People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
University of 8 Mai 1945 Guelma

جامعة 8 ماي 1945 قالمة
UNIVERSITE 8 MAI 1945 GUELMA

Faculty: Mathematics, Computer Science and Science of Matter
Department of Mathematics

Lecture Notes

3rd Year Degree in Mathematics

Title

Unconstrained Optimization

Presented by:
Dr. Rabah DEBBAR

Academic year
2023/2024

Semestre : 5
Unité d'enseignement : Méthodologie
Matière : Optimisation sans contraintes
Crédits : 5
Coefficient : 2

Objectifs de l'enseignement : Ce cours traite les principaux outils et résultats de l'optimisation reposant sur des techniques d'analyse convexe et de dualité. Il présente les bases de la programmation dynamique.
Connaissances préalables recommandées : Bases d'analyse fonctionnelles, de topologie et d'algèbre linéaire. l'UE "Analyse convexe" est fortement recommandée.

Contenu de la matière :

Chapitre1 : Quelques rappels de calcul différentiel, Convexité

1.1 Différentiable, gradient, matrice hessienne
1.2 Développement de Taylor
1.3 Fonctions convexes

Chapitre2 : Minimisation sans contraintes

2.1 Résultats d'existence et d'unicité
2.2 Conditions d'optimalité du 1er ordre
2.3 Conditions d'optimalité du 2ème ordre

Chapitre3 : Algorithmes

3.1 Méthode du gradient
3.2 Méthode du gradient conjugué
3.3 Méthode de Newton
3.4 Méthode de relaxation
3.5 Travaux pratiques

Mode d'évaluation : Examen (60%), contrôle continu (40%)

Références :

1. M. Bierlaire, Introduction à l'optimisation différentiable, PPUR, 2006.
2. J-B. Hiriart-Urruty, Optimisation et analyse convexe, exercices corrigés, EDP Sciences, 9009.

جامعة 8 ماي 1945 قالمة
قسم الرياضيات
كلية الرياضيات، الإعلام الآلي وعلوم المادة

الجمـــــهورية الجزائريـــــة الديمقراطيـــة الشعبيـــة
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université 8 mai 1945 Guelma
Faculté des Mathématiques et de l'Informatique et des Sciences de la Matière

وزارة التعليم العالــي و البحث العلمي
جامعة 8 مـــاي 1945 قالمـــة
كليـــة الرياضيـات و الإعلام الآلـي
و علــوم المــادة

Département : Mathématiques

SYLLABUS

Unité d'Enseignement : UEM5.1.1, **Matière** : Optimisation sans contraintes

Domaine/Filière : 3ème année mathématiques

Semestre : 05 **Année Universitaire** : 2023-2024
Crédits : 05, **Coefficient** : 02
Volume Horaire Hebdomadaire Total : 67h30

- Cours Magistral (*Nombre d'heures par semaine*) : 01h30
- Travaux Dirigés (*Nombre d'heures par semaine*) : 01h30
- Travaux Pratiques (*Nombre d'heures par semaine*) : 01h30

Langue d'enseignement: Français
Enseignant responsable de la matière : Dr. Rabah. DEBBAR **Grade** : MCA
Bureau : E 5.9 B
Email : rabah.debbar@yahoo.fr **Téléphone**
Périodes de consultation :
Objectifs :
Le module propose une introduction à l'optimisation sans contraintes. Un étudiant ayant suivi ce cours saura reconnaître les outils et résultats de base en optimisation ainsi que les principales méthodes utilisées dans la pratique. Des séances de travaux pratiques sont proposées pour être notamment implémentés sous le logiciel de calcul scientifique Matlab et ce, afin d'assimiler les notions théoriques des algorithmes vues en cours.
Programme du cours théorique :
Chapitre1 : Quelques rappels de calcul différentiel, Convexité
1.1 Différentiabilité, gradient, matrice hessienne
1.2 Développement de Taylor
1.3 Fonctions convexes
Chapitre2 : Minimisation sans contraintes
2.1 Résultats d'existence et d'unicité
2.2 Conditions d'optimalité du 1er ordre
2.3 Conditions d'optimalité du 2nd ordre
Chapitre3 : Algorithmes
3.1 Méthode du gradient
3.2 Méthode du gradient conjugué
3.3 Méthode de Newton
3.4 Méthode de relaxation
3.5 Travaux pratiques
Evaluation : Contrôles des connaissances & Pondérations

Contrôle	Pondération (%)
Examen final	60%
Travaux Dirigés (Présence & Participation)	40%
Micro-Intérrogations	
Devoirs à Domicile	
Total	100%

Références bibliographiques (Livres et polycopiés, sites internet, etc.).
1. M. Bierlaire, Introduction à l'optimisation différentiable, PPUR, 2006.
2. J-B. Hiriart-Urruty, Optimisation et analyse convexe, exercices corrigés, EDP sciences, 2009.

Date & Signature
Rabah Debbar 20/09/2023

Les paroles s’envolent mais les écrits restent...
À cet effet ce polycopié!

To my thesis director Mr. Abdelkader DEHICI
and To my parents
May God have mercy on them
and make them dwell in His spacious gardens

Contents

Introduction

Optimization is central to any problem involving decision making, whether in engineering or in economics. The task of decision making entails choosing between various alternatives. This choice is governed by our desire to make the "best" decision. The measure of goodness of the alternatives is described by an objective function or performance index. Optimization theory and methods deal with selecting the best alternative in the sense of the given objective function.

The area of optimization has received enormous attention in recent years, primarily because of the rapid progress in computer technology, including the development and availability of user-friendly software, high-speed and parallel processors, and artificial neural networks. A clear example of this phenomenon is the wide accessibility of optimization software tools such as the Optimization Toolbox of MATLAB1 and the many other commercial software packages. There are currently several excellent graduate textbooks on optimization theory and methods (e.g., [1], [5], [6], [8], [9], [10], [12], [15]), as well as undergraduate textbooks on the subject with an emphasis on engineering design (e.g., [1]). However, there is a need for an introductory textbook on optimization theory and methods at a senior undergraduate or beginning graduate level. The present text was written with this goal in mind. The material is an outgrowth of our lecture notes for a one-semester course in optimization methods for seniors and beginning

0.1 Type of Optimization

The classification of optimization is not well established and there is some confusion in literature, especially about the use of some terminologies. Here we will use the most widely used terminologies. However, we do not intend to be rigorous in classifications; rather we would like to introduce all relevant concepts in a concise manner. Loosely speaking, classification can be carried out in terms of the number of objectives, number of constraints, function forms, landscape of the objective functions, type of design variables, uncertainty in values, and computational effort (see Figure 0.1 [16]).

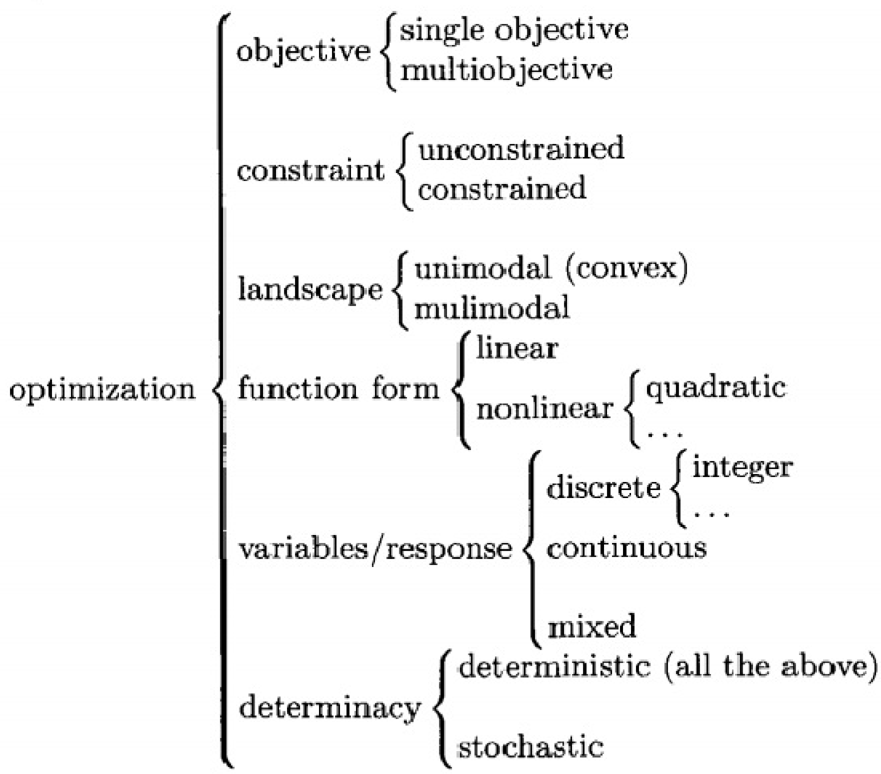


Figure 0.1: Classification of optimization problems.

Chapter 1

Basic Concepts of Unconstrained Optimization

1.1 Euclidean Space $\mathbb{R}^n$

The vector space $\mathbb{R}^n$ is the set of n-dimensional column vectors with real components endowed with the component-wise addition operator

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

and the scalar-vector product

$$\lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix},$$

where in the above $x_1, x_2, \dots x_n, \lambda$ are real numbers. Throughout the Handout we will be mainly interested in problems over $\mathbb{R}^n$, although other vector spaces will be considered in a few cases. We will denote the standard basis of $\mathbb{R}^n$ by $e_1, e_2, \dots e_n$, where e_i is the n-length column vector whose ith component is one while all the others are zeros. The column vectors of all ones and all zeros will be denoted by e and 0, respectively, where the length of the vectors will be clear from the context.

For given $x, y \in \mathbb{R}^n$, the closed line segment between x and y is a subset of $\mathbb{R}^n$ denoted by $[x, y]$ and defined as

$$[x, y] = x + \alpha(y - x) : \alpha \in [0, 1].$$

The open line segment (x, y) is similarly defined as

$$[x, y] = x + \alpha(y - x) : \alpha \in (0, 1).$$

1.1.1 Inner Products and Norms

Definition 1.1.1. *(inner product). An inner product on $\mathbb{R}^n$ is a map $\langle .,. \rangle : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ with the following properties:*

1. (symmetry) $\langle x, y\rangle = \langle y, x\rangle$ for any $x, y \in \mathbb{R}^n$.

2. (additivity) $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$ for any $x, y, z \in \mathbb{R}^n$.

3. (homogeneity) $\langle \lambda x, y\rangle = \lambda\langle y, x\rangle$ for any $\lambda \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$.

4. (positive definiteness) $\langle x, x\rangle > 0$ for any $x \in \mathbb{R}$ and $\langle x, x\rangle = 0$ if and only if $x = 0$.

Example 1.1.1. *Perhaps the most widely used inner product is the so-called dot product defined by*

$$\langle x, y\rangle = x^T y = \sum_{i=1}^{n} x_i y_i \quad x, y \in \mathbb{R}^n.$$

Since this is in a sense the "standard" inner product.

Definition 1.1.2. *(norm). A norm $\|.\|$ on $\mathbb{R}^n$ is a function $\|.\| : \mathbb{R}^n \longrightarrow \mathbb{R}$ satisfying the following:*

1. (nonnegativity) $\|x\| \geq 0$ *for any* $x \in \mathbb{R}^n$ *and* $\|x\| = 0$ *if and only if* $x = 0$.

2. (positive homogeneity) $\|\lambda x\| = |\lambda|\|x\|$ *for any* $x \in \mathbb{R}^n$ *and* $\lambda \in \mathbb{R}$.

3. (triangle inequality) $\|x + y\| \leq \|x\| + \|y\|$ *for any* $x, y \in \mathbb{R}^n$.

One natural way to generate a norm on Rn is to take any inner product $\langle ., . \rangle$ on $\mathbb{R}^n$ and define the associated norm

$$\|x\| = \sqrt{\langle x, x \rangle} \text{ for all} x \in \mathbb{R}^n,$$

which can be easily seen to be a norm. If the inner product is the dot product, then the associated norm is the so-called Euclidean norm or l_2-norm :

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \text{ for all} x \in \mathbb{R}^n.$$

The Euclidean norm belongs to the class of $l_p$norm (for $p \geq 1$) defined by

$$\|x\|_p =^p \sqrt{\sum_{i=1}^{n} |x_i|^p} \text{ for all} x \in \mathbb{R}^n.$$

Another important norm is the l_∞norm given by

$$\|x\|_\infty = \max_{i=1,2,\cdots n} |x_i| \text{ for all} x \in \mathbb{R}^n.$$

Lemma 1.1.1. *(Cauchy-Schwarz inequality). For any* $x, y \in \mathbb{R}^n$,

$$|x^T y| \leq \|x\|_2 \cdot \|y\|_2.$$

Equality is satisfied if and only if x and y are linearly dependent.

1.2 Matrices

1.2.1 Positive and Negative Definite or Semi Definite Matrix

Definition 1.2.1. *An* $n \times n$ *symmetric real matrix* M *and* x *of order* $n \times 1$ *column vector,* M *is said to be:*

1. positive definite if $x^T Mx > 0$ *for all* $x \neq 0$

2. negative definite if $x^T Mx < 0$*for all* $x \neq 0$

3. positive semidefinite if $x^T Mx \geq 0$ *for all* x

4. negative semidefinite if $x^T Mx \leq 0$ *for all* x

5. indefinite if it is neither positive nor negative semidefinite(i.e.if $x^T Mx > 0$ *for some* x *and* $x^T Mx < 0$ *for some* x*).*

Remark 1.2.1. *Test for Positive and Negative (Definite or Semi Definite) Matrix*

1. A matrix M *is positive definite if it is Symmetric and all its eigenvalues are positive*

2. All Upper Left (Leading) determinants are positive

3. A matrix M *is positive definite if it is Symmetric and all its pivots are positive*

4. $S = M^T M$ *Independent Columns (Means No Zero Column)*

1.3 Topology

Definition 1.3.1. ***(Open ball).*** *Let* $a \in \mathbb{R}^n$ *and* $\epsilon > 0$*. The open ball of radius* ϵ *centered at* a *is*

$$B_\epsilon(a) := \{x \in \mathbb{R}^n : \|x - a\| < \epsilon\}.$$

Definition 1.3.2. ***(Open sets).*** *A set* $U \subseteq \mathbb{R}^n$ *is open if*

$$\forall a \in U, \exists \epsilon > 0 \quad such \quad that \quad B_\epsilon(a) \subseteq U.$$

In other words, U *is open if every point of* U *is the center of an open ball contained in* U*.*

Definition 1.3.3. ***(closed sets).*** *A set* $U \subseteq \mathbb{R}^n$ *is said to be* ***closed*** *if it contains all the limits of convergent sequences of points in* U*; that is,* U *is closed if for every sequence of points* $\{x_i\}_{i\geq 1} \subseteq U$ *satisfying* $x_i \to x^*$ *as* $i \to \infty$*, it holds that* $x^* \in U$*.*

Definition 1.3.4. ***(Boundary).*** *Let $A \subseteq \mathbb{R}^n$. The boundary of A is the set of all points $a \in \mathbb{R}^n$ such that,*

$$\forall \epsilon > 0 \quad (B_\epsilon(a) \cap A \neq \emptyset \quad and \quad B_\epsilon(a) \setminus A \neq \emptyset.)$$

We denote the boundary of A by ∂A.

Definition 1.3.5. ***(boundedness and compactness).***

1. *A set $U \subseteq \mathbb{R}^n$ is called **bounded** if there exists $M > 0$ for which $U \subseteq B(O, M)$.*

2. *A set $U \subseteq \mathbb{R}^n$ is called **compact** if it is closed and bounded.*

Examples of compact sets are closed balls and line segments. The positive orthant is not compact since it is unbounded, and open balls are not compact since they are not closed.

1.4 Differentiability

1.4.1 Partial derivative

Definition 1.4.1. *For a real-valued function $f : U \to \mathbb{R}$ defined on an open set U in $\mathbb{R}^n$ and a point $\boldsymbol{a}$ of U: If $i = 1, 2, ..., n$, the **partial derivative** of f at $\boldsymbol{a}$ with respect to x_i is defined by:*

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \to 0} \frac{f(a + he_i) - f(a)}{h}$$

Note that $a + he_i = (a_1, ..., a_i + h, ..., a_n)$, so $a + he_i$ and $\boldsymbol{a}$ differ only in the ith coordinate. Thus the partial derivative is defined by the one-variable difference quotient for the derivative with variable x_i. Other common notations for the partial derivative are $f_{x_i}(a)$, $(D_i f)(a)$ and $\nabla_i f(a)$.

Geometric interpretation

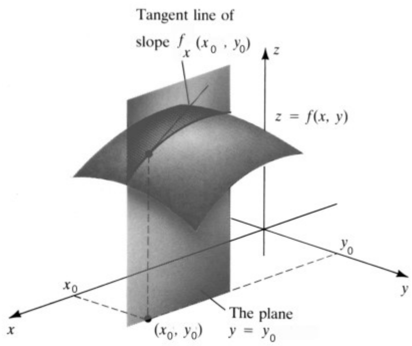


Figure: Graph of $z = f(x, y)$ and geometric interpretation of $\partial_x f(x_0, y_0)$.

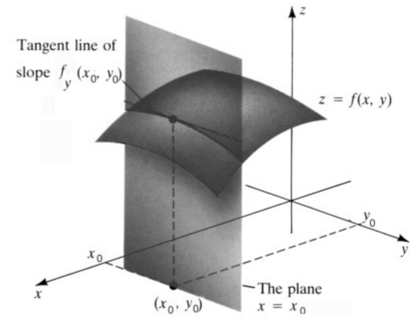


Figure: Graph of $z = f(x, y)$ and geometric interpretation of $\partial_y f(x_0, y_0)$.

Example 1.4.1. *Let*

$$f(x_1, x_2) = x_1^3 + x_2^2 + 4x_1x_2^2$$

Then, since $\dfrac{\partial}{\partial x_1}$ *treats* x_2 *as a constant,*

$$\frac{\partial f}{\partial x_1} = 3x_1^2 + 4x_2^2$$

and, since $\dfrac{\partial}{\partial x_2}$ *treats* x_1 *as a constant,*

$$\frac{\partial f}{\partial x_2} = 2x_2 + 8x_1x_2$$

In particular, at $(x_1, x_2) = (1, 0)$ *these partial derivatives take the values*

$$\frac{\partial f}{\partial x_1}(1, 0) = 3$$

$$\frac{\partial f}{\partial x_2}(1, 0) = 0$$

1.4.2 The Gradient

Definition 1.4.2. *Let $\Omega \subset \mathbb{R}^n$ be the domain of a real-valued functions $f : \Omega \to \mathbb{R}$ If f is differentiable we define the* ***gradient*** *of f to be the vector field $\nabla f : \Omega \to \mathbb{R}^n$ defined by*

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x) e_i.$$

The notation $grad f = \nabla f$ is also common.

Remark 1.4.1. *Since the gradient is a vector it can be written as either a row or a column unless it is used in conjunction with matrix multiplication. In that case it is assumed to be a column or an $n \times 1$ matrix. Note the relationship between the gradient and the total derivative, the $1 \times n$ (row) matrix*

$$Df(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)$$

We can think of the gradient as the transpose of the total derivative

$$\nabla f = Df^T.$$

Example 1.4.2. *Let*

$$f(x_1, x_2) = x_1^3 + x_2^2 + 4x_1x_2^2$$

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \end{pmatrix} = \begin{pmatrix} 3x_1^2 + 4x_2^2 \\ 2x_2 + 8x_1x_2 \end{pmatrix}$$

$$\nabla f(1,0) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(1,0) \\ \frac{\partial f}{\partial x_2}(1,0) \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

1.4.3 Hessian Matrix

Definition 1.4.3. *The Hessian Matrix, $H(x)$ or $\nabla^2 f(x)$ is defined to be the square matrix of second partial derivatives:*

$$H(x) = \nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ & & \vdots & \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{pmatrix},$$

We can also obtain the Hessian by applying the gradient operator on the gradient transpose,

$$H(x) = \nabla^2 f(x) = \nabla(\nabla f(x)^T) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix} \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right)$$

The Hessian is a symmetric matrix. The Hessian matrix gives us information about the curvature of a function, and tells us how the gradient is changing.

Example 1.4.3. *Let*

$$f(x_1, x_2) = x_1^3 + x_2^2 + 4x_1 x_2^2$$

$$H(x) = \nabla^2 f(x) = \begin{pmatrix} 6x_1 & 8x_2 \\ 8x_2 & 8x_1 + 2 \end{pmatrix},$$

$$H(1,0) = \nabla^2 f(1,0) = \begin{pmatrix} 6 & 0 \\ 0 & 10 \end{pmatrix},$$

1.5 Directional Derivatives

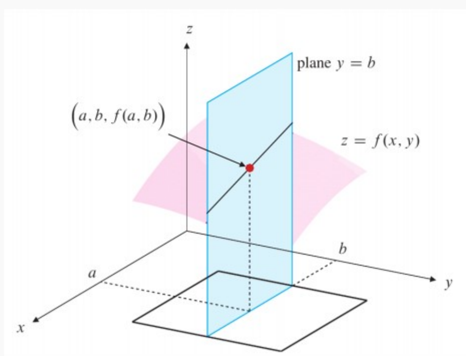
The gradient can be used to define a generalization of the partial derivative called the directional derivative(see [14]).

Definition 1.5.1. *Let $\Omega \in \mathbb{R}^n$ be the domain of a real-valued functions $f : \Omega \longrightarrow \mathbb{R}$, and let $v i n\mathbb{R}^n$ be a unit vector. If f is differentiable we define the **directional derivative** of f at $x \in \Omega$ in the direction v to be*

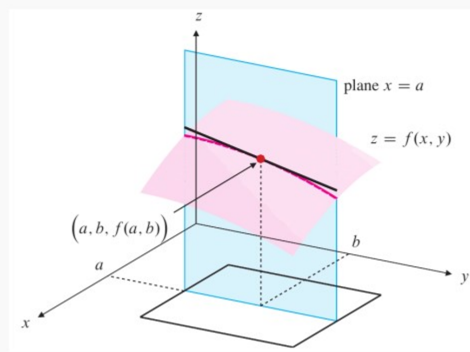
$$D_v f(x) = \frac{d}{dt} f(x+tv)\Big|_{t_0} = \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t}.$$

Partial derivatives are also directional derivatives

The partial derivative $\frac{\partial f}{\partial x} = D_{\mathbf{u}} f$ for $\mathbf{u} = (1,0)$, i.e., when $\mathbf{u}$ is the unit vector along the direction of x axis.



The partial derivative $\frac{\partial f}{\partial y} = D_{\mathbf{u}} f$ for $\mathbf{u} = (0,1)$, i.e., when $\mathbf{u}$ is the unit vector along the direction of y axis.



The following theorem gives us an easy way to calculate directional derivatives.

Theorem 1.5.1. *Let $\Omega \in \mathbb{R}^n$ be the domain of a real-valued functions $f : \Omega \longrightarrow \mathbb{R}$, and let $v \in \mathbb{R}^n$ be a unit vector. If f is differentiable then*

$$D_v f(x) = \nabla f(x) \cdot v.$$

Proof. For $x \in \Omega$ and any unit vector $v \in \mathbb{R}^n$ define $g : \Omega \longrightarrow \mathbb{R}$ by

$$g(t) = x + tv.$$

Note that $D_g = v, \quad g(0) = x$, and that $f(x+tv) = f(g(t))$. Thus, using the chain rule for mappings and the relationship between the total derivative and the gradient, we can compute

$$\begin{aligned} D_v f(x) &= \frac{d}{dt} f(g(t)) \\ &= Df(g(t))Dg(t)|_{t=0} \\ &= Df(x)\cdot v \\ &= \nabla f(x)\cdot v. \end{aligned}$$

Example 1.5.1. *Note that when v is one of the standard basis vectors e_i we get*

$$D_{e_i} f(x) = \frac{\partial f}{\partial x_i}(x).$$

Thus, partial derivatives are special cases of the directional derivative.
The following theorem gives us some geometric information about the gradient.

Theorem 1.5.2. *Suppose $f : \Omega \longrightarrow \mathbb{R}$ is a differentiable function and $\nabla f(x) \neq 0$. Then the directional derivative is maximized when v points in the direction of $\nabla f(x)$ and is minimized when v points in the direction of $-\nabla f(x)$. That is, $\nabla f(x)$ points in the direction of steepest increase of f while $-\nabla f(x)$ points in the direction of steepest decrease.*

Proof. Using the fact that v is a unit vector, we get

$$D_v f(x) = \nabla f(x)\cdot v = \cos\theta \|\nabla f(x)\| \cdot \|v\|.$$

where θ is the angle between $\nabla f(x)$ and v. Thus $D_v f(x)$ depends on v only through the angle θ. Thus, $D_v f(x)$ is maximized when the cosine is maximized ($\theta = 0$, v in the direction of $\nabla f(x)$) and minimized when the cosine is minimized ($\theta = \pi$, v in the direction of $-\nabla f(x)$). The next theorem describes the relationship between the gradient of a function and the level sets of that function.

Theorem 1.5.3. *Suppose $f : \Omega \longrightarrow \mathbb{R}$is differentiable. Then $\nabla f(x_0)$ is normal to the level surface of f at $x_0 \in \Omega$. That is, suppose $f(x_0) = c$, and $g(t)$ is a curve that lies entirely in the level set $f(x) = c$. If $g(t_0) = x_0$ then $\nabla f(x_0)$ is orthogonal to the tangent vector $g'(t_0)$.*

Proof. Suppose $f(g(t)) = c$ and $g(t_0) = x_0$: Since the composition is constant, its derivative is zero. Thus, using the chain rule we get

$$\begin{aligned} 0 &= \frac{d}{dt} f(g(t))\Big|_{t=t_0} \\ &= Df(g(t))Dg(t)|_{t=0} \\ &= Df(x_0)g'(t_0) \\ &= \nabla f(x_0)^T \cdot g'(t_0). \end{aligned}$$

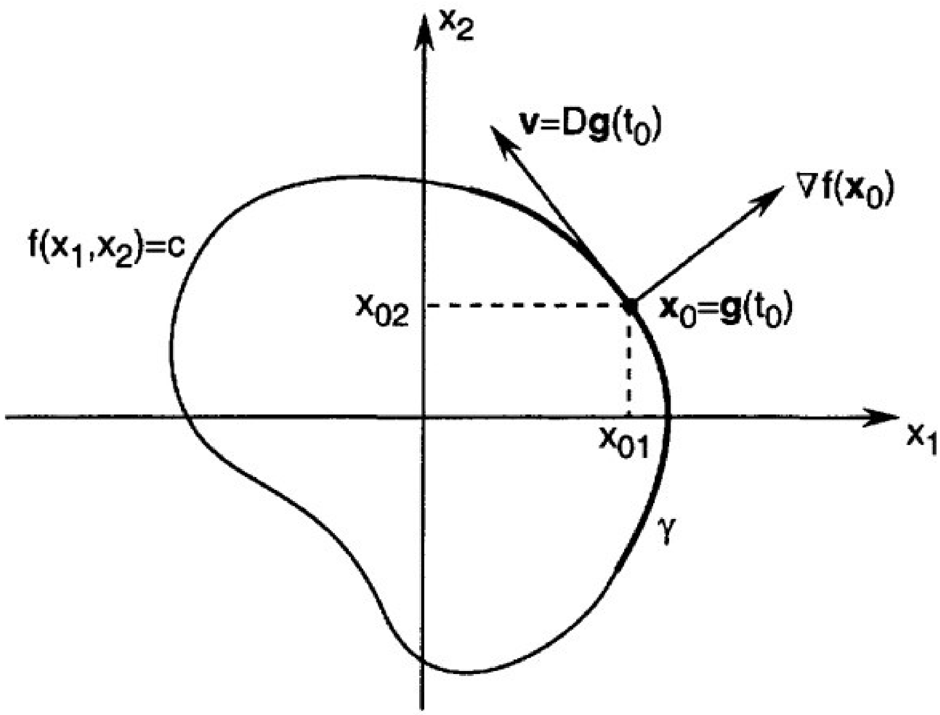


Figure 1.1 Orthogonality of the gradient to the level set

Example 1.5.2. *To find the equation for the tangent plane to the sphere*

$$x^2 + y^2 + z^2 = 14.$$

at the point $\boldsymbol{x}_0 = (x_0, y_0, z_0) = (1,2,3)$ *we calculate the gradient of* $f(x;y;z) = x^2 + y^2 + z^2$

$$\nabla f = (2x, 2y, 2z).$$

We evaluate this at the point $(1,2,3)$ *to get the normal vector* $n = (2,4,6)$*, and use this to derive the equation for the tangent plane*

$$0 = n(x - x_0) = \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} \cdot \begin{pmatrix} x-1 \\ y-2 \\ z-3 \end{pmatrix} = 2x + 4y + 6z - 28,$$

or $2x + 4y + 6z = 28$.

We can use the gradient to give a version of the Mean Value Theorem for scalar functions on $\mathbb{R}^n$.

Theorem 1.5.4. *Let* $\Omega \in \mathbb{R}^n$ *contain the entire line connecting* $x_1 \in \Omega$ *to* $x_2 \in \Omega$*, and suppose* $f : \Omega \longrightarrow \mathbb{R}$ *is* $\mathscr{C}^1$*. Then there is a point* $\widehat{x} \in \Omega$ *on the line segment between* x_1 *and* x_2 *such that*

$$f(x_2) - f(x_1) = \nabla f(\widehat{x}) \cdot (x_2 - x_1).$$

Proof. We define a real valued function of a single variable by

$$g(t) = f(tx_2 + (1-t)x_1), \quad t \in [0,1].$$

We note that this function is $\mathscr{C}^1$ and therefore the mean value theorem for real valued functions of a single variable says there exists $\widehat{t} \in (0,1)$ such that

$$g(1) - g(0) = g'(\widehat{t})(1-0).$$

Note that $g(1) = f(x_2)$ and $g(0) = f(x_1)$. The chain rule gives us

$$g'(t) = f(tx_2 + (1-t)x_1) \cdot (x_2 - x_1).$$

So if we let

$$\widehat{x} = \widehat{t}x_2 + (1-\widehat{t})x_1$$

this gives us the desired result.

1.6 Descent Direction

Definition 1.6.1. *Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a continuously differentiable function over $\mathbb{R}^n$. A vector $0 \neq d \in \mathbb{R}^n$ is called* ***a descent direction*** *of f at x if the directional derivative $D_d f(x)$ is negative, meaning that*

$$D_d f(x) = \nabla f(x) \cdot d < 0.$$

The most important property of descent directions is that taking small enough steps along these directions lead to a decrease of the objective function.

Lemma 1.6.1. ***(descent property of descent directions)****. Let f be a continuously differentiable function over $\mathbb{R}^n$, and let $x \in \mathbb{R}^n$. Suppose that d is a descent direction of f at x. Then there exists $\varepsilon > 0$ such that*

$$f(x + td) < f(x)$$

for any $t \in (0, \varepsilon]$.

Proof. Since $D_v f(x) < 0$, it follows from the definition of the directional derivative that

$$\lim_{t\to 0} \frac{f(x+td)-f(x)}{t} = D_d f(x) < 0.$$

Therefore, there exists an $\varepsilon > 0$ such that

$$\frac{f(x+td)-f(x)}{t} < 0.$$

for any $t \in (0,\varepsilon]$, which readily implies the desired result.

1.7 Multivariate Taylor Expansion

We now turn to the Taylor series expansion of a real-valued function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ about the point $x_0 \in \mathbb{R}^n$. Suppose $f \in \mathscr{C}^2$. Let x and x_0 be points in $\mathbb{R}^n$, and let $z(\alpha) = x_0 + \alpha(x - x_0)/\|x - x_0\|$. Define $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ by:

$$\phi(\alpha) = f((\alpha)) = f(x_0 + \alpha(x - x_0)/\|x - x_0\|).$$

Using the chain rule, we obtain

$$\begin{aligned}
\phi'(\alpha) &= \frac{d\phi}{d\alpha}(\alpha) \\
&= Df(z(\alpha))Dz(\alpha) \\
&= Df(z(\alpha))\frac{(x-x_0)}{\|x-x_0\|} \\
&= \frac{1}{\|x-x_0\|}(x-x_0)^T Df(z(\alpha))^T,
\end{aligned}$$

and

$$\begin{aligned}
\phi''(\alpha) &= \frac{d^2\phi}{d^2\alpha}(\alpha) \\
&= \frac{d}{d\alpha}\left(\frac{d\phi}{d\alpha}\right)(\alpha) \\
&= Df(z(\alpha))\frac{(x-x_0)}{\|x-x_0\|} \\
&= \frac{(x-x_0)^T}{\|x-x_0\|}\frac{d}{d\alpha}Df(z(\alpha))^T \\
&= \frac{(x-x_0)^T}{\|x-x_0\|}D(Df)z(\alpha))^T\frac{dz}{d\alpha}(\alpha) \\
&= \frac{1}{\|x-x_0\|}(x-x_0)^T D^2 f(z(\alpha))^T(x-x_0) \\
&= \frac{1}{\|x-x_0\|}(x-x_0)^T D^2 f(z(\alpha))(x-x_0),
\end{aligned}$$

$D^2 f = (D^2 f)^T$ since $f \in \mathscr{C}^2$. Observe that

$$\begin{aligned}
f(x) &= \phi(\|x-x_0\|) \\
&= \phi(0) + \frac{\|x-x_0\|}{1!}\phi'(0) + \frac{\|x-x_0\|^2}{2!}\phi''(0) + o(\|x-x_0\|^2).
\end{aligned}$$

Hence,

$$\begin{aligned}
f(x) &= f(x_0) + \frac{1}{1!}Df(x_0)(x-x_0) \\
&\quad + \frac{1}{2!}(x-x_0)^T D^2 f(x_0)(x-x_0) + o(\|x-x_0\|^2).
\end{aligned}$$

$$\lim_{x\to x_0}\frac{o(\|x-x_0\|)^2}{\|x-x_0\|^2} = 0$$

Theorem 1.7.1. ***(Taylor's Theorem)[11].*** *Suppose that $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is continuously differentiable and that $p \in \mathbb{R}^n$. Then we have that*

$$f(x+p) = f(x) + \nabla f(x+tp)^T p,$$

for some $t \in (0,1)$. Moreover, if f is twice continuously differentiable, we have that

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)p\,dt,$$

and that

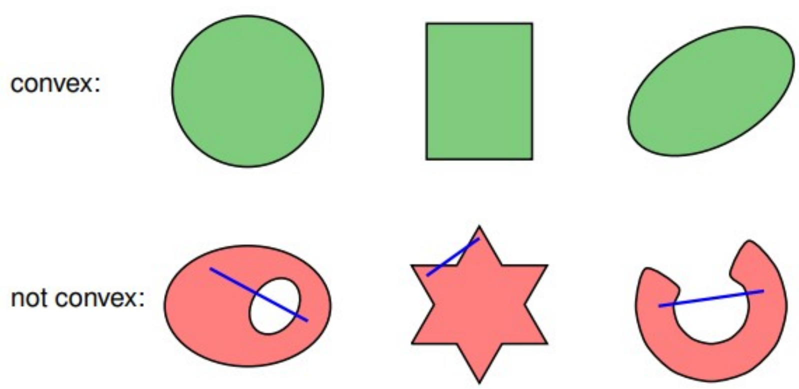
$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+tp)p,$$

for some $t \in (0,1)$.

1.8 Convex functions of several variables

1.8.1 Convex Sets

Definition 1.8.1. *A set $S \subseteq \mathbb{R}^n$ is called a convex set if for every choice of $X_1, X_2 \in S$, the points $\lambda X_1 + (1-\lambda)X_2 \quad \forall \lambda \in [0,1]$ lies in S i.e., if $X_1, X_2 \in S$ then line segment joining the points X_1 and X_2 must lie inside S.*



1.8.2 Conex Combination(Generalization of line segment)

Definition 1.8.2. *Convex combination of points* $X_1, X_2, \ldots, X_n \in \mathbb{R}^n$ *is given by*

$$X = \sum_{i=1}^{n} \lambda_i X_i, \quad \forall \lambda_i \geq 0 \quad and \quad \sum_{i=1}^{n} \lambda_i = 1.$$

i.e., A linear combination become a convex combination if all the Scalar's are non-negative and are such that their sum is equal to 1.

Remark 1.8.1. *1. Empty set, singleton set and whole of* $\mathbb{R}^n$ *are trivially convex sets ,*

2. Triangles, circles, ellipse, parabola with their interior are also convex sets,

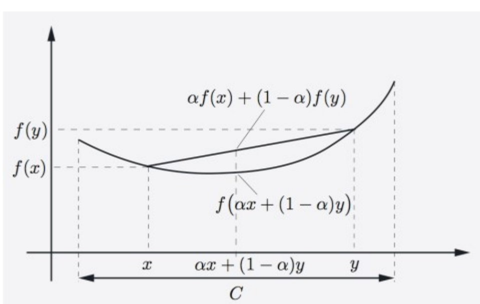
3. Some convex sets in $\mathbb{R}^2$ *are shown below.*

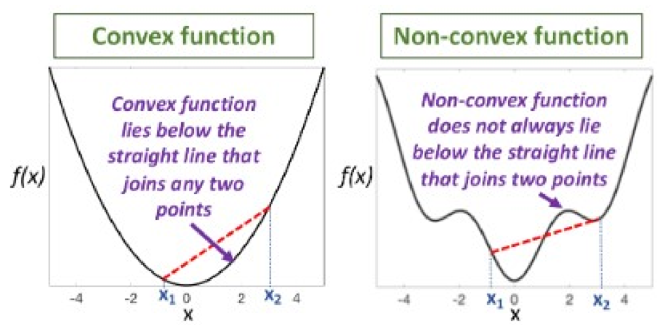
1.8.3 Convex Function

Definition 1.8.3. *Let* $f : S \to \mathbb{R}$ *be a function, where* S *is a non-empty convex set in* $\mathbb{R}^n$*. Then* f *is said to be a convex function on the set* S *if*

$$f(\lambda X_1 + (1-\lambda)X_2) \leq \lambda f(X_1) + (1-\lambda) f(X_2)$$

For all $X_1, X_2 \in S$ *and for each* $\lambda \in (0,1)$.

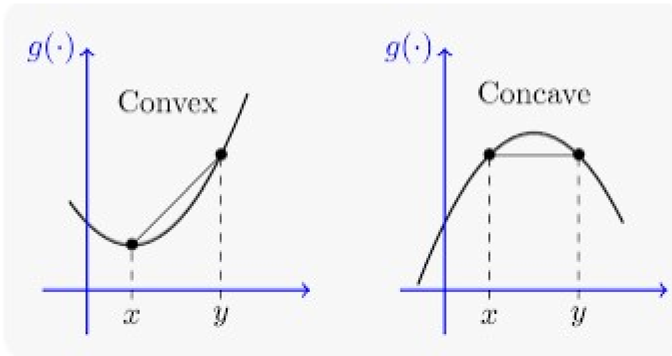


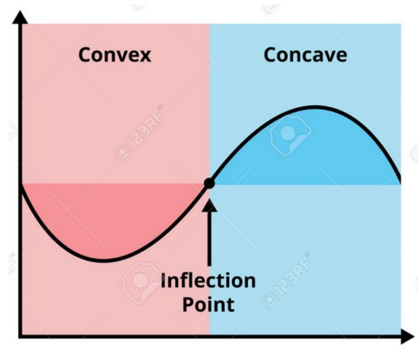


Remark 1.8.2. *1.* f *is said to be a concave function on the set* S *if*

$$f(\lambda X_1 + (1-\lambda)X_2) \geq \lambda f(X_1) + (1-\lambda) f(X_2)$$

For all $X_1, X_2 \in S$ *and for each* $\lambda \in (0,1)$,





2. f *is said to be strictly convex function on* S *if*

$$f(\lambda X_1 + (1-\lambda)X_2) < \lambda f(X_1) + (1-\lambda) f(X_2)$$

for all $X_1, X_2 \in S, \quad X_1 \neq X_2$ *and* $\lambda \in (0,1)$.

Properties 1.8.1. *1) If* $f(x)$ *is (strictly) convex, then* $-f(x)$ *is (strictly) concave (and vice versa).*

2) If $f_1(x), ..., f_k(x)$ *are convex (concave) functions and* $a_1, ..., a_k > 0$, *then*

$$g(x) = a_1 f_1(x) + ... + a_k f_k(x)$$

is also convex (concave).

3) If (at least) one of the functions $f_i(x)$ *is strictly convex (strictly concave), then* $g(x)$ *is strictly convex (strictly concave).*

1.8.4 Strongly Convex Function

Definition 1.8.4. f *is strongly convex with parameter* $m > 0$ *if*

$$f(tx + (1-t)y) \leq t f(x) + (1-t) f(y) - \frac{1}{2} m t(1-t) \|x-y\|_2^2$$

$$for \quad all \quad x, y \in S, \quad t \in [0,1].$$

Remark 1.8.3. *If f strongly convex (with any parameter $m > 0$), then f is strictly convex. The converse is not true: for example, the function $f(x) = exp(x)$ is strictly convex but not strongly convex .*

Example 1.8.1. *1-The function $f(x) = |x|$, $x \in \mathbb{R}$ f is convex bat is not strictly convex*
2-Every affine function $f(x) = ax + b$, $x \in \mathbb{R}$ is convex, but not strictly convex
3- $f(x) = x^2$, $x \in \mathbb{R}$ is strictly convex.

1.8.5 First-Order and Second-Order Characterization of Convex Functions

Differentiable Functions

Definition 1.8.5. *f is differentiable (i.e., its gradient ∇f exists at each point in dom f, which is open). at $\widehat{x} \in \mathbb{R}^n$, we write:*

$$\forall x \in \mathbb{R}^n, \quad f(x) = f(\widehat{x}) + \nabla f(\widehat{x})^\top (x - \widehat{x}) + o(\|x - \widehat{x}\|)$$

where by definition:

$$\lim_{x \to \widehat{x}} \frac{o(\|x - \widehat{x}\|)}{\|x - \widehat{x}\|} = 0$$

Twice Differentiable Function

Definition 1.8.6. *f is twice differentiable, that is, its Hessian or second derivative $\nabla^2 f$ exists at each point in dom f, which is open. at $\widehat{x} \in \mathbb{R}^n$, we write:*

$$\forall x \in \mathbb{R}^n, \quad f(x) = f(\widehat{x}) + \nabla f(\widehat{x})^\top (x - \widehat{x}) + \frac{1}{2}(x - \widehat{x})^\top H_f(\widehat{x})(x - \widehat{x}) + o(\|x - \widehat{x}\|^2)$$

where by definition:

$$\lim_{x \to \widehat{x}} \frac{o(\|x - \widehat{x}\|^2)}{\|x - \widehat{x}\|^2} = 0$$

Theorem 1.8.1. *Let $S \subseteq \mathbb{R}^n$ be convex and open. Then, for a function $f : S \to \mathbb{R}$, the following are equivalent.*

i) f is convex;

ii) for all $x, y \in S$,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

iii) for all $x, y \in S$,(monotonicity)

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$$

Proof.

$$\begin{aligned}
& i) \Longrightarrow ii) \quad Let \quad x, y \in S, \quad 0 \leq \lambda \leq 1 \\
& \Longrightarrow f((1-\lambda)x + \lambda y) \leq (1-\lambda) f(x) + \lambda f(y) \\
& \Longrightarrow f(x + \lambda(y-x)) - f(x) \leq \lambda(f(y) - f(x)) \\
& \Longrightarrow \frac{f(x + \lambda(y-x)) - f(x)}{\lambda} \leq f(y) - f(x) \\
& \Longrightarrow \lambda \to 0 \quad \langle \nabla f(x), y - x \rangle \leq f(y) - f(x).
\end{aligned}$$

$$\begin{aligned}
& ii) \Longrightarrow iii) \quad Let \quad x, y \in S \\
& \Longrightarrow f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \quad and \quad f(x) \geq f(y) - \langle \nabla f(y), y - x \rangle \\
& \Longrightarrow \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0.
\end{aligned}$$

$iii) \Longrightarrow i)$ Let $x, y \in S, 0 \leq \lambda \leq 1$

$f(x + \lambda(y-x)) - f(x) = \displaystyle\int_0^\lambda \frac{d}{dt} f(x + t(y-x))\,dt$

$= \int_0^\lambda \langle \nabla f(x+t(y-x)), y-x \rangle dt$

$\leq \int_0^\lambda \langle \nabla f(x+\lambda(y-x), y-x \rangle dt$

Because: $\langle \nabla f(x+\lambda(y-x)) - \nabla f(x+t(y-x)), \underbrace{(\lambda - t)}_{\geq 0}(y-x) \rangle \underbrace{\geq 0}_{leq(iii)}$

$$= \lambda \langle \nabla f(x+\lambda(y-x)), y-x \rangle.$$

$\underbrace{Analogously:}_{(x \leftrightarrow y \quad and \quad \lambda \leftrightarrow 1-\lambda)}$ $f(x+\lambda(y-x)) - f(y) \leq (1-\lambda)\langle \nabla f(x+\lambda(y-x)), x-y \rangle.$

Multiply the first ineq, with $(1-\lambda)$ the 2nd with λ.

$f(x+\lambda(y-x)) - (1-\lambda)f(x) - \lambda f(y) \leq 0.$

Theorem 1.8.2. *Let $S \subseteq \mathbb{R}^n$ be convex and open, and let $f : S \to \mathbb{R}$ be twice differentiable then f is convex if and only if $\nabla^2 f(x)$ is positive semidefinite for all $x \in S$*

Proof.

Let f be convex, let $d \in \mathbb{R}^n$

$\nabla^2 f(x) d = \lim_{t \to 0} \frac{\nabla f(x+td) - \nabla f(x)}{t}$

$\Rightarrow \langle d, \nabla^2 f(x) d \rangle = \lim_{t \to 0} \frac{1}{t} \langle \nabla f(x+td) - \nabla f(x), (x+td) - x \rangle$

$\Rightarrow \geq 0$

bay property (iii)of the previous thm.

Let $\nabla^2 f(x)$ be positlve semidefinite for all $x \in S$, by Taylor's thm,

$\forall x, y \in S : f(y) = f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2} \langle y-x, \nabla^2 f(z)(y-x) \rangle$

With $z = (1-\lambda)x + \lambda y$ for some $0 < \lambda < 1$

$f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle$

$\Rightarrow$ f is convex.

(ii) of the previous thm

Chapter 2

Unconstrained Optimization Theory

2.1 Introduction

In this chapter, we consider the optimization problem

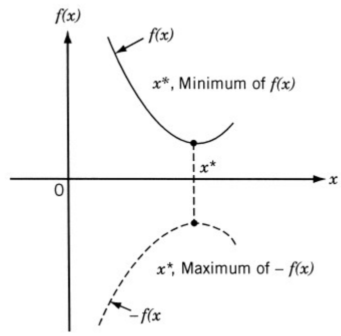
$$\begin{cases} \textbf{minimize} & f(x) \\ \textbf{subject to} & x \in \Omega. \end{cases}$$

The function $f : \mathbb{R}^n \to \mathbb{R}$ that we wish to minimize is a real-valued function, and is called the **objective function**, or cost function. The vector x is an n-vector of independent variables, that is, $x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$. The variables $x_1, x_2, \ldots, x_n$ are often referred to as decision variables. The set Ω is a subset of $\mathbb{R}^n$, called the **constraint set** or feasible set.

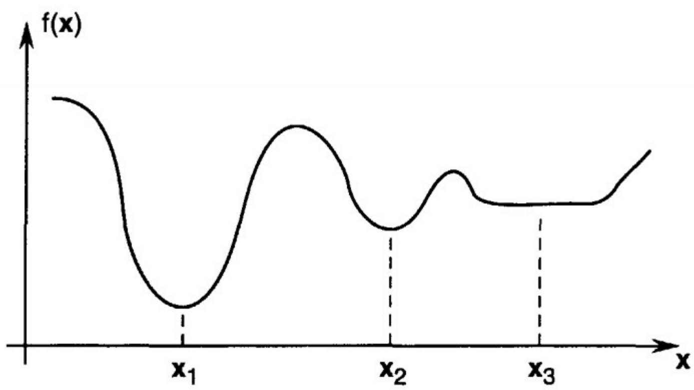
The optimization problem above can be viewed as a decision problem that involves finding the "best" vector x of the decision variables over all possible vectors in Ω. By the "best" vector we mean the one that results in the smallest value of the objective function. This vector is called the minimizer of f over Ω. It is possible that there may be many minimizers. In this case, finding any of the minimizers will suffice.

There are also optimization problems that require maximization of the objective function. These problems, however, can be represented in the above form because maximizing f

is equivalent to minimizing $-f$. Therefore, we can confine our attention to minimization problems without loss of generality(see [4],[13],[2]).



The above problem is a general form of a constrained optimization problem, because the decision variables are constrained to be in the constraint set Ω. If $\Omega = \mathbb{R}^n$, then we refer to the problem as an **unconstrained optimization problem**. In this chapter, we discuss basic properties of the general optimization problem above,



Examples of minimizers: x_1: strict global minimizer; x_2: strict local minimizer; x_3: local (not strict) minimizer

Definition 2.1.1. ***Local minimizer***. *Suppose that* $f : \mathbb{R}^n \to \mathbb{R}$ *is a real-valued function defined on some set* $\Omega \subset \mathbb{R}^n$. *A point* $\widehat{x}$ *is a local minimizer of f over* Ω *if there exists* $\epsilon > 0$ *such* *that $f(\hat{x}) \leq f(x)$ for all $x \in \Omega \setminus \{\hat{x}\}$ and $\|x - \hat{x}\| < \epsilon$.*

Definition 2.1.2. ***Global minimizer***. *A point $\hat{x} \in \Omega$, is a global minimizer of f over Ω if $f(\hat{x}) \leq f(x)$ for all $x \in \Omega \setminus \{\hat{x}\}$.*

Remark 2.1.1. *If, in the above definitions, we replace "$\leq$" with "$<$", then we have a strict local minimizer and a strict global minimizer, respectively.*

Remark 2.1.2. *Of course, a global minimum (maximum) point is also a local minimum (maximum) point. As with global minimum and maximum points, we will also use the terminology local minimizer and local maximizer for local minimum and maximum points, respectively.*

Another important issue is the one of deciding on whether a function actually has a global minimizer or maximizer. This is the issue of attainment or existence. A very well known result is due to Weierstrass, stating that a continuous function attains its minimum and maximum over a compact set.

2.2 Existence and Uniqueness of Optimal Solutions

Theorem 2.2.1. ***(Weierstrass theorem).*** *Let f be a continuous function defined over a nonempty and compact set $\Omega \subseteq \mathbb{R}^n$. Then there exists a global minimum point off over Ω and a global maximum point off over Ω.*

When the underlying set is not compact, the Weierstrass theorem does not guarantee the attainment of the solution, but certain properties of the function f can imply attainment of the solution even in the noncompact setting. One example of such a property is coerciveness.

Definition 2.2.1. ***(coerciveness).*** *Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a continuous function defined over $\mathbb{R}^n$.*

*The function f is called **coercive** if*

$$\lim_{\|x\|\to\infty} f(x) = \infty.$$

The important property of coercive functions that will be frequently used in this lecture-notes is that a coercive function always attains a global minimum point on any closed set.

Theorem 2.2.2. ***(attainment under coerciveness).*** *Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a continuous and coercive function and let $S \subseteq \mathbb{R}$ be a nonempty closed set. Then f has a global minimum point over S.*

Proof. Let $x_0 \in S$ be an arbitrary point in S. Since the function is coercive, it follows that there exists an $M > 0$ such that

$$f(x) > f(x_0) \quad \textit{for any} \quad x \quad \textit{such that} \quad \|x\| > M. \tag{2.1}$$

Since any global minimizer x^* off over S satisfies $f(x^*) < f(x_0)$, it follows from (2.1) that the set of global minimizers off over S is the same as the set of global minimizers of$f$ over $S \cap B[O, M]$. The set $S \cap B[O, M]$ is compact and nonempty, and thus by the Weierstrass theorem, there exists a global minimizer off over $S \cap B[O, M]$ and hence also over S.

Theorem 2.2.3. ***(strict convexity and uniqueness of sptimal solutions).*** *where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is strictly convex on Ω and Ω is a convex set. Then the optimal solution (assuming it exists) must be unique.*

Proof. Suppose there were two optimal solutions $x, y \in \mathbb{R}^n$. This means that $x, y \in \Omega$ and

$$f(x) = f(y) \leq f(z), \quad \forall z \in \Omega. \tag{2.2}$$

But consider $z = \dfrac{x+y}{2}$. By convexity of Ω , we have $z \in \Omega$. By strict convexity, we have

$$\begin{aligned} f(z) &= f\left(\frac{x+y}{2}\right) \\ &< \frac{1}{2}f(x) + \frac{1}{2}f(y) \\ &= \frac{1}{2}f(x) + \frac{1}{2}f(x) \\ &= f(x). \end{aligned}$$

But this contradicts (2.2)

2.3 Conditions for optimality

Definition 2.3.1. *A point $\widehat{x} \in \mathbb{R}^n$ at which $\nabla f(\widehat{x}) = 0$ is called a* ***stationary point****.*

2.3.1 Necessary optimality conditions

Theorem 2.3.1. *[3] Let x_{min} be a local minimum of a function $f : \mathbb{R}^n \to \mathbb{R}$. If f is differentiable in an open neighborhood V of x_{min}, then,*

$$\nabla f(x_{min}) = 0. \tag{2.3}$$

If, in addition, f is twice differentiable on V, then

$$\nabla^2 f(x_{min}) \text{ is positive semidefinite.} \tag{2.4}$$

Condition (2.1) is said to be a first-order necessary condition, and condition (2.2) is said to be a second-order necessary condition.

Proof. We recall that $-\nabla f(\widehat{x})$ is the direction of the steepest descent in $\widehat{x}$ (Lemma 1.6.1) and assume by contradiction that $\nabla f(\widehat{x}) \neq 0$. We can then use Theorem 1.5.2 with the descent direction $d = -\nabla f(\widehat{x})$ to obtain ε such that

$$f(\widehat{x}-t\nabla f(\widehat{x}))<f(\widehat{x}), \quad \forall t\in]0,\varepsilon],$$

which contradicts the optimality of $\widehat{x}$ and demonstrates the first-order condition. To demonstrate the second-order condition, we invoke Taylor's theorem in $\widehat{x}$, with an arbitrary direction d and an arbitrary step $t>0$ such that $\widehat{x}+td\in V$.

As

$$f(\widehat{x}+td)-f(\widehat{x})=td^T\nabla f(\widehat{x})+\frac{1}{2}t^2d^T\nabla^2 f(\widehat{x})d+0(\|td\|^2)$$

we have

$$\begin{aligned} f(\widehat{x}+td)-f(\widehat{x}) &= \frac{1}{2}t^2d^T\nabla^2 f(\widehat{x})d+0(\|td\|^2) \quad \textit{from } (2.3)\\ &= \frac{1}{2}t^2d^T\nabla^2 f(\widehat{x})d+0(t^2) \quad \|d\| \quad \textit{does not depend ont}\\ &\geq 0 \qquad\qquad \widehat{x} \quad \textit{is optimal.}\end{aligned}$$

When we divide by t^2, we get

$$\frac{1}{2}d^T\nabla^2 f(\widehat{x})d+\frac{0(t^2)}{t^2}\geq 0$$

Intuitively, as the second term can be made as small as desired, the result must hold.
More formally, let us assume by contradiction that $d^T\nabla^2 f(\widehat{x})d$ is negative and that its value is -2η, with $\eta>0$. According to the Landau notation $o(.)$,
for all $\eta>0$, there exists ε such that

$$\frac{|0(t^2)|}{t^2}<\eta, \qquad \forall 0<t\leq\varepsilon,$$

and

$$\frac{1}{2}d^T\nabla^2 f(\widehat{x})d+\frac{0(t^2)}{t^2}\leqslant\frac{1}{2}d^T\nabla^2 f(\widehat{x})d+\frac{|0(t^2)|}{t^2}<-\frac{1}{2}2\eta+\eta=0,$$

which contradicts and proves that $d^T\nabla^2 f(\widehat{x})d\geq 0$. Since d is an arbitrary direction, $\nabla^2 f(\widehat{x})$ is positive semidefinite

2.3.2 Sufficient optimality conditions

Theorem 2.3.2. *Consider a function* $f : \mathbb{R}^n \to \mathbb{R}$ *twice differentiable in an open subset* V *of* $\mathbb{R}^n$ *and let* $\widehat{x} \in V$ *satisfy the conditions*

$$\nabla f(\widehat{x}) = 0. \tag{2.5}$$

and

$$\nabla^2 f(\widehat{x}) \text{ is positive definite.} \tag{2.6}$$

In this case, $\widehat{x}$ *is a local minimum of* f.

Proof.

We assume by contradiction that there exists a direction d and $\varepsilon > 0$ such that, for any $0 < t \le \varepsilon$, $f(\widehat{x} + td) < f(\widehat{x})$. With an identical approach to the proof of Theorem 2.3.1, we have

$$\frac{f(\widehat{x}+td) - f(\widehat{x})}{t^2} = \frac{1}{2} d^T \nabla^2 f(\widehat{x}) d + \frac{o(t^2)}{t^2}$$

and

$$\frac{1}{2} d^T \nabla^2 f(\widehat{x}) d + \frac{o(t^2)}{t^2} < 0$$

or

$$\frac{1}{2} d^T \nabla^2 f(\widehat{x}) d + \frac{o(t^2)}{t^2} + \eta = 0$$

with $\eta > 0$.According to the definition of the Landau notation $o(.)$

there exists $\overline{\varepsilon}$ such that

$$\frac{|o(t^2)|}{t^2} < \eta, \qquad \forall t, 0 < t \le \overline{\varepsilon},$$

and then, for any $t \le min(\varepsilon, \overline{\varepsilon})$, we have

$$-\frac{o(t^2)}{t^2} \leq \frac{|o(t^2)|}{t^2} < \eta,$$

such that

$$\frac{1}{2} d^T \nabla^2 f(\widehat{x}) d = -\frac{o(t^2)}{t^2} - \eta < 0,$$

which contradicts the fact that $\nabla^2 f(\widehat{x})$ is positive definite.

Chapter 3

Unconstrained Optimization Methods

3.1 Steepest Descent (CAUCHY) Method

The use of the negative of the gradient vector as a direction for minimization was first made by Cauchy in 1847 [6.12]. In this method we start from an initial trial point X_1 and iteratively move along the steepest descent directions until the optimum point is found. The steepest descent method can be summarized by the following steps:

1. Start with an arbitrary initial point X_1. Set the iteration number as $i = 1$.

2. Find the search direction S_i as

$$S_i = -\nabla f_i = -\nabla f(X_i) \tag{3.1}$$

3. Determine the optimal step length $\widehat{\lambda}_i$ i in the direction S_i and set

$$X_{i+1} = X_i + \widehat{\lambda}_i S_i = X_i - \widehat{\lambda}_i \nabla f_i \tag{3.2}$$

4. Test the new point, X_{i+1}, for optimality. If X_{i+1} is optimum, stop the process. Otherwise, go to step 5.

5. Set the new iteration number $i = i + 1$ and go to step 2.

The method of steepest descent may appear to be the best unconstrained minimization technique since each one-dimensional search starts in the "best" direction. However, owing to the fact that the steepest descent direction is a local property, the method is not really effective in most problems.

Example 3.1.1. *Minimize* $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ *starting from the point* $X_1 = (0,0)$.

SOLUTION

Iteration 1

The gradient of f is given by

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(x) \\ \dfrac{\partial f}{\partial x_2}(x) \end{pmatrix} = \begin{pmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{pmatrix}$$

$$\nabla f_1 = \nabla f(X_1) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Therefore,

$$S_1 = -\nabla f_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

To find X_2, we need to find the optimal step length $\widehat{\lambda}_1$. For this, we minimize $f(X_1 + \lambda_1 S_1) = f(-\lambda_1, \lambda_1) = \lambda_1^2 - 2\lambda_1$ with respect to λ_1. Since $df/d\lambda_1 = 0$ at $\widehat{\lambda}_1 = 1$, we obtain

$$X_2 = X_1 + \widehat{\lambda}_1 S_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

As $\nabla f_2 = \nabla f(X_2) = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, X_2 is not optimum.

Iteration 2

$$S_2 = -\nabla f_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

To minimize

$$f(X_2 + \lambda_2 S_2) = f(-1+\lambda_2, 1+\lambda_2) = 5\lambda_2^2 - 2\lambda_2 - 1$$

we set $df/\lambda_2 = 0$. This gives $\widehat{\lambda}_2 = \frac{1}{5}$, and hence

$$X_3 = X_2 + \widehat{\lambda}_2 S_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \frac{1}{5}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.8 \\ 1.2 \end{pmatrix}$$

Since the components of the gradient at X_3, $\nabla f_3 = \begin{pmatrix} 0.2 \\ -0.2 \end{pmatrix}$, are not zero, we proceed to the next iteration.

Iteration 3

$$S_3 = -\nabla f_3 = \begin{pmatrix} -0.2 \\ 0.2 \end{pmatrix}$$

As

$$f(X_3 + \lambda_3 S_3) = f(-0.8 + 0.2\lambda_3, 1.2 + 0.2\lambda_3) = 0.04\lambda_3^2 - 0.08\lambda_3 - 1.2.$$

$\frac{df}{d\lambda_3} = 0$ at $\widehat{\lambda}_3 = 1.0$

Therefore,

$$X_4 = X_3 + \widehat{\lambda}_3 S_3 = \begin{pmatrix} -0.8 \\ 1.2 \end{pmatrix} + 1.0\begin{pmatrix} -0.2 \\ 0.2 \end{pmatrix} = \begin{pmatrix} -1.0 \\ 1.4 \end{pmatrix}$$

The gradient at X_4 is given by

$$\nabla f_4 = \begin{pmatrix} -0.20 \\ -0.20 \end{pmatrix}$$

Since $\nabla f_4 \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $X_4$is not optimum and hence we have to proceed to the next iteration. This process has to be continued until the optimum point, $\widehat{X} = \begin{pmatrix} -1.0 \\ 1.5 \end{pmatrix}$, is found.

Convergence Criteria : The following criteria can be used to terminate the iterative process.

1. When the change in function value in two consecutive iterations is small:

$$\left| \frac{f(X_{i+1}) - f(X_i)}{f(X_i)} \right| \leq \varepsilon_1 \tag{3.3}$$

2. When the partial derivatives (components of the gradient) of f are small:

$$\left| \frac{\partial f}{\partial x_i} \right| \leq \varepsilon_2 \tag{3.4}$$

3. When the change in the design vector in two consecutive iterations is small:

$$|X_{i+1} - X_i| \leq \varepsilon_3 \tag{3.5}$$

3.2 Conjugate Gradient (FLETCHER-REEVES) Method

The convergence characteristics of the steepest descent method can be improved greatly by modifying it into a conjugate gradient method (which can be considered as a conjugate directions method involving the use of the gradient of the function). That any minimization method that makes use of the conjugate directions is quadratically convergent. This property of quadratic convergence is very useful because it ensures that the method will minimize a quadratic function in n steps or less. Since any general function can be approximated reasonably well by a quadratic near the optimum point, any quadratically convergent method is expected to find the optimum point in a finite number of iterations.

We have seen that Powell's conjugate direction method requires n single-variable minimizations per iteration and sets up a new conjugate direction at the end of each iteration. Thus it requires, in general, n^2 single-variable minimizations to find the minimum of a quadratic function. On the other hand, if we can evaluate the gradients of the objective function, we can set up a new conjugate direction after every one-dimensional minimization, and hence we can achieve faster convergence. The construction of conjugate directions and development of the Fletcher-Reeves method are discussed in this section.

3.2.1 Development of the Fletcher-Reeves Method

The Fletcher-Reeves method is developed by modifying the steepest descent method to make it quadratically convergent. Starting from an arbitrary point X_1, the quadratic function

$$f(X) = \frac{1}{2} X^T [A] X + B^T X + C \tag{3.6}$$

can be minimized by searching along the search direction $S_1 = -\nabla f_1$(steepest descent direction)

$$\widehat{\lambda}_1 = -\frac{S_1^T \nabla f_1}{S_1^T A S_1} \tag{3.7}$$

The second search direction S_2 is found as a linear combination of S_1 and $-\nabla f_2$:

$$S_2 = -\nabla f_2 + \beta_2 S_1 \tag{3.8}$$

where the constant β_2 can be determined by making S_1 and S_2 conjugate with respect to $[A]$.

$$\beta_2 = -\frac{\nabla f_2^T \nabla f_2}{\nabla f_1^T S_1} = \frac{\nabla f_2^T \nabla f_2}{\nabla f_1^T \nabla f_1} \tag{3.9}$$

This process can be continued to obtain the general formula for the ith search direction as

$$S_i = -\nabla f_i + \beta_i S_{i-1} \tag{3.10}$$

where

$$\beta_i = \frac{\nabla f_i^T \nabla f_i}{\nabla f_{i-1}^T \nabla f_{i-1}} \tag{3.11}$$

Thus the Fletcher-Reeves algorithm can be stated as follows.

3.2.2 Fletcher-Reeves Method

The iterative procedure of Fletcher-Reeves method can be stated as follows:

1. Start with an arbitrary initial point X_1.

2. Set the first search direction $S_1 = -\nabla f(X_1) = -\nabla f_1$.

3. Find the point X_2 according to the relation

$$X_2 = X_1 + \widehat{\lambda}_1 S_1 \tag{3.12}$$

where $\widehat{\lambda}_1$ is the optimal step length in the direction S_1. Set $i = 2$ and go to the next step.

4. Find $\nabla f_i = \nabla f(X_i)$, and set

$$S_i = -\nabla f_i + \frac{|\nabla f_i|^2}{|\nabla f_{i-1}|^2} S_{i-1} \tag{3.13}$$

5. Compute the optimum step length $\widehat{\lambda}_i$ in the direction S_i, and find the new point

$$X_{i+1} = X_i + \widehat{\lambda}_i S_i \tag{3.14}$$

6. Test for the optimality of the point X_{i+1}. If X_{i+1} is optimum, stop the process. Otherwise, set the value of $i = i + 1$ and go to step 4.

Example 3.2.1. *Minimize* $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ *starting from the point* $X_1 = (0, 0)$.

SOLUTION

Iteration 1

The gradient of f is given by

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(x) \\ \dfrac{\partial f}{\partial x_2}(x) \end{pmatrix} = \begin{pmatrix} 1+4x_1+2x_2 \\ -1+2x_1+2x_2 \end{pmatrix}$$

$$\nabla f_1 = \nabla f(X_1) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

The search direction is taken as

$$S_1 = -\nabla f_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

To find the optimal step length $\widehat{\lambda}_1$ along S_1, we minimize $f(X_1+\lambda_1 S_1)$ with respect to λ_1. Here

$$f(X_1+\lambda_1 S_1) = f(-\lambda_1,\lambda_1) = \lambda_1^2 - 2\lambda_1$$

$$\frac{df}{d\lambda_1} = 0 \quad \textbf{at} \quad \widehat{\lambda}_1 = 1$$

Therefore,

$$X_2 = X_1 + \widehat{\lambda}_1 S_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + 1\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Iteration 2

Since $\nabla f_2 = \nabla f(X_2) = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$, Eq. (3.13) gives the next search direction as

$$S_2 = -\nabla f_2 + \frac{|\nabla f_2|^2}{|\nabla f_1|^2} S_1$$

where

$$|\nabla f_1|^2 = 2 \quad \textbf{and} \quad |\nabla f_2|^2 = 2$$

Therefore,

$$S_2 = -\begin{pmatrix} -1 \\ -1 \end{pmatrix} + \frac{2}{2}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ +2 \end{pmatrix}$$

To find $\widehat{\lambda}_2$, we minimize

$$\begin{aligned} f(X_2 + \lambda_2 S_2) = f(-1, 1 + 2\lambda_2) \\ &= -1 - (1 + 2\lambda_2) + 2 - 2(1 + 2\lambda_2) + (1 + 2\lambda_2)^2 \\ &= 4\lambda_2^2 - 2\lambda_2 - 1 \end{aligned}$$

with respect to λ_2. As $df/d\lambda_2 = 8\lambda_2 - 2 = 0$ at $\widehat{\lambda}_2 = \frac{1}{4}$, we obtain

$$X_3 = X_2 + \widehat{\lambda}_2 S_2 = \begin{pmatrix} -1 \\ -1 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1.5 \end{pmatrix}$$

Thus the optimum point is reached in two iterations. Even if we do not know this point to be optimum, we will not be able to move from this point in the next iteration. This can be verified as follows.

Iteration 3

Now

$$\nabla f_3 = \nabla f(X_3) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad |\nabla f_2|^2 = 2, \quad \textbf{and} \quad |\nabla f_3|^2 = 0$$

Thus

$$S_3 = -\nabla f_3 + (|\nabla f_3|^2 / |\nabla f_2|^2) S_2 = -\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{0}{2}\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This shows that there is no search direction to reduce f further, and hence X_3 is optimum.

3.3 NEWTON'S Method

Newton's method can be extended for the minimization of multivariable functions. For this, consider the quadratic approximation of the function $f(X)$ at $X = X_i$ using the Taylor's series expansion

$$f(X) = f(X_i) + \nabla f_i^T (X - X_i) + \frac{1}{2}(X - X_i)^T [J_i](X - X_i) \tag{3.15}$$

where $[J_i] = [J]|X_i$ is the matrix of second partial derivatives (Hessian matrix) of f evaluated at the point X_i. By setting the partial derivatives of Eq. (3.15) equal to zero for the minimum of $f(X)$, we obtain

$$\frac{\partial f}{\partial x_j} = 0, \quad j = 1, 2, \ldots, n \tag{3.16}$$

Equations (3.16) and (3.15) give

$$\nabla f = \nabla f_i [J_i](X - X_i) = 0 \tag{3.17}$$

If $[J_i]$ is nonsingular, Eqs. (3.17) can be solved to obtain an improved approximation $(X = X_{i+1})$ as

$$X_{i+1} = X_i - [J_i]^{-1} \nabla f_i \tag{3.18}$$

Since higher-order terms have been neglected in Eq. (3.15), Eq. (3.18) is to be used iteratively to find the optimum solution $\widehat{X}$.

The sequence of points $X_1, X_2, ..., X_{i+1}$ can be shown to converge to the actual solution $\widehat{X}$ from any initial point X_1 sufficiently close to the solution $\widehat{X}$, provided that $[J_1]$ is nonsingular. It can be seen that Newton's method uses the second partial derivatives of the objective function (in the form of the matrix $[J_i]$) and hence is a second-order method.

Example 3.3.1. *Show that the Newton's method finds the minimum of a quadratic function in one iteration.*

SOLUTION

Let the quadratic function be given by

$$f(X) = \frac{1}{2} X^T [A] X + B^T X + C$$

The minimum of $f(X)$ is given by

$$\nabla f = [A] X + B = 0$$

or

$$\widehat{X} = -[A]^{-1} B$$

The iterative step of Eq. (3.18) gives

$$X_{i+1} = X_i - [A]^{-1}([A] X_i + B) \qquad (E_1)$$

where X_i is the starting point for the ith iteration. Thus Eq. (E_1) gives the exact solution

$$X_{i+1} = \widehat{X} = -[A]^{-1} B$$

Figure 3.01 illustrates this process.

Example 3.3.2. *Minimize* $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1 x_2 + x_2^2$ *by taking the starting point as* $X_1 = (0, 0)$.

SOLUTION

To find X_2 according to Eq. (3.18), we require $[J_1]^{-1}$, where

$$[J_1] = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} \end{pmatrix}_{X_1} = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$$

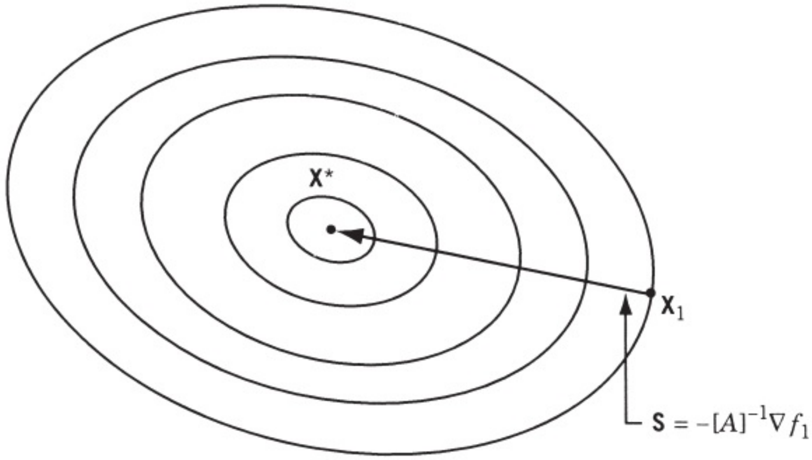


Figure 3.01 Minimization of a quadratic function in one step.

Therefore,

$$[J_1]^{-1} = \frac{1}{4}\begin{pmatrix} +2 & -2 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} +\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}$$

As

$$g_1 = \begin{pmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \end{pmatrix}_{X_1} = \begin{pmatrix} 1+4x_1+2x_2 \\ -1+2x_1+2x_2 \end{pmatrix}_{(0,0)} = \begin{pmatrix} +1 \\ -1 \end{pmatrix}$$

Equation (3.18) gives

$$X_2 = X_1 - [J_1]^{-1}g_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} +\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ \frac{3}{2} \end{pmatrix}$$

To see whether or not X_2 is the optimum point, we evaluate

$$g_2 = \begin{pmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \end{pmatrix}_{X_2} = \begin{pmatrix} 1+4x_1+2x_2 \\ -1+2x_1+2x_2 \end{pmatrix}_{(-1,3/2)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

As $g_2 = 0$, X_2 is the optimum point. Thus the method has converged in one iteration for this quadratic function.

If $f(X)$ is a nonquadratic function, Newton's method may sometimes diverge, and it may converge to saddle points and relative maxima. This problem can be avoided by modifying Eq. (3.18) as

$$X_{i+1} = X_i + \widehat{\lambda}_i S_i = X_i - \widehat{\lambda}_i [J_i]^{-1} \nabla f_i \tag{3.19}$$

where $\widehat{\lambda}_i$ is the minimizing step length in the direction $S_i = -\widehat{\lambda}_i [J_i]^{-1} \nabla f_i$. The modification indicated by Eq. (3.19) has a number of advantages. First, it will find the minimum in lesser number of steps compared to the original method. Second, it finds the minimum point in all cases, whereas the original method may not converge in some cases. Third, it usually avoids convergence to a saddle point or a maximum. With all these advantages, this method appears to be the most powerful minimization method. Despite these advantages, the method is not very useful in practice, due to the following features of the method:

1. It requires the storing of the $n \times n$ matrix $[J_i]$.

2. It becomes very difficult and sometimes impossible to compute the elements of the matrix $[J_i]$.

3. It requires the inversion of the matrix $[J_i]$ at each step.

4. It requires the evaluation of the quantity $[J_i]^{-1} \nabla f_i$ at each step.

These features make the method impractical for problems involving a complicated objective function with a large number of variables.

Chapter 4

Practical Work

4.1 TP No. 01

TP1RABAH

Extreme point analysis

For the function f:

$\Re^2 \longrightarrow \Re$ Définie $f(x, y) = 2x^3 + 2y^3 - 9x^2 + 3y^2 - 12y$

To find the critical points, we use the symbolic variables syms of Matlab(Symbolic Toolbox) which make it easy to find partial derivatives1

```
Command Window
>> syms x y
f=2*x^3+2*y^3-9*x^2+3*y^2-12*y;
fx=diff(f,x)
fy=diff(f,y)

fx =

6*x^2 - 18*x


fy =

6*y^2 + 6*y - 12
```

We use the solve command to find the place where the partial drifts are simultaneously equal to zero

```
>> S=solve(fx,fy)

S =

   x: [4x1 sym]
   y: [4x1 sym]
```

To examine the S camps, we write…

```
>> [S.x,S.y]

ans =

[ 0,  1]
[ 0, -2]
[ 3,  1]
[ 3, -2]
```

To classify the points we use the second derivative test, which consists of evaluating the sign of the determinant of the Hessian matrix

$$|H| = f_{xx}(x, y) * f_{yy}(x, y) - f^2{}_{xy}(x, y)$$

We define

1

```
>> fxx=diff(fx,x)
fyy=diff(fy,y)
fxy=diff(fx,y)

fxx =

12*x - 18


fyy =

12*y + 6

fxy =

0
```

And we evaluate them at each point found previously. As f is defined as a value
symbolic, the command to write is:

```
>> a=subs(f,[x,y],[0,1])

a =

    -7

>> a=subs(f,[x,y],[0,-2])

a =

    20
```

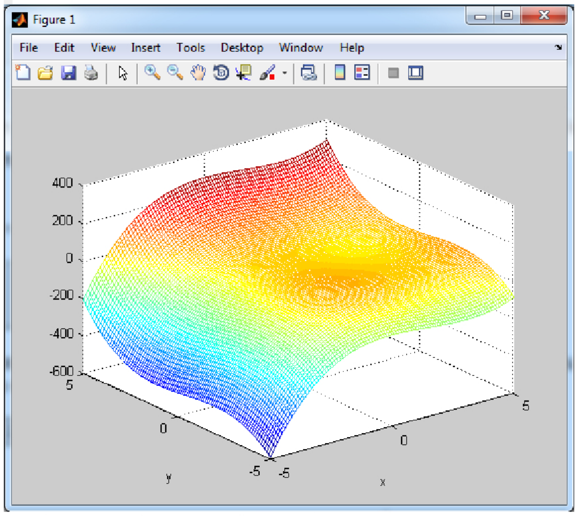
```
>> a=subs(f,[x,y],[3,1])

a =

   -34

>> a=subs(f,[x,y],[3,-2])

a =

    -7
```

(x, y)	$f_{xx}(x, y)$	$f_{xx}(x, y) * f_{yy}(x, y) - f^2_{xy}(x, y)$	**Classification**
(0,1)	-7	-324	Saddle point
(0,-2)	20	324	Local minimum
(3,1)	-34	324	Local maximum
(3,-2)	-7	-324	Saddle point

To visualize the solutions and ensure the correct classification of the points, we create a mesh of points, and we define the function:

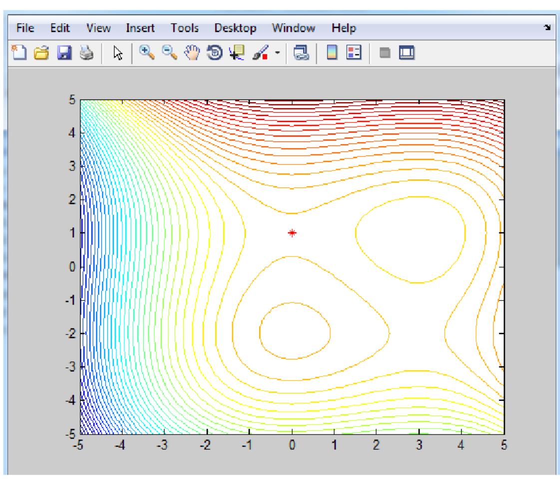
2

```
>> [x,y]=meshgrid(-5:0.1:5);
z=2*x.^3+2*y.^3-9*x.^2+3*y.^2-12*y;
mesh(x,y,z)
xlabel('x')
ylabel('y')
zlabel(z='f(x,y)')
```



To better locate the points we use a contour map and locate the points there:

```
>> contour(x,y,z,50)
hold on
plot(0,1,'r*');
```

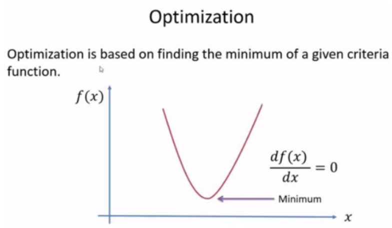


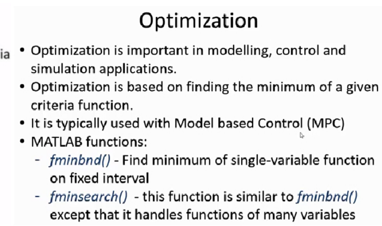
Example: For the function f: $f(x,y) = (1-x)^2 + 100(y - x^2)^2$
Find the critical points.

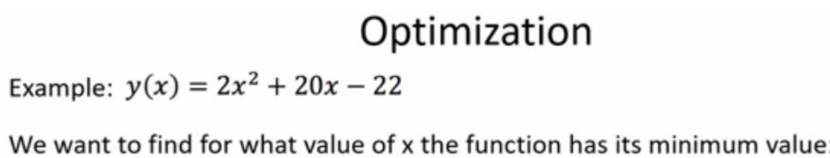
3

4.2 TP No. 02

TP2RABAH

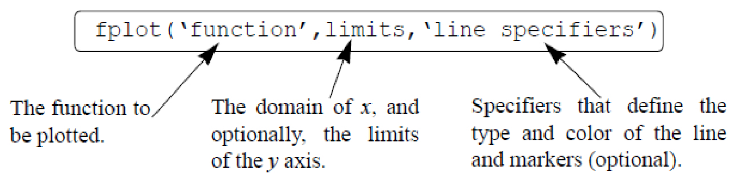






THE fplot COMMAND

The `fplot` command plots a function with the form $y = f(x)$ between specified limits. The command has the form:



Méthode1

```
>> fplot('2*x^2+20*x-22',[-20 20], '--r')
```

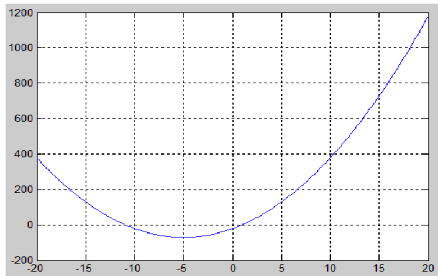
Méthode2

```
>> x=[-20:0.1:20];
y=2.*x.^2+20.*x-22;
plot(x,y)
```

Méthode3

Editor Window (TP2RABAH1.m) and Figure Window

1




```
x=[-20:0.1:20];
y=2.*x.^2+20.*x-22;
plot(x,y)
grid
```

```
x=[-20:0.1:20];
y=2.*x.^2+20.*x-22;
plot(x,y)
grid
i=1;
while (y(i) > y(i+1))
    i=i+1;
end
x(i)
y(i)
```

Command Window:

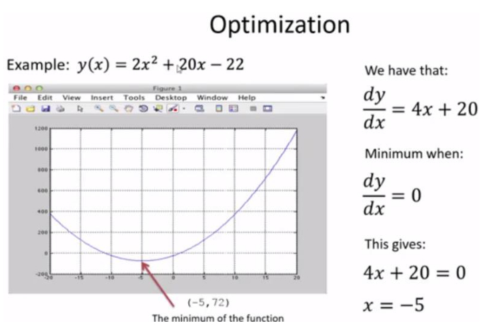
```
ans =

    -5


ans =

   -72

fx >>
```

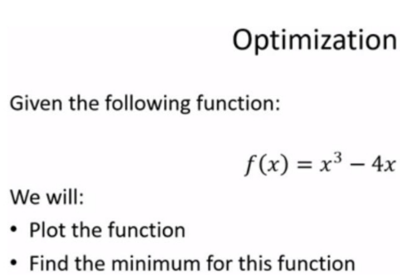
Mysimplefunc.m and TP2RABAH2.m

```
function f = mysimplefunc(x)
f = 2.*x.^2+20.*x-22;
```

```
x = -20:0.1:20
f = 2.*x.^2+20.*x-22
plot(x, f)

xmin = fminbnd(@mysimplefunc, -20, 20)

ymin = mysimplefunc(xmin)
```

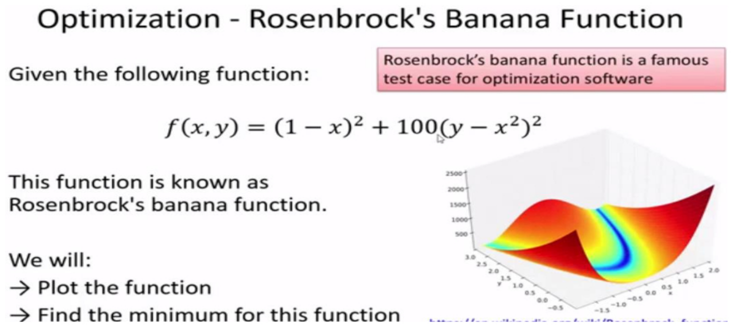
```
    0.5060    0.5128    0.5197    0.5266    0.5335    0.5405    0.5475    0.5546    0.5617    0.5688    0.5760    0.5832    0.5905    0.5978    0.6051    0.6125
  Columns 337 through 352
    0.6199    0.6274    0.6349    0.6424    0.6500    0.6576    0.6653    0.6730    0.6807    0.6885    0.6963    0.7042    0.7121    0.7200    0.7280    0.7360
  Columns 353 through 368
    0.7441    0.7522    0.7603    0.7685    0.7767    0.7850    0.7933    0.8016    0.8100    0.8184    0.8269    0.8354    0.8439    0.8525    0.8611    0.8698
  Columns 369 through 384
    0.8785    0.8872    0.8960    0.9048    0.9137    0.9226    0.9315    0.9405    0.9495    0.9586    0.9677    0.9768    0.9860    0.9952    1.0045    1.0138
  Columns 385 through 400
    1.0231    1.0325    1.0419    1.0514    1.0609    1.0704    1.0800    1.0896    1.0993    1.1090    1.1187    1.1285    1.1383    1.1482    1.1581    1.1680
  Column 401
    1.1780

xmin =
    -5

ymin =
   -72
fx >>
```





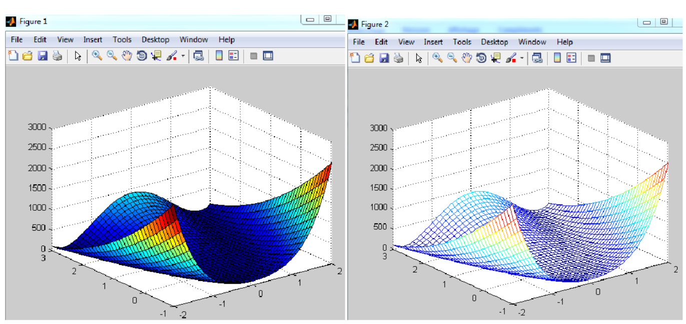
4.3 TP No. 03

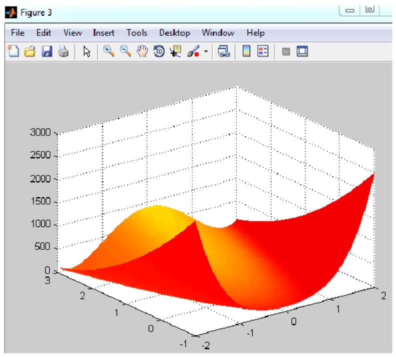
TP3RABAH



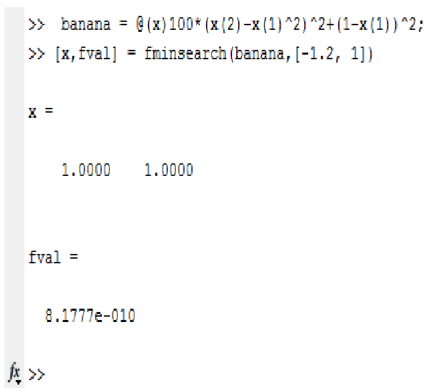

```
clear,clc

[x,y] = meshgrid(-2:0.1:2, -1:0.1:3);

f = (1-x).^2 +100.*(y-x.^2).^2;

figure(1)
surf(x,y,f)

figure(2)
mesh(x,y,f)

figure(3)
surfl(x,y,f)
shading interp;
colormap(hot);
```

Banana_plot.m

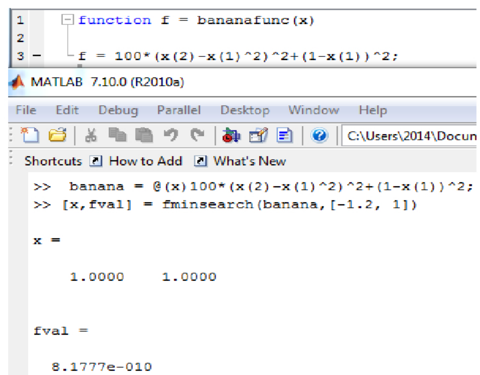




Méthode 1 et 2



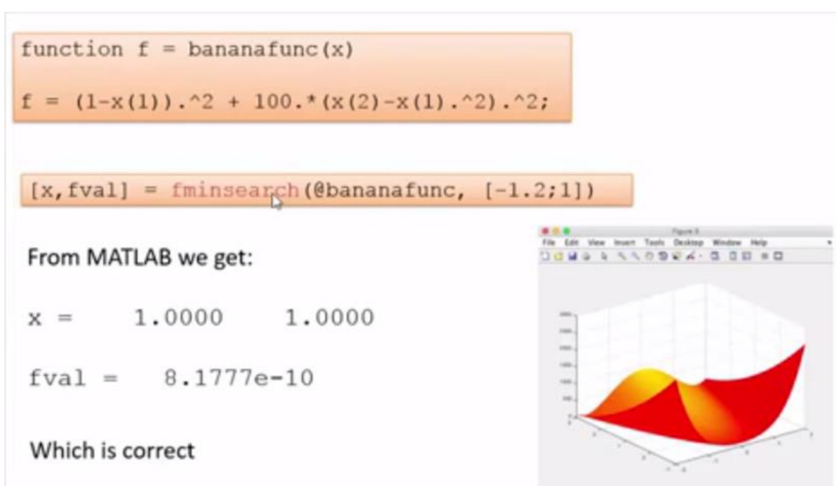
```
>> banana = @(x)100*(x(2)-x(1)^2)^2+(1-x(1))^2;
>> [x,fval] = fminsearch(banana,[-1.2, 1])

x =

    1.0000    1.0000


fval =

  8.1777e-010

>>
```



```
1   function f = bananafunc(x)
2
3 -   f = 100*(x(2)-x(1)^2)^2+(1-x(1))^2;
```



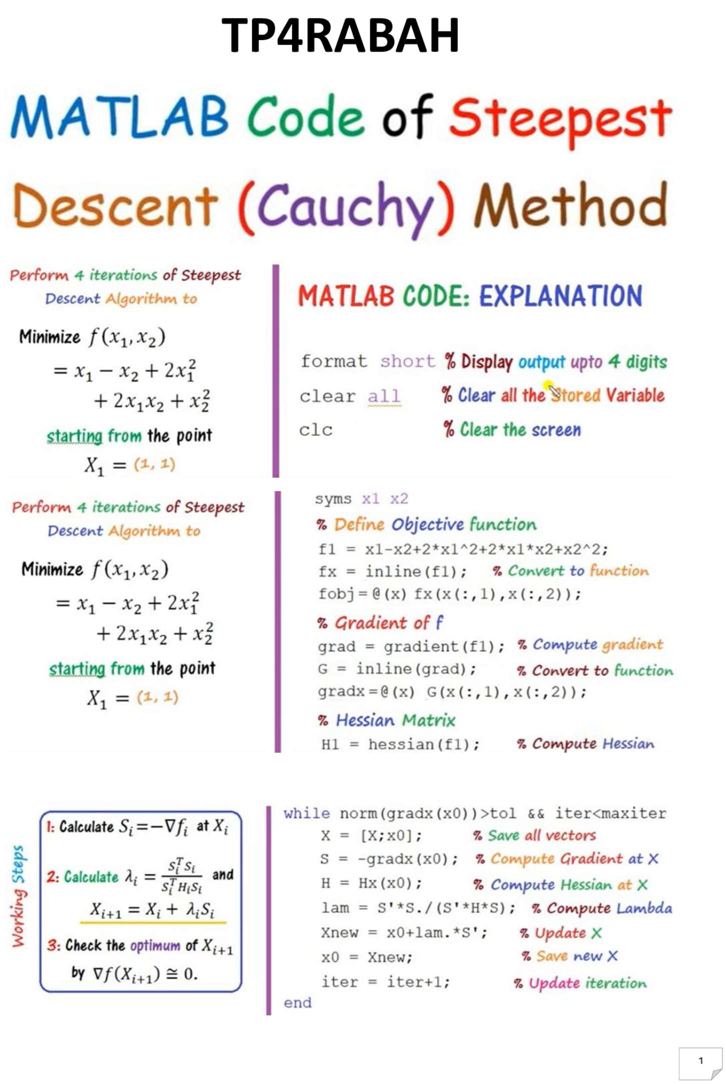
```
>> banana = @(x)100*(x(2)-x(1)^2)^2+(1-x(1))^2;
>> [x,fval] = fminsearch(banana,[-1.2, 1])

x =

    1.0000    1.0000


fval =

  8.1777e-010
```



2

4.4 TP No. 04

(see [7])

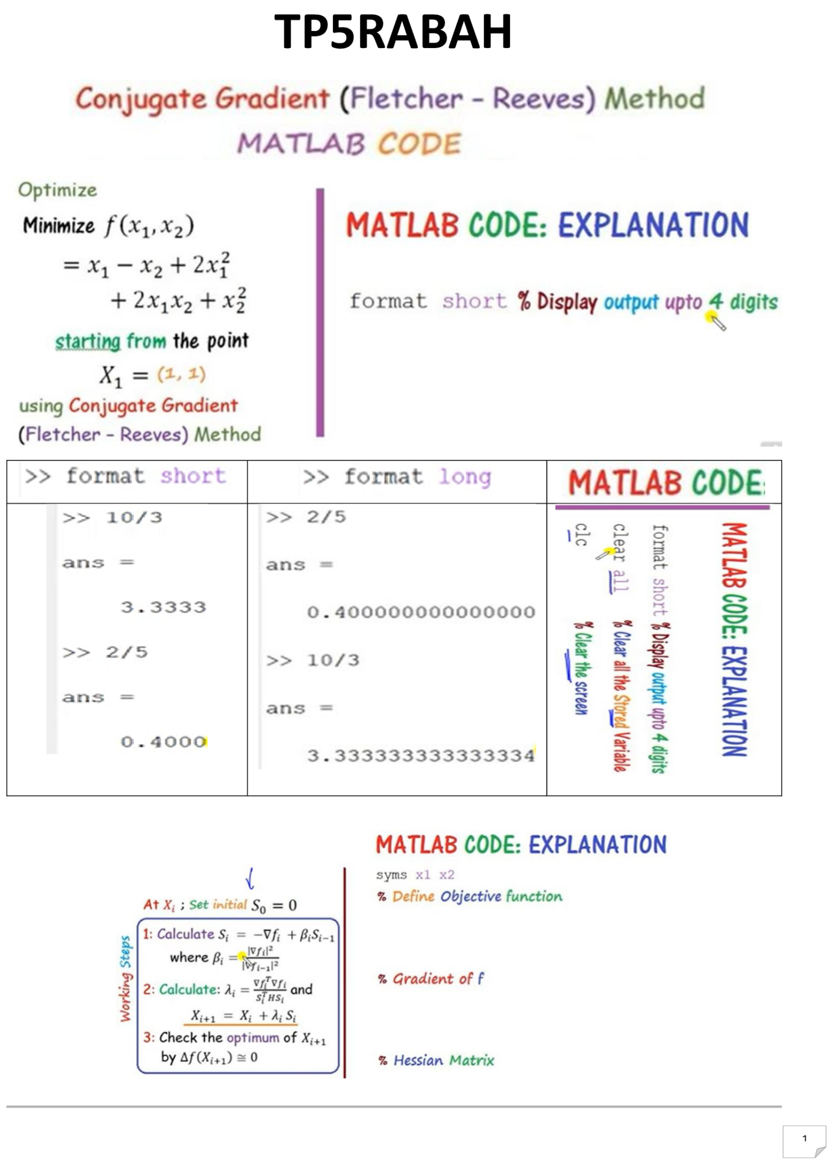


```
%%% PRINT the Solution
fprintf('Optimal Solution x = [%f, %f]\n',x0(1), x0(2));
fprintf('Optimal value f(x) = %f \n',fobj(x0));

========== END of the CODE =============
```

```
Optimal Solution x = [-0.981216, 1.495304]
Optimal value f(x) = -1.249449
>> X

X =

    1.0000    1.0000
   -0.3624    0.4161
   -0.8062    1.4515
   -0.9382    1.3950

>>
```

4.5 TP No. 05

(see [7])



```
syms x1 x2
% Define Objective function
f1 = x1-x2+2*x1^2+2*x1*x2+x2^2;
```

```
>> f1

f1 =

2*x1^2 + 2*x1*x2 + x1 + x2^2 - x2
```

<table>
<tr><td><pre>% Gradient of f
grad = gradient(f1); % Compute gradient</pre></td><td><pre>% Hessian Matrix
H1 = hessian(f1);   % Compute Hessia</pre></td></tr>
<tr><td><pre>>> gradient(f1)

ans =

 4*x1 + 2*x2 + 1
 2*x1 + 2*x2 - 1

>> grad=gradient(f1)

grad =

 4*x1 + 2*x2 + 1
 2*x1 + 2*x2 - 1</pre></td><td><pre>>> hessian(f1)

ans =

[4, 2]
[2, 2]</pre></td></tr>
<tr><td><pre>% Define Objective function
f1 = x1-x2+2*x1^2+2*x1*x2+x2^2;
fx = inline(f1);   % Convert to function
fobj=@(x) fx(x(:,1),x(:,2));</pre></td><td><pre>>> fx=inline(f1)

fx =

     Inline function:
     fx(x1,x2) = x1-x2+x1.*x2.*2.0+x1.^2.*2.0+x2.^2

>> fx(0,1)

ans =

     0
>> fx(2,7)

ans =

    80

>> fx(2,-6)

ans =

    28
>> fobj

fobj =

     @(x)fx(x(:,1),x(:,2))</pre></td></tr>
<tr><td><pre>% Gradient of f
grad = gradient(f1); % Compute gradient
G = inline(grad);   % Convert to function
gradx=@(x) G(x(:,1),x(:,2));</pre></td><td><pre>>> grad
grad =
4*x1 + 2*x2 + 1
2*x1 + 2*x2 - 1</pre></td></tr>
</table>

`% Hessian Matrix`
`H1 = hessian(f1);   % Compute Hessian`
`Hx = inline(H1);    % Convert to function`	

`x0 = [1 1];   % Set initial vector`
`maxiter = 4;   % Set maximum iteratio`
`tol = 1e-3;   % maximum tolerance`
`iter = 1;   % initial counter`
`X = [];   % initial vector array empty`
	`>> x0`
	`x0 =`
	`-1.0000    1.5000`
	`>> S`
	`S =`
	`-0.4964`
	`0.8438`

Algorithm

At X_i ; Set initial $S_0 = 0$

Working Steps

1: Calculate $S_i = -\nabla f_i + \beta_i S_{i-1}$ where $\beta_i = \frac{|\nabla f_i|^2}{|\nabla f_{i-1}|^2}$

2: Calculate: $\lambda_i = \frac{\nabla f_i^T \nabla f_i}{S_i^T H S_i}$ and $X_{i+1} = X_i + \lambda_i S_i$

3: Check the optimum of X_{i+1} by $\Delta f(X_{i+1}) \cong 0$

MATLAB CODE: EXPLANATION

```
S = 0;                       % Initial Search Direction
Gpr = -gradx(x0);            % Compute initial ∇f_{i-1}
while norm(gradx(x0))>tol && iter<maxiter
    X = [X;x0];              % Save all vectors
    Gi = -gradx(x0);         % Compute Gradient at X
    H = Hx(x0);              % Compute Hessian at X
    bet = norm(Gi).^2./norm(Gpr).^2;
    S = Gi + bet.*S;         % Compute direction "S"
    lam = Gi'*Gi./(S'*H*S);  % Compute Lambda
    Xnew = x0+lam.*S';       % Update X
    x0 = Xnew;               % Save new X
    Gpr = Gi;                % Update ∇f_{i-1}
    iter = iter+1;           % Update iteration
end
```

```
%%% PRINT the Solution
fprintf('Optimal Solution x = [%f, %f]\n',x0(1), x0(2));
fprintf('Optimal value f(x) = %f \n',fobj(x0));

========== END of the CODE =============

%%% Conjugate Gradient (Flecter-Reeves) METHOD (Quadratic function only)
%%%  MATLAB CODE
```

```
format short
clear all
clc

syms x1 x2
x0 = [1 1];
tol = 1e-3;
maxiter = 4;
    % Objective function:
%     f1 = x1.^2-x1.*x2+3.*x2.^2;
```

```
    % Gradient of f
    grad = gradient(f1);
    G = inline(grad);
    gradx = @(x) G(x(:,1), x(:,2));
% Hessian matrix
    H1 = hessian(f1);
    Hx = inline(H1);


%%% MAIN CODE Fleetcher-Reeves method
X = [];
S = 0;       % initial S_0 = 0
iter = 1;        %for iteration
Gpr = -gradx(x0);        % initial Gradient at i-1
if norm(Gpr)==0
```

```
     disp('Change x0');
     x0 = input('Provide New x0=');
     Gpr = -gradx(x0);
 end
 while norm(gradx(x0))>tol && iter<maxiter
     X = [X;x0];
     Gi = -gradx(x0);
```

```
Optimal Solution x = [-1.000000, 1.500000]
Optimal value f(x) = -1.250000
>> X

X =

    1.0000    1.0000
   -0.3624    0.4161
```

```
Change x0
Provide New x0=[0 0.6]
Optimal Solution x = [-1.000000, 1.500000]
Optimal value f(x) = -1.250000
>> X

X =

         0    0.6000
   -0.5064    0.5540
```

```
Change x0
Provide New x0=[ 9 2]
Optimal Solution x = [-1.000000, 1.500000]
Optimal value f(x) = -1.250000
>> X

X =

    9.0000    2.0000
    1.1265   -2.0328
```

4

Chapter 5

Tutorials

5.1 TD Series No. 01

Exercise 5.1.1. *1. Calculate the gradient of $f(x,y,z)$ in the following cases.*

a. $f(x,y,z) = x^2 + y^3 + z^4$.

b. $f(x,y,z) = x^2 y^3 z^4$.

c. $f(x,y,z) = e^x \sin y \ln z$.

2. Determine the stationary points of the function f of two variables defined by

$$f(x,y) = x(x+1)^2 - y^2.$$

3. Calculate the derivative or gradient of (gof) by two methods in the following cases

a. $f(x,y) = \exp(x) + \cos(y), \quad g(x) = 4x + 1.$

b. $f(x) = (\exp(x), \cos(x)), \quad g(x,y) = 4x + 2y.$

Exercise 5.1.2. *1. Show that*

a.

$$\nabla(f.g) = g.\nabla f + f.\nabla g$$

b.

$$\nabla\left(\frac{f}{g}\right) = \frac{g.\nabla f - f.\nabla g}{g^2}$$

2. Show the following equality

$$\nabla^2 f(x)h = \nabla\langle\nabla f(x), h\rangle; \quad x \in Df \subset \mathbb{R}^n \quad \forall h \in \mathbb{R}^n.$$

Exercise 5.1.3. *1. Calculate the directional derivative of* $f(x,y) := e^{xy^2}$ *at the point* $(1,2)$ *in the direction forming a angle of 30° with the positive x-axis.*
2. Let $T(x,y) = x^3 + y^2 - 2xy + 1$ *be the temperature at point (x,y). In which direction to the point (1, 3), the temperature T*

a. *is it increasing the fastest and at what rate ?*

b. *is it decreasing the fastest and at what rate ?*

Exercise 5.1.4. *Determine the Taylor expansion of the following functions*

a. $f(x,y) = -\cos x \cos y$ *in* $(0,0)$ *and* $(\frac{\pi}{2}, \frac{\pi}{2})$ *to order "2"*

b. $f(x,y) = e^x \cos y$ *in* $(0,0)$ *to order "2"*

Exercise 5.1.5. *Calculate the directional derivative of the following functions at the points indicated.*

a. $f(x,y) = x + y$ *in* $(0,0)$ *and* $d = (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})^T$.

b. $f(x,y) = x + y^2 + 2 \quad in(1,-2) \ and \ d = (3,-4)^T.$

c. $f(x,y) = e^x cosy \quad in(0,0) \ and \ d = (-1,1)^T.$

Exercise 5.1.6. *Calculate the gradient, the Hessian matrix and the Directional derivative*

1. $f_1 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_1(x) = a.$

2. $f_2 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_2(x) = \langle a, x\rangle + b \quad a \in \mathbb{R}^n, b \in \mathbb{R}.$

3. $f_3 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_3(x) = a\langle b, x\rangle + c \quad b \in \mathbb{R}^n, a \ and \ c \in \mathbb{R}.$

4. $f_4 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_4(x) = a\langle x, x\rangle + b \quad a \ and \ b \in \mathbb{R}.$

5. $f_5 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_5(x) = \sum_{i=1}^{m} g_i(x)$ *such as*
$g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ *is twice differentiable.*

6. $f_6 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_6(x) = \sum_{i=1}^{m} (g_i(x))^2$ *such as*
$g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ *is twice differentiable.*

Exercise 5.1.7. *we assume that it exists* $L > 0$ *such that* $\forall x, y \in \mathbb{R}^n$, *we have*
$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$ *i.e.*∇f *is Lipschitzian or* f *is class* $\mathscr{C}^1(\mathbb{R}^n)$
Then

$$|f(x+h) - f(x) - \langle \nabla f(x), h\rangle| \leq \frac{L}{2}\|h\|^2 \quad \forall x, h \in \mathbb{R}^n$$

5.2 TD Series No. 02

Exercise 5.2.1. *Show that a norm is convex.*

Exercise 5.2.2. *Show that the indicator function; of a set* Ω *defined by*

$$1_\Omega = \begin{cases} 0 & if \quad x \in \Omega \\ +\infty & if \quad x \notin \Omega \end{cases}$$

is convex if and only if Ω *is convex.*

Exercise 5.2.3. *Let* U *be a convex part of a vector space* V*. Show that* $f : U \subset V \longrightarrow \mathbb{R}$ *is convex if and only if the following set:*

$$epi(f) = \{(v, \alpha) \subset U \times \mathbb{R} \diagup \alpha \geq f(v)\}$$

is a convex part of $U \times \mathbb{R}$*.*

Exercise 5.2.4. *Let* F *be a function from* $\mathbb{R}^n$ *in* $\mathbb{R}$*. we define the following function from* $\mathbb{R}_+^*$ *to* $\mathbb{R}$*:*

$$\forall \alpha > 0, \quad \forall (u, v) \in \mathbb{R}^n \times \mathbb{R}^n \qquad \Phi(\alpha) = \frac{F(u + \alpha v) - F(u)}{\alpha}$$

Show that if F *is convex then* Φ *is increasing.*

Exercise 5.2.5. *Let* $(f_i)_{i \in I}$ *be any family of convex functions of* $U \subset V \to \mathbb{R}$*. Prove that the function* $\sup\limits_{x \in \mathbb{R}^n} f_i$ *is convex.*

Exercise 5.2.6. *Show Young's inequality* $\forall a, b > 0 \quad \forall p, q \in \mathbb{N}$ *such as* $\frac{1}{p} + \frac{1}{q} = 1$

$$ab \leq \frac{1}{p} a^p + \frac{1}{q} b^q$$

Exercise 5.2.7. *Let* f *be a convex function from* $\mathbb{R}^n$ *to* $\mathbb{R}$*. To show that:*

$\forall (\lambda_i)_{1 \leq i \leq p} \in (\mathbb{R}^n)^p$ *such as* $\sum\limits_{i=1}^{p} \lambda_i = 1, \quad \forall (x_i)_{1 \leq i \leq p} \in (\mathbb{R}^n)^p; f\left(\sum\limits_{i=1}^{p} \lambda_i x_i\right) \leq \sum\limits_{i=1}^{p} \lambda_i f(x_i).$

Exercise 5.2.8. *(Characterization of convexity)*

Let $\Omega \in \mathbb{R}^n$ *be an open one* $U \subset \Omega$ *with* U *convex and* $f : \Omega \to \mathbb{R}$ *R a function of class* $\mathscr{C}^1$. *Then the following 3 propositions are equivalent: 1.* f *is convex on* U

2. $f(y) \geq f(x) + \langle \nabla f(x); y - x \rangle \quad \forall x, y \in U$

3. ∇f *is monotonous on* U

Exercise 5.2.9. *Let* f *is of class* $\mathscr{C}^2$ *then* f *is convex on* U *(convex) if and only if*

$$\langle \nabla^2 f(x)(y - x); y - x \rangle; \quad \forall x, y \in U$$

5.3 TD Series No. 03

Exercise 5.3.1. *Show that if $\widehat{x}$ is a max (local or global) of f, then $\widehat{x}$ is a min (local or global) of* $-f$

Exercise 5.3.2. *Are the following functions coercive?*

1. $f_1 : \mathbb{R} \longrightarrow \mathbb{R}; x \mapsto f_1(x) = x^3 - x^2 + 5.$

2. $f_2 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_2(x) = \langle a, x\rangle + b \quad a \in \mathbb{R}^n, b \in \mathbb{R}.$

3. $f_3 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_3(x) = a\langle x, x\rangle + b \quad a \text{ and } b \in \mathbb{R}.$

4. $f_4 : \mathbb{R}^2 \longrightarrow \mathbb{R}; x \mapsto f_4(x) = 2x_1^2 + x_2 - 5$

5. $f_5 : \mathbb{R}^2 \longrightarrow \mathbb{R}; x \mapsto f_5(x) = x_1^2 + 2x_2^3 + x_2^2 - x_1$

6. $f_6 : \mathbb{R}^2 \longrightarrow \mathbb{R}; x \mapsto f_6(x) = x_1^2 + 2x_1 + x_2^2$

7. $f_7 : \mathbb{R}^2 \longrightarrow \mathbb{R}; x \mapsto f_7(x) = x_1^2 + x_2^2 - 3x_2 - 5$

8. $f_8 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_8(x) = \langle x, x\rangle + \langle a, x\rangle + b \quad a \in \mathbb{R}^n, b \in \mathbb{R}$

Exercise 5.3.3. *We consider the function f defined on $\mathbb{R}^2$ by*

$$f(x, y) = x^4 + y^4 - 2(x - y)^2$$

1. Show that there exists $(\alpha, \beta) \in \mathbb{R}_+^2$ such that

$$f(x, y) \geq \alpha \|(x, y)\|^2 + \beta \quad (x, y) \in \mathbb{R}^2$$

Deduce that the following problem has at least one solution,

$$(P_1)\, min_{(x,y)\in\mathbb{R}^2} f(x, y)$$

f is it convex on$\mathbb{R}^2$?

3. Solve the problem (P_1).

Exercise 5.3.4. *Let$f : \mathbb{R}^2 \longrightarrow \mathbb{R}; x \longmapsto f(x,y) = x^2 + y^2 + ax + by + c$*

We consider the problem

$$(P_2) min_{(x,y)\in\mathbb{R}^2} f(x,y)$$

1) Show that f is elliptical.

2) Solve the problem (P_2).

Exercise 5.3.5. *Consider a cloud of n points $M_i(t_i, x_i) \in \mathbb{R}^2 \quad i = 1,2,...,10$ given by the table following*

t_i	1	2	3	4	5	6	7	8	9	10	$\sum_{i=1}^{10} t_i =$
x_i	0	-3	6	-3	6	3.8	5	-2	1.4	8	$\sum_{i=1}^{10} x_i =$
t_i^2											$\sum_{i=1}^{10} t_i^2 =$

We are looking for the regression line of this cloud. For this we use the method of least squares, as we do not have $x_i = at_i + b$ for all $i = 1,2,...,10$, we seek to minimize the square of differences. We therefore want to find a pair of reals (a,b) solution of

$$(P_3) = \begin{cases} min \mathscr{J}(a,b) \\ (a,b) \in \mathbb{R}^2 \end{cases}$$

Or $\mathscr{J}(a,b) = \sum_{i=1}^{10} (x_i - at_i - b)^2$.

1. Complete the table.

2. Calculate the gradient and the Hessian matrix of the function$\mathscr{J}$.

3. Does the problem (P_3) have a solution? Is it unique?

4. Solve the problem (P_3) , deduce the equation of the regression line.

Exercise 5.3.6. *We consider the following minimization problem*

$$(P_4) = \begin{cases} min \mathscr{J}(v) \\ v \in \mathbb{R}^n \end{cases}$$

Or $\mathscr{J}(v) = \dfrac{1}{2}\langle Av, v\rangle - \langle b, v\rangle$. *and* A *is a positive definite symmetric matrix of* $\mathbb{R}^n$ *in* $\mathbb{R}^n$ *and* $v \in \mathbb{R}^n$.

1. Demonstrate that

a. *The function* $\mathscr{J}$ *is strictly convex.*

b. $\mathscr{J}$ *is a coercive function.*

2. Calculate the gradient and the Hessian matrix of the function $\mathscr{J}$.

3. Show that the problem (P_4) *admits a single solution.*

4. Solve the problem (P_4), *deduce the minimum value of* $\mathscr{J}$.

Chapter 6

Corrected Tutorials

6.1 TD Series No. 01 Corrected

Answer 6.1.1. **a.** $f(x,y,z) = x^2 + y^3 + z^4$.

$$\nabla f(x,y,z) = \begin{pmatrix} \frac{\partial f}{\partial x}(x,y,z) \\ \frac{\partial f}{\partial y}(x,y,z) \\ \frac{\partial f}{\partial z}(x,y,z) \end{pmatrix} = \begin{pmatrix} 2x \\ 3y \\ 4z \end{pmatrix}$$

b. $f(x,y,z) = x^2 y^3 z^4$.

$$\nabla f(x,y,z) = \begin{pmatrix} \frac{\partial f}{\partial x}(x,y,z) \\ \frac{\partial f}{\partial y}(x,y,z) \\ \frac{\partial f}{\partial z}(x,y,z) \end{pmatrix} = \begin{pmatrix} 2xy^3z^4 \\ 3x^2y^2z^4 \\ 4x^2y^3z^3 \end{pmatrix}$$

c. $f(x,y,z) = e^x \sin y \ln z$.

$$\nabla f(x,y,z) = \begin{pmatrix} \frac{\partial f}{\partial x}(x,y,z) \\ \frac{\partial f}{\partial y}(x,y,z) \\ \frac{\partial f}{\partial z}(x,y,z) \end{pmatrix} = \begin{pmatrix} e^x \sin y \ln z \\ e^x \cos y \ln z \\ \frac{e^x \sin y}{z} \end{pmatrix}$$

2. $f(x,y) = x(x+1)^2 - y^2$

$$\nabla f(x,y) = \begin{pmatrix} \frac{\partial f}{\partial x}(x,y \\ \frac{\partial f}{\partial y}(x,y) \end{pmatrix} = \begin{pmatrix} 3x^2+4x+1 \\ -2y \end{pmatrix}$$

$$\nabla f(x,y) = 0 \Longrightarrow \begin{pmatrix} 3x^2+4x+1 \\ -2y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$(x,y) = (-1,0) \vee (\frac{-1}{3},0)$$

3. Calculate the derivative or gradient of (gof) *by two methods in the following cases*

a. $f(x,y) = \exp(x) + \cos(y), \quad g(x) = 4x+1.$

$$\begin{aligned} (gof)(x,y) &= g(f(x,y)) \\ &= g(\exp(x)+\cos(y)) \\ &= 4(\exp(x)+\cos(y))+1. \end{aligned}$$

$$\begin{aligned}\nabla(gof)(x,y) &= g'(f(x,y))\nabla f(x,y)\\ &= g'(f(x,y))\begin{pmatrix}\frac{\partial f}{\partial x}(x,y)\\ \frac{\partial f}{\partial y}(x,y)\end{pmatrix}\\ &= g'(\exp(x)+\cos(y))\begin{pmatrix}\exp(x)\\ -\sin(y)\end{pmatrix}, \quad (g'(x)=4)\\ &= 4\begin{pmatrix}\exp(x)\\ -\sin(y)\end{pmatrix}.\end{aligned}$$

2nd method

$$\begin{aligned}\nabla(gof)(x,y) &= \begin{pmatrix}\frac{\partial gof}{\partial x}(x,y)\\ \frac{\partial gof}{\partial y}(x,y)\end{pmatrix}\\ &= \begin{pmatrix}4\exp(x)\\ -4\sin(y)\end{pmatrix}.\end{aligned}$$

b. $f(x) = (\exp(x), \cos(x)), \quad g(x,y) = 4x+2y.$

$$\begin{aligned}(gof)(x) &= g(f(x))\\ &= g(f_1(x), f_2(x))\\ &= g(\exp(x), \cos(x))\\ &= 4\exp(x)+2\cos(x).\end{aligned}$$

$$(gof)'(x) = 4\exp(x) - 2\sin(x)$$

2nd method

$$\begin{aligned}
(gof)'(x) &= f'(x)\nabla g(f(x)) \\
&= (f_1'(x), f_1'(x)) \begin{pmatrix} \dfrac{\partial g}{\partial x}(f(x)) \\ \dfrac{\partial g}{\partial y}(f(x)) \end{pmatrix} f'(x) \\
&= (\exp(x), -\sin(x)) \begin{pmatrix} 4 \\ 2 \end{pmatrix} \\
&= 4\exp(x) - 2\sin(x).
\end{aligned}$$

Answer 6.1.2. *1.*

a.

$$\begin{aligned}
\nabla(f.g) &= \begin{pmatrix} \dfrac{\partial(f.g)}{\partial x_1}(x) \\ \dfrac{\partial(f.g)}{\partial x_2}(x) \\ \vdots \\ \dfrac{\partial(f.g)}{\partial x_n}(x) \end{pmatrix} \\
&= g \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(x) \\ \dfrac{\partial f}{\partial x_2}(x) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(x) \end{pmatrix} + f \begin{pmatrix} \dfrac{\partial g}{\partial x_1}(x) \\ \dfrac{\partial g}{\partial x_2}(x) \\ \vdots \\ \dfrac{\partial g}{\partial x_n}(x) \end{pmatrix} \\
&= g.\nabla f + f.\nabla g
\end{aligned}$$

b.

$$\nabla\left(\frac{f}{g}\right)=\begin{pmatrix}\dfrac{\partial\left(\frac{f}{g}\right)}{\partial x_1}(x)\\ \dfrac{\partial\left(\frac{f}{g}\right)}{\partial x_2}(x)\\ \vdots\\ \dfrac{\partial\left(\frac{f}{g}\right)}{\partial x_n}(x)\end{pmatrix}=\begin{pmatrix}\dfrac{g\frac{\partial f}{\partial x_1}-f\frac{\partial g}{\partial x_1}}{g^2}(x)\\ \dfrac{g\frac{\partial f}{\partial x_2}-f\frac{\partial g}{\partial x_2}}{g^2}(x)\\ \vdots\\ \dfrac{g\frac{\partial f}{\partial x_n}-f\frac{\partial g}{\partial x_n}}{g^2}(x)\end{pmatrix}=\frac{1}{g^2}\begin{pmatrix}g\frac{\partial f}{\partial x_1}(x)-f\frac{\partial g}{\partial x_1}(x)\\ g\frac{\partial f}{\partial x_2}(x)-f\frac{\partial g}{\partial x_2}(x)\\ \vdots\\ g\frac{\partial f}{\partial x_n}(x)-f\frac{\partial g}{\partial x_n}(x)\end{pmatrix}$$

$$\vdots$$

$$=\frac{g.\nabla f-f.\nabla g}{g^2}$$

2.

$$\nabla^2 f(x)h=\nabla\langle\nabla f(x),h\rangle;\quad x\in Df\subset\mathbb{R}^n\quad\forall h\in\mathbb{R}^n.$$

$$\begin{aligned}\nabla^2 f(x)h&=\nabla\nabla^T f(x)h\\&=\nabla\langle\nabla f(x),h\rangle\end{aligned}$$

Answer 6.1.3. 1. $f:\mathbb{R}^2\longrightarrow\mathbb{R}$, *and let* $v\in\mathbb{R}^2$ *be a unit vector.*

$v=r(\cos 30^\circ i+\sin 30^\circ j)=r(\frac{\sqrt{3}}{2}i+\frac{1}{2})$

be a unit vector $\Longrightarrow\quad r=1$

$$\nabla f(x,y)=\begin{pmatrix}\frac{\partial f}{\partial x}(x,y)\\ \frac{\partial f}{\partial y}(x,y)\end{pmatrix}=\begin{pmatrix}y^2e^{xy^2}\\ 2ye^{xy^2}\end{pmatrix}$$

$$\begin{aligned}D_v f(x)&=\nabla f(x)\cdot v\\&=\frac{\sqrt{3}}{2}\frac{\partial f}{\partial x}(1.2)+\frac{\sqrt{1}}{2}\frac{\partial f}{\partial y}(1.2)\\&=2e^4(\sqrt{3}+1).\end{aligned}$$

2.

$$\nabla T(x,y)=\begin{pmatrix}\dfrac{\partial T}{\partial x}(x,y)\\ \dfrac{\partial T}{\partial y}(x,y)\end{pmatrix}=\begin{pmatrix}3x^2-2y\\ 2y-2x\end{pmatrix}$$

$$\nabla T(1,3)=\begin{pmatrix}-3\\ 4\end{pmatrix}$$

a. *increasing the fastest*

$$\frac{\nabla T(1,3)}{\|\nabla T(1,3)\|}=\begin{pmatrix}\dfrac{-3}{5}\\ \dfrac{4}{5}\end{pmatrix}\quad \textit{and the rate}\quad \|\nabla T(1,3)\|=5$$

b. *decreasing the fastest*

$$-\frac{\nabla T(1,3)}{\|\nabla T(1,3)\|}=\begin{pmatrix}\dfrac{3}{5}\\ \dfrac{-4}{5}\end{pmatrix}\quad \textit{and the rate}\quad -\|\nabla T(1,3)\|=-5$$

Answer 6.1.4. a. $f(x,y)=-\cos x\cos y$ *in* $(0,0)$ *and* $(\frac{\pi}{2},\frac{\pi}{2})$ *to order "2"*

$$f(x,y)=f(0,0)+x\frac{\partial f}{\partial x}(0,0)+y\frac{\partial f}{\partial y}(0,0)+\frac{x^2}{2}\frac{\partial^2 f}{\partial^2 x}(0,0)+\frac{y^2}{2}\frac{\partial^2 f}{\partial^2 y}(0,0)+\frac{xy}{2}\frac{\partial^2 f}{\partial x\partial y}(0,0)+(x^2+y^2)\varepsilon(x,y)$$

$$f(x,y)=-1+\frac{x^2}{2}+\frac{y^2}{2}+(x^2+y^2)\varepsilon(x,y)\quad \textit{such that}\quad \varepsilon(x,y)\underset{(x,y)\longrightarrow(0,0)}{\longrightarrow}0$$

$$f(x+\frac{\pi}{2},y+\frac{\pi}{2})=-xy+(x^2+y^2)\varepsilon(x,y)\quad \textit{such that}\quad \varepsilon(x,y)\underset{(x,y)\longrightarrow(0,0)}{\longrightarrow}0$$

b. $f(x,y)=e^x\cos y$ *in* $(0,0)$ *to order "2"*

$$f(x,y)=1+x+\frac{x^2}{2}-\frac{y^2}{2}+(x^2+y^2)\varepsilon(x,y)\quad \textit{such that}\quad \varepsilon(x,y)\underset{(x,y)\longrightarrow(0,0)}{\longrightarrow}0$$

Answer 6.1.5. **a.** $f(x,y) = x + y \quad in\ (0,0)\ and\ d = (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})^T.$

$v in \mathbb{R}^2$ *be a unit vector* $(\| \| = 1)$

$$\begin{aligned}
D_v f(x) &= \frac{d}{dt} f(x+tv)\Big|_{t_0} \\
&= \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t} \\
&= \lim_{t\to 0} \frac{f(\frac{\sqrt{2}}{2}t, -\frac{\sqrt{2}}{2}t) - f(0,0)}{t} \\
&= \lim_{t\to 0} \frac{\frac{\sqrt{2}}{2}t - \frac{\sqrt{2}}{2}t - 0}{t}
\end{aligned}$$

2nd method

$$\begin{aligned}
D_v f(0,0) &= \langle \nabla f(0.0).v \rangle \\
&= 1.\frac{\sqrt{2}}{2} + 1(-\frac{\sqrt{2}}{2}) \\
&= 0.
\end{aligned}$$

b. $f(x,y) = x + y^2 + 2 \quad in (1,-2)\ and\ d = (3,-4)^T.$

$v = \frac{d}{\|d\|} = (\frac{3}{5}, \frac{-4}{5})^T$

$$\begin{aligned}
D_v f(1,-2) &= \langle \nabla f(1,-2).v \rangle \\
&= \frac{3}{5}.1 + \frac{-4}{5}(-4) \\
&= \frac{19}{5}.
\end{aligned}$$

c. $f(x,y) = e^x cos y \quad in (0,0)\ and\ d = (-1,1)^T.$

$v = \frac{d}{\|d\|} = (\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T$

$$
\begin{aligned}
D_v f(0,0) &= \langle \nabla f(0,0).v \rangle \\
&= \frac{-1}{\sqrt{2}}.\frac{1}{\sqrt{2}} + \frac{-1}{\sqrt{2}}.0 \\
&= \frac{-1}{2}.
\end{aligned}
$$

Answer 6.1.6. **1.** $f_1 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \longmapsto f_1(x) = a.$

$$
\nabla f_1(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) \\ \frac{\partial f_1}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f_1}{\partial x_n}(x) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0, \quad 0 \in \mathbb{R}^n
$$

$$
H(x) = \nabla^2 f_1(x) = \begin{pmatrix} \frac{\partial^2 f_1}{\partial x_1 \partial x_1}(x) & \frac{\partial^2 f_1}{\partial x_1 \partial x_2}(x) & \dots & \frac{\partial^2 f_1}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f_1}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f_1}{\partial x_2 \partial x_2}(x) & \dots & \frac{\partial^2 f_1}{\partial x_2 \partial x_n}(x) \\ & \vdots & & \\ \frac{\partial^2 f_1}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f_1}{\partial x_n \partial x_2}(x) & \dots & \frac{\partial^2 f_1}{\partial x_n \partial x_n}(x) \end{pmatrix} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & 0 \end{pmatrix}.
$$

2. $f_2 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \longmapsto f_2(x) = \langle a, x \rangle + b \quad a \in \mathbb{R}^n, b \in \mathbb{R}.$

$f_2(x) = \sum_{i=1}^{n} a_i x_i + b$

$$
\nabla f_2(x) = \begin{pmatrix} \frac{\partial f_2}{\partial x_1}(x) \\ \frac{\partial f_2}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f_2}{\partial x_n}(x) \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a, \quad a \in \mathbb{R}^n
$$

$$H(x) = \nabla^2 f_2(x) = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

3. $f_3 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_3(x) = a\langle b, x\rangle + c \quad b \in \mathbb{R}^n, a \text{ and } c \in \mathbb{R}$.

$$f_3(x) = a\sum_{i=1}^{n} b_i x_i + c$$

$$\nabla f_3(x) = \begin{pmatrix} \frac{\partial f_3}{\partial x_1}(x) \\ \frac{\partial f_3}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f_3}{\partial x_n}(x) \end{pmatrix} = \begin{pmatrix} ab_1 \\ ab_2 \\ \vdots \\ ab_n \end{pmatrix} = ab \quad a \in \mathbb{R}, \quad b \in \mathbb{R}^n$$

$$H(x) = \nabla^2 f_3(x) = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

4. $f_4 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_4(x) = a\langle x, x\rangle + b \quad a \text{ and } b \in \mathbb{R}$.

$$f_4(x) = a\sum_{i=1}^{n} x_i^2 + b$$

$$\nabla f_4(x) = \begin{pmatrix} \frac{\partial f_4}{\partial x_1}(x) \\ \frac{\partial f_4}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f_4}{\partial x_n}(x) \end{pmatrix} = \begin{pmatrix} 2ax_1 \\ 2ax_2 \\ \vdots \\ 2ax_n \end{pmatrix} = 2ax, \quad a \in \mathbb{R} \quad x \in \mathbb{R}^n.$$

$$H(x) = \nabla^2 f_4(x) = \begin{pmatrix} 2a & 0 & \dots & 0 \\ 0 & 2a & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & 2a \end{pmatrix}.$$

5. $f_5 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_5(x) = \sum_{i=1}^{m} g_i(x)$ *such as*

$g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ *is twice differentiable.*

$$f_5(x) = \sum_{i=1}^{m} g_i(x)$$

$$\nabla f_5(x) = \sum_{i=1}^{m} \nabla g_i(x)$$

$$H(x) = \nabla^2 f_5(x) = \sum_{i=1}^{m} \nabla^2 g_i(x)$$

6. $f_6 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_6(x) = \sum_{i=1}^{m} (g_i(x))^2$ *such as*

$g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ *is twice differentiable.*

$$\nabla f_6(x) = 2\sum_{i=1}^{m} g_i(x) \nabla g_i(x)$$

$$H(x) = \nabla^2 f_6(x) = 2\sum_{i=1}^{m} g_i(x) \nabla g_i(x)$$

Answer 6.1.7. $f(x+h) = f(x) + \int_0^1 \langle \nabla f(x+th), h\rangle dt$

$$\begin{aligned}
f(x+h) - f(x) - \left\langle \nabla f(x), h \right\rangle &= \int_0^1 \left\langle \nabla f(x+th) - \nabla f(x), h \right\rangle dt \\
\left| f(x+h) - f(x) - \left\langle \nabla f(x), h \right\rangle \right| &= \left| \int_0^1 \left\langle \nabla f(x+th) - \nabla f(x), h \right\rangle dt \right|. \\
&\leq \int_0^1 \left| \left\langle \nabla f(x+th) - \nabla f(x), h \right\rangle dt \right|. \\
&\leq \int_0^1 \left\| \nabla f(x+th) - \nabla f(x) \right\| \|h\| dt \\
&\leq \int_0^1 L \|x+th-x\| \|h\| dt \\
&= \int_0^1 Lt\|h\|^2 dt \\
&= L\|h\|^2 \int_0^1 t dt \\
&= \frac{L}{2}\|h\|^2.
\end{aligned}$$

6.2 TD Series No. 02 Corrected

Answer 6.2.1. *Let* $\|.\| : \mathbb{R}^n \to \mathbb{R}$ *be a norm, Then* $\|.\|$ *is said to be a convex if*

$$\|\lambda X_1 + (1-\lambda)X_2\| \le \lambda\|X_1\| + (1-\lambda)\|X_2\|$$

For all $X_1, X_2 \in S$ *and for each* $\lambda \in (0,1)$.

$$\begin{aligned}\|\lambda X_1 + (1-\lambda)X_2\| &\le \|\lambda X_1\| + \|(1-\lambda)X_2\| \quad \textit{(triangle inequality)}\\ &\le |\lambda|\|X_1\| + |(1-\lambda)|\|X_2\| \quad \textit{(positive homogeneity)}\\ &\le \lambda\|X_1\| + (1-\lambda)\|X_2\|.\end{aligned}$$

Answer 6.2.2. 1_Ω *is convex* $\Longrightarrow \Omega$ *is convex* ?

$$\begin{aligned}1_\Omega \quad \textit{is convex} &\Longrightarrow 1_\Omega(\lambda X_1 + (1-\lambda)X_2) \le \lambda 1_\Omega(X_1) + (1-\lambda)1_\Omega(X_2)\\ &\Longrightarrow \textit{For all} \quad X_1, X_2 \in \Omega \quad \left(1_\Omega(X_1) = 1_\Omega(X_2) = 0\right) \quad \textit{and for each} \quad \lambda \in [0,1]\\ &\Longrightarrow 0 \le 1_\Omega(\lambda X_1 + (1-\lambda)X_2) \le 0 + 0\\ &\Longrightarrow 1_\Omega(\lambda X_1 + (1-\lambda)X_2) = 0\\ &\Longrightarrow \lambda X_1 + (1-\lambda)X_2 \in \Omega\\ &\Longrightarrow \Omega \quad \textit{is convex.}\end{aligned}$$

Ω *is convex* $\Longrightarrow 1_\Omega$ *is convex* ?

$$\begin{aligned}\Omega \quad \textit{is convex} &\Longrightarrow \textit{For all} \quad X_1, X_2 \in \Omega \quad \textit{and for each} \quad \lambda \in [0,1] \quad \lambda X_1 + (1-\lambda)X_2 \in \Omega\\ &\Longrightarrow \textit{For all} \quad X_1, X_2 \in \Omega \quad \left(1_\Omega(X_1) = 1_\Omega(X_2) = 0\right), \quad \left(1_\Omega(\lambda X_1 + (1-\lambda)X_2) = 0\right)\\ &\Longrightarrow 0 \le 0 + 0\\ &\Longrightarrow 1_\Omega(\lambda X_1 + (1-\lambda)X_2) \le \lambda 1_\Omega(X_1) + (1-\lambda)1_\Omega(X_2)\\ &\Longrightarrow 1_\Omega \quad \textit{is convex}\end{aligned}$$

$$\Omega \quad \textit{is convex} \Longrightarrow \textit{For all} \quad X_1, X_2 \notin \Omega \quad \Big(1_\Omega(X_1) = 1_\Omega(X_2) = \infty\Big),$$
$$\Big(1_\Omega(\lambda X_1 + (1-\lambda)X_2) = 0\Big) or \Big(1_\Omega(\lambda X_1 + (1-\lambda)X_2) = \infty\Big)$$
$$\Longrightarrow \Big(0 \leq \infty + \infty\Big) or \Big(\infty \leq \infty + \infty\Big)$$
$$\Longrightarrow 1_\Omega \quad \textit{is convex}$$

$$\Omega \quad \textit{is convex} \Longrightarrow \textit{For all} \quad X_1 \in \Omega, X_2 \notin \Omega \quad \Big(1_\Omega(X_1) = 0, \quad 1_\Omega(X_2) = \infty\Big),$$
$$\Big(1_\Omega(\lambda X_1 + (1-\lambda)X_2) = 0\Big) or \Big(1_\Omega(\lambda X_1 + (1-\lambda)X_2) = \infty\Big)$$
$$\Longrightarrow \Big(0 \leq 0 + \infty\Big) or \Big(\infty \leq 0 + \infty\Big)$$
$$\Longrightarrow 1_\Omega \quad \textit{is convex}$$

Answer 6.2.3. $f \quad \textit{is convex} \Longrightarrow epi(f) \quad \textit{is convex} \qquad ?$

$$f \quad \textit{is convex} \Longrightarrow \textit{For all} \quad (u,\alpha),(v,\beta) \in epi(f), \quad \Big(f(u) \leq \alpha, \quad f(v) \leq \beta\Big),$$
$$\Longrightarrow f(tu + (1-t)v) \leq tf(u) + (1-t)f(v)$$
$$\Longrightarrow f(tu + (1-t)v) \leq t\alpha + (1-t)\beta, \quad \Big(tu + (1-t)v \in U\, convex\Big)$$
$$\Longrightarrow \Big(tu + (1-t)v, t\alpha + (1-t)\beta\Big) \in epi(f)$$
$$\Longrightarrow t(u,\alpha) + (1-t)(v,\beta) \in epi(f)$$
$$\Longrightarrow epi(f) \quad \textit{is convex.}$$

$epi(f) \quad \textit{is convex} \Longrightarrow f \quad \textit{is convex} \qquad ?$

$$\begin{aligned}
epi(f) \quad \textit{is convex} \Longrightarrow & (u, f(u)), (v, f(v)) \in epi(f) \\
& t(u, f(u)) + (1-t)(v, f(v)) \in epi(f) \\
& \Longrightarrow \Big(tu + (1-t)v, tf(u) + (1-t)f(v) \Big) \in epi(f) \\
& \Longrightarrow f(tu + (1-t)v) \leq tf(u) + (1-t)f(v) \\
& \Longrightarrow f \quad \textit{is convex}
\end{aligned}$$

Answer 6.2.4. $F \quad \textit{is convex} \Longrightarrow \Phi \quad \textit{is increasing} \qquad ?$

$$F \quad \textit{is convex} \Longrightarrow \textit{let} \quad t_2 \geq t_1 > 0 \quad \textit{on pose} \quad t = \frac{t_1}{t_2} \in (0,1]$$

$$\begin{aligned}
& F(u + t_1 v) = F(u + tt_2 v) = F(u + tu - tu + tt_2 v) = F((1-t)u + t(u + t_2 v)) \\
& \leq (1-t)F(u) + tF(u + t_2 v) \\
& \Longrightarrow f \quad \textit{is convex} \\
& \Longrightarrow F(u + t_1 v) - F(u) \leq t \Big[F(u + t_2 v) - F(u) \Big] = \frac{t_1}{t_2} \Big[F(u + t_2 v) - F(u) \Big] \\
& \Longrightarrow \frac{\Big[F(u + t_1 v) - F(u) \Big]}{t_1} \leq \frac{\Big[F(u + t_2 v) - F(u) \Big]}{t_2} \\
& \Longrightarrow \Phi(t_1) \leq \Phi(t_2) \\
& \Longrightarrow \Phi \quad \textit{is increasing}
\end{aligned}$$

Answer 6.2.5. $(f_i)_{i \in I} \quad \textit{be any family of convex} \Longrightarrow \sup_{x \in \mathbb{R}^n} f_i \quad \textit{is convex} \qquad ?$

$$f_i(x) \le \sup_{x\in\mathbb{R}^n} f_i(x) \Longrightarrow t f_i(x) \le t \sup_{x\in\mathbb{R}^n} f_i(x)$$
$$f_i(y) \le \sup_{y\in\mathbb{R}^n} f_i(y) \Longrightarrow (1-t) f_i(y) \le (1-t) \sup_{y\in\mathbb{R}^n} f_i(y)$$
$$f_i(tx+(1-t)y) \le t f_i(x) + (1-t) f_i(y) \le t \sup_{x\in\mathbb{R}^n} f_i(x) + (1-t) \sup_{y\in\mathbb{R}^n} f_i(y)$$
$$f_i \textit{ convex}$$
$$\sup_{x,y\in\mathbb{R}^n} f_i(tx+(1-t)y) \le t \sup_{x\in\mathbb{R}^n} f_i(x) + (1-t) \sup_{y\in\mathbb{R}^n} f_i(y)$$

Answer 6.2.6. $ab \le \frac{1}{p}a^p + \frac{1}{q}b^q$?

$$ab = \exp^{\ln ab} = \exp^{\ln a + \ln b} = \exp^{\frac{1}{p}\ln a^p + \frac{1}{q}\ln b^q} \le \frac{1}{p}\exp^{\ln a^p} + \frac{1}{q}\exp^{\ln b^q} = \frac{1}{p}a^p + \frac{1}{q}b^q$$
$\exp$ *is convex.*

Answer 6.2.7. ***Reasoning by recurrence***

a. $\mathscr{P}(2)$ *(verifies the property): this is the initialization (or base) of the recurrence;*

b. *For any integer p, $\mathscr{P}(p) \Longrightarrow \mathscr{P}(p+1)$): this is heredity (we say that $\mathscr{P}$ is hereditary).*

a) $p = 2$

$\forall(\lambda_i)_{1\le i\le 2} \in (\mathbb{R}^n)^2$ *such as* $\sum_{i=1}^{2}\lambda_i = 1$, $\forall(x_i)_{1\le i\le 2} \in (\mathbb{R}^n)^2; f\left(\sum_{i=1}^{2}\lambda_i x_i\right) \le \sum_{i=1}^{2}\lambda_i f(x_i)$

(f be a convex function $\lambda_2 = 1-\lambda_1$ $\mathscr{P}(2)$, *is true).*

b) $\mathscr{P}(p) \Longrightarrow \mathscr{P}(p+1))$?

$\forall(\lambda_i)_{1\le i\le p+1} \in (\mathbb{R}^n)^{p+1}$ *such as* $\sum_{i=1}^{p+1}\lambda_i = 1$, *and let* $i_0 \in \{1,2,\dots,p+1\}$ *be such that*

$$\sum_{i=1, i\ne i_0}^{p+1}\lambda_i \ne 0 \textit{ laid } \sum_{i=1, i\ne i_0}^{p+1}\lambda_i = \mu. \textit{ So } \mu + \lambda_{i_0} = 1 \textit{and } \mu > 0, \lambda_{i_0} \ge 0$$

$\sum_{i=1, i\ne i_0}^{p+1}\lambda_i \ne 0$ *then there exists* $x \in \mathbb{R}^n$ *(Barycenter)* $\sum_{i=1, i\ne i_0}^{p+1}\lambda_i x_i = \mu x$

$$
\begin{aligned}
f \ convex &\Longrightarrow f(\lambda_{i_0} x_{i_0} + \mu x) \leq \lambda_{i_0} f(x_{i_0}) + \mu f(x) \\
&\Longrightarrow f(\sum_{i=1}^{p+1} \lambda_i x_i) \leq \lambda_{i_0} f(x_{i_0}) + \mu f(x) \\
f(x) = f(\sum_{i=1, i\neq i_0}^{p+1} \frac{\lambda_i}{\mu} x_i) &\leq \sum_{i=1, i\neq i_0}^{p+1} \frac{\lambda_i}{\mu} f(x_i) \quad (\mathscr{P}(p), is\ true) \\
&\Longrightarrow f(\sum_{i=1}^{p+1} \lambda_i x_i) \leq \sum_{i=1}^{p+1} \lambda_i f(x_i) \\
&\Longrightarrow (\mathscr{P}(p+1), is\ true)
\end{aligned}
$$

Answer 6.2.8. *(See theorem 1.8.1)*

Answer 6.2.9. *(See theorem 1.8.2)*

6.3 TD Series No. 03 Corrected

Answer 6.3.1. *Let $\widehat{x}$ is a max (local or global) of f then*

$f(\widehat{x}) = max\{f(x), \quad x \in \mathbb{R}^n (or\, x \in v)\} \quad v \in V(\widehat{x})$

$$\begin{aligned}
&\Longleftrightarrow f(x) \le f(\widehat{x}), \quad \forall x \in \mathbb{R}^n \quad (x \in v)\\
&\Longleftrightarrow -f(\widehat{x}) \le -f(x), \quad \forall x \in \mathbb{R}^n \quad (x \in v)\\
&\Longleftrightarrow -f(\widehat{x}) = \min\{-f(x), \quad \forall x \in \mathbb{R}^n \quad (x \in v)\}\\
&\Longleftrightarrow f(\widehat{x}) = -\min\{-f(x), \quad \forall x \in \mathbb{R}^n \quad (x \in v)\}
\end{aligned}$$

Answer 6.3.2. **1.** $f_1 : \mathbb{R} \longrightarrow \mathbb{R}; x \mapsto f_1(x) = x^3 - x^2 + 5.$

$$\lim_{\|x\|\to\infty} f_1(x) = \begin{cases} \lim\limits_{x\to+\infty} f_1(x) \\ \lim\limits_{x\to-\infty} f_1(x) \end{cases} = \begin{cases} \lim\limits_{x\to+\infty} x^3 = +\infty \\ \lim\limits_{x\to-\infty} x^3 = -\infty \end{cases} \textit{ is not coercive}$$

2. $f_2 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_2(x) = \langle a, x\rangle + b \quad a \in \mathbb{R}^n, b \in \mathbb{R}.$

$$\lim_{\|x\|\to\infty} f_2(x) = \begin{cases} b & if \quad a = 0 \\ -\infty & if \quad a \neq 0 \end{cases} \textit{ is not coercive}$$

$a \neq 0 \Longrightarrow \exists i_0 \neq 0$ *such that* $a = (0 \cdots a_{i_0} \cdots 0) \quad x_k = (0 \cdots - k a_{i_0} \cdots 0)$

$f_2(x_k) = -k a_{i_0}^2 + b \quad \|x_k\| \to +\infty \quad f_2(x_k) \to -\infty$

3. $f_3 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_3(x) = a\langle x, x\rangle + b \quad b \in \mathbb{R}^n$,*a and* $b \in \mathbb{R}$.

$$\lim_{\|x\|\to\infty} f_3(x) = \lim_{\|x\|\to\infty} (a\|x\|^2 + b) = \begin{cases} -\infty & if \quad a < 0 \quad \textit{is not coercive} \\ b & if \quad a = 0 \quad \textit{is not coercive} \\ +\infty & if \quad a > 0 \quad \textit{is coercive} \end{cases}$$

4. $f_4 : \mathbb{R}^2 \longrightarrow \mathbb{R}; x \mapsto f_4(x) = 2x_1^2 + x_2 - 5$

we take the sequence $x_n = (0. - n), \quad n \ge 0$

$\|x_n\| = n \to +\infty \quad f(x_n) = -n - 5 \to -\infty$ *then* f_4 *is not coercive*

5. $f_5 : \mathbb{R}^2 \longrightarrow \mathbb{R}; x \mapsto f_5(x) = x_1^2 + 2x_2^3 + x_2^2 - x_1$

we take the sequence $x_n = (0. - n), \quad n \geq 0$

$\|x_n\| = n \to +\infty \quad f(x_n) = -2n^3 + n^2 \to -\infty$ *then* f_5 *is not coercive*

6. $f_6 : \mathbb{R}^2 \longrightarrow \mathbb{R}; x \mapsto f_6(x) = x_1^2 + 2x_1 + x_2^2$

We have $\quad (x_1+2)^2 \geq 0 \Longrightarrow 2x_1 \geq -\frac{1}{2}x_1^2 - 2$

$f_6(x) \geq \frac{1}{2}x_1^2 + x_2^2 - 2 \geq \frac{1}{2}(x_1^2 + x_2^2) - 2, \quad (x_2^2 \geq \frac{1}{2}x_2^2)$

$f_6(x) \geq \frac{1}{2}\|(x_1,x_2)\|^2 - 2, \quad \|(x_1,x_2)\| \longrightarrow +\infty \Longrightarrow f_6(x) \longrightarrow +\infty,$ *then* f_6 *is coercive*

7. $f_7 : \mathbb{R}^2 \longrightarrow \mathbb{R}; x \mapsto f_7(x) = x_1^2 + x_2^2 - 3x_2 - 5$

We have $\quad (x_2-3)^2 \geq 0 \Longrightarrow -3x_2 \geq -\frac{1}{2}x_2^2 - \frac{9}{2}$

$f_7(x) \geq x_1^2 + \frac{1}{2}x_2^2 - \frac{9}{2} \geq \frac{1}{2}(x_1^2 + x_2^2) - \frac{9}{2}, \quad (x_1^2 \geq \frac{1}{2}x_1^2)$

$f_7(x) \geq \frac{1}{2}\|(x_1,x_2)\|^2 - \frac{9}{2}, \quad \|(x_1,x_2)\| \longrightarrow +\infty \Longrightarrow f_7(x) \longrightarrow +\infty,$ *then* f_7 *is coercive*

8. $f_8 : \mathbb{R}^n \longrightarrow \mathbb{R}; x \mapsto f_8(x) = \langle x, x\rangle + \langle a, x\rangle + b \quad a \in \mathbb{R}^n, b \in \mathbb{R}$

$f_8(x) = \|x\|^2 + \sum_{i=1}^{n} a_i x_i + b$

We have $\quad (x_i + a_i)^2 \geq 0 \Longrightarrow a_i x_i \geq -\frac{1}{2}x_i^2 - \frac{1}{2}a_i^2$

$\sum_{i=1}^{n} a_i x_i \geq -\frac{1}{2}\sum_{i=1}^{n} x_i^2 - \frac{1}{2}\sum_{i=1}^{n} a_i^2 = -\frac{1}{2}\|x\|^2 - \frac{1}{2}\|a\|^2$

$f_8(x) \geq \frac{1}{2}\|x\|^2 - \frac{1}{2}\|a\|^2 + b, \quad \|x\| \longrightarrow +\infty \Longrightarrow f_8(x) \longrightarrow +\infty,$ *then* f_8 *is coercive*

Answer 6.3.3. *1. We have* $\quad \forall (x,\varepsilon) \quad (x^2-\varepsilon)^2 \geq 0 \Longrightarrow x^4 \geq 2\varepsilon x^2 - \varepsilon^2 \qquad (1)$

and $\quad \forall (y,\varepsilon) \quad (y^2-\varepsilon)^2 \geq 0 \Longrightarrow y^4 \geq 2\varepsilon y^2 - \varepsilon^2 \qquad (2)$

and $\quad (x+y)^2 \geq 0 \Longrightarrow xy \geq -\frac{1}{2}(x^2+y^2) \qquad (3)$

bay 1,2 and 3 We have $f(x,y) \geq (2\varepsilon - 4)(x^2+y^2) - 2\varepsilon^2$

there exists $(\alpha,\beta) \in \mathbb{R}_+^2$ *such that* $(\alpha,\beta) = (2\varepsilon - 4, -2\varepsilon^2)$

$\|(x,y)\| \longrightarrow +\infty \Longrightarrow f(x,y) \longrightarrow +\infty,$ *then* $f(x,y)$ *is coercive*

$f(x,y)$ be a continuous and coercive function defined on all $\mathbb{R}^2$, Then $f(x,y)$ has at least one global minimizer.

2.

f is convex if and only if $\nabla^2 f(x,y)$ is positive semidefinite for all $(x,y)\in\mathbb{R}^2$

$$H(x)=\nabla^2 f(x,y)=4\begin{pmatrix}3x^2-1 & 1\\ 1 & 3y^2-1\end{pmatrix},$$

$$\nabla^2 f(0,0)=4\begin{pmatrix}-1 & 1\\ 1 & -1\end{pmatrix}\Longrightarrow \nabla^2 f(0,0)-\lambda I=\begin{pmatrix}-4-\lambda & 4\\ 4 & -4-\lambda\end{pmatrix}$$

$det(\nabla^2 f(0,0)-\lambda I)=\lambda(\lambda+8)=0\Longrightarrow \lambda=0 \quad or \quad \lambda=-8$

$\lambda=-8<0\Longrightarrow \nabla^2 f(0,0)$ *is not positive semidefinite* $\Longrightarrow f$ *is not convex.*

3. $\nabla f=0\Longrightarrow \begin{pmatrix}4x^3-4(x-y)\\ 4y^3+4(x-y)\end{pmatrix}=0\Longrightarrow (x,y)=(0,0)\vee(\sqrt{2},-\sqrt{2})\vee(-\sqrt{2},\sqrt{2})$

a. $(0,0),\quad det\nabla^2 f(0,0)=0$ *saddel point.*

b. $(\sqrt{2},-\sqrt{2}),\quad det\nabla^2 f(\sqrt{2},-\sqrt{2})=384>0$ *and* $f_{xx}=20>0 \quad min_{(x,y)\in\mathbb{R}^2} f(x,y)=f(\sqrt{2},-\sqrt{2})=-8$

c. $(-\sqrt{2},\sqrt{2}),\quad det\nabla^2 f(-\sqrt{2},\sqrt{2})=384>0$ *and* $f_{xx}=20>0 \quad min_{(x,y)\in\mathbb{R}^2} f(x,y)=f(-\sqrt{2},\sqrt{2})=-8$

Answer 6.3.4. $\nabla f(x,y)=\begin{pmatrix}2x+a\\ 2y+b\end{pmatrix}\Longrightarrow \nabla^2 f(x,y)=\begin{pmatrix}2 & 0\\ 2 & 0\end{pmatrix}$

$\left\langle \nabla^2 f(x,y)\begin{pmatrix}u\\ v\end{pmatrix},\begin{pmatrix}u\\ v\end{pmatrix}\right\rangle=\left\langle \begin{pmatrix}2 & 0\\ 2 & 0\end{pmatrix}\begin{pmatrix}u\\ v\end{pmatrix},\begin{pmatrix}u\\ v\end{pmatrix}\right\rangle=2(u^2+v^2)=2\|(u,v)\|^2\geq\alpha\|(u,v)\|^2$
such that $\alpha\in]0,2]$. *Then f is elliptical*

2. f is elliptical $\Longrightarrow$ *f is coercive and strictly convex*

. $\Longrightarrow$ *the problem (P_2) have a solution unique.*

$$\nabla f(x,y)=0 \Longrightarrow \begin{pmatrix} 2x+a \\ 2y+b \end{pmatrix}=0 \Longrightarrow (x,y)=(-\frac{a}{2},-\frac{b}{2}).$$

Answer 6.3.5. *1.*

t_i	1	2	3	4	5	6	7	8	9	10	$\sum_{i=1}^{10} t_i=55$
x_i	0	-3	6	-3	6	3.8	5	-2	1.4	8	$\sum_{i=1}^{10} x_i=22.2$
t_i^2	1	4	9	16	25	36	49	64	81	100	$\sum_{i=1}^{10} t_i^2=385$

$$\mathscr{J}(a,b)=\sum_{i=1}^{10}(x_i-at_i-b)^2.$$

$\mathscr{J}$ *she is diff* $\frac{\partial \mathscr{J}}{\partial a}(a,b)=2\sum_{i=1}^{10}(-t_i)((x_i-at_i-b)=2a\sum_{i=1}^{10}t_i^2+2b\sum_{i=1}^{10}t_i-2\sum_{i=1}^{10}t_ix_i$

$$\frac{\partial \mathscr{J}}{\partial b}(a,b)=2\sum_{i=1}^{10}(-1)((x_i-at_i-b)=2a\sum_{i=1}^{10}t_i+20b-2\sum_{i=1}^{10}x_i$$

$$\nabla \mathscr{J}(a,b)=\begin{pmatrix} \frac{\partial \mathscr{J}}{\partial a}(a,b) \\ \frac{\partial \mathscr{J}}{\partial b}(a,b) \end{pmatrix}=\begin{pmatrix} 2a\sum_{i=1}^{10}t_i^2+2b\sum_{i=1}^{10}t_i-2\sum_{i=1}^{10}t_ix_i \\ 2a\sum_{i=1}^{10}t_i+20b-2\sum_{i=1}^{10}x_i \end{pmatrix}.$$

$\frac{\partial^2 \mathscr{J}}{\partial a^2}(a,b)=2\sum_{i=1}^{10}t_i^2$

$\frac{\partial^2 \mathscr{J}}{\partial a\partial b}(a,b)=2\sum_{i=1}^{10}t_i$

$\frac{\partial^2 \mathscr{J}}{\partial b^2}(a,b)=20$ *It is clear that* $\mathscr{J}$ *is twice diff (polyane in a and b)*

$$H\mathscr{J}(a,b)=\begin{pmatrix} \frac{\partial^2 \mathscr{J}}{\partial a^2}(a,b) & \frac{\partial^2 \mathscr{J}}{\partial a\partial b}(a,b) \\ \frac{\partial^2 \mathscr{J}}{\partial a\partial b}(a,b) & \frac{\partial^2 \mathscr{J}}{\partial b^2}(a,b) \end{pmatrix}=\begin{pmatrix} 2\sum_{i=1}^{10}t_i^2 & 2\sum_{i=1}^{10}t_i \\ 2\sum_{i=1}^{10}t_i & 20 \end{pmatrix}=\begin{pmatrix} 770 & 110 \\ 110 & 20 \end{pmatrix}.$$

The Hessian matrix is positive semi-definite because $2T^2 \geq 0$.

(the positive eigenvalues)$\Longrightarrow \mathscr{J}$ *is strictly convex (convex) then the solution is unique (global).*

We have f *is diff and convex then any stationary point is a global min* $\Longrightarrow$ *the pb admits a single solution.*

$$\nabla \mathscr{J}(a,b) = 0 \Leftrightarrow \begin{cases} 2aT^2 + 2bT - 2TX = 0 \\ 2aT + 20b - 2X = 0 \end{cases}$$

$T^2 = \sum_{i=1}^{10} t_i^2 \quad T = \sum_{i=1}^{10} t_i \quad TX = \sum_{i=1}^{10} t_i x_i \quad X = \sum_{i=1}^{10} x_i$

The system admits a unique solution if

$$\begin{vmatrix} T^2 & T \\ T & 10 \end{vmatrix} = 10T^2 - TT \neq 0 \quad (TT = (\sum_{i=1}^{10} t_i)^2)$$

$a = \dfrac{\begin{vmatrix} T & XT \\ 10 & X \end{vmatrix}}{10T^2 - TT}$

$b = \dfrac{\begin{vmatrix} T^2 & XT \\ T & X \end{vmatrix}}{10T^2 - TT}$

So the general case if $\quad 10T^2 - TT \neq 0 \Longrightarrow A^{-1}$ *exists* $\Longrightarrow$ *the pb admits a solution.*

Answer 6.3.6. *Let* $x, y \in \mathbb{R}^n$ *such that* $x \neq y$ *and* $t \in]0,1[$

1.

a. $\mathscr{J}(tu + (1-t)v) - t\mathscr{J}(u) - (1-t)\mathscr{J}(v) = \dfrac{t(t-1)}{2}\langle A(u-v), u-v\rangle > 0 \quad (t(t-1) > 0$ *and* A *is a positive definite)* $\Rightarrow \mathscr{J}$ *is strictly convex.*

b. *A is symmetric there exists an orthonormal base $(u_i)_{1\leq i\leq n}$ and A positive definite therefore the associated eigenvalues are all strictly positive therefore*

$$x = \sum_{i=1}^{n} x_i u_i, \quad x_i = \langle x, u_i \rangle$$

$$Ax = \sum_{i=1}^{n} x_i A u_i = \sum_{i=1}^{n} \lambda_i x_i u_i$$

$$\langle Ax, x \rangle = \sum_{i=1}^{n} \lambda_i x_i x_j \langle u_i, u_j \rangle = \sum_{i=1}^{n} \lambda_i x_i^2 \geq \min\{\lambda_i\} \sum_{i=1}^{n} x_i^2$$

$$\frac{1}{2}\langle Ax, x \rangle \geq \lambda \|x\|^2 \quad (\lambda = \frac{\min\{\lambda_i\}}{2} > 0)$$

$$\langle b, x \rangle \leq \|b\|.\|x\| \Rightarrow -\langle b, x \rangle \geq -\|b\|.\|x\|$$

$$\mathscr{J}(x) \geq \lambda \|x\|^2 - \|b\|.\|x\| = \|x\|^2(\lambda - \frac{b}{\|x\|}) \to +\infty \quad \|x\| \to +\infty$$

$\mathscr{J}$ is a coercive function.

2. *View the course*

$\mathscr{J}$ is differentiable.

$$\nabla \mathscr{J}(x) = Ax - b$$

$$H\mathscr{J}(x) = A$$

3. *we have*

$\mathscr{J}$ is strictly convex and coercive so (P_4) admits only one solution.

4.

$$\nabla \mathscr{J}(x) = 0 \Longrightarrow Ax - b = 0 \Longrightarrow x = A^{-1}b.$$

A^{-1} *exists because* A *is positive definite and* $\det A \neq 0 \Longrightarrow A^{-1}$ *exists*

A^{-1} *exists* $\Leftrightarrow$ *We are not an eigenvalue of* A *and* A *defines positive* $\Leftrightarrow$ *all non-zero eigenvalues* $\Longrightarrow 0$ *is not a vp*

Even if A *negative definite and* $\det A \neq 0 \Longrightarrow 0$ *We're not vp* $\Longrightarrow A^{-1}$ *exists.*

Chapter 7

Final Exam

7.1 Final Exam 2017-2018

Exercice 1 (Examen et interrogation) (05.00 points)
Etudiez les solutions optimales locales de f, défnie par

$$f(x,y) = x^3 + y^3 - 3xy$$

.

Exercice 2 (Examen et interrogation) (05.00 points)
On considere la fonction

$$f(x,y) = 2x^2 - 2xy + y^2$$

En partant du point initial $(x_0; y_0) = (1; 1)$ et en appliquant la méthode du gradient avec ρ_k optimale, calculez $(x_1; y_1)$; $(x_2; y_2)$ et $(x_3; y_3)$. Puis Images correspondant à sous points par f .

Exercice 3 (04.00 points)
Soit $J : C \subset \mathbb{H} \longrightarrow \mathbb{R}$, Gâteaux différentiable sur C, avec C convexe. J est convexe si et seulement si

$$\forall (u,v) \in C \times C \qquad J(v) \geq J(u) + \langle \nabla J(u), v - u \rangle$$

Exercice 4 (06.00 points)
Une firme aéronautique fabrique des avions qu'elle vend sur deux marchés étrangers. Soit q_1 le nombre d'avions vendus sur le premier marché et q_2 le nombre d'avions vendus sur le deuxième marché. Les fonctions de demande dans les deux marchés respectifs sont :

$$p_1 = 60 - 2q_1$$

$$p_2 = 80 - 4q_2$$

p_1 et p_2 sont les deux prix de vente. La fonction de coût total de la firme est : $C = 50 + 40q$ où q est le nombre total d'avions produits. Il faut trouver le nombre d'avions que la firme doit vendre sur chaque marché pour maximiser son bénéfice.

7.2 Final Exam 2018-2019

Exercice 1 (05.00 points)
Soit la fonction $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ et les points $a = (1, 1)$ et $b = (-1, 2)$.
a) Calculer $f(a), f(b), \nabla f(a)$ et $\nabla f(b)$
b) Discuter les conditions d'optimalité en a et en b sur la base des résultats obtenus en a).
c) La direction $d = a - b$ est-elle une direction de descente en b ? Justifier.

Exercice 2 (05.00 points)
On considere la fonction

$$f(x, y) = 4x^2 - 4xy + 2y^2$$

En partant du point initial $(x_0; y_0) = (2; 3)$ et en appliquant la méthode du gradient avec ρ_k optimale, calculez $(x_1; y_1)$; $(x_2; y_2)$ et $(x_3; y_3)$. Puis Images correspondant à sous points par f .

Exercice 3 (05.00 points)
Soit $f \in C^1(\mathbb{R}^N, \mathbb{R})(N \geq 1)$. On suppose que f vérifie :

$$\exists \alpha > 0 \quad tq. \quad (\nabla f(x) - \nabla f(y)).(x - y) \geq \alpha \mid x - y \mid^2, \quad \forall x, y \in \mathbb{R}^N, \tag{1}$$

$$\exists M > 0 \quad tq. \quad \mid \nabla f(x) - \nabla f(y) \mid \leq M \mid x - y \mid, \quad \forall x, y \in \mathbb{R}^N. \tag{2}$$

1. Montrer que $f(y) - f(x) \geq \nabla f(x).(y - x) + \frac{\alpha}{2} \mid y - x \mid^2, \quad \forall x, y \in \mathbb{R}^N$.
2. Montrer que f est strictement convexe et que $f(x) \longrightarrow \infty$quand $\mid x \mid \longrightarrow \infty$.En déduire qu'il existe un et un seul $\bar{x} \in \mathbb{R}^N$tq. $f(\bar{x}) \leq f(x)$ pour tout $x \in \mathbb{R}^N$.
3. Soient $\rho \in]0, (2\alpha/M^2)[$ et $x_0 \in \mathbb{R}^N$. Montrer que la suite $(x_n)_{n \in \mathbb{N}}$ définie par $x_{n+1} = x_n - \rho \nabla f(x_n)$ (pour $n \in \mathbb{N}$) converge vers $\bar{x}$.

Exercice 4 (05.00 points)
Un industriel produit simultanément 2 biens A et B dont il a le monopole de la production et de la vente dans un pays. Soit x la quantité produit du premier bien et y la quantité produite du second. Les prix p_A et p_B auxquels il vend les bien A et B sont fonction des quantités écoulées selon les relations :

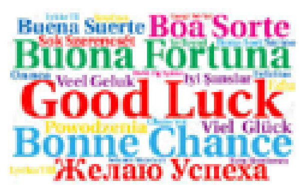
$$\begin{cases} p_A = f(x) \\ p_B = g(y) \end{cases}$$

Le coût de production total des quantités x et y est une fonction $c(x, y)$.

Le Bénéfice de l'entreprise si elle vend les quantités x et y est donc la fonction
$\pi(x, y) = xf(x) + yg(y) - c(x, y)$
Trouvez les quantités qui maximisent le bénéfice de l'entreprise, la valeur maximale du bénéfice ainsi que les prix de vente de chacun des biens

$$\begin{cases} p_A = 1 - x \\ p_B = 1 - y \\ c(x, y) = xy \end{cases}$$

1.jpg 

2.jpeg
3.jpeg 

7.3 Final Exam 2019-2020

Exercice 1 (03.00 points)
Soit $f : I \longrightarrow \mathbb{R}$ une fonction convexe et strictement croissante. Étudier la convexité de $f^{-1} : f(I) \longrightarrow I$.

Exercice 2 (07.00 points)
On considere la fonction

$$f(x, y) = (2x - y)^2 + y^2$$

1- Trouver l'extremum local $X^{(*)}$.
2- En partant du point initial $X^{(0)} = (x_0; y_0) = (0; 1)$. Calculez $X^{(1)}; X^{(2)}$ et $X^{(3)}$. Puis les images $f(X^{(k)}) \quad k = *, 0, 1, 2, 3$ et comparez-les. Appliquant

1. Méthode du gradient à pas constant $\rho = \frac{1}{10}$.
2. Méthode du gradient à pas optimal ρ_k.

3- Déduire $lim X^{(k)} et f(X^{(k)}), k \longrightarrow +\infty$

Exercice 3 (06.00 points)
Considérons la fonction $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ définie par

$$f(x, y) = \frac{1}{2}x^2 + x cos(y)$$

1. Trouvez les points stationnaires.
2. Trouvez les points qui vérifient la condition suffisante d'optimalité.
3. Trouvez les solutions minimales locales strictes.

Exercice 4 (04.00 points)
Considérer la fonction suivante :

$$f(x, y) = x^2 - xy + 2y^2 - 2x + e^{x+y}$$

1. Est-ce que $X^{(0)} = (0; 0)$ est un minimum local de la fonction f ? Justifier.
2. Si oui, est-ce aussi un minimum global ? Si non, trouver une direction de descente pour f en $X^{(0)}$.

1.jpg 

2.jpeg 3.jpeg 

Bibliography

[1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, Nonlinear Programming: Theory and Algorithms. New York: Wiley, second ed.,1993.

[2] A. Beck, Introduction to Nonlinear Optimization - Theory, Algorithms and Applications, 2014.

[3] Miche. Bierlaire, Optimization: Principles and Algorithms, 2018.

[4] Edwin K. P. Chong, and Stanislaw Zak, An Introduction to Optimization, 2nd Edition, 2002.

[5] J. E. Dennis, Jr. and R. B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[6] R. Fletcher, Practical Methods of Optimization. Chichester: Wiley, second ed., 1987.

[7] Harish Garg sites.google.com/site/harishg58iitr.

[8] P. E. Gill, W. Murray, and M. H. Wright, Practical Optimization. London: Academic Press, 1981.

[9] D. G. Luenberger, Linear and Nonlinear Programming. Reading, MA: Addison-Wesley, second ed., 1984

[10] D. G. Luenberger, Optimization by Vector Space Methods. New York: Wiley, 1969.

[11] Jorge Nocedal , Stephen J. Wright, Numerical Optimization 2006.

[12] A. L. Peressini, F. E. Sullivan, and J. J. Uhl, Jr.,The Mathematics of Nonlinear Programming. New York: Springer-Verlag, 1988.

[13] Singiresu S. Rao, Engineering Optimization Theory and Practice 2009.

[14] Robert C. Rogers , The Calculus of Several Variables, 2011.

[15] P. P. Varaiya, Notes on Optimization. New York: Van Nostrand Reinhold Co., 1972.

[16] Xin-She Yang, Engineering Optimization: An Introduction with Metaheuristic Applications, (2011).