

17. Unconstrained Optimization

A quick summary of this chapter is that

1. A necessary condition for \mathbf{x}_0 to be an interior optimal point of the \mathcal{C}^1 function $f(\mathbf{x})$ is that $Df(\mathbf{x}_0) = \mathbf{0}$. This equation is called the *first order necessary condition*.¹
2. Appropriate conditions on the second derivative together with the first order necessary condition are sufficient for an optimum

¹When $Df = \mathbf{0}$ involves more than one equation, we sometimes refer to *first order necessary conditions*.

17.1 Maxima and Minima

11/10/22

We will use the terms *max*, *maximizer*, and *maximum point* more or less interchangeably. The same is true of *min*, *minimizer*, and *minimum point*. They refer to points where a function takes its maximum or minimum value. If it is unlikely to cause confusion, we may identify the maximum or minimum point with the maximum or minimum value of the function, both of which can be called the max or min. Finally, we use *optimum* to refer to either a maximum or minimum, as appropriate.

Types of Maxima (and Minima). Let $f: U \rightarrow \mathbb{R}$ be a real-valued function with domain $U \subset \mathbb{R}^m$.

- A point $\mathbf{x}^* \in U$ is a *maximum* or *global maximum* of f over U if $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x} \in U$.
- A point $\mathbf{x}^* \in U$ is a *strict maximum* or *strict global maximum* of f over U if $f(\mathbf{x}^*) > f(\mathbf{x})$ for all $\mathbf{x} \neq \mathbf{x}^*$ with $\mathbf{x} \in U$.
- A point $\mathbf{x}^* \in U$ is a *local maximum* or *relative maximum* of f if there is a ball $B_\epsilon(\mathbf{x}^*)$ so that $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x} \in B_\epsilon(\mathbf{x}^*)$.
- A point $\mathbf{x}^* \in U$ is a *strict local maximum* or *strict relative maximum* of f if there is a ball $B_\epsilon(\mathbf{x}^*)$ so that $f(\mathbf{x}^*) > f(\mathbf{x})$ for all $\mathbf{x} \neq \mathbf{x}^*$ with $\mathbf{x} \in B_\epsilon(\mathbf{x}^*)$.

Similar terminology applies to minimum points.

17.2 Maxima and Minima

If the graph of a function has several hills, the hilltops represent local maxima. The top of the highest hill is a global maximum. Valley bottoms are minima, with the bottom of the lowest valley the global minimum.

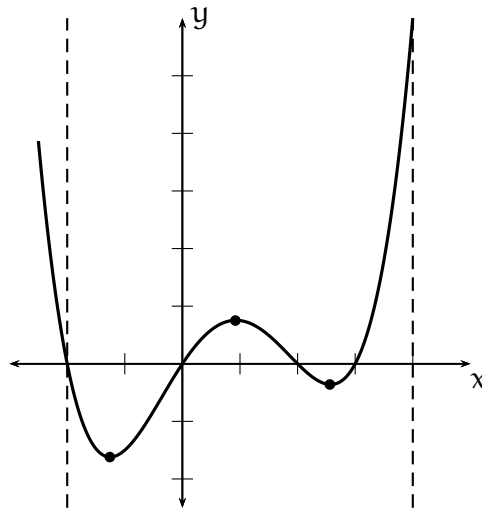


Figure 17.2.1: This function has three critical points on the interval $[-2, +4]$. The leftmost is a global minimum, the center one is a local maximum, and the rightmost is a local minimum. Here “global” refers to the interval $[-2, +4]$, so the global maximum is at the right endpoint. The dashed lines indicate the limits of the interval.

17.3 First Order Conditions

As with functions from \mathbb{R} to \mathbb{R} , our main technique for finding optima will start by rounding up the usual suspects. In the case of optimization, the usual suspects are the *critical points*, points \mathbf{x}^* where $Df(\mathbf{x}^*) = \mathbf{0}$.

Critical points are not guaranteed to be optimal, but every interior optimal point must be a critical point. For interior optimal points, it is **necessary** that they be critical points. All interior optima are critical points. But this is **not sufficient** to show they are in fact optima. Some interior critical points may not be optima. See the following example.

► **Example 17.3.1: Non-Optimal Critical Point.** Define $f: \mathbb{R} \rightarrow \mathbb{R}$ by $f(x) = x^3$. Then $f'(x) = 3x^2$, so $x = 0$ is the only critical point. It is neither a maximum nor a minimum because for any $x > 0$, $f(x) > 0 = f(0)$ and for any $x < 0$, $f(x) < 0$.

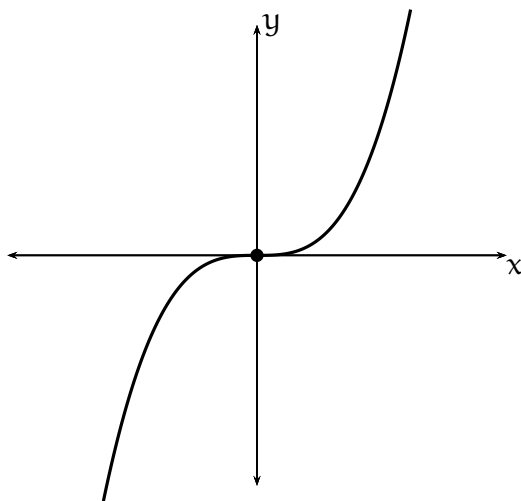


Figure 17.3.2: The function $f(x) = x^3$ is plotted here. Although $x = 0$ is a critical point, it is neither a maximizer nor minimizer.



17.4 Interior Optima are Critical Points

Theorem 17.4.1 establishes that all interior optima are critical points.

Theorem 17.4.1. *Let $f: \mathcal{U} \rightarrow \mathbb{R}$ be a real-valued \mathcal{C}^1 function defined on an subset $\mathcal{U} \subset \mathbb{R}^m$. If a point \mathbf{x}^* is in the interior of \mathcal{U} and is either a local maximum or local minimum of f , then $Df(\mathbf{x}^*) = \mathbf{0}$. That is,*

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0 \quad \text{for } i = 1, \dots, m.$$

Proof. Consider the difference quotient. We will consider the case of a maximizer \mathbf{x}^* . Then $f(\mathbf{x}^* + h\mathbf{e}_i) \leq f(\mathbf{x}^*)$ for all $h \neq 0$ and $i = 1, \dots, m$. Then the difference quotients obey

$$\begin{aligned} \frac{f(\mathbf{x}^* + h\mathbf{e}_i) - f(\mathbf{x}^*)}{h} &\leq 0 && \text{when } h > 0 \\ 0 &\leq \frac{f(\mathbf{x}^* + h\mathbf{e}_i) - f(\mathbf{x}^*)}{h} && \text{when } h < 0. \end{aligned}$$

Because $f \in \mathcal{C}^1$, the limit of the difference quotients exists as $h \rightarrow 0$. Moreover, it is the same from both directions. That means

$$0 \leq \frac{\partial f}{\partial x_i}(\mathbf{x}^*) \leq 0.$$

This holds for all $i = 1, \dots, m$, so

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0 \quad \text{for } i = 1, \dots, m.$$

■

17.5 Optima on the Boundary

There are versions of Theorem 17.4.1 that apply at boundary points. For some intuition about this, consider an interval $[a, b] \subset \mathbb{R}$. If a differentiable function has a local maximum at a we can consider $a + h$ for $h > 0$ but not $a + h$ for $h < 0$, which is out of the set $[a, b]$. By the argument in Theorem 17.4.1, it follows that $f'(a) \leq 0$.

A similar argument shows that a maximum at b must obey $f'(b) \geq 0$. Of course, the signs reverse for minima.

17.6 Toward the Second Order Conditions

We can use the Theorem 17.4.1 to identify possible optimal points. We need more information to determine whether such points are optima.

We turn to Taylor's formula for help in that effort. Let \mathbf{x}^* be a critical point, then there is a point \mathbf{y} in the line segment $\ell(\mathbf{x}, \mathbf{x}^*)$ with

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{x}^*) + Df(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T [D^2f(\mathbf{y})](\mathbf{x} - \mathbf{x}^*) \\ &= f(\mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T [D^2f(\mathbf{y})](\mathbf{x} - \mathbf{x}^*) \end{aligned} \quad (17.6.1)$$

because $Df(\mathbf{x}^*) = \mathbf{0}$.

We can rewrite this as

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{x}^*) &= \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T [D^2f(\mathbf{y})](\mathbf{x} - \mathbf{x}^*) \\ &= \frac{1}{2} [D^2f(\mathbf{y})](\mathbf{x} - \mathbf{x}^*)^{\otimes 2} \end{aligned} \quad (17.6.2)$$

The left-hand side of equation (17.6.2) will always be non-positive if the critical point \mathbf{x}^* is a maximum and non-negative if \mathbf{x}^* is a minimum. If the optimum is strict, that changes to negative for a strict minimum and positive for a strict maximum.

Unless it is zero, the behavior of the second derivative, the Hessian, will determine whether a critical point is a maximum or minimum.

17.7 What if the Hessian is Zero?

If the second derivative is zero, we need to check higher order terms, if possible. To get a little feel for that, if both first and second derivatives are zero, a non-zero third derivative ensures that an interior critical point is neither a maximum nor minimum. If the third derivative is zero, we then check the fourth derivative to see if it is positive or negative definite. If it is zero, we continue to the fifth derivative, etc.

An example is $f(x) = x^4$. This has a minimum at $x = 0$. Now $f'(x) = 4x^3$, yielding $x = 0$ as a critical point. However, $f''(0) = 0$, so the second order conditions are not helpful. The third derivative is $f'''(x) = 24x$, which is again 0 at $x = 0$. Finally, the fourth derivative is $f^{(4)}(x) = 24 > 0$, indicating a minimum at $x = 0$.

On \mathbb{R}^m , we can consider the tensors of the form

$$F(\mathbf{z}) = D^4f(\mathbf{x})(\mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z}),$$

which are positive definite if $F(\mathbf{z}) \geq 0$ and $F(\mathbf{z}) = 0$ if and only if $\mathbf{z} = \mathbf{0}$.

17.8 Global Sufficient Conditions: Max and Min

We turn to the global sufficient conditions next because they exhibit the basic argument in its purest form.

Suppose f is defined on U and \mathbf{x}^* is an interior critical point. To use equation (17.6.2), we will need $\ell(\mathbf{x}, \mathbf{x}^*) \subset U$ for $\mathbf{x} \in U$. The weakest condition that ensures this always happens is that the domain U of f is star-shaped with respect to \mathbf{x}^* . After all, U being star-shaped with respect to \mathbf{x}^* requires precisely that $\ell(\mathbf{x}, \mathbf{x}^*) \subset U$ for all $\mathbf{x} \in U$.

Of course, we also need \mathbf{x}^* in the interior of U to be sure that optima are critical points, the basis for equation (17.6.2). With that in mind, we can now state a theorem describing sufficient conditions for a maximum or minimum.

Theorem 17.8.1. *Let $f: U \rightarrow \mathbb{R}$ be a real-valued \mathcal{C}^2 function defined on an subset $U \subset \mathbb{R}^m$. Suppose \mathbf{x}^* is a critical point of f in the interior of U and that U is star-shaped with respect to \mathbf{x}^* .*

1. *If the Hessian $D^2f(\mathbf{x})$ is negative semidefinite for every $\mathbf{x} \in U$ then \mathbf{x}^* maximizes f over U .*
2. *If the Hessian $D^2f(\mathbf{x})$ is positive semidefinite for every $\mathbf{x} \in U$ then \mathbf{x}^* minimizes f over U .*

17.9 Proof of Global Sufficient Conditions: Max and Min

Proof. We start with equation (17.6.2):

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T [D^2f(\mathbf{y})](\mathbf{x} - \mathbf{x}^*). \quad (17.6.2)$$

where $\mathbf{y} \in \ell(\mathbf{x}^*, \mathbf{x})$. The line segment $\ell(\mathbf{x}^*, \mathbf{x}) \subset \mathcal{U}$ because \mathcal{U} is star-shaped with respect to \mathbf{x}^* . It follows that $D^2f(\mathbf{y})$ is negative semidefinite, so for all $\mathbf{x} \in \mathcal{U}$,

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{x}^*) &= \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T [D^2f(\mathbf{y})](\mathbf{x} - \mathbf{x}^*) \\ &\leq 0 \end{aligned}$$

showing that $f(\mathbf{x}) \leq f(\mathbf{x}^*)$ for every $\mathbf{x} \in \mathcal{U}$. The point \mathbf{x}^* maximizes f over \mathcal{U} .

The proof of case (2) is the same, except that $D^2f(\mathbf{y})$ is positive semidefinite, making $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{U}$. ■

17.10 Saddlepoints: Neither Max nor Min

Before continuing, we define saddlepoints, which are neither maxima nor minima.

Saddlepoint. A point $\mathbf{x}^* \in \text{dom } f$ a *saddlepoint* of f if the Hessian D^2f is indefinite at \mathbf{x}^* .

Saddlepoints are neither local maxima nor minima. There are directions where f increases as you move away from \mathbf{x}^* and there are directions where it decreases. If the function maps $\mathbb{R}^2 \rightarrow \mathbb{R}$, as is the case $f(\mathbf{x}) = x_1^2 - x_2^2$, the graph is similar to a saddle shape in the vicinity of $\mathbf{x}^* = 0$. In higher dimensions, the shapes can be more complex, but there will still be three-dimensional slices where it looks locally like a saddle, as shown in Figure 17.10.1.²

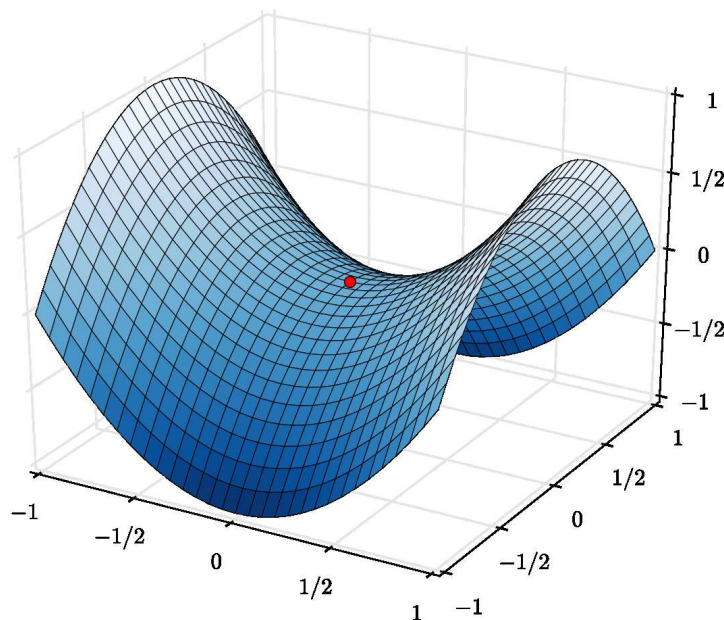


Figure 17.10.1: The red dot indicates the saddlepoint. Notice how the surface rises in some directions from the red dot, and falls in other directions. This is the defining characteristic of a saddlepoint. You'll notice, although it is missing a number of pieces, it looks a lot like a saddle one would place on a horse.

² The diagram was created by the Wikipedia user Nicoguardo, and is licensed under the Creative Commons Attribution 3.0 Unported license: <https://creativecommons.org/licenses/by/3.0/legalcode>.

17.11 Global Sufficient Conditions: Strict Max and Min

For strict optima, we need definite Hessians, not semidefinite Hessians. The proofs of cases (1) and (2) in Theorem 17.11.1 are barely modified from cases (1) and (2) in Theorem 17.8.1.

Case (3) is only sufficient in a negative sense. It shows that the critical point \mathbf{x}^* cannot be either a maximum or minimum. Proving it takes some extra work. We need to use the fact that \mathbf{x}^* is an interior point to shrink down two vectors where the quadratic term of equation (17.6.2) takes opposite signs. This yields two vectors that are actually in \mathcal{U} where the quadratic term takes opposite signs. We need them in \mathcal{U} so that the corresponding \mathbf{y} where D^2f is evaluated is also in \mathcal{U} .

Theorem 17.11.1. *Let $f: \mathcal{U} \rightarrow \mathbb{R}$ be a real-valued \mathcal{C}^2 function defined on an subset $\mathcal{U} \subset \mathbb{R}^m$. Suppose \mathbf{x}^* is a critical point of f in the interior of \mathcal{U} and that \mathcal{U} is star-shaped with respect to \mathbf{x}^* .*

1. *If the Hessian $D^2f(\mathbf{x})$ is negative definite for $\mathbf{x} \in \mathcal{U}$ then \mathbf{x}^* strictly maximizes f over \mathcal{U} .*
2. *If the Hessian $D^2f(\mathbf{x})$ is positive definite for $\mathbf{x} \in \mathcal{U}$ then \mathbf{x}^* strictly minimizes f over \mathcal{U} .*
3. *If the Hessian $D^2f(\mathbf{x}^*)$ is indefinite then \mathbf{x}^* is a saddlepoint of f over \mathcal{U} .*

Part (3) of the theorem doesn't involve a global condition on the Hessian. We only need to know the Hessian at \mathbf{x}^* to show that \mathbf{x}^* is neither a maximum nor a minimum.

17.12 Proof of Theorem 17.11.1

Proof of Theorem 17.11.1. We again start with equation (17.6.2):

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T [D^2f(\mathbf{y})](\mathbf{x} - \mathbf{x}^*). \quad (17.6.2)$$

where $\mathbf{y} \in \ell(\mathbf{x}^*, \mathbf{x})$. The line segment $\ell(\mathbf{x}^*, \mathbf{x}) \subset \mathcal{U}$ because \mathcal{U} is star-shaped with respect to \mathbf{x}^* . It follows that $D^2f(\mathbf{y})$ is negative definite, so for all $\mathbf{x} \in \mathcal{U}$ with $\mathbf{x} \neq \mathbf{x}^*$,

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{x}^*) &= \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T [D^2f(\mathbf{y})](\mathbf{x} - \mathbf{x}^*) \\ &< 0 \end{aligned}$$

showing that $f(\mathbf{x}) < f(\mathbf{x}^*)$ for every $\mathbf{x} \in \mathcal{U}$ other than $\mathbf{x} = \mathbf{x}^*$. This proves \mathbf{x}^* strictly maximizes f over \mathcal{U} .

The proof of case (2) is the same, except that $D^2f(\mathbf{y})$ is positive definite, making $f(\mathbf{x}^*) < f(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{U}$, so \mathbf{x}^* strictly minimizes f over \mathcal{U} .

In case (3), the Hessian $\mathbf{H} = D^2f(\mathbf{x}^*)$ is indefinite. Find unit vectors \mathbf{u}_i with $\mathbf{u}_1^T \mathbf{H} \mathbf{u}_1 < 0$ and $\mathbf{u}_2^T \mathbf{H} \mathbf{u}_2 > 0$. The two mappings, $\mathbf{y} \mapsto \mathbf{u}_i^T [D^2f(\mathbf{y})] \mathbf{u}_i$, $i = 1, 2$, are both continuous in \mathbf{y} . Choose $\varepsilon > 0$ small enough that $B_\varepsilon(\mathbf{x}^*) \subset \mathcal{U}$ and both values of the quadratic form have the same sign at $\mathbf{y} \in B_\varepsilon(\mathbf{x}^*)$ as they do at \mathbf{x}^* .

Then for $\delta < \varepsilon$, let

$$\mathbf{x}_i = \delta \mathbf{u}_i + \mathbf{x}^* \in B_\varepsilon(\mathbf{x}^*) \subset \mathcal{U}.$$

It follows that $f(\mathbf{x}_1) < f(\mathbf{x}^*)$ and $f(\mathbf{x}_2) > f(\mathbf{x}^*)$, showing that \mathbf{x}^* is neither a maximum nor a minimum. ■

17.13 Global Optima and Hessian Minors I

We can use the Semidefinite Matrices Theorems to rewrite Theorem 17.8.1 in terms of the minors of the Hessian.

Theorem 17.8.1 (Using Minors). *Let $f: \mathcal{U} \rightarrow \mathbb{R}$ be a real-valued \mathcal{C}^2 function defined on an subset $\mathcal{U} \subset \mathbb{R}^m$. Suppose \mathbf{x}^* is a critical point of f in the interior of \mathcal{U} and that \mathcal{U} is star-shaped with respect to \mathbf{x}^* .*

1. *If for every $\mathbf{x} \in \mathcal{U}$, the even order principal minors of the Hessian $D^2f(\mathbf{x})$ are non-negative and the odd order principal minors are non-positive, then \mathbf{x}^* maximizes f over \mathcal{U} .*
2. *If for every $\mathbf{x} \in \mathcal{U}$, every principal minor of the Hessian $D^2f(\mathbf{x})$ is non-negative, then \mathbf{x}^* minimizes f over \mathcal{U} .*

17.14 Global Optima and Hessian Minors II

Similarly, we can also use the Definite Matrices Theorems to restate Theorem 17.11.1 in terms of the minors of the Hessian.

Theorem 17.11.1 (Using Minors). *Let $f: \mathcal{U} \rightarrow \mathbb{R}$ be a real-valued \mathcal{C}^2 function defined on an subset $\mathcal{U} \subset \mathbb{R}^m$. Suppose \mathbf{x}^* is a critical point of f in the interior of \mathcal{U} and that \mathcal{U} is star-shaped with respect to \mathbf{x}^* .*

1. *If for every $\mathbf{x} \in \mathcal{U}$, the leading principal minors of the Hessian $\mathbf{H} = \mathbf{D}^2f(\mathbf{x})$ obey*

$$(-1)^k \det \mathbf{H}_k > 0 \quad \text{for all } k = 1, \dots, m$$

then \mathbf{x}^ strictly maximizes f over \mathcal{U} .*

2. *If for every $\mathbf{x} \in \mathcal{U}$, the leading principal minors of the Hessian $\mathbf{H} = \mathbf{D}^2f(\mathbf{x})$ obey*

$$\det \mathbf{H}_k > 0 \quad \text{for all } k = 1, \dots, m$$

then \mathbf{x}^ strictly minimizes f over \mathcal{U} .*

3. *If the Hessian $\mathbf{D}^2f(\mathbf{x}^*)$ has leading principal minors that are non-zero and violate the sign patterns in parts (1) and (2), then \mathbf{x}^* is a saddlepoint of f over \mathcal{U} .*

17.15 Second Order Sufficient Conditions

The next set of second order conditions includes the second order sufficient conditions, points (1) and (2) in the theorem. In these cases, we are able to show that a critical point \mathbf{x}^* is a local maximum or minimum.

Theorem 17.15.1. *Let $f: \mathcal{U} \rightarrow \mathbb{R}$ be a real-valued \mathcal{C}^2 function defined on an subset $\mathcal{U} \subset \mathbb{R}^m$. Suppose \mathbf{x}^* is a critical point of f in the interior of \mathcal{U} .*

1. *If the Hessian $D^2f(\mathbf{x}^*)$ is negative definite then \mathbf{x}^* is a strict local maximizer of f .*
2. *If the Hessian $D^2f(\mathbf{x}^*)$ is positive definite then \mathbf{x}^* is a strict local minimizer of f .*

Proof. As usual, we start with equation (17.6.2):

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T [D^2f(\mathbf{y})](\mathbf{x} - \mathbf{x}^*). \quad (17.6.2)$$

where $\mathbf{y} \in \ell(\mathbf{x}^*, \mathbf{x})$.

Keeping in mind that Hessians must be symmetric, Corollary 35.5.1, the shows that the set of positive (negative) definite matrices is an open set within the symmetric matrices. Choose $\varepsilon > 0$ so that $B_\varepsilon(\mathbf{x}^*)$ is contained in the inverse image of that open set. Then choose $\mathbf{x} \in B_\varepsilon(\mathbf{x}^*)$. The line segment $\ell(\mathbf{x}, \mathbf{x}^*) \in B_\varepsilon(\mathbf{x}^*)$, so $D^2f(\mathbf{y})$ will be positive (negative) definite for $\mathbf{x} \in B_\varepsilon(\mathbf{x}^*)$.

Then for $\mathbf{x} \in B_\varepsilon(\mathbf{x}^*)$ with $\mathbf{x} \neq \mathbf{x}^*$ and any $\mathbf{y} \in \ell(\mathbf{x}, \mathbf{x}^*)$,

$$(\mathbf{x} - \mathbf{x}^*)^T [D^2f(\mathbf{y})](\mathbf{x} - \mathbf{x}^*) \begin{cases} > 0 & \text{when } D^2f(\mathbf{x}^*) \text{ is positive definite} \\ < 0 & \text{when } D^2f(\mathbf{x}^*) \text{ is negative definite.} \end{cases}$$

Now we apply equation (17.6.2).

It follows that when $D^2f(\mathbf{x}^*)$ is positive definite, for all $\mathbf{x} \in B_\varepsilon(\mathbf{x}^*)$ with $\mathbf{x} \neq \mathbf{x}^*$, $f(\mathbf{x}) > f(\mathbf{x}^*)$. The point \mathbf{x}^* is a local minimizer, proving item (2).

It also follows that when $D^2f(\mathbf{x}^*)$ is negative definite, for all $\mathbf{x} \in B_\varepsilon(\mathbf{x}^*)$ with $\mathbf{x} \neq \mathbf{x}^*$, $f(\mathbf{x}) < f(\mathbf{x}^*)$. The point \mathbf{x}^* is a local maximizer, proving item (1). ■

17.16 Second Order Necessary Conditions

Although we will not directly cite Corollary 35.6.1, the fact that the semidefinite matrices are a closed set within the symmetric matrices plays a key role in proving the second order necessary conditions. We use a limiting argument instead of the corollary. That limiting argument could also be adapted to provide an alternative proof to Corollary 35.6.1.

Theorem 17.16.1. *Let $f: \mathcal{U} \rightarrow \mathbb{R}$ be a real-valued \mathcal{C}^2 function defined on an subset $\mathcal{U} \subset \mathbb{R}^m$. Suppose $\mathbf{x}^* \in \mathcal{U}^0$ is a local maximum or minimum. Then \mathbf{x}^* is a critical point of f , and:*

1. *If \mathbf{x}^* is a local maximizer, the Hessian $D^2f(\mathbf{x}^*)$ is negative semidefinite.*
2. *If \mathbf{x}^* is a local minimizer, the Hessian $D^2f(\mathbf{x}^*)$ is positive semidefinite.*

Proof. That \mathbf{x}^* is a critical point of f is Theorem 17.4.1. Since \mathbf{x}^* is a critical point, (17.6.2) holds:

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T [D^2f(\mathbf{y})](\mathbf{x} - \mathbf{x}^*). \quad (17.6.2)$$

where \mathbf{y} is a point in $\ell(\mathbf{x}^*, \mathbf{x})$.

If \mathbf{x}^* is a local max, we can take $\varepsilon > 0$ so that $f(\mathbf{x}) \leq f(\mathbf{x}^*)$ for $\mathbf{x} \in B_\varepsilon(\mathbf{x}^*)$. Then

$$0 \geq f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T [D^2f(\mathbf{y})](\mathbf{x} - \mathbf{x}^*)$$

for all $\mathbf{x} \in B_\varepsilon(\mathbf{x}^*)$. Let \mathbf{u} be any unit vector and define $\mathbf{x}_n = \mathbf{x}^* + (1/n)\mathbf{u}$, so $\mathbf{x}_n - \mathbf{x}^* = (1/n)\mathbf{u}$. For n large, $\mathbf{x}_n \in B_\varepsilon(\mathbf{x}^*)$ and

$$0 \geq (\mathbf{x}_n - \mathbf{x}^*)^T [D^2f(\mathbf{y}_n)](\mathbf{x}_n - \mathbf{x}^*) = \frac{1}{n^2}\mathbf{u}^T [D^2f(\mathbf{y}_n)]\mathbf{u}$$

where $\mathbf{y}_n \in \ell(\mathbf{x}_n, \mathbf{x}^*)$. Multiply by n^2 and then let $n \rightarrow \infty$ so that $\mathbf{y}_n \rightarrow \mathbf{x}^*$. Since D^2f is continuous, this yields

$$0 \geq \mathbf{u}^T [D^2f(\mathbf{x}^*)]\mathbf{u}.$$

This is true for any unit vector \mathbf{u} , so $D^2f(\mathbf{x}^*)$ is negative semidefinite.

The case where \mathbf{x}^* is a local maximum is similar. ■

17.17 Local Optima and Hessian Minors I

The Definite Matrices Theorems lets us restate the sufficient conditions of Theorem 17.15.1 in terms of the minors of the Hessian.

Theorem 17.15.1 (Using Minors). *Let $f: \mathcal{U} \rightarrow \mathbb{R}$ be a real-valued \mathcal{C}^2 function defined on an subset $\mathcal{U} \subset \mathbb{R}^m$. Suppose \mathbf{x}^* is a critical point of f in the interior of \mathcal{U} .*

1. *If the leading principal minors of the Hessian $\mathbf{H} = D^2f(\mathbf{x}^*)$ obey*

$$(-1)^k \det \mathbf{H}_k > 0 \quad \text{for all } k = 1, \dots, m$$

then \mathbf{x}^ is a strict local maximizer of f .*

2. *If the leading principal minors of the Hessian $\mathbf{H} = D^2f(\mathbf{x}^*)$ obey*

$$\det \mathbf{H}_k > 0 \quad \text{for all } k = 1, \dots, m$$

then \mathbf{x}^ is a strict local minimizer of f .*

17.18 Local Optima and Hessian Minors II

The Semidefinite Matrices Theorems can be used to restate the necessary conditions from Theorem 17.16.1 for a local optimum in terms of the minors of the Hessian.

Theorem 17.16.1 (Using Minors). *Let $f: \mathcal{U} \rightarrow \mathbb{R}$ be a real-valued \mathcal{C}^2 function defined on an subset $\mathcal{U} \subset \mathbb{R}^m$. Suppose $\mathbf{x}^* \in \mathcal{U}^0$ is a local maximum or minimum. Then \mathbf{x}^* is a critical point of f , and:*

1. *If \mathbf{x}^* is a local maximizer, the even order principal minors of the Hessian $\mathcal{D}^2f(\mathbf{x}^*)$ are non-negative and the odd order principal minors are non-positive.*
2. *If \mathbf{x}^* is a local minimizer, every principal minor of the Hessian $\mathcal{D}^2f(\mathbf{x}^*)$ is non-negative.*

November 11, 2022