Differential Properties of Functions Defined on $\mathbb{R}^n$

Prof. Dan A. Simovici

UMB

Let S be an open subset of $\mathbb{R}^n$ and let $f : S \longrightarrow \mathbb{R}$ be a function.
The partial derivatives of f are denoted by $\frac{\partial f}{\partial x_i}$ for $1 \leqslant \leqslant n$.
The second order partial derivatives are denoted by

$$\frac{\partial^2 f}{\partial x_i\, \partial x_j}.$$

Definition

The linear form

$$f'(\boldsymbol{x}, \boldsymbol{h}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} h_i$$

is the first differential of f at $\boldsymbol{x}$.
It is also the derivative at $t = 0$ of the function $g(t) = f(\boldsymbol{x} + t\boldsymbol{h})$.
$f'(\boldsymbol{x}, \boldsymbol{h})$ can be interpreted as the derivative of f at $\boldsymbol{x}$ in the direction $\boldsymbol{h}$.

Definition

The gradient of f at $\boldsymbol{x}$ is the vector

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \end{pmatrix}.$$

We have $f'(\boldsymbol{x}, \boldsymbol{h}) = \nabla f(\boldsymbol{x})'\boldsymbol{h}$.

The quadratic form

$$f''(\boldsymbol{x}, \boldsymbol{h}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \, \partial x_j} h_i h_j$$

is the second order differential of f at $\boldsymbol{x}$.
It is also the second order derivative of the function $g(t) = f(\boldsymbol{x} + t\boldsymbol{h})$; accordingly, it is the second order derivative of f at $\boldsymbol{x}$ in the direction $\boldsymbol{h}$.
The matrix of second derivatives of f is

$$f''(\boldsymbol{x}) = \left(\frac{\partial^2 f}{\partial x_i \, \partial x_j} \right)$$

is called the *Hessian* of f at $\boldsymbol{x}$.

Taylor Theorem for One Argument Functions

Theorem

(Taylor's Theorem for One-Argument Functions) *Let* $f : [a, b] \longrightarrow \mathbb{R}$ *be a function such that* f *together with its first* $n - 1$ *derivatives* $f^{(1)}, \ldots, f^{(n-1)}$ *are continuous on the interval* $[a, b]$. *If* $f^{(n)}$ *exists on* (a, b), *then there exists* $c \in (a, b)$ *such that*

$$\begin{aligned} f(b) &= f(a) + f^{(1)}(a)(b-a) + \frac{f^{(2)}(a)}{2!}(b-a)^2 + \cdots \\ &\quad + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \frac{f^{(n)}(c)}{n!}(b-a)^n. \end{aligned}$$

Proof

Let $\phi : [a, b] \longrightarrow \mathbb{R}$ be the function defined by

$$\begin{aligned} \phi(x) &= f(b) - f(x) - f^{(1)}(x)(b-x) - \frac{f^{(2)}(x)}{2!}(b-x)^2 - \cdots \\ &\quad - \frac{f^{(n-1)}(x)}{(n-1)!}(b-x)^{n-1}. \end{aligned}$$

The derivative of ϕ exists for any $x \in (a, b)$ and is easily seen to be

$$\phi'(x) = -\frac{f^{(n)}(x)}{(n-1)!}(b-x)^{n-1}.$$

Define the function $g : [a, b] \longrightarrow \mathbb{R}$ as

$$g(x) = \phi(x) - \left(\frac{b-x}{b-a}\right)^n \phi(a).$$

Proof (cont'd)

Since $g(a) = g(b) = 0$, by Rolle's Theorem from elementary analysis, there exists $c \in (a, b)$ such that $g'(c) = 0$. Note that

$$g'(x) = \phi'(x) + n\frac{(b-x)^{n-1}}{(b-a)^n}\phi(a),$$

so

$$\phi'(c) + n\frac{(b-c)^{n-1}}{(b-a)^n}\phi(a) = 0.$$

Proof (cont'd)

Since $\phi'(c) = -\frac{f^{(n)}(c)}{(n-1)!}(b-c)^{n-1}$, it follows that

$$\phi(a) = (b-a)^n n! f^{(n)}(c),$$

which implies the desired equality

$$f(b) - f(a) - f^{(1)}(a)(b-a) - \frac{f^{(2)}(a)}{2!}(b-a)^2 - \cdots - \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} \\ = \frac{(b-a)^n}{n!} f^{(n)}(c).$$

Notations

Let $f : S \longrightarrow \mathbb{R}$ be a function, where $S \subseteq \mathbb{R}^n$. If all partial derivatives $\frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}$ exist for $1 \leqslant k \leqslant n$ and $\boldsymbol{h} \in \mathbb{R}^n$, define the expression

$$f^{[k]}(\boldsymbol{x};\boldsymbol{h}) = \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(\boldsymbol{x}) h_{i_1} \cdots h_{i_k}.$$

This notation is needed for formulating Taylor's Theorem for functions of n variables. Note that

$$f^{[k]}(\boldsymbol{x};\lambda\boldsymbol{h}) = \lambda^k f^{[k]}(\boldsymbol{x};\boldsymbol{h}).$$

The function $f^{[1]}(\boldsymbol{x};\boldsymbol{h})$ is $f'(\boldsymbol{x},\boldsymbol{h})$, the *differential* of the function f.

Notations

For $k = 2$, we have

$$\begin{aligned} f^{[2]}(\boldsymbol{x};\boldsymbol{h}) &= \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} \frac{\partial^2 f}{\partial x_{i_1}\,\partial x_{i_2}}(\boldsymbol{x}) h_{i_1} h_{i_2} \\ &= \boldsymbol{h}' H_f(\boldsymbol{x})\boldsymbol{h}, \end{aligned}$$

where $H_f(\boldsymbol{x}) \in \mathbb{R}^{n\times n}$, the matrix defined by

$$H_f(\boldsymbol{x}) = \left(\frac{\partial^2 f}{\partial x_i\,\partial x_j}\right).$$

is the as the *Hessian matrix* of the function f at $\boldsymbol{x}$.
$f^{[2]}(\boldsymbol{x};\boldsymbol{h})$ is $f''(\boldsymbol{x},\boldsymbol{h})$ the second order differential of f at $\boldsymbol{x}$.

Taylor's Theorem for Functions of Several Arguments

Theorem

Let $f : S \longrightarrow \mathbb{R}$ be a function, where $S \subseteq \mathbb{R}^n$ is an open set. If f and all its partial derivatives of order less or equal to m are differentiable on S, $\boldsymbol{a}, \boldsymbol{b} \in S$ such that $[\boldsymbol{a}, \boldsymbol{b}] \subseteq S$, then there exists a point $\boldsymbol{c} \in [\boldsymbol{a}, \boldsymbol{b}]$ such that

$$f(\boldsymbol{b}) \quad = \quad f(\boldsymbol{a}) + \sum_{k=1}^{m-1} \frac{1}{k!} f^{[k]}(\boldsymbol{a}, \boldsymbol{b} - \boldsymbol{a}) + \frac{1}{m!} f^{[m]}(\boldsymbol{c}, \boldsymbol{b} - \boldsymbol{a}).$$

Proof

Let $g : \mathbb{R} \longrightarrow \mathbb{R}$ be the function defined by $g(t) = f(\boldsymbol{p}(t))$, where $\boldsymbol{p}(t) = \boldsymbol{a} + t(\boldsymbol{b} - \boldsymbol{a}))$ for $t \in [0, 1]$. We have $g(0) = f(\boldsymbol{a})$ and $g(1) = f(\boldsymbol{b})$. The Taylor's formula applied to g yields the existence of $c \in (0, 1)$ such that

$$g(1) = g(0) + \sum_{k=1}^{m-1} \frac{1}{k!} g^{(k)}(0) + g^{(m)}(c).$$

We claim that

$$g^{(m)}(t) = f^{[m]}(\boldsymbol{p}(t), \boldsymbol{b} - \boldsymbol{a}).$$

Indeed, for $m = 1$, by applying the chain rule we have

$$g'(t) = \sum_{j=1}^{m} \frac{\partial f}{\partial x_j}(\boldsymbol{p}(t))(b_j - a_j) = f^{[1]}(\boldsymbol{x}; \boldsymbol{b} - \boldsymbol{a}).$$

Proof (cont'd)

Suppose that the equality holds for m. Then

$$g^{(m+1)}(t) = (f^{[m]}(\boldsymbol{p}(t), \boldsymbol{b} - \boldsymbol{a}))'$$
$$= \left(\sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} \frac{\partial^m f}{\partial x_{i_1} \cdots \partial x_{i_m}} (\boldsymbol{p}(t))(b_{i_1} - a_{i_1}) \cdots (b_{i_m} - a_{i_m}) \right)'$$
$$= \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} \sum_{i_{m+1}=1}^{n} \frac{\partial^{m+1} f}{\partial x_{i_1} \cdots \partial x_{i_{m+1}}} (\boldsymbol{p}(t))(b_{i_1} - a_{i_1}) \cdots (b_{i_m} - a_{i_m})(b_{i_{m+1}} -$$
$$= f^{[m+1]}(\boldsymbol{p}(t), \boldsymbol{b} - \boldsymbol{a}).$$

When the values of the derivatives of g are substituted we obtain the equality of the theorem.

Proof

Example

For $m = 2$, Taylor's Theorem yields the existence of $\boldsymbol{c} \in [\boldsymbol{a}, \boldsymbol{b}]$ such that

$$
\begin{aligned}
f(\boldsymbol{b}) &= f(\boldsymbol{a}) + f^{[1]}(\boldsymbol{a}, \boldsymbol{b} - \boldsymbol{a}) + \frac{1}{2}(\boldsymbol{b} - \boldsymbol{a})' H_f(\boldsymbol{c})(\boldsymbol{b} - \boldsymbol{a}) \\
&= f(\boldsymbol{a}) + (\nabla f)'_{\boldsymbol{a}}(\boldsymbol{b} - \boldsymbol{a}) + \frac{1}{2}(\boldsymbol{b} - \boldsymbol{a})' H_f(\boldsymbol{c})(\boldsymbol{b} - \boldsymbol{a}).
\end{aligned}
$$

Example

Let $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$ be the function given by $f(\boldsymbol{x}) = \| \boldsymbol{x} \|$ for $\boldsymbol{x} \in \mathbb{R}^3$. We have

$$\frac{\partial f}{\partial x_1} = \frac{x_1}{\sqrt{x_1^2 + x_2^2 + x_3^2}} = \frac{x_1}{\| \boldsymbol{x} \|},$$

and similar expressions for $\frac{\partial f}{\partial x_2}$ and $\frac{\partial f}{\partial x_3}$ and we have

$$(\nabla f)_{\boldsymbol{x}_0} = \frac{1}{\| \boldsymbol{x}_0 \|} \boldsymbol{x}_0$$

if $\boldsymbol{x}_0 \neq \mathbf{0}_3$. The gradient $\nabla(f)_{\boldsymbol{x}_0}$ is a unit vector for every $\boldsymbol{x}_0 \in \mathbb{R}^3 - \{\mathbf{0}_3\}$.

Example (cont'd)

The function $R(\boldsymbol{x}_0, \boldsymbol{x})$ is given by

$$
\begin{aligned}
R(\boldsymbol{x}_0, \boldsymbol{x}) &= \frac{f(\boldsymbol{x}) - f(\boldsymbol{x}_0) - (\nabla f)'_{\boldsymbol{x}_0}(\boldsymbol{x} - \boldsymbol{x}_0)}{\parallel \boldsymbol{x} - \boldsymbol{x}_0 \parallel} \\
&= \frac{\parallel \boldsymbol{x} \parallel - \parallel \boldsymbol{x}_0 \parallel - \frac{1}{\|\boldsymbol{x}_0\|}\boldsymbol{x}'_0(\boldsymbol{x} - \boldsymbol{x}_0)}{\parallel \boldsymbol{x} - \boldsymbol{x}_0 \parallel}
\end{aligned}
$$

Therefore, the function f is differentiable in $\boldsymbol{x}_0$ if $\boldsymbol{x}_0 \neq \mathbf{0}_3$.

Example

Let $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$ be the function given by $f(\boldsymbol{x}) = \parallel \boldsymbol{x} \parallel^2 =$ for $\boldsymbol{x} \in \mathbb{R}^3$. We have

$$\frac{\partial f}{\partial x_1} = 2x_1, \frac{\partial f}{\partial x_2} = 2x_2, \frac{\partial f}{\partial x_3} = 2x_3,$$

so $(\nabla f)_{\boldsymbol{x}} = 2\boldsymbol{x}$ and this function is differentiable for all $\boldsymbol{x} \in \mathbb{R}^3$.

Definition

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a function.
A point $\boldsymbol{x}_0 \in \mathbb{R}^n$ is a local minimum for f if there exists a closed sphere $B(\boldsymbol{x}_0, \epsilon)$ such that $f(\boldsymbol{x}_0) \leqslant f(\boldsymbol{x})$ for every $\boldsymbol{x} \in B(\boldsymbol{x}_0, \epsilon)$. If we have $f(\boldsymbol{x}_0) < f(\boldsymbol{x})$ for every $\boldsymbol{x} \in B(\boldsymbol{x}_0, \epsilon) - \{\boldsymbol{x}_0\}$, then $\boldsymbol{x}_0$ is a strict local minimum.
A global minimum for f is a point $\boldsymbol{x}_0$ such that $f(\boldsymbol{x}_0) \leqslant f(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^n$; $\boldsymbol{x}_0$ is a strict global minimum if $f(\boldsymbol{x}_0) < f(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^n - \{\boldsymbol{x}_0\}$.

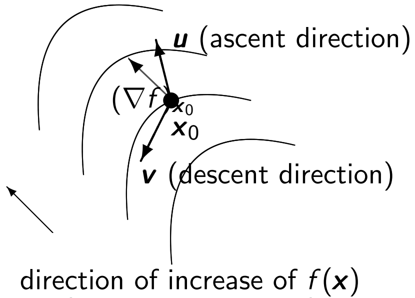
Similar definitions can be formulated for local maxima, strict local maxima, global maxima, and strict global maxima:
$\boldsymbol{x}_0 \in \mathbb{R}^n$ is a local maximum if there exists a closed sphere $B(\boldsymbol{x}_0, \epsilon)$ such that $f(\boldsymbol{x}_0) \geqslant f(\boldsymbol{x})$ for every $\boldsymbol{x} \in B(\boldsymbol{x}_0, \epsilon)$. If $f(\boldsymbol{x}_0) > f(\boldsymbol{x})$ for every $\boldsymbol{x} \in B(\boldsymbol{x}_0, \epsilon) - \{\boldsymbol{x}_0\}$, then $\boldsymbol{x}_0$ is a strict local maximum.
A global maximum for f is a point $\boldsymbol{x}_0$ such that $f(\boldsymbol{x}_0) \geqslant f(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^n$; $\boldsymbol{x}_0$ is a strict global maximum if $f(\boldsymbol{x}_0) > f(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^n - \{\boldsymbol{x}_0\}$.
A local minimum or maximum of f is said to be a *local extremum*.

An unconstraint optimization problem consists in finding a local minimum or a local maximum of a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, when such a minimum exists. The function f is referred to as the objective function.
Finding a local minimum of a function f is equivalent to finding a local maximum for the function $-f$.

Definition

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a function.
An ascent direction for f in $\boldsymbol{x}_0$ is a vector $\boldsymbol{h} \in \mathbb{R}^n$ such that there exists a positive number ϵ for which $0 < t < \epsilon$ implies $f(x_0 + t\boldsymbol{h}) \geqslant f(\boldsymbol{x}_0)$.
A descent direction for f in $\boldsymbol{x}_0$ is a vector $\boldsymbol{h} \in \mathbb{R}^n$ such that exists a positive number ϵ for which $0 < t < \epsilon$ implies $f(x_0 + t\boldsymbol{h}) \leqslant f(\boldsymbol{x}_0)$.

If $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a differentiable function, the existence of the gradient provides new instruments for computing extrema. The vector $(\nabla f)_{\boldsymbol{x}}$ points in the direction of increased values of the function f.



An ascent direction $\boldsymbol{u}$ in $\boldsymbol{x}_0$ makes an acute angle with the vector $(\nabla f)_{\boldsymbol{x}_0}$, while a descent direction $\boldsymbol{v}$ makes an obtuse angle with the same vector, as we see in the next statement.

Theorem

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *be a differentiable function at* $\boldsymbol{x}_0$*.*
If $(\nabla f)'_{\boldsymbol{x}_0}\boldsymbol{w} < 0$*, then* $\boldsymbol{w}$ *is a descent direction for* f *in* $\boldsymbol{x}_0$*.*
If $(\nabla f)'_{\boldsymbol{x}_0}\boldsymbol{w} > 0$*, then* $\boldsymbol{w}$ *is a ascent direction for* f *in* $\boldsymbol{x}_0$*.*

Proof

Since f is differentiable in $\boldsymbol{x}_0$ we can write

$$f(\boldsymbol{x}_0 + t\boldsymbol{w}) = f(\boldsymbol{x}_0) + t(\nabla f)'_{\boldsymbol{x}_0}\boldsymbol{w} + t \parallel \boldsymbol{w} \parallel R(\boldsymbol{x}_0, \boldsymbol{x}_0 + t\boldsymbol{w}),$$

for $t > 0$, where $\lim_{t \to 0} R(\boldsymbol{x}_0, \boldsymbol{x}_0 + t\boldsymbol{w}) = 0$. This implies

$$\frac{f(\boldsymbol{x}_0 + t\boldsymbol{w}) - f(\boldsymbol{x}_0)}{t} = (\nabla f)'_{\boldsymbol{x}_0}\boldsymbol{w} + \parallel \boldsymbol{w} \parallel R(\boldsymbol{x}_0, \boldsymbol{x}_0 + t\boldsymbol{w}).$$

Since $(\nabla f)'_{\boldsymbol{x}_0}\boldsymbol{w} < 0$ and $\lim_{t \to 0} R(\boldsymbol{x}_0, \boldsymbol{x}_0 + t\boldsymbol{w}) = 0$, there exists $\epsilon > 0$ such that $0 < t < \epsilon$ implies $f(\boldsymbol{x}_0 + t\boldsymbol{w}) - f(\boldsymbol{x}_0) \leqslant 0$, so $\boldsymbol{w}$ is a descent direction. The argument for the second part of the theorem is similar.

Theorem

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a function differentiable at $\mathbf{x}_0$. If $\mathbf{x}_0$ is a local minimum or a local maximum, then $(\nabla f)_{\mathbf{x}_0} = \mathbf{0}_n$.

Proof

Let $\boldsymbol{x}_0$ be a local minimum of f. Suppose that $(\nabla f)_{\boldsymbol{x}_0} \neq \mathbf{0}_n$ and let $\boldsymbol{w} = -(\nabla f)_{\boldsymbol{x}_0}$. We have $(\nabla f)'_{\boldsymbol{x}_0}\boldsymbol{w} < 0$, so, by Theorem **??**, $\boldsymbol{w}$ is a descent direction, that is, there exists a positive number ϵ such that $0 < t < \epsilon$ implies $f(x_0 + t\boldsymbol{u}) \leqslant f(\boldsymbol{x}_0)$, which contradicts the initial assumption concerning $\boldsymbol{x}_0$. Therefore, $(\nabla f)_{\boldsymbol{x}_0} = \mathbf{0}_n$.
The case when $\boldsymbol{x}_0$ is a local maximum can be treated similarly.

Definition

A *stationary point* of a differentiable function $f : S \longrightarrow \mathbb{R}$ (where $S \subseteq \mathbb{R}^n$) is a point $\boldsymbol{x} \in S$ such that $(\nabla f)_{\boldsymbol{x}} = \mathbf{0}_n$.

Observe that a stationary point of a function is not necessarily a local extremum of the function, as the next example shows.

Example

Consider the function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by $f(x_1, x_2) = x_1 x_2$ for $(x_1, x_2) \in \mathbb{R}^2$. We have

$$(\nabla f)_{\boldsymbol{x}} = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix},$$

which shows that $\mathbf{0}_2$ is a stationary point of f. Note that in any sphere $B(\mathbf{0}_2, \epsilon)$ there are both positive and negative numbers. Since $f(\mathbf{0}_2) = 0$, it is clear that although $\mathbf{0}_2$ is a stationary point of f, $\mathbf{0}_2$ is not a local extremum.

For functions that are twice differentiable it is possible to give characterizations of local extrema.
The next theorem gives sufficient conditions for the existence of a local minimum.

Theorem

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a twice differentiable function at $\boldsymbol{x}_0$.
If $(\nabla f)_{\boldsymbol{x}_0} = \mathbf{0}_n$, and the Hessian matrix $H_f(\boldsymbol{x}_0)$ is positive definite, then $\boldsymbol{x}_0$ is a local minimum of f.

By Taylor's Theorem,taking into account that $(\nabla f)_{\boldsymbol{x}_0} = \mathbf{0}_n$, we have

$$\begin{aligned} f(\boldsymbol{x}) &= f(\boldsymbol{x}_0) + (\boldsymbol{x} - \boldsymbol{x}_0)' H_f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \\ &\quad + \parallel \boldsymbol{x} - \boldsymbol{x}_0 \parallel^2 R(\boldsymbol{x}_0, \boldsymbol{x} - \boldsymbol{x}_0), \end{aligned}$$

where $\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} R(\boldsymbol{x}_0, \boldsymbol{x} - \boldsymbol{x}_0) = 0$.
Suppose that $\boldsymbol{x}_0$ is not a minimum. Then, there exists a sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n, \ldots$ such that $\lim_{n \to \infty} \boldsymbol{x}_n = \boldsymbol{x}_0$ such that $f(\boldsymbol{x}_n) < f(\boldsymbol{x}_0)$ for each $n \geqslant 1$. Let $\boldsymbol{r}_n$ be the unit vector $\boldsymbol{r}_n = \frac{1}{\|\boldsymbol{x}_n - \boldsymbol{x}_0\|}(\boldsymbol{x}_n - \boldsymbol{x}_0)$.

Proof (cont'd)

We have

$$\begin{aligned} f(\boldsymbol{x}_n) &= f(\boldsymbol{x}_0) + \| \boldsymbol{x}_n - \boldsymbol{x}_0 \|^2 \boldsymbol{r}_n' H_f(\boldsymbol{x}_0)\boldsymbol{r}_n \\ &\quad + \| \boldsymbol{x}_n - \boldsymbol{x}_0 \|^2 R(\boldsymbol{x}_0, \boldsymbol{x}_n - \boldsymbol{x}_0) < f(\boldsymbol{x}_0), \end{aligned}$$

which implies

$$\boldsymbol{r}_n' H_f(\boldsymbol{x}_0)\boldsymbol{r}_n + R(\boldsymbol{x}_0, \boldsymbol{x}_n - \boldsymbol{x}_0) < 0.$$

The sequence $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n, \ldots)$ is bounded since it consists of unit vectors and, therefore, it contains a subsequence convergent subsequence $(\boldsymbol{r}_{i_1}, \ldots, \boldsymbol{r}_{i_m}, \ldots)$ such that $\lim_{m\to\infty} \boldsymbol{r}_{i_m} = \boldsymbol{r}$ and $\| \boldsymbol{r} \| = 1$. This implies $\boldsymbol{r}' H_f(\boldsymbol{x}_0)\boldsymbol{r} \leqslant 0$, which contradicts the fact that $H_f(\boldsymbol{x}_0)$ is positive definite. Therefore, $\boldsymbol{x}_0$ is a local minimum.