AARHUS
UNIVERSITET

Optimization

- Elements of Calculus

- Basics of constrained and unconstrained optimization

Henrik Karstoft
@
Aarhus University, Department of Engineering

In this lecture, we will talk about methods to find the extrema of a function.
These methods are also useful if you don't have the function

Optimization – the general problem

Optimization

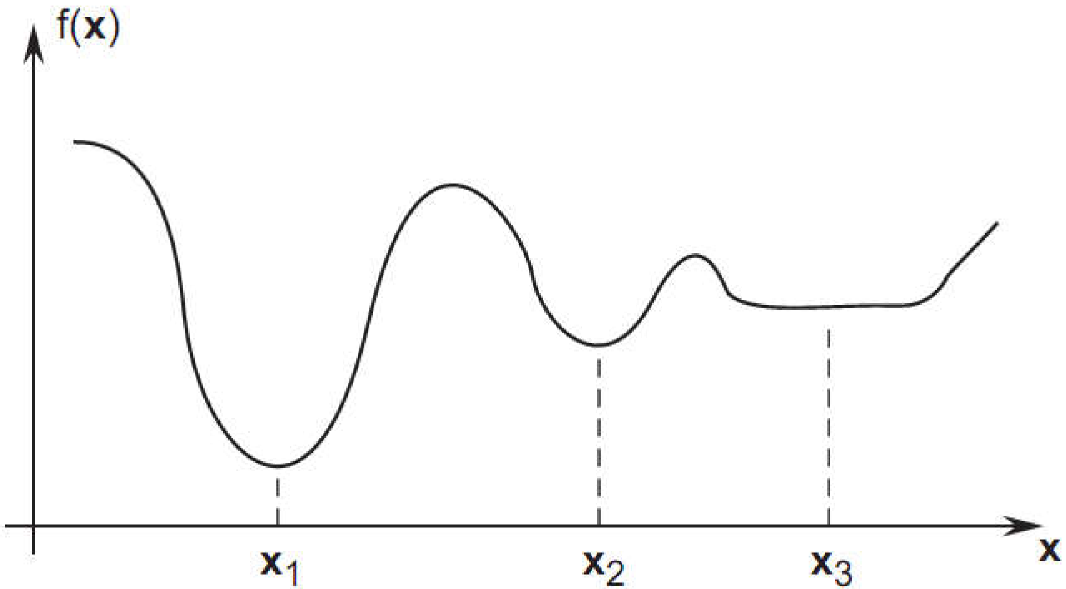
minimize $f(x)$,
subject to $x \in \Omega$

- $x = (x_1, \dots, x_n)$, optimization variables, or decision variables
- $f\colon R^n \to R$, objective function
- Ω, is the constraint set or feasible set, sometimes $\Omega = \{x \in R^n |\ f_i(x) \le b_i$, where $i = 1, \dots, m\}$
- $f_i(x)\ R^n \to R$, where $i = 1, \dots, m$: constraint functions
- x **optimal solution** or **minimizer,** is the smallest value of f among $x = (x_1, \dots, x_n)$, satisfying the constraint

Henrik Karstoft
Aarhus University, Department of Engineering

Local minimizer

Definition 6.1 Suppose $f\colon R^n \to R$, is a real-valued function defined on the set $\Omega \subseteq R^n$. A point x is a local minimizer of f over Ω if there exists $\varepsilon > 0$ such that $f(x) \geq f(x^*)$ for all $x \in \Omega$ and $\| x - x^* \| < \varepsilon$. A point x^* is a global minimizer of f over Ω if $f(x) \geq f(x\ )$ for all $x \in \Omega$. omega is the set of contraints



There are many minimisers around x3

Partiel derivatives and the Hessian matrix

Gradient of $f : \mathbb{R}^n \to \mathbb{R}$ is a column vector

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

First derivative—row vector

$$Df(\boldsymbol{x}) = \nabla f(\boldsymbol{x})^\top = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

$$\begin{aligned} \boldsymbol{F}(\boldsymbol{x}) &= D^2 f(\boldsymbol{x}) = \begin{bmatrix} D\left(\frac{\partial f}{\partial x_1}\right) \\ D\left(\frac{\partial f}{\partial x_2}\right) \\ \vdots \\ D\left(\frac{\partial f}{\partial x_n}\right) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \end{aligned}$$

Example 6.1 Let $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$. Then,

$$Df(\boldsymbol{x}) = (\nabla f(\boldsymbol{x}))^\top = \left[\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \frac{\partial f}{\partial x_2}(\boldsymbol{x})\right] = [5 + x_2 - 2x_1, 8 + x_1 - 4x_2]$$

and

$$\boldsymbol{F}(\boldsymbol{x}) = D^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -4 \end{bmatrix}.$$

■

Directional derivatives

The directional derivative of f at $x \in \mathbb{R}^n$ in the direction d is denoted

$$\frac{\partial f(x)}{\partial d}$$

$$\begin{aligned}\frac{\partial f(x)}{\partial d} &= \lim_{\alpha \to 0} \frac{f(x+\alpha d) - f(x)}{\alpha} \\ &= \frac{d}{d\alpha} f(x+\alpha d)\Big|_{\alpha=0} \\ &= \begin{bmatrix} & Df(x) & \end{bmatrix} \begin{bmatrix} \\ d \\ \\ \end{bmatrix}\end{aligned}$$

This is a composition of two functions like f(g(x))

Example 6.2 Define $f : \mathbb{R}^3 \to \mathbb{R}$ by $f(\boldsymbol{x}) = x_1 x_2 x_3$, and let

$$\boldsymbol{d} = \left[\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right]^\top .$$

The directional derivative of f in the direction $\boldsymbol{d}$ is

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \nabla f(\boldsymbol{x})^\top \boldsymbol{d} = [x_2 x_3, x_1 x_3, x_1 x_2] \begin{bmatrix} 1/2 \\ 1/2 \\ 1/\sqrt{2} \end{bmatrix} = \frac{x_2 x_3 + x_1 x_3 + \sqrt{2} x_1 x_2}{2}.$$

Level sets and contour plots

The *level set* of a function $f : \mathbb{R}^n \to \mathbb{R}$ at level c is the set of points

$$S = \{\boldsymbol{x} : f(\boldsymbol{x}) = c\}.$$

Gradient is always orthogonal to the level set

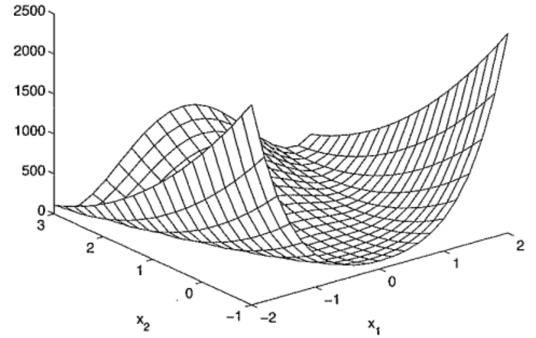
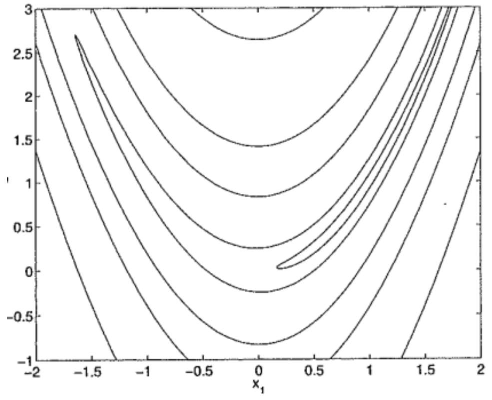


Figure 5.2 Graph of Rosenbrock's function.



$$f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2,$$

MatLab demo: ContoursInMatlab.m

Henrik Karstoft
Aarhus University, Department of Engineering

Exercise on Derivatives

Find a point (x,y) where the gradient parallel to the y-axis for the following function:

Set the x=0 to find the gradient parallel to the y-axis

$$f(x, y) = 3\mathrm{x}^2 \quad 4\mathrm{xy} + 7\mathrm{y}^2$$

In the point you found above, what is the directional derivation in the direction (1,0)?

Feasible directions

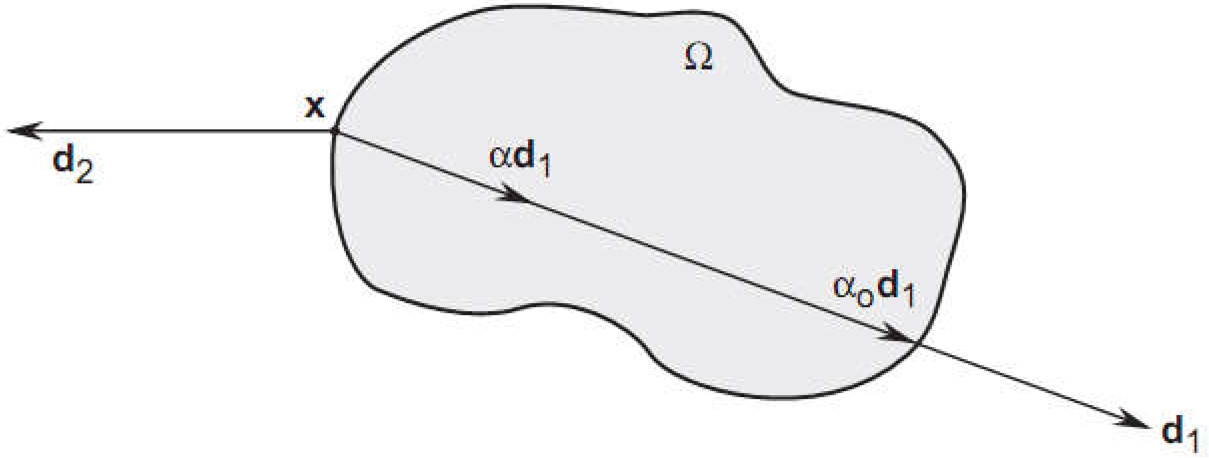
are used in constrained optimisation. A feasible direction is associated with a specific point.

$\boldsymbol{d} \in \mathbb{R}^n$ is a feasible direction at $\boldsymbol{x} \in \Omega$ if there is $\alpha_0 > 0$ such that

$$\boldsymbol{x} + \alpha \boldsymbol{d} \in \Omega \quad \text{for all} \quad \alpha \in [0, \alpha_0]$$

alpha can very small`

In other words, d is a feasible direction at x if x is added to d with a very small value, the resulting point is still within the feasible set.



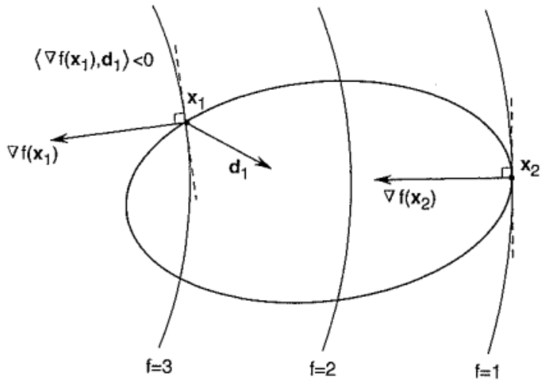
First Order Necessary Condition (FONC)

Theorem 6.1 ***First-Order Necessary Condition (FONC).*** *Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-valued function on Ω. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω, then for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$, we have*

$$\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*) \geq 0.$$

□

This allows us to determine whether a point can be a local minimiser in the case where the point is at the border of the feasible set



Less important

Corollary 6.1 ***Interior Case.*** *Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-valued function on Ω. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω and if $\boldsymbol{x}^*$ is an interior point of Ω, then*

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}.$$

□

More important

Henrik Karstoft
Aarhus University, Department of Engineering

FONC example (Thm. 6.1)

Example 6.3 Consider the problem

$$\begin{array}{ll} \text{minimize} & x_1^2 + 0.5x_2^2 + 3x_2 + 4.5 \\ \text{subject to} & x_1, x_2 \geq 0. \end{array}$$

a. Is the first-order necessary condition (FONC) for a local minimizer satisfied at $\boldsymbol{x} = [1, 3]^\top$?

b. Is the FONC for a local minimizer satisfied at $\boldsymbol{x} = [0, 3]^\top$?

c. Is the FONC for a local minimizer satisfied at $\boldsymbol{x} = [1, 0]^\top$?

d. Is the FONC for a local minimizer satisfied at $\boldsymbol{x} = [0, 0]^\top$?

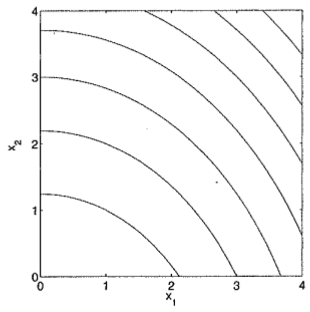


Figure 6.4 Level sets of the function in Example 6.3.

FONC example (Cor. 6.1)

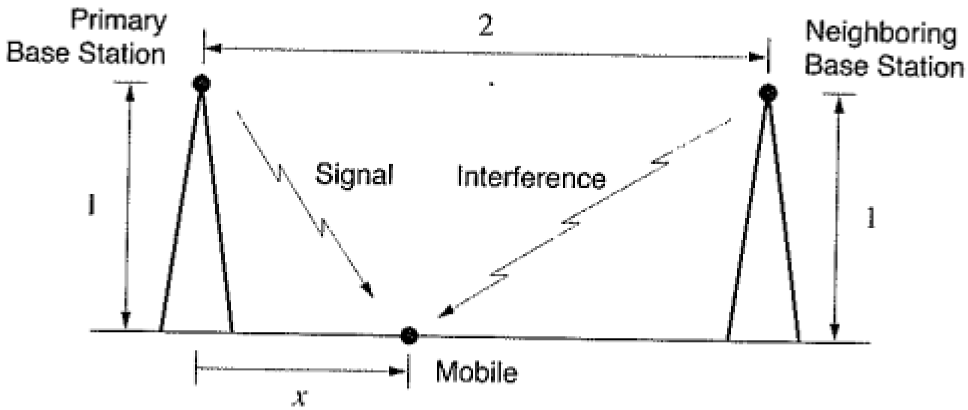


Figure 6.5 Simplified cellular wireless system in Example 6.4.

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

$$\begin{aligned} f'(x) &= \frac{-2(2 - x)(1 + x^2) - 2x(1 + (2 - x)^2)}{(1 + x^2)^2} \\ &= \frac{4(x^2 - 2x - 1)}{(1 + x^2)^2}. \end{aligned}$$

Quadratic forms – using matrices

EXAMPLE 1 Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Compute $\mathbf{x}^T A \mathbf{x}$ for the following matrices:

a. $A = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}$ b. $A = \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}$

a. $\mathbf{x}^T A \mathbf{x} = [\, x_1 \quad x_2 \,] \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [\, x_1 \quad x_2 \,] \begin{bmatrix} 4x_1 \\ 3x_2 \end{bmatrix} = 4x_1^2 + 3x_2^2.$

b. There are two -2 entries in A. Watch how they enter the calculations. The $(1, 2)$-entry in A is in boldface type.

$$\begin{aligned} \mathbf{x}^T A \mathbf{x} &= [\, x_1 \quad x_2 \,] \begin{bmatrix} 3 & \mathbf{-2} \\ -2 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [\, x_1 \quad x_2 \,] \begin{bmatrix} 3x_1 - \mathbf{2}x_2 \\ -2x_1 + 7x_2 \end{bmatrix} \\ &= x_1(3x_1 - \mathbf{2}x_2) + x_2(-2x_1 + 7x_2) \\ &= 3x_1^2 - \mathbf{2}x_1x_2 - 2x_2x_1 + 7x_2^2 \\ &= 3x_1^2 - 4x_1x_2 + 7x_2^2 \end{aligned}$$

Examples from Lay sec. 7.2

Quadratic forms – using matrices

EXAMPLE 2 For $\mathbf{x}$ in $\mathbb{R}^3$, let $Q(\mathbf{x}) = 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1x_2 + 8x_2x_3$. Write this quadratic form as $\mathbf{x}^TA\mathbf{x}$.

Solution The coefficients of x_1^2, x_2^2, x_3^2 go on the diagonal of A. To make A symmetric, the coefficient of x_ix_j for $i \neq j$ must be split evenly between the (i, j)- and (j, i)-entries in A. The coefficient of x_1x_3 is 0. It is readily checked that

$$Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x} = [\,x_1 \quad x_2 \quad x_3\,]\begin{bmatrix} 5 & -1/2 & 0 \\ -1/2 & 3 & 4 \\ 0 & 4 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Examples from Lay sec. 7.2

Qudratic forms – using matrices

EXAMPLE 3 Let $Q(\mathbf{x}) = x_1^2 - 8x_1x_2 - 5x_2^2$. Compute the value of $Q(\mathbf{x})$ for $\mathbf{x} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \end{bmatrix}$, and $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$.

Solution

$$
\begin{aligned}
Q(-3, 1) &= (-3)^2 - 8(-3)(1) - 5(1)^2 = 28 \\
Q(2, -2) &= (2)^2 - 8(2)(-2) - 5(-2)^2 = 16 \\
Q(1, -3) &= (1)^2 - 8(1)(-3) - 5(-3)^2 = -20
\end{aligned}
$$

In some cases, quadratic forms are easier to use when they have no cross-product terms—that is, when the matrix of the quadratic form is a diagonal matrix. Fortunately, the cross-product term can be eliminated by making a suitable change of variable.

Examples from Lay sec. 7.2

Qudratic forms – change of varibles

EXAMPLE 4 Make a change of variable that transforms the quadratic form in Example 3 into a quadratic form with no cross-product term.

Solution The matrix of the quadratic form in Example 3 is

$$A = \begin{bmatrix} 1 & -4 \\ -4 & -5 \end{bmatrix}$$

The first step is to orthogonally diagonalize A. Its eigenvalues turn out to be $\lambda = 3$ and $\lambda = -7$. Associated unit eigenvectors are

$$\lambda = 3: \begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}; \qquad \lambda = -7: \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

These vectors are automatically orthogonal (because they correspond to distinct eigenvalues) and so provide an orthonormal basis for $\mathbb{R}^2$. Let

$$P = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}, \qquad D = \begin{bmatrix} 3 & 0 \\ 0 & -7 \end{bmatrix}$$

Examples from Lay sec. 7.2

Henrik Karstoft
Aarhus University, Department of Engineering

Qudratic forms – and extremal points

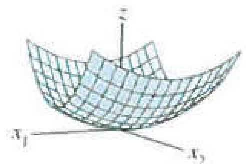
DEFINITION

A quadratic form Q is:

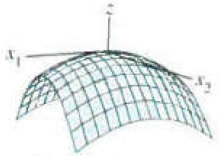
a. **positive definite** if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$,

b. **negative definite** if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$,

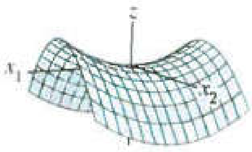
c. **indefinite** if $Q(\mathbf{x})$ assumes both positive and negative values.



Positive definite



Negative definite



Indefinite

THEOREM 5

Quadratic Forms and Eigenvalues

Let A be an $n \times n$ symmetric matrix. Then a quadratic form $\mathbf{x}^T A\mathbf{x}$ is:

a. positive definite if and only if the eigenvalues of A are all positive,

b. negative definite if and only if the eigenvalues of A are all negative, or

c. indefinite if and only if A has both positive and negative eigenvalues.

a. Minimum

b. Maximum

c. No min. No max

mples from Lay se

Henrik Karstoft
Aarhus University, Department of Engineering

Exercise on Qudratic forms

Classify the quadratic forms in Exercises 9–18. Then make a change of variable, $\mathbf{x} = P\mathbf{y}$, that transforms the quadratic form into one with no cross-product term. Write the new quadratic form. Construct P using the methods of Section 7.1.

9. $3x_1^2 - 4x_1x_2 + 6x_2^2$

10. $9x_1^2 - 8x_1x_2 + 3x_2^2$

11. $2x_1^2 + 10x_1x_2 +$ [illegible]

12. $-5x_1^2 + 4x_1x_2 - 2x_2^2$

13. $x_1^2 - 6x_1x_2 + 9$[illegible]

[illegible] $+ 6x_1x_2$

15. **[M]** $-2x_1^2 - 6$[illegible]$x_1x_2 + 4x_1x_3 + 4x_1x_4 + 6x_3x_4$

16. **[M]** $4x_1^2 + 4$[illegible]$x_3x_4 - 4x_1x_4 + 4x_2x_3$

17. **[M]** $x_1^2 + x_2^2 + x_3^2 + x_4^2$ [illegible] $2x_2x_3 + 9x_3x_4$

18. **[M]** $11x_1^2 - x_2^2 - 12x_1x_2$ [illegible] $2x_3x_4$

Examples from Lay sec. 7.2

Optimization

Theorem 6.2 ***Second-Order Necessary Condition (SONC).*** *Let $\Omega \subset \mathbb{R}^n$, $f \in \mathcal{C}^2$ a function on Ω, $\boldsymbol{x}^*$ a local minimizer of f over Ω, and $\boldsymbol{d}$ a feasible direction at $\boldsymbol{x}^*$. If $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$, then*

$$\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0,$$

where $\boldsymbol{F}$ is the Hessian of f. □

Corollary 6.2 ***Interior Case.*** *Let $\boldsymbol{x}^*$ be an interior point of $\Omega \subset \mathbb{R}^n$. If $\boldsymbol{x}^*$ is a local minimizer of $f : \Omega \to \mathbb{R}$, $f \in \mathcal{C}^2$, then*

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0},$$

This is important.

and $\boldsymbol{F}(\boldsymbol{x}^*)$ is positive semidefinite ($\boldsymbol{F}(\boldsymbol{x}^*) \geq 0$); that is, for all $\boldsymbol{d} \in \mathbb{R}^n$,

$$\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0.$$

This is the same as saying that all eigenvalues are positive

□

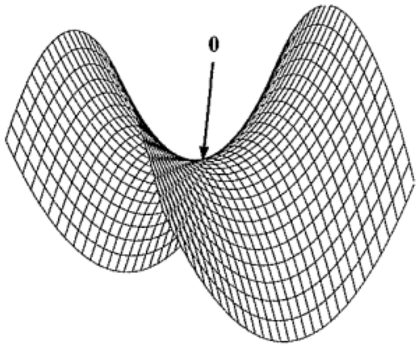
Optimization

Example 6.6 Consider a function of one variable $f(x) = x^3$, $f : \mathbb{R} \to \mathbb{R}$. Because $f'(0) = 0$, and $f''(0) = 0$, the point $x = 0$ satisfies both the FONC and SONC. However, $x = 0$ is not a minimizer (see Figure 6.6). ■

Example 6.7 Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$, where $f(\boldsymbol{x}) = x_1^2 - x_2^2$. The FONC requires that $\nabla f(\boldsymbol{x}) = [2x_1, -2x_2]^\top = \mathbf{0}$. Thus, $\boldsymbol{x} = [0, 0]^\top$ satisfies the FONC. The Hessian matrix of f is

$$\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}.$$

Here, the eigenvalues are easy to read off the matrix.



Optimization

Theorem 6.3 ***Second-Order Sufficient Condition (SOSC), Interior Case.*** *Let $f \in \mathcal{C}^2$ be defined on a region in which $\boldsymbol{x}^*$ is an interior point. Suppose that*

1. $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$.
2. $\boldsymbol{F}(\boldsymbol{x}^*) > 0$.

Then, $\boldsymbol{x}^*$ is a strict local minimizer of f. □