Quadratic forms

Each $A \in \mathbb{R}^{n\times n}$ gives rise to a quadratic form, via

$$q_A : \mathbb{R}^n \to \mathbb{R} : x \mapsto x^{\rm t}Ax.$$

However, as we already observed, the quadratic form 'sees' only the **symmetric part**

$$(A + A^{\rm t})/2$$

of A, i.e.,

$$\forall x \in \mathbb{R}^n \quad x^{\rm t}Ax = x^{\rm t}(A + A^{\rm t})/2\ x.$$

For this reason, in discussions of the quadratic form q_A, we will always assume that A is real symmetric.

The Taylor expansion for q_A is very simple. One computes

$$q_A(x+h) = (x+h)^{\rm t}A(x+h) = x^{\rm t}Ax + x^{\rm t}Ah + h^{\rm t}Ax + h^{\rm t}Ah = q_A(x) + 2(Ax)^{\rm t}h + h^{\rm t}Ah,$$

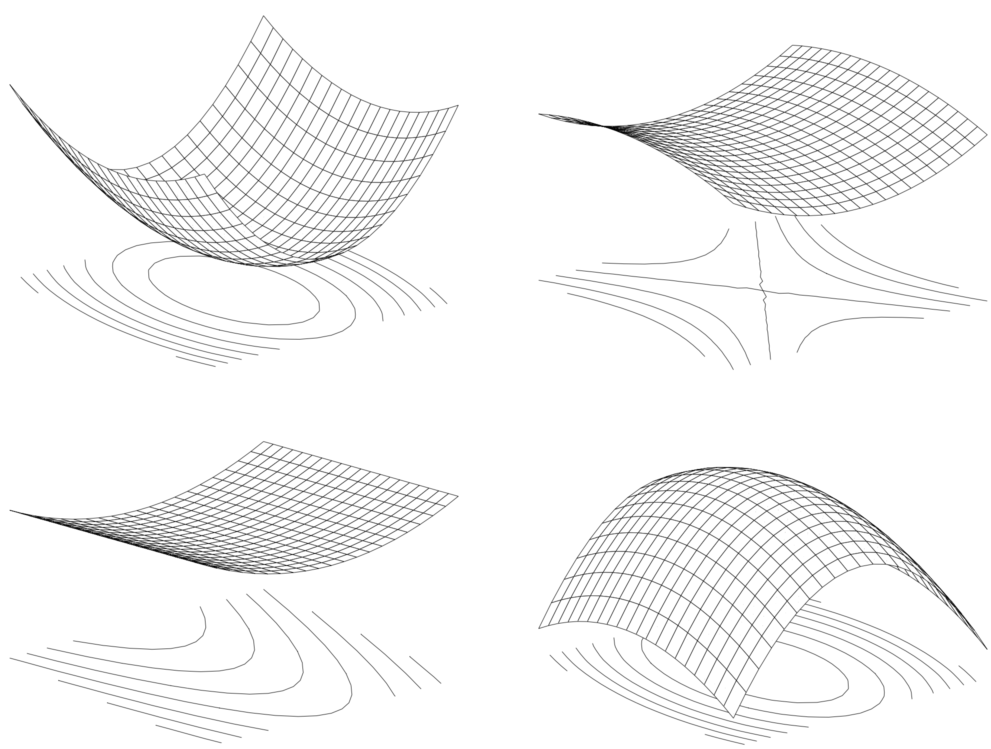
using the fact that $A^{\rm t} = A$, thus $h^{\rm t}Ax = x^{\rm t}Ah = (Ax)^{\rm t}h$, hence

$$Dq_A(x) = 2Ax, \quad D^2q_A(x) = 2A.$$

It follows that, for any 1-1 A, 0 is the only critical point of q_A. The sought-for classification of critical points of smooth functions has led to the following classification of quadratic forms:

$$A \text{ is } \begin{array}{c}\textbf{positive}\\ \textbf{positive semi-}\\ \textbf{negative semi-}\\ \textbf{negative}\end{array}\textbf{definite} := \begin{array}{ll}\forall x \ne 0 & x^{\rm t}Ax > 0\\ \forall x & x^{\rm t}Ax \ge 0\\ \forall x & x^{\rm t}Ax \le 0\\ \forall x \ne 0 & x^{\rm t}Ax < 0\end{array} \iff 0 \text{ is } \begin{array}{l}\textbf{the unique minimizer}\\ \textbf{a minimizer}\\ \textbf{a maximizer}\\ \textbf{the unique maximizer}\end{array} \text{ for } q_A.$$

If none of these conditions obtains, i.e., if there exist x and y so that $x^{\rm t}Ax < 0 < y^{\rm t}Ay$, then q_A is called **indefinite** and, in this case, 0 is a **saddle point** for q_A.



(15.1) Figure. Local behavior near a critical point.

(15.1)Figure shows three quadratic forms near their unique critical point. One is a minimizer, another is a saddle point, and the last one is a maximizer. Also shown is a quadratic form with a whole straight line of critical points. The figure (generated by the `MATLAB` command `meshc`) also shows some **contour lines** or **level lines**, i.e., lines in the domain $\mathbb{R}^2$ along which the function is constant. The contour plots are characteristic: Near an extreme point, be it a maximum or a minimum, the level lines are ellipses, with the extreme point their center, while near a saddle point, the level lines are hyperbolas, with the extreme point their center and with two level lines actually crossing at the saddle point.

There is an intermediate case between these two, also shown in (15.1)Figure, in which the level lines are parabolas and, correspondingly, there is a whole line of critical points. In this case, the quadratic form is semidefinite. Note, however, that the *definition* of semidefiniteness does not exclude the possibility that the quadratic form is actually definite.

Since, near any critical point x, a smooth f behaves like its quadratic term $h \mapsto h^{\mathrm{t}}(D^2 f(x)/2)h$, we can be sure that a contour plot for f near an extremum would approximately look like concentric ellipses while, near a saddle point, it would look approximately like concentric hyperbolas.

These two patterns turn out to be the only two possible ones for definite quadratic forms on $\mathbb{R}^2$. On $\mathbb{R}^n$, there are only $\lceil (n+1)/2 \rceil$ possible distinct patterns, as follows from the fact that, *for every quadratic form q_A, there are o.n. coordinate systems U for which*

$$q_A(x) = \sum_{i=1}^{n} d_i \, (U^{\mathrm{c}} x)_i^2.$$

15.1 For each of the following three functions on $\mathbb{R}^2$, compute the Hessian $D^2 f(0)$ at 0 and use it to determine whether 0 is a (local) maximum, minimum, or neither. (In an effort to make the derivation of the Hessians simple, I have made the problems so simple that you could tell by inspection what kind of critical point $0 = (0,0) \in \mathbb{R}^2$ is; nevertheless, give your answer based on the spectrum of the Hessian.)

(a) $f(x,y) = (x-y)\sin(x+y)$

(b) $f(x,y) = (x+y)\sin(x+y)$

(b) $f(x,y) = (x+y)\cos(x+y)$.

Reduction of a quadratic form to a sum of squares

Consider the effects of a *change of basis*. Let $V \in \mathbb{R}^n$ be a basis for $\mathbb{R}^n$ and consider the map

$$f := q_A \circ V.$$

We have $f(x) = (Vx)^{\mathrm{t}} A V x = x^{\mathrm{t}}(V^{\mathrm{t}} A V)x$, hence

$$q_A \circ V = q_{V^{\mathrm{t}} A V}.$$

This makes it interesting to look for bases V for which $V^{\mathrm{t}} A V$ is as simple as possible. Matrices A and B for which $B = V^{\mathrm{t}} A V$ are said to be **congruent** to each other. Note that congruent matrices are not necessarily similar; in particular, their spectra can be different. However, by Sylvester's Law of Inertia (see (15.9) below), congruent *hermitian* matrices have the same number of positive, of zero, and of negative, eigenvalues. This is not too surprising in view of the following *reduction to a sum of squares* which is possible for any quadratic form.

(15.2) Proposition: Every quadratic form q_A on $\mathbb{R}^n$ can be written in the form

$$q_A(x) = \sum_{j=1}^{n} d_j ({u_j}^{\mathrm{t}} x)^2,$$

for some suitable o.n. basis $U = [u_1, u_2, \ldots, u_n]$ for which $U^{\mathrm{t}} A U = \mathrm{diag}(d_1, \ldots, d_n) \in \mathbb{R}^n$.

Proof: Since A is hermitian, there exists, by (12.2)Corollary, some o.n. basis $U = [u_1, u_2, \ldots, u_n]$ for $\mathbb{F}^n$ for which $U^{\mathrm{t}}AU = \mathrm{diag}(d_1, d_2, \ldots, d_n) \in \mathbb{R}^{n\times n}$. Now use the facts that $U^{\mathrm{t}}U = \mathrm{id}_n$ and therefore $q_A(x) = q_{U^{\mathrm{t}}AU}(U^{\mathrm{t}}x)$ to obtain for $q_A(x)$ the displayed expression. □

What about the classification introduced earlier, into positive or negative (semidefinite)? The proposition permits us to visualize $q_A(x)$ as a weighted sum of squares (with real weights $d_1, d_2, \ldots, d_n$) and $U^{\mathrm{t}}x$ an arbitrary n-vector (since U is a basis), hence permits us to conclude that q_A is definite if and only if all the d_j are strictly of one sign, semidefinite if and only if all the d_j are of one sign (with zero possible), and indefinite if and only if there are both positive and negative d_j.

`MATLAB` readily provides these numbers d_j by the command `eig(A)`.

Consider specifically the case $n = 2$ for which we earlier provided some pictures. Assume without loss that $d_1 \le d_2$. If $0 < d_1$, then A is positive definite and, correspondingly, the contour line

$$c_r := \{x \in \mathbb{R}^2 : q_A(x) = r\} = \{x \in \mathbb{R}^2 : d_1({u_1}^{\mathrm{t}}x)^2 + d_2({u_2}^{\mathrm{t}}x)^2 = r\}$$

for $r > 0$ is an ellipse, with axes parallel to u_1 and u_2. If $0 = d_1 < d_2$, then these ellipses turn into straight lines. Similarly, if $d_2 < 0$, then the contour line

$$c_r := \{x \in \mathbb{R}^2 : q_A(x) = r\} = \{x \in \mathbb{R}^2 : d_1({u_1}^{\mathrm{t}}x)^2 + d_2({u_2}^{\mathrm{t}}x)^2 = r\}$$

for $r < 0$ is an ellipse, with axes parallel to u_1 and u_2. Finally, if $d_1 < 0 < d_2$, then, for any r, the contour line

$$c_r := \{x \in \mathbb{R}^2 : q_A(x) = r\} = \{x \in \mathbb{R}^2 : d_1({u_1}^{\mathrm{t}}x)^2 + d_2({u_2}^{\mathrm{t}}x)^2 = r\}$$

is a hyperbola, with axes parallel to u_1 and u_2.

Note that such an o.n. basis U is Cartesian, i.e., its columns are orthogonal to each other (and are normalized). This means that we can visualize the change of basis, from the natural basis to the o.n. basis U, as a rigid motion, involving nothing more than rotations and reflections.

Rayleigh quotient

This section is devoted to the proof and exploitation of the following remarkable

> **Fact:** The eigenvectors of a hermitian matrix A are the critical points of the corresponding **Rayleigh quotient**
>
> $$R_A(x) := \langle Ax, x\rangle / \langle x, x\rangle,$$
>
> and $R_A(x) = \mu$ in case $Ax = \mu x$.

This fact has many important consequences concerning how the eigenvalues of a hermitian matrix depend on that matrix, i.e., how the eigenvalues change when the entries of the matrix are changed, by round-off or for other reasons.

This perhaps surprising connection has the following intuitive explanation: Suppose that $Ax \notin \mathrm{ran}[x]$. Then the error $h := Ax - R_A(x)x$ in the least-squares approximation to Ax from $\mathrm{ran}[x]$ is not zero, and is perpendicular to $\mathrm{ran}[x]$. Consequently, $\langle Ax, h\rangle = \langle h, h\rangle > 0$, and therefore the value

$$\langle A(x + th), x + th\rangle = \langle Ax, x\rangle + 2t\langle Ax, h\rangle + t^2\langle Ah, h\rangle$$

of the numerator of $R_A(x + th)$ grows linearly for positive t, while its denominator

$$\langle x + th, x + th\rangle = \langle x, x\rangle + t^2\langle h, h\rangle$$

grows only quadratically, i.e., much less fast for t near zero. It follows that, in this situation, $R_A(x+th) > R_A(x)$ for all 'small' positive t, hence x cannot be a critical point for R_A. – To put it differently, for any critical point x for R_A, we necessarily have $Ax \in \mathrm{ran}[x]$, therefore $Ax = R_A(x)x$. Of course, that makes any such x an eigenvector with corresponding eigenvalue $R_A(x)$. □

Next, recall from (12.2) that a hermitian matrix is unitarily similar to a real diagonal matrix. This means that we may assume, after some reordering if necessary, that

$$A = UDU^{\mathrm{c}}$$

with U unitary and with $\mathrm{M} = \mathrm{diag}(\mu_1, \ldots, \mu_n)$ where

$$\mu_1 \le \mu_2 \le \cdots \le \mu_n.$$

At times, we will write, more explicitly,

$$\mu_j(A)$$

to denote the jth eigenvalue of the hermitian matrix A in this ordering. Note that there may be coincidences here, i.e., $\mu_j(A)$ is the jth smallest eigenvalue of A *counting multiplicities*. Note also that, in contrast to the singular values (and in contrast to most books), we have put here the eigenvalues in *increasing* order.

Now recall that a unitary basis has the advantage that it preserves angles and lengths since $\langle Ux, Uy\rangle = \langle x, y\rangle$ for any orthonormal U. Thus

$$\langle Ax, x\rangle = \langle U\mathrm{M}U^{\mathrm{c}}x, x\rangle = \langle \mathrm{M}(U^{\mathrm{c}}x), U^{\mathrm{c}}x\rangle,$$

and $\langle x, x\rangle = \langle U^{\mathrm{c}}x, U^{\mathrm{c}}x\rangle$. Therefore

$$R_A(x) = \langle Ax, x\rangle/\langle x, x\rangle = \langle \mathrm{M}(U^{\mathrm{c}}x), U^{\mathrm{c}}x\rangle/\langle U^{\mathrm{c}}x, U^{\mathrm{c}}x\rangle = R_{\mathrm{M}}(U^{\mathrm{c}}x).$$

This implies that

$$\begin{matrix}\max_x\\ \min_x\end{matrix} R_A(x) = \begin{matrix}\max_y\\ \min_y\end{matrix} R_{\mathrm{M}}(y).$$

On the other hand, since M is diagonal, $\langle \mathrm{M}y, y\rangle = \sum_j \mu_j |y_j|^2$, therefore

$$R_{\mathrm{M}}(y) = \sum_j \mu_j |y_j|^2 / \sum_j |y_j|^2,$$

and this shows that

$$\min_x R_A(x) = \min_y R_{\mathrm{M}}(y) = \mu_1, \qquad \max_x R_A(x) = \max_y R_{\mathrm{M}}(y) = \mu_n.$$

This is **Rayleigh's Principle**. It characterizes the extreme eigenvalues of a hermitian matrix. The intermediate eigenvalues are the solution of more subtle extremum problems. This is the content of the **Courant-Fischer minimax Theorem** and the **?.?. maximin Theorem**. It seems most efficient to combine both in the following

(15.3) MMM (or, **maximinimaxi**) **Theorem:** Let A be a hermitian matrix of order n, hence $A = U\mathrm{M}U^{\mathrm{c}}$ for some unitary U and some real diagonal matrix $\mathrm{M} = \mathrm{diag}(\cdots, \mu_j, \ldots)$ with $\mu_1 \le \cdots \le \mu_n$. Then, for $j = 1{:}n$,

$$\max_{\dim G < j} \min_{x \perp G} R_A(x) = \mu_j = \min_{j \le \dim H} \max_{x \in H} R_A(x),$$

with G and H otherwise arbitrary linear subspaces.

Proof: If $\dim G < j \le \dim H$, then one can find $y \in H\backslash 0$ with $y \perp G$ (since, with V a basis for G and W a basis for H, this amounts to finding a nontrivial solution to the equation $V^{\mathrm{c}}W? = 0$, and this system is homogeneous with more unknowns than equations). Therefore

$$\min_{x \perp G} R_A(x) \le R_A(y) \le \max_{x \in H} R_A(x).$$

Hence,

$$\max_{\dim G < j} \min_{x \perp G} R_A(x) \le \min_{j \le \dim H} \max_{x \in H} R_A(x).$$

On the other hand, for $G = \mathrm{ran}[u_1, \ldots, u_{j-1}]$ and $H = \mathrm{ran}[u_1, \ldots, u_j]$,

$$\min_{x \perp G} R_A(x) = \mu_j(A) = \max_{x \in H} R_A(x).$$

□

The MMM theorem has various useful (and immediate) corollaries.

(15.4) Interlacing Theorem: If the matrix B is obtained from the hermitian matrix A by crossing out the kth row and column (i.e., $B = A(I,I)$ with $I := (1{:}k-1, k+1{:}n)$), then

$$\mu_j(A) \le \mu_j(B) \le \mu_{j+1}(A), \quad j < n.$$

Proof: It is sufficient to consider the case $k = n$, since we can always achieve this situation by interchanging rows k and n, and columns k and n, of A, and this will not change $\operatorname{spec}(A)$. Let $J : \mathbb{F}^{n-1} \to \mathbb{F}^n : x \mapsto (x, 0)$. Then $R_B(x) = R_A(Jx)$ and $\operatorname{ran} J = \operatorname{ran}[e_n]\perp$, therefore also $J(G\perp) = (JG + \operatorname{ran}[e_n])\perp$ and $\{JG + \operatorname{ran}[e_n] : \dim G < j, G \subset \mathbb{F}^{n-1}\} \subset \{\tilde G : \dim \tilde G < j+1, \tilde G \subset \mathbb{F}^n\}$. Hence

$$\mu_j(B) = \max_{\dim G<j} \min_{x\perp G} R_A(Jx) = \max_{\dim G<j} \min_{y\perp JG+\operatorname{ran}[e_n]} R_A(y) \le \max_{\dim \tilde G<j+1} \min_{y\perp \tilde G} R_A(y) = \mu_{j+1}(A).$$

Also, since $\{JH : j \le \dim H, H \subset \mathbb{F}^{n-1}\} \subset \{\tilde H : j \le \dim \tilde H, \tilde H \subset \mathbb{F}^n\}$,

$$\mu_j(B) = \min_{j\le \dim H} \max_{x\in H} R_A(Jx) = \min_{j\le \dim H} \max_{y\in JH} R_A(y) \ge \min_{j\le \dim \tilde H} \max_{y\in \tilde H} R_A(y) = \mu_j(A).$$

□

(15.5) Corollary: If $A = \begin{bmatrix} B & C \\ D & E \end{bmatrix} \in \mathbb{F}^{n\times n}$ is hermitian, and $B \in \mathbb{F}^{r\times r}$, then at least r eigenvalues of A must be $\le \max\operatorname{spec}(B)$ and at least r eigenvalues of A must be $\ge \min\operatorname{spec}(B)$.

In particular, if the spectrum of B is negative and the spectrum of E is positive, then A has exactly r negative, and $n-r$ positive, eigenvalues.

A different, simpler, application of the MMM theorem is based on the following observation: If

$$f(t) \le g(t) \quad \forall t,$$

then this inequality persists if we take on both sides the maximum or minimum over the same set T, i.e., then

$$\max_{t\in T} f(t) \le \max_{t\in T} g(t), \qquad \min_{t\in T} f(t) \le \min_{t\in T} g(t).$$

It even persists if we further take the minimum or maximum over the same family $\mathbf{T}$ of subsets T, e.g., then also

$$\max_{T\in\mathbf{T}} \min_{t\in T} f(t) \le \max_{T\in\mathbf{T}} \min_{t\in T} g(t).$$

Consequently,

(15.6) Corollary: If A, B are hermitian, and $R_A(x) \le R_B(x) + c$ for some constant c and all x, then

$$\mu_j(A) \le \mu_j(B) + c, \quad \forall j.$$

This gives

(15.7) Weyl's inequalities: If $A = B + C$, with A, B, C hermitian, then

$$\mu_j(B) + \mu_1(C) \le \mu_j(A) \le \mu_j(B) + \mu_n(C), \quad \forall j.$$

Proof: Since $\mu_1(C) \le R_C(x) \le \mu_n(C)$ (by Rayleigh's principle), while $R_B(x) + R_C(x) = R_A(x)$, the preceding corollary provides the proof. □

A typical *application of Weyl's Inequalities* is the observation that, for $A = BB^c + C \in \mathbb{F}^{n\times n}$ with $B \in \mathbb{F}^{n\times k}$ and A hermitian (hence also C hermitian), $\mu_1(C) \le \mu_j(A) \le \mu_n(C)$ for all $j < (n-k)$, since $\operatorname{rank} BB^c \le \operatorname{rank} B \le k$, hence $\mu_j(BB^c)$ must be zero for $j < (n-k)$.

Since $C = A - B$, Weyl's inequalities imply that

$$|\mu_j(A) - \mu_j(B)| \le \max\{|\mu_1(A-B)|, |\mu_n(A-B)|\} = \rho(A-B).$$

Therefore, with the substitutions $A \leftarrow A + E$, $B \leftarrow A$, we obtain

(15.8) max-norm Wielandt-Hoffman: If A and E are both hermitian, then

$$\max_j |\mu_j(A+E) - \mu_j(A)| \le \max_j |\mu_j(E)|.$$

A corresponding statement involving 2-norms is valid but much harder to prove.

Finally, a totally different application of the MMM Theorem is

(15.9) Sylvester's Law of Inertia: Any two *congruent* hermitian matrices have the same number of positive, zero, and negative eigenvalues.

Proof: It is sufficient to prove that if $B = V^c A V$ for some hermitian A and some invertible V, then $\mu_j(A) > 0$ implies $\mu_j(B) > 0$. For this, we observe that, by the MMM Theorem, $\mu_j(A) > 0$ implies that R_A is positive somewhere on every j-dimensional subspace, while (also by the MMM Theorem), for some j-dimensional subspace H,

$$\mu_j(B) \;=\; \max_{x\in H} R_B(x) = \max_{x\in H} R_A(Vx) R_{V^cV}(x),$$

and this is necessarily positive, since $\dim VH = j$ and $R_{V^cV}(x) = \|Vx\|^2/\|x\|^2$ is positive for any $x \ne 0$. □

It follows that we don't have to diagonalize the real symmetric matrix A (as we did in the proof of (15.2)Proposition) in order to find out whether or not A or the corresponding quadratic form q_A is definite. Assuming that A is invertible, hence has no zero eigenvalue, it is sufficient to use Gauss elimination without pivoting to obtain the factorization $A = LDL^c$, with L unit lower triangular. By Sylvester's Law of Inertia, the number of positive (negative) eigenvalues of A equals the number of positive (negative) diagonal entries of D.

This fact can be used to locate the eigenvalues of a real symmetric matrix by *bisection*. For, the number of positive (negative) diagonal entries in the diagonal matrix D_μ obtained in the factorization $L_\mu D_\mu L_\mu{}^c$ for $(A - \mu\,\mathrm{id})$ tells us the number of eigenvalues of A to the right (left) of μ, hence makes it easy to locate and refine intervals that contain just one eigenvalue of A.

