Unconstrained Quadratic Optimization

Thomas R. Cameron

1 Introduction

In the previous two lectures, we studied quadratic programming and the geometry of quadratic functions. We saw that the gradient of a quadratic function is linear, that the Hessian matrix is constant, and that the eigenvalues and eigenvectors of the Hessian determine the shape of the graph and its level sets. In this lecture, we specialize to the unconstrained problem. Our aim is to completely classify the behavior of a quadratic objective of the form

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x},$$

where $Q \in \mathbb{R}^{n\times n}$ is symmetric and $\mathbf{c} \in \mathbb{R}^n$. Since there are no constraints, all of the interesting behavior comes from the interaction between the linear term $\mathbf{c}^T\mathbf{x}$ and the spectral structure of Q. We seek to understand the existence and uniqueness of the minimizers or maximizers of f.

2 Critical Points and the Linear System

Recall that

$$\nabla f(\mathbf{x}) = Q\mathbf{x} + \mathbf{c}.$$

Therefore, a critical point of f is a point $\mathbf{x}^*$ satisfying

$$Q\mathbf{x}^* = -\mathbf{c}.$$

If Q is invertible, then there is a unique critical point,

$$\mathbf{x}^* = -Q^{-1}\mathbf{c}.$$

If Q is singular, then two possibilities arise: either there are infinitely many critical points or no critical points.

Example

Consider the quadratic $f(\mathbf{x})$ described by

$$Q = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} -2 \\ -4 \end{pmatrix}.$$

The eigenvalues of Q are

$$\lambda_1 = 3 - \sqrt{2}, \qquad \lambda_2 = 3 + \sqrt{2},$$

which implies that Q is invertible. Hence, this quadratic has a unique critical point

$$\mathbf{x}^* = -Q^{-1}\mathbf{c} = \begin{bmatrix} 10/7 \\ 8/7 \end{bmatrix}.$$

Furthermore, since the eigenvalues of Q are both positive it follows that Q is positive definite. Therefore, the quadratic $f(\mathbf{x})$ is convex and it follows that $\mathbf{x}^*$ is an absolute minimum.

Example

Consider the quadratic $f(\mathbf{x})$ described by

$$Q = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \qquad \mathbf{c} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

The eigenvalues of Q are

$$\lambda_1 = 0, \qquad \lambda_2 = 2,$$

which implies that Q is singular. Hence, the quadratic either has no critical points or infinitely many critical points. In this case, the linear system

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

has infinitely many solutions which can be written in the form

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

where $t \in \mathbb{R}$. Furthermore, since Q is positive semidefinite, the quadratic $f(\mathbf{x})$ is convex. Therefore, each one of these critical points is an absolute minimum.

Example

Consider the quadratic $f(\mathbf{x})$ described by

$$Q = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \qquad \mathbf{c} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}.$$

Again, Q is singular. This time, the linear system

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

has no solution. Therefore the quadratic $f(\mathbf{x})$ has no local extrema.

3 Singular Matrices

Linear systems with singular matrices have many important applications and occur frequently in quadratic programming. Let $A \in \mathbb{R}^{n \times n}$ and $\mathbf{b} \in \mathbb{R}^n$. Consider the linear system

$$A\mathbf{x} = \mathbf{b}.$$

This system has a solution if and only if $\mathbf{b}$ is in the column space of A, which we denote by

$$\text{col}(A) = \{A\mathbf{x} \colon \mathbf{x} \in \mathbb{R}^n\}.$$

It is known that A is invertible if and only if the columns of A are linearly independent which is true if and only if $\text{col}(A) = \mathbb{R}^n$.

The null space of a matrix is defined by

$$\text{nul}(A) = \{\mathbf{x} \colon A\mathbf{x} = 0\}$$

Both col(A) and nul(A) are examples of a subspace of $\mathbb{R}^n$. A subspace is a collection of vectors that is closed under linear combinations. Given a subspace $S \subseteq \mathbb{R}^n$, we let dim(S) denotes the dimension of the space, which is the largest collection of linearly independent vectors from that space. The rank-nullity theorem states that

$$\dim(\text{col}(A)) + \dim(\text{nul}(A)) = n$$

Therefore, A is invertible if and only if $\dim(\text{col}(A)) = n$ which is true if and only if $\dim(\text{nul}(A)) = 0$.

Suppose that A is symmetric. Then the column space and null space are orthogonal. Indeed, let $\mathbf{x} \in \text{nul}(A)$ and $\mathbf{y} \in \text{col}(A)$. Then, $\mathbf{y} = A\mathbf{w}$ for some $\mathbf{w} \in \mathbb{R}^n$ and

$$\begin{aligned} \mathbf{y}^T\mathbf{x} &= (A\mathbf{w})^T \mathbf{x} \\ &= \mathbf{w}^T A^T \mathbf{x} \\ &= \mathbf{w}^T A\mathbf{x} = 0. \end{aligned}$$

Since A is symmetric, the spectral theorem guarantees that there exists an orthogonal matrix $V \in \mathbb{R}^{n\times n}$ and a diagonal matrix $D \in \mathbb{R}^{n\times n}$ such that

$$A = VDV^T,$$

where the columns of V are eigenvectors of A corresponding to the eigenvalues along the main diagonal of D. If A is singular, then it has a zero eigenvalue. Without loss of generality, let $\lambda_1, \dots, \lambda_k$ be the zero eigenvalues of A. Then, $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ forms a basis for nul(A) and $\{\mathbf{v}_{k+1}, \dots, \mathbf{v}_n\}$ forms a basis for col(A). Note that the linear system $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b} \in \text{col}(A)$ which is true if and only if $\mathbf{b}$ is orthogonal to every vector in nul(A) (in particular, the eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_k$).

3.1 The Minimum Norm Solution

Suppose that A is singular and $\lambda_1, \dots, \lambda_k$ are the zero eigenvalues of A. Also, suppose that the linear system $A\mathbf{x} = \mathbf{b}$ has a solution. Then, $\mathbf{b} \in \text{col}(A)$ and it follows that $\mathbf{b}$ is orthogonal to $\mathbf{v}_1, \dots, \mathbf{v}_k$.

We rewrite the linear system using the spectral decomposition of A, that is,

$$VDV^T\mathbf{x} = \mathbf{b} \quad \Longleftrightarrow \quad DV^T\mathbf{x} = V^T\mathbf{b} \quad \Longleftrightarrow \quad D\mathbf{y} = V^T\mathbf{b},$$

where $\mathbf{y} = V^T\mathbf{x}$. Since $\mathbf{b}$ is orthogonal to every vector in nul(A), the first k entries of $V^T\mathbf{b}$ are zero. Similarly, the first k entries of $D\mathbf{y}$ are zero. Hence, we can select any value we like for $y_1, \dots, y_k$. Since we desire the minimum norm solution, we set $y_1 = \cdots = y_k = 0$. The remaining values of $\mathbf{y}$ can be solved directly by dividing by the corresponding non-zero eigenvalue. Note that $\mathbf{x} = V\mathbf{y}$ is the final minimum norm solution where

$$y_i = \begin{cases} 0 & \text{if } \lambda_i = 0, \\ \frac{\mathbf{v}_i^T\mathbf{b}}{\lambda_i} & \text{if } \lambda_i \neq 0. \end{cases}$$

It is worth noting that since V is orthogonal, the norm of $\mathbf{y}$ and the norm of $\mathbf{x}$ are the same. Hence, minimizing the norm of $\mathbf{y}$ is equivalent to minimizing the norm of $\mathbf{x}$.

Example

Consider the linear system $A\mathbf{x} = \mathbf{x}$, where

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Then, spectral decomposition $A = VDV^T$ is given by

$$V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \qquad D = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}.$$

Since A is singular and $\mathbf{b} \in \mathrm{col}(A)$, we know there are infinitely many solutions. The minimum norm solution is given by $\mathbf{x} = V\mathbf{y}$, where $D\mathbf{y} = V^T\mathbf{b}$, that is

$$\begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{2}{\sqrt{2}} \end{bmatrix}.$$

Therefore, $y_1 = 0$, $y_2 = 1/\sqrt{2}$, and

$$\mathbf{x} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

3.2 The Least Squares Solution

Suppose that A is singular and $\lambda_1, \ldots, \lambda_k$ are the zero eigenvalues of A. Also, suppose that the linear system $A\mathbf{x} = \mathbf{b}$ does not have a solution. The least squares solution, or nearest solution, is an $\mathbf{x}$ that minimize the residual

$$\mathbf{r} = \mathbf{b} - A\mathbf{x}.$$

Recall that $A = VDV^T$. Since V is orthogonal, $V^T\mathbf{r}$ has the same norm as $\mathbf{r}$. Furthermore,

$$\begin{aligned} V^T\mathbf{r} &= V^T\mathbf{b} - A\mathbf{x} \\ &= V^T\mathbf{b} - DV^T\mathbf{x} \\ &= \begin{bmatrix} \hat{\mathbf{r}} \\ \hat{V}^T\mathbf{b} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \hat{D}\hat{V}^T\mathbf{x} \end{bmatrix}, \end{aligned}$$

where $\hat{V}$ denotes the $n \times (n-k)$ matrix whose columns correspond to the eigenvectors $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$, $\hat{D}$ denotes the diagonal $(n-k) \times (n-k)$ matrix whose main diagonal entries correspond to the non-zero eigenvalues $\lambda_{k+1}, \ldots, \lambda_n$, and

$$\hat{r}_i = \mathbf{v}_i^T\mathbf{b}, \quad 1 \leq i \leq k,$$

denotes the uncontrolled error in the solution. Therefore, the least squares solution is given by

$$\hat{D}\hat{V}^T\mathbf{x} = \hat{V}^T\mathbf{b} \quad \Longleftrightarrow \quad \hat{D}\hat{\mathbf{y}} = \hat{V}^T\mathbf{b},$$

where $\hat{\mathbf{y}} = \hat{V}^T\mathbf{x}$. Since the diagonal entries of $\hat{D}$ are non-zero, the above system has a unique solution

$$\hat{y}_i = \frac{\mathbf{v}_i^T\mathbf{b}}{\lambda_i}, \quad k+1 \leq i \leq n.$$

It is worth noting that $\mathbf{x} = \hat{V}\hat{\mathbf{y}}$ is the least squares solution of minimum norm.

Example

Consider the linear system $A\mathbf{x} = \mathbf{x}$, where

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Again, spectral decomposition $A = VDV^T$ is given by

$$V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \qquad D = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}.$$

Since A is singular and $\mathbf{b} \notin \text{col}(A)$, we seek the least squares solution. To this end, note that $\hat{D} = [2]$ and

$$\hat{V} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The least squares solution is given by $\mathbf{x} = \hat{V}\hat{\mathbf{y}}$, where $\hat{D}\hat{\mathbf{y}} = \hat{V}^T\mathbf{b}$, that is,

$$2y_2 = \frac{3}{\sqrt{2}}.$$

Therefore, $y_2 = \frac{3}{2\sqrt{2}}$ and

$$\mathbf{x} = \begin{bmatrix} \frac{3}{4} \\ \frac{3}{4} \end{bmatrix}.$$

4 Computational Remarks

The spectral theorem gives the cleanest theoretical understanding of unconstrained quadratic optimization, but it is not the tool one would normally use to solve a problem numerically. In computation, one usually distinguishes three tasks: solving the linear system when it is nonsingular, recognizing definiteness, and computing minimum-norm or least-squares solutions in singular or inconsistent cases.

If Q is positive definite, then the preferred direct method is the Cholesky factorization

$$Q = R^T R.$$

This is both efficient and numerically stable. Moreover, a successful Cholesky factorization certifies that Q is positive definite. If Q is symmetric but not known to be positive definite, then an LDL^T factorization is the standard tool. Its pivot structure reveals whether the matrix is positive definite, negative definite, semidefinite, or indefinite.

When one wants a minimum-norm solution or a nearest least-squares solution, one solves the associated least-squares problem by QR decomposition, or, in rank-deficient situations, by singular value decomposition (spectral decomposition in symmetric case). This parallels exactly what one does in numerical linear algebra for inconsistent or underdetermined systems. Regularization provides yet another practical strategy: instead of solving the singular problem directly, one solves

$$(Q + \varepsilon I)\mathbf{x} = -\mathbf{c}$$

for a small $\varepsilon > 0$.

5 Summary

The unconstrained quadratic problem is completely determined by the pair $(Q, \mathbf{c})$. First, one asks whether the critical point equation $Q\mathbf{x} = -\mathbf{c}$ has a solution. If it does not, then the quadratic has no critical point and is unbounded. If it does, then one studies the signs of the eigenvalues of Q. Positive definite matrices give unique global minima, negative definite matrices give unique global maxima, semidefinite matrices lead to nonunique extrema along flat directions, and indefinite matrices produce saddle points. When singularity causes nonuniqueness, one may select the minimum-norm solution. When inconsistency prevents the existence of a critical point, one may instead solve the nearest least-squares problem. The following diagram summarizes all cases.

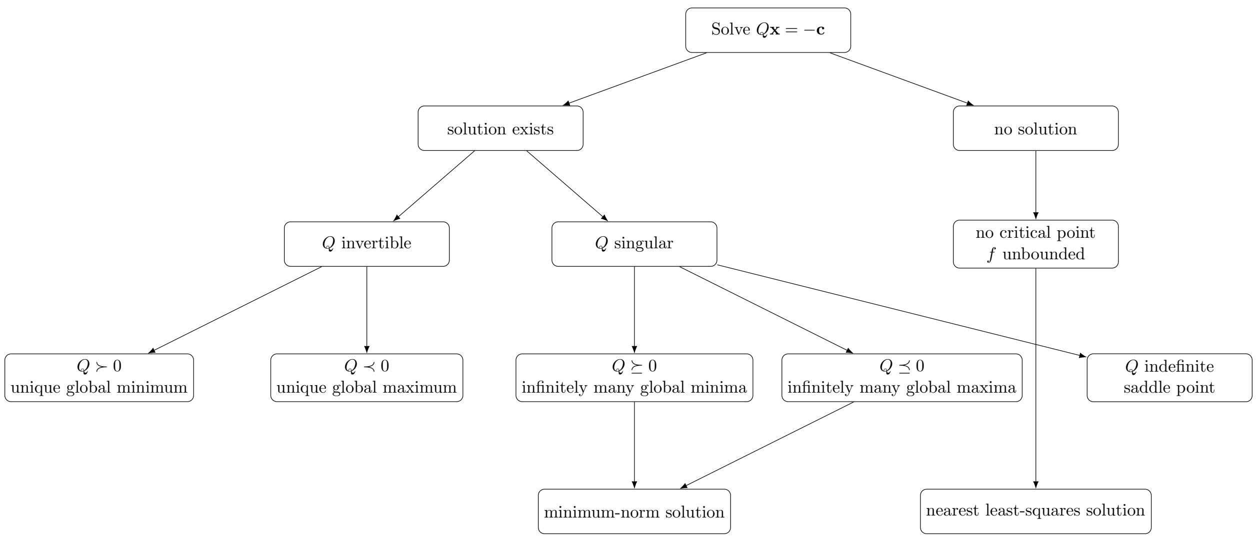

Figure 1: Classification of unconstrained quadratic optimization.