Quadratic Forms

Each of

$$x^2 - \sqrt{2}x + 7, \qquad 3x^{18} - \frac{2}{\pi}x^{11}, \qquad \text{and} \qquad 0$$

is a **polynomial** in the single variable x. But polynomials can involve more than one variable. For instance,

$$\sqrt{5}x_1^8 - \frac{1}{3}x_1x_2^4 + x_1^3x_2 \qquad \text{and} \qquad 3x_1^2 + 2x_1x_2 + 7x_2^2$$

are polynomials in the two variables x_1, x_2; products between powers of variables in terms are permissible, but all exponents in such powers must be nonnegative integers to fit the classification *polynomial*. The degrees of the terms of

$$\sqrt{5}x_1^8 - \frac{1}{3}x_1x_2^4 + x_1^3x_2$$

are 8, 5 and 4, respectively. When all terms in a polynomial are of the same degree k, we call that polynomial a k-form. Thus,

$$3x_1^2 + 2x_1x_2 + 7x_2^2$$

is a 2-form (also known as a **quadratic form**) in two variables, while the dot product of a constant vector $\mathbf{a}$ and a vector $\mathbf{x} \in \mathbb{R}^n$ of unknowns

$$\mathbf{a} \cdot \mathbf{x} \;=\; \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \;=\; a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

is a 1-form, or **linear form** in the n variables found in $\mathbf{x}$. The quadratic form in variables x_1, x_2

$$ax_1^2 + bx_1x_2 + cx_2^2 \;=\; \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \;=\; \langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle,$$

for

$$\mathbf{A} = \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix} \qquad \text{and} \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Similarly, a quadratic form in variables x_1, x_2, x_3 like

$$2x_1^2 - 3x_1x_2 - x_2^2 + 4x_1x_3 + 5x_2x_3 \qquad \text{can be written as} \qquad \langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle,$$

where

$$\mathbf{A} = \begin{bmatrix} 2 & -1.5 & 2 \\ -1.5 & -1 & 2.5 \\ 2 & 2.5 & 0 \end{bmatrix} \qquad \text{and} \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Note that the coefficient 5 of the x_2x_3 term could have been partitioned between entries a_{23} and a_{32} of $\mathbf{A}$ in any manner which sums to 5, but to put half as a_{23} and the rest as a_{32} (and similarly with other terms) turns $\mathbf{A}$ into a symmetric matrix. In fact, we can simply define a quadratic form to be any expression of the form

$$\langle \mathbf{x}, \mathbf{A}\mathbf{x} \rangle = \mathbf{x}^{\mathrm{T}} \mathbf{A} \mathbf{x} = \langle \mathbf{A}^{\mathrm{T}} \mathbf{x}, \mathbf{x} \rangle,$$

where $\mathbf{A}$ is symmetric.

Approximating Functions of Multiple Variables

Polynomials are easy to differentiate and evaluate, and we like to use them to approximate other functions. This is the content of Taylor's theorem, encountered in Calculus:

Theorem 4 (Taylor's Theorem for Real-Valued Functions): Suppose $f : (a-R, a+R) \to \mathbb{R}$, where I is the open interval $(a-R, a+R)$ centered at a of some positive radius $R > 0$. Suppose also that $k > 0$ be an integer such that $f^{(k)}$ is continuous on I and $f^{(k+1)}$ exists throughout I. Given any $x \in I$ there exists a number c between x and a such that

$$f(x) = \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!} (x-a)^j + \frac{f^{(k+1)}(c)}{\ell!} (x-a)^{k+1}.$$

The expression

$$T_k(x) = \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!} (x-a)^j$$

is a k^{th} degree polynomial known as the **k^{th} degree Taylor polynomial of f centered at a.** It's form arises from being the only polynomial of degree k which has an identical value as f at $x = a$, an identical 1^{st} derivative value as f at $x = a$, right up to an identical k^{th} derivative value as f at $x = a$. In particular, the 2^{nd} degree Taylor polynomial looks like

$$T_2(x) = f(a) + f'(a)(x-a) + (x-a)\frac{f''(a)}{2}(x-a).$$

Example 2:

Find the 2^{nd} degree Taylor polynomial of $\sin(x)$ at $(-\pi/4)$. The answer is

$$T_2(x) = \frac{\sqrt{2}}{4}\left(x + \frac{\pi}{4}\right)^2 + \frac{\sqrt{2}}{2}\left(x + \frac{\pi}{4}\right) - \frac{\sqrt{2}}{2}.$$

See the video at https://www.youtube.com/watch?v=44PeKBY_ySQ for details, if interested.

■

When $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ is a sufficiently smooth, real-valued function of n real variables $x_1, x_2, \ldots, x_n$, there is a version of Taylor's theorem which guides the approximation of f by polynomials in $x_1, x_2, \ldots, x_n$. I will not state that theorem here. But I will point out that if, as above, we focus on T_2, the Taylor polynomial in $x_1, \ldots, x_n$ centered at $\mathbf{x} = \mathbf{a}$ whose terms are of degree two or less, it has a particularly nice formula:

$$T_2(\mathbf{x}) \; = \; f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^{\mathrm{T}} H_f(\mathbf{a})(\mathbf{x} - \mathbf{a}),$$

where the **gradient** vector and **Hessian** matrix are

$$\nabla f(\mathbf{a}) \; = \; \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{a}) \\ \frac{\partial f}{\partial x_2}(\mathbf{a}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{a}) \end{bmatrix} \qquad \text{and} \qquad H_f(\mathbf{a}) \; = \; \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_1 \partial x_3}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2 \partial x_3}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_3 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_3 \partial x_2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_3^2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_3 \partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_n \partial x_3}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{a}) \end{bmatrix}.$$

Notice that, under smoothness conditions discussed in Calculus, cross-partial derivatives $\frac{\partial^2 f}{\partial x_i \partial x_j}$ and $\frac{\partial^2 f}{\partial x_j \partial x_i}$ with respect to the same two variables are equal, making the Hessian matrix symmetric. Hence, setting $\mathbf{h} = \mathbf{x} - \mathbf{a}$, the expressions

$$\nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) \; = \; \nabla f(\mathbf{a}) \cdot \mathbf{h} \qquad \text{and} \qquad \frac{1}{2}(\mathbf{x} - \mathbf{a})^{\mathrm{T}} H_f(\mathbf{a})(\mathbf{x} - \mathbf{a}) \; = \; \mathbf{h}^{\mathrm{T}} \left(\frac{1}{2} H_f(\mathbf{a}) \right) \mathbf{h}$$

are linear and quadratic forms in the variables of $\mathbf{h}$, respectively, so we have

$$T_2(\mathbf{a} + \mathbf{h}) \; = \; f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{h} + \frac{1}{2}\mathbf{h}^{\mathrm{T}} H_f(\mathbf{a})\mathbf{h} \; = \; f(\mathbf{a}) + \langle \nabla f(\mathbf{a}), \mathbf{h} \rangle + \left\langle \frac{1}{2} H_f(\mathbf{a})\mathbf{h}, \mathbf{h} \right\rangle \tag{3}$$

as an approximation to values of $f(\mathbf{a} + \mathbf{h})$ when $\|\mathbf{h}\|$ is small.

Example 3:

Find the gradient vector and Hessian matrix of

$$f(x, y, z) = x^3 z + y z^2$$

at the point $\mathbf{a} = (1, 2, 3)$.

The three 1^{st} partial derivatives of f are

$$\frac{\partial f}{\partial x} = 3x^2 z, \qquad \frac{\partial f}{\partial y} = z^2, \qquad \frac{\partial f}{\partial z} = x^3 + 2yz,$$

so

$$\nabla f(1,2,3) \;=\; \begin{bmatrix} 9 \\ 9 \\ 13 \end{bmatrix}.$$

As cross-partial derivatives are equal, we list only 6 different 2^{nd} partial derivatives:

$$\frac{\partial^2 f}{\partial x \partial y} = 0, \qquad \frac{\partial^2 f}{\partial x \partial z} = 3x^2, \qquad \frac{\partial^2 f}{\partial y \partial z} = 2z, \qquad \frac{\partial^2 f}{\partial x^2} = 6xz, \qquad \frac{\partial^2 f}{\partial y^2} = 0, \qquad \frac{\partial^2 f}{\partial z^2} = 2y.$$

Thus, the Hessian matrix generally uses formulas

$$\begin{bmatrix} 6xz & 0 & 3x^2 \\ 0 & 0 & 2z \\ 3x^2 & 2z & 2y \end{bmatrix} \qquad \text{and} \qquad H_f(1,2,3) \;=\; \begin{bmatrix} 18 & 0 & 3 \\ 0 & 0 & 6 \\ 3 & 6 & 4 \end{bmatrix}.$$

Now, the f in this example is already a polynomial, chosen to be so in order to make the calculation of derivatives simple. In practice, you would probably find a 2^{nd} degree polynomial approximation when f is *not* a polynomial. But to carry out the approximation of f near $\mathbf{a} = (1,2,3)$, we have

$$\begin{aligned} f(1+h_1, 2+h_2, 3+h_3) \;&\approx\; f(1,2,3) + \nabla f(1,2,3) \cdot \mathbf{h} + \frac{1}{2}\mathbf{h}^{\mathrm{T}} H_f(1,2,3)\mathbf{h} \\ &=\; 21 + \begin{bmatrix} 9 \\ 9 \\ 13 \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} + \begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix} \begin{bmatrix} 9 & 0 & 1.5 \\ 0 & 0 & 3 \\ 1.5 & 3 & 2 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} \\ &=\; 9h_1^2 + 2h_3^2 + 3h_1h_3 + 6h_2h_3 + 9h_1 + 9h_2 + 13h_3 + 21. \end{aligned}$$

Compare the function value at $(1.1, 1.95, 3.08)$

$$f(1.1, 1.95, 3.08) \;=\; (1.1)^3(3.08) + (1.95)(3.08)^2 \;\doteq\; 22.598,$$

with the estimate at $\mathbf{h} = (0.1, -0.05, 0.08)$

$$9(0.1)^2 + 2(0.08)^2 + 3(0.1)(0.08) + 6(-0.05)(0.08) + 9(0.1) + 9(-0.05) + 13(0.08) + 21 \;\doteq\; 22.593.$$

■