MATH3027: Optimization 2022

Week 2: Unconstrained Optimization

Prof. Richard Wilkinson
Please send any comments or mistakes to r.d.wilkinson@nottingham.ac.uk

This week we will study unconstrained optimization problems in $\mathbb{R}^n$. We first define global and local minima (and maxima) of a function, and then we give a first characterization of stationary points of a function. Since the characterization of minima and maxima of a function is related to its Hessian, we need to study some relevant matrix properties, such as positive definiteness and diagonal dominance. These properties will be useful for stating second order and global optimality conditions. We conclude this week by studying a very important class of problems where the cost function is a quadratic form.

Global Minimum and Maximum

Definition (Global Minimum)**.** *Let $f : S \to \mathbb{R}$ be defined on a set $S \subseteq \mathbb{R}^n$. Then:*

1. *$\mathbf{x}^* \in S$ is a global minimum point of f over S if $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for any $\mathbf{x} \in S$.*
2. *$\mathbf{x}^* \in S$ is a strict global minimum point of f over S if $f(\mathbf{x}) > f(\mathbf{x}^*)$ for any $\mathbf{x}^* \neq \mathbf{x} \in S$.*

The global maximum is defined similarly. In addition, the minimal value of f over S is

$$\inf\{f(\mathbf{x}) : \mathbf{x} \in S\}.$$

and the maximal value is $\sup\{f(\mathbf{x}) : \mathbf{x} \in S\}$. Note: we don't generally use 'inf' and 'sup' in these notes, but 'min' and 'max' where usage does not implies that the max and min are achieved. Note also that the minimal and maximal values are always unique.

In this chapter, optimization is over the entire domain of the function, but later we'll look at constrained problems. The set S over which optimization is performed is called the *feasible set* and $x \in S$ a *feasible solution.*

A constrained example: Find the global minimum and maximum points of $f(x_1, x_2) = x_1 + x_2$ over the unit ball in $\mathbb{R}^2$, $S = B[0,1] = \left\{(x_1, x_2)^T : x_1^2 + x_2^2 \leq 1\right\}$.

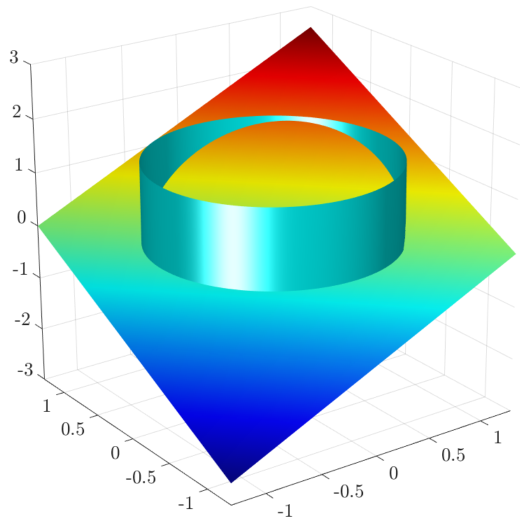

Figure 1: The function $f(x_1, x_2) = x_1 + x_2$ constrained over the unit ball.

Local Minima and Maxima

Definition (Local Minimum and Maximum)**.** *Let $f : S \longrightarrow \mathbb{R}$ be defined on a set $S \subseteq \mathbb{R}^n$. Then $\mathbf{x}^* \in S$ is a local minimum of f over S if there exists $r > 0$ for which $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for any $\mathbf{x} \in S \cap B(\mathbf{x}^*, r)$. We say $\mathbf{x}^*$ is a strict local minimum if the inequality is strict.*

Of course, a global minimum (maximum) point is also a local minimum (maximum) point. In optimization, by convention we only consider minimization problems. That is because we can always write any maximization problem as a minimization problem, i.e., maximizing $f(x)$ is the same as minimizing $-f(x)$.

Example: Identify the different (strict) local minima and maxima of the function

$$f(x) = \begin{cases} (x-1)^2 + 2, & -1 \le x \le 1 \\ 2, & 1 \le x \le 2 \\ -(x-2)^2 + 2, & 2 \le x \le 2.5 \\ (x-3)^2 + 1.5, & 2.5 \le x \le 4 \\ -(x-5)^2 + 3.5, & 4 \le x \le 6 \\ -2x + 14.5, & 6 \le x \le 6.5 \\ 2x - 11.5, & 6.5 \le x \le 8 \end{cases} \tag{1}$$

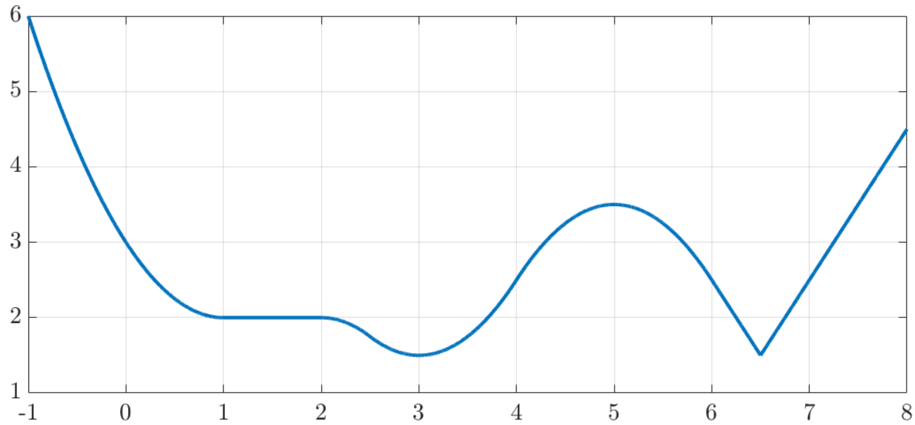

Figure 2: The function described in (1) has several minima and maxima.

Throughout the module we will look at different kinds of conditions for a point x^* to be an optimum of f. We will distinguish between **necessary** and **sufficient** conditions, and between **first** and **second** order conditions. A first order condition only uses information about the first derivative of f, whereas a second order condition uses the second derivative.

For one-dimensional differentiable functions $f : [a, b] \longrightarrow \mathbb{R}$, you've seen that if $x^* \in (a, b)$ is a local optima then $f'(x) = 0$. This is a first order necessary condition (it is not a sufficient condition). This is known as Fermat's theorem. The multidimensional extension of this is:

Theorem (Fermat's Theorem: Necessary First Order Optimality Conditions)**.** *Let $f : U \longrightarrow \mathbb{R}$ be a function defined on a set $U \subseteq \mathbb{R}^n$. Suppose that $\mathbf{x}^* \in \operatorname{int}(U)$ is a local optimum point and that all the partial derivatives of f exist at $\mathbf{x}^*$. Then $\nabla f(\mathbf{x}^*) = 0$.*

Proof. Let $i \in \{1, 2, \ldots, n\}$ and consider the 1-D function $g(t) = f(\mathbf{x}^* + t\mathbf{e}_i)$. Then g is differentiable at $t = 0$ and $g'(0) = \frac{\partial f}{\partial x_i}(\mathbf{x}^*)$. Because $\mathbf{x}^*$ is a local optimum point of f, we can see that $t = 0$ is a local optimum of g. Hence $g'(0) = \frac{\partial f}{\partial x_i}(\mathbf{x}^*) = 0$. □

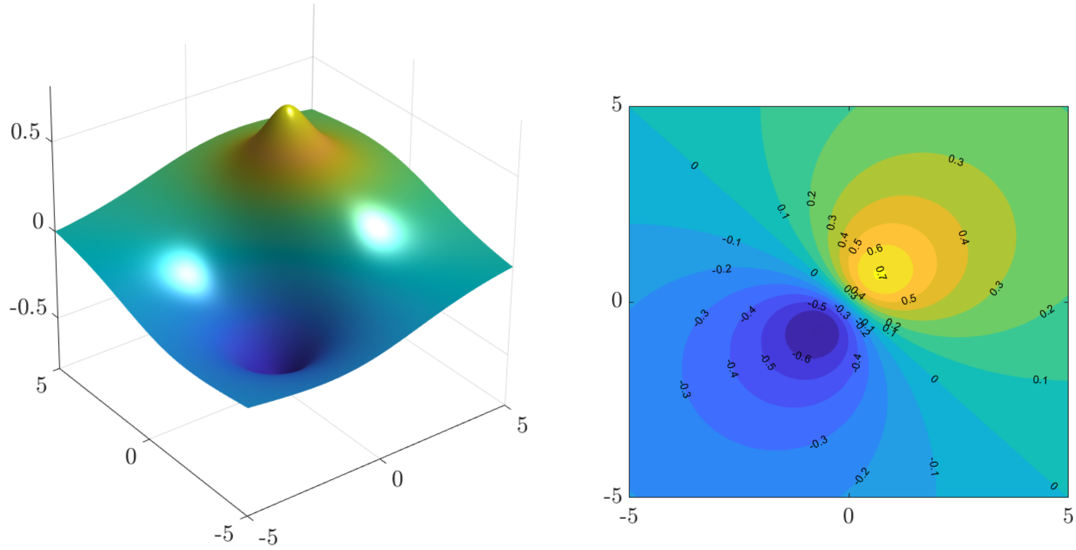

Figure 3: The function $f(x_1, x_2) = \frac{x_1+x_2}{x_1^2+x_2^2+1}$ (left) and its contour plot (right).

Note that $\nabla f(\mathbf{x}^*) = 0$ is a **necessary** condition for $\mathbf{x}^*$ to be a local optima, not a **sufficient** condition. It is possible for $\nabla f'(\mathbf{x}^*) = 0$ and $\mathbf{x}^*$ to not be a local optima. In other words,

$$\mathbf{x}^* \text{a local optima} \Rightarrow \nabla f(\mathbf{x}^*) = 0$$

but the converse is not true.

Definition (Stationary Points[1]). *Let $f : U \to \mathbb{R}$ be a function defined on a set $U \subseteq \mathbb{R}^n$. Suppose that $\mathbf{x}^* \in \text{int}(U)$ and that all the partial derivatives of f are defined at $\mathbf{x}^*$. Then $\mathbf{x}^*$ is called a stationary point of f if* $\nabla f(\mathbf{x}^*) = 0$.

Example: Find the global maximizers and minimizers in $\mathbb{R}^2$ of

$$f(x_1, x_2) = \frac{x_1 + x_2}{x_1^2 + x_2^2 + 1} \quad ?$$

Let's first compute the gradient of f

$$\nabla f(x_1, x_2) = \frac{1}{\left(x_1^2+x_2^2+1\right)^2} \begin{pmatrix} \left(x_1^2 + x_2^2 + 1\right) - 2(x_1 + x_2)x_1 \\ \left(x_1^2 + x_2^2 + 1\right) - 2(x_1 + x_2)x_2 \end{pmatrix}$$

The stationary points satisfy

$$\begin{aligned} -x_1^2 - 2x_1x_2 + x_2^2 &= -1 \\ x_1^2 - 2x_1x_2 - x_2^2 &= -1 \,. \end{aligned}$$

Hence, the stationary points are $(1/\sqrt{2}, 1/\sqrt{2})$ and $(-1/\sqrt{2}, -1/\sqrt{2})$, with $(1/\sqrt{2}, 1/\sqrt{2})$ a global maximum, and $(-1/\sqrt{2}, -1/\sqrt{2})$ a global minimum.

[1] When we study constrained optimization problems, we will introduce a more general definition of stationarity.

Classification of Matrices

Definition (Positive Definiteness)**.** *A symmetric matrix* $\mathbf{A} \in \mathbb{R}^{n \times n}$ *is called* ***positive semidefinite****, denoted by* $\mathbf{A} \geq \mathbf{0}$, *if* $\mathbf{x}^T \mathbf{A}\mathbf{x} \geq 0$ *for every* $\mathbf{x} \in \mathbb{R}^n$.

We say $\mathbf{A}$ *is* ***positive definite*** *if* $\mathbf{x}^T A\mathbf{x} > 0$ *for every non-zero* $\mathbf{x} \in \mathbb{R}^n$ *and write* $\mathbf{A} \succ \mathbf{0}$.

Negative definite and negative semidefinite are defined similarly (change the direction of the inequality). Note that

- $\mathbf{A}$ is negative (semi)definite if and only if $-\mathbf{A}$ is positive (semi)definite.
- The sum of two positive(negative) (semi)definite matrices is positive(negative) (semi)definite.

Proposition. *Let* $\mathbf{A}$ *be a positive definite (semidefinite) matrix. Then the diagonal elements of* $\mathbf{A}$ *are positive (nonnegative).*

Proof. If $A_{ii} < 0$, consider $\mathbf{e}_i^\top \mathbf{A}\mathbf{e}_i$.

Definition (Indefinite matrices)**.** *A symmetric matrix* $\mathbf{A} \in \mathbb{R}^{n \times n}$ *is called indefinite if there exist* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *such that* $\mathbf{x}^T \mathbf{A}\mathbf{x} > 0$ *and* $\mathbf{y}^T \mathbf{A}\mathbf{y} < 0$.

Some remarks:

- A matrix is indefinite if and only if it is neither positive semidefinite nor negative semidefinite.
- A symmetric matrix that has positive and negative elements in the diagonal is indefinite.

Some of the optimal conditions we look at will depend on whether the Hessian matrix is positive/negative semidefinite/definite. Thus it will be important that we are able to correctly classify matrices as such.

Theorem (Eigenvalue Characterization)**.** *Let* $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix. Then:*

a) $\mathbf{A}$ *is positive definite iff all its eigenvalues are positive.*

b) $\mathbf{A}$ *is positive semidefinite iff all its eigenvalues are nonnegative.*

c) $\mathbf{A}$ *is negative definite iff all its eigenvalues are negative.*

d) $\mathbf{A}$ *is negative semidefinite iff all its eigenvalues are nonpositive.*

e) $\mathbf{A}$ *is indefinite iff it has at least one positive eigenvalue and at least one negative eigenvalue.*

Proof. By the Spectral Factorization theorem, there exists orthogonal $\mathbf{U} \in \mathbb{R}^{n\times n}$ such that

$$\mathbf{U}^T\mathbf{A}\mathbf{U} = \mathbf{D} \equiv \operatorname{diag}(d_1, d_2, \ldots, d_n) \ ,$$

where $d_i = \lambda_i(\mathbf{A})$, the i^{th} eigenvalue of $\mathbf{A}$. Making the linear change of variables $\mathbf{x} = \mathbf{U}\mathbf{y}$, we have

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{y}^T\mathbf{U}^T\mathbf{A}\mathbf{U}\mathbf{y} = \mathbf{y}^T\mathbf{D}\mathbf{y} = \sum_{i=1}^{n} d_i y_i^2 \ .$$

The result then follows. For example, for part a) $\mathbf{x}^T A\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$ iff

$$\sum_{i=1}^{n} d_i y_i^2 > 0 \text{ for any } \mathbf{y} \neq 0 \ .$$

The latter holds iff $d_i > 0$ for all i. □

Trace and Determinant. Let $\mathbf{A}$ be a positive semidefinite (definite) matrix. Then the trace $Tr(\mathbf{A})$ and the determinant $\det(\mathbf{A})$ are nonnegative (positive). This follows directly recalling that the trace of a matrix is the sum of its eigenvalues and the determinant its product.

For a 2×2 matrix $\mathbf{A}$, we can show that $\mathbf{A}$ is positive semidefinite (definite) if and only if $\operatorname{Tr}(\mathbf{A}), \det(\mathbf{A}) \geq 0$ ($\operatorname{Tr}(\mathbf{A}), \det(\mathbf{A}) > 0$). See the exercises.

The Principal Minors Criteria

Given an $n \times n$ matrix, the determinant of the upper left $k \times k$ submatrix is called the k-th principal minor and is denoted by $D_k(\mathbf{A})$. For example, if

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$\text{then } D_1(\mathbf{A}) = a_{11}, D_2(\mathbf{A}) = \det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, D_3(\mathbf{A}) = \det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Proposition. *Let $\mathbf{A}$ be an $n \times n$ symmetric matrix. Then $\mathbf{A}$ is positive definite if and only if*

$$D_1(\mathbf{A}) > 0, D_2(\mathbf{A}) > 0, \ldots, D_n(\mathbf{A}) > 0 \ .$$

Corollary. *Let $\mathbf{A}$ be an $n \times n$ symmetric matrix. Then $\mathbf{A}$ is negative definite if and only if*

$$(-1)^k D_k(\mathbf{A}) > 0 \ , \quad \textit{for all } \ k = 1, \ldots, n.$$

This is equivalent to checking that $-\mathbf{A}$ is positive definite.

Diagonal Dominance

Definition (Diagonal Dominance). *Let A be a symmetric $n \times n$ matrix.*

a) $\mathbf{A}$ *is called diagonally dominant if*

$$|\mathbf{A}_{ii}| \geq \sum_{j \neq i} |\mathbf{A}_{ij}| \;\; \forall i = 1, 2, \ldots, n$$

b) $\mathbf{A}$ *is called strictly diagonally dominant if*

$$|\mathbf{A}_{ii}| > \sum_{j \neq i} |\mathbf{A}_{ij}| \;\; \forall i = 1, 2, \ldots, n$$

Theorem (Positive definiteness of diagonally dominant matrices). *a) If the matrix* $\mathbf{A}$ *is symmetric, diagonally dominant with nonnegative diagonal elements, then* $\mathbf{A}$ *is positive semidefinite.*

b) If $\mathbf{A}$ *is symmetric, strictly diagonally dominant with positive diagonal elements, then* $\mathbf{A}$ *is positive definite.*

Second Order Optimality Conditions

We now consider necessary and sufficient conditions that are useful in classifying stationary points.

Theorem (Necessary Second Order Optimality Conditions). *Let $f : U \to \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable and that* $\mathbf{x}^*$ *is a stationary point. Then*

1. $\mathbf{x}^*$ *is a local minimum point* $\Rightarrow \nabla^2 f(\mathbf{x}^*) \succeq 0$.
2. $\mathbf{x}^*$ *is a local maximum point* $\Rightarrow \nabla^2 f(\mathbf{x}^*) \preceq 0$.

Proof. We prove 1. Assume $\mathbf{x}^*$ is a local minimum, then there exists a ball $B(\mathbf{x}^*, r) \subseteq U$ for which $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in B(\mathbf{x}^*, r)$. Next, let $\mathbf{d} \in \mathbb{R}^n$ be a nonzero vector. For any $0 < \alpha < \frac{r}{\|\mathbf{d}\|}$, we have $\mathbf{x}_\alpha^* \equiv \mathbf{x}^* + \alpha \mathbf{d} \in B(\mathbf{x}^*, r)$ and for any such α, $f(\mathbf{x}_\alpha^*) \geq f(\mathbf{x}^*)$.

On the other hand, by the linear approximation theorem there exists a vector $\mathbf{z}_\alpha \in [\mathbf{x}^*, \mathbf{x}_\alpha^*]$ such that

$$\begin{aligned} f(\mathbf{x}_\alpha^*) - f(\mathbf{x}^*) &= \alpha \nabla f(\mathbf{x}^*)^\top \mathbf{d} + \frac{\alpha^2}{2} \mathbf{d}^T \nabla^2 f(\mathbf{z}_\alpha) \mathbf{d} \\ &= \frac{\alpha^2}{2} \mathbf{d}^T \nabla^2 f(\mathbf{z}_\alpha) \mathbf{d}. \end{aligned}$$

This implies that for any $\alpha \in \left(0, \frac{r}{\|\mathbf{d}\|}\right)$ the inequality $\mathbf{d}^T \nabla^2 f(\mathbf{z}_\alpha) \mathbf{d} \geq 0$ holds. Since $\mathbf{z}_\alpha \to \mathbf{x}^*$ as $\alpha \to 0^+$, we obtain that $\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} \geq 0$ (as $\nabla^2 f$ is continuous so $\nabla^2 f(\mathbf{z}_\alpha) \to \nabla^2 f(\mathbf{x}^*)$), which leads to $\nabla^2 f(\mathbf{x}^*) \succeq 0$. The proof of 2. follows analogously. □

This is a **necessary** condition for optimality. We now state a **sufficient** condition for **strict** local optimality.

Theorem (Sufficient Second Order Optimality Conditions)**.** *Let $f : U \to \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable and that $\mathbf{x}^*$ is a stationary point. Then*

1. $\nabla^2 f(\mathbf{x}^*) \succ 0 \Rightarrow \mathbf{x}^*$ *is a strict local minimum.*
2. $\nabla^2 f(\mathbf{x}^*) \prec 0 \Rightarrow \mathbf{x}^*$ *is a strict local maximum.*

Proof. We prove 1. Let $\mathbf{x}^*$ be a stationary point and suppose that $\nabla^2 f(\mathbf{x}^*) \succ 0$. Since $\nabla^2 f$ is continuous and U is open, there exists a ball $B(\mathbf{x}^*, r) \subseteq U$ for which $\nabla^2 f(\mathbf{x}) \succ 0$ for any $\mathbf{x} \in B(\mathbf{x}^*, r)$. Then, by the Linear Approximation Theorem, there exists a vector $\mathbf{z_x} \in [\mathbf{x}^*, \mathbf{x}]$ (and hence $\mathbf{z_x} \in B(\mathbf{x}^*, r)$) for which

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^T \nabla^2 f(\mathbf{z_x}) (\mathbf{x} - \mathbf{x}^*) \ .$$

Since $\nabla^2 f(\mathbf{z_x}) \succ 0$, it follows that for any $\mathbf{x} \in B(\mathbf{x}^*, r)$ such that $\mathbf{x} \neq \mathbf{x}^*$, the inequality $f(\mathbf{x}) > f(\mathbf{x}^*)$ holds. Therefore $\mathbf{x}^*$ is a strict local minimum point of f.

2. follows by considering $-f$. □

Note that the converses to these theorems are false:

- It is possible to have $\nabla^2 f(\mathbf{x}^*) \succeq 0$ and $\mathbf{x}^*$ not be a local optima.
- It is also possible to have $\mathbf{x}^*$ be a strict local optima and not have $\nabla^2 f(\mathbf{x}^*) \succ 0$.

Saddle Points

Definition (Saddle Point)**.** *Let $f : U \to \mathbb{R}$ be a continuously differentiable function defined on an open set $U \subseteq \mathbb{R}^n$. A stationary point $\mathbf{x}^* \in U$ is called a saddle point of f if it is neither a local minimum point nor a local maximum point of f.*

Theorem (Sufficient Condition for Saddle Points)**.** *Let $f : U \to \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable and that $\mathbf{x}^*$ is a stationary point. If $\nabla^2 f(\mathbf{x}^*)$ is an indefinite matrix, then $\mathbf{x}^*$ is a saddle point of f.*

$$\nabla f(\mathbf{x}^*) = 0 \text{ and } \nabla^2 f(\mathbf{x}^*) \text{ indefinite} \quad \Rightarrow \quad \mathbf{x}^* \text{ is a saddle point.}$$

Proof. Since $\nabla^2 f(\mathbf{x}^*)$ is indefinite, it has at least one positive eigenvalue $\lambda > 0$. Let $\mathbf{v}$ be the corresponding unit eigenvector. Since U is open, there exists a radius $r > 0$ such that $\mathbf{x}^* + \alpha\mathbf{v} \in U$ for any $\alpha \in (0, r)$. By the Quadratic Approximation Theorem (week 1), there exists a function $g : \mathbb{R}_{++} \to \mathbb{R}$ satisfying

$$\frac{g(t)}{t} \to 0 \text{ as } t \to 0\,,$$

such that for any $\alpha \in (0, r)$

$$f\left(\mathbf{x}^* + \alpha\mathbf{v}\right) = f\left(\mathbf{x}^*\right) + \frac{\lambda\alpha^2}{2}\|\mathbf{v}\|^2 + g\left(\|\mathbf{v}\|^2\alpha^2\right)\,.$$

Since $||\mathbf{v}|| = 1$, we write

$$f\left(\mathbf{x}^* + \alpha\mathbf{v}\right) = f\left(\mathbf{x}^*\right) + \frac{\lambda\alpha^2}{2} + g\left(\alpha^2\right)$$

By the properties of g, there exists an $\varepsilon_1 \in (0, r)$ such that

$$\left|\frac{g(t)}{t}\right| < \frac{\lambda}{2} \text{ for } t < \varepsilon_1^2$$

and thus $g\left(\alpha^2\right) > -\frac{\lambda}{2}\alpha^2$ for all $\alpha \in (0, \varepsilon_1)$ and hence $f\left(\mathbf{x}^* + \alpha\mathbf{v}\right) > f\left(\mathbf{x}^*\right)$ for all $\alpha \in (0, \varepsilon_1)$. This shows that $\mathbf{x}^*$ cannot be a local maximum point of f. A similar argument-exploiting an eigenvector of $\nabla^2 f\left(\mathbf{x}^*\right)$ corresponding to a negative eigenvalue-shows that $\mathbf{x}^*$ cannot be a local minimum point of f, establishing the desired result that $\mathbf{x}^*$ is a saddle point. $\square$

Attainment of Minimal/Maximal Points

On compact domains, f attains the global maxima and minima:

Theorem (Weierstrass' Theorem). *Let f be a continuous function defined over a nonempty compact*[2] *set $C \subseteq \mathbb{R}^n$. Then there exists a global minimum point of f over C and a global maximum point of f over C.*

When the underlying set is not compact, Weierstrass theorem does not guarantee the attainment of the solution, but certain properties of the function f can imply attainment of the solution

Definition (Coerciveness). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function over $\mathbb{R}^n$. f is called coercive if*

$$\lim_{\|\mathbf{x}\| \to \infty} f(\mathbf{x}) = \infty.$$

[2] Closed and bounded.

Theorem (Attainment of Global Optima Points for Coercive Functions)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous and coercive function and let $S \subseteq \mathbb{R}^n$ be a nonempty closed set. Then f attains a global minimum point on S.*

Proof. Let $\mathbf{x}_0 \in S$ be an arbitrary point in S. Since the function is coercive, it follows that there exists an $M > 0$ such that

$$f(\mathbf{x}) > f(\mathbf{x}_0) \text{ for any } \mathbf{x} \text{ such that } \|\mathbf{x}\| > M\,.$$

Since any global minimizer $\mathbf{x}^*$ of f over S satisfies $f(\mathbf{x}^*) \leq f(\mathbf{x}_0)$, it follows that the set of global minimizers of f over S is the same as the set of global minimizers of f over $S \cap B[0, M]$. The set $S \cap B[0, M]$ is compact and nonempty, and thus by the Weierstrass theorem, there exists a global minimizer of f over $S \cap B[0, M]$ and hence also over S. □

Global Optimality Conditions

In the previous section we looked at conditions for local optimality. Since these conditions only used local information (gradient and Hessian at the stationary point), they can't imply global optimality. We must use global information to infer global optimality.

Theorem (Global Optimality Condition)**.** *Let f be a twice continuously differentiable function defined over $\mathbb{R}^n$. Suppose*[3] *that $\nabla^2 f(\mathbf{x}) \succeq 0$ for any $\mathbf{x} \in \mathbb{R}^n$. Let $\mathbf{x}^* \in \mathbb{R}^n$ be a stationary point of f. Then $\mathbf{x}^*$ is a global minimum point of f.*

Proof. By the Linear Approximation Theorem, it follows that for any $\mathbf{x} \in \mathbb{R}^n$, there exists a vector $\mathbf{z}_\mathbf{x} \in [\mathbf{x}^*, \mathbf{x}]$ for which

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \nabla^2 f(\mathbf{z}_\mathbf{x})(\mathbf{x} - \mathbf{x}^*)\,.$$

Since $\nabla^2 f(\mathbf{z}_\mathbf{x}) \succeq 0$, we have that $f(\mathbf{x}) \geq f(\mathbf{x}^*)$, establishing the fact that $\mathbf{x}^*$ is a global minimum point of f. □

Quadratic Functions

A **quadratic function** over $\mathbb{R}^n$ is a function of the form

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A}\mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

Recall that

$$\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x} + 2\mathbf{b}$$
$$\nabla^2 f(\mathbf{x}) = 2\mathbf{A}$$

[3] We will study convexity later in the module. But for now, note that this condition is equivalent to saying that f is a convex function. I.e., a twice-differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if $\nabla^2 f(\mathbf{x}) \succeq 0$ for all $\mathbf{x}$.

Proposition. *Let $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c$, with $\mathbf{A} \in \mathbb{R}^{n\times n}$ symmetric, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Then*

1. *$\mathbf{x}$ is a stationary point of f iff $\mathbf{A}\mathbf{x} = -\mathbf{b}$.*
2. *if $\mathbf{A} \succeq 0$, then $\mathbf{x}$ is a global minimum point of f iff $\mathbf{A}\mathbf{x} = -\mathbf{b}$.*
3. *if $\mathbf{A} \succ 0$, then $\mathbf{x} = -\mathbf{A}^{-1}\mathbf{b}$ is a strict global minimum point of f.*

Proof.

1. The proof follows immediately from the formula of the gradient of f.
2. since $\nabla^2 f(\mathbf{x}) = 2\mathbf{A} \succeq 0$, it follows from global optimality conditions that the global minimum points are exactly the stationary points, which combined with part 1. implies the result.
3. When $\mathbf{A} \succ 0$, the vector $\mathbf{x} = -\mathbf{A}^{-1}\mathbf{b}$ is the unique solution to $\mathbf{A}\mathbf{x} = -\mathbf{b}$, and hence by parts 1. and 2., it is the unique global minimum point of f.

□

Checklist

The idea of this checklist is to help you to self-evaluate your progress and understanding of the subject, and to give you some guidance on where to focus. If you can tick all the boxes it means you're doing alright, otherwise you need to study a bit more, grab a book, watch the videos, or seek help from classmates, the lecturers, or the demonstrators. Try to fill as many gaps as quickly as possible.

And remember to do the 

's!

Learning Outcome	Check
I understand the difference between global and local maximum (minimum).	
I understand the difference between strict and non-strict optima.	
I know how to compute stationary points.	
I know the definition of a positive definite matrix.	
I understand the relation between positive definiteness (and variants) and eigenvalues of a matrix.	
I know 3 different ways to check whether a matrix is positive definite.	
I understand $\nabla f(\mathbf{x}) = 0$ is necessary but not enough for optimality, and that I need to check a second-order condition on the Hessian.	
I can see the difference between a necessary and a sufficient optimality condition.	
I can see that a saddle point makes things complicated and I reviewed the video example discussing it.	
I understand the difference between local and global optimality conditions.	
I understand global optimality conditions for quadratic functions.	

Exercises

1. Make a table listing all of the second order necessary and sufficient conditions.
2. Find a counter-example for the converse of Fermat's theorem?
3. Let $\mathbf{A}$ be a symmetric 2×2 matrix. Prove that $\mathbf{A}$ is positive semidefinite (definite) if and only if $\mathrm{Tr}(\mathbf{A}), \det(\mathbf{A}) \geq 0$ ($\mathrm{Tr}(\mathbf{A}), \det(\mathbf{A}) > 0$).
4. Classify the matrices

$$\mathbf{A} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \mathbf{B} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

$$\mathbf{C} = \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix}, \mathbf{D} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0.1 \end{pmatrix}. \mathbf{E} = \begin{pmatrix} 4 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 2 & 4 \end{pmatrix},$$

$$\mathbf{F} = \begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & -1 \end{pmatrix}, \ \mathbf{G} = \begin{pmatrix} -4 & 1 & 1 \\ 1 & -4 & 1 \\ 1 & 1 & -4 \end{pmatrix}.$$

5. Find counter-examples to the converse of the second-order necessary and sufficient optimal condition theorems.
6. Find a counter example for the converse of the sufficient condition for saddle points theorem, i.e., find a function f such that x^* is a saddle point but where $\nabla^2 f(x^*)$ is not indefinite.
7. Find a function f where the optima are not attained.
8. Find and classify the stationary points of the function $f(x_1, x_2) = -2x_1^2 + x_1 x_2^2 + 4x_1^4$ (see Fig. 4).
9. Minimize the function

$$f(x_1, x_2) = x_1^2 + x_2^2$$

over the set $C = \{(x_1, x_2) : x_1 + x_2 \leq -1\}$.

10. Consider the function

$$f(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_1x_3 + x_2x_3 + \left(x_1^2 + x_2^2 + x_3^2\right)^2$$

. Show that $\mathbf{x}^* = 0$ is the global minimum of f.

11. Find the global minimum of

$$f(x, y) = 2x^2 + 2xy + y^2 + x + y + 1.$$

12. Let $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} x + 2\mathbf{b}^T \mathbf{x} + c$ where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Prove that f is coercive if and only if $\mathbf{A} \succ \mathbf{0}$.

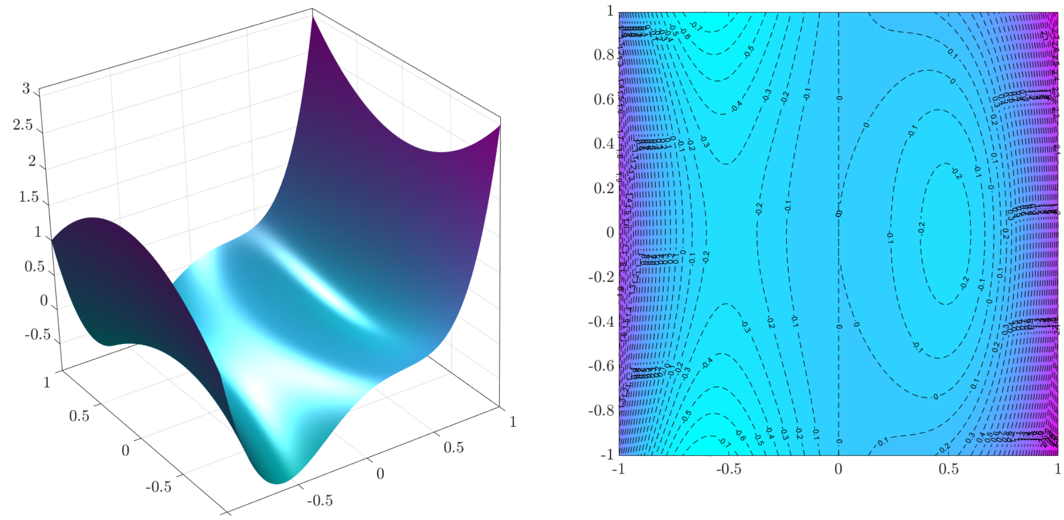

Figure 4: The function $f(x_1, x_2) = -2x_1^2 + x_1x_2^2 + 4x_1^4$ and its contour plot.