Chapter 12

Applications of Euclidean Geometry to Various Optimization problems

12.1 Applications of the SVD and QR-Decomposition to Least Squares Problems

The method of least squares is a way of "solving" an overdetermined system of linear equations

$$Ax = b,$$

i.e., a system in which A is a rectangular $m \times n$-matrix with more equations than unknowns (when $m > n$).

Historically, the method of least square was used by Gauss and Legendre to solve problems in astronomy and geodesy.

The method was first published by Legendre in 1805 in a paper on methods for determining the orbits of comets.

However, Gauss had already used the method of least squares as early as 1801 to determine the orbit of the asteroid Céres, and he published a paper about it in 1810 after the discovery of the asteroid Pallas. Incidentally, it is in that same paper that Gaussian elimination using pivots is introduced.

The reason why more equations than unknowns arise in such problems is that repeated measurements are taken to minimize errors.

This produces an overdetermined and often inconsistent system of linear equations.

For example, Gauss solved a system of eleven equations in six unknowns to determine the orbit of the asteroid Pallas.

As a concrete illustration, suppose that we observe the motion of a small object, assimilated to a point, in the plane.

From our observations, we suspect that this point moves along a straight line, say of equation $y = dx + c$.

Suppose that we observed the moving point at three different locations (x_1, y_1), (x_2, y_2), and (x_3, y_3).

Then, we should have

$$\begin{aligned} c + dx_1 &= y_1, \\ c + dx_2 &= y_2, \\ c + dx_3 &= y_3. \end{aligned}$$

If there were no errors in our measurements, these equations would be compatible, and c and d would be determined by only two of the equations.

However, in the presence of errors, the system may be inconsistent. Yet, we would like to find c and d!

The idea of the method of least squares is to determine (c, d) so that it minimizes the sum of the squares of the errors, namely

$$(c + dx_1 - y_1)^2 + (c + dx_2 - y_2)^2 + (c + dx_3 - y_3)^2.$$

In general, for an overdetermined $m \times n$ system $Ax = b$, what Gauss and Legendre discovered is that there are solutions x minimizing

$$\|Ax - b\|^2$$

and that these solutions are given by the square $n \times n$ system

$$A^\top Ax = A^\top b.$$

Furthermore, when the columns of A are linearly independent, it turns out that $A^\top A$ is invertible, and so x is unique and given by

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

Note that $A^\top A$ is a symmetric matrix, one of the nice features of the so-called normal equations of a least squares problem. For instance, the normal equations for the above problem are

$$\begin{pmatrix} 3 & x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 & x_1^2 + x_2^2 + x_3^2 \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} y_1 + y_2 + y_3 \\ x_1 y_1 + x_2 y_2 + x_3 y_3 \end{pmatrix}.$$

In fact, given any real $m \times n$-matrix A, there is always a unique x^+ of minimum norm that minimizes $\|Ax - b\|^2$, even when the columns of A are linearly dependent. How do we prove this, and how do we find x^+?

Theorem 12.1.1 *Every linear system $Ax = b$, where A is an $m \times n$-matrix, has a unique least-squares solution x^+ of smallest norm.*

Proof. Geometry offers a nice proof of the existence and uniqueness of x^+.

Indeed, we can interpret b as a point in the Euclidean (affine) space $\mathbb{R}^m$, and the image subspace of A (also called the column space of A) as a subspace U of $\mathbb{R}^m$ (passing through the origin).

Then, we claim that x minimizes $\|Ax - b\|^2$ iff Ax is the orthogonal projection p of b onto the subspace U, which is equivalent to $\mathbf{pb} = b - Ax$ being orthogonal to U.

First of all, if $U^\perp$ is the vector space orthogonal to U, the affine space $b + U^\perp$ intersects U in a unique point p (this follows from lemma 2.9.3).

Next, for any point $y \in U$, the vectors $\mathbf{py}$ and $\mathbf{bp}$ are orthogonal, which implies that

$$\|\mathbf{by}\|^2 = \|\mathbf{bp}\|^2 + \|\mathbf{py}\|^2 .$$

Thus, p is indeed the unique point in U that minimizes the distance from b to any point in U.

To show that there is a unique x^+ of minimum norm minimizing $\|Ax - b\|^2$, we use the fact that

$$\mathbb{R}^n = \mathrm{Ker}\, A \oplus (\mathrm{Ker}\, A)^{\perp}.$$

Indeed, every $x \in \mathbb{R}^n$ can be written uniquely as $x = u + v$, where $u \in \mathrm{Ker}\, A$ and $v \in (\mathrm{Ker}\, A)^{\perp}$, and since u and v are orthogonal,

$$\|x\|^2 = \|u\|^2 + \|v\|^2 .$$

Furthermore, since $u \in \mathrm{Ker}\, A$, we have $Au = 0$, and thus $Ax = p$ iff $Av = p$, which shows that the solutions of $Ax = p$ for which x has minimum norm must belong to $(\mathrm{Ker}\, A)^{\perp}$.

However, the restriction of A to $(\mathrm{Ker}\, A)^{\perp}$ is injective.

This shows that there is a unique x of minimum norm minimizing $\|Ax - b\|^2$, and that it must belong to $(\mathrm{Ker}\, A)^{\perp}$. $\square$

The proof also shows that x minimizes $\|Ax - b\|^2$ iff $\mathbf{pb} = b - Ax$ is orthogonal to U, which can be expressed by saying that $b - Ax$ is orthogonal to every column of A. However, this is equivalent to

$$A^{\top}(b - Ax) = 0, \quad \text{i.e.} \quad A^{\top}Ax = A^{\top}b.$$

Finally, it turns out that the minimum norm least squares solution x^+ can be found in terms of the pseudo-inverse A^+ of A, which is itself obtained from the SVD of A.

If $A = VDU^\top$, with

$$D = diag(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0),$$

where D is an $m \times n$ matrix and $\lambda_i > 0$, letting

$$D^+ = diag(1/\lambda_1, \ldots, 1/\lambda_r, 0, \ldots, 0),$$

an $n \times m$ matrix, the pseudo-inverse of A is defined as

$$A^+ = UD^+V^\top.$$

Actually, it seems that A^+ depends on the specific choice of U and V in an SVD (U, D, V) for A, but the next lemma shows that this is not so.

Theorem 12.1.2 *The least-squares solution of smallest norm of the linear system $Ax = b$, where A is an $m \times n$-matrix, is given by*

$$x^+ = A^+b = UD^+V^\top b.$$

Proof. First, assume that A is a (rectangular) diagonal matrix D, as above. Then, since x minimizes $\|Dx - b\|^2$ iff Dx is the projection of b onto the image subspace F of D, it is fairly obvious that $x^+ = D^+b$.

Otherwise, we can write

$$A = VDU^\top,$$

where U and V are orthogonal. However, since V is an isometry,

$$\|Ax - b\| = \left\|VDU^\top x - b\right\| = \left\|DU^\top x - V^\top b\right\|.$$

Letting $y = U^\top x$, we have $\|x\| = \|y\|$ since U is an isometry, and since U is surjective, $\|Ax - b\|$ is minimized iff $\left\|Dy - V^\top b\right\|$ is minimized, and we showed that the least solution is

$$y^+ = D^+ V^\top b.$$

Since $y = U^\top x$, with $\|x\| = \|y\|$, we get

$$x^+ = U D^+ V^\top b = A^+ b.$$

Thus, the pseudo-inverse provides the optimal solution to the least-squares problem. $\square$

By Lemma 12.1.2 and Theorem 12.1.1, A^+b is uniquely defined by every b, and thus, A^+ depends only on A.

The following properties due to Penrose characterize the pseudo-inverse of a matrix, and give another justification of the uniqueness of A:

Lemma 12.1.3 *Given any $m \times n$-matrix A (real or complex), the pseudo-inverse A^+ of A is the unique $n \times m$-matrix satisfying the following properties:*

$$\begin{aligned} AA^+A &= A, \\ A^+AA^+ &= A^+, \\ (AA^+)^\top &= AA^+, \\ (A^+A)^\top &= A^+A. \end{aligned}$$

If A is an $m \times n$-matrix of rank n (and so, $m \geq n$), it is immediately shown that the QR-decomposition in terms of Householder transformations applies as follows:

There are n $m \times m$-matrices $H_1, \ldots, H_n$, Householder matrices or the identity, and an upper triangular $m \times n$-matrix R or rank n, such that

$$A = H_1 \cdots H_n R.$$

Then, because each H_i is an isometry,

$$\|Ax - b\| = \|Rx - H_n \cdots H_1 b\|,$$

and the least-square problem $Ax = b$ is equivalent to the system

$$Rx = H_n \cdots H_1 b.$$

Now, the system

$$Rx = H_n \cdots H_1 b$$

is of the form

$$\begin{pmatrix} R_1 \\ 0_{m-n} \end{pmatrix} x = \begin{pmatrix} c \\ d \end{pmatrix},$$

where R_1 is an invertible $n \times n$-matrix (since A has rank n), $c \in \mathbb{R}^n$, and $d \in \mathbb{R}^{m-n}$, and the least square solution of smallest norm is

$$x^+ = R_1^{-1} c.$$

Since R_1 is a triangular matrix, it is very easy to invert R_1.

12.2 Minimization of Quadratic Functions Using Lagrange Multipliers

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.

For example, a small ball placed on top of a sphere is in an unstable equilibrium position. A small motion causes the ball to roll down.

On the other hand, a ball placed inside and at the bottom of a sphere is in a stable equilibrium position, because the potential energy is minimal.

The simplest kind of energy function is a quadratic function.

Such functions can be conveniently defined in the form

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

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

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

The question is, under which conditions (on A) does $P(x)$ have a unique global minimum?

It turns out that if A is symmetric positive definite, then $P(x)$ has a unique global minimum precisely when

$$Ax = b.$$

Recall that a symmetric positive definite matrix is a matrix whose eigenvalues are strictly positive. An equivalent criterion is given in the following lemma:

Lemma 12.2.1 *Given any Euclidean space E of dimension n, every self-adjoint linear map $f\colon E \to E$ is positive definite iff*

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

for all $x \neq 0$.

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 $m \geq n$), then $A^{\top}CA$ is symmetric positive definite.

We can now prove that

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

has a global minimum when A is symmetric positive definite.

Lemma 12.2.2 *Given a quadratic function*

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

if A is symmetric positive definite, then $P(x)$ has a unique global minimum for the solution of the linear system $Ax = b$. The minimum value of $P(x)$ is

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

Remarks:

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

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

but the definition using $x^\top b$ is more convenient for the proof of Lemma 12.2.2.

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

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

the proof of Lemma 12.2.2 still shows that $P(x)$ has a unique global minimum for $x = A^{-1}b$, but the minimal value is

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

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

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

where A is symmetric positive definite, finding the global minimum of $P(x)$ 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 satisfied for all admissible solutions. For instance, we may want to minimize the quadratic function

$$Q(y_1, y_2) = \frac{1}{2}(y_1^2 + y_2^2)$$

subject to the constraint

$$2y_1 - y_2 = 5.$$

The solution for which $Q(y_1, y_2)$ is minimum is no longer $(y_1, y_2) = (0, 0)$, but instead, $(y_1, y_2) = (2, -1)$, as will be shown later.

Geometrically, the graph of the function defined by $z = Q(y_1, y_2)$ in $\mathbb{R}^3$ is a paraboloid of revolution P with axis of revolution Oz.

The constraint

$$2y_1 - y_2 = 5$$

corresponds to the vertical plane H parallel to the z-axis and containing the line of equation $2y_1 - y_2 = 5$ 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*.

But first, let us define precisely what kind of minimization problems we intend to solve.

Definition 12.2.3 The *quadratic constrained minimization problem* consists in minimizing a quadratic function

$$Q(y) = \frac{1}{2} y^\top C^{-1} y - b^\top y$$

subject to the linear constraints

$$A^\top y = f,$$

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

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

Since C and C^{-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 $A^\top y = f$ into the quadratic function $Q(y)$, by introducing extra variables $\lambda = (\lambda_1, \ldots, \lambda_n)$ called *Lagrange multipliers*, one for each constraint.

We form the *Lagrangian*

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

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

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

which can be written in matrix form as

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

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

$$C^{-1}y + A\lambda = b,$$

we get

$$y = C(b - A\lambda),$$

and substituting into the second equation, we get

$$A^{\top}C(b - A\lambda) = f,$$

that is,

$$A^{\top}CA\lambda = A^{\top}Cb - f.$$

However, by a previous remark, since C is symmetric positive definite and the columns of A are linearly independent, $A^{\top}CA$ is symmetric positive definite, and thus invertible.

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

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

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

is equivalent to the system

$$\begin{aligned} e &= b - A\lambda, \\ y &= Ce, \\ A^{\top} y &= f. \end{aligned}$$

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

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

This includes spring mass systems, electrical networks, and trusses, which are structures built from elastic bars.

In each case, y, e, b, C, λ, f, and $K = A^{\top}CA$ have a physical interpretation.

The matrix $K = A^{\top}CA$ is usually called the *stiffness matrix*. Again, the reader is referred to Strang [**?**].

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

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

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

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

by Lemma 12.2.2 the global minimum (with respect to y) of $L(y, \lambda)$ is obtained for the solution y of

$$C^{-1} y = b - A\lambda,$$

that is, when

$$y = C(b - A\lambda),$$

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

$$\min_y L(y, \lambda) = -\frac{1}{2}(A\lambda - b)^\top C(A\lambda - b) - \lambda^\top f.$$

Letting

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

we claim that the solution of the constrained minimization of $Q(y)$ subject to $A^\top y = f$ is equivalent to the unconstrained maximization of $-P(\lambda)$.

Of course, since we minimized $L(y, \lambda)$ w.r.t. y, we have

$$L(y, \lambda) \geq -P(\lambda)$$

for all y and all λ. However, when the constraint $A^\top y = f$ holds, $L(y, \lambda) = Q(y)$, and thus for any admissible y, which means that $A^\top y = f$, we have

$$\min_y Q(y) \geq \max_\lambda -P(\lambda).$$

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

Lemma 12.2.4 *The quadratic constrained minimization problem of Definition 12.2.3 has a unique solution (y, λ) given by the system*

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

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

Remarks:

(1) There is a form of duality going on in this situation.

The constrained minimization of $Q(y)$ subject to $A^\top y = f$ is called the *primal problem*, and the unconstrained maximization of $-P(\lambda)$ is called the *dual problem*.

Duality is the fact stated slightly loosely as

$$\min_y Q(y) = \max_\lambda -P(\lambda).$$

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

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

we can also write

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

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 $-P(\lambda)$).

(2) It is immediately verified that the equations of Lemma 12.2.4 are equivalent to the equations stating that the partial derivatives of the Lagrangian $L(y, \lambda)$ are null:

$$\frac{\partial L}{\partial y_i} = 0, \quad i = 1, \ldots, m,$$
$$\frac{\partial L}{\partial \lambda_j} = 0, \quad j = 1, \ldots, n.$$

Thus, the constrained minimum of $Q(y)$ subject to $A^\top y = f$ is an extremum of the Lagrangian $L(y, \lambda)$.

As we showed in Lemma 12.2.4, this extremum corresponds to simultaneously minimizing $L(y, \lambda)$ with respect to y and maximizing $L(y, \lambda)$ with respect to λ.

Geometrically, such a point is a *saddle point* for $L(y, \lambda)$.

(3) The Lagrange multipliers sometimes have a natural physical meaning. For example, in the spring mass system they correspond to node displacements.

In some general sense, Lagrange multipliers are correction terms needed to satisfy equilibrium equations and the price paid for the constraints. For more details, see Strang [**?**].

Going back to the constrained minimization of $Q(y_1, y_2) = \frac{1}{2}(y_1^2 + y_2^2)$ subject to $2y_1 - y_2 = 5$, the Lagrangian is

$$L(y_1, y_2, \lambda) = \frac{1}{2}(y_1^2 + y_2^2) + \lambda(2y_1 - y_2 - 5),$$

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

$$\begin{aligned} y_1 + 2\lambda &= 0, \\ y_2 - \lambda &= 0, \\ 2y_1 - y_2 - 5 &= 0. \end{aligned}$$

We obtain the solution $(y_1, y_2, \lambda) = (2, -1, -1)$.

Much more should be said about the use of Lagrange multipliers in optimization or variational problems. This is a vast topic.

Least squares methods and Lagrange multipliers are used to tackle many problems in computer graphics and computer vision; see Trucco and Verri [**?**], Metaxas [**?**], Jain, Katsuri, and Schunck [**?**], Faugeras [**?**], and Foley, van Dam, Feiner, and Hughes [**?**].

For a lucid introduction to optimization methods, see Ciarlet [**?**].