle le Lagrangien. Il s'agit de faire rentrer les contraintes dans la fonction objective et de pondérer chacune d'entre elles par une variable duale : $L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i [y_i(\langle w, x_i \rangle + b) - 1]$.

Les variables duales α_i intervenant dans le Lagrangien sont appelées multiplicateurs de Lagrange. Elles représentent la "force" avec laquelle la solution appuie sur la contrainte i .

Ainsi, un hyperplan qui violerait la contrainte pour x_i (il classe cet exemple du mauvais côté) rendrait α_i très grand ce qui ferait fortement augmenter la fonction objective (L). Cette solution ne pourra donc être retenue comme solution optimale. Notons que L doit être minimisé par rapport aux variables primales et maximisé par rapport aux variables duales.

A présent, nous introduisons les conditions Karush Kuhn et Tucker (KKT) statuant sur l'optimalité d'une solution.

Théorème 4.6.1. (KKT) pour les problèmes différentiables convexes. considérons un problème d'optimisation de la forme :

Minimiser $g(x)$

$$\text{tel que } \begin{cases} c_i(x) \leq 0 \quad \forall i = 1, \dots, n \\ e_j(x) = 0 \quad \forall j = 1, \dots, n' \end{cases}$$

avec $g(\cdot)$, $c_i(\cdot)$, $e_j(\cdot)$ convexes et différentiables.

Le Lagrangien est formé comme suit :

$$L(x, \alpha, \beta) = g(x) + \sum_{i=1}^n \alpha_i c_i(x) + \sum_{j=1}^{n'} \beta_j e_j(x).$$

La solution \bar{x} est optimale si est seulement s'il existe $\bar{\alpha} \in \mathbb{R}^n$ avec $\bar{\alpha}_i \geq 0 \quad \forall i = 1, \dots, n$ et $\bar{\beta} \in \mathbb{R}^{n'}$ tels que :

$$\partial_x L(\bar{x}, \bar{\alpha}, \bar{\beta}) = \partial_x g(\bar{x}) + \sum_{i=1}^n \bar{\alpha}_i \partial_x c_i(\bar{x}) + \sum_{j=1}^{n'} \bar{\beta}_j \partial_x e_j(\bar{x}) = 0.$$

$$\begin{aligned}\partial_{\alpha_i} L(\bar{x}, \bar{\alpha}, \bar{\beta}) &= c_i(\bar{x}) \leq 0 \\ \partial_{\beta_j} L(\bar{x}, \bar{\alpha}, \bar{\beta}) &= e_j(\bar{x}) = 0 \\ \bar{\alpha}_i c_i(\bar{x}) &= 0 \quad \forall i = 1, \dots, n\end{aligned}$$

Ce théorème fondamental en optimisation mathématique, nous fournit une condition suffisante et nécessaire pour l'optimalité d'une solution dans le cadre de problèmes différentiables convexes (ce qui est notre cas).

La dernière condition est souvent appelée conditions KKT complémentaire.

Ces conditions expriment deux choses. Pour le voir prenons la première :

- 1) $\alpha_i = 0$: Dans ce cas la solution n'est pas "sur la contrainte". Il n'y a donc rien à imposer au niveau de la solution.
- 2) $\alpha_i \neq 0$: La solution est "sur la contrainte".

Dans ce cas, la nullité du produit impose que la solution ne dépasse pas la contrainte (elle reste faisable).

Déterminons à présent les conditions KKT de notre problème d'optimisation *QP1*.

$$\begin{cases} \partial_w L(w, b, \alpha) = w - \sum_{i=1}^n \alpha_i y_i x_i = 0 & (1) \\ \partial_b L(w, b, \alpha) = \sum_{i=1}^n \alpha_i y_i = 0 & (2) \\ \partial_{\alpha_i} L(w, b, \alpha) = -y_i (< w, x_i > + b) + 1 \leq 0 & (3) \\ \alpha_i (y_i (< w, x_i > + b) - 1) = 0 \quad \forall i = 1, \dots, n & (4) \end{cases}$$

L'équation (1), permet de réexprimer w : $w = \sum_{i=1}^n \alpha_i y_i x_i$ (5)

Remarquons qu'avec cette formulation, on peut calculer w en fixant seulement n paramètres.

L'idée va donc être de formuler un problème dans lequel w est remplacé par sa formulation (5).

De cette façon, le nombre de paramètre à fixer est relatif au nombre d'exemples du training set et non plus à la dimension de l'espace d'entrée (supposé très élevée).

Pour ce faire, nous substituons (2) et (5) dans le Lagrangien :

$$\begin{aligned}L(w, b, \alpha) &= \frac{1}{2} \langle w, w \rangle - \sum_{i=1}^n \alpha_i [y_i (< w, x_i > + b) - 1], \\ &= \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle - \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle + \sum_{i=1}^n \alpha_i, \\ &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle.\end{aligned}$$

A partir de quoi nous pouvons formuler le problème dual :

$$QP2 \text{ Maximiser } W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle.$$

$$\text{telque } \begin{cases} \sum_{i=1}^n \alpha_i y_i = 0 \text{ (stationnarité).} \\ \alpha_i \geq 0 \quad \forall i = 1, \dots, n \text{ (admissibilité duale).} \end{cases}$$

La résolution du dual permet donc de calculer le vecteur w à moindre coût, cependant cette formulation ne fait à aucun moment apparaître le terme b . Pour calculer ce dernier nous devons utiliser les variables primales :

$$b = -\frac{\max_{y_i=-1}(\langle w, x_i \rangle) + \min_{y_i=+1}(\langle w, x_i \rangle)}{2}.$$

Nous avons à présent tous les éléments nécessaires pour exprimer la fonction de décision de notre classificateur linéaire :

$$f(x) = \sum_{i=1}^n \alpha_i y_i \langle x, x_i \rangle + b.$$

Notons qu'un grand nombre de termes de cette somme sont nuls. En effet seuls les α_i correspondants aux exemples se trouvant sur les hyperplans canoniques (sur la contrainte) sont non nuls.

Ces exemples sont appelés vecteurs supports (SV). On peut les voir comme les représentants de leurs catégories car si l'ensemble d'apprentissage n'était constitué que des SV, l'hyperplan optimal que l'on trouverait serait identique. Les vecteurs qui ne sont pas de support ($\alpha_i = 0$) n'ont aucune influence dans la solution.

La fonction de décision pour la classification de vecteurs inconnus x est donnée par :

$$f(x) = \text{sign}\left(\sum_{i=1}^m \alpha_i y_i \langle x, x_i \rangle + b\right).$$

où m est le nombre de vecteurs supports.

4.7 Les avantages et les inconvénients des SVM

Avantages :

SVM est une méthode de classification intéressante car le champ de ses applications est large, parmi ses avantages nous avons :

- Très efficaces en dimension élevée.
- Pour la décision, n'utilisent pas tous les échantillons d'apprentissage, mais seulement une partie (les vecteurs de support). En conséquence, ces algorithmes demandent moins de mémoire.
- Décision rapide. La classification d'un nouvel exemple consiste à voir le signe de la fonction de décision.
- La résolution du problème est convertie en résolution d'un problème quadratique convexe dont la solution est unique et donnée par des méthodes mathématiques classiques de programmation quadratique.

Inconvénients :

L'inconvénient majeur du classificateur SVM est qu'il est désigné ou conçu pour la classification binaire (la séparation entre deux classes une $+1$ et l'autre -1).

Chapitre 5

Algorithme d'optimisation du problème

5.1 L'algorithme proposé

Soient \bar{x}_+ , \bar{x}_- deux éléments de X^+ , X^- respectivement ; tels que $\|\bar{x}_+ - \bar{x}_-\| = \inf_{\substack{x_- \in X^- \\ x_+ \in X^+}} \|x_+ - x_-\|$.

Etudions le cas où $\overset{\circ}{X}^+ \neq \emptyset$, $\overset{\circ}{X}^- \neq \emptyset$. Posons :

$$(p) : \begin{cases} \max\{-\|w\|^2\} \\ wx_+ + b \geq \frac{1}{2} \\ wx_- + b \leq -\frac{1}{2} \\ x^+ \in X^+, x_- \in X^- \end{cases}$$

et soit \bar{w} , \bar{b} la solution de ce problème. On a :

$$i) \{w \in \mathbb{R}^n, b \in \mathbb{R} : wx_+ + b \geq \frac{1}{2}, x_+ \in X^+, wx_- + b \leq -\frac{1}{2}, x_- \in X^-\}$$

$$\subset \{w \in \mathbb{R}^n : w(x_+ - \bar{x}_-) \geq 1, x_+ \in X^+, w(x_- - \bar{x}_-) \leq 0, x_- \in X^-\}.$$

En effet,

$$\forall (w, b) \in \mathbb{R}^{n+1} : \begin{cases} wx_+ + b \geq \frac{1}{2} \\ wx_- + b \leq -\frac{1}{2} \end{cases} \Rightarrow \begin{cases} wx_+ + b \geq \frac{1}{2} \\ -wx_- - b \geq \frac{1}{2} \end{cases}$$

en particulier pour $wx_- + b \leq -\frac{1}{2} \Rightarrow w\bar{x}_- + b = -\frac{1}{2} \leq -\frac{1}{2}$
alors on a : $wx_+ + b - wx_- - b \geq \frac{1}{2} + \frac{1}{2} \Rightarrow w(x_+ - \bar{x}_-) \geq 1, \forall x_+ \in X^+$.

Considérons alors le problème (p_+) suivant :

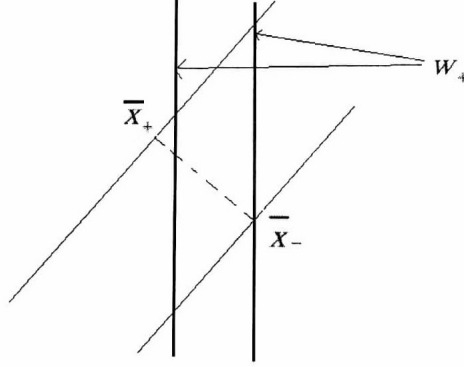
$$(p_+) \begin{cases} \max\{-\|w\|^2\} \\ wx_+ + b \geq \frac{1}{2}, \forall x^+ \in X^+ \\ wx_- + b = -\frac{1}{2} \\ w(x_- - \bar{x}_-) \leq 0, \forall x_- \in X^- \end{cases}$$

5.1. L'ALGORITHME PROPOSÉ

C'est la séparation à vaste marge de X^+ et $\{\bar{x}_-\}$, comme $b = -\frac{1}{2} - w\bar{x}_-$, (p_+) devient :

$$(p_+) \begin{cases} \max\{-\|w\|^2\} \\ w(x_+ - \bar{x}_-) \geq 1, \forall x_+ \in X^+ \\ w(x_- - \bar{x}_-) \leq 0, \forall x_- \in X^- \end{cases}$$

Soit w_+ la solution de (p_+)



$$\begin{aligned} & ii) \{(w, b) \in \mathbb{R}^{n+1} : wx_+ + b \geq \frac{1}{2}, x_+ \in X^+, wx_- + b \leq -\frac{1}{2}, x_- \in X^-\} \\ & \subset \{w \in \mathbb{R}^n : w(x_- - \bar{x}_+) \geq -1, \forall x_- \in X^-, w(\bar{x}_+ - x_+) \leq 0, \forall x_+ \in X^+\}. \end{aligned}$$

Considérons le problème suivant :

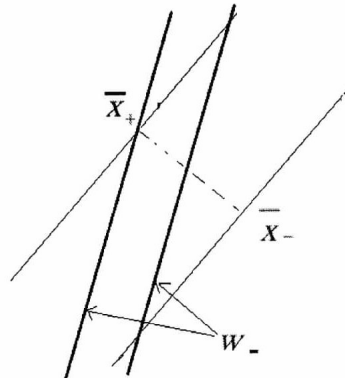
$$(p_-) \begin{cases} \max\{-\|w\|^2\} \\ w\bar{x}_+ + b = \frac{1}{2} \\ wx_- + b \leq -\frac{1}{2}, \forall x_- \in X^- \\ w(\bar{x}_+ - x_+) \leq 0, \forall x_+ \in X^+ \end{cases}$$

C'est la séparation à vaste marge de X_- et $\{\bar{x}_+\}$.

Comme $b = \frac{1}{2} - w\bar{x}_+$,

$$(p_-) \begin{cases} \max\{-\|w\|^2\} \\ w(x_- - \bar{x}_+) \leq -1, \forall x_- \in X^- \\ w(\bar{x}_+ - x_+) \leq 0, \forall x_+ \in X^+ \end{cases}$$

Soit w_- la solution de (p_-) .



5.1. L'ALGORITHME PROPOSÉ

Comme \bar{w} est une solution de (p_+) et (p_-) , on a :

$$\frac{1}{\|w_+\|} \geq \frac{1}{\|\bar{w}\|} \text{ et } \frac{1}{\|w_-\|} \geq \frac{1}{\|\bar{w}\|}, \text{ car } \bar{w} \text{ n'est pas optimal pour } (p_+) \text{ et } (p_-).$$

D'après (i) et (ii), on a :

$$iii) \{(w, b) \in \mathbb{R}^{n+1} : wx_+ + b \geq \frac{1}{2}, \forall x_+ \in X^+, wx_- + b \leq -\frac{1}{2}, \forall x_- \in X^-\} \subset D.$$

où :

$$D = \{w \in \mathbb{R}^n : w(x_+ - \bar{x}_-) \geq 1, \forall x_+ \in X^+\} \cap \{w \in \mathbb{R}^n : w(x_- - \bar{x}_+) \leq -1, \forall x_- \in X^-\}$$

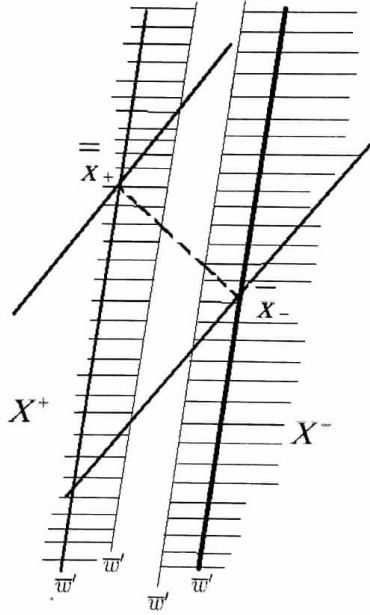
. Considérons alors le problème (p') suivant :

$$(p') \begin{cases} \max\{-\|w\|^2\} \\ w(x_+ - \bar{x}_-) \geq 1, \forall x_+ \in X^+ \\ w(x_- - \bar{x}_+) \leq -1, \forall x_- \in X^- \end{cases}$$

Soit \bar{w}' la solution de (p') . D'après (iii), on a $\frac{1}{\|\bar{w}'\|} \geq \frac{1}{\|\bar{w}\|}$.

Comme la solution de (p) est unique (projection de 0 sur les hyperplans) alors $\|\bar{w}'\| = \|\bar{w}\|$.

Il reste à déterminer \bar{b} .



Calcul de \bar{b}

On sait que la projection de zéro sur (D) est un point frontière D car $0 \notin D$. On a alors :

i $\exists \bar{x}_+ \in X^+ : \bar{w}(\bar{x}_+ - \bar{x}_-) = 1$, si non on a :

ii $\exists \bar{x}_- \in X^- : \bar{w}(\bar{x}_- - \bar{x}_+) = -1$. Car \bar{w} est un point frontière se D .

On i et ii ou même. \bar{x}_+ (ou bien \bar{x}_-) est un vecteur support,

et $\bar{w}\bar{x}_+ + b = \frac{1}{2} \Rightarrow b = \frac{1}{2} - \bar{w}\bar{x}_+$ (ou bien $b = -\frac{1}{2} - \bar{w}\bar{x}_-$).

La solution de (p') est par la technique de projection, on peut même utiliser les algorithmes connus (par passage au dual,...)

5.1. L'ALGORITHME PROPOSÉ

Notes que la solution de (p_+) et (p_-) permet de déterminer les vecteurs supports ; si $\overset{\circ}{X}^+ = \overset{\circ}{X}^- = \emptyset$; la marge est magnée par $\|\bar{x}_+ - \bar{x}_-\|$.
Le problème (p') ne contient plus la constante b .

$$\overset{\circ}{X}^+ = \overset{\circ}{X}^- = \emptyset \Rightarrow w_+ = w_- = (\bar{x}_+ - \bar{x}_-).$$

Soit $\|\bar{x}_+ - \bar{x}_-\| = \inf_{\substack{x_- \in X^- \\ x_+ \in X^+}} \|x_+ - x_-\|$; s'il existe plusieurs \bar{x}_+ et \bar{x}_- ; il suffit de choisir \bar{x}_+ et \bar{x}_- de manière quelconque.

Considérons le problème suivant :

$$(p_1) \begin{cases} \max\{-\|w\|^2\} \\ w(x_+ - \bar{x}_-) \geq 1, x_+ \in X^+ \end{cases}$$

C'est a dire, on veut séparer X^+ et $\{\bar{x}_-\}$; en effet,

$$w\bar{x}_- + b = -\frac{1}{2} \Rightarrow b = -\frac{1}{2} - w\bar{x}_-, \text{ et}$$

$$wx_+ + b \geq \frac{1}{2} \Rightarrow w(x_+ - \bar{x}_-) \geq 1, x_+ \in X^+$$

Notons par w_- la solution de ce problème, et si \bar{w} est la solution du problème

$$(p_2) : \begin{cases} \max\{-\|w\|^2\} \\ wx_+ + b \geq \frac{1}{2}, x_+ \in X^+ \\ wx_- + b \leq -\frac{1}{2}, x_- \in X^- \end{cases}$$

Comme w_- est la solution de (p_1) . Alors $\frac{1}{\|w_-\|} \geq \frac{1}{\|\bar{w}\|}$. Si $w_-x_- + (-\frac{1}{2} - w_-\bar{x}_-) \leq -\frac{1}{2}$, $\forall x_- \in X^-$. Alors w_- est la solution de (p_2) , et $\|w_-\| = \|\bar{w}\|$.

Sinon, il existe $\bar{x}_+ \in X^+ : w_-(\bar{x}_+ - \bar{x}_-) = 1$.

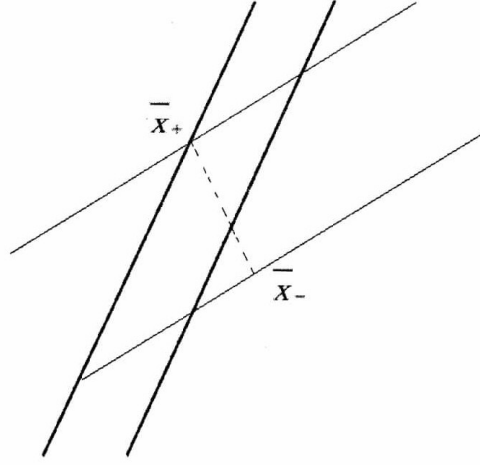
On considère le problème suivant :

$$(p_3) \begin{cases} \max\{-\|w\|^2\} \\ w(x_- - \bar{x}_+) \leq -1, x_- \in X^- \end{cases}$$

C'est à dire, on veut séparer X^- et $\{\bar{x}_+\}$; en effet,

$$w\bar{x}_+ + b = \frac{1}{2} \Rightarrow b = \frac{1}{2} - w\bar{x}_+, \text{ et}$$

$$\begin{aligned} wx_- + b &\leq -\frac{1}{2} \Rightarrow wx_- + \frac{1}{2} - w\bar{x}_+ \leq -\frac{1}{2} \\ &\Rightarrow w(x_- - \bar{x}_+) \leq -1. \end{aligned}$$



Notons par w_+ la solution de ce problème. Alors $\frac{1}{\|w_+\|} \geq \frac{1}{\|\bar{w}\|}$.

Si $w_+x_+ + (\frac{1}{2} - w_+\bar{x}_+) \geq \frac{1}{2}$ c'est à dire $w_+x_+ - w_+\bar{x}_+ \geq 0, \forall x_+ \in X^+$, alors w_+ est solution de (p_2) , et $\|w_+\| = \|\bar{w}\|$.

Sinon, il existe $\bar{x}_- \in X^- : w_+(\bar{x}_- - \bar{x}_+) = -1$.

Car \bar{w} est solution du (p_1) et (p_3) et dans la marge est pour (p_2) , (p_2) a plus le problème :

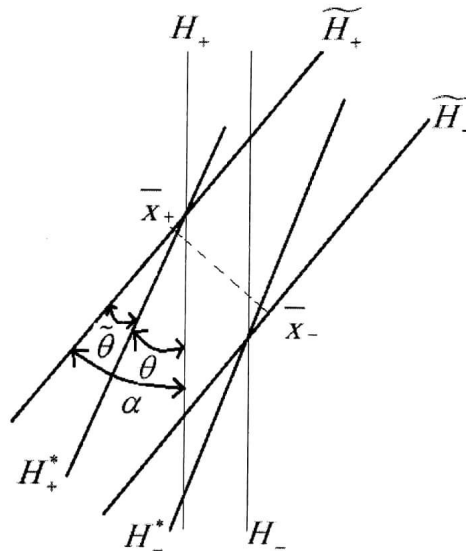
$$(p_0) \begin{cases} \max\{-\|w\|^2\} \\ w(x_+ - \bar{x}_-) \geq 1, \forall x_+ \in X^+ \\ w(x_- - \bar{x}_+) \leq -1, \forall x_- \in X^- \end{cases}$$

C'est la séparation de X^+ et $\{\bar{x}_-\}$ et X^- et $\{\bar{x}_+\}$, une solution de ce problème est \bar{w} .

On note par w_0 la solution de (p_0) , alors $\frac{1}{\|w_0\|} \geq \frac{1}{\|\bar{w}\|}$.

Soit α l'angle fermé entre H_+ et \widetilde{H}_+ ; $0 < \alpha < \frac{\pi}{2}$.

Posons $\tilde{\theta}$ l'angle fermé entre \widetilde{H}_+ et H_+^* , et θ l'angle fermé entre H_+^* , et H_+ ; $\tilde{\theta} + \theta = \alpha$.



Montrons qu'il existe $\lambda_1 \geq 0, \lambda_2 \geq 0$ tels que si H_+^* a pour équation $w^*x + b = 0$, alors :

$$w^* = \lambda_1 w_+ + \lambda_2 \bar{w}.$$

$$\langle w^*, w_+ \rangle = \langle \lambda_1 w_+ + \lambda_2 w, w_+ \rangle = \|w^*\| \|w_+\| \cos \tilde{\theta}.$$

Car l'angle fermé entre deux plans est l'angle fermé entre les deux normales aux plans.

$$\begin{aligned} \|w^*\| \|w_+\| \cos \tilde{\theta} &= \lambda_1 \|w_+\|^2 + \lambda_2 \|w\| \|w_+\| \cos \alpha \\ \Rightarrow \lambda_1 \|w_+\| + \lambda_2 \|w\| \cos \alpha &= \|w^*\| \cos \tilde{\theta} \quad (1). \end{aligned}$$

$$\begin{aligned} \langle w^*, w \rangle &= \langle \lambda_1 w_+ + \lambda_2 w, w \rangle = \|w^*\| \|w\| \cos \theta \\ \|w^*\| \|w\| \cos \theta &= \lambda_2 \|w_+\| \|w\| \cos \alpha + \lambda_2 \|w\|^2 \\ \Rightarrow \lambda_1 \|w_+\| \cos \alpha + \lambda_2 \|w\| &= \|w^*\| \cos \theta \quad (2). \end{aligned}$$

D'où le système :

$$\begin{cases} \lambda_1 \|w_+\| + \lambda_2 \|w\| \cos \alpha = \|w^*\| \cos \tilde{\theta} \\ \lambda_1 \|w_+\| \cos \alpha + \lambda_2 \|w\| = \|w^*\| \cos \theta \end{cases}$$

$$\begin{aligned} \Delta &= \begin{vmatrix} \|w_+\| & \|w\| \cos \alpha \\ \|w_+\| \cos \alpha & \|w\| \end{vmatrix} = \|w_+\| \|w\| \begin{vmatrix} 1 & \cos \alpha \\ \cos \alpha & 1 \end{vmatrix} \\ \Rightarrow \Delta &= \|w_+\| \|w\| (1 - \cos^2 \alpha); \alpha \neq 0 \Rightarrow \Delta > 0. \end{aligned}$$

$$\Delta_{\lambda_1} = \begin{vmatrix} \|w^*\| \cos \tilde{\theta} & \|w\| \cos \alpha \\ \|w^*\| \cos \theta & \|w\| \end{vmatrix} = \|w^*\| \|w\| (\cos \tilde{\theta} - \cos \alpha \cos \theta)$$

Mais

$$\tilde{\theta} = \alpha - \theta \Rightarrow \cos \tilde{\theta} = \cos(\alpha - \theta) = \cos \alpha \cos \theta + \sin \alpha \sin \theta.$$

D'où

$$\begin{aligned} \Delta_{\lambda_1} &= \|w^*\| \|w\| \sin \alpha \sin \theta; \quad 0 \leq \theta < \frac{\pi}{2}, \quad 0 < \alpha < \frac{\pi}{2}. \\ \Rightarrow \Delta_{\lambda_1} &\geq 0. \text{ D'où } \lambda_1 = \frac{\Delta_{\lambda_1}}{\Delta} \geq 0. \end{aligned}$$

De même

$$\Delta_{\lambda_2} = \begin{vmatrix} \|w_+\| & \|w^*\| \cos \tilde{\theta} \\ \|w_+\| \cos \alpha & \|w^*\| \cos \theta \end{vmatrix} = \|w_+\| \|w^*\| (\cos \theta - \cos \tilde{\theta} \cos \alpha)$$

Mais

$$\theta = \alpha - \tilde{\theta} \Rightarrow \cos \theta = \cos(\alpha - \tilde{\theta}) = \cos \alpha \cos \tilde{\theta} + \sin \alpha \sin \tilde{\theta}.$$

$$\begin{aligned} \Delta_{\lambda_2} &= \|w_+\| \|w^*\| \sin \alpha \sin \tilde{\theta}; \quad 0 \leq \tilde{\theta} < \frac{\pi}{2}, \quad 0 < \alpha < \frac{\pi}{2}. \\ \Rightarrow \Delta_{\lambda_2} &\geq 0. \text{ D'où } \lambda_2 = \frac{\Delta_{\lambda_2}}{\Delta} \geq 0. \end{aligned}$$

$$\Delta > 0 \Rightarrow \exists! \lambda_1, \exists! \lambda_2 : w^* = \lambda_1 w_+ + \lambda_2 w, \quad \lambda_1 \geq 0 \text{ et } \lambda_2 \geq 0.$$

pour chaque valeur de $\tilde{\theta}$ et θ . Notons que si

$$w_+ x_+ + b_+ \geq \frac{1}{2}, \quad \forall x_+ \in X^+ / \widetilde{X^+}$$

$$w_+x_+ + b \geq \frac{1}{2}, \forall x_+ \in X^+/\widetilde{X}^+$$

Alors

$$\exists b^* \in \mathbb{R} : w^*x_+ + b^* \geq \frac{1}{2}, \forall x_+ \in X^+/\widetilde{X}^+.$$

En effet

$$\left. \begin{array}{l} \lambda_1 > 0 \Rightarrow \lambda_1 w_+ x_+ + \lambda_1 b_+ \geq \frac{1}{2} \lambda_1 \\ \lambda_2 > 0 \Rightarrow \lambda_2 w_+ x_+ + \lambda_2 b \geq \frac{1}{2} \lambda_2 \end{array} \right\} \Rightarrow$$

$$(\lambda_1 w_+ + \lambda_2 w) x_+ + \lambda_1 b_+ + \lambda_2 b \geq \frac{1}{2} (\lambda_1 + \lambda_2) \Leftrightarrow$$

$$\left(\frac{\lambda_1}{\lambda_1 + \lambda_2} w_+ + \frac{\lambda_2}{\lambda_1 + \lambda_2} w \right) x_+ + \frac{\lambda_1 b_+ + \lambda_2 b}{\lambda_1 + \lambda_2} \geq \frac{1}{2}, \forall x_+ \in X^+/\widetilde{X}^+$$

$\left(\frac{\lambda_1}{\lambda_1 + \lambda_2} w_+ + \frac{\lambda_2}{\lambda_1 + \lambda_2} w, \frac{\lambda_1 b_+ + \lambda_2 b}{\lambda_1 + \lambda_2} \right)$ est un angle séparateur de X^+/\widetilde{X}^+
où encore $(\lambda_1 w_+ + \lambda_2 w, \lambda_1 b_+ + \lambda_2 b)$.

(w^*, b^*) où $b^* = \lambda_1 b_+ + \lambda_2 b$.

Conclusion

H_+ et H_- forment une marge de séparation X^+/\widetilde{X}^+ et X^-/\widetilde{X}^- .

\widetilde{H}_+ et \widetilde{H}_- forment une marge de séparation X^+/\widetilde{X}^+ et X^-/\widetilde{X}^- .

H_+^* et H_-^* forment une marge de séparation X^+/\widetilde{X}^+ et X^-/\widetilde{X}^- .

On détermine alors l'hyperplan H_+^* et H_-^* qui séparent \widetilde{X}^+ et \widetilde{X}^- .

Comme ils séparent aussi X^+/\widetilde{X}^+ et X^-/\widetilde{X}^- , sa marge ne dépasse pas celle de H_+ , H_- .

5.2 Applications

Exemple 5.2.1.

$$X^+ = \{(1, 3)\}.$$

$$X^- = \{(2, 2), (2, 1)\}.$$

1)

$$x_+ = \{(1, 3)\}.$$

$$x_- = \{(2, 2)\}.$$

$$\begin{cases} w_{11} + 3w_{12} = 1 \\ 2w_{11} + 2w_{12} = -1 \end{cases} \Rightarrow \begin{cases} w_{11} = -\frac{5}{4} \\ w_{12} = \frac{3}{4} \end{cases} \Rightarrow -\frac{5}{4}x_1 + \frac{3}{4}x_2 = 0 \quad (H)$$

$$\|w\|^2 = \frac{25}{16} + \frac{9}{16} = \frac{34}{16} = 2.125$$

$$-\frac{5}{4}x_1 + \frac{3}{4}x_2 = 1 \quad (H_+) \text{ et } -\frac{5}{4}x_1 + \frac{3}{4}x_2 = -1 \quad (H_-)$$

2)

$$x_+ = \{(1, 3)\}.$$

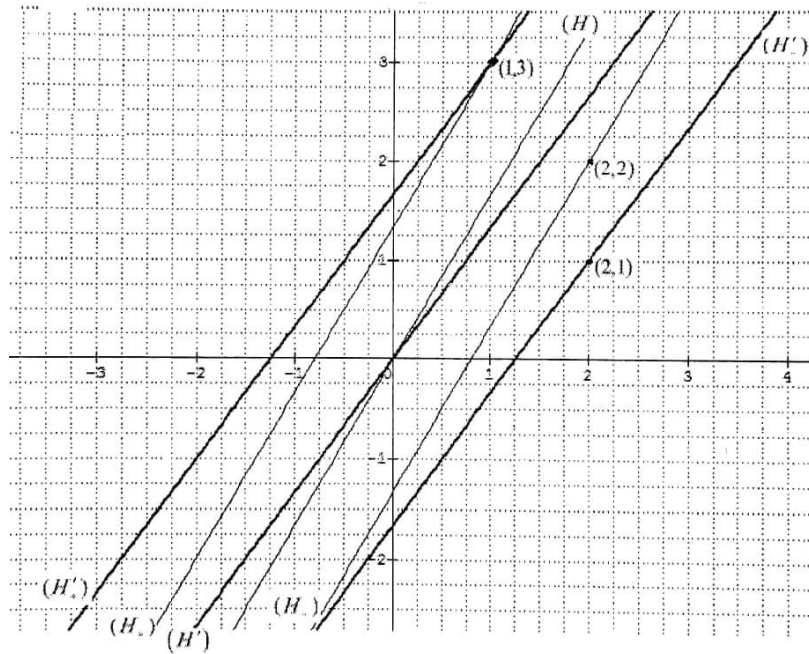
$$x_- = \{(2, 1)\}.$$

$$\begin{cases} w_{11} + 3w_{12} = 1 \\ 2w_{11} + 1w_{12} = -1 \end{cases} \Rightarrow \begin{cases} w_{11} = -\frac{4}{5} \\ w_{12} = \frac{3}{5} \end{cases} \Rightarrow -\frac{4}{5}x_1 + \frac{3}{5}x_2 = 0 \quad (H')$$

Alors $\|w'\|^2 = \frac{16}{25} + \frac{9}{25} = 1$

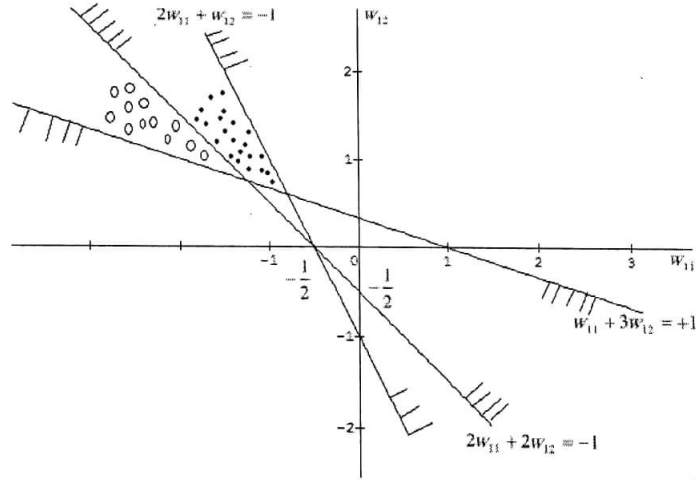
$$-\frac{4}{5}x_1 + \frac{3}{5}x_2 = 1 \quad (H'_+) \text{ et } -\frac{4}{5}x_1 + \frac{3}{5}x_2 = -1 \quad (H'_-)$$

Alors on a $\|w\|^2 > \|w'\|^2 \Rightarrow \frac{1}{\|w\|^2} < \frac{1}{\|w'\|^2}$ donc le séparateur est (H') .



$$X^+ = \{(1, 3)\}; \quad X^- = \{(2, 2), (2, 1)\}.$$

$$\begin{cases} w_{11} + 3w_{12} \geq +1 \\ 2w_{11} + 2w_{12} \leq -1 \\ 2w_{11} + w_{12} \leq -1 \end{cases}$$



$$\inf_{\substack{w_{11}+3w_{12} \geq +1 \\ 2w_{11}+2w_{12} \leq -1 \\ 2w_{11}+w_{12} \leq -1}} \|w\|^2 = \inf_{\substack{2w_{11}+2w_{12} = -1 \\ w_{11}+3w_{12} = +1}} \|w\|^2$$

L'hyperplan séparateur est $(H) : -\frac{5}{4}x_1 + \frac{3}{4}x_2 = 0$ mais il n'est pas maximal $\|w\|^2 = 2.125$.

Exemple 5.2.2. Supposons que l'hyperplan séparateur à vaste marge (SVM) passe par le point \bar{x}_+ .

Les inégalités $wx_+ + b \geq 1$, $wx_- + b \leq -1$ peuvent s'écrire sur la forme $\frac{w}{2}x_+ + \frac{b}{2} \geq \frac{1}{2}$, $\frac{w}{2}x_- + \frac{b}{2} \leq \frac{-1}{2}$ donc l'hyperplan séparateur $wx + b = 0$ est $\frac{w}{2}x + \frac{b}{2} = 0$.

Le problème devient alors :

$$\inf_{\substack{wx_+ + b \geq \frac{1}{2} \\ wx_- + b \leq -\frac{1}{2}}} \|w\|^2.$$

Comme $w\bar{x}_+ + b = \frac{1}{2} \Rightarrow b = \frac{1}{2} - w\bar{x}_+$.

D'où

$$\begin{aligned} wx_+ + b \geq \frac{1}{2} &\Leftrightarrow wx_+ + \left(\frac{1}{2} - w\bar{x}_+\right) \geq \frac{1}{2} \\ &\Leftrightarrow w(x_+ - \bar{x}_+) \geq 0. \end{aligned}$$

et $wx_- + b \leq -\frac{1}{2} \Leftrightarrow w(x_- - \bar{x}_+) \leq -1$.

D'où $\inf_{\substack{w(x_+ - \bar{x}_+) \geq 0 \\ w(x_- - \bar{x}_+) \leq -1}} \|w\|^2 = - \max_{\substack{w(x_+ - \bar{x}_+) \geq 0 \\ w(x_- - \bar{x}_+) \leq -1}} \{-\|w\|^2\}$

Le problème devient à chercher :
$$\begin{cases} \max\{-\|w\|^2\} \\ w(\bar{x}_+ - x_+) \leq 0 \\ w(x_- - \bar{x}_+) \leq -1. \end{cases}$$

$\Omega = \{w \in \mathbb{R}^2 / w(\bar{x}_+ - x_+) \leq 0 \text{ et } w(x_- - \bar{x}_+) \leq -1\}$ est un polyèdre convexe et $(0, 0) \notin \Omega$. Donc

$$\begin{cases} \max\{-\|w\|^2\} = p_\Omega(0). \\ w(\bar{x}_+ - x_+) \leq 0 \\ w(x_- - \bar{x}_+) \leq -1. \end{cases}$$

5.2. APPLICATIONS

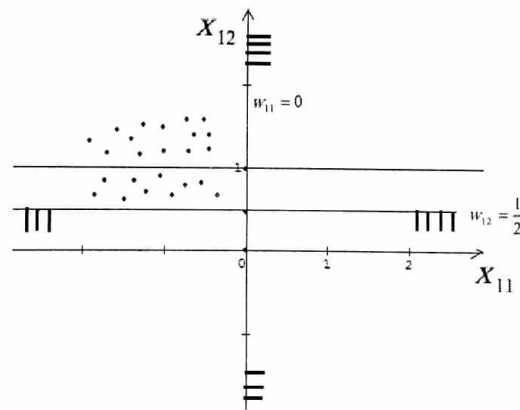
$$x_+ = (2, 3).$$

$$w(-2, 3) + (2, 3) = 0$$

$$w(-1, 3) + (2, 3) \leq 0 \Rightarrow w_{11} \leq 0$$

$$w((2, 2) - (2, 3)) \leq -1 \Rightarrow 0 - w_{12} \leq -1 \Rightarrow w_{12} \geq 1$$

$$w((2, 1) - (2, 3)) \leq -1 \Rightarrow -2w_{12} \leq -1 \Rightarrow w_{12} \geq \frac{1}{2}$$



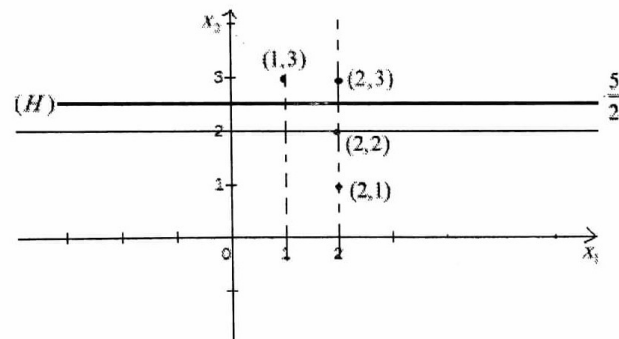
Donc $p_\Omega(0) = (0, 1)$.

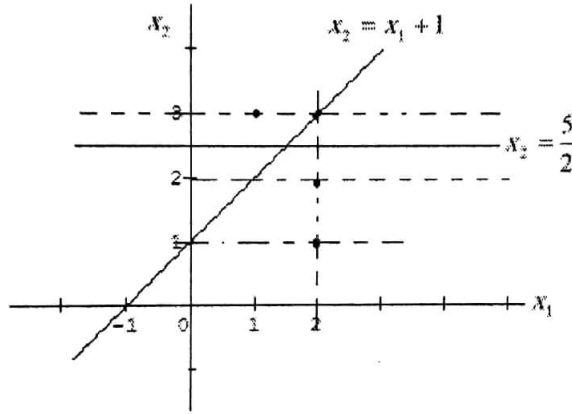
$$\|w\|^2 = 1 \Rightarrow \|w\| = 1.$$

$$b = \frac{1}{2} - (0, 1)(2, 3) = \frac{1}{2} - 3 = -\frac{5}{2}.$$

$$(0, 1)(x_1, x_2) - \frac{5}{2} = 0$$

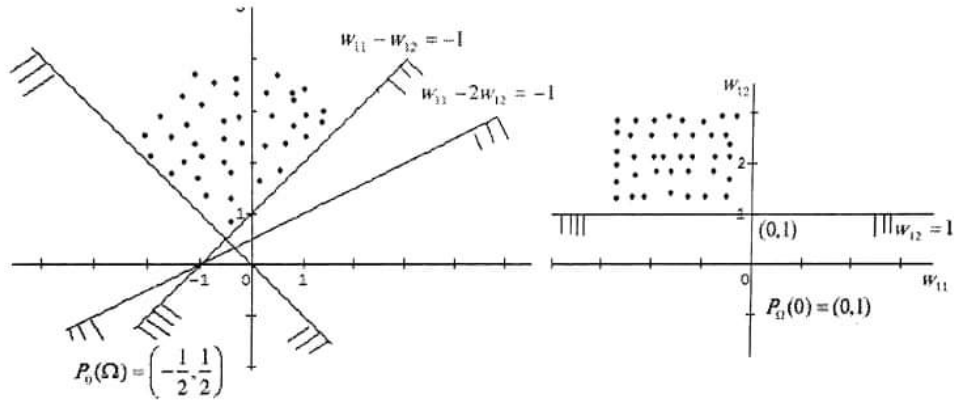
$$x_2 - \frac{5}{2} = 0 \Rightarrow x_2 = \frac{5}{2}.$$





$$\inf_{\substack{x_+ \in X^+ \\ x_- \in X^-}} \|x_+ - x_-\| = \|(2, 3) - (2, 2)\| = 1.$$

Domaine :



Exemple 5.2.3. Si la bande contient x'_- , alors $(\bar{x}_+ - x_-)$ et $(\bar{x}_+ - x'_-)$ sont indépendants.

$$\begin{vmatrix} \bar{x}_+^1 - x_-^1 & \bar{x}_+^1 - x_-^{1'} \\ \bar{x}_+^2 - x_-^2 & \bar{x}_+^2 - x_-^{2'} \end{vmatrix} \neq 0$$

Posons $A_1 = (\bar{x}_+^1 - x_-^1)_{x_- \in X^-}$ et $A_2 = (\bar{x}_+^2 - x_-^2)_{x_- \in X^-}$.

Si $A_1 = \lambda A_2$ donc $\begin{cases} \bar{x}_+^1 - x_-^1 = \lambda(\bar{x}_+^2 - x_-^2) \\ \bar{x}_+^1 - x_-^{1'} = \lambda(\bar{x}_+^2 - x_-^{2'}) \end{cases}$

$$\Rightarrow \begin{vmatrix} \bar{x}_+^1 - x_-^1 & \bar{x}_+^2 - x_-^2 \\ \bar{x}_+^1 - x_-^{1'} & \bar{x}_+^2 - x_-^{2'} \end{vmatrix} = \begin{vmatrix} \bar{x}_+^1 - x_-^1 & \bar{x}_+^1 - x_-^{1'} \\ \bar{x}_+^2 - x_-^2 & \bar{x}_+^2 - x_-^{2'} \end{vmatrix} = 0$$

contradiction. Donc on a :

$$\begin{aligned} w(\bar{x}_+ - x_+) &\leq 0, x_+ \in \widetilde{X}^+. \\ w(x_- - \bar{x}_+) &\leq -1, x_- \in \widetilde{X}^-. \end{aligned}$$

$$\begin{cases} w_{11}(\bar{x}_+^1 - x_{+1}^1) + w_{12}(\bar{x}_+^2 - x_{+1}^2) + w_3 = 0 \\ \vdots \\ w_{11}(\bar{x}_+^1 - x_{+p}^1) + w_{12}(\bar{x}_+^2 - x_{+p}^2) + w_{p+2} = 0 \\ w_{11}(\bar{x}_{-1}^1 - \bar{x}_+^1) + w_{12}(x_{-1}^2 - \bar{x}_+^2) + w_{p+3} = -1 \\ \vdots \\ w_{11}(\bar{x}_{-k}^1 - \bar{x}_+^1) + w_{12}(x_{-k}^2 - \bar{x}_+^2) + w_{p+(k-1)} = -1 \end{cases}$$

Posons $\tilde{A}_1 = (\bar{x}_+^1 - x_{+1}^1, \dots, \bar{x}_+^1 - x_{-k}^1)$ et $\tilde{A}_2 = (\bar{x}_+^2 - x_{+1}^2, \dots, \bar{x}_+^2 - x_{-k}^2)$.

Exemples

i)

$$X^+ = \{(1, 3)\}.$$

$$X^- = \{(2, 2), (2, 1)\}.$$

$$\bar{x}_+ = (1, 3), A_1 = (1, 1), A_2 = (-1, -2) \text{ et } \begin{vmatrix} 1 & 1 \\ -1 & -2 \end{vmatrix} \neq 0$$

ii)

$$X^+ = \{(1, 3), (2, 3)\}.$$

$$X^- = \{(2, 2), (2, 1)\}.$$

$$\bar{x}_+ = (1, 3), A_1 = (0, 0, 1) \text{ et } A_2 = (-1, -2, 0).$$

$$\bar{x}_+ = (2, 3), A_1 = (0, 0, 0) \text{ et } A_2 = (-1, -2, -1).$$

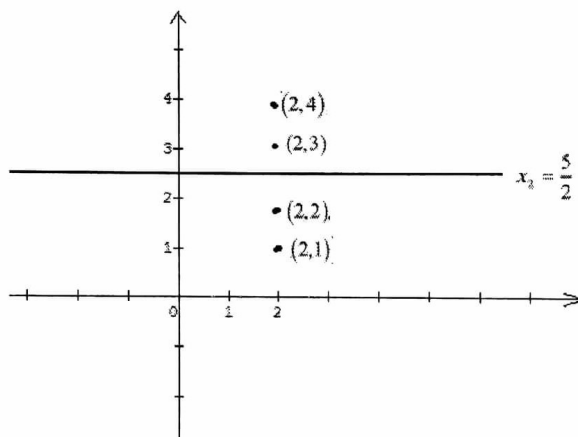
$$(-1, -2, -1)^t w_{12} \leq \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix}$$

$$(-1, -2, -1)^t w_{12} \leq \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} -1 \\ -2 \\ -1 \end{pmatrix} w_{12} \leq \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix} \Rightarrow \begin{cases} w_{12} \geq 1 \\ w_{12} \geq \frac{1}{2} \\ w_{12} \geq 0 \end{cases} \Rightarrow w_{12} \geq 1 \Rightarrow w_{12} = 1 \text{ et } w_{11} = 0.$$

$$w_{12} = 1, w_{11} = 0 \Rightarrow b = \frac{1}{2} - (0, 1)(2, 3) = \frac{1}{2} - 3 = -\frac{5}{2}.$$

$$(0, 1)(x_1, x_2) - \frac{5}{2} = 0 \Rightarrow x_2 = \frac{5}{2}.$$



Notes alors si A^1 et A^2 sont non nuls, alors

$$w((2,3) - (2,4)) \leq 0 \Rightarrow -w_{12} \leq 0$$

$$w((2,2) - (2,3)) \leq -1 \Rightarrow -w_{12} \leq -1$$

$$w((2,1) - (2,3)) \leq -1 \Rightarrow -2w_{12} \leq -1$$

donc $w_{12} \geq 1$. *iv*)

$$X^+ = \{(-1, 3)\}.$$

$$X^- = \{(-2, 2), (-2, 1)\}.$$

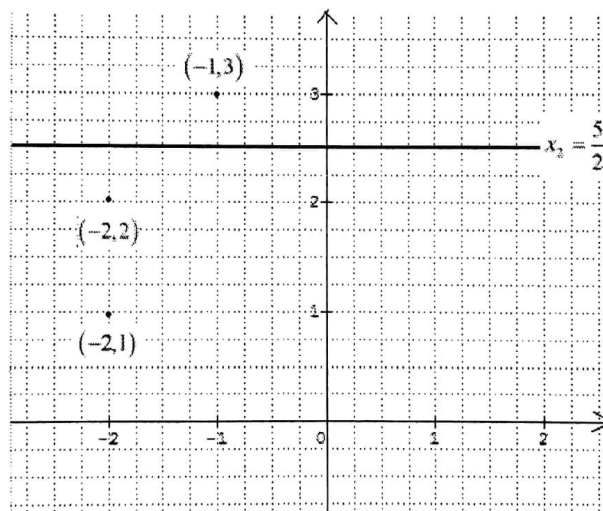
$$\bar{x}_+ = (-1, 3), A_1 = (-1, -1), A_2 = (-1, -2) \text{ et } \begin{vmatrix} -1 & -1 \\ -1 & -2 \end{vmatrix} \neq 0$$

$$\begin{pmatrix} -1 & -1 \\ -1 & -2 \end{pmatrix} \begin{pmatrix} w_{11} \\ w_{12} \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \Rightarrow \begin{cases} w_{11} + w_{12} = 1. \\ w_{11} + 2w_{12} = 1. \end{cases}$$

Donc $w_{11} = 1$ et $w_{12} = 0$.

$$b = \frac{1}{2} - (0, 1)(-1, 3) = \frac{1}{2} - 3 = -\frac{5}{2}.$$

$$(0, 1)(x_1, x_2) - \frac{5}{2} = 0 \Rightarrow x_2 = \frac{5}{2}.$$



Conclusion

Dans ce mémoire, nous avons tenté de présenter de manière simple le concept de système d'apprentissage introduit par Vladimir Vapnik. Cette méthode de classification est basée sur la recherche d'un hyperplan qui permet de séparer au mieux des ensembles de données, elle est applicable pour des tâches de classification à deux classes, mais il existe des extensions pour la classification multiclasse.

Les machines à vecteurs de supports ont l'avantage de ne pas nécessiter le réglage de beaucoup de paramètres et ont été appliquées avec grand succès dans de nombreux domaines :

- Finance (évolution des prix, valeurs en bourse, etc.)
- Structure des protéines (protein folding).
- Génomique (microarray gene expression data).
- Reconnaissance de visages.
- Détections des catastrophes, forecasting.
- Images satellite et surveillance.
- Reconnaissance de l'écriture manuscrite.
- Diagnostic médical.
- Le traitement d'images, la catégorisation textes...

Bibliographie

- [1] A.Cornuéjols, L.Miclet, *Apprentissage artificiel : Concepts et algorithmes (2^e édition)*, Eyrolles, 2009.
- [2] A.Kolmogorov, S.Fomine, *Éléments de la théorie des fonctions et de l'analyse fonctionnelle (2^e édition)*, Mir, Moscou, 1977.
- [3] C.Bishop, *Pattern recognition and machine learning*, Springer, 2002.
- [4] H.Brézis, *Analyse fonctionnelle : Théorie and applications*, Masson, Paris, 1983.
- [5] I.Steinwart, A.Christmann, *Support vector machines*, Springer-Verlag, 2008.
- [6] P.Caron, A.Juhel, F.Vandeveld, *Programmation linéaire*, Dunod, Paris, 1988.
- [7] R.Duda, P.Hart and D.Stork, *Pattern classification (second edition)*, Wiley, New York, 2001.
- [8] S.Delabriere, Y.Raynaud, *Analyse convexe*, Université Pierre et Marie Curie-Paris 6, 2000-2001.
- [9] T.Chenini, *Thèse de magister : Séparation des convexes par des hyperplans à vaste marge*, École normale supérieure de Laghouat, 2018-2019.
- [10] V.N.Vapnik, *Statistical learning theory*, Wiley, New York, 1998.
- [11] V.N.Vapnik, *The nature of statistical learning theory*, Springer-Verlag, 1995.