Optimization Notes

Robert Carlson
Department of Mathematics
University of Colorado at Colorado Springs
rcarlson@uccs.edu

Contents

Chapter 1

Introduction

We begin with some examples of optimization problems.

1) Astronaut Armstrong throws a baseball up from the surface of the moon with an initial vertical velocity component of 85 mph. What is the maximum height reached by the ball. Repeat experiment 1 on earth.

2) We want to predict weight W as a function of height H based on a set of data $\{(H_i, W_i), i = 1 \ldots, N\}$. Find the line $W = aH + b$ so that the sum of the squared errors

$$E(a, b) = \sum_i [(aH_i + b) - W_i]^2$$

is minimized.

3) Suppose A is an $N \times N$ real symmetric matrix, $X \in \mathbb{R}^N$, and $\|X\| = 1$. Minimize $\langle X, AX \rangle$.

4) (Linear program) You have a 1000 acres of farmland, and your potential crops are rice, corn, and soybeans. Each crop i generates profit P_i per acre, requires total labor L_i per acre, water W_i per acre, and fertilizer F_i per acre. Total labor, water, and fertilizer are constrained. Determine the acreage in each crop which maximizes profit subject to the constraints.

5) (Combinatorial optimization) Find the cheapest air route from Bangalore, India to Huntsville, Alabama.

6) (Calculus of variations) Find the shortest path along the ground from Colorado Springs to Grand Junction.

7) (Bioinformatics) Given a set of mitochondrial DNA records from a human population sample, find the evolutionary tree that explains the data and has the maximum probability.

Problems 1-4 illustrate types of problems we will consider in this course. The function to be optimized is a real valued function of N real variables. The domain of the function is typically all of $\mathbb{R}^N$, or an open subset of $\mathbb{R}^N$ together with its boundary.

In problem 5 we cannot vary the airport positions or the costs, so it is not obvious that calculus can help. In problem 6 the variables are paths along the ground, which come from an infinite dimensional space. In problem 7 we have another combinatorial problem, this time finding an optimal tree.

Chapter 2

Calculus

2.1 One variable

We begin with a review of some ideas from Calculus 1.

An interval (open, closed, etc.) will be denoted by ι. Suppose $f(x)$ is a real-valued function on ι, written $f : \iota \to \mathbb{R}$.

Recall the following terms: global minimizer, strict global minimizer, local minimizer, strict local minimizer, critical point, extreme point or extremizer. Saying that x^* is a local minimizers means there is a $\delta > 0$ such that $f(x^*) \leq f(x)$ for all $x \in \iota \cap (x^* - \delta, x^* + \delta)$. Saying that the local minimizer is strict means $f(x) < f(x^*)$ for these x if $x \neq x^*$.

Our text defines a critical point as a point x^* such that $f'(x^*)$ exists and $f(x^*) = 0$. Note that other authors might toss in points where $f'(x)$ does not exist.

Theorem 2.1.1. *Suppose f is differentiable on ι. If x^* is a local minimizer or maximizer, then either x^* is an endpoint or $f'(x^*) = 0$.*

Proof. If x^* is not an endpoint, compute

$$f'(x^*) = \lim_{h \to 0^+} \frac{f(x^* + h) - f(x^*)}{h} \geq 0,$$

$$f'(x^*) = \lim_{h \to 0^-} \frac{f(x^* + h) - f(x^*)}{h} \leq 0.$$

Since these agree, $f'(x^*) = 0$. $\square$

In Calculus one often sees the baseball problem. Suppose

$$h(t) = \frac{1}{2}at^2 + v_0 t + h_0.$$

The ball reaches its maximum height when

$$h'(t) = at + v_0 = 0, \quad t = -\frac{v_0}{a}.$$

The next two results often help.

Theorem 2.1.2. *(Rolle's Theorem) Suppose f is differentiable on (α, β), continuous on $[\alpha, \beta]$, and $f(\alpha) = f(\beta) = 0$. Then there is some $c \in (\alpha, \beta)$ such that $f'(c) = 0$.*

Proof. Either f is constant or there is an interior max or min. □

Theorem 2.1.3. *(Mean Value Theorem) Suppose f is differentiable on (α, β), continuous on $[\alpha, \beta]$. Then there is some $c \in (\alpha, \beta)$ such that $f'(c) = \frac{f(\beta)-f(\alpha)}{\beta-\alpha}$.*

Proof. Apply Rolle's Theorem to

$$g(t) = f(t) - f(\alpha) - \frac{f(\beta) - f(\alpha)}{\beta - \alpha}(t - \alpha).$$

□

The following first order version of Taylor's Theorem is also very helpful. Our hypotheses are stricter than the text's to make the proof simpler.

Theorem 2.1.4. *(Taylor's Theorem) Suppose f'' is continuous on $[\alpha, \beta]$. If $x \neq x^*$ are in $[\alpha, \beta]$, then there is a z strictly between x and x^* such that*

$$f(x) = f(x^*) + f'(x^*)(x - x^*) + \frac{f''(z)}{2}(x - x^*)^2.$$

Proof. By the Fundamental Theorem of Calculus,

$$f(x) - f(x^*) = \int_{x^*}^{x} f'(t)\ dt,$$

or

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

Integration by parts gives

$$f(x) = f(x^*) + \int_{x^*}^{x} (t-x)' f'(t)\, dt$$

$$= f(x^*) + (t-x)f'(t)\Big|_{x^*}^{x} - \int_{x^*}^{x} (t-x) f''(t)\, dt$$

$$= f(x^*) + (x-x^*)f'(x^*) + \int_{x^*}^{x} (x-t) f''(t)\, dt$$

Suppose $x > x^*$ so $x - t \geq 0$. We have

$$f(x) - f(x^*) - (x-x^*)f'(x^*) = \int_{x^*}^{x} (x-t) f''(t)\, dt.$$

Let z_{max} and z_{min} extremize $f''(z)$ for $x^* \leq z \leq x$. Then

$$\int_{x^*}^{x} (x-t) f''(z_{min})\, dt \leq \int_{x^*}^{x} (x-t) f''(t)\, dt \leq \int_{x^*}^{x} (x-t) f''(z_{max})\, dt. \quad (2.1.1)$$

The function

$$g(z) = \int_{x^*}^{x} (x-t) f''(z)\, dt$$

is continuous on $[x^*, x]$ so by the Intermediate Value Theorem there is a z such that

$$\int_{x^*}^{x} (x-t) f''(t)\, dt = \int_{x^*}^{x} (x-t) f''(z)\, dt = -f''(z)\frac{(x-t)^2}{2}\Big|_{x^*}^{x} = f''(z)\frac{(x^*-x)^2}{2}.$$

The only time we have equality in (2.1.1) is when $z_{min} = z_{max}$, that is f'' is constant. In every case then we can assume $x^* < z < x$. The case $x < x^*$ is similar.

□

Here is a proof of the version in the text.

Theorem 2.1.5. *(Taylor's Theorem) Suppose f'' exists on $[\alpha, \beta]$. If $x \neq x^*$ are in $[\alpha, \beta]$, then there is a z strictly between x and x^* such that*

$$f(x) = f(x^*) + f'(x^*)(x-x^*) + \frac{f''(z)}{2}(x-x^*)^2.$$

Proof. Construct the auxiliary function

$$g(t) = f(t) - f(x^*) - f'(x^*)(t - x^*) - k(t - x^*)^2,$$

where k is chosen so $g(x) = 0$. Since $g(x^*)$ is also 0, there is a point c_1 strictly between x^* and x with $g'(c_1) = 0$. But we also have $g'(x^*) = 0$. Applying Rolle's Theorem to $g'(t)$ we find a z strictly between x^* and c_1 with $g''(z) = 0$. Then

$$0 = g''(z) = f''(z) - 2k,$$

which is the desired result. □

Our treatment of single variable calculus results wraps up with the next theorem.

Theorem 2.1.6. *Suppose f, f', f'' are continuous on an interval ι and $f'(x^*) = 0$.*

(a) If $f''(x) \geq 0$ for all $x \in \iota$, then x^* is a global minimizer of f on ι.

(b) If $f''(x) > 0$ for all $x \in \iota$, then x^* is a strict global minimizer of f on ι.

(c) If $f''(x^*) \geq 0$, then x^* is a strict local minimizer of f.

Proof. Since $f'(x^*) = 0$ we have

$$f(x) - f(x^*) = \frac{f''(z)}{2}(x - x^*)^2, \quad z \neq x, x^*.$$

For (c) note that $f''(x) > 0$ in some interval ι containing x^*. □

Here are a few simple examples.

(i) The first is

$$f(x) = \exp(x^2), \quad f'(x) = 2x\exp(x^2), \quad f''(x) = 2\exp(x^2) + 4x^2\exp(x^2),$$

so $x = 0$ is a strict global minimizer.

(ii) Our second is

$$g(x) = \sin(x), \quad g'(x) = \cos(x), \quad g''(x) = -\sin(x).$$

Critical points are at

$$x = \frac{\pi}{2} + n\pi.$$

(iii) For the third examples, compare

$$f(x) = x^4,$$

$$g(x) = \left\{ \begin{array}{cc} x^4, & x \leq 0, \\ 0, & 0 \leq x \leq 1, \\ (x-1)^4, & 1 \leq x. \end{array} \right\}$$

Note that $g''(x) \geq 0$, but there is no strict minimizer.

2.2 Several variables

The Euclidean space $\mathbb{R}^N$ is the set of ordered N-tuples of real numbers, denoted

$$X = \begin{pmatrix} x_1, \\ x_2, \\ ,\vdots \\ x_N \end{pmatrix}.$$

For typographical convenience we adopt the Matlab notation, writing a column vector as

$$X = [x_1, x_2, \dots, x_N].$$

We have the usual arithmetic and other features:

$$X \bullet Y = \sum_{k=1}^{N} x_k y_k, \quad \|X\| = [X \bullet X]^{1/2}.$$

The open ball of radius r centered at X^* is

$$B(X^*, r) = \{X, \|X - X^*\| < r\}.$$

We will commonly use D for the domain of functions. Recall that $X \in D$ in an *interior point* if there is an r_X such that $B(X, r_X) \subset D$ A set D is *open* if every point of D is an interior point. Using $B(X^*, r)$ we can define local minimizers, etc for functions $f : D \to \mathbb{R}$.

Here is more vocabulary from Calculus.

Suppose we have a sequence (list) $\{X_k\}$ from $\mathbb{R}^N$. Say that the sequence has *limit* X^* if for every $\epsilon > 0$ there is a K such that

$$\|X_k - X^*\| < \epsilon, \quad k \geq K.$$

Say that a set $\subset \mathbb{R}^N$ is *closed* if $X^* \in F$ for every point $X^* \in \mathbb{R}^N$ which is the limit of a sequence from F. This is equivalent to saying the complement of F is open. As examples, note that

$$\overline{B(X^*, r)} = \{X, \|X - X^*\| \leq r\}$$

is closed.

A function $f : D \subset \mathbb{R}^N \to \mathbb{R}$ is *continuous* if

$$\lim_{k\to\infty} f(X_k) = f(X^*),$$

whenever $X_k \in D$, $X_k \to X^*$, $X^* \in D$. Thus $\sin(1/x)$ is continuous on $(0, 1)$, but has no continuous extension to $[0, 1]$.

We say D is *bounded* if there is some number M such that

$$\|X\| \leq M, \quad X \in D.$$

Here is a simple example of minimization in several variables that helps set the direction to follow. Let

$$f(X) = a_1 x_1^2 + \dots a_N x_N^2 = \sum_{n=1}^{N} a_n x_n^2, \quad a_n > 0. \tag{2.2.1}$$

Say that X^* is a critical point of $f : D \to \mathbb{R}$ if $\partial f/\partial x_k(X^*) = 0$.

Theorem 2.2.1. *Suppose all first partials of $f : D \to \mathbb{R}$ exist. If X^* is an interior point of D that is a local minimizer of f, then X^* is a critical point.*

Proof. For each $n = 1, \dots, N$ the function

$$g_n(t) = f(x_1^*, \dots, x_{n-1}^*, t, x_{n+1}^*, \dots, x_N) : \mathbb{R}^1 \to \mathbb{R}^1$$

has an interior local minimizer at $t = x_n^*$. Thus

$$\frac{dg_n}{dt}(x_n^*) = \frac{\partial f}{\partial x_n}(X^*) = 0.$$

□

Returning to the example (2.2.1), the only critical point is at $X = 0$ since

$$\nabla f = [2a_1x_1, \dots, 2a_Nx_N].$$

of course $f(X) \ge 0$ for all $X \in \mathbb{R}^N$, so this must be a global minimum. It is also easy to find critical points for an general example like

$$g(x_1, x_2, x_3) = a_1x_1^2 + a_2x_1x_2 + a_3x_2^2 + a_4x_1x_3 + a_5x_2x_3 + a_6x_3^2, \quad a_n \in \mathbb{R}.$$

but determining if the critical points are maxima, minima, or neither is more challenging. To obtain a version of the second derivative test in several variables, we again need Taylor's Theorem.

Here are some examples to consider as we develop some techniques.

1. $f_1(x_1, x_2) = x_1^2 + 8x_1x_2 + x_2^2$

2 $f_2(x_1, x_2) = x_1^2 - 2x_1x_2 + 4x_2^2$

3. $f_3(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - x_1x_2 + x_2x_3 - x_1x_3$

4. $f_4(x, y, z) = e^{x-y} + e^{y-x} + e^{x^2} + z^2.$

Taylor's Theorem for a real valued function $f : \mathbb{R}^N \to \mathbb{R}$ of N real variables is a bit more complex. One approach is to reduce the problem to functions of one variable. Suppose $X, X^* \in \mathbb{R}^N$. For $0 \le t \le 1$ let

$$Y(t) = X^* + t(X - X^*), \quad y_n = x_n^* + t(x_n - x_n^*).$$

The curve $Y(t)$ is a line in $\mathbb{R}^N$ starting at X_n^* when $t = 0$ and ending at X when $t = 1$.

Consider the function $\phi : \mathbb{R} \to \mathbb{R}$ given by

$$\phi(t) = f(X^* + t(X - X^*)).$$

Apply the one-variable Taylor's Theorem to ϕ to get

$$\phi(1) = \phi(0) + \phi'(0)(1 - 0) + \frac{1}{2}\phi''(s)(1 - 0)^2 = \phi(0) + \phi'(0) + \frac{1}{2}\phi''(s),$$

where s is strictly between 0 and 1. Of course $\phi(0) = f(X^*)$. By the chain rule

$$\phi'(t) = \frac{d}{dt} f(X^* + t(X - X^*)) = \sum_{n=1}^{N} \frac{\partial f}{\partial y_n}(Y(t)) \frac{dy_n}{dt},$$

so

$$\phi'(0) = \sum_{n=1}^{N} \frac{\partial f}{\partial y_n}(X^*)(x_n - x_n^*) = \nabla f(X^*) \bullet (X - X^*).$$

Apply the chain rule again to compute the second derivative.

$$\phi''(t) = \frac{d}{dt}[\sum_{n=1}^{N} \frac{\partial f}{\partial y_n}(Y(t)) \frac{dy_n}{dt}]$$

$$= \sum_{m=1}^{N}[\sum_{n=1}^{N} \frac{\partial^2 f}{\partial y_m \partial y_n}(Y(t)) \frac{dy_m}{dt} \frac{dy_n}{dt} + \frac{\partial f}{\partial y_n}(Y(t)) \frac{d^2 y_n}{dt^2}].$$

Since $dy_n(t)/dt$ is a constant, the last terms are 0, and

$$\phi''(s) = \sum_{m=1}^{N}[\sum_{n=1}^{N} \frac{\partial^2 f}{\partial y_m \partial y_n}(Y(s))(x_m - x_m^*)(x_n - x_n^*)].$$

The second derivative can be expressed more compactly using vector and matrix notation. Let

$$Z = X^* + s(X - X^*),$$

and introduce the Hessian matrix-valued function

$$Hf(Z) = \begin{pmatrix} \frac{\partial^2 f}{\partial y_1 \partial y_1} & \cdots & \frac{\partial^2 f}{\partial y_1 \partial y_N} \\ \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial y_N \partial y_1} & \cdots & \frac{\partial^2 f}{\partial y_N \partial y_N} \end{pmatrix} (Z) = \left(\frac{\partial^2 f}{\partial y_m \partial y_n}(Z). \right)$$

If the (mixed) second partial derivatives are all continuous, then

$$\frac{\partial^2 f}{\partial y_m \partial y_n} = \frac{\partial^2 f}{\partial y_n \partial y_m}$$

and the Hessian is a real symmetric matrix. We summarize with this statement of Taylor's Theorem.

Theorem 2.2.2. *(Taylor's Theorem) Suppose $D \subset \mathbb{R}^N$ is open, $f : D \to \mathbb{R}$, and the first and second partial derivatives of f are continuous on D. If $X, X^* \in \mathbb{R}^N$ are distinct, and $Y(t) = X^* + t(X - X^*) \in D$ for $0 \le t \le 1$, then there is a Z on the line segment strictly between X and X^* such that*

$$f(X) = f(X^*) + \nabla f(X^*)(X - X^*) + \frac{1}{2}(X - X^*) \bullet Hf(Z)(X - X^*).$$

Just as for functions of one variable, a sign condition on the quadratic term can tell us whether critical points are local or global extremizers.

Theorem 2.2.3. *Suppose X^* is a critical point of $f : \mathbb{R}^N \to \mathbb{R}$ and f has continuous first and second partial derivatives on all of $\mathbb{R}^N$. Then*
(a) X^* is a global minimizer for f if $(X - X^*) \bullet Hf(Z)(X - X^*) \ge 0$ for all $X, Z \in \mathbb{R}^N$,
(b) X^* is a strict global minimizer for f if $(X - X^*) \bullet Hf(Z)(X - X^*) > 0$ for all $X \ne X^* \in \mathbb{R}^N$, and all $Z \in \mathbb{R}^N$,

There is also a local result based on the positivity of $(X - X^*) \bullet Hf(X^*)(X - X^*)$ which we will return to later.

Here is a simple example that helps set the direction to follow. Let

$$f(X) = a_1 x_1^2 + \dots a_N x_N^2 = \sum_{n=1}^{N} a_n x_n^2, \quad a_n > 0.$$

The only critical point is at $X = 0$ since

$$\nabla f = [2a_1 x_1, \dots, 2a_N x_N].$$

The Hessian is

$$Hf(X) = \mathrm{diag}[a_1, \dots, a_N] = \begin{pmatrix} a_1 & 0 & 0 & \dots & 0 \\ 0 & a_2 & 0 & \dots & 0 \\ 0 & 0 & \vdots & \vdots & 0 \\ 0 & 0 & 0 & \dots & a_N \end{pmatrix}.$$

For every X, X^* we find

$$(X - X^*) \bullet Hf(Z)(X - X^*) = \sum_{n=1}^{N} a_n (x_n - x_n^*)^2 \ge 0.$$

We noted above that under modest continuity hypotheses the Hessian

$$Hf(Z) = \begin{pmatrix} \frac{\partial^2 f}{\partial y_1 \partial y_1} & \dots & \frac{\partial^2 f}{\partial y_1 \partial y_N} \\ \vdots & \dots & \vdots \\ \frac{\partial^2 f}{\partial y_N \partial y_1} & \dots & \frac{\partial^2 f}{\partial y_N \partial y_N} \end{pmatrix} (Z)$$

is a real symmetric matrix function. If A is a real symmetric matrix, then

$$Q_A(Y) = Y \bullet AY$$

is called the quadratic form associated to A. The conditions we want for the Hessian to detect local or global extrema are properties of its quadratic form.

For the example f_1, the matrix $A = Hf_1(X)$ is constant.

$$Hf_1 = \begin{pmatrix} 2 & 8 \\ 8 & 2 \end{pmatrix}, \quad Q_A(X) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \bullet A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2x_1^2 + 16x_1x_2 + 2x_2^2.$$

The real symmetric matrix A and its quadratic form $Q_A(X)$ are
(a) positive semidefinite if

$$Q_A(X) = X \bullet AX \ge 0, \quad X \in \mathbb{R}^N,$$

(b) positive definite if

$$Q_A(X) = X \bullet AX > 0, \quad X \ne 0 \in \mathbb{R}^N,$$

(c) negative semidefinite if

$$Q_A(X) = X \bullet AX \le 0, \quad X \in \mathbb{R}^N,$$

(d) negative definite if

$$Q_A(X) = X \bullet AX < 0, \quad X \ne 0 \in \mathbb{R}^N,$$

(e) indefinite if $Q_A(X_1) < 0$ for some $X_1 \in \mathbb{R}^N$ and $Q_A(X_2) > 0$ for some other $X_2 \in \mathbb{R}^N$.

In these terms we are asking if the Hessian is positive or negative (semi) definite, or indefinite.

2.3 Tests for quadratic forms

The first thing to observe is that it is NOT sufficient to check the sign of the matrix entries. Here are some examples. The first, coming from f_1, has positive entries, but is not positive semidefinite.

$$A = \begin{pmatrix} 1 & 4 \\ 4 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ -1 \end{pmatrix} \bullet A \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \bullet \begin{pmatrix} 1-4 \\ 4-1 \end{pmatrix} = -6.$$

The second, coming from f_2, has some negative entries, but is positive definite.

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \bullet A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$
$$= x_1^2 - 2x_1x_2 + 4x_2^2 = (x_1 - x_2)^2 + 3x_2^2 > 0.$$

The third example is a general observation that for diagonal matrices the entry sign does determine the positivity.

$$A = \begin{pmatrix} a_1 & 0 & 0 & \dots & 0 \\ 0 & a_2 & 0 & \dots & 0 \\ 0 & 0 & \vdots & \vdots & 0 \\ 0 & 0 & 0 & \dots & a_N \end{pmatrix}, \quad Q_A(X) = \sum_{n=1}^N a_n x_n^2.$$

We want to develop some effective tests for positivity of a real symmetric matrix. Let's start with the 2×2 case, with

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

We want

$$X \bullet AX \geq 0, \quad X \in \mathbb{R}^2.$$

In particular the value must be nonnegative for $X = [1, 0]$, forcing $a_{11} > 0$.

If $x_2 \neq 0$ then we can write $x_1 = tx_2$, and we can consider the optimization problem of minimizing the quadratic form. Here

$$Q_A(X) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 = [a_{11}t^2 + 2a_{12}t + a_{22}]x_2^2.$$

The quadratic polynomial has a positive leading coefficient, so has a minimum. Minimizing this requires

$$2a_{11}t + 2a_{12} = 0, \quad t = -a_{12}/a_{11}.$$

Since A is symmetric, the minimum value is

$$Q_A(X^*) = [a_{11}\frac{a_{12}^2}{a_{11}^2} - 2a_{12}\frac{a_{12}}{a_{11}} + a_{22}]x_2^2$$

$$= [-a_{12}a_{21} + a_{11}a_{22}]\frac{x_2^2}{a_{11}} = \det(A)\frac{x_2^2}{a_{11}}.$$

This establishes an effective test for 2×2 matrices.

Theorem 2.3.1. *A 2×2 real symmetric matrix A is positive definite if and only if $a_{11} > 0$ and $\det(A) > 0$.*

A 2×2 real symmetric matrix A is negative definite if and only if $a_{11} < 0$ and $\det(A) > 0$.

This result has a generalization to $N \times N$ real symmetric matrices $A = (a_{mn})$. The proof takes some work, so it is not provided yet. The *k-th principle minor* Δ_k of A is the determinant of the upper left $k \times k$ block of A,

$$\Delta_k = \det(a_{mn}), \quad 1 \le m, n \le k.$$

Theorem 2.3.2. *If A is an $N \times N$ real symmetric matrix, then*

A is positive definite if and only if $\Delta_k > 0$ for $k = 1, \dots, N$, and

A is negative definite if and only if $(-1)^k\Delta_k > 0$ for $k = 1, \dots, N$,

It is a bit surprising that this test does not extend to the semidefinite case. Consider the matrix

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1/2 \end{pmatrix},$$

with $\Delta_1 = 1$, $\Delta_2 = 0$, $\Delta_3 = 0$. A test shows that

$$\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \bullet A \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \bullet \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -2.$$

Here are some examples.

Minimize the function

$$f = x_1^2 + x_2^2 + x_3^2 - x_1x_2 + x_2x_3 - x_1x_3.$$

The critical points where all partials vanish satisfy

$$2x_1 - x_2 - x_3 = 0,$$

$$-x_1 + 2x_2 + x_3 = 0,$$

$$-x_1 + x_2 + 2x_3 = 0.$$

In this case the Hessian is the same as the coefficient matrix for the critical points,

$$Hf(X) = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}.$$

The principle minors are $\Delta_1 = 2$, $\Delta_2 = 3$, $\Delta_3 = 4$. Thus the only critical point is $X = 0$ and the constant Hessian is positive definite, so we have a global minimizer at $X = 0$.

Show that $(x, y, z) = (0, 0, 0)$ is the global minimizer of

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

The critical points satisfy

$$e^{x-y} - e^{y-x} + 2xe^{x^2} = 0,$$

$$-e^{x-y} + e^{y-x} = 0,$$

$$2z = 0.$$

The only critical point is $(x, y, z) = (0, 0, 0)$. The Hessian is

$$Hf(X) = \begin{pmatrix} e^{x-y} + e^{y-x} + 4x^2e^{x^2} + 2e^{x^2} & -e^{x-y} - e^{y-x} & 0 \\ -e^{x-y} - e^{y-x} & e^{x-y} + e^{y-x} & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

It is easy to see that $\Delta_1 > 0$. Then

$$\Delta_2 = [e^{x-y} + e^{y-x}]^2 + [e^{x-y} + e^{y-x}][4x^2e^{x^2} + 2e^{x^2}] - [e^{x-y} + e^{y-x}]^2$$

$$= [e^{x-y} + e^{y-x}][4x^2e^{x^2} + 2e^{x^2}] > 0.$$

Finally, $\Delta_3 = 2\Delta_2$, so Hf is positive definite everywhere and the critical point is a global minimizer.

There are corresponding tests for local minima.

Theorem 2.3.3. *Suppose $f : \mathbb{R}^N \to \mathbb{R}$ has continuous first and second partials on $D \subset \mathbb{R}^N$. Suppose X^* is an interior critical point of D. Then*
(a) X^* is a strict local minimizer if $Hf(X^*)$ is positive definite,
(b) X^* is a strict local maximizer if $Hf(X^*)$ is negative definite.

Proof. By continuity of the second partials, the principle minors Δ_k are all positive in some ball B centered at X^*. Thus $Hf(X)$ is positive definite on B and Taylor's Theorem completes the result. □

In some cases the Hessian can also tell us a critical point is neither a min nor max. The test is finding $X, Y \in \mathbb{R}^N$ such that

$$X \bullet Hf(X^*)X > 0, Y \bullet Hf(X^*)Y < 0.$$

Once again this test is inconclusive if $Hf(X^*)$ is semidefinite as the examples $f = x^4 + y^4$ and $g = x^4 - y^4$ show.