ADVANCED CALCULUS
SPRING 2019

LESSON 2: Optimization

This lesson considers the following topics.

1. Taylor polynomials.
2. Optimization: unconstrained extrema, constrained extrema, Lagrange multipliers.

It supplements Chapter 3 of the Marsden-Tromba text.

1 Introduction

Optimization is a matter of practical concern in any human endeavor, be it business, design, manufacturing, sport, or just living: how to maximize or minimize an outcome given a range of options from which to choose. Mathematically, the simplest version of optimization is determining the peaks and valleys (local maximum and minimum values) of a function in a given domain.

Once a point where a function may attain a locally optimal state is identified, further examination of the function in the vicinity of the point is required to determine whether optimality is in fact achieved. Since such an examination is typically restricted to a small neighborhood, it is convenient to forego a consideration of the full function (especially if it is complicated and expensive to compute) in favor of a reasonably accurate and simple approximation to the function near the candidate point. Polynomials provide a particularly useful class of approximating functions; they are easy to evaluate, differentiate and integrate. A systematic procedure for constructing polynomial approximations is based on Taylor's Theorem. We begin by recalling Taylor-polynomial approximations of functions $f : \mathcal{R} \to \mathcal{R}$, and then extend the procedure of constructing such approximations to multivariate functions $f : \mathcal{R}^n \to \mathcal{R}$.

2 Taylor polynomials for $f : \mathcal{R}^n \to \mathcal{R}$

2.1 $n = 1$

We begin our discussion with functions of a single variable, for which Taylor's theorem is introduced in Calculus I and is reviewed below.

Let $f(x)$ be a function in class C^{m+1}, *i.e.*, a function with continuous derivatives till the derivative of order $m+1$, in an open interval I. Let $x_0 \in I$. For $x \in I$ we can write

$$f(x) \quad = \quad f(x_0) + \int_{x_0}^{x} f'(t)\, dt.$$

An integration by parts leads to

$$\begin{aligned} f(x) &= f(x_0) + f'(t)(t-x)\Big|_{x_0}^{x} - \int_{x_0}^{x} (t-x) f''(t)\, dt \\ &= f(x_0) + (x-x_0)f'(x_0) - \int_{x_0}^{x} (t-x) f''(t)\, dt. \end{aligned}$$

Integrating by parts again we obtain

$$\begin{aligned} f(x) &= f(x_0) + (x-x_0)f'(x_0) - \left[f''(t)\frac{(t-x)^2}{2}\Big|_{x_0}^{x} - \int_{x_0}^{x} \frac{(t-x)^2}{2} f'''(t)\, dt \right] \\ &= f(x_0) + (x-x_0)f'(x_0) + \frac{(x-x_0)^2}{2} f''(x_0) + (-1)^2 \int_{x_0}^{x} \frac{(t-x)^2}{2} f'''(t)\, dt. \end{aligned}$$

Repeated integration by parts yields the result

$$\begin{aligned} f(x) &= f(x_0) + (x-x_0)f'(x_0) + \frac{(x-x_0)^2}{2!} f''(x_0) + \cdots + \frac{(x-x_0)^m}{m!} f^{(m)}(x_0) + R_m(x, x_0) \\ &= P_m(x) + R_m(x, x_0), \end{aligned} \tag{2.1}$$

where

$$P_m(x) = f(x_0) + (x-x_0)f'(x_0) + \frac{(x-x_0)^2}{2!} f''(x_0) + \cdots + \frac{(x-x_0)^m}{m!} f^{(m)}(x_0) \tag{2.2}$$

is called the *Taylor polynomial of degree* m *for* $f(x)$, *centered at* x_0, and

$$R_m(x, x_0) = (-1)^m \int_{x_0}^{x} \frac{(t-x)^m}{m!} f^{(m+1)}(t)\, dt \tag{2.3}$$

is known as the *remainder.*

Remarks.

- The remainder represents the error if we approximate $f(x)$ by the polynomial $P_m(x)$. For the approximation to be good for values of x near x_0 we must estimate the size of the remainder to make sure that it is small and hence safe to neglect. The expression (2.3) is not particularly suitable for estimation but can be rendered more convenient by applying the *Second Mean-Value Theorem* for integrals, stated below.

  Theorem: Consider
  $$I = \int_a^b F(t)G(t)\,dt,$$
  where F and G are continuous. If $F(t)$ does not change sign in $[a,b]$, then there exists a $t^* \in [a,b]$ such that
  $$I = G(t^*)\int_a^b F(t)\,dt.$$

  Since the function $(t-x)^m$ does not change sign for t between x_0 and x, application of the Second Mean-Value theorem to the integral in (2.3) results in the alternate expression

  $$\begin{aligned} R_m(x,x_0) &= f^{(m+1)}(t^*)(-1)^m \frac{(t-x)^{m+1}}{(m+1)!}\Big|_{x_0}^{x} \\ &= \frac{(x-x_0)^{m+1}}{(m+1)!} f^{(m+1)}(t^*). \end{aligned} \tag{2.4}$$

  Thus (2.1) can be rewritten as

  $$f(x) = P_m(x) + R_m(x,x_0) = P_m(x) + \frac{(x-x_0)^{m+1}}{(m+1)!} f^{(m+1)}(t^*), \tag{2.5}$$

  where t^* is not explicitly known but lies between x_0 and x. This expression shows that

  $$\lim_{x\to x_0} \frac{f(x)-P_m(x)}{(x-x_0)^m} = \lim_{x\to x_0} \frac{R_m(x,x_0)}{(x-x_0)^m} = 0, \tag{2.6}$$

  that is, the error R_m goes to zero at a rate faster than $(x-x_0)^m$.

- The Taylor polynomial of degree one is

  $$P_1(x) = f(x_0) + (x-x_0)f'(x_0). \tag{2.7}$$

  This is the familiar linear approximation or the tangent-line approximation studied earlier. The result (2.6) for $m = 1$ confirms what we know from previous work, namely, that the error of the linear approximation goes to zero faster than $x - x_0$ as $x \to x_0$.

  The linear approximation allows us to write

  $$f(x) - f(x_0) \approx (x-x_0)f'(x_0).$$

  On setting $x - x_0 = \Delta x$ and $f(x) - f(x_0) = \Delta f$ we get

  $$\Delta f \approx f'(x_0)\Delta x.$$

  The quantity on the RHS is the increment along the tangent line corresponding the increment Δx in x, and is known as the *differential* of f, written as df. Therefore

  $$\Delta f \approx df = f'(x_0)\Delta x.$$

For the special case $f(x) = x$ the above reduces to $dx = 1\ \Delta x = \Delta x$, thus allowing us to write the differential of f at $x = x_0$ as

$$df = f'(x_0)dx.$$

The differential is a practically useful quantity for a quick estimate of the increment in f as x is incremented by a small amount. For example, let $f(x) = \tan x$ and let x be incremented from its base value $x_0 = \pi/4$ by $dx = 0.01$. Then the corresponding increase in the value of $\tan x$ from the base value $\tan \pi/4 = 1$ is given approximately by $df = f'(x_0)dx = \sec^2(\pi/4)(0.01) = 2(0.01) = 0.02$.

- At $x = x_0$ the Taylor polynomial $P_m(x)$ has the same value as $f(x)$, and its derivatives to order m have the same values as the corresponding derivatives of $f(x)$, *i.e.*,

$$\frac{d^k}{dx^k} f(x_0) = \frac{d^k}{dx^k} P_m(x_0), \quad k = 0, 1, 2, \cdots, m.$$

- Consider $f(x) = e^x \cos 2x$. The first two Taylor polynomials centered at $x = 0$ are easily found to be $P_1(x) = 1 + x$, $P_2(x) = 1 + x - 3x^2/2$ and are plotted in Figure 1.

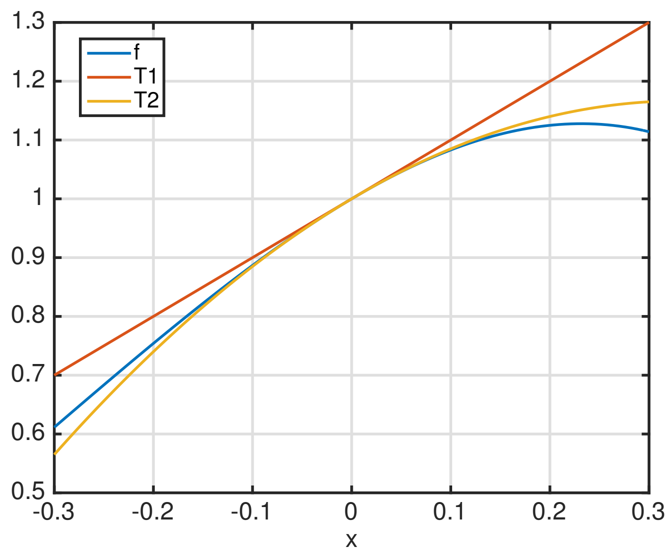

Figure 1: $f(x) = e^x \cos 2x$ and its Taylor polynomials. Note increasing accuracy with degree of polynomial.

2.2 $n > 1$

Taylor polynomials can also be obtained for multivariate functions. We consider the twice continuously differentiable function $f : \mathcal{R}^n \to \mathcal{R}$, and seek its Taylor polynomials centered at $\boldsymbol{x}_0$. Attention is restricted to $P_1(\boldsymbol{x})$ and $P_2(\boldsymbol{x})$, the first and second degree Taylor polyomials, as these are the ones of most relevance in optimization problems. The analysis extends the $n = 1$ case in a conceptually straightforward manner.

2.2.1 $P_1(\boldsymbol{x})$

The Taylor polynomial $P_1(\boldsymbol{x})$ has the same value and the same derivative as $f(\boldsymbol{x})$ at $\boldsymbol{x}_0$. It is the linear approximation studied in the previous lesson and is given by

$$P_1(\boldsymbol{x}) = f(\boldsymbol{x}_0) + Df(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0). \tag{2.8}$$

(Note the similarity with the $n = 1$ case in (2.7).) Here we have elected to use the matrix notation. Thus $Df(\boldsymbol{x}_0)$ is the $1 \times n$ row vector whose components are the first partial derivatives of f evaluated at $\boldsymbol{x}_0$, *i.e.*,

$$Df(\boldsymbol{x}_0) = \left[\frac{\partial f}{\partial x_1}(\boldsymbol{x}_0),\ \frac{\partial f}{\partial x_2}(\boldsymbol{x}_0),\ \cdots,\ \frac{\partial f}{\partial x_n}(\boldsymbol{x}_0)\right], \tag{2.9}$$

and $\boldsymbol{x} - \boldsymbol{x}_0$ is the $n \times 1$ column vector whose components are the increments in the coordinates from $\boldsymbol{x}_0$ to $\boldsymbol{x}_1$, *i.e.*,

$$\boldsymbol{x} - \boldsymbol{x}_0 = \begin{bmatrix} x_1 - x_{1_0} \\ x_2 - x_{2_0} \\ \cdots\cdots \\ x_n - x_{n_0} \end{bmatrix}.$$

After the matrix multiplication is carried out the polynomial can be written in an alternate form as

$$P_1(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\boldsymbol{x}_0)\,(x_k - x_{k_0})\,. \tag{2.10}$$

Remarks.

- For $n = 2$, $\boldsymbol{x} = (x, y)$ and the linear approximation is

$$P_1(x, y) = f(x_0, y_0) + (x - x_0) f_x(x_0, y_0) + (y - y_0) f_y(x_0, y_0). \tag{2.11}$$

- The goodness of the linear approximation to $f(\boldsymbol{x})$ is quantified by

$$\lim_{\|\boldsymbol{x}-\boldsymbol{x}_0\| \to 0} \frac{f(\boldsymbol{x}) - P_1(\boldsymbol{x})}{\|\boldsymbol{x} - \boldsymbol{x}_0\|} = 0.$$

2.2.2 $P_2(\boldsymbol{x})$

The Taylor polynomial $P_2(\boldsymbol{x})$ has the same value and the same first and second derivatives as $f(\boldsymbol{x})$ at $\boldsymbol{x}_0$. Let us first examine it for $n = 2$, where $\boldsymbol{x} = (x, y)$. Then

$$\begin{aligned} P_2(x, y) \quad = \quad & f(x_0, y_0) + (x - x_0) f_x(x_0, y_0) + (y - y_0) f_y(x_0, y_0) \\ & + \frac{1}{2}\left\{(x - x_0)^2 f_{xx}(x_0, y_0) + 2(x - x_0)(y - y_0) f_{xy}(x_0, y_0) + (y - y_0)^2 f_{yy}(x_0, y_0)\right\}. \end{aligned} \tag{2.12}$$

It is convenient to write it in a matrix form as

$$\begin{aligned} P_2(x, y) \quad = \quad & f(x_0, y_0) + \left[f_x(x_0, y_0),\ f_y(x_0, y_0)\right] \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \\ & + \frac{1}{2}\left[x - x_0,\ y - y_0\right] \begin{bmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{yx}(x_0, y_0) & f_{yy}(x_0, y_0) \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}. \end{aligned} \tag{2.13}$$

Remark. The matrix of second partial derivatives appearing in (2.13) is known as the Hessian and is denoted by $Hf(x_0, y_0)$. Since f is assumed to be twice continuously differentiable, the mixed partial derivatives are identical ($f_{xy} = f_{yx}$) so that the Hessian is a symmetric matrix.

Example 2.1. Consider

$$f(x, y) = e^{x+2y}$$

and let us compute $P_2(x, y)$ centered at $P(0, 0)$. We have

$$f_x = e^{x+2y},\ f_y = 2e^{x+2y},\ f_{xx} = e^{x+2y},\ f_{xy} = 2e^{x+2y},\ f_{yy} = 4e^{x+2y}.$$

At P,

$$f = 1,\ f_x = 1,\ f_y = 2,\ f_{xx} = 1,\ f_{xy} = 2,\ f_{yy} = 4.$$

Therefore,

$$P_2(x, y) = 1 + x + 2y + \frac{1}{2}(x^2 + 4xy + 4y^2) = 1 + x + 2y + \frac{1}{2}x^2 + 2xy + 2y^2.$$

For general n, $P_2(\boldsymbol{x})$ can be written in matrix form as

$$P_2(\boldsymbol{x}) = f(\boldsymbol{x}_0) + Df(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^T Hf(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0). \tag{2.14}$$

Here, $Df(\boldsymbol{x}_0)$ is the $1 \times n$ derivative matrix (or gradient) defined in (2.9), $(\boldsymbol{x} - \boldsymbol{x}_0)^T$ the $1 \times n$ row vector, $(\boldsymbol{x} - \boldsymbol{x}_0)$ the $n \times 1$ column vector and $Hf(\boldsymbol{x}_0)$ the $n \times n$ Hessian matrix evaluated at $\boldsymbol{x}_0$. The Hessian has the (i, j) component

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}_0).$$

The equality of mixed second partial derivatives again renders the Hessian to be a symmetric matrix. In an expanded form,

$$Hf(\boldsymbol{x}_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

Equation (2.14) is the generalization of (2.13) to arbitrary n.

Remarks.

- The goodness of $P_2(\boldsymbol{x})$ as an approximation to $f(\boldsymbol{x})$ is quantified by

$$\lim_{\|\boldsymbol{x} - \boldsymbol{x}_0\|} \frac{f(\boldsymbol{x}) - P_2(\boldsymbol{x})}{\|\boldsymbol{x} - \boldsymbol{x}_0\|^2} = 0.$$

- Equation (2.13) can be written in the following alternate form,

$$P_2(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \sum_{i=1}^{n}(x_i - x_{i_0})\frac{\partial f}{\partial x_i}(\boldsymbol{x}_0) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_{i_0})(x_j - x_{j_0})\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}_0). \tag{2.15}$$

Example 2.2. Consider

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

and let us compute $P_2(x, y, z)$ centered at $P(0, 0, 0)$. We have

$$f_x = e^{x+2y+3z},\ f_y = 2e^{x+2y+3z},\ f_z = 3e^{x+2y+3z},$$

and

$$f_{xx} = e^{x+2y+3z},\ f_{xy} = 2e^{x+2y+3z},\ f_{xz} = 3e^{x+2y+3z},$$
$$y_{yx} = 2e^{x+2y+3z},\ f_{yy} = 4e^{x+2y+3z},\ f_{yz} = 6e^{x+2y+3z},$$
$$f_{zx} = 3e^{x+2y+3z},\ f_{zy} = 6e^{x+2y+3z},\ f_{zz} = 9e^{x+2y+3z}.$$

Evaluation of the above quantities at $P(0,0,0)$ yields

$$f=1,\ f_x=1,\ f_y=2,\ f_z=3,$$

and

$$\begin{aligned} f_{xx}=1,\ f_{xy}=2,\ f_{xz}=3,\\ y_{yx}=2,\ f_{yy}=4,\ f_{yz}=6,\\ f_{zx}=3,\ f_{zy}=6,\ f_{zz}=9. \end{aligned}$$

Then

$$Df(0,0,0)=[1,\ 2,\ 3], \quad Hf(0,0,0)=\begin{bmatrix} 1 & 2 & 3\\ 2 & 4 & 6\\ 3 & 6 & 9 \end{bmatrix}$$

. We can now write

$$\begin{aligned} P_2(x,y,z) &= f(0,0,0)+Df(0,0,0)\begin{bmatrix} x\\ y\\ z \end{bmatrix}+\begin{bmatrix} x & y & z \end{bmatrix}Hf(0,0,0)\begin{bmatrix} x\\ y\\ z \end{bmatrix}\\ &= 1+x+2y+3z+x^2+4y^2+9z^2+4xy+12yz+6zx. \end{aligned}$$

3 Optimization

With the above background at hand, we are ready to examine multivariate functions to determine where they take local extreme values, or extrema. (*Extrema* is a collective term for maxima and minima). We shall see that the derivative Df (or equivalently, the Taylor polynomial $P_1(\boldsymbol{x})$) identifies possible locations of local extrema and the Hessian Hf (or equivalently, the Taylor polynomial $P_2(\boldsymbol{x})$) classifies the extrema.

3.1 Critical points

Let $\mathcal{D}\subseteq\mathcal{R}^n$ be an open set. A function $f:\mathcal{D}\subseteq\mathcal{R}^n\to\mathcal{R}$ has a local minimum at $\boldsymbol{x}_0$ if there exists a neighborhood U of $\boldsymbol{x}_0$ such that $f(\boldsymbol{x})-f(\boldsymbol{x}_0)\geq 0$ for all $\boldsymbol{x}\in U$. The local minimum is strict if the inequality is strict, *i.e.*, if $f(\boldsymbol{x})-f(\boldsymbol{x}_0)>0$ for all $\boldsymbol{x}\in U$. An analogous definition applies to local maximum.

For $n=2$ the graph of f is a surface. Then the local maximum corresponds to a peak and a local minimum to a pit in the surface.

We shall focus on differentiable functions. Then a necessary condition for $f(\boldsymbol{x})$ to have an extremum at $\boldsymbol{x}_0$ is $Df(\boldsymbol{x}_0)=0$. This result is derived most easily by considering points $\boldsymbol{x}\in U$ of the form $\boldsymbol{x}=\boldsymbol{x}_0+d\boldsymbol{x}=\boldsymbol{x}_0+t\boldsymbol{h}$ where $\boldsymbol{h}$ is an arbitrary vector. Then $f(\boldsymbol{x})$ has a local minimum at $\boldsymbol{x}_0$ if the single-variable function $g(t)=f(\boldsymbol{x}_0+t\boldsymbol{h})$ has a local minimum at $t=0$. From single-variable calculus the necessary condition is $g'(t)=0$ at $t=0$. According to the chain rule,

$$\begin{aligned} g'(0)=\frac{d}{dt}f(\boldsymbol{x}_0+t\boldsymbol{h})\Big|_{t=0} &= Df(\boldsymbol{x}_0)\boldsymbol{h}\\ &= \left[\frac{\partial f}{\partial x_1}(\boldsymbol{x}_0),\ \frac{\partial f}{\partial x_2}(\boldsymbol{x}_0),\ \cdots,\frac{\partial f}{\partial x_n}(\boldsymbol{x}_0)\right]\begin{bmatrix} h_1\\ h_2\\ \cdots\\ h_n \end{bmatrix}\\ &= \frac{\partial f}{\partial x_1}(\boldsymbol{x}_0)h_1+\frac{\partial f}{\partial x_2}(\boldsymbol{x}_0)h_2+\cdots+\frac{\partial f}{\partial x_n}(\boldsymbol{x}_0)h_n\\ &= 0. \end{aligned}$$

Since the h_i are arbitrary, $Df(\boldsymbol{x}_0)\boldsymbol{h}=0$ implies that

$$\frac{\partial f}{\partial x_k}(\boldsymbol{x}_0)=0, \quad k=1,2,\cdots,n,$$

or,

$$Df(\boldsymbol{x}_0) = 0.$$

The points $\boldsymbol{x}_0$ satisfying the above equation are candidate points for extrema and are known as *critical points*.

3.2 Classification

Once a critical point $\boldsymbol{x}_0$ of $f(\boldsymbol{x})$ has been located, the next step is to examine the behavior of $f(\boldsymbol{x})$ in a neighborhood of $\boldsymbol{x}_0$ to determine whether the function attains a local minimum, a local maximum or neither at the critical point. This is done most economically by approximating $f(\boldsymbol{x})$ by $P_2(\boldsymbol{x})$, its Taylor polynomial of degree 2 centered at $\boldsymbol{x}_0$.

3.2.1 $n = 1$

For this familiar case we have

$$f(x) \approx P_2(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2}(x - x_0)^2 f''(x_0).$$

Since $f'(x_0) = 0$,

$$f(x) - f(x_0) \approx P_2(x) - f(x_0) = \frac{1}{2}(x - x_0)^2 f''(x_0).$$

If we set $x = x_0 + h$ then the above reduces to

$$f(x_0 + h) - f(x_0) \approx \frac{1}{2}h^2 f''(x_0).$$

A local minimum at x_0 is achieved if $f(x_0 + h) - f(x_0) \geq 0$, which requires $f''(x_0) > 0$. Similarly, a local maximum at x_0 is achieved if $f''(x_0) < 0$. If $f''(x_0) = 0$, then the analysis is inconclusive at this stage and one must go on to examine the next Taylor polynomial, $P_3(x)$, which provides the approximation

$$f(x_0 + h) - f(x_0) \approx \frac{1}{3!}h^3 f'''(x_0).$$

For $f'''(x_0) \neq 0$, $f(x_0 + h) - f(x_0)$ changes sign with h, and the critical point is neither a local maximum nor a local minimum, but a point of inflection. If $f'''(x_0)$ also vanishes then one must go on to $P_4(x)$, and proceed in a by-now-obvious manner.

3.2.2 $n = 2$

The critical point (x_0, y_0) is classified by examining the approximation

$$\begin{aligned} f(x_0 + h_1, y_0 + h_2) - f(x_0, y_0) &\approx P_2(x_0 + h_1, y_0 + h_2) - f(x_0, y_0) \\ &= \frac{1}{2}[h_1,\ h_2] \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \\ &= \frac{1}{2}Q(h_1, h_2) \equiv \frac{1}{2}[H_{11}h_1^2 + 2H_{12}h_1h_2 + H_{22}h_2^2]. \end{aligned}$$

Here we have employed the expression (2.13) for P_2, where the entries $H_{ij}(x_0, y_0)$ of the (symmetric) Hessian matrix are identified. Note that no linear terms appear, as they vanish at the critical point. The quantity $Q(h_1, h_2)$ appearing on the RHS above is a quadratic expression in h_1, h_2 and is known as the quadratic form associated with the Hessian matrix. Being an approximation to the increment $f(x_0 + h_1, y_0 + h_2) - f(x_0, y_0)$, the quadratic form plays the same role for classification of the critical point as did the second derivative $f''(x_0)$ for $n = 1$; it is required to be positive definite for the critical point to be a local minimum and negative definite for the critical point to be a local maximum.

The quadratic form is said to be *positive definite* if $Q(h_1, h_2) > 0$ and *negative definite* if $Q(h_1, h_2) < 0$ for arbitrary, but not all zero, h_1 and h_2. Note that by completing squares Q can be manipulated into the expression

$$\begin{aligned} Q &= H_{11}\left[h_1^2 + 2\frac{H_{12}}{H_{11}}h_1h_2 + \left(\frac{H_{12}}{H_{11}}\right)^2 h_2^2\right] + \left[H_{22} - \frac{H_{12}^2}{H_{11}}\right] h_2^2 \\ &= H_{11}\left[h_1 + \frac{H_{12}}{H_{11}}h_2\right]^2 + \frac{1}{H_{11}}[H_{11}H_{22} - H_{12}^2]h_2^2 \end{aligned}$$

Positive-definiteness is guaranteed if $H_{11} > 0$ and $D \equiv H_{11}H_{22} - H_{12}^2 > 0$, and negative-definiteness if $H_{11} < 0$ and $D > 0$.

When $D < 0$ the quadratic form is *indefinite*, *i.e.*, its sign can change depending upon the choices for h_1 and h_2. In this case the critical point is neither a maximum nor a minimum but a saddle. The case $D = 0$ is inconclusive (much like the case $f''(x_0) = 0$ for $n = 1$) and then one either needs to consider the polynomial P_3 or examine the function f directly in the vicinity of the critical point.