# Appendix B: The Kuhn–Tucker theorem

These notes give a short overview of some key results on optimisation under constraints. In particular, they present the Kuhn–Tucker theorem on characterising optima for such problems. Note that this does not give existence of an optimiser, but only equivalent descriptions which can then be used to derive other properties.

We consider an optimisation problem under constraints which is given by

$$\text{maximise } U(y) \text{ over the set } \{y \in Y \mid g(y) \le 0\}, \tag{B.1}$$

where $Y$ is a subset of some vector space (e.g. $Y \subseteq \mathbb{R}^n$) and the constraints are given by a mapping $g : Y \to \mathbb{R}^m$.

**Definition.** (B.1) is called a *concave program* if

- **(i)** $Y$ is convex.
- **(ii)** $U : Y \to \mathbb{R}$ is concave.
- **(iii)** $g : Y \to \mathbb{R}^m$ is convex.

We can associate to (B.1) a *Lagrange function* $L : Y \times \mathbb{R}^m_+ \to \mathbb{R}$ by

$$L(y, \lambda) := U(y) - \lambda \bullet g(y).$$

**Definition.** A point $(y_0, \lambda_0) \in Y \times \mathbb{R}^m_+$ is called a *saddle point* of $L$ if

$$L(y, \lambda_0) \le L(y_0, \lambda_0) \le L(y_0, \lambda) \qquad \text{for all } (y, \lambda).$$

The vector $\lambda_0$ is then called a *Lagrange multiplier* for (B.1).

**Lemma B.1.** *If $(y_0, \lambda_0)$ is a saddle point of $L$, then $y_0$ is a solution of* (B.1).

**Proof.** For all $\lambda \in \mathbb{R}^m_+$, we have from the saddle point property that

$$(\lambda - \lambda_0) \bullet g(y_0) = U(y_0) - \lambda_0 \bullet g(y_0) - \big(U(y_0) - \lambda \bullet g(y_0)\big) = L(y_0, \lambda_0) - L(y_0, \lambda) \le 0.$$

This implies that $g(y_0) \le 0$ so that the constraint is satisfied in $y_0$. For all $y$ with $g(y) \le 0$, we moreover have

$$\begin{aligned} U(y) - U(y_0) &= L(y, \lambda_0) + \lambda_0 \bullet g(y) - \big(L(y_0, \lambda) + \lambda \bullet g(y_0)\big) \\ &= L(y, \lambda_0) - L(y_0, \lambda) + \lambda_0 \bullet g(y) - \lambda \bullet g(y_0) \\ &\le \lambda_0 \bullet g(y) - \lambda \bullet g(y_0) \\ &\le -\lambda \bullet g(y_0), \end{aligned}$$

where the first inequality again uses the saddle point property together with $\lambda_0 \ge 0$ and $g(y) \le 0$. Taking $\lambda = 0$ shows that $y_0$ is optimal for (B.1) and proves the result. **q.e.d.**

In order to obtain a converse result, we need an extra condition and some more properties.

**Definition.** We say that (B.1) satisfies the *Slater condition* if there exists some $\hat{y} \in Y$ with $g(\hat{y}) \in I\!R^m_{--}$, i.e. $g(\hat{y})$ is strictly negative in all coordinates.

**Theorem B.2.** *Suppose that* (B.1) *is a concave program which satisfies the Slater condition. If $y_0$ is a solution of* (B.1), *there exists some $\lambda_0 \in I\!R^m_+$ such that $(y_0, \lambda_0)$ is a saddle point of $L$. Moreover, the pair $(y_0, \lambda_0)$ then also satisfies the complementary slackness condition that $\lambda_0 \bullet g(y_0) = 0$.*

**Proof.** The idea of the proof is to use in a judicious way the separation theorem for well-chosen convex sets.

Let

$$\begin{aligned} A &:= \{(r, z) \in I\!R \times I\!R^m \mid U(y) \ge r \text{ and } z \ge g(y) \text{ for some } y \in Y\}, \\ B &:= \{(r, z) \in I\!R \times I\!R^m \mid U(y_0) \le r \text{ and } z \le 0\}. \end{aligned}$$

The sets $A$ and $B$ are convex because $U$ is concave and $g$ is convex. Moreover, the interior $B^\circ$ of $B$ is nonempty, and $A \cap B^\circ = \emptyset$ because $y_0$ is a solution of (B.1). So a separation theorem for convex sets (see Theorem A.4) yields the existence of a linear mapping $f : I\!R \times I\!R^m \to I\!R$,

$f \not\equiv 0$, with $f(u) \le f(v)$ for all $u \in A$ and $v \in B$. We write $f$ as $f(r,z) = \alpha r + \bar\lambda \bullet z$ with $\alpha \in I\!R$ and $\bar\lambda \in I\!R^m$. For fixed $u \in A$, we thus have

$$f(u) \le f(r,z) = \alpha r + \bar\lambda \bullet z \qquad \text{for all } r \ge U(y_0) \text{ and all } z \le 0,$$

and this implies that $\alpha \ge 0$ and $\bar\lambda \le 0$, since otherwise the right-hand side above would go to $-\infty$. If we had $\alpha = 0$, then taking $u_0 := (U(\hat y), g(\hat y))$ would give

$$f(u_0) = \bar\lambda \bullet g(\hat y) \le \bar\lambda \bullet z \qquad \text{for all } z \le 0$$

and hence for $z := \frac{1}{2} g(\hat y)$ also $\bar\lambda \bullet g(\hat y) \le 0$. But we know that $\bar\lambda \le 0$ and $g(\hat y) \in I\!R^m_{--}$ by the Slater condition; so we should get $\bar\lambda = 0$ and hence $f \equiv 0$, which is a contradiction. Therefore we must have $\alpha > 0$.

Now define $\lambda_0 := -\frac{\bar\lambda}{\alpha}$ so that $\lambda_0 \in I\!R^m_+$. Since $g(y_0) \le 0$, the point $(U(y_0), 0)$ is both in $A$ and in $B$, and because $(U(y_0), g(y_0))$ is in $A$, we obtain

$$\alpha U(y_0) + \bar\lambda \bullet g(y_0) = f\big(U(y_0), g(y_0)\big) \le f\big(U(y_0), 0\big) = \alpha U(y_0).$$

Because also $\bar\lambda \le 0$ and $g(y_0) \le 0$, we thus obtain $0 \ge \bar\lambda \bullet g(y_0) \ge 0$, hence $\bar\lambda \bullet g(y_0) = 0$ and also $\lambda_0 \bullet g(y_0) = 0$. For arbitrary $y$ and $\lambda \ge 0$, this now gives, using $g(y_0) \le 0$ and $\alpha > 0$, that

$$\begin{aligned}
\alpha L(y, \lambda_0) = \alpha U(y) + \bar\lambda \bullet g(y) &= f\big(U(y), g(y)\big) \\
&\le f\big(U(y_0), 0\big) = \alpha U(y_0) + \bar\lambda \bullet g(y_0) \\
&= \alpha L(y_0, \lambda_0) \\
&\le \alpha U(y_0) - \alpha\lambda \bullet g(y_0) \\
&= \alpha L(y_0, \lambda).
\end{aligned}$$

This shows that $(y_0, \lambda_0)$ is a saddle point for $L$. **q.e.d.**

If $U$ and $g$ are differentiable, the saddle point property can also be described in a different way by using the derivatives of $U$ and $g$.

**Lemma B.3.** *Suppose that $U$ and $g$ are differentiable on $Y$. If $(y_0, \lambda_0)$ is a saddle point of $L$ and $y_0$ an interior point of $Y$, then*

$$U'(y_0) - \lambda_0 \bullet g'(y_0) = 0,$$

*where $'$ denotes the gradient. Written out, we thus have*

$$0 = \operatorname{grad} U(y_0) - \sum_{i=1}^{m} \lambda_0^i \operatorname{grad} g^i(y_0).$$

**Proof.** Because $(y_0, \lambda_0)$ is a saddle point of $L$, the function $y \mapsto L(y, \lambda_0)$ has a minimum in $y$. Since $y_0$ is an interior point of $Y$, the derivative $\frac{\partial L}{\partial y}$ must therefore vanish in the point $(y_0, \lambda_0)$. But this derivative is easily computed to be $\frac{\partial L}{\partial y} = U' - \lambda \bullet g'$ since $U$ and $g$ are differentiable. **q.e.d.**

**Remarks. 1)** If for instance $Y$ is open, then any $y$ is an interior point of $Y$.

**2)** A similar statement for $\lambda_0$ instead of $y_0$ will not hold in general. The derivative $\frac{\partial L}{\partial \lambda}$ need not vanish in a saddle point; and because $\lambda \geq 0$, it is also not natural to assume that $\lambda_0$ is an interior point.

**Corollary B.4.** *Suppose that* (B.1) *is a concave program which satisfies the Slater condition, and assume that $U$ and $g$ are differentiable. If $y_0$ is a solution of* (B.1) *and an interior point of $Y$, then there exists $\lambda_0 \in \mathbb{R}_+^m$ with*

$$U'(y_0) - \lambda_0 \bullet g'(y_0) = 0,$$
$$\lambda_0 \bullet g(y_0) = 0.$$

**Proof.** This immediately follows from Theorem B.2 and Lemma B.3. **q.e.d.**