Chapter 18

Quadratic Optimization Problems

18.1 Quadratic Optimization: The Positive Definite Case

In this chapter, we consider two classes of quadratic optimization problems that appear frequently in engineering and in computer science (especially in computer vision):

1. Minimizing
$$Q(x) = \frac{1}{2}x^\top Ax - x^\top b$$
over all $x \in \mathbb{R}^n$, or subject to linear or affine constraints.

2. Minimizing
$$Q(x) = \frac{1}{2}x^\top Ax - x^\top b$$
over the unit sphere.

In both cases, A is a symmetric matrix. We also seek necessary and sufficient conditions for f to have a global minimum.

Many problems in physics and engineering can be stated as the *minimization of some energy function*, with or without constraints.

Indeed, it is a fundamental principle of mechanics that nature acts so as to minimize energy.

Furthermore, if a physical system is in a stable state of equilibrium, then the energy in that state should be minimal.

The simplest kind of energy function is a quadratic function.

Such functions can be conveniently defined in the form

$$Q(x) = x^\top A x - x^\top b,$$

where A is a symmetric $n \times n$ matrix, and x, b, are vectors in $\mathbb{R}^n$, viewed as column vectors.

Actually, for reasons that will be clear shortly, it is preferable to put a factor $\frac{1}{2}$ in front of the quadratic term, so that

$$Q(x) = \frac{1}{2} x^\top A x - x^\top b.$$

The question is, under what conditions (on A) does $Q(x)$ have a global minimum, preferably unique?

We give a complete answer to the above question in two stages:

1. In this section, we show that if A is symmetric positive definite, then $Q(x)$ has a unique global minimum precisely when
$$Ax = b.$$
2. In Section 18.2, we give necessary and sufficient conditions in the general case, in terms of the pseudo-inverse of A.

We begin with the matrix version of Definition 16.2.

Definition 18.1. A symmetric *positive definite matrix* is a matrix whose eigenvalues are strictly positive, and a symmetric *positive semidefinite matrix* is a matrix whose eigenvalues are nonnegative.

Equivalent criteria are given in the following proposition.

Proposition 18.1. *Given any Euclidean space E of dimension n, the following properties hold:*

(1) Every self-adjoint linear map $f\colon E \to E$ is positive definite iff

$$\langle f(x), x\rangle > 0$$

for all $x \in E$ with $x \neq 0$.

(2) Every self-adjoint linear map $f\colon E \to E$ is positive semidefinite iff

$$\langle f(x), x\rangle \geq 0$$

for all $x \in E$.

Some special notation is customary (especially in the field of convex optimization) to express that a symmetric matrix is positive definite or positive semidefinite.

Definition 18.2. Given any $n \times n$ symmetric matrix A we write $A \succeq 0$ if A is positive semidefinite and we write $A \succ 0$ if A is positive definite.

It should be noted that we can define the relation

$$A \succeq B$$

between any two $n \times n$ matrices (symmetric or not) iff $A - B$ is symmetric positive semidefinite.

It is easy to check that this relation is actually a partial order on matrices, called the *positive semidefinite cone ordering*; for details, see Boyd and Vandenberghe [8], Section 2.4.

If A is symmetric positive definite, it is easily checked that A^{-1} is also symmetric positive definite.

Also, if C is a symmetric positive definite $m \times m$ matrix and A is an $m \times n$ matrix of rank n (and so $m \geq n$ and the map $x \mapsto Ax$ is injective), then $A^\top CA$ is symmetric positive definite.

We can now prove that

$$Q(x) = \frac{1}{2}x^\top Ax - x^\top b$$

has a global minimum when A is symmetric positive definite.

Proposition 18.2. *Given a quadratic function*

$$Q(x) = \frac{1}{2}x^\top Ax - x^\top b,$$

if A is symmetric positive definite, then $Q(x)$ has a unique global minimum for the solution $x_0 = A^{-1}b$ of the linear system $Ax = b$. The minimum value of $Q(x)$ is

$$Q(A^{-1}b) = -\frac{1}{2}b^\top A^{-1}b.$$

Remarks:

(1) The quadratic function $Q(x)$ is also given by

$$Q(x) = \frac{1}{2}x^\top Ax - b^\top x,$$

but the definition using $x^\top b$ is more convenient for the proof of Proposition 18.2.

(2) If $Q(x)$ contains a constant term $c \in \mathbb{R}$, so that

$$Q(x) = \frac{1}{2}x^\top Ax - x^\top b + c,$$

the proof of Proposition 18.2 still shows that $Q(x)$ has a unique global minimum for $x = A^{-1}b$, but the minimal value is

$$Q(A^{-1}b) = -\frac{1}{2}b^\top A^{-1}b + c.$$

Thus, when the energy function $Q(x)$ of a system is given by a quadratic function

$$Q(x) = \frac{1}{2}x^\top Ax - x^\top b,$$

where A is symmetric positive definite, *finding the global minimum of $Q(x)$ is equivalent to solving the linear system $Ax = b$*.

Sometimes, it is useful to recast a linear problem $Ax = b$ as a variational problem (finding the minimum of some energy function).

However, very often, a minimization problem comes with extra *constraints* that must be satisfied for all admissible solutions.

For instance, we may want to minimize the quadratic function

$$Q(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2)$$

subject to the constraint

$$2x_1 - x_2 = 5.$$

The solution for which $Q(x_1, x_2)$ is minimum is no longer $(x_1, x_2) = (0, 0)$, but instead, $(x_1, x_2) = (2, -1)$, as will be shown later.

Geometrically, the graph of the function defined by $z = Q(x_1, x_2)$ in $\mathbb{R}^3$ is a paraboloid of revolution P with axis of revolution Oz.

The constraint

$$2x_1 - x_2 = 5$$

corresponds to the vertical plane H parallel to the z-axis and containing the line of equation $2x_1 - x_2 = 5$ in the xy-plane.

Thus, the constrained minimum of Q is located on the parabola that is the intersection of the paraboloid P with the plane H.

A nice way to solve constrained minimization problems of the above kind is to use the method of *Lagrange multipliers*.

Definition 18.3. The *quadratic constrained minimization problem* consists in minimizing a quadratic function

$$Q(x) = \frac{1}{2}x^\top A^{-1}x - b^\top x$$

subject to the linear constraints

$$B^\top x = f,$$

where A^{-1} is an $m \times m$ symmetric positive definite matrix, B is an $m \times n$ matrix of rank n (so that $m \geq n$), and where $b, x \in \mathbb{R}^m$ (viewed as column vectors), and $f \in \mathbb{R}^n$ (viewed as a column vector).

The reason for using A^{-1} instead of A is that the constrained minimization problem has an interpretation as a set of equilibrium equations in which the matrix that arises naturally is A (see Strang [31]).

Since A and A^{-1} are both symmetric positive definite, this doesn't make any difference, but it seems preferable to stick to Strang's notation.

The method of Lagrange consists in *incorporating the n constraints $B^\top x = f$ into the quadratic function $Q(x)$*, by introducing extra variables $\lambda = (\lambda_1, \ldots, \lambda_n)$ called *Lagrange multipliers*, one for each constraint. We form the *Lagrangian*

$$\begin{aligned} L(x, \lambda) &= Q(x) + \lambda^\top (B^\top x - f) \\ &= \frac{1}{2} x^\top A^{-1} x - (b - B\lambda)^\top x - \lambda^\top f. \end{aligned}$$

We shall prove in Proposition 18.3 that our constrained minimization problem has a unique solution given by the system of linear equations

$$\begin{aligned} A^{-1}x + B\lambda &= b, \\ B^\top x &= f, \end{aligned}$$

which can be written in matrix form as

$$\begin{pmatrix} A^{-1} & B \\ B^\top & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} b \\ f \end{pmatrix}.$$

Note that the matrix of this system is symmetric. Eliminating x from the first equation

$$A^{-1}x + B\lambda = b,$$

we get

$$x = A(b - B\lambda),$$

and substituting into the second equation, we get

$$B^\top A(b - B\lambda) = f,$$

that is,

$$B^\top AB\lambda = B^\top Ab - f.$$

However, by a previous remark, since A is symmetric positive definite and the columns of B are linearly independent, $B^\top AB$ is symmetric positive definite, and thus invertible. Thus we obtain the solution

$$\lambda = (B^\top AB)^{-1}(B^\top Ab - f), \qquad x = A(b - B\lambda).$$

Note that this way of solving the system requires solving for the Lagrange multipliers first.

Letting $e = b - B\lambda$, we also note that the system

$$\begin{pmatrix} A^{-1} & B \\ B^\top & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} b \\ f \end{pmatrix}$$

is equivalent to the system

$$\begin{aligned} e &= b - B\lambda, \\ x &= Ae, \\ B^\top x &= f. \end{aligned}$$

The latter system is called the *equilibrium equations* by Strang [31].

Indeed, Strang shows that the equilibrium equations of many physical systems can be put in the above form.

In order to prove that our constrained minimization problem has a unique solution, we proceed to prove that the *constrained minimization* of $Q(x)$ subject to $B^\top x = f$ is equivalent to the *unconstrained maximization* of another function $-G(\lambda)$.

We get $G(\lambda)$ by *minimizing the Lagrangian* $L(x, \lambda)$ *treated as a function of x alone*.

Since A^{-1} is symmetric positive definite and

$$L(x, \lambda) = \frac{1}{2}x^\top A^{-1}x - (b - B\lambda)^\top x - \lambda^\top f,$$

by Proposition 18.2 the global minimum (with respect to x) of $L(x, \lambda)$ is obtained for the solution x of

$$A^{-1}x = b - B\lambda,$$

and the minimum of $L(x, \lambda)$ is

$$\min_x L(x, \lambda) = -\frac{1}{2}(B\lambda - b)^\top A(B\lambda - b) - \lambda^\top f.$$

Letting

$$G(\lambda) = \frac{1}{2}(B\lambda - b)^\top A(B\lambda - b) + \lambda^\top f,$$

we show in Proposition 18.3 that the solution of the *constrained minimization* of $Q(x)$ subject to $B^\top x = f$ is equivalent to the *unconstrained maximization* of $-G(\lambda)$.

In order to prove that the unique minimum of the constrained problem $Q(x)$ subject to $B^\top x = f$ is the unique maximum of $-G(\lambda)$, we compute $Q(x) + G(\lambda)$.

Proposition 18.3. *The quadratic constrained minimization problem of Definition 18.3 has a unique solution (x, λ) given by the system*

$$\begin{pmatrix} A^{-1} & B \\ B^\top & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} b \\ f \end{pmatrix}.$$

Furthermore, the component λ of the above solution is the unique value for which $-G(\lambda)$ is maximum.

Remarks:

(1) There is a form of *duality* going on in this situation. The constrained minimization of $Q(x)$ subject to $B^\top x = f$ is called the *primal problem*, and the unconstrained maximization of $-G(\lambda)$ is called the *dual problem*. Duality is the fact stated slightly loosely as

$$\min_x Q(x) = \max_\lambda -G(\lambda).$$

Recalling that $e = b - B\lambda$, since

$$G(\lambda) = \frac{1}{2}(B\lambda - b)^\top A(B\lambda - b) + \lambda^\top f,$$

we can also write

$$G(\lambda) = \frac{1}{2}e^\top Ae + \lambda^\top f.$$

This expression often represents the total *potential energy* of a system. Again, the optimal solution is the one that minimizes the potential energy (and thus maximizes $-G(\lambda)$).

(2) It is immediately verified that the equations of Proposition 18.3 are equivalent to the equations stating that the partial derivatives of the Lagrangian $L(x, \lambda)$ are null:

$$\begin{aligned} \frac{\partial L}{\partial x_i} &= 0, \quad i = 1, \ldots, m, \\ \frac{\partial L}{\partial \lambda_j} &= 0, \quad j = 1, \ldots, n. \end{aligned}$$

Thus, the constrained minimum of $Q(x)$ subject to $B^\top x = f$ is an extremum of the Lagrangian $L(x, \lambda)$. As we showed in Proposition 18.3, this extremum corresponds to simultaneously *minimizing* $L(x, \lambda)$ with respect to x and *maximizing* $L(x, \lambda)$ with respect to λ. Geometrically, such a point is a *saddle point* for $L(x, \lambda)$.

(3) The Lagrange multipliers sometimes have a natural physical meaning.

Going back to the constrained minimization of $Q(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2)$ subject to

$$2x_1 - x_2 = 5,$$

the Lagrangian is

$$L(x_1, x_2, \lambda) = \frac{1}{2}\big(x_1^2 + x_2^2\big) + \lambda(2x_1 - x_2 - 5),$$

and the equations stating that the Lagrangian has a saddle point are

$$\begin{aligned} x_1 + 2\lambda &= 0, \\ x_2 - \lambda &= 0, \\ 2x_1 - x_2 - 5 &= 0. \end{aligned}$$

We obtain the solution $(x_1, x_2, \lambda) = (2, -1, -1)$.

18.2 Quadratic Optimization: The General Case

In this section, we complete the study initiated in Section 18.1 and give necessary and sufficient conditions for the quadratic function $\frac{1}{2}x^\top Ax - x^\top b$ to have a global minimum.

We begin with the following simple fact:

Proposition 18.4. *If A is an invertible symmetric matrix, then the function*

$$f(x) = \frac{1}{2}x^\top Ax - x^\top b$$

has a minimum value iff $A \succeq 0$, in which case this optimal value is obtained for a unique value of x, namely $x^* = A^{-1}b$, and with

$$f(A^{-1}b) = -\frac{1}{2}b^\top A^{-1}b.$$

Let us now consider the case of an arbitrary symmetric matrix A.

Proposition 18.5. *If A is a symmetric matrix, then the function*

$$f(x) = \frac{1}{2}x^\top Ax - x^\top b$$

has a minimum value iff $A \succeq 0$ and $(I - AA^+)b = 0$, in which case this minimum value is

$$p^* = -\frac{1}{2}b^\top A^+ b.$$

Furthermore, if $A = U^\top \Sigma U$ is an SVD of A, then the optimal value is achieved by all $x \in \mathbb{R}^n$ of the form

$$x = A^+ b + U^\top \begin{pmatrix} 0 \\ z \end{pmatrix},$$

for any $z \in \mathbb{R}^{n-r}$, where r is the rank of A.

The problem of minimizing the function

$$f(x) = \frac{1}{2}x^\top Ax - x^\top b$$

in the case where we add either linear constraints of the form $C^\top x = 0$ or affine constraints of the form $C^\top x = t$ (where $t \neq 0$) can be reduced to the unconstrained case using a QR-decomposition of C or N.

Let us show how to do this for linear constraints of the form $C^\top x = 0$.

If we use a QR decomposition of C, by permuting the columns, we may assume that

$$C = Q^\top \begin{pmatrix} R & S \\ 0 & 0 \end{pmatrix} \Pi,$$

where R is an $r \times r$ invertible upper triangular matrix and S is an $r \times (m-r)$ matrix (C has rank r).

Then, if we let

$$x = Q^\top \begin{pmatrix} y \\ z \end{pmatrix},$$

where $y \in \mathbb{R}^r$ and $z \in \mathbb{R}^{n-r}$, then, after some calculations, our original problem becomes

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}(y^\top \; z^\top) QAQ^\top \begin{pmatrix} y \\ z \end{pmatrix} + (y^\top \; z^\top) Qb \\ \text{subject to} \quad & y = 0, \; y \in \mathbb{R}^r, \; z \in \mathbb{R}^{n-r}. \end{aligned}$$

Thus, the constraint $C^\top x = 0$ has been eliminated, and if we write

$$QAQ^\top = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}$$

and

$$Qb = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad b_1 \in \mathbb{R}^r, \; b_2 \in \mathbb{R}^{n-r},$$

our problem becomes

$$\text{minimize} \, \frac{1}{2} z^\top G_{22} z + z^\top b_2, \quad z \in \mathbb{R}^{n-r},$$

the problem solved in Proposition 18.5.

Constraints of the form $C^\top x = t$ (where $t \neq 0$) can be handled in a similar fashion.

In this case, we may assume that C is an $n \times m$ matrix with full rank (so that $m \leq n$) and $t \in \mathbb{R}^m$.

18.3 Maximizing a Quadratic Function on the Unit Sphere

In this section we discuss various quadratic optimization problems mostly arising from computer vision (image segmentation and contour grouping).

These problems can be reduced to the following basic optimization problem: Given an $n \times n$ real symmetric matrix A

$$\begin{aligned} &\text{maximize} && x^\top A x \\ &\text{subject to} && x^\top x = 1,\ x \in \mathbb{R}^n. \end{aligned}$$

In view of Proposition 17.10, the maximum value of $x^\top A x$ on the unit sphere is equal to the largest eigenvalue λ_1 of the matrix A, and it is achieved for any unit eigenvector u_1 associated with λ_1.

A variant of the above problem often encountered in computer vision consists in minimizing $x^\top Ax$ on the *ellipsoid* given by an equation of the form

$$x^\top Bx = 1,$$

where B is a symmetric positive definite matrix.

Since B is positive definite, it can be diagonalized as

$$B = QDQ^\top,$$

where Q is an orthogonal matrix and D is a diagonal matrix,

$$D = \mathrm{diag}(d_1, \ldots, d_n),$$

with $d_i > 0$, for $i = 1, \ldots, n$.

If we define the matrices $B^{1/2}$ and $B^{-1/2}$ by

$$B^{1/2} = Q\,\mathrm{diag}\left(\sqrt{d_1}, \ldots, \sqrt{d_n}\right) Q^\top$$

and

$$B^{-1/2} = Q\,\mathrm{diag}\left(1/\sqrt{d_1}, \ldots, 1/\sqrt{d_n}\right) Q^\top,$$

it is clear that these matrices are symmetric, that $B^{-1/2}BB^{-1/2} = I$, and that $B^{1/2}$ and $B^{-1/2}$ are mutual inverses.

Then, if we make the change of variable

$$x = B^{-1/2}y,$$

the equation $x^\top Bx = 1$ becomes $y^\top y = 1$, and the optimization problem

$$\begin{aligned} &\text{maximize} && x^\top Ax \\ &\text{subject to} && x^\top Bx = 1,\ x \in \mathbb{R}^n, \end{aligned}$$

is equivalent to the problem

$$\begin{aligned} &\text{maximize} && y^\top B^{-1/2}AB^{-1/2}y \\ &\text{subject to} && y^\top y = 1,\ y \in \mathbb{R}^n, \end{aligned}$$

where $y = B^{1/2}x$ and where $B^{-1/2}AB^{-1/2}$ is symmetric.

The complex version of our basic optimization problem in which A is a Hermitian matrix also arises in computer vision. Namely, given an $n \times n$ complex Hermitian matrix A,

$$\begin{aligned} &\text{maximize} && x^*Ax \\ &\text{subject to} && x^*x = 1,\ x \in \mathbb{C}^n. \end{aligned}$$

Again by Proposition 17.10, the maximum value of x^*Ax on the unit sphere is equal to the largest eigenvalue λ_1 of the matrix A and it is achieved for any unit eigenvector u_1 associated with λ_1.

It is worth pointing out that if A is a *skew-Hermitian* matrix, that is, if $A^* = -A$, then x^*Ax is *pure imaginary or zero*.

In particular, if A is a real matrix and if A is *skew-symmetric*, then

$$x^\top Ax = 0.$$

Thus, for any real matrix (symmetric or not),

$$x^\top Ax = x^\top H(A)x,$$

where $H(A) = (A + A^\top)/2$, the symmetric part of A.

There are situations in which it is necessary to add linear constraints to the problem of maximizing a quadratic function on the sphere.

This problem was completely solved by Golub [16] (1973).

The problem is the following: Given an $n \times n$ real symmetric matrix A and an $n \times p$ matrix C,

$$\begin{array}{ll} \text{minimize} & x^\top Ax \\ \text{subject to} & x^\top x = 1,\ C^\top x = 0,\ x \in \mathbb{R}^n. \end{array}$$

Golub shows that the linear constraint $C^\top x = 0$ can be eliminated as follows: If we use a QR decomposition of C, by permuting the columns, we may assume that

$$C = Q^\top \begin{pmatrix} R & S \\ 0 & 0 \end{pmatrix} \Pi,$$

where R is an $r \times r$ invertible upper triangular matrix and S is an $r \times (p - r)$ matrix (assuming C has rank r).

Then if we let

$$x = Q^\top \begin{pmatrix} y \\ z \end{pmatrix},$$

where $y \in \mathbb{R}^r$ and $z \in \mathbb{R}^{n-r}$, then, after some calculations, our original problem becomes

$$\begin{aligned} \text{minimize} \quad & (y^\top \; z^\top) Q A Q^\top \begin{pmatrix} y \\ z \end{pmatrix} \\ \text{subject to} \quad & z^\top z = 1, \; z \in \mathbb{R}^{n-r}, \\ & y = 0, \; y \in \mathbb{R}^r. \end{aligned}$$

Thus, the constraint $C^\top x = 0$ has been eliminated, and if we write

$$QAQ^\top = \begin{pmatrix} G_{11} & G_{12} \\ G_{12}^\top & G_{22} \end{pmatrix},$$

our problem becomes

$$\begin{array}{ll} \text{minimize} & z^\top G_{22} z \\ \text{subject to} & z^\top z = 1,\ z \in \mathbb{R}^{n-r}, \end{array}$$

a standard eigenvalue problem.

Observe that if we let

$$J = \begin{pmatrix} 0 & 0 \\ 0 & I_{n-r} \end{pmatrix},$$

then

$$JQAQ^\top J = \begin{pmatrix} 0 & 0 \\ 0 & G_{22} \end{pmatrix},$$

and if we set

$$P = Q^\top J Q,$$

then

$$PAP = Q^\top JQAQ^\top JQ.$$

Now, $Q^\top JQAQ^\top JQ$ and $JQAQ^\top J$ have the same eigenvalues, so PAP and $JQAQ^\top J$ also have the same eigenvalues.

It follows that the solutions of our optimization problem are among the eigenvalues of $K = PAP$, and at least r of those are 0.

Using the fact that CC^+ is the projection onto the range of C, where C^+ is the pseudo-inverse of C, it can also be shown that

$$P = I - CC^+,$$

the projection onto the kernel of $C^\top$.

In particular, when $n \geq p$ and C has full rank (the columns of C are linearly independent), then we know that $C^+ = (C^\top C)^{-1}C^\top$ and

$$P = I - C(C^\top C)^{-1}C^\top.$$

This fact is used by Cour and Shi [10] and implicitly by Yu and Shi [37].

The problem of adding affine constraints of the form $N^\top x = t$, where $t \neq 0$, also comes up in practice.

At first glance, this problem may not seem harder than the linear problem in which $t = 0$, but it is.

This problem was extensively studied in a paper by Gander, Golub, and von Matt [15] (1989).

Gander, Golub, and von Matt consider the following problem:

Given an $(n+m) \times (n+m)$ real symmetric matrix A (with $n > 0$), an $(n+m) \times m$ matrix N with full rank, and a nonzero vector $t \in \mathbb{R}^m$ with $\left\|(N^\top)^\dagger t\right\| < 1$ (where $(N^\top)^\dagger$ denotes the pseudo-inverse of $N^\top$),

$$\begin{array}{ll} \text{minimize} & x^\top A x \\ \text{subject to} & x^\top x = 1,\ N^\top x = t,\ x \in \mathbb{R}^{n+m}. \end{array}$$

The condition $\left\|(N^\top)^\dagger t\right\| < 1$ ensures that the problem has a solution and is not trivial.

The authors begin by proving that the affine constraint $N^\top x = t$ can be eliminated.

One way to do so is to use a QR decomposition of N.

It turns out that we get a simplified problem of the form

$$\begin{array}{ll} \text{minimize} & z^\top C z + 2z^\top b \\ \text{subject to} & z^\top z = s^2, \; z \in \mathbb{R}^m. \end{array}$$

Unfortunately, if $b \neq 0$, Proposition 17.10 is no longer applicable.

It is still possible to find the minimum of the function $z^\top C z + 2z^\top b$ using Lagrange multipliers, but such a solution is too involved to be presented here.

Interested readers will find a thorough discussion in Gander, Golub, and von Matt [15].

Bibliography

[1] Michael Artin. *Algebra*. Prentice Hall, first edition, 1991.

[2] Marcel Berger. *Géométrie 1*. Nathan, 1990. English edition: Geometry 1, Universitext, Springer Verlag.

[3] Marcel Berger. *Géométrie 2*. Nathan, 1990. English edition: Geometry 2, Universitext, Springer Verlag.

[4] J.E. Bertin. *Algèbre linéaire et géométrie classique*. Masson, first edition, 1981.

[5] Nicolas Bourbaki. *Algèbre, Chapitres 1-3*. Eléments de Mathématiques. Hermann, 1970.

[6] Nicolas Bourbaki. *Algèbre, Chapitres 4-7*. Eléments de Mathématiques. Masson, 1981.

[7] Nicolas Bourbaki. *Espaces Vectoriels Topologiques*. Eléments de Mathématiques. Masson, 1981.

[8] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, first edition, 2004.

[9] P.G. Ciarlet. *Introduction to Numerical Matrix Analysis and Optimization*. Cambridge University Press, first edition, 1989. French edition: Masson, 1994.

[10] Timothée Cour and Jianbo Shi. Solving markov random fields with spectral relaxation. In Marita Meila and Xiaotong Shen, editors, *Artifical Intelligence and Statistics*. Society for Artificial Intelligence and Statistics, 2007.

[11] James W. Demmel. *Applied Numerical Linear Algebra*. SIAM Publications, first edition, 1997.

[12] Jean Dieudonné. *Algèbre Linéaire et Géométrie Elémentaire*. Hermann, second edition, 1965.

[13] David S. Dummit and Richard M. Foote. *Abstract Algebra*. Wiley, second edition, 1999.

[14] Jean H. Gallier. *Geometric Methods and Applications, For Computer Science and Engineering*. TAM, Vol. 38. Springer, second edition, 2011.

[15] Walter Gander, Gene H. Golub, and Urs von Matt. A constrained eigenvalue problem. *Linear Algebra and its Applications*, 114/115:815–839, 1989.

[16] Gene H. Golub. Some modified eigenvalue problems. *SIAM Review*, 15(2):318–334, 1973.

[17] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, third edition, 1996.

[18] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, second edition, 2009.

[19] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, first edition, 1990.

[20] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, first edition, 1994.

[21] Hoffman Kenneth and Kunze Ray. *Linear Algebra*. Prentice Hall, second edition, 1971.

[22] D. Kincaid and W. Cheney. *Numerical Analysis*. Brooks/Cole Publishing, second edition, 1996.

[23] Serge Lang. *Algebra*. Addison Wesley, third edition, 1993.

[24] Serge Lang. *Real and Functional Analysis.* GTM 142. Springer Verlag, third edition, 1996.

[25] Peter Lax. *Linear Algebra and Its Applications.* Wiley, second edition, 2007.

[26] Saunders Mac Lane and Garrett Birkhoff. *Algebra.* Macmillan, first edition, 1967.

[27] Carl D. Meyer. *Matrix Analysis and Applied Linear Algebra.* SIAM, first edition, 2000.

[28] Laurent Schwartz. *Analyse I. Théorie des Ensembles et Topologie.* Collection Enseignement des Sciences. Hermann, 1991.

[29] Denis Serre. *Matrices, Theory and Applications.* GTM No. 216. Springer Verlag, second edition, 2010.

[30] G.W. Stewart. On the early history of the singular value decomposition. *SIAM review*, 35(4):551–566, 1993.

[31] Gilbert Strang. *Introduction to Applied Mathematics.* Wellesley-Cambridge Press, first edition, 1986.

[32] Gilbert Strang. *Linear Algebra and its Applications.* Saunders HBJ, third edition, 1988.

[33] Claude Tisseron. *Géométries affines, projectives, et euclidiennes.* Hermann, first edition, 1994.

[34] L.N. Trefethen and D. Bau III. *Numerical Linear Algebra.* SIAM Publications, first edition, 1997.

[35] B.L. Van Der Waerden. *Algebra, Vol. 1.* Ungar, seventh edition, 1973.

[36] J.H. van Lint and R.M. Wilson. *A Course in Combinatorics.* Cambridge University Press, second edition, 2001.

[37] Stella X. Yu and Jianbo Shi. Grouping with bias. In Thomas G. Dietterich, Sue Becker, and Zoubin Ghahramani, editors, *Neural Information Processing Systems, Vancouver, Canada, 3-8 Dec. 2001.* MIT Press, 2001.