Math 2270-004 Week 15 notes
We will not necessarily finish the material from a given day's notes on that day. We may also add or subtract some material as the week progresses, but these notes represent an in-depth outline of what we plan to cover. These notes cover material for Monday. I'll add course review material for Tuesday later.

Mon Apr 23

- 7.2 Second derivative test, and maybe another conic diagonalization example.

Announcements:

Warm-up Exercise:

From last week ....

Spectral Theorem Let A be an $n \times n$ symmetric matrix. Then all of the eigenvalues of A are real, and there exists an orthonormal eigenbasis $B = \{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots \underline{\boldsymbol{u}}_n\}$ consisting of eigenvectors for A. Eigenspaces with different eigenvalues are automatically orthogonal to each other. If any eigenspace has dimension greater than 1, its orthonormal basis may be constructed via *Gram − Schmidt*.

Diagonalization of quadratic forms: Let

$$Q(\underline{\boldsymbol{x}}) = \sum_{i,j=1}^{n} a_{ij} x_i x_j = \underline{\boldsymbol{x}}^T A \underline{\boldsymbol{x}}$$

for a symmetric matrix A, with real entries. A symmetric $\Rightarrow$ by the spectral theorem there exists an orthonormal eigenbasis $B = \{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots \underline{\boldsymbol{u}}_n\}$.

For the corresponding orthogonal matrix

$$P = [\underline{\boldsymbol{u}}_1 | \underline{\boldsymbol{u}}_2 | \cdots | \underline{\boldsymbol{u}}_n]$$

$$D = P^T A P,$$

where D is the diagonal matrix of eigenvalues corresponding to the eigenvectors in P. And we have

$$\underline{\boldsymbol{x}} = P \underline{\boldsymbol{y}}$$

where $\underline{\boldsymbol{y}} = [\underline{\boldsymbol{x}}]_B$ and $P = P_{E \leftarrow B}$. Thus

$$Q(\underline{\boldsymbol{x}}) = \underline{\boldsymbol{x}}^T A \underline{\boldsymbol{x}}$$

$$= \underline{\boldsymbol{y}}^T P^T A P \underline{\boldsymbol{y}} = \underline{\boldsymbol{y}}^T D \underline{\boldsymbol{y}}$$

$$= \sum_{i=1}^{n} \lambda_i y_i^2 .$$

So by the orthogonal change of variables all cross terms have been removed. Applications include conic curves, quartic surfaces, multivariable second derivative test, principal component analysis (PCA) in statistics, singular value matrix decomposition (SVD) in geometry and computer science, and more.

Definition: The quadratic form $Q(\underline{\mathbf{x}}) = \sum_{i,j=1}^{n} a_{ij} x_i x_j = \underline{\mathbf{x}}^T A \underline{\mathbf{x}}$ (for A a symmetric matrix) is called *positive definite* if

$$Q(\underline{\mathbf{x}}) > 0 \quad \text{for all } \underline{\mathbf{x}} \neq \underline{\mathbf{0}}.$$

We see that this is the same as saying that all of the eigenvalues of A are positive.

Definition: The quadratic form $Q(\underline{\mathbf{x}}) = \sum_{i,j=1}^{n} a_{ij} x_i x_j = \underline{\mathbf{x}}^T A \underline{\mathbf{x}}$ (for A a symmetric matrix) is called *negative definite* if

$$Q(\underline{\mathbf{x}}) < 0 \quad \text{for all } \underline{\mathbf{x}} \neq \underline{\mathbf{0}}.$$

We see that this is the same as saying that all of the eigenvalues of A are negative.

First and second derivative tests from multivariable calculus, revisited. It turns out that a lot of multivariable calculus is easier to understand once you know linear algebra. This is just one example of where that happens. (Math majors will see this, and quite a bit more, in Math 3220.) Let

$$f : \mathbb{R}^n \to \mathbb{R}.$$

$$f(x_1, x_2, \ldots, x_n) = f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right).$$

$$\underline{\boldsymbol{x}}_0 \in \mathbb{R}^n,\ \underline{\boldsymbol{u}} \in \mathbb{R}^n \text{ a unit vector.}$$

Then

$$D_{\underline{\boldsymbol{u}}} f(\underline{\boldsymbol{x}}_0) := \frac{d}{dt}\left(f(\underline{\boldsymbol{x}}_0 + t\,\underline{\boldsymbol{u}})\right)\Big|_{t=0}$$

is the rate of change of f in the direction of $\underline{\boldsymbol{u}}$, at x_0. ("The directional derivative of f, at $\underline{\boldsymbol{x}}_0$, in the direction of $\underline{\boldsymbol{u}}$". This generalizes pure partial derivatives, which are rates of change in the standard coordinate-directions.) Using the multivariable version of the chain rule we compute

$$\frac{d}{dt} f(x_0 + t\,u) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\underline{\boldsymbol{x}}_0 + t\,\underline{\boldsymbol{u}}) \frac{d}{dt}(\underline{\boldsymbol{x}}_{0,i} + t\,\underline{\boldsymbol{u}}_i) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\underline{\boldsymbol{x}}_0 + t\,\underline{\boldsymbol{u}})\, u_i\,.$$

$$= \nabla f(\underline{\boldsymbol{x}}_0 + t\,\underline{\boldsymbol{u}}) \cdot \underline{\boldsymbol{u}}.$$

So at $t = 0$ this rate of change is computed via

$$D_{\underline{\boldsymbol{u}}} f(\underline{\boldsymbol{x}}_0) = \nabla f(\underline{\boldsymbol{x}}_0) \cdot \underline{\boldsymbol{u}}\,.$$

<u>Definition:</u> Let f be a differentiable function as above. Then $\underline{\boldsymbol{x}}_0$ is a *critical point* for f if and only if

$$\nabla f(\underline{\boldsymbol{x}}_0) := \left[\frac{\partial f}{\partial x_1}(\underline{\boldsymbol{x}}_0),\ \frac{\partial f}{\partial x_1}(\underline{\boldsymbol{x}}_0),\ \ldots.\ \frac{\partial f}{\partial x_n}(\underline{\boldsymbol{x}}_0)\right] = \underline{\boldsymbol{0}}\,.$$

In other words, a critical point is a point at which *all* directional derivatives are zero. Local extrema of differentiable functions will only occur at critical points, but not all critical points are the locations of local extrema. The ways in which things can go wrong are more interesting than in the single-variable case, where we used the second derivative test.

In your first multivariable calculus class you were probably shown a second derivative test for functions of (only) two variables. The one you were probably shown obscures what's really going on, which is actually simpler to understand in general once you know linear algebra. Here's what you were probably shown (taken from the beginning of the Wikipedia article on this topic):

https://en.wikipedia.org/wiki/Second_partial_derivative_test

The test [edit]

Functions of two variables [edit]

Suppose that $f(x, y)$ is a differentiable real function of two variables whose second partial derivatives exist. The Hessian matrix H of f is the 2×2 matrix of partial derivatives of f:

$$H(x,y) = \begin{pmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{pmatrix}.$$

Define $D(x, y)$ to be the determinant

$$D(x,y) = \det(H(x,y)) = f_{xx}(x,y)f_{yy}(x,y) - (f_{xy}(x,y))^2,$$

of H. Finally, suppose that (a, b) is a critical point of f (that is, $f_x(a, b) = f_y(a, b) = 0$). Then the second partial derivative test asserts the following:[1]

1. If $D(a,b) > 0$ and $f_{xx}(a,b) > 0$ then (a,b) is a local minimum of f.
2. If $D(a,b) > 0$ and $f_{xx}(a,b) < 0$ then (a,b) is a local maximum of f.
3. If $D(a,b) < 0$ then (a,b) is a saddle point of f.
4. If $D(a,b) = 0$ then the second derivative test is inconclusive, and the point (a, b) could be any of a minimum, maximum or saddle point.

Note that other equivalent versions of the test are possible. For example, some texts may use the trace $f_{xx} + f_{yy}$ in place of the value f_{xx} in the first two cases above.[citation needed] Such variations in the procedure applied do not alter the outcome of the test.

Continuing the discussion about directional derivatives,

Definition: Let f, $\underline{\mathbf{x}}_0$, $\underline{\boldsymbol{u}}$ be as above. The *second derivative of f* at $\underline{\mathbf{x}}_0$, in the $\underline{\boldsymbol{u}}$ direction is defined by

$$\mathrm{D}_{\underline{\boldsymbol{u}}\,\underline{\boldsymbol{u}}}(f(x_0)) := \left.\frac{d^2}{dt^2} f(\underline{\mathbf{x}}_0 + t\,\underline{\boldsymbol{u}})\right|_{t=0}.$$

We compute this expression with the chain rule, starting with our expression for the first directional derivative, from the previous pages:

$$\frac{d^2}{dt^2} f(\underline{\mathbf{x}}_0 + t\,\underline{\boldsymbol{u}}) = \frac{d}{dt}\left(\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\underline{\mathbf{x}}_0 + t\,\underline{\boldsymbol{u}})\, u_i\right)$$

$$= \sum_{i=1}^{n} \frac{d}{dt}\frac{\partial f}{\partial x_i}(\underline{\mathbf{x}}_0 + t\,\underline{\boldsymbol{u}})\, u_i$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(\underline{\mathbf{x}}_0 + t\,\underline{\boldsymbol{u}})\, u_j\, u_i\ .$$

At $t = 0$ and recalling that $\frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{\mathbf{x}}_0) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{\mathbf{x}}_0)$ this reads

$$\mathrm{D}_{\underline{\boldsymbol{u}}\,\underline{\boldsymbol{u}}}(f(\underline{\mathbf{x}}_0)) := \left.\frac{d^2}{dt^2} f(\underline{\mathbf{x}}_0 + t\,\underline{\boldsymbol{u}})\right|_{t=0} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{\mathbf{x}}_0)\, u_i\, u_j\ .$$

Definition: The *Hessian matrix* of f at $\underline{\mathbf{x}}_0$, $[\mathrm{D}^2 f(\underline{\mathbf{x}}_0)]$ is the $n \times n$ (symmetric) matrix of second partial derivatives; $entry_{ij}\,[\mathrm{D}^2 f(\underline{\mathbf{x}}_0)] = \frac{\partial^2 f}{\partial x_j \partial x_i}(\underline{\mathbf{x}}_0) = f_{x_i x_j}(\underline{\mathbf{x}}_0)$:

$$[\mathrm{D}^2 f(\underline{\mathbf{x}}_0)] = \begin{bmatrix} f_{x_1 x_1}(\underline{\mathbf{x}}_0) & f_{x_1 x_2}(\underline{\mathbf{x}}_0) & \cdots & f_{x_1 x_n}(\underline{\mathbf{x}}_0) \\ f_{x_2 x_1}(\underline{\mathbf{x}}_0) & f_{x_2 x_2}(\underline{\mathbf{x}}_0) & \cdots & f_{x_2 x_n}(\underline{\mathbf{x}}_0) \\ \vdots & \vdots & \cdots & \cdots \\ f_{x_n x_1}(\underline{\mathbf{x}}_0) & f_{x_n x_2}(\underline{\mathbf{x}}_0) & \cdots & f_{x_n x_n}(\underline{\mathbf{x}}_0) \end{bmatrix}.$$

So,

$$\mathrm{D}_{\underline{\boldsymbol{u}}\,\underline{\boldsymbol{u}}}(f(\underline{\mathbf{x}}_0)) = \underline{\boldsymbol{u}}^T\,[\mathrm{D}^2 f(\underline{\mathbf{x}}_0)]\,\underline{\boldsymbol{u}}.$$

Theorem The function f is concave up in every direction $\underline{\boldsymbol{u}}$ at $\underline{\boldsymbol{x}}_0$ if and only if the Hessian matrix $\left[D^2 f(\underline{\boldsymbol{x}}_0)\right]$ is positive definite. The function f is concave down in every direction $\underline{\boldsymbol{u}}$ at $\underline{\boldsymbol{x}}_0$ if and only if the Hessian matrix $\left[D^2 f(\underline{\boldsymbol{x}}_0)\right]$ is negative definite. The first case happens if and only if all of the eigenvalues of $\left[D^2 f(\underline{\boldsymbol{x}}_0)\right]$ are positive, and the second case happens if and only if they are all negative. If $\underline{\boldsymbol{x}}_0$ is a critical point for f, then in the first case $f(\underline{\boldsymbol{x}}_0)$ is a local minimum value; and in the second case it is a local maximum value. If the Hessian has some negative and some positive eigenvalues, then $f(\underline{\boldsymbol{x}}_0)$ is neither a local minimum nor a local maximum. If all the eigenvalues are non-negative, or if they are all non-positive, but some are zero, then further work is required to determine whether $f(\underline{\boldsymbol{x}}_0)$ is a local extreme value.

Functions of many variables [edit]

For a function *f* of two or more variables, there is a generalization of the rule above. In this context, instead of examining the determinant of the Hessian matrix, one must look at the eigenvalues of the Hessian matrix at the critical point. The following test can be applied at any critical point *a* for which the Hessian matrix is invertible:

1. If the Hessian is positive definite (equivalently, has all eigenvalues positive) at *a*, then *f* attains a local minimum at *a*.
2. If the Hessian is negative definite (equivalently, has all eigenvalues negative) at *a*, then *f* attains a local maximum at *a*.
3. If the Hessian has both positive and negative eigenvalues then *a* is a saddle point for *f* (and in fact this is true even if *a* is degenerate).

In those cases not listed above, the test is inconclusive.[2]

The test [edit]

Functions of two variables [edit]

Suppose that *f*(*x*, *y*) is a differentiable real function of two variables whose second partial derivatives exist. The Hessian matrix *H* of *f* is the 2 × 2 matrix of partial derivatives of *f*:

$$H(x,y) = \begin{pmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{pmatrix}.$$

Define *D*(*x*, *y*) to be the determinant

$$D(x,y) = \det(H(x,y)) = f_{xx}(x,y)f_{yy}(x,y) - (f_{xy}(x,y))^2,$$

of *H*. Finally, suppose that (*a*, *b*) is a critical point of *f* (that is, $f_x(a, b) = f_y(a, b) = 0$). Then the second partial derivative test asserts the following:[1]

1. If $D(a,b) > 0$ and $f_{xx}(a,b) > 0$ then (a,b) is a local minimum of *f*.
2. If $D(a,b) > 0$ and $f_{xx}(a,b) < 0$ then (a,b) is a local maximum of *f*.
3. If $D(a,b) < 0$ then (a,b) is a saddle point of *f*.
4. If $D(a,b) = 0$ then the second derivative test is inconclusive, and the point (*a*, *b*) could be any of a minimum, maximum or saddle point.

Note that other equivalent versions of the test are possible. For example, some texts may use the trace $f_{xx} + f_{yy}$ in place of the value f_{xx} in the first two cases above.[citation needed] Such variations in the procedure applied do not alter the outcome of the test.

Exercise 1 Explain the (more complicated) second derivative test you were taught in multivariable calculus for functions of just two variables, as a special case of the the more general one that uses eigenvalues. Hint:

$$det\begin{bmatrix} f_{xx} - \lambda & f_{xy} \\ f_{yx} & f_{yy} - \lambda \end{bmatrix} = \lambda^2 - (f_{xx} + f_{yy})\lambda + \left(f_{xx} f_{yy} - f_{xy}^{\ 2}\right)$$

has roots

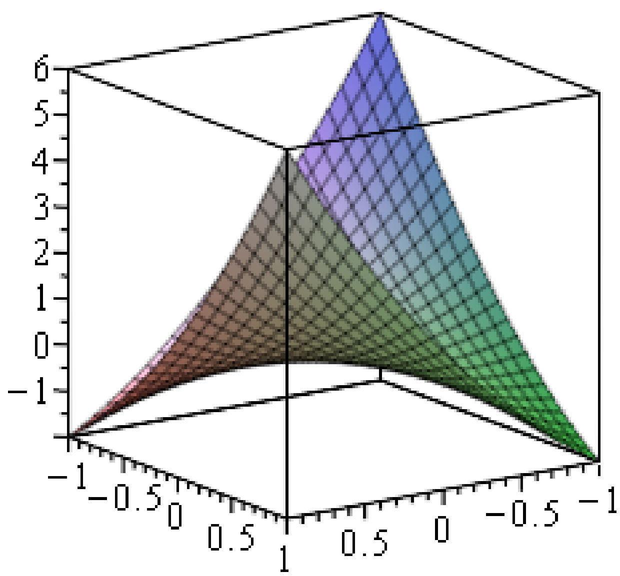
$$\lambda = \frac{(f_{xx} + f_{yy}) \pm \sqrt{(f_{xx} + f_{yy})^2 - 4\left(f_{xx} f_{yy} - f_{xy}^{\ 2}\right)}}{2}.$$

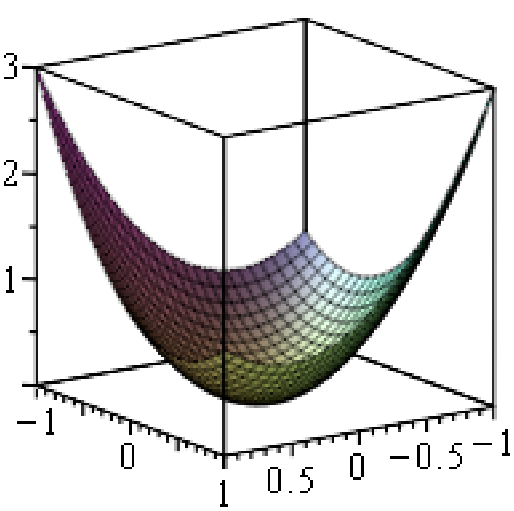
Exercise 2) Which of the following functions has a local minimum at the origin, if any? Could you diagonalize the associated quadratic forms and sketch level sets?

2a) $f(x, y) = x^2 + 4xy + y^2$

2b) $f(x, y) = x^2 - xy + y^2$.

> $with(plots):$
$plot3d(x^2 + 4\cdot x\cdot y + y^2, x=-1..1, y=-1..1);$
$plot3d(x^2 - x\cdot y + y^2, x=-1..1, y=-1..1);$





>