

Handout 9 - ECON703 (Fall 2023)

1 Derivatives of Functions with Multiple Arguments

Definition 1.1 (Gradient). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function. The gradient of f at $a \in \mathbb{R}^n$, denoted $\nabla f(a)$, is defined by

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

The gradient generalizes the derivative for functions with multiple arguments.

Later, we will define the **Jacobian matrix** for vector-valued functions. The gradient is its equivalent for scalar-valued functions.

Example 1.1. Consider $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ given by $f(x, y, z) = xy + z$. Then the gradient of f at a point (a, b, c) is

$$\nabla f(a, b, c) = \begin{bmatrix} b \\ a \\ 1 \end{bmatrix}$$

Definition 1.2 (Hessian). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function for which each second-order partial derivative exists. The Hessian of f at $a \in \mathbb{R}^n$, denoted $H_f(a)$ or $D_f^2(a)$, is defined as

$$H_f(a) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(a) & \frac{\partial^2 f}{\partial x_2^2}(a) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(a) & \frac{\partial^2 f}{\partial x_n \partial x_2}(a) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(a) \end{bmatrix}$$

It is the natural analogue of the second derivative for scalar-valued functions.

Example 1.2. Consider $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ given by $f(x, y, z) = xy + z$. Then the Hessian of f at a point (a, b, c) is

$$H_f(a, b, c) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

2 Extrema

Definition 2.1 (Global Extrema of Functions). Let $A \subset \mathbb{R}^n$ and $f : A \rightarrow \mathbb{R}$. We say that $f(\bar{a})$ is a global maximum of f on A if $\bar{a} \in A$ and $\forall x \in A : f(\bar{a}) \geq f(x)$.

Analogously, we say that $f(\underline{a})$ is a global minimum of f on A if $\underline{a} \in A$ and $\forall x \in A : f(\underline{a}) \leq f(x)$.

Definition 2.2 (Global Maximizer/Minimizer of a Function). Let $A \subset \mathbb{R}^n$ and $f : A \rightarrow \mathbb{R}$. We say that $\bar{a} \in A$ maximizes f on A , denoted $\bar{a} \in \arg \max_{x \in A} f(x)$, if $\forall x \in A : f(\bar{a}) \geq f(x)$.

Analogously, we say that $\underline{a} \in A$ minimizes f on A , denoted $\underline{a} \in \arg \min_{x \in A} f(x)$, if $\forall x \in A : f(\underline{a}) \leq f(x)$.

Note that both $\arg \max$ and $\arg \min$ are technically set-valued. However, in an abuse of notation, we often write statements such as $\bar{a} = \arg \max_{x \in A} f(x)$ if the $\arg \max$ is a singleton.

Definition 2.3 (Local Extrema of Functions). Let $A \subset \mathbb{R}^n$ and $f : A \rightarrow \mathbb{R}$. We say that $f(\bar{a})$ is a local maximum of f on A if $\bar{a} \in A$ and $\exists r > 0$ s.t. $\forall x \in B_r(\bar{a}) \cap A : f(\bar{a}) \geq f(x)$.

Analogously, we say that $f(\underline{a})$ is a local minimum of f on A if $\underline{a} \in A$ and $\exists r > 0$ s.t. $\forall x \in B_r(\underline{a}) \cap A : f(\underline{a}) \leq f(x)$.

Definition 2.4 (Local Maximizer/Minimizer of a Function). Let $A \subset \mathbb{R}^n$ and $f : A \rightarrow \mathbb{R}$. We say that $\bar{a} \in A$ locally maximizes f on A , if $\exists r > 0$ s.t. $\forall x \in B_r(\bar{a}) \cap A : f(\bar{a}) \geq f(x)$.

Analogously, we say that $\underline{a} \in A$ locally minimizes f on A , if $\exists r > 0$ s.t. $\forall x \in B_r(\underline{a}) \cap A : f(\underline{a}) \leq f(x)$.

Definition 2.5 (Critical Points / First-Order Conditions). Let $f : A \rightarrow \mathbb{R}$ for $A \subset \mathbb{R}^n$. A point $x^* \in \text{int}(A)$ is called a critical point of f if $\nabla f(x^*) = 0$. $\nabla f(x^*) = 0$ is often called the **first-order condition**.

3 Karush-Kuhn-Tucker Optimization

To make sense of KKT optimization theoretically, we would have to consider the question of **Dual Problems**. For those of you interested in the theoretical background, search for

- Theorem of Lagrange
- Theorem of Kuhn and Tucker
- Primal Problems and Dual Problems
- Duality Gap and Strong Duality

Sundaram, R. (1996). A First Course in Optimization Theory.

Cambridge: Cambridge University Press. doi:10.1017/CBO9780511804526

However, here we will think about how to apply the technique instead.

Definition 3.1 (The KKT Optimization Problem). Let $X \subset \mathbb{R}^n$ be a convex subset of Euclidean space. Let $f : X \rightarrow \mathbb{R}$ be the objective function. Let $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be the function corresponding to the inequality constraints. Let $h : \mathbb{R}^n \rightarrow \mathbb{R}^l$ be the function corresponding to the equality constraints.

The KKT optimization problem is as follows:

$$\max_{x \in X} f(x) \quad \text{subject to: } g(x) \leq 0 \quad \text{and} \quad h(x) = 0$$

We can form the corresponding Lagrangian function as follows:

$$\mathcal{L}(x, \mu, \lambda) = f(x) - \mu'g(x) - \lambda'h(x)$$

Definition 3.2 (The KKT Conditions). The Karush-Kuhn-Tucker conditions for a maximum at x^* are

- **Stationarity:**

$$0 = \nabla f(x^*) - \sum_{j=1}^l \lambda_j \nabla h_j(x^*) - \sum_{i=1}^m \mu_i \nabla g_i(x^*)$$

- **Primal Feasibility:**

$$\forall j = 1, \dots, l : h_j(x^*) = 0 \quad \forall i = 1, \dots, m : g_i(x^*) \leq 0$$

- **Dual Feasibility:**

$$\forall i = 1, \dots, m : \mu_i \geq 0$$

- **Complementary Slackness:**

$$\forall i = 1, \dots, m : \mu_i g_i(x^*) = 0$$

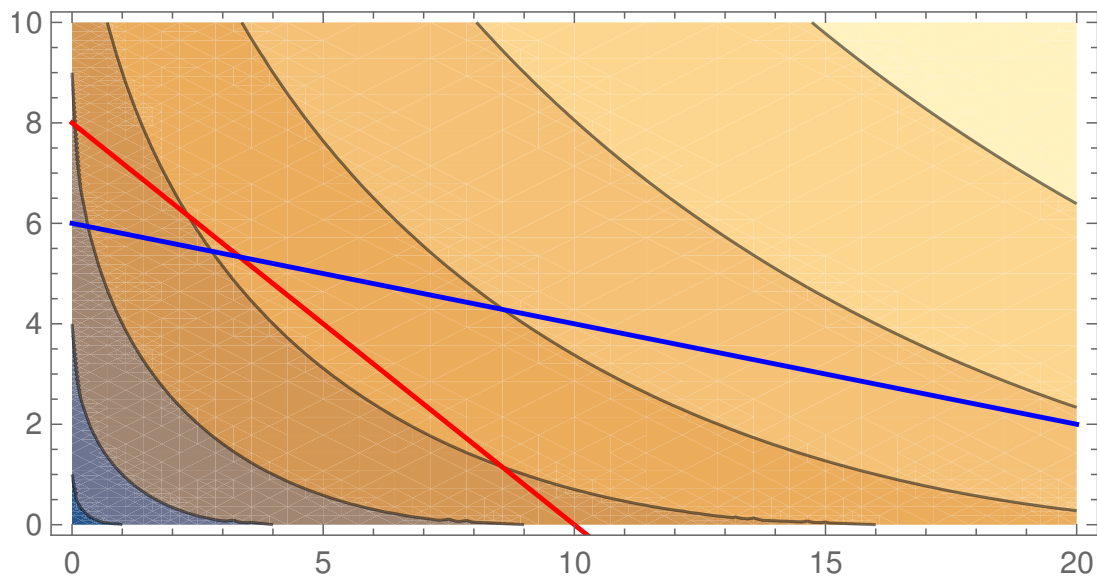


Figure 1: A visualization of a simple KKT problem

Example 3.1. Consider the following optimization problem:

$$\begin{aligned} \max_{(x,y) \in \mathbb{R}^2} \quad & xy \quad \text{s.t.} \quad x \geq 0 \quad \wedge \quad y \geq 0 \\ & \wedge \quad x + y \leq 5 \end{aligned}$$

This gives us the Lagrangian

$$\begin{aligned} \mathcal{L} &= xy - \lambda_1(-x) - \lambda_2(-y) - \lambda_3(x + y - 5) \\ &= xy + \lambda_1 x + \lambda_2 y - \lambda_3(x + y - 5) \end{aligned}$$

Forming the KKT conditions gives us:

$$\begin{array}{ll} \textbf{Stationarity:} & 0 = y + \lambda_1 - \lambda_3, & 0 = x + \lambda_2 - \lambda_3, \\ \textbf{Primal Feasibility:} & x \geq 0, \quad y \geq 0, & x + y \leq 5, \\ \textbf{Dual Feasibility:} & \lambda_1 \geq 0, \quad \lambda_2 \geq 0, & \lambda_3 \geq 0, \\ \textbf{Complementary Slackness:} & \lambda_1 x = 0, \quad \lambda_2 y = 0, & \lambda_3(x + y - 5) = 0. \end{array}$$

With these conditions, we can solve for a candidate.

$$\begin{array}{ll} x \neq 0 \quad \wedge \quad y \neq 0 & \text{since } (x, y) = (1, 1) \text{ is feasible and gives a higher value.} \\ \implies \lambda_1 = 0 \quad \wedge \quad \lambda_2 = 0 & \text{by complementary slackness} \\ \implies y = \lambda_3 \quad \wedge \quad x = \lambda_3 & \text{by stationarity} \\ \implies x = y = \lambda_3 > 0 & \\ \implies x + y = 5 & \text{by complementary slackness} \\ \implies x = y = 2.5 & \end{array}$$

So $(x, y) = (2.5, 2.5)$ is our solution candidate.

Definition 3.3 (Constraint Qualification). There are several conditions called **constraint qualifications** that tell us settings in which an optimizer must satisfy the KKT conditions. They vary in complexity and applicability and tell us when **strong duality** holds.

Wikipedia: KKT regularity conditions

Three common constraint qualifications are the following:

Definition 3.4 (Linear Constraint Qualification). If all constraints are affine functions, no further criteria need to be met.

Definition 3.5 (Linear Independence Constraint Qualification). The gradients of the active inequality constraints (i.e. those binding in x^*) and the gradients of the equality constraints are linearly independent at x^* .

Definition 3.6 (Slater's Condition). For a convex problem, i.e., minimizing a convex function (maximizing a concave function) under convex inequality and linear equality constraints, there exists a point such that $h(x) = 0$ and $g_i(x) < 0$. In other words, the feasible region must have an interior point.

Example 3.2. In the previous example, all constraints were affine, so a solution to the maximization problem must satisfy the KKT conditions by the linear constraint qualification. To verify that our solution candidate is the maximizer, note that every feasible point satisfies $x + y \leq 5$ and therefore $xy \leq \left(\frac{x+y}{2}\right)^2 \leq \left(\frac{5}{2}\right)^2$.

A good source for the intuition behind KKT optimization is this Arizona Math Camp video. I highly recommend this video by the channel “**Arizona Math Camp**”.