The Karush-Kuhn-Tucker Conditions

We'll be looking at nonlinear optimization with constraints:

$$\begin{array}{ll} \text{maximize} & f(x_1, \dots x_n) \\ \text{subject to} & g_i(x_1, \dots x_n) \le b_i \text{ for } i = 1 \dots m \end{array}$$

The text does both minimize and maximize, but it's simpler just to say we'll make any minimize problem into a maximize problem.

We'll start with an example:

$$\begin{array}{lll} \text{maximize} & f(x_1, x_2) & = x_1 + x_2 \\ \text{subject to} & g_1(x_1, x_2) & = x_1^2 + x_2^2 \le b_1 = 2 \end{array}$$

The feasible region is a disk of radius $\sqrt{2}$ centred at the origin. The global maximum (which is the only local maximum) is at $\mathbf{p}_0 = (1, 1)$. Suppose you're at some other point. How can you tell it's not a local maximum? Because there's some direction you can move that increases f and stays within the feasible region. If you're at a local maximum you can't do that.

Case 1: From a point $\mathbf{p}$ in the interior of the disk, you can go in the direction of the gradient $\nabla f(\mathbf{p})$. As long as that gradient is not $\mathbf{0}$, f increases in that direction. On the other hand, if there was a point $\mathbf{p}$ with $\nabla f(\mathbf{p}) = 0$ we might have a local maximum there.

Case 2: From a point $\mathbf{p}$ on the circle, you might not be able to go in the direction of the gradient, but you can go in the direction of some vector $\mathbf{v}$ that points into the circle. In order for f to increase in that direction, we want $\mathbf{v} \cdot \nabla f(p) > 0$. In order to make sure the vector points into rather than out of the circle, we want $\mathbf{v} \cdot \nabla g_1(p) < 0$.

At the maximum $\mathbf{p}_0$, there's no such $\mathbf{v}$. Why not? $\nabla f(\mathbf{p}_0) = (1, 1)$ and $\nabla g_1(\mathbf{p}_0) = (2, 2) = 2\nabla f(\mathbf{p}_0)$. Clearly if $\nabla f(\mathbf{p}) = \lambda \nabla g(\mathbf{p})$ with $\lambda \ge 0$, there can't be a vector $\mathbf{v}$ with $\mathbf{v} \cdot \nabla f(p) > 0$ and $\mathbf{v} \cdot \nabla g_1(p) < 0$. And this is the only way it can happen: if there is no vector $\mathbf{v}$ with $\mathbf{v} \cdot \nabla f(\mathbf{p}) > 0$ and $\mathbf{v} \cdot \nabla g_1(\mathbf{p}) \le 0$, $\nabla f(\mathbf{p})$ must be $\lambda \nabla g_1(\mathbf{p})$ for some $\lambda \ge 0$.

You may have noticed a slight change in the last paragraph: I started with $\mathbf{v} \cdot \nabla g_1(p) < 0$ and then changed that $<$ to $\le$. In this case, the justification is this: if there was a vector $\mathbf{v}$ with $\mathbf{v} \cdot \nabla f(\mathbf{p}) > 0$ and $\mathbf{v} \cdot \nabla g_1(\mathbf{p}) = 0$ you could move it a little (at least if $\nabla g_1(\mathbf{p}) \ne 0$) to make $\mathbf{v} \cdot \nabla g_1(\mathbf{p}) > 0$ and still have $\mathbf{v} \cdot \nabla f(\mathbf{p}) > 0$. On the other hand, we could be in trouble in other examples if $\nabla g_1(\mathbf{p}) = 0$, because then you couldn't use $\nabla g_1(\mathbf{p})$ to tell you whether a certain direction goes into the feasible set or not. This slight quibble is going to re-emerge when we talk about "constraint qualification".

We can combine the two cases: for a local maximum we need $\nabla f(\mathbf{p}) = \lambda \nabla g_1(\mathbf{p})$ with $\lambda \ge 0$ and $\lambda(b_1 - g_1(\mathbf{p})) = 0$. This might remind you of a complementary slackness condition.

What if there's more than one constraint? Let's add the constraint $g_2(x_1, x_2) = x_1 \le b_2 = 0$. Now the maximum is at $(0, \sqrt{2})$.

How can we tell $(0, \sqrt{2})$ is a maximum? This is a point $\mathbf{p}_1$ where both $g_1(\mathbf{p}_1) = b_1$ and $g_2(\mathbf{p}_1) = b_2$; $\nabla f(\mathbf{p}_1) = (1,1)$, $\nabla g_1(\mathbf{p}_1) = (0, 2\sqrt{2})$ and $\nabla g_2(\mathbf{p}_1) = (1,0)$. Could there be a vector $\mathbf{v}$ with $\mathbf{v} \cdot \nabla f(\mathbf{p}_1) > 0$, $\mathbf{v} \cdot \nabla g_1(\mathbf{p}_1) \le 0$ and $\mathbf{v} \cdot \nabla g_2(\mathbf{p}_1) \le 0$? No, because $\nabla f(\mathbf{p}_1) = \frac{1}{2\sqrt{2}} \nabla g_1(\mathbf{p}_1) + \nabla g_2(\mathbf{p}_2)$.

On the other hand, $\mathbf{p}_2 = (0, -\sqrt{2})$ also has $g_1(\mathbf{p}_2) = b_1$ and $g_2(\mathbf{p}_2) = b_2$; but $\nabla f(\mathbf{p}_2) = (1,1)$, $\nabla g_1(\mathbf{p}_2) = (0, -2\sqrt{2})$ and $\nabla g_2(\mathbf{p}_2) = (1,0)$. There is a vector $\mathbf{v}$ in this case, e.g. $(0,1)$, so $\mathbf{p}_2$ is not a maximum. Notice that you can't write $\nabla f(\mathbf{p}_2)$ as a linear combination of $\nabla g_1(\mathbf{p}_2)$ and $\nabla g_2(\mathbf{p}_2)$ with coefficients ≥ 0.

Theorem: Suppose $\mathbf{a}_1, \ldots \mathbf{a}_m$ and $\mathbf{c}$ are vectors in $\mathbf{R}^n$. Then the following are equivalent:

(a): there are no vectors $\mathbf{x}$ with $\mathbf{x} \cdot \mathbf{c} > 0$ and all $\mathbf{x} \cdot \mathbf{a}_i \le 0$

(b): There are $\lambda_1, \ldots \lambda_m$ with $\mathbf{c} = \lambda_1 \mathbf{a}_1 + \ldots \lambda_m \mathbf{a}_m$ and all $\lambda_i \ge 0$.

Proof: Consider the linear programming problem P:

$$
\begin{array}{ll}
\text{maximize} & z = \mathbf{x} \cdot \mathbf{c} \\
\text{subject to} & \mathbf{x} \cdot \mathbf{a}_i \le 0 \text{ for all } i \\
& \text{all } x_j \text{ URS}
\end{array}
$$

This is certainly feasible ($\mathbf{x} = 0$ satisfies the constraints). There are two possibilities:

(i) (a) is true, and P has an optimal solution: the optimal value is 0.

(ii) (a) is false, and P is unbounded (because if $\mathbf{x}$ satisfies (a), so does $2\mathbf{x}$ with a larger value of z).

By duality, in case (i) the dual problem D also has an optimal solution, while in case (ii) D is infeasible. But D is this:

$$
\begin{array}{ll}
\text{minimize} & 0 \\
\text{subject to} & \displaystyle\sum_i y_i \mathbf{a}_i = \mathbf{c} \\
& \textit{all } y_i \ge 0
\end{array}
$$

In case (i), an optimal solution of D has $y_i = \lambda_i$ satisfying (b). In case (ii), saying D is infeasible just says no such λ_i exist.

Theorem: Suppose the problem

$$
\begin{array}{ll}
\text{maximize} & f(\mathbf{x}) \\
\text{subject to} & g_i(\mathbf{x}) \le b_i \text{ for } i = 1 \ldots m
\end{array}
$$

has a local maximum at $\mathbf{x} = \mathbf{p}$, and that a constraint qualification (to be specified) is satisfied at $\mathbf{p}$. Then there are $\lambda_1, \ldots \lambda_m$ such that

$$
\begin{aligned}
\nabla f(\mathbf{p}) - \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{p}) &= 0 \\
\lambda_i (b_i - g_i(\mathbf{p})) &= 0, \; i = 1, \ldots, m \\
\lambda_i &\ge 0, \; i = 1, \ldots, m \\
g_i(\mathbf{p}) &\le b_i, \; i = 1, \ldots, m
\end{aligned}
$$

Those equations (the first is really n, one for each coordinate) and inequalities are called the Karush-Kuhn-Tucker (KKT) conditions. Note that I'm including the inequalities $g_i(\mathbf{p}) \le b_i$ of the problem itself as part of the KKT conditions, just to make sure we don't forget them. Also, if we require $x_i \ge 0$, we treat that as just one other constraint (in the form $-x_i \le 0$), rather than have a special version of the KKT conditions as the text does.

We can also deal with equality constraints as well as inequalities, with the following modification: for an equality constraint $g_i(\mathbf{x}) = b_i$, of course we require $g_i(\mathbf{x}) = b_i$, but we don't care about the sign of the corresponding λ_i.

Worked Example:

$$
\begin{array}{lll}
\text{maximize} & f(x_1,x_2) = (x_1-1)^4 + (x_2-2)^2 & \\
\text{subject to} & g_1(x_1,x_2) = x_1 + x_2 & \le 2 \\
 & g_2(x_1,x_2) = -x_1 + x_2 & \le 2 \\
 & g_3(x_1,x_2) = x_1 - x_2 & \le 2 \\
 & g_4(x_1,x_2) = -x_1 - x_2 & \le 2
\end{array}
$$

Write the KKT conditions and show that $\mathbf{p}_1 = (2,0)$ satisfies them, but $\mathbf{p}_2 = (0,2)$ doesn't.

$$
\begin{aligned}
4(x_1-1)^3 &= \lambda_1 - \lambda_2 + \lambda_3 - \lambda_4 \\
2(x_2-2) &= \lambda_1 + \lambda_2 - \lambda_3 - \lambda_4 \\
\lambda_1(2 - x_1 - x_2) &= 0 \\
\lambda_2(2 + x_1 - x_2) &= 0 \\
\lambda_3(2 - x_1 + x_2) &= 0 \\
\lambda_4(2 + x_1 + x_2) &= 0 \\
x_1 + x_2 &\le 2 \\
-x_1 + x_2 &\le 2 \\
x_1 - x_2 &\le 2 \\
-x_1 - x_2 &\le 2 \\
\lambda_1, \lambda_2, \lambda_3, \lambda_4 &\ge 0
\end{aligned}
$$

For $\mathbf{p}_1$ we have $g_1(\mathbf{p}_1) = g_3(\mathbf{p}_1) = 2$ while $g_2(\mathbf{p}_1) < 2$ and $g_4(\mathbf{p}_1) < 2$, so $\lambda_2 = \lambda_4 = 0$. The first two equations then say

$$
\begin{aligned}
4 &= \lambda_1 + \lambda_3 \\
-4 &= \lambda_1 - \lambda_3
\end{aligned}
$$

The solution of these is $\lambda_1 = 0$, $\lambda_3 = 4$, and these are both ≥ 0.

For $\mathbf{p}_2$ we have $g_1(\mathbf{p}_2) = g_2(\mathbf{p}_2) = 0$ while $g_3(\mathbf{p}_2) < 2$ and $g_4(\mathbf{p}_2) < 2$, so $\lambda_3 = \lambda_4 = 0$. The first two equations say

$$
\begin{aligned}
-4 &= \lambda_1 - \lambda_2 \\
0 &= \lambda_1 + \lambda_2
\end{aligned}
$$

The only solution of these is $\lambda_1 = -2$, $\lambda_2 = 2$. Since $-2 < 0$, we can't satisfy the KKT conditions here.

One of several possible constraint qualifications is the Linear Independence Constraint Qualification (LICQ). A constraint $g_i(\mathbf{x}) \leq b_i$ is said to be **binding** at $\mathbf{x} = \mathbf{p}$ if $g_i(\mathbf{p}) = b_i$. We say the LICQ holds at $\mathbf{x} = \mathbf{p}$ if the gradients of the g_i for the constraints that are binding at $\mathbf{p}$ are linearly independent. For example, in the last example each of $\mathbf{p}_1$ and $\mathbf{p}_2$ had two binding constraints, and the gradients of the corresponding g_i were linearly independent.

The KKT conditions are **necessary conditions** for a local maximum. They don't guarantee that a point satisfying them is actually a local maximum. In this example, $(2, 0)$ is actually not a local maximum.