Extra Credit 1: Quadratic forms

Definition. A *quadratic form* in the variables $x_1, \dots, x_n$ is a polynomial of the form

$$Q(x_1, \dots, x_n) = \sum_{1 \le i \le j \le n} a_{ij} x_i x_j.$$

For example, a quadratic form in two variables has the form $ax^2 + bxy + cy^2$ for some constants a, b, c; in 3 variables,

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz.$$

Note that something like $2x^2 - xy + y^2 + x - 2$ is *not* a quadratic form, because terms of degree other than 2 appear.

These objects arise naturally, and there are (at least) two ways they're relevant for us.

1. They describe common geometric objects. For example, in 2 variables, the graph of $ax^2 + by^2$ is an elliptic paraboloid, and its level curves are ellipses (if $a, b > 0$); the graph of $ax^2 - by^2$ is a hyperbolic paraboloid, and its level curves are hyperbolas.

2. Consider the problem of classifying critical points of multivariable functions, points p such that $\nabla f(p) = 0$; how can we tell if p is a local minimum, a maximum, or neither? Take a real-valued function $f(x, y)$ whose third partial derivatives are continuous, defined in a neighborhood of a critical point $p = (a, b)$. Taylor's theorem lets us write

$$\begin{aligned} f(a+h, b+k) = f(a,b) &+ df(a, b; h, k) \\ &+ \frac{1}{2!}(h^2 f_{xx}(a,b) + 2hk f_{xy}(a,b) + k^2 f_{yy}(a,b)) + \epsilon(h,k) \end{aligned}$$

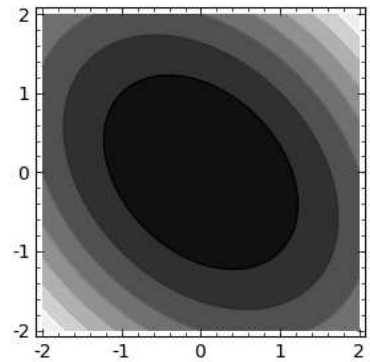
with some control over the error $\epsilon(h, k)$. Since $p = (a, b)$ is a critical point, $df(a, b; h, k) = 0$, and so we get

$$f(a+h, b+k) = f(a,b) + \frac{1}{2!}(h^2 f_{xx}(a,b) + 2hk f_{xy}(a,b) + k^2 f_{yy}(a,b)) + \epsilon(h,k).$$

Notice that, except for the error $\epsilon(h, k)$, this is a quadratic form in variables h, k, shifted by $f(a, b)$. In most cases we expect the error $\epsilon(h, k)$ to be dominated by the preceding terms, so it's reasonable to expect that the type of critical point for f is the same as the type of the critical point for the quadratic form. For the most part this turns out to work, and so it's important to understand the critical points of quadratic forms. See 7.6 of the text for the details of this argument, and below for a precise statement of when this works.

Problem 1. (1 pt). Show that a quadratic form (in any number of variables) has a critical point at the origin. Can there be any more?

In some cases, it's easy to see exactly what type the critical point of a quadratic form is. For example, $Q(x, y) = x^2 + y^2$ has a minimum at $(0, 0)$ and no other critical points. How about $Q(x, y) = 3x^2 + 2xy + 3y^2$? Here's a contour plot, where darker values are smaller:



The picture makes it clear that Q has a minimum at $(0, 0)$, and that the graph of Q is an elliptic paraboloid. In fact, if we change variables by $x = \frac{u+v}{\sqrt{2}}, y = \frac{u-v}{\sqrt{2}}$, we get $4u^2 + 2v^2$, whose graph is an elliptic paraboloid opening upwards. This transformation is clockwise rotation by 45 degrees.

1 Matrices and quadratic forms

Definition 1. A quadratic form Q is in *normal form* if $Q(x_1, \ldots, x_n) = a_1x_1^2 + \cdots + a_nx_n^2$ for some real numbers $a_1, \ldots, a_n$.

Problem 2. (1 pt). Let $Q(x_1, \ldots, x_n) = a_1x_1^2 + \cdots + a_nx_n^2$ be in normal form. Show that if $a_1, \ldots, a_n \geq 0$, then Q has a local minimum at the origin; if $a_1, \ldots, a_n \leq 0$, then Q has a local maximum at the origin; and if some a_i's are positive and some negative, then Q has neither a local maximum nor a local minimum at the origin.

This problem plus the last example demonstrates that if we can rotate a quadratic form Q about the origin to obtain one in normal form, then we'll immediately be able to see what sort of critical point Q has (certainly rotation won't change the type of critical point).

The key step in finding such a rotation is to connect quadratic forms to linear algebra. Observe:

$$
\begin{aligned}
ax^2 + bxy + cy^2 &= \left(ax^2 + \frac{1}{2}bxy\right) + \left(\frac{1}{2}bxy + cy^2\right) \\
&= x\left(ax + \frac{1}{2}by\right) + y\left(\frac{1}{2}bx + cy\right) \\
&= \langle x, y\rangle \cdot \langle ax + \frac{1}{2}by, \frac{1}{2}bx + cy\rangle \\
&= \langle x, y\rangle \cdot \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.
\end{aligned}
$$

Here we've used the fact that $v \cdot w = v^T w$ for (column) vectors v, w, where ^{T} is transpose.

Let $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, and let $A = \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix}$, given $Q(x, y) = ax^2 + bxy + cy^2$. We've shown that $Q(x, y) = \mathbf{x}^T A\mathbf{x}$. More generally, given a quadratic form $Q(x, y) = \sum_{1\le i\le j\le n} a_{ij} x_i x_j$, if we set $\mathbf{x} = (x_1\ x_2\ \cdots\ x_n)^T$, and let A be the matrix whose (i, j) entry is

$$
A_{ij} = \begin{cases} a_{ij} & \text{if } i = j \\ \frac{1}{2}a_{ij} & \text{if } i < j \end{cases},
$$

then the same kind of calculation shows that $Q(x_1, \ldots, x_n) = \mathbf{x}^T A\mathbf{x}$. Thus, every quadratic form has an associated matrix. Notice that this matrix is *symmetric*, meaning $A^T = A$. Conversely, any symmetric matrix leads to a quadratic form.

2 Putting a quadratic form into normal form

What happens to the matrix A of a quadratic form Q when we apply a change of variables? Suppose $R : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation represented as a matrix, such as a rotation about the origin. We'd like to know what the matrix of $Q \circ R = Q(R(x_1, \ldots, x_n))$ is, in terms of A and R. In the example above with $Q(x, y) = 3x^2 + 2xy + 3y^2$, R was the transformation

$$
R(u, v) = ((u + v)/\sqrt{2}, (u - v)/\sqrt{2}).
$$

Given that $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, we have

$$
Q(R\mathbf{x}) = (R\mathbf{x})^T A(R\mathbf{x}) = \mathbf{x}R^T AR\mathbf{x},
$$

using the fact that $(AB)^T = B^T A^T$ for any matrices A, B. That is:

Proposition. If a quadratic form Q has matrix A, and R is a linear transformation, then the quadratic form $Q \circ R$ has matrix $R^T AR$.

Notice that a quadratic form is in normal form exactly when its matrix is *diagonal*; for example, $ax^2 + by^2$ has matrix $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$. Therefore we can restate the problem of putting a quadratic form into normal form as: given a symmetric matrix A, find a linear transformation R with the property that $R^T AR$ is diagonal. This problem is dealt with by the following useful theorem:

Theorem (Spectral theorem)**.** Any real symmetric matrix A is diagonalizable: there is a matrix R such that $R^{-1}AR$ is diagonal. Moreover, we can take R to be an *orthogonal* matrix, meaning that $R^{-1} = R^T$.

Corollary. For any quadratic form Q, there's an orthogonal linear transformation R such that $Q \circ R$ is in normal form.

As $Q \circ R$ is simply Q in different coordinates, and R is invertible, the critical point at 0 for Q is of the same type as for $Q \circ R$. The geometric meaning of a linear transformation R being orthogonal is that it's a *rigid motion* of space, i.e. it doesn't distort distances: $|Rv - Rw| = |v - w|$. In 2 dimensions, the orthogonal matrices are all either rotations or reflections.

Example 1. Put the quadratic form $Q(x, y) = 3x^2 + 2xy + 3y^2$ in normal form.

Solution. We need to diagonalize the matrix $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$. So, compute its eigenvalues: the characteristic polynomial is

$$\det \begin{pmatrix} 3 - \lambda & 1 \\ 1 & 3 - \lambda \end{pmatrix} = (3 - \lambda)^2 - 1 = \lambda^2 - 6\lambda + 8 = (\lambda - 2)(\lambda - 4),$$

and the eigenvalues of A are its roots $\lambda = 2, 4$. This means there's some orthogonal matrix R such that

$$R^T AR = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix},$$

so $3x^2 + 2xy + 3y^2$ is a rotation/reflection of $2x^2 + 4y^2$. Finding the matrix R amounts to finding eigenvectors of A, but I won't do that here. □

Notice that the normal form isn't unique: $3x^2 + 2xy + 3y^2$ is also a rotation of $4x^2 + 2y^2$.

Problem 3. (1 pt). Find a normal form for $Q(x, y) = -2x^2 + 4xy - 5y^2$.

3 Critical points of functions

We've said that most of the time, a critical point p for a function $f(x, y)$ has the same type as the critical point 0 for the quadratic form $Q(x, y) = f_{xx}(p)h^2 + 2f_{xy}(p)hk + f_{yy}(p)k^2$. Notice that the associated matrix is

$$\begin{pmatrix} f_{xx}(p) & f_{xy}(p) \\ f_{yx}(p) & f_{yy}(p) \end{pmatrix}.$$

This is called the *Hessian* of f at p, and we'll write it H_pf. Notice that by equality of mixed partials, the Hessian is symmetric (we're assuming f has continuous second partials).

More generally, for a function $f(x_1, \ldots, x_n)$, the second-order terms in the Taylor polynomials for $f(x_1 + h_1 + \cdots + x_n + h_n)$, if $p = (x_1, \ldots, x_n)$, are

$$\left(h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n}\right)^2 \Big|_p f = \sum_{i=1}^{n} f_{x_i x_i}(p) h_i^2 + 2 \sum_{1 \le i < j \le n} f_{x_i x_j}(p) h_i h_j.$$

The matrix of this quadratic form, i.e. the Hessian of f at p, is simply the matrix of all second partials:

$$\begin{pmatrix} f_{x_1x_1}(p) & f_{x_2x_1}(p) & \cdots & f_{x_nx_1}(p) \\ f_{x_1x_2}(p) & f_{x_2x_2}(p) & \cdots & f_{x_nx_2}(p) \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_1x_n}(p) & f_{x_2x_n}(p) & \cdots & f_{x_nx_n}(p) \end{pmatrix}$$

Here's a precise statement classifying critical points, using the Hessian.

Theorem. Let $U \subseteq \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}$ have continuous partials of order 3. If p is a critical point of f and $\det H_pf \neq 0$, then the critical point p for f is of the same type as the critical point 0 for the quadratic form associated to H_pf.

Problem 2, and the section on normal forms of quadratic forms, tells us that if Q is a quadratic form with matrix A, then 0 is a local maximum if all eigenvalues of A are negative; a local minimum if all eigenvalues of A are positive; and neither (a saddle point) if some eigenvalues are negative and some are positive. So we can rephrase the theorem above in terms of eigenvalues:

Theorem. (Second derivative test). Let $U \subseteq \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}$ have continuous partials of order 3. If p is a critical point of f and $\det H_pf \neq 0$, then the critical point p is: a local maximum if the Hessian $H_p(f)$ has all nonpositive eigenvalues; a local minimum if it has all nonnegative eigenvalues; and neither if some eigenvalues are negative and some are positive.

Problem 4. (1 pt). The function $f(x,y) = x^3y + xy^3 - xy + x^2 - y^2$ has a critical point at $(0,0)$. Use the second derivative test to decide if it's a local minimum, a local maximum, or neither (or both??).

Problem 5. (2 pt). Show that when $\det H_p f = 0$, the second derivative test gives you no information, by finding two functions f, g of 2 variables such that $\det H_p f = \det H_q g = 0$, with p, q critical points of f, g respectively, but where p is a local maximum of f whereas q is neither a local maximum nor minimum of g.

Problem 6. (2 pt). Suppose p is a critical point of $f(x,y)$, and

$$H_p f = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

Show that if $ac - b^2 > 0$ and $a + c < 0$, then p is a local maximum of f; if $ac - b^2 > 0$ and $a + c > 0$, then p is a local minimum of f; if $ac - b^2 < 0$, then p is neither a local maximum nor minimum.

Problem 7. (3 pt). A function $T : \mathbb{R}^n \to \mathbb{R}^n$ with the property that $|T(v) - T(w)| = |v - w|$ for all v, w in $\mathbb{R}^n$ is called an *isometry*; these are transformations which don't distort distances. Rotations and reflections are examples, as we mentioned earlier (calling them "rigid motions"); another example is translation: $T(x,y) = (x + a, y + b)$ for some fixed a, b.

Note that any nontrivial translation isn't a linear transformation in the sense of linear algebra, since $T(0) = 0$ fails. However, show that if T is any isometry of $\mathbb{R}^n$ with the property that $T(0) = 0$, then T is a linear transformation, meaning $T(v + w) = T(v) + T(w)$ for any v, w in $\mathbb{R}^n$ and $T(cv) = cT(v)$ if c is a scalar. This is a large step in saying exactly what the isometries of $\mathbb{R}^n$ are.

[This is a little harder than the previous problems. But it's entirely elementary and can be done just by thinking about basic geometry and drawing pictures, although I'd be happy to see a slick proof that uses fancier machinery too.]