Nonlinear Programming
and the
Kuhn-Tucker Conditions

We typically begin studying constrained optimization analysis with just a single, binding constraint (an equation), and with variables that are otherwise unrestricted. This rules out situations where there are multiple constraints, where some constraints may be non-binding, and where non-negativity constraints may be binding (*i.e.*, where some variables may be zero at the optimum).

The Kuhn-Tucker Conditions provide a unified treatment of constrained optimization in which

- there may be any number of constraints;
- constraints may be binding or not binding at the solution;
- boundary solutions (some x_i's $= 0$) are permitted;
- non-negativity and structural constraints are treated in the same way;
- dual variables (also called Lagrange multipliers) are shadow values (*i.e.*, marginal values).

The Kuhn-Tucker Conditions are simply the first-order conditions for a constrained optimization problem – a generalization of the first-order conditions we're familiar with, a generalization that can handle the situations described above. A special case covered by the Kuhn-Tucker Conditions is Linear Programming.

The Kuhn-Tucker Conditions

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable functions, and let $\mathbf{b} \in \mathbb{R}^m$. We want to characterize those vectors $\widehat{\mathbf{x}} \in \mathbb{R}^n$ that satisfy

($*$) $\widehat{\mathbf{x}}$ is a solution of the problem

(**P**) Maximize $f(\mathbf{x})$ subject to $\mathbf{x} \geqq \mathbf{0}$ and $G(\mathbf{x}) \leqq \mathbf{b}$,

i.e., subject to $x_1, x_2, \ldots, x_n \geqq 0$ and to $G^i(\mathbf{x}) \leqq b_i$ for $i = 1, \ldots, m$.

The Kuhn-Tucker Conditions are the first-order conditions that characterize the vectors $\widehat{\mathbf{x}}$ that satisfy ($*$) (when appropriate second-order conditions are satisfied, which we'll see momentarily):

$\exists\, \lambda_1, \ldots, \lambda_m \in \mathbb{R}_+$ such that

(KT1) For $j = 1, \ldots, n$: $\quad \dfrac{\partial f}{\partial x_j} \leqq \sum_{i=1}^m \lambda_i \dfrac{\partial G^i}{\partial x_j}$, $\quad$ with equality if $\widehat{x}_j > 0$;

(KT2) For $i = 1, \ldots, m$: $\quad G^i(\widehat{\mathbf{x}}) \leqq b_i$, $\quad$ with equality if $\lambda_i > 0$,

where the partial derivatives are evaluated at $\widehat{\mathbf{x}}$.

The Kuhn-Tucker Conditions given above are in *partial derivative form*. An equivalent statement of the conditions is in *gradient form*:

$\exists\, \lambda \in \mathbb{R}^m_+$ such that

$$\begin{array}{llll} \text{(KT1)} & \nabla f \leqq \sum_{i=1}^m \lambda_i \nabla G^i & \text{and} & \widehat{\mathbf{x}} \cdot (\nabla f - \sum_{i=1}^m \lambda_i \nabla G^i) = 0\ ; \\ \text{(KT2)} & G(\widehat{\mathbf{x}}) \leqq \mathbf{b} & \text{and} & \lambda \cdot (\mathbf{b} - G(\widehat{\mathbf{x}})) = 0\ , \end{array}$$

where gradients are evaluated at $\widehat{\mathbf{x}}$.

The Kuhn-Tucker Theorems

The first theorem below says that the Kuhn-Tucker Conditions are *sufficient* to guarantee that $\widehat{\mathbf{x}}$ satisfies $(*)$, and the second theorem says that the Kuhn-Tucker Conditions are *necessary* for $\widehat{\mathbf{x}}$ to satisfy $(*)$. Taken together, the two theorems are called the Kuhn-Tucker Theorem.

Theorem 1: Assume that each G^i is quasiconvex; that either (a) f is concave or (b) f is quasiconcave and $\nabla f \neq \mathbf{0}$ at $\widehat{\mathbf{x}}$; and that f and each G^i are differentiable. If $\widehat{\mathbf{x}}$ satisfies the Kuhn-Tucker Conditions then $\widehat{\mathbf{x}}$ satisfies $(*)$.
[Briefly, (KT) $\Rightarrow$ $(*)$.]

Theorem 2: Assume that f is quasiconcave; that each G^i is quasiconvex and the constraint set $\{\mathbf{x} \in \mathbb{R}^n \mid G(\widehat{\mathbf{x}}) \leqq \mathbf{b}\}$ satisfies one of the constraint qualifications (to be described shortly); and that f and each G^i are differentiable. If $\widehat{\mathbf{x}}$ satisfies $(*)$ then $\widehat{\mathbf{x}}$ satisfies the Kuhn-Tucker Conditions.
[Briefly, $(*)$ $\Rightarrow$ (KT).]

The next theorem tells us how changes in the values of the b_i's affect the value of the objective function f. For the nonlinear programming problem defined by f, G, and $\mathbf{b}$, define the **value function** $v : \mathbb{R}^m \to \mathbb{R}$ as follows:

$$\forall\, \mathbf{b} \in \mathbb{R}^m : \ v(\mathbf{b}) \text{ is the value of } f(\widehat{\mathbf{x}}) \text{ where } \widehat{\mathbf{x}} \text{ satisfies } (*).$$

Theorem 3: If $(*)$ and (KT) are both satisfied at $\widehat{\mathbf{x}}$, then $\lambda_i = \dfrac{\partial v}{\partial b_i}$ for each i.

In other words, λ_i is the "shadow value" of the i^{th} constraint, the marginal value to the objective function of relaxing or tightening the constraint by one unit.

Note that second-order (curvature/convexity/concavity) conditions are required in order for the Kuhn-Tucker (first-order) conditions to be either necessary or sufficient for $\widehat{\mathbf{x}}$ to be a solution to the nonlinear programming problem.

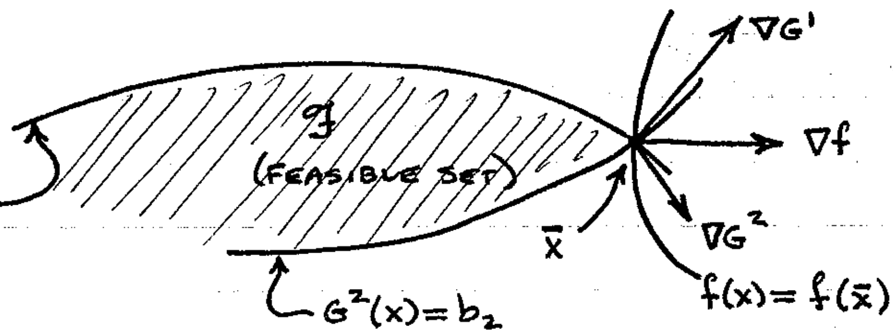
THE FOLLOWING GEOMETRICAL EXAMPLES WILL BE HELPFUL FOR UNDERSTANDING THE KUHN-TUCKER CONDITIONS AND THEIR RELATION TO CONSTRAINED OPTIMIZATION.

(1) SUPPOSE THAT EACH $\bar{x}_j > 0$ AND THAT (KT) IS SATISFIED AT $\bar{x}$. THEN

$$\nabla f = \sum_{i=1}^{m} \lambda_i \nabla G^i \quad \text{FOR SOME } \lambda_1, \ldots, \lambda_m \geqq 0;$$

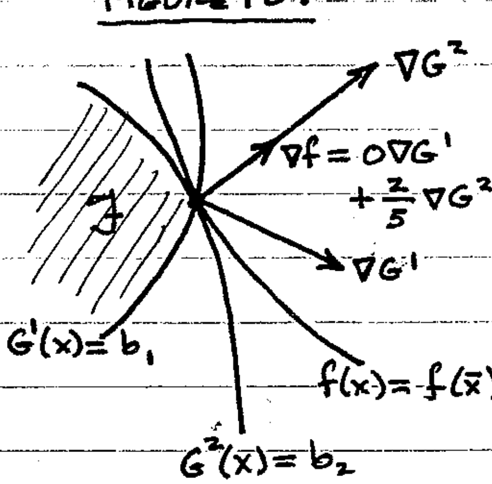
i.e., ∇f LIES IN THE CONE FORMED BY THE GRADIENTS OF THE CONSTRAINTS (IT IS A NON-NEGATIVE LINEAR COMBINATION OF THEM), AS IN FIGURE 1a.

FIGURE 1a:



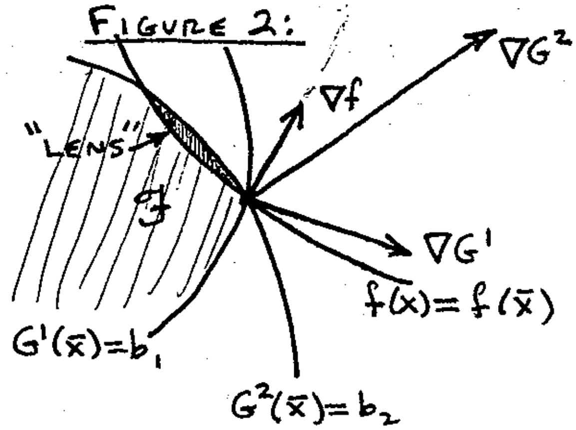
NOTICE THAT EACH $G^i(\bar{x}) = b_i$ IN THIS FIRST EXAMPLE, AND THAT EACH $\lambda_i > 0$. IT IS, HOWEVER, POSSIBLE TO HAVE $\lambda_i = 0$ FOR SOME i's, EVEN WHEN $G^i(\bar{x}) = b_i$; THIS WOULD STILL LEAVE ∇f IN THE CONE FORMED BY THE ∇G^i's, AS IN FIGURE 1b.

FIGURE 1b:



(2) Now suppose that f and/or G^2 in Figure 1b were perturbed slightly, giving us Figure 2. Now ∇f no longer lies in the cone of the ∇G^i's. But notice that there is also a "lens" formed between the f and G^2 contours through $\bar{x}$ and that this lens contains points x that are feasible (i.e., both $G^1(x) \leqq b_1$ and $G^2(x) \leqq b_2$) and give larger values of f than does $\bar{x}$ — i.e., $f(x) > f(\bar{x})$. In other words, $\bar{x}$ no longer satisfies (*), just as it no longer satisfies (KT).

Figure 2:



This idea of the "lens" between contours is a valuable idea. When f is quasi-concave and each G^i is quasi-convex, the condition that ∇f lies in the cone of the ∇G^is is precisely that any lens between the f contour and a G^i contour must lie "outside" of some other G^i contour — i.e., that the f contour is "tangent" (in the generalized sense) to the feasible set at $\bar{x}$, and thus that f is maximized at $\bar{x}$.

Note that so far we have always had each $x_j > 0$ and each $G^i(\bar{x}) = b_i$; now we consider some $x_j = 0$ and some $G^i(\bar{x}) < b_i$:

③ IF $G^i(\bar{x}) < b_i$, THEN (KT) REQUIRES THAT $\lambda_i = 0$ — i.e., ∇f LIES IN THE CONE OF THE OTHER ∇G^k's. THIS IS BECAUSE, AS IN FIGURE 3, G^i CONTRIBUTES NOTHING TO DEFINING THE FEASIBLE SET AT $\bar{x}$, AND THUS ANY LENS BETWEEN THE f CONTOUR AND ANOTHER G^k CONTOUR NEED ONLY LIE OUTSIDE THE SET FORMED BY JUST THOSE OTHER G^k's.

④ SUPPOSE THAT $\frac{\partial f}{\partial x_2} < \sum_{i=1}^{m} \lambda_i \frac{\partial G^i}{\partial x_2}$; THEN (KT) REQUIRES THAT $\bar{x}_2 = 0$, AS IN FIGURE 4. NOW WE DON'T HAVE ∇f IN THE CONE OF THE ∇G^i's, SO THERE IS A LENS BETWEEN f AND SOME G^i THAT LIES "INSIDE" ALL G^i CONTOURS. BUT THE FEASIBLE SET $\mathcal{F}$ IS TRUNCATED BY THE AXES (HERE, BY $x_2 = 0$), SO THIS LENS IS NOT ACTUALLY IN $\mathcal{F}$. ANALYTICALLY, THIS IS CAPTURED BY THE FACT THAT IT IS x_2 THAT IS ZERO AND IT IS THE x_2-COMPONENT IN WHICH EQUALITY FAILS IN $\nabla f \leqq \sum \lambda_i \nabla G^i$. NOTICE IN FIGURE 4 THAT IT IS POSSIBLE TO HAVE $\nabla f \leqq \lambda \nabla G$ (i.e., $\lambda \nabla G$ LIES "NORTHEAST" OF ∇f) AND ALSO $\frac{\partial f}{\partial x_1} = \lambda \frac{\partial G}{\partial x_1}$.

FIGURE 3:

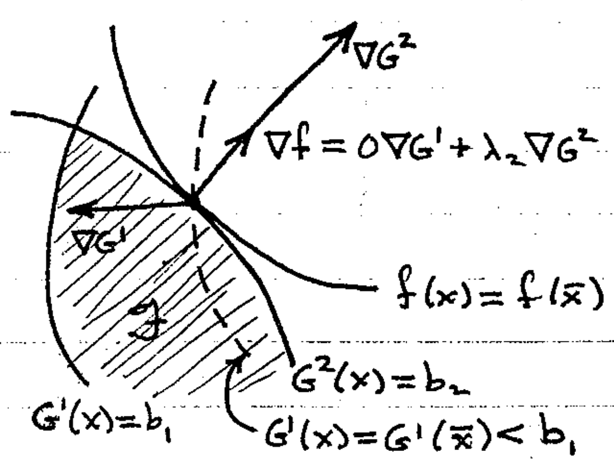
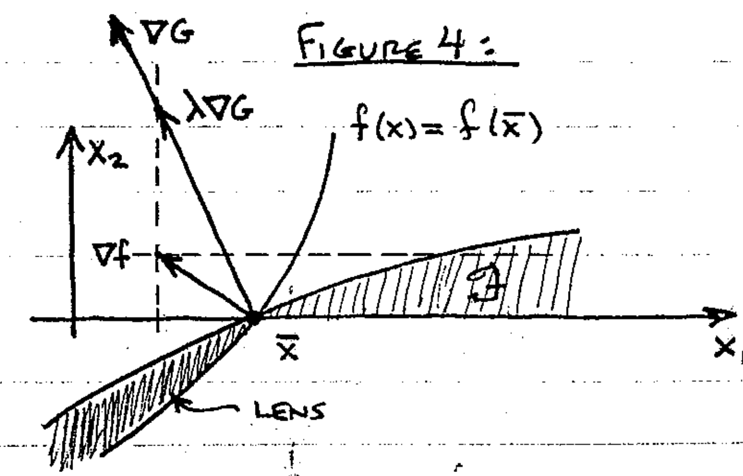


FIGURE 4:



⑤ IN ORDER THAT (KT) BE SUFFICIENT FOR (*), WE MUST INCLUDE THE CONVEXITY (i.e., SECOND-ORDER, OR CURVATURE) CONDITIONS ON f AND G, AS FIGURE 5 DEMONSTRATES. THE FIGURE SHOWS A SITUATION IN WHICH (KT) IS SATISFIED, BUT (*) IS NOT — BECAUSE G IS NOT QUASI-CONVEX.

⑥ IF WE RECALL EXAMPLE 2, IT SEEMS AS IF (KT) MUST BE NECESSARY IF (*) IS TO HOLD: OTHERWISE, IT SEEMS, THERE WOULD BE A LENS PARTLY IN THE FEASIBLE SET. WITHOUT SECOND-ORDER CONDITIONS OF SOME KIND, HOWEVER, THIS IS NOT QUITE SO, AS FIGURE 6 DEMONSTRATES. HERE THE GRADIENTS ∇G^1 AND ∇G^2 ARE OPPOSITE TO ONE ANOTHER, SO $\bar{x}$ IS THE ONLY POINT IN THE FEASIBLE SET, AND IT THEREFORE CLEARLY SATISFIES (*). BUT ∇f IS NOT IN THE CONE OF THE ∇G^i's, SO (KT) IS NOT SATISFIED.

FIGURE 5:

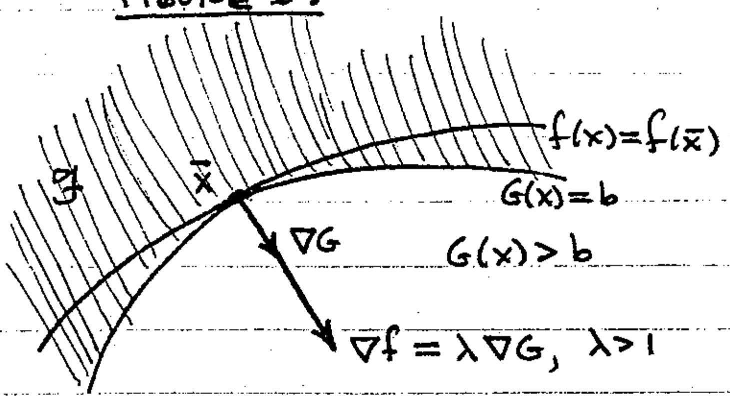
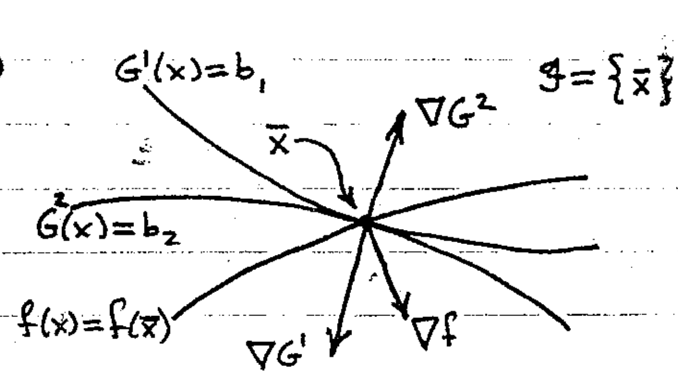


FIGURE 6:

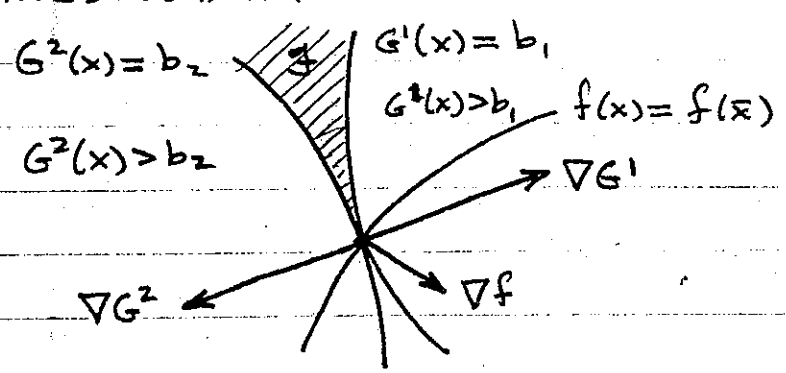


(7) It seems as if we could rule out situations like the one in Example 6 by ruling out degenerate (singleton) constraint sets. But Figure 7 shows that we need a stronger condition to ensure that (KT) is necessary — i.e., that (*) implies (KT). Again, we have ∇G^1 and ∇G^2 opposite to one another, but here we have quasi-concave constraint functions G^i. We can rule out these kinds of situations by imposing both a curvature condition and a condition that the feasible set be of full dimension (i.e., have a nonempty interior) as in the following "Constraint Qualification."

Constraint Qualification: There is a program $x \in \mathbb{R}^n_+$ such that $G^i(x) < b_i$, $i = 1, \dots, m$; and each G^i is either convex or else quasi-convex with a non-trivial gradient (i.e., $\nabla G^i \neq (0, 0, \dots, 0)$) at $\bar{x}$.

Theorem 2 says that if G satisfies (CQ) and (*) holds, then (KT) holds — i.e., with (CQ) the Kuhn-Tucker Conditions are necessary for a constrained maximum.

Figure 7:



NOTE THAT IN BOTH OF THE DEGENERATE EXAMPLES IN WHICH (KT) IS NOT NECESSARY FOR (*) (FIGURES 6 AND 7), THE TWO CONSTRAINT GRADIENTS ∇G^1 AND ∇G^2 WERE LINEARLY DEPENDENT. THIS CAN BE GENERALIZED TO A MORE STRAIGHTFORWARD CONSTRAINT QUALIFICATION:

<u>CONSTRAINT QUALIFICATION:</u> THE GRADIENTS $\nabla G^1(\bar{x}), \ldots, \nabla G^m(\bar{x})$ ARE LINEARLY INDEPENDENT, i.e., THE JACOBIAN MATRIX $DG(\bar{x})$ HAS RANK m.