16. Quadratic Forms and Definite Matrices

Quadratic forms play a key role in optimization theory. They are the simplest functions where optimization (maximization or minimization) is an interesting problem. Much of the theory of quadratic optimization carries over to general $\mathcal{C}^2$ functions via Taylor's Formula. The key point is that we can approximate any $\mathcal{C}^2$ function near any point by a combination of a constant term, linear term, and quadratic form.

16.1 Simple Optimization Problems

The simplest type of optimization problem is to maximize or minimize the constant function $f(\mathbf{x}) = a$ for any $\mathbf{x}$ in some domain D. All points of D both maximize and minimize the function f.

The next simplest type of function is linear, $f(\mathbf{x}) = \mathbf{p}\cdot\mathbf{x}$ for some $\mathbf{p} \neq \mathbf{0}$. Suppose that we try to maximize this over some inner product space. Since $\mathbf{p} \neq \mathbf{0}$, $f(n\mathbf{p}) = n\|\mathbf{p}\|^2$ and $f(-n\mathbf{p}) = -n\|\mathbf{p}\|^2$. As $n \to \infty$, the former gets ever larger, and the latter every smaller. There is neither a maximum value nor a minimum value. The same is true of affine functions such as $f(\mathbf{x}) = a + \mathbf{p}\cdot\mathbf{x}$ for $\mathbf{p} \neq \mathbf{0}$.

That brings us to quadratic forms. If f is defined on $\mathbb{R}$, quadratic forms are written $f(x) = ax^2$. If $a > 0$, there is no maximum as $f(x) = ax^2 \to +\infty$ as $x \to +\infty$. There is a minimum as $f(x) \geq 0$ and $ax^2 = 0$ if and only if $x = 0$. Such a form is called *positive definite*. When $a < 0$, the situation is reversed. There is no minimum, and $f(x) \leq 0$ with $f(x) = 0$ if and only if $x = 0$, what we call a *negative definite* form. There is a unique maximum at $x = 0$.

Finally, if $f(x) = ax^2 + bx + c$ is a function on $\mathbb{R}$, we can complete the square to rewrite f as $a(x + b/2a)^2 + (c - b^2/4a^2)$. This is like $f(x) = a^2$, except that the minimum when $a > 0$ has moved to $x = -b/2a$, while if $a < 0$ the maximum is now at $x = -b/2a$. The situation when f is defined on $\mathbb{R}^m$ is a bit more complex.

16.2 Quadratic Forms

Recall that a bilinear form from $\mathbb{R}^{2m} \to \mathbb{R}$ can be written $f(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{A} \mathbf{y}$ where $\mathbf{A}$ is an $m \times m$ matrix. We can use this to define a quadratic form,

$$Q_{\mathbf{A}}(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{ij=1}^{m} a_{ij} x_i x_j = \sum_{i=1}^{m} a_{ii} x_i^2 + \sum_{i<j} (a_{ij} + a_{ji}) x_i x_j. \tag{16.2.1}$$

Quadratic Form. A *quadratic form* on $\mathbb{R}^m$ is a real-valued function of the form

$$Q(\mathbf{x}) = \sum_{i \le j} a_{ij} x_i x_j. \tag{16.2.2}$$

If a quadratic form is generated by a matrix that is not symmetric, there is another matrix that is symmetric that generates the same form. To see this, recall that $(\mathbf{x}^T \mathbf{B} \mathbf{x})^T = \mathbf{x}^T \mathbf{B} \mathbf{x}$ because we are taking the transpose of a real numbers. The first term can also be written $\mathbf{x}^T \mathbf{B}^T \mathbf{x}$, so $\mathbf{B}$ and $\mathbf{B}^T$ generate the same quadratic form. But then $\mathbf{A} = \frac{1}{2}(\mathbf{B} + \mathbf{B}^T)$ also generates the same form and is symmetric.

The key thing about quadratic forms is that each term of $Q(\mathbf{x})$ is of degree two in the x_i's. To match equation (16.2.2) with the matrix version, suppose $\mathbf{A}$ is symmetric with

$$\mathbf{A} = \begin{pmatrix} a_{11} & \frac{1}{2} a_{12} & \cdots & \frac{1}{2} a_{1m} \\ \frac{1}{2} a_{21} & a_{22} & \cdots & \frac{1}{2} a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2} a_{m1} & \frac{1}{2} a_{m2} & \cdots & a_{mm}. \end{pmatrix}$$

Note how factors of $1/2$ are placed everywhere except the diagonal in order to match equation (16.2.2).

Then

$$Q_{\mathbf{A}}(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i \le j} a_{ij} x_i x_j.$$

It's easy to see that $DQ_{\mathbf{A}}(\mathbf{x}) = \mathbf{x}^T(\mathbf{A} + \mathbf{A}^T) = 2\mathbf{x}^T \mathbf{A}$ since $\mathbf{A}$ is symmetric. Furthermore, $D^2 Q_{\mathbf{A}} = 2\mathbf{A}$. When it appears in Taylor's formula, $D^2 Q_{\mathbf{A}}$ is divided by 2, so we see $\mathbf{A}$ there. For example, consider the following Taylor expansion of $Q_{\mathbf{A}}(\mathbf{x})$ around $\mathbf{a}$:

$$Q_{\mathbf{A}}(\mathbf{x}) = Q_{\mathbf{A}}(\mathbf{a}) + 2\mathbf{a}^T \mathbf{A}(\mathbf{x} - \mathbf{a}) + (\mathbf{x} - \mathbf{a})^T \mathbf{A} (\mathbf{x} - \mathbf{a}).$$

We see a constant, a linear term, and a quadratic term. In fact, any $\mathcal{C}^2$ function can be approximated by this type of equation using Taylor's formula.

16.3 Definite and Non-definite Forms

Of course, $Q(\mathbf{0}) = 0$ for any quadratic form. We call a form *definite* if the zero vector is the only vector where the form is zero. If the form is defined on $\mathbb{R}$, there aren't many possibilities. If $a > 0$, then $Q(x) = ax^2 \geq 0$ with $Q(x) = 0$ if and only if $x = 0$. We call such a form *positive definite* and $x = 0$ minimizes the form over $\mathbb{R}$. If $a < 0$, then $Q(x) = ax^2 \leq 0$ with $Q(x) = 0$ if and only if $x = 0$. We call such a form *negative definite* and $x = 0$ maximizes the form over $\mathbb{R}$.

► Example 16.3.1: Definite Forms in $\mathbb{R}^2$. When a quadratic form is defined on $\mathbb{R}^2$, there are more possibilities. There are still the two basic cases. A form can be *positive definite*, as is $Q(\mathbf{x}) = \|\mathbf{x}\|^2$, or it can be *negative definite*, as with $Q(\mathbf{x}) = -\|\mathbf{x}\|^2$. ◄

However, there are still more possibilities.

► Example 16.3.2: Semidefinite Forms in $\mathbb{R}^2$. Consider the form $Q(\mathbf{x}) = x_1^2 + 2x_1x_2 + x_2^2 = (x_1 + x_2)^2$. This form is always non-negative, but it is not definite because $Q(1, -1) = 0$. Such a form, that is always non-negative, but not definite, is called *positive semidefinite*.

If we flip the sign, we get the form

$$Q(\mathbf{x}) = -x_1^2 - 2x_1x_2 - x_2^2 = -(x_1 + x_2)^2.$$

This form is never positive, but not definite. We refer to it as *negative semidefinite*. ◄

Some forms don't fit into any of the above categories.

► Example 16.3.3: Indefinite Forms. Define

$$Q(\mathbf{x}) = \mathbf{x}^T \begin{pmatrix} -1 & 0 \\ 0 & +1 \end{pmatrix} \mathbf{x} = -x_1^2 + x_2^2.$$

Then $Q(2, 1) = -3$ while $Q(1, 2) = +3$. This form is sometimes positive and sometimes negative. A quadratic form that sometimes takes negative values and sometimes takes positive values is called *indefinite*. ◄

16.4 Definite, Semidefinite, and Indefinite Matrices

We classify the associated symmetric matrices in a similar fashion. Let's consider the general case where the form $Q(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ is defined on $\mathbb{R}^m$.

Definite Matrices. Let $\mathbf{A} \neq \mathbf{0}$ be an $m \times m$ symmetric matrix. Then $\mathbf{A}$ is:

(a) *positive definite* if $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$.
(b) *negative definite* if $\mathbf{x}^T\mathbf{A}\mathbf{x} < 0$ for all $\mathbf{x} \neq \mathbf{0}$.
(c) *positive semidefinite* if $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^m$.
(d) *negative semidefinite* if $\mathbf{x}^T\mathbf{A}\mathbf{x} \leq 0$ for all $\mathbf{x} \in \mathbb{R}^m$.
(e) *indefinite* if there is $\mathbf{x}_0 \in \mathbb{R}^m$ with $\mathbf{x}_0^T\mathbf{A}\mathbf{x}_0 < 0$. and $\mathbf{x}_1 \in \mathbb{R}^m$ with $\mathbf{x}_0^T\mathbf{A}\mathbf{x}_0 > 0$.

Positive (negative) definite matrices are automatically positive (negative) semidefinite. Other than that, the categories do not overlap.

To clarify the status of indefinite forms, which only exist if $m \geq 2$, we show that no indefinite form is definite.

Theorem 16.4.1. *Indefinite quadratic forms are not definite. That is, there is a* $\mathbf{x}_0 \neq \mathbf{0}$ *with* $Q(\mathbf{x}_0) = 0$.

Proof. To see this, suppose $Q(\mathbf{x})$ is indefinite. Then there are $\mathbf{x}_1$ and $\mathbf{x}_2$ with $Q(\mathbf{x}_1) > 0$ and $Q(\mathbf{x}_2) < 0$. Define $f(t) = Q\big((1-t)\mathbf{x}_1 + t\mathbf{x}_2\big)$. Now f is continuous, $f(0) = Q(\mathbf{x}_1) > 0$, and $f(1) = Q(\mathbf{x}_2) < 0$. By the Intermediate Value Theorem, there is a $t^* \in (0,1)$ with $f(t^*) = 0$.

Set $\mathbf{x}_0 = (1-t^*)\mathbf{x}_1 + t^*\mathbf{x}_2$. **Suppose this is zero**. Then solve for $\mathbf{x}_2 = [(t^*-1)/t^*]\mathbf{x}_1$. But then

$$0 > Q(\mathbf{x}_2) = \left(\frac{t^*-1}{t^*}\right)^2 Q(\mathbf{x}_1) > 0.$$

This is impossible. It contradicts the fact that the form has opposite signs on $\mathbf{x}_1$ and $\mathbf{x}_2$.

Therefore $\mathbf{x}_0 \neq \mathbf{0}$, showing that Q is not definite. ∎

Our classification of matrices will be useful for stating the second order optimization conditions.

16.5 Properties of Definite and Semidefinite Matrices

Positive definiteness is not to be confused the associated matrix having positive entries. It's perfectly possible to have a matrix with positive entries that defines a quadratic form that is neither positive definite nor positive semidefinite. This is true even for symmetric matrices. Consider the matrix

$$\begin{pmatrix} 1 & 4 \\ 4 & 1 \end{pmatrix}.$$

The associated form is $x_1^2 + 8x_1x_2 + x_2^2 = (x_1 + x_2)^2 + 6x_1x_2$. Then $Q(1, -1) = -6$ while $Q(1, 1) = 10$, showing that Q is indefinite in spite of all the positive elements.

What we can show is that any positive (negative) semidefinite matrix has non-negative (non-positive) diagonal elements. Any positive (negative) definite matrix has strictly positive (negative) diagonal elements. This follows immediately from considering $Q(\boldsymbol{e}_i) = \boldsymbol{e}_i^T \mathbf{A} \boldsymbol{e}_i = a_{ii}$.

One other useful property is that both definite matrices, both positive and negative, are invertible.

Theorem 16.5.1. *Let $\mathbf{A}$ be a symmetric $m \times m$ matrix. If $\mathbf{A}$ is positive (or negative) definite, it is non-singular.*

Proof. We will prove the negative definite case. The positive definite case is similar. Let $\mathbf{A}$ be negative definite and **suppose it is singular**. Then there is $\boldsymbol{x} \neq \mathbf{0}$ with $\mathbf{A}\boldsymbol{x} = \mathbf{0}$. It follows that $Q(\boldsymbol{x}) = \boldsymbol{x}^T \mathbf{A} \boldsymbol{x} = 0$, showing that Q is not definite. This **contradiction shows** that **$\mathbf{A}$ must be non-singular**. ∎

Semidefinite matrices need not be invertible. The same is true for indefinite matrices when $m > 2$.

16.6 Principal Submatrices and Minors

We need criteria that will determine how to classify a given symmetric matrix. Is it positive definite? Negative semidefinite? Indefinite?

We start by considering principal submatrices and their associated minors.

Given an $m \times m$ matrix $\mathbf{A}$, a k^{th} *order principal minor*, $\mathbf{A}_k$ is a $k \times k$ submatrix formed by removing the same $m - k$ rows and columns from $\mathbf{A}$. If rows $i_1, \ldots, i_{m-k}$ are removed, columns $i_1, \ldots, i_{m-k}$ are also removed. The determinant of a k^{th} order principal submatrix is called a k^{th} order *principal minor*.

Given a 4×4 matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix},$$

there is 1 fourth submatrix, the entire matrix $\mathbf{A}$. There are 4 third order submatrices.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{12} & a_{14} \\ a_{21} & a_{22} & a_{24} \\ a_{41} & a_{42} & a_{44} \end{pmatrix}$$
$$\begin{pmatrix} a_{11} & a_{13} & a_{14} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{pmatrix} \quad \begin{pmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{pmatrix}.$$

The 6 second order submatrices are

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{14} \\ a_{41} & a_{44} \end{pmatrix}$$
$$\begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} \quad \begin{pmatrix} a_{22} & a_{24} \\ a_{42} & a_{44} \end{pmatrix} \quad \begin{pmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{pmatrix}.$$

and the 4 first order submatrices are

$$(a_{11}) \quad (a_{22}) \quad (a_{33}) \quad (a_{44}).$$

The determinants of these submatrices are the corresponding principal minors.

16.7 Minors and Positive Definite Matrices

In general, there are

$$\binom{m}{k} = \frac{m!}{(m-k)!k!} = \frac{m(m-1)\cdots(m-k+1)}{k(k-1)\cdots 2}$$

k^{th} order principal submatrices/minors of an $m \times m$ matrix and a total of $2^m - 1$ principal submatrices and minors.

The *leading principal submatrix* of order k is a submatrix where the the last $(m-k)$ rows and columns have been removed. We will use $\mathbf{A}_k$ to denote the k^{th} leading principal submatrix of $\mathbf{A}$. The k^{th} order leading principal minor is then $\det \mathbf{A}_k$. There are only m leading principal submatrices/minors of an $m \times m$ matrix.

We can check whether a matrix is positive or negative definite by looking at the signs of its leading principal minors.

Theorem (Definite Matrices). *Let* $\mathbf{A}$ *be an* $m \times m$ *symmetric matrix. Then*

(a) *The matrix* $\mathbf{A}$ *is positive definite if and only if all of its leading principal minors are positive.*

$$\det \mathbf{A}_k > 0 \quad \textit{for all } k = 1, \ldots, m$$

(b) *The matrix* $\mathbf{A}$ *is negative definite if and only if its leading principal minors alternate in sign with* $\det \mathbf{A}_1 < 0$. *That is,*

$$(-1)^k \det \mathbf{A}_k > 0 \quad \textit{for all } k = 1, \ldots, m$$

(c) *If non-zero leading principal minors violate both (a) and (b), then* $\mathbf{A}$ *is indefinite.*

Part (b) follows from part (a) by applying part (a) to the matrix $-\mathbf{A}$.

To test whether a matrix is semidefinite, we must check every principal minor, not just the leading ones.

Theorem (Semidefinite Matrices). *Let* $\mathbf{A}$ *be an* $m \times m$ *symmetric matrix. Then*

(a) *The matrix* $\mathbf{A}$ *is positive semidefinite if and only if every principal minor is non-negative.*

(b) *The matrix* $\mathbf{A}$ *is negative semidefinite if and only if every odd order principal minor is non-positive and every even order principal minor is non-negative.*

Here too, part (b) follows from part (a) by applying part (a) to $-\mathbf{A}$.

16.8 Definite Diagonal Matrices

10/22/20

Suppose $\mathbf{A}$ is a diagonal matrix. The associated quadratic form is

$$Q(\mathbf{x}) = \sum_{i=1}^{m} a_{ii} x_i^2.$$

It is easy to see that $Q(\mathbf{x})$ is positive definite if and only if all the a_{ii} are positive, and that it is negative definite if and only if all of the a_{ii} are negative. This yields the sign pattern mentioned in the Definite Matrices Theorem.

In the positive definite case

$$\det \mathbf{A}_1 = a_{11} > 0, \quad \det \mathbf{A}_2 = a_{11}a_{22} > 0, \quad \det \mathbf{A}_3 = a_{11}a_{22}a_{33} > 0, \quad \text{etc.}$$

which is equivalent to $a_{11} > 0, a_{22} > 0, \ldots, a_{mm} > 0$.

In the negative definite case

$$\det \mathbf{A}_1 = a_{11} < 0, \quad \det \mathbf{A}_2 = a_{11}a_{22} > 0, \quad \det \mathbf{A}_3 = a_{11}a_{22}a_{33} < 0, \quad \text{etc.}$$

which is equivalent to $a_{11} < 0, a_{22} < 0, \ldots, a_{mm} < 0$.

Similarly, $Q(\mathbf{x})$ is positive semidefinite if and only if all of the $a_{ii} \geq 0$, and that it is negative semidefinite if and only if all of the $a_{ii} \leq 0$. Finally, $Q(\mathbf{x})$ is indefinite if and only there is at least one i with $a_{ii} > 0$ and at least one j with $a_{jj} < 0$.

Later, we will see that for any symmetric matrix, there is a basis where the matrix has a diagonal form and these results apply.

This is a good time to see why we need to look at all of the principal minors to show a form is semidefinite. Consider the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

The leading principal minors are $a_{11} = 1$, $a_{11}a_{22} = 0$, and $a_{11}a_{22}a_{33} = 0$. If we only checked whether the leading principal minors were non-zero, we would conclude that this matrix is positive semidefinite. Yet this is an indefinite matrix. Checking all principal minors includes looking at $a_{33} = -1$, showing the matrix is not positive semidefinite. With non-diagonal matrices the second order minors also come into play.

16.9 Definite Matrices: The Two By Two Case

Let
$$\mathbf{A} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \quad \text{so} \quad Q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2.$$

Then if $a = 0$, $Q(1,0) = 0$, so Q is not definite.

If $a \neq 0$, we can complete the square, obtaining

$$\begin{aligned} Q(x_1, x_2) &= ax_1^2 + 2bx_1x_2 + cx_2^2 \\ &= a\left(x_1^2 + \frac{2b}{a}x_1x_2 + \frac{b^2}{a^2}x_2^2\right) - \frac{b^2}{a}x_2^2 + cx_2^2 \\ &= a\left(x_1 + \frac{b}{a}x_2\right)^2 + \frac{ac - b^2}{a}x_2^2. \end{aligned}$$

Now $Q(1,0) = a$ and $Q(-b, a) = a(ac - b^2) = a \det \mathbf{A}_2$.

If Q is positive definite, $|\mathbf{A}_1| = a > 0$ and $|\mathbf{A}_2| = (ac - b^2) > 0$. Conversely, if those two numbers are positive, Q is non-negative. Moreover $Q(\mathbf{x}) = 0$ implies $x_2 = 0$ and $x_1 - bx_2/a = 0$, so $x_1 = x_2 = 0$, meaning that Q is positive definite.

If Q is negative definite, $|\mathbf{A}_1| = a < 0$ and $|\mathbf{A}_2| = (ac - b^2) > 0$. Conversely, if $|\mathbf{A}_1| = a < 0$ and $|\mathbf{A}_2| = (ac - b^2) > 0$, Q is non-negative. Moreover $Q(\mathbf{x}) = 0$ implies $x_2 = 0$ and $x_1 - bx_2/a = 0$, so $x_1 = x_2 = 0$.

Finally, if either $a > 0$ and $(ac - b^2) < 0$ or $a < 0$ and $(ac - b^2) < 0$ (the only possible non-zero violations of the patterns), then $Q(1,0)$ and $Q(-b/a, 1)$ have opposite signs and Q is indefinite.

This proves the Definite Matrices Theorem for matrices of size 2. The general case can be proved by induction. Section 16.4 of Simon and Blume shows how.

16.10 Norms for Matrices

It would be helpful to have a norm on the set of matrices, allowing us to think of the $m \times m$ matrices as a normed vector space. There's one special property that would be useful, we would like the norm to obey

$$\|\mathbf{A} \times \mathbf{B}\| \leq \|\mathbf{A}\| \, \|\mathbf{B}\|.$$

We call such a norm a *compatible matrix norm*. Fortunately, there is a surfeit of such norms. One way to get such a norm is to consider a norm on $\mathbb{R}^m$, and then define

$$\|\mathbf{A}\| = \sup_{x \neq 0} \frac{\|\mathbf{A}x\|}{\|x\|} = \sup_{\|x\|=1} \|\mathbf{A}x\|.$$

With such a norm, it follows that

$$\|\mathbf{A}x\| \leq \|\mathbf{A}\| \, \|x\|.$$

Consider

$$\begin{aligned} \|\mathbf{AB}x\| &\leq \|\mathbf{A}\| \, \|\mathbf{B}x\| \\ &\leq \|\mathbf{A}\| \, \|\mathbf{B}\| \, \|x\|. \end{aligned}$$

Take x to be a unit vector and take the supremum over all unit vectors x. We then obtain

$$\|\mathbf{A} \times \mathbf{B}\| \leq \|\mathbf{A}\| \, \|\mathbf{B}\|.$$

One inequality that applies to compatible norms is the the norm of $\mathbf{A}^n$ is no greater than the n^{th} power of $\|\mathbf{A}\|$.

$$\|\mathbf{A}^n\| \leq \|\mathbf{A}\|^n \quad \text{for } n = 0, 1, \ldots.$$

16.11 The ℓ_∞ Matrix Norm

So what does this norm look like in practice? Let's consider one case, where we start with the ℓ_∞ norm on $\mathbb{R}^m$. We'll call the resulting matrix norm, $\|\mathbf{A}\|_\infty$. This does not mean we are taking the max of all the entries of $\mathbf{A}$. Rather it is the matrix norm we generated from the $\|\cdot\|_\infty$ norm on $\mathbb{R}^m$.

$$\begin{aligned}\left|(\mathbf{A}\mathbf{x})_i\right| &= \left|\sum_{j=1}^m a_{ij}x_j\right| = \sum_{j=1}^m |a_{ij}|\,|x_j| \\ &\le \sum_{j=1}^m |a_{ij}|\,\|\mathbf{x}\|_\infty \\ &= \left(\sum_{j=1}^m |a_{ij}|\right)\|\mathbf{x}\|_\infty.\end{aligned}$$

Using the definition of $\|\mathbf{A}\|_\infty$ implies

$$\|\mathbf{A}\|_\infty \le \max_{i=1,\dots,m}\sum_{j=1}^m |a_{ij}|.$$

To see that the maximum can be attained by a unit vector $\mathbf{x}$, focus on a row i that attains the maximum. Set $x_j = +1$ if $a_{ij} > 0$ and $x_j = -1$ if $a_{ij} < 0$. Then $\|\mathbf{A}\mathbf{x}\|_\infty = \max_i \sum_j |a_{ij}|$. We conclude that

$$\|\mathbf{A}\|_\infty = \max_{i=1,\dots,m}\sum_{j=1}^m |a_{ij}|$$

is a norm on the set of $m \times m$ matrices that is compatible with matrix multiplication.

Moreover, the topology it generates is the same as the topology where convergence of a sequence $\mathbf{A}_n \to \mathbf{A}$ means that each entry converges, that for every pair (i,j), $a^n_{ij} \to a_{ij}$.

16.12 Determinants are Continuous

Now that we have a topology on the set of $m \times m$ matrices, we can consider continuity. What about the determinant, is it continuous?

Theorem 16.12.1. *The mapping* $\mathbf{A} \to \det \mathbf{A}$ *is continuous on the set of* $m \times m$ *matrices with the norm* $\|\mathbf{A}\|_\infty$.

Proof. Using the formula

$$\det \mathbf{A} = \sum_{\sigma \in P_n} (\operatorname{sgn} \sigma) \left(\prod_{i=1}^{n} a_{i\sigma(i)} \right).$$

we see that the determinant of $\mathbf{A}$ is a sum of products of (± 1) and entries of the matrix $\mathbf{A}$. Addition and multiplication are continuous operations, and matrix limits are defined entrywise, so the determinant is continuous. ∎

Corollary 16.12.2. *The set of invertible* $m \times m$ *matrices is an open set.*

Proof. Recall that a matrix is invertible if and only if its determinant is non-zero. That means that the set of invertible matrices is $\det^{-1}\{x : x \neq 0\}$. As the continuous inverse image of an open set, the set of invertible matrices is open. ∎

16.13 The Sets of Definite and Semidefinite Matrices

We can use the criteria in the Definite Matrices Theorem to show:

Corollary 16.13.1. *The sets of positive definite, negative definite, and indefinite* $m \times m$ *matrices are all open sets.*

Proof. Consider the negative definite $m \times m$ matrices. They are defined by the inequalities

$$(-1)^k \det \mathbf{A}_k > 0 \quad \text{for all } k = 1, \ldots, m.$$

where $\mathbf{A}_k$ is the k^{th} leading principal submatrix. Define $f_k(\mathbf{A}) = (-1)^k \det \mathbf{A}_k$. The f_k are all continuous functions since the determinant is continuous. The set of negative definite matrices is then

$$\bigcap_{k=1}^{m} f_k^{-1}(0, +\infty).$$

As the intersection of a finite collection of open sets, themselves the inverse images of open sets, the set of negative definite matrices is open. The proof for the positive definite is similar, with $f_k(\mathbf{A}) = \det \mathbf{A}_k$.

The indefinite case works a little differently. Suppose $\mathbf{A}$ is an indefinite matrix. Then there are i and j with $(-1)^i \det \mathbf{A}_i < 0$ (breaking the negative definite sign pattern) and $\det \mathbf{A}_j < 0$ (breaking the positive definite sign pattern). The numbers i and j may be the same if i is even (the matrix in section 16.5 is an example). Again, the set of matrices with $(-1)^i \det \mathbf{A}_i < 0$ and $\det \mathbf{A}_j > 0$ is open. Denote this open set by U_{ij}. Then the set of indefinite matrices is the union

$$\bigcup_{i,j=1}^{m} U_{ij}$$

where i and j are not necessarily distinct. As the union of open sets, the set of indefinite matrices is open. ∎

Similarly, the criteria in the Semidefinite Matrices Theorem can be used to show:

Corollary 16.13.2. *The set of positive (negative) semidefinite* $m \times m$ *matrices is an closed set.*

16.14 Infinite Series of Matrcies

We can consider matrix series of the form

$$\mathbf{S} = \sum_{n=0}^{\infty} a_n \mathbf{A}^n.$$

Define the k^{th} *partial sum* $\mathbf{S}_k$ by

$$\mathbf{S}_k = \sum_{n=0}^{k} a_n \mathbf{A}^n.$$

As with infinite series of real numbers, the infinite sum $\mathbf{S}$ is defined as $\lim_{k\to\infty} \mathbf{S}_k$, provided the limit exists.

Theorem 16.14.1. *Suppose the infinite series of real numbers $\sum_{n=0}^{\infty} |a_n| K^n$ converges and $\|\mathbf{A}\|_\infty \le K$ for an $m \times m$ matrix $\mathbf{A}$. Then $\sum_{n=0}^{\infty} a_n \mathbf{A}^n$ converges.*

Proof. First, we show that $\mathbf{S}_k$ forms a Cauchy sequence. Let $\ell > k$. Then

$$\begin{aligned}
\|\mathbf{S}_\ell - \mathbf{S}_k\|_\infty &= \left\| \sum_{n=k+1}^{\ell} a_n \mathbf{A}^n \right\|_\infty \\
&\le \sum_{n=k+1}^{\ell} |a_n| \|\mathbf{A}^n\| \le \sum_{n=k+1}^{\ell} |a_n| \|\mathbf{A}\|^n \\
&\le \sum_{n=k+1}^{\ell} |a_n| K^n.
\end{aligned}$$

Since the last series is a convergent sequence of real numbers, it is a Cauchy sequence. That implies that $\{\mathbf{S}_k\}$ is also a Cauchy sequence. We know convergence in the ∞-norm topology on the $m \times m$ matrices is equivalent to entrywise convergence of the matrices, so the series $\sum_{n=0}^{\infty} a_n \mathbf{A}^n$ has a limit.

Taking the limit as $\ell \to \infty$, we find

$$\|\mathbf{S}_k - \mathbf{S}\|_\infty \le \sum_{n=k+1}^{\infty} |a_n| K^n.$$

The series on the right converges, so $\mathbf{S}_k \to \mathbf{S}$ in the ∞-norm. ∎

16.15 Norms and Inverses

We can also show that any matrix that is close enough to the identity matrix must be invertible. More precisely, if the distance to the identity matrix is less than one in any compatible matrix norm, the matrix is invertible.

Theorem 16.15.1. *Suppose* $\|\mathbf{I}-\mathbf{A}\|_\infty < 1$. *Then* $\mathbf{A}$ *is invertible.*

Proof. Define a matrix $\mathbf{S}$ by

$$\mathbf{S} = \sum_{n=0}^{\infty} (\mathbf{I}-\mathbf{A})^n.$$

Because $\|\mathbf{I}-\mathbf{A}\| < 1$, this sum converges by Theorem 16.14.1.

Now

$$(\mathbf{I}-\mathbf{A})\mathbf{S} = \sum_{n=1}^{\infty} (\mathbf{I}-\mathbf{A})^n = \mathbf{S} - (\mathbf{I}-\mathbf{A})^0 = \mathbf{S} - \mathbf{I}.$$

It follows that $\mathbf{AS} = \mathbf{I}$, showing that $\mathbf{S} = \mathbf{A}^{-1}$. ∎

One consequence of this is that the ball $B_1(\mathbf{I}_m) \subset \{m \times m \text{ matrices}\}$ is a set consisting solely of invertible matrices.

16.16 Matrix Power Series

One application of the matrix norm is to show that we can define certain functions of matrices using power series.

Theorem 16.16.1. *Suppose the series* $\sum_{n=0}^{\infty} a_n x^n$ *converges absolutely and uniformly for* $|x| \le K$. *Then*

$$\sum_{n=0}^{\infty} a_n \mathbf{A}^n$$

converges absolutely for $\|\mathbf{A}\|_\infty \le K$.

Proof. This follows immediately from Theorem 16.14.1. ∎

This lets us define matrix functions such as

$$\exp(t\mathbf{A}) = \sum_{n=0}^{\infty} \frac{t^n \mathbf{A}^n}{n!}. \tag{16.16.3}$$

The corresponding real-valued power series

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = \exp x$$

converges absolutely and uniformly on any compact interval $[-K, K]$. It follows that equation (16.16.3) converges absolutely and uniformly for $|t|\|\mathbf{A}\|_\infty \le K$ for any $K > 0$. Since K can be any positive number, that means that $\exp(t\mathbf{A})$ is a continuous function of $t \in \mathbb{R}$. In fact, for fixed $\mathbf{A}$, you can even show that $\exp(t\mathbf{A})$ is $\mathcal{C}^\infty$ for all $t \in \mathbb{R}$.

Because $\mathbf{A}$ commutes with itself, it is straightforward to show that $\exp\big((s+t)\mathbf{A}\big) = (\exp s\mathbf{A})(\exp t\mathbf{A})$ in the same way we do for real numbers. If $\mathbf{A}$ and $\mathbf{B}$ commute, you can also show $\exp(\mathbf{A} + \mathbf{B}) = (\exp \mathbf{A})(\exp \mathbf{B})$.

16.17 Constrained Optimization

When we're trying to maximize or minimize a quadratic form $Q(\mathbf{x})$ our classification of matrices (and forms) tells us all we need to know. A positive definite matrix has a unique minimum at zero (and no maximum). A negative definite matrix has a unique maximum at zero (and no minimum). Indefinite matrices have neither maxima nor minima. Positive (negative) semidefinite matrices have a minimum (maximum) at zero, but unless the matrix is definite, the optimum is not unique. Simple enough.

But a lot of economics is about optimization under constraint. What if we place a linear constraint (e.g., a budget constraint) on our optimization problem. What can we say then?

16.18 Constrained Optimization in $\mathbb{R}^2$, part I

For example, $Q(\mathbf{x}) = -2x_1^2 + x_2^2 + x_1x_2$ is an indefinite form. Can we find a maximum if we also require $x_2 = 0$? In that case, the form reduces to $-2x_1^2$, which has a maximum at $\mathbf{x} = \mathbf{0}$. Under the constraint $x_1 = 2x_2$, the form reduces to $-5x_2^2$, which has a maximum at $\mathbf{x} = \mathbf{0}$. Under some constraints there is a minimum. The constraint $x_1 = 0$ yields the reduced form x_2^2, which has a minimum at $\mathbf{x} = \mathbf{0}$.

We will consider the problem of maximizing

$$Q(\mathbf{x}) = \mathbf{x}^T \begin{pmatrix} a & b \\ b & c \end{pmatrix} \mathbf{x} = ax_1^2 + 2bx_1x_2 + cx_2^2$$

under a single linear constraint

$$Ax_1 + Bx_2 = 0.$$

We do not consider the case of two linear constraints as it would add nothing. With two constraints in $\mathbb{R}^2$, there are three things that can happen. The second constraint could be redundant (changing nothing), or the intersection of the constraints might be a single point (which would be both maximum and minimum), or the intersection could be empty, leaving the problem without a solution.

The following theorem describes the solution.

Theorem 16.18.1. *The quadratic form* $Q(\mathbf{x}) = ax_1^2 + 2bx_1x_2 + cx_2^2$ *is positive (negative) definite on the constraint set,* $L = \{\mathbf{x} : Ax_1 + Bx_2 = 0\}$ *if and only if the determinant*

$$-\begin{vmatrix} 0 & A & B \\ A & a & b \\ B & b & c \end{vmatrix}$$

is positive (negative).

16.19 Constrained Optimization in $\mathbb{R}^2$, part II

Theorem 16.19.1. *The quadratic form* $Q(\mathbf{x}) = ax_1^2 + 2bx_1x_2 + cx_2^2$ *is positive (negative) definite on the constraint set,* $L = \{\mathbf{x} : Ax_1 + Bx_2 = 0\}$ *if and only if the determinant*

$$-\begin{vmatrix} 0 & A & B \\ A & a & b \\ B & b & c \end{vmatrix}$$

is positive (negative).

Proof. At least one of A and B must be non-zero to even have a constraint. If $A = 0$, $B \neq 0$, which implies $x_2 = 0$ and $Q(\mathbf{x}) = ax_1^2$. In that case, there is a unique maximum at $\mathbf{x} = \mathbf{0}$ if $a < 0$, and a unique minimum there if $a > 0$. Finally, if $a = 0$, $Q(\mathbf{x}) = 0$ whenever the constraint is satisfied. All points on the line $\{(x_1, 0)\}$ are both maximizers and minimizers. If $A = 0$, the determinant becomes $-aB^2$, so its sign is the sign of $-a$.

Now suppose $A \neq 0$, so $x_1 = -(B/A)x_2$. Then

$$\begin{aligned} Q(\mathbf{x}) &= ax_1^2 + 2bx_1x_2 + cx_2^2 \\ &= a\left(\frac{B^2}{A^2}\right)^2 x_2^2 - \frac{2bB}{A}x_2^2 + cx_2^2 \\ &= \frac{aB^2}{A^2}x_2^2 - \frac{2bAB}{A^2}x_2^2 + \frac{cA^2}{A^2}x_2^2 \\ &= \left(aB^2 - 2bAB + cA^2\right)\left(\frac{x_2}{A}\right)^2. \end{aligned}$$

Everything depends on the sign of $aB^2 - 2baAB + cA^2$. We can write this in a somewhat different fashion as

$$aB^2 - 2bAB + cA^2 = -\begin{vmatrix} 0 & A & B \\ A & a & b \\ B & b & c \end{vmatrix}. \tag{16.19.4}$$

Let's rewrite the constraint as $(A, B)\mathbf{x} = 0$. We formed the matrix on the right-hand side of equation (16.19.4) by adding a border to $\mathbf{A}$ consisting of (A, B) on the top, and its transpose on the side, while adding a zero to fill out the new matrix.

The calculations above prove the theorem for $A \neq 0$. ∎

Constraints with Definite Matrices. If $\mathbf{A}$ is positive or negative definite, we don't need to the use the bordered matrix. We already know there is a unique maximum or minimum. The linear constraint does not exclude it, no can it change the value of the form elsewhere, and it remains a maximum or minimum. In fact,

$$aB^2 - 2AB + cA^2 = \begin{pmatrix} B & -A \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} B \\ -A \end{pmatrix}$$

which is positive when $\mathbf{A}$ is positive definite and negative when $\mathbf{A}$ is negative definite. The bordered matrix adds no information in this case.

16.20 Bordered Hessians and Relatives

Let $\mathbf{A}$ be an $m \times m$ symmetric matrix. It defines a quadratic form $Q_{\mathbf{A}}$ in the usual way:

$$Q_{\mathbf{A}}(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}.$$

Suppose we have a set of linear constraints given by an $k \times m$ matrix $\mathbf{B}$,

$$\mathbf{B}\mathbf{x} = \mathbf{0}.$$

We will be interested in the problem of determining the definiteness of $Q_{\mathbf{A}}$ on the linear constraint set $L = \{\mathbf{x} \in \mathbb{R}^m : \mathbf{B}\mathbf{x} = \mathbf{0}\}$.

To that end, we construct the bordered matrix $\mathbf{H}$ by taking $\mathbf{A}$, putting $\mathbf{B}$ as a border above it, $\mathbf{B}^T$ as a border to the left, and filling in the rest of the $(k+m) \times (k+m)$ matrix by zeros. The result is

$$\mathbf{H} = \begin{pmatrix} \mathbf{0}_k & \mathbf{B} \\ \mathbf{B}^T & \mathbf{A} \end{pmatrix} = \left(\begin{array}{ccc|ccc} 0 & \cdots & 0 & b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & b_{k1} & \cdots & b_{km} \\ \hline b_{11} & \cdots & b_{k1} & a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ b_{1m} & \cdots & b_{km} & a_{1m} & \cdots & a_{mm} \end{array}\right) \tag{16.20.5}$$

Since $\mathbf{A}$ is symmetric, we have written a_{1m} for a_{m1}.

16.21 Bordered Matrices and Definite Quadratic Forms

Theorem 16.21.1. *Suppose a quadratic form on* $\mathbb{R}^m$ *can be written* $Q(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ *for a symmetric* $m \times m$ *matrix* $\mathbf{A}$. *To determine the definiteness of* Q *on a linear constraint set* $L = \{\mathbf{x} \in \mathbb{R}^m : \mathbf{B}\mathbf{x} = \mathbf{0}\}$ *where* $\mathbf{B}$ *is a* $k \times m$ *matrix with* $m > k$ *and* $\operatorname{rank}\mathbf{B} = k$, *form the bordered matrix* $\mathbf{H}$ *by*

$$\mathbf{H} = \begin{pmatrix} \mathbf{0}_k & \mathbf{B} \\ \mathbf{B}^T & \mathbf{A} \end{pmatrix}$$

and check the signs of the last $m - k$ *leading principal minors of* $\mathbf{H}$*:*

$$\det \mathbf{H}_{m+k},\ \det \mathbf{H}_{m+k-1},\ \dots,\ \det \mathbf{H}_{2k+1}.$$

(a) *If* $\det \mathbf{H}(-1)^m = \det \mathbf{H}_{m+k}(-1)^m > 0$ *and the leading principal minors* $\det \mathbf{H}_j$, *for* $j = 2k+1, \dots, k+m$ *alternate in sign, then* Q *is negative definite on the constraint set* L. *Moreover,* $\mathbf{x} = \mathbf{0}$ *is the unique global maximizer of* Q *on* L.

(b) *If* $\det \mathbf{H}(-1)^k = \det \mathbf{H}_{m+k}(-1)^k > 0$ *and the leading principal minors* $\det \mathbf{H}_j$, *for* $j = 2k+1, \dots, k+m$ *have the same sign, then* Q *is positive definite on the constraint set* L. *Moreover,* $\mathbf{x} = \mathbf{0}$ *is the unique global minimizer of* Q *on* L.

(c) *If both (a) and (b) are violated by non-zero leading principal minors* $\det \mathbf{H}_j$ *for* $j = 2k+1, \dots, k+m$, *then* Q *is indefinite on the constraint set* L. *Moreover, there is neither a global maximum nor global minimum of* Q *in the constraint set* L.

There are other forms of this theorem. In particular, the bordered matrix is sometimes written

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0}_k \end{pmatrix}.$$

This affects the way the sign conditions are written, as detailed in Simon and Blume (pg. 391-392), but the differences are only cosmetic.

16.22 What About the First $2k$ Minors?

You might wonder why we ignore the first $2k$ minors in Theorem 16.21.1. The first k are easy. These are the determinants of submatrices of the zero matrix. They're all zero. But what about the rest?

Because of all the zeros in the upper right hand corner of the bordered matrix, the next $(k-1)$ minors must also be zero. You can see how it works in the example below. The example also looks at the $(2k)$ minor, which contains no terms from $\mathbf{A}$.

▶ Example 16.22.1: The Ignored Minors. Consider the bordered matrix

$$\begin{pmatrix} 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 4 & 9 \\ 1 & 1 & a_{11} & a_{12} & a_{13} \\ 2 & 4 & a_{12} & a_{22} & a_{23} \\ 3 & 9 & a_{13} & a_{23} & a_{33} \end{pmatrix}$$

Here $k = 2$ and $m = 3$. So $2k - 1 = 3$, leaving us only the last minor to check $(m - k = 1)$. The first two leading principal minors are zero as all the terms in the matrix are zero. As for the third leading principal minor,

$$\mathbf{H}_3 = \begin{vmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & a_{11} \end{vmatrix} = 1 \begin{vmatrix} 0 & 1 \\ 0 & 1 \end{vmatrix} = 0$$

because there was still a zero column remaining in the $(3,1)$ cofactor of $\mathbf{H}_3$.

Finally, $\det \mathbf{H}_4$ works a bit differently. It doesn't necessarily reduce to a matrix with a zero column. But it doesn't depend on $\mathbf{A}$ either.

$$\mathbf{H}_4 = \begin{vmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 1 & a_{11} & a_{12} \\ 2 & 4 & a_{12} & a_{22} \end{vmatrix} = \begin{vmatrix} 0 & 1 & 2 \\ 0 & 1 & 4 \\ 4 & a_{12} & a_{22} \end{vmatrix} - 2 \begin{vmatrix} 0 & 1 & 2 \\ 0 & 1 & 4 \\ 1 & a_{11} & a_{12} \end{vmatrix} = 8 - 4 = 4.$$

Notice how the zeros ensured that the a_{ij} would not appear in the final result. ◀

16.23 The First $(2k-1)$ Minors are Zero

In example 16.22.1, the first $(2k-1)$ minors in Theorem 16.21.1 were zero. In fact, they always are!

Theorem 16.23.1. *Let* $\mathbf{A}$ *be a symmetric* $m \times m$ *matrix* $\mathbf{A}$ *and* $\mathbf{B}$ *a* $k \times m$ *matrix with* $k < m$. *Then the first* $(2k-1)$ *minors of the bordered matrix* $\mathbf{H}$

$$\mathbf{H} = \begin{pmatrix} 0_k & \mathbf{B} \\ \mathbf{B}^T & \mathbf{A} \end{pmatrix}$$

are zero.

Proof. The first k principal submatrices are submatrices of the zero matrix. Their determinants are zero. When we expand the other determinants, we will eventually encounter a zero column. Notice that for $2k-1>0$, we must have $k>1$.

To see how the expansion works, first consider $\mathbf{H}_{k+1}$. We expand along the first column, obtaining a sum of $k \times k$ cofactor. Each cofactor is based on a submatrix that has lost the first column and last row. This submatrix has a $k \times (k-1)$ zero block in the upper left. The fact that one or more columns is zero means the cofactors are zero.

The $(2k-1)$ case is the toughest, but it happens even there. The $(2k-1)$ submatrix has a $k \times k$ zero matrix in the upper left-hand corner. All of the other entries are potentially non-zero. Let's expand the determinant along the first column. The first k terms are zero, so we focus on the last $(k-1)$ terms. These involve cofactors that include a $k \times (k-1)$ zero block. Expand by cofactors along the second column. These include a $k \times (k-2)$ zero block. This continues a total of $(k-1)$ times, eventually yielding cofactors with a $k \times 1$ zero block. At this point, the cofactors come from $k \times k$ submatrices, and the zero column makes the determinant zero. Smaller principal submatrices would have reached the zero column point sooner. ∎

As for $\mathbf{H}_{2k}$, which is also ignored in Theorem 16.21.1, it is a $(2k) \times (2k)$ matrix with the form

$$\begin{pmatrix} 0_k & \mathbf{B}_k^T \\ \mathbf{B}_k^T & \mathbf{A}_k \end{pmatrix}.$$

Where $\mathbf{B}_k$ is the submatrix consisting of the first k columns of $\mathbf{B}$ and $\mathbf{A}_k$ is the k^{th} leading principal minor of $\mathbf{A}$. The presence of the $k \times k$ zero block means that when we expand its determinant as above, we obtain zeroes until we have eliminated the block. By that time, we have also eliminated the k rows containing $\mathbf{A}_k$. This means that $\det \mathbf{H}_k$ doesn't contain any entries from $\mathbf{A}$, and so cannot inform us about the definiteness of $\mathbf{A}$. The minor $\mathbf{H}_{2k+1}$ is the first leading principal minor that contains any entries from $\mathbf{A}$. That is way it is the first minor mentioned in Theorem 16.21.1.

November 15, 2020