

KKT conditions for general nonlinear optimization

These short notes are about the most general problem we have ever considered in this course, namely this nonlinear constrained optimization problem over $x \in \mathbb{R}^n$:

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && g_i(x) = 0, \quad i = 1, \dots, \ell \\ & && h_i(x) \geq 0, \quad i = 1, \dots, m \end{aligned}$$

These notes supplement section 14.5 of the textbook,¹ where the KKT conditions are given but only for equality or inequality constraints separately.

Define $g(x)$ to be the vector formed from $g_1(x), \dots, g_\ell(x)$, and $h(x)$ from $h_1(x), \dots, h_m(x)$. In these terms we can write

$$(1) \quad \begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && g(x) = 0 \\ & && h(x) \geq 0 \end{aligned}$$

We will need two definitions from section 14.5. All functions f, g_i, h_i are assumed to be as differentiable as needed.

Definition. (i) For a feasible point x , let $\tilde{h}(x)$ be the vector of length \tilde{m} formed from the active constraints $h_i(x)$ at x , i.e. for which $h_i(x) = 0$.

(ii) A feasible point x_* is a regular point of the constraints if the matrix

$$[\nabla g_1(x_*) \quad \dots \quad \nabla g_\ell(x_*) \quad \nabla \tilde{h}_1(x_*) \quad \dots \quad \nabla \tilde{h}_{\tilde{m}}(x_*)]$$

has linearly independent columns. This matrix has n rows and $\ell + \tilde{m}$ columns.

In definition (ii), note that the gradient of a scalar-valued function is a column vector.

Conceptually, at a regular point each active constraint provides new information; there are no redundancies. The inactive inequality constraints are ignored in this definition. The definition of a regular point is called the *linearly independent constraint qualification (LIQC)* in some books.

The Lagrangian for problem (1) is

$$\begin{aligned} \mathcal{L}(x, \lambda, \mu) &= f(x) - \sum_{i=1}^{\ell} \lambda_i g_i(x) - \sum_{j=1}^m \mu_j h_j(x) \\ &= f(x) - \lambda^\top g(x) - \mu^\top h(x) \end{aligned}$$

where $\lambda \in \mathbb{R}^\ell$ and $\mu \in \mathbb{R}^m$ are column vectors. In the second form we regard $g(x)$ and $h(x)$ as column vectors. Note that the full $h(x)$, not $\tilde{h}(x)$, is used here.

¹Griva, Nash, and Sofer, *Linear and Nonlinear Optimization*, 2nd ed., SIAM Press 2009.

The following KKT theorem² states the first-order necessary conditions.

Theorem (Karush-Kuhn-Tucker, 1939 & 1951). *Suppose x_* is a local minimizer of problem (1), and assume it is a regular point of the constraints. Then there exist vectors $\lambda_* \in \mathbb{R}^\ell$ and $\mu_* \in \mathbb{R}^m$ so that*

$$\begin{aligned} \nabla_x \mathcal{L}(x_*, \lambda_*, \mu_*) &= 0 && \text{stationarity} \\ g(x_*) &= 0 && \text{primal feasibility: equality constraints} \\ h(x_*) &\geq 0 && \text{primal feasibility: inequality constraints} \\ \mu_* &\geq 0 && \text{dual feasibility} \\ \mu_*^\top h(x_*) &= 0 && \text{complementary slackness} \end{aligned}$$

The stationarity condition can be expanded:

$$\begin{aligned} \nabla f(x_*) &= \nabla g(x_*)\lambda_* + \nabla h(x_*)\mu_* \\ &= (\lambda_*)_1 \nabla g_1(x_*) + \cdots + (\lambda_*)_\ell \nabla g_\ell(x_*) \\ &\quad + (\mu_*)_1 \nabla h_1(x_*) + \cdots + (\mu_*)_m \nabla h_m(x_*) \end{aligned}$$

This says that the gradient of f at the solution x_* can be written as a linear combination of the gradients of the constraints. Some of the inequality constraints can be inactive, however, so complementary slackness says that the corresponding multipliers μ_i are zero. Thus, in fact, the $\nabla f(x_*)$ is a linear combination of the gradients of the *active* constraints. The fact that x_* is a regular point implies that the linear combination is unique, thus that the Lagrange multipliers λ_*, μ_* are unique.

It will be helpful to see one example with both equality and inequality constraints. The example has two inequality constraints, so that we can see complementary slackness in action.

Example.

$$\begin{aligned} \text{minimize} \quad & f(x) = \frac{1}{2}(x_1^2 + x_2^2) \\ \text{subject to} \quad & x_1 + x_2 = 1 \\ & x_1 \geq 1 \\ & x_2 \geq -1 \end{aligned}$$

Here $g_1(x) = x_1 + x_2 - 1$, $h_1(x) = x_1 - 1$, and $h_2(x) = x_2 + 1$.

The feasible set is a closed line segment between the points $(1, 0)$ and $(2, -1)$ in the x_1, x_2 plane. The objective is essentially the square of the distance to the origin. Thus, in this simple example problem, $x_* = (1, 0)$ is the solution. At this point the first inequality constraint is active while the second is not.

The Lagrangian is

$$\mathcal{L}(x, \lambda, \mu) = \frac{1}{2}(x_1^2 + x_2^2) - \lambda(x_1 + x_2 - 1) - \mu_1(x_1 - 1) - \mu_2(x_2 + 1),$$

²en.wikipedia.org/wiki/Karush-Kuhn-Tucker_conditions

with $\lambda \in \mathbb{R}^1$ and $\mu \in \mathbb{R}^2$. The KKT theorem gives 8 scalar facts that must be true at the solution:

$$\begin{array}{ll}
 x_1 - \lambda_1 - \mu_1 = 0 & \textit{stationarity} \\
 x_2 - \lambda_1 - \mu_2 = 0 & \\
 x_1 + x_2 - 1 = 0 & \textit{primal feasibility: equality constraint} \\
 x_1 - 1 \geq 0 & \textit{primal feasibility: inequality constraints} \\
 x_2 + 1 \geq 0 & \\
 \mu_1 \geq 0 & \textit{dual feasibility} \\
 \mu_2 \geq 0 & \\
 \mu_1(x_1 - 1) + \mu_2(x_2 + 1) = 0 & \textit{complementary slackness}
 \end{array}$$

Since $x = (1, 0)$ is the solution, and only the first inequality constraint is active, it follows that $\mu_2 = 0$. The complementary slackness condition is $\mu_1(x_1 - 1) = 0$, so $x_1 = 1$. (We will check that $\mu_1 > 0$ at the solution, so this logic is correct.) The system of remaining *equations* is now

$$\begin{array}{l}
 1 - \lambda_1 - \mu_1 = 0 \\
 x_2 - \lambda_1 = 0 \\
 x_2 = 0
 \end{array}$$

This is a set of 3 equations for x_2, λ_1, μ_1 . It is easy to see $x_2 = 0$, $\lambda_1 = 0$, and $\mu_1 = 1$. That is,

$$(x_1, x_2, \lambda_1, \mu_1, \mu_2) = (1, 0, 0, 1, 0).$$

It is easy to check that with these values all 8 KKT conditions hold. In particular, complementary slackness says $1 \cdot 0 + 0 \cdot 1 = 0$, and strict complementarity holds.