

Lecture 25

Optimality Conditions

25.1 Lecture Objectives

- Define local and global minimizers.
- Understand the first order conditions for local minimizers in both unconstrained and constrained optimization problems.
- Understand the second order conditions for local minimizers.

25.2 Local and Global Minimizers

25.2.1 Local minimizers

Consider an optimization problem

$$\min_{\mathbf{x}} f(\mathbf{x}), \text{ s.t. } \mathbf{x} \in \Omega \quad (25.1)$$

where Ω defines the constraint set for $\mathbf{x}$ (ie: either the entire $\mathbb{R}^N$ for an unconstrained problem, or some subset determined by equality or inequality constraints otherwise). A point $\mathbf{x}^*$ is a local minimizer if there exists an $\epsilon > 0$ such that:

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \text{ for all } \mathbf{x} \in \Omega \text{ such that } \|\mathbf{x} - \mathbf{x}^*\| < \epsilon \quad (25.2)$$

Further, if the inequality above is a strict inequality, then $\mathbf{x}^*$ is a *strict* local minimizer, ie:

$$f(\mathbf{x}^*) < f(\mathbf{x}), \text{ for all } \mathbf{x} \in \Omega \text{ such that } \|\mathbf{x} - \mathbf{x}^*\| < \epsilon \quad (25.3)$$

25.2.2 Global minimizers

A point $\mathbf{x}^*$ is a global minimizer if:

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \text{ for all } \mathbf{x} \in \Omega \quad (25.4)$$

Further, if the inequality above is a strict inequality, then $\mathbf{x}^*$ is a *strict* global minimizer, ie:

$$f(\mathbf{x}^*) < f(\mathbf{x}), \text{ for all } \mathbf{x} \in \Omega \quad (25.5)$$

25.3 Optimality Conditions for Various Problems

25.3.1 Unconstrained Optimization

The gradient $\nabla f(\mathbf{x})$ is defined as:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_N} \end{bmatrix} \quad (25.6)$$

A necessary condition for a minimizer $\mathbf{x}^*$ of an unconstrained optimization problem ($\min_{\mathbf{x}} f(\mathbf{x})$) with a defined gradient $\nabla f(\mathbf{x})$ is as follows:¹

$$\nabla f(\mathbf{x}^*) = \mathbf{0} \quad (25.7)$$

Note that this is just a necessary condition (assuming the gradient exists everywhere), but it is not sufficient. For instance, a point that satisfies this zero gradient constraint may be a saddle point, or a local maximizer.

The Hessian $\nabla^2 f(\mathbf{x})$ is defined as:

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_N} \\ \vdots & & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_N \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_N^2} \end{bmatrix} \quad (25.8)$$

A sufficient set of conditions condition for a minimizer (given a twice differentiable f) is that: i) the gradient is zero (as described above), and ii) the Hessian is positive definite matrix, ie: a matrix such that:

$$\mathbf{d}^T [\nabla^2 f(\mathbf{x})] \mathbf{d} > 0 \quad (25.9)$$

for any non-zero $\mathbf{d} \in \mathbb{R}^N$.

In the one-dimensional case, these conditions reduce to: i) derivative equal to zero, and ii) second derivative greater than zero. These concepts are summarized for one-dimensional optimization (ie: where $\mathbf{x}$ is a scalar) in figure 25.1.

¹For functions that may not have a defined gradient, optimality conditions can be written in terms of the “subgradient”. See, for example, <https://www.stat.cmu.edu/~ryantibs/convexopt-F16/scribes/sg-method-scribed.pdf>.

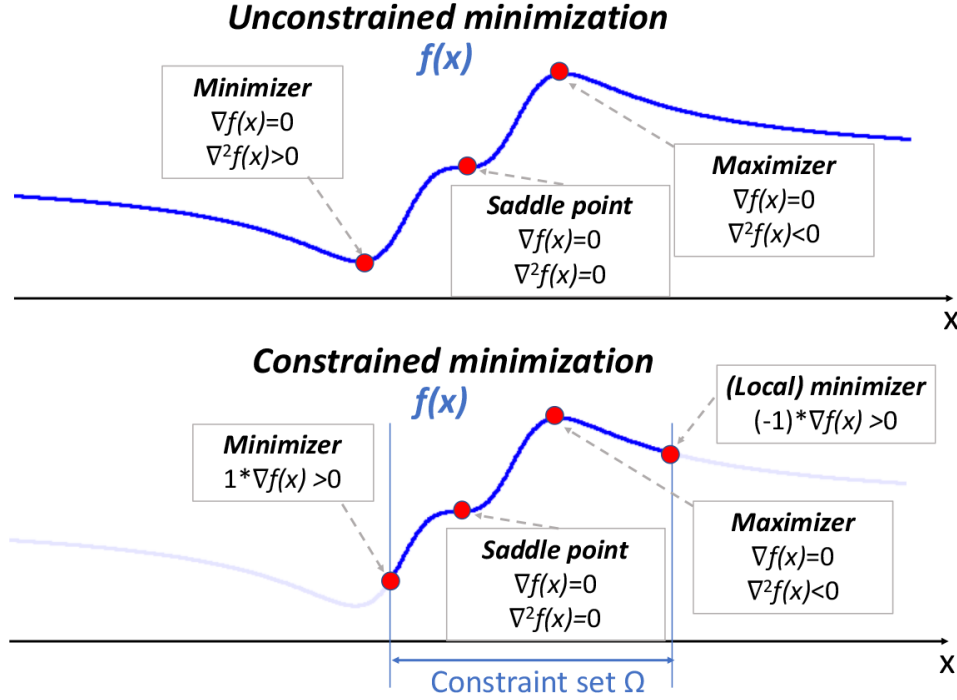


Figure 25.1: One-dimensional examples of unconstrained and constrained optimization, with various minimizers, a saddle point, and a maximizer.

25.3.2 Constrained Optimization

In the presence of constraints, the conditions above are not necessary for a local minimizer.

In this case, a necessary condition for a local minimizer is that, for any feasible direction $\mathbf{d}$ (ie: a direction in which we do not immediately exit the feasible set Ω):

$$\mathbf{d}^T \nabla f(\mathbf{x}^*) \geq 0 \quad (25.10)$$

In other words, at a local minimizer $\mathbf{x}^*$ there is “no way but up” (or constant) within the feasible set. This condition is depicted graphically for a 1D example in figure 25.1, and for a 2D example in figure 25.2.

Similarly to the unconstrained case, a second order condition is that, for any feasible direction $\mathbf{d}$, if $\mathbf{d}^T \nabla \mathbf{x}^* = 0$, then:

$$\mathbf{d}^T [\nabla^2 f(\mathbf{x}^*)] \mathbf{d} > 0 \quad (25.11)$$

25.3.3 Karush-Kuhn-Tucker Conditions

For general optimization problems including equality and inequality constraints

$$\begin{aligned} & \text{minimize } f(\mathbf{x}) \\ & \text{such that } g_k(\mathbf{x}) \leq b_k, \text{ for } k = 1, \dots, K \\ & \text{such that } h_l(\mathbf{x}) = c_l, \text{ for } l = 1, \dots, L \end{aligned} \quad (25.12)$$

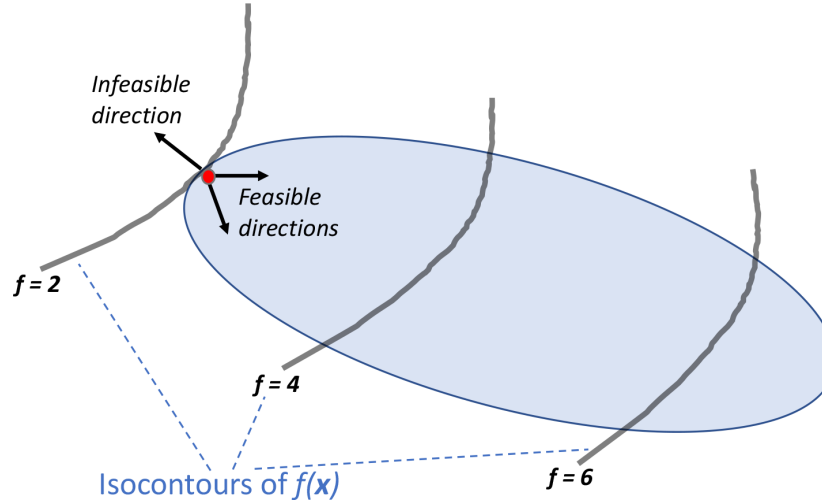


Figure 25.2: Geometric interpretation of the solution to a constrained optimization problem. In this case, the minimizer $\mathbf{x}^*$ is marked by the red dot, with a value $f(\mathbf{x}^*) = 2$.

necessary conditions for a point $\mathbf{x}^*$ to be a minimizer were described by W. Karush in his (unpublished) master's thesis in 1939, and later published by H. W. Kuhn and A. W. Tucker in 1951. These KKT conditions are as follows:

$$\begin{aligned}
 \nabla \left(f(\mathbf{x}^*) + \sum_{k=1}^K u_k g_k(\mathbf{x}^*) + \sum_{l=1}^L v_l h_l(\mathbf{x}^*) \right) &= 0 \text{ (stationarity condition)} \\
 u_k (g_k(\mathbf{x}^*) - b_k) &= 0, \text{ for } k = 1, \dots, K \text{ (complementary slackness)} \\
 g_k(\mathbf{x}^*) &\leq b_k, \text{ for } k = 1, \dots, K \text{ (primal feasibility: inequality conditions)} \\
 h_l(\mathbf{x}^*) &= c_l, \text{ for } l = 1, \dots, L \text{ (primal feasibility: equality conditions)} \\
 u_k &\geq 0, \text{ for } k = 1, \dots, K \text{ (dual feasibility)}
 \end{aligned} \tag{25.13}$$

where ∇ denotes the gradient with respect to $\mathbf{x}$. Note that these conditions assume the existence of the gradient². A few notes for intuitive interpretation of the KKT conditions:

- For unconstrained problems, the KKT conditions reduce to the familiar condition $\nabla(f(\mathbf{x}^*)) = 0$
- For problems including linear equality constraints $\mathbf{H}\mathbf{x} = \mathbf{c}$, the KKT conditions reduce to

$$\begin{aligned}
 \nabla (f(\mathbf{x}^*) + \mathbf{v}^T \mathbf{H}\mathbf{x}^*) &= 0 \text{ (stationarity condition, satisfied for some vector } \mathbf{v}) \\
 \mathbf{H}\mathbf{x} &= \mathbf{c} \text{ (primal feasibility: equality conditions)}
 \end{aligned}$$

²If the gradient does not exist, the condition that the gradient at $\mathbf{x}^*$ be equal to zero can be rewritten in terms of the subgradient at $\mathbf{x}^*$ containing the value zero.

(25.14)

which can be viewed as a solution $\mathbf{x}^*$ that satisfies the equality constraints, and such that the gradient of $f(\mathbf{x}^*)$ is orthogonal to the constraint set (i.e., $\nabla f(\mathbf{x}^*) + \mathbf{H}^T \mathbf{v} = 0$). Note that, for any matrix $\mathbf{H}$, the range space of $\mathbf{H}^T$ is orthogonal to the null space of $\mathbf{H}$. In other words, for any choice of $\mathbf{v}$, the vector $\mathbf{H}^T \mathbf{v}$ will be orthogonal to the space of vectors $\mathbf{x}$ that satisfy the constraint $\mathbf{H}\mathbf{x} = \mathbf{c}$ (in the sense that for any two vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ that satisfy the constraint, their difference $\mathbf{x}_1 - \mathbf{x}_2$ will be in the null space of $\mathbf{H}$, and so $\mathbf{H}^T \mathbf{v}$ will be orthogonal to $\mathbf{x}_1 - \mathbf{x}_2$).

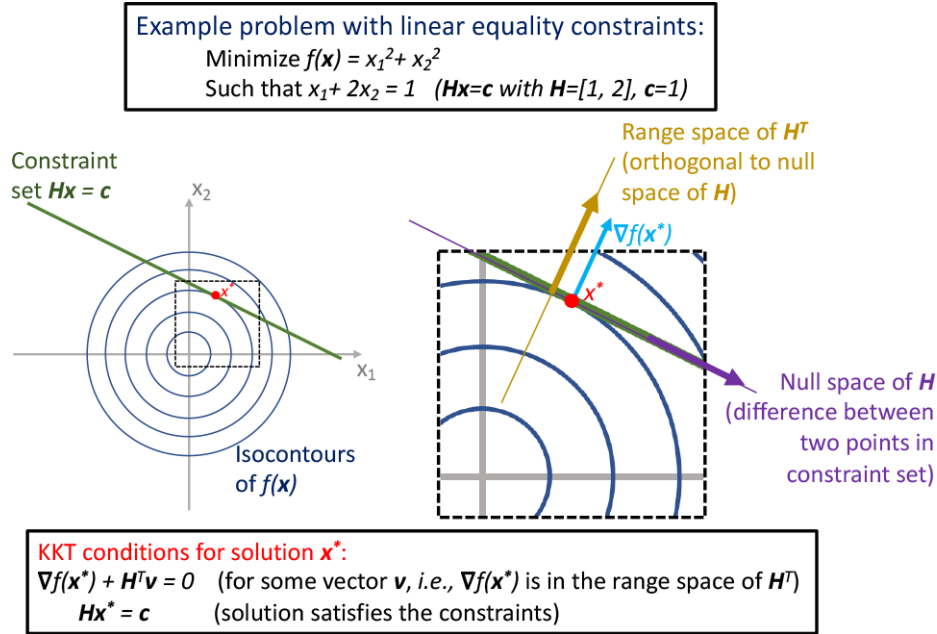


Figure 25.3: Representation of the KKT conditions for an optimization problem with linear equality constraints.

- For problems including inequality constraints: for each constraint $g_k(\mathbf{x}) \leq b_k$, a solution $\mathbf{x}^*$ will either be in the interior or at the boundary of the set defined by this constraint. If the solution is in the interior of the set, the inequality constraint does not impose any conditions on the solution (other than feasibility). If the solution is at the boundary of the constraint set (i.e., the constraint is “active”), then the KKT conditions require that the gradient of the cost function $\nabla f(\mathbf{x}^*)$ be anti-parallel to the gradient of the constraint function $\nabla g_k(\mathbf{x}^*)$ (assuming a single constraint, i.e., $-\nabla f(\mathbf{x}^*) = u_k \nabla g_k(\mathbf{x}^*)$). If there are multiple active constraints, then the negative of the cost function gradient $-\nabla f(\mathbf{x}^*)$ needs to be expressible as a linear combination of the gradients of the active constraints $\nabla g_k(\mathbf{x}^*)$, with non-negative coefficients, i.e. $-\nabla f(\mathbf{x}^*) = \sum_k u_k \nabla g_k(\mathbf{x}^*)$ with $u_k \geq 0$. Intuitively, the direction of steepest descent of the cost function ($-\nabla f(\mathbf{x}^*)$) needs to be “shooting straight out” of the constraint set.

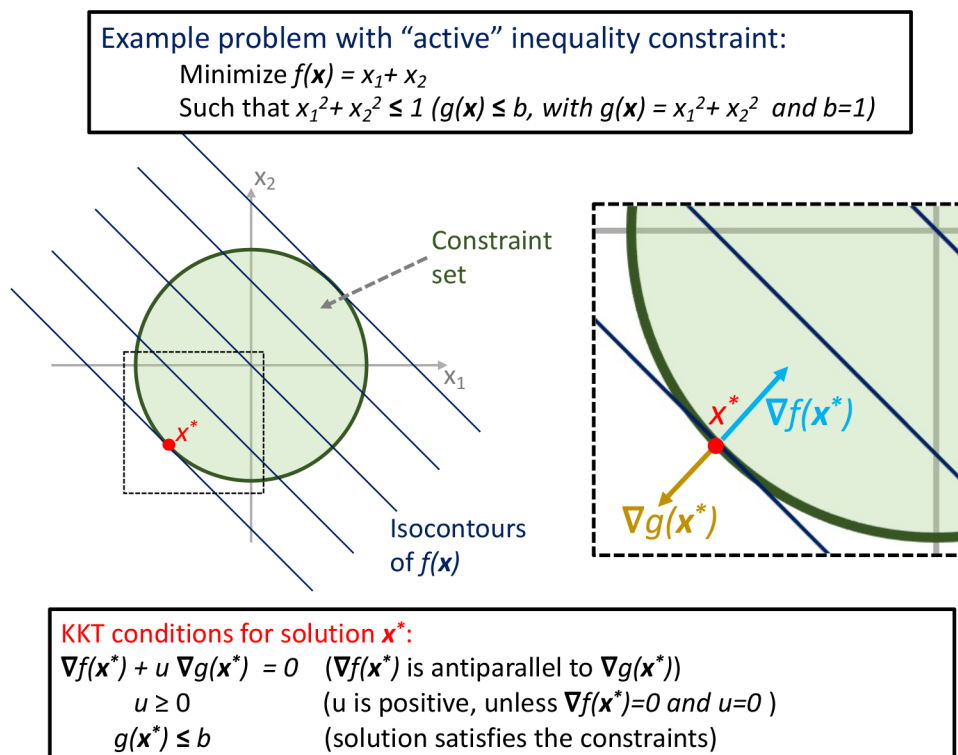


Figure 25.4: Representation of the KKT conditions for an optimization problem with inequality constraints.

25.3.4 Brief Historical Perspective

The idea that maxima and minima can be found at locations where the gradient is zero (ie: where the tangent is flat) is quite old, and was formulated by Pierre de Fermat in his treatise entitled “Methodus ad Disquirendam Maximam et Minimam” (Fermat, 1637)³.

³See “Method for the Study of Maxima and Minima”, <https://science.larouchepac.com/fermat/fermat-maxmin.pdf> for an English translation from the original Latin