Math 4571 (Advanced Linear Algebra)

Lecture #31

Quadratic Forms (Part 2):

- The Second Derivatives Test in $\mathbb{R}^n$
- Sylvester's Law of Inertia

This material represents §5.2.3 + §5.2.4 from the course notes.

Overview

In the previous lecture, we introduced quadratic forms and how to diagonalize quadratic forms on $\mathbb{R}^n$ using the spectral theorem.

We now continue with a few additional applications of the study of quadratic forms to analysis and geometry in $\mathbb{R}^n$.

Critical Points, I

Our next application is to establish a general version of the second derivative(s) test from calculus.

- A very basic application of calculus is to find minimum and maximum values of functions (of one or several variables).
- For a function f, the partial derivative f_x is the rate of change of a function as x varies and the other variables are held fixed.
- Thus, any local minimum or maximum of a differentiable function must occur at a critical point, where all the partial derivatives are zero.
- However, once all critical points are identified, it can be difficult to determine whether the critical points are actually local minima or local maxima, especially for functions of more than one or two variables.
- The purpose of the second derivatives test is to allow us to classify critical points easily.

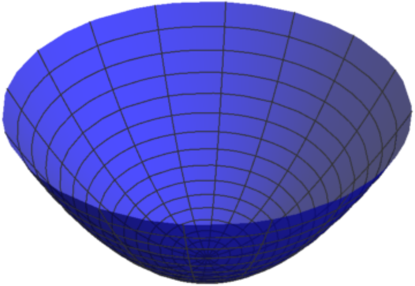
Critical Points, II

We briefly review the terminology for critical points for a function f of several variables.

- A local minimum is a critical point where f nearby is always bigger.
- A local maximum is a critical point where f nearby is always smaller.
- A saddle point is a critical point where f nearby is bigger in some directions and smaller in others.
- Formally, the "nearby" condition means that for any $\epsilon > 0$ there exists a point within a distance ϵ of the critical point with the given property.

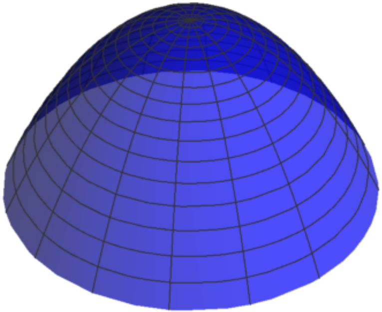
Critical Points, III

<u>Examples</u>: The function $f(x, y) = x^2 + y^2$ has a local minimum at its critical point $(0, 0)$.



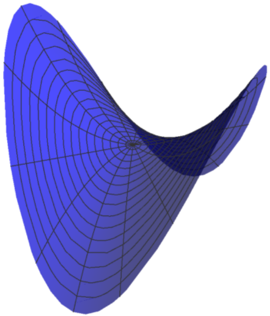
Critical Points, IV

Examples: The function $f(x, y) = -x^2 - y^2$ has a local maximum at its critical point $(0, 0)$.



Critical Points, V

<u>Examples</u>: The function $h(x, y) = x^2 - y^2$ has a saddle point at its critical point $(0, 0)$.



Saddle points are so named because their graphs have a similar shape to this "saddle surface": along one direction the point looks like a local minimum, and along another direction the point looks like a local maximum.

Second Derivatives Test, I

We can use quadratic forms to prove the famous "second derivatives test" from multivariable calculus:

Theorem (Second Derivatives Test in $\mathbb{R}^n$)

Suppose f is a function of n variables $x_1, \dots, x_n$ that is twice-differentiable and P is a critical point of f, so that $f_{x_i}(P) = 0$ for each i. Let H be the Hessian matrix, whose (i, j)-entry is the second-order partial derivative $f_{x_i x_j}(P)$.

- *If all eigenvalues of H are positive then P is a local minimum.*
- *If all eigenvalues of H are negative then P is a local maximum.*
- *If H has at least one eigenvalue of each sign then P is a saddle point.*
- *In all other cases (where H has at least one zero eigenvalue and does not have one of each sign) the test is inconclusive.*

Second Derivatives Test, II

Proof:

- By translating appropriately, assume that P is at the origin.
- By the multivariable version of Taylor's theorem in $\mathbb{R}^n$, the function $f(x_1, \dots, x_n) - f(P)$ will be closely approximated by its degree-2 Taylor polynomial T, which has the form $T = \sum_{1 \le i \le j \le n} a_{i,j} x_i x_i$, where $a_{i,j} = \begin{cases} f_{x_i,x_i}(P)/2 & \text{for } i = j \\ f_{x_i,x_j}(P) & \text{for } i \neq j \end{cases}$.
- Now observe T is a quadratic form whose associated bilinear form has matrix $H/2$.
- We can then classify the behavior of f at the critical point by diagonalizing this quadratic form. The only needed information is the values of the eigenvalues, since they will determine the local behavior of f near P.

Second Derivatives Test, III

<u>Proof</u> (continued):

- With new coordinates $x_1', \ldots, x_n'$, $f(x_1, \ldots, x_n) - f(P) = \frac{1}{2}\lambda_1(x_1')^2 + \cdots + \frac{1}{2}\lambda_n(x_n')^2 + O[(x_1')^2 + \cdots + (x_n')^2]$.
- If all $\lambda_i > 0$, the error term is smaller than the remaining terms. Then $f(x_1, \ldots, x_n) - f(P) > 0$ sufficiently close to P, so P is a local minimum.
- Likewise, if all $\lambda_i < 0$, we see $f(x_1, \ldots, x_n) - f(P) < 0$ sufficiently close to P, so P is a local maximum.
- If $\lambda_i > 0$ and $\lambda_j < 0$, then approaching P along $\mathbf{x}_i$ (respectively, along $\mathbf{x}_j$) yields values of f greater than at P (respectively, less than at P), so P is a saddle point.
- The other cases are inconclusive because we can take (for example) the functions $f = x_1^2 + x_2^4$ and $g = x_1^2 - x_2^4$: then H has a single nonzero eigenvalue (corresponding to x_1), but f has a local minimum while g has a saddle point.

Second Derivatives Test, IV

Example: Classify the critical point at $(0, 0)$ for the function $f(x, y) = 2x^2 + xy + 4y^2$.

Second Derivatives Test, IV

<u>Example</u>: Classify the critical point at $(0, 0)$ for the function $f(x, y) = 2x^2 + xy + 4y^2$.

- We compute the Hessian matrix: we have $f_{xx} = 4$, $f_{xy} = f_{yx} = 1$, and $f_{yy} = 8$, so evaluating these at $(0, 0)$ yields $H = \begin{bmatrix} 4 & 1 \\ 1 & 8 \end{bmatrix}$.
- The characteristic polynomial of H is $p(t) = \det(tI_2 - H) = t^2 - 12t + 31$, whose roots are $\lambda = 3 \pm \sqrt{2}$.
- Since the eigenvalues are both positive, the critical point is a $\boxed{\text{local minimum}}$.

Second Derivatives Test, V

Example: Classify the critical point at $(0,0)$ for the function $f(x,y) = x^2 + 3xy - 6y^2 + x^5y^3$.

Second Derivatives Test, V

<u>Example</u>: Classify the critical point at $(0,0)$ for the function $f(x,y) = x^2 + 3xy - 6y^2 + x^5y^3$.

- We compute the Hessian matrix: we have $f_{xx} = 2 + 20x^3y^3$, $f_{xy} = f_{yx} = 3 + 15x^4y^2$, and $f_{yy} = -6 + 6x^5y$, so evaluating these at $(0,0)$ yields $H = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$.
- The characteristic polynomial of H is $p(t) = \det(tI_2 - H) = t^2 - 4t - 21 = (t-7)(t+3)$.
- Since the eigenvalues are -7 and 3, there is an eigenvalue of each sign, so the critical point is a $\boxed{\text{saddle point}}$.

Second Derivatives Test, VI

Example: Classify the critical point at $(0,0,0)$ for the function $f(x,y,z) = 3x^2 + 2xy - xz + y^2 - yz + z^2$.

Second Derivatives Test, VI

<u>Example</u>: Classify the critical point at $(0, 0, 0)$ for the function $f(x, y, z) = 3x^2 + 2xy - xz + y^2 - yz + z^2$.

- We compute the Hessian matrix: we have $f_{xx} = 6$, $f_{xy} = f_{yx} = 2$, $f_{xz} = f_{zx} = -1$, $f_{yy} = 2$, $f_{yz} = f_{zy} = -1$, and $f_{zz} = 2$, so $H = \begin{bmatrix} 6 & 2 & -1 \\ 2 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$.
- The characteristic polynomial of H is $p(t) = \det(tI_2 - H) = t^3 - 10t^2 + 22t - 12 = (t-2)(t^2 - 8t + 6)$, whose roots are $\lambda = 2, 4 \pm \sqrt{10}$.
- Since the eigenvalues are all positive, the critical point is a $\boxed{\text{local minimum}}$.

Positive-Definiteness, I

A fundamental component of the classification in the second derivatives test was the behavior of the quadratic form (and in particular, whether it was "always positive" or "always negative" for nonzero inputs).

This behavior is quite important and we will record it:

Definition

A quadratic form on a real vector space is positive definite if $Q(\mathbf{v}) > 0$ for every nonzero vector $\mathbf{v} \in V$, and it is negative definite if $Q(\mathbf{v}) < 0$ for every nonzero vector $\mathbf{v} \in V$.

Example:

- If V is a real inner product space, then the square of the norm $||\mathbf{v}||^2 = \langle \mathbf{v}, \mathbf{v} \rangle$ is a positive-definite quadratic form on V.
- Indeed, the positive-definiteness condition is precisely axiom [I3] from the definition of the inner product.

Positive-Definiteness, II

We can in fact detect positive-definiteness for quadratic forms on finite-dimensional spaces by calculating the eigenvalues of the associated matrix.

- As noted in the proof of the second derivatives test, if a real quadratic form is positive definite, then all the diagonal entries in its diagonalization are positive. Likewise, if a real quadratic form is negative definite, then all the diagonal entries in its diagonalization are negative.
- It is not hard to see that the converse of this statement holds also, by considering the diagonalization: a real quadratic form is positive definite if and only all its eigenvalues are positive, while it is positive semidefinite if and only if all its eigenvalues are nonnegative.

Positive-Definiteness, III

There are also useful weaker versions of these conditions:

Definition

We say Q is positive semidefinite if $Q(\mathbf{v}) \geq 0$ for all $\mathbf{v} \in V$ and negative semidefinite if $Q(\mathbf{v}) \leq 0$ for all $\mathbf{v} \in V$.

We can also detect these properties by examining eigenvalues: positive semidefinite quadratic forms have nonnegative eigenvalues, while negative semidefinite quadratic forms have nonpositive eigenvalues.

Positive-Definiteness, IV

Example: Determine whether the quadratic form $Q[(x, y)] = 2x^2 - 4xy + 4y^2$ on $\mathbb{R}^2$ is positive definite.

Positive-Definiteness, IV

Example: Determine whether the quadratic form $Q[(x,y)] = 2x^2 - 4xy + 4y^2$ on $\mathbb{R}^2$ is positive definite.

- The associated matrix for the bilinear form is $\begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}$ whose eigenvalues are $\lambda = 4 \pm \sqrt{10}$. Since these are both positive, Q is $\boxed{\text{positive definite}}$.

Sylvester's Law of Inertia, I

As our final topic, we will discuss the possible diagonal entries for the diagonalization of a real quadratic form.

- By making different choices of basis (e.g., by rescaling it or selecting different row operations), we may obtain different diagonalizations of a given real quadratic form.
- For example, with the quadratic form $Q(x, y) = x^2 + 2y^2$, which is already diagonal, if we change basis to $x' = x/2$, $y' = y/3$, then we obtain $Q(x, y) = 4(x')^2 + 18(y')^2$.
- Indeed, by rescaling, we may change any positive coefficient to an arbitrary positive value and any negative coefficient to an arbitrary negative value.
- Our claim is that this is essentially the only possible change we may make to the diagonalization over $\mathbb{R}$.

Sylvester's Law of Inertia, II

As our final topic, we will discuss the possible diagonal entries for the diagonalization of a real quadratic form.

Theorem (Sylvester's Law of Inertia)

Suppose V is a finite-dimensional real vector space and Q is a quadratic form on V. Then the numbers of positive diagonal entries, zero diagonal entries, and negative diagonal entries in any diagonalization of Q is independent of the diagonalization.

The idea is: we decompose V as a direct sum of three spaces, one on which Q acts as a positive-definite quadratic form (giving positive diagonal entries), one on which Q acts as the zero map (giving zero diagonal entries), and one on which Q acts as a negative-definite quadratic form (giving negative diagonal entries). These dimensions then depend only on Q, hence are invariant.

Sylvester's Law of Inertia, III

Proof:

- Since $\mathrm{char}(F) \neq 2$, we may equivalently work with the symmetric bilinear form Φ associated to Q.
- Let V_0 be the subspace of V given by $V_0 = \{\mathbf{v}_0 \in V : \Phi(\mathbf{v}_0, \mathbf{v}) = 0 \text{ for all } \mathbf{v} \in V\}$: then Φ acts as the zero map on V_0. Now write $V = V_0 \oplus V_1$: we claim that Φ is nondegenerate on V_1.
- To see this, suppose $\mathbf{y} \in V_1$ has $\Phi(\mathbf{y}, \mathbf{v}_1) = 0$ for all $\mathbf{v}_1 \in V$: then for any $\mathbf{w} \in V$ we may write $\mathbf{w} = \mathbf{v}_0 + \mathbf{v}_1$ for $\mathbf{v}_i \in V_i$, in which case $\Phi(\mathbf{y}, \mathbf{w}) = \Phi(\mathbf{y}, \mathbf{v}_0) + \Phi(\mathbf{y}, \mathbf{v}_1) = 0$. But this would imply $\mathbf{y} \in V_0$ whence $\mathbf{y} = \mathbf{0}$.
- Thus, Φ is nondegenerate on V_1.

Sylvester's Law of Inertia, IV

Proof (continued more):

- Now we show that if Φ is nondegenerate on V_1, then V_1 decomposes as a direct sum $V_1 = V_+ \oplus V_-$, where Φ is positive-definite on V_+ and negative-definite on V_-.
- Let V_+ be the maximal subspace of V_1 on which Φ is positive-definite (since the condition is defined only on individual vectors, this subspace is well-defined), and define $V_- = \{\mathbf{w} \in V : \Phi(\mathbf{v}_+, \mathbf{w}) = 0 \text{ for all } \mathbf{v}_+ \in V_+\}$.
- Then by an application of Gram-Schmidt (which in fact holds for bilinear forms) via Φ, we see that $V_1 = V_+ \oplus V_-$.

Sylvester's Law of Inertia, V

<u>Proof</u> (continued even yet still more):

- It remains to show that Φ is negative-definite on V_-, so let $\mathbf{z} \in V_-$ be nonzero. Then by assumption, Φ is not positive-definite on $V_+ \oplus \langle \mathbf{z} \rangle$, so there exists some nonzero $\mathbf{v} = \mathbf{v}_+ + \alpha \mathbf{z}$ with $\mathbf{v}_+ \in V_+$ and $\alpha \in \mathbb{R}$ such that $\Phi(\mathbf{v}, \mathbf{v}) \leq 0$.
- We cannot have $\alpha = 0$ since then positive-definiteness would imply $\mathbf{v}_+ = 0$. Since $\Phi(\mathbf{v}, \mathbf{v}) = \Phi(\mathbf{v}_+, \mathbf{v}_+) + 2\alpha\Phi(\mathbf{v}_+, \mathbf{z}) + \alpha^2\Phi(\mathbf{z}, \mathbf{z}) = \Phi(\mathbf{v}_+, \mathbf{v}_+) + \alpha^2\Phi(\mathbf{z}, \mathbf{z})$, we have $\Phi(\mathbf{z}, \mathbf{z}) = \dfrac{1}{\alpha^2}[\Phi(\mathbf{v}, \mathbf{v}) - \Phi(\mathbf{v}_+, \mathbf{v}_+)]$.
- Then both terms are less than or equal to zero, and both cannot be zero. Hence $\Phi(\mathbf{z}, \mathbf{z}) < 0$ for all nonzero $\mathbf{z} \in V_-$ and so Φ is negative-definite on V_-.

Sylvester's Law of Inertia, VI

Proof (continued even further yet still additionally more also):

- We now establish the result using from the direct sum decomposition $V = V_0 \oplus V_+ \oplus V_-$.
- If we select any diagonalization, the restriction to the subspace generated by the basis vectors with diagonal entries 0, positive, negative is trivial, positive-definite, negative-definite (respectively), and thus the number of such diagonal elements is at least $\dim(V_0)$, $\dim(V_+)$, $\dim(V_-)$ (respectively).
- But since the total number of diagonal elements is $\dim(V) = \dim(V_0) + \dim(V_+) + \dim(V_-)$, we must have equality everywhere.
- Hence the numbers of positive diagonal entries, zero diagonal entries, and negative diagonal entries in any diagonalization of Q is independent of the choice of diagonalization, as claimed.

Sylvester's Law of Inertia, VII

We will also mention that there is some classical terminology associated with Sylvester's law of inertia.

Definition

If Q is a quadratic form on $\mathbb{R}^n$, the index of Q is the number of positive diagonal entries (in any diagonalization) and the signature is the difference between the number of positive and negative diagonal entries.

Examples:

- $Q(x, y, z) = x^2 - y^2 - z^2$ on $\mathbb{R}^3$ has index 1 and signature -1.
- $Q(x, y, z) = x^2 - z^2$ on $\mathbb{R}^3$ has index 1 and signature 0.
- $Q(x, y, z) = 5x^2 + 4xy + 6y^2 + 4yz + 7z^2$ on $\mathbb{R}^3$ has index 3 and signature 3, since we computed its diagonalization to have diagonal entries $3, 6, 9$.

Sylvester's Law of Inertia, VIII

From our discussion of the spectral theorem, the index is equal to the number of positive eigenvalues of the matrix associated to the symmetric bilinear form, while the signature is the difference between the number of positive eigenvalues and the number of negative eigenvalues.

- Some authors instead refer to the triple $(\dim V_+, \dim V_-, \dim V_0)$, or some appropriate permutation, as the signature of the quadratic form. These three values themselves are called the <u>invariants</u> of the form, and the value of any two of them (along with the dimension of the ambient space V) is sufficient to find the value of the other one.
- For nondegenerate forms, where there are no 0 entries (so $\dim V_0 = 0$), the dimension of the space along with the value of $\dim V_+ - \dim V_-$ is sufficient to find $\dim V_+$ and $\dim V_-$.

Sylvester's Law of Inertia, IX

As a corollary of Sylvester's law of inertia, we can read off the shape of a conic section or quadric surface (in all nondegenerate cases, and also in many degenerate cases) simply by examining the signs of the eigenvalues of the underlying quadratic form.

Examples:

- The eigenvalues of the symmetric matrix associated to the quadratic form $Q(x, y) = x^2 - 4xy$ are $\lambda = 4, -1$. Since one eigenvalue is positive and the other is negative, the conic $Q(x, y) = 1$ is a hyperbola.
- The eigenvalues of the symmetric matrix associated to the quadratic form $Q(x, y) = 3x^2 - 2xy + 3y^2$ are $\lambda = 2, 4$. Since both eigenvalues are positive, the conic $Q(x, y) = 1$ is an ellipse.

Sylvester's Law of Inertia, X

Example: Determine the shape of the quadric surface
$13x^2 - 4xy + 10y^2 - 8xz + 4yz + 13z^2 = 1$.

Sylvester's Law of Inertia, X

Example: Determine the shape of the quadric surface $13x^2 - 4xy + 10y^2 - 8xz + 4yz + 13z^2 = 1$.

- If $Q(x, y, z)$ is the quadratic form above, the bilinear form has associated matrix $A = \begin{bmatrix} 13 & -2 & -4 \\ -2 & 10 & 2 \\ -4 & 2 & 13 \end{bmatrix}$.
- The characteristic polynomial is $p(t) = \det(tI_3 - A) = t^3 - 144t^2 + 6480t - 93312 = (t-36)^2(t-72)$.
- This means, upon diagonalizing $Q(x, y, z)$, we will obtain the equation $36(x')^2 + 36(y')^2 + 72(z')^2 = 1$. This is the equation of an ellipsoid.
- Again, the only information we really needed here, to see that the quadratic variety is an ellipsoid, was that all the eigenvalues were positive.

Quadratic Forms, Summary

We will end the lecture by remarking that the study of quadratic forms touches on nearly every branch of mathematics.

- We have already discussed some ties to linear algebra (in the guise of bilinear forms and diagonalization), analysis (in the classification of critical points), and geometry (in the analysis of quadratic varieties and matrices acting on quadratic forms).
- Also, the study of quadratic forms over $\mathbb{Q}$ turns out to be intimately tied with many topics in number theory.
- One very classical problem is to characterize integers that can be written as the sum of two squares, which is a question about integers represented by $Q(x, y) = x^2 + y^2$.
- This family of problems is intimately related to a number of very deep results in modern number theory (in particular, it greatly motivated the development of class field theory).

It's The End of Linear Algebra!

We reviewed local minima, local maxima, and saddle points.

We proved the second derivatives test for classifying critical points of real-valued functions on $\mathbb{R}^n$.

We proved Sylvester's law of inertia and its implications for classifications of quadratic varieties, and also discussed the signature and index of a quadratic form.

But There's Always More Algebra To Learn...

We're now at the end of the course (except of course for the final).

I hope you enjoyed learning a bunch of linear algebra with me this semester as much as I enjoyed teaching it. Linear algebra is an incredibly fundamental subject, and I hope you've both gotten an appreciation for the theoretical side, and also enjoyed the variety of applications we've been able to discuss.

If you did in fact enjoy the course, I would greatly appreciate it if you took the time to fill out the TRACE evaluations and mention that fact.

Thanks, and good luck on the final!