

Nonlinear Programming and the Kuhn-Tucker Conditions

Our constrained optimization analysis has dealt only with situations in which there is a single constraint, and in which all the variables are generally positive at the optimum. This rules out situations where there are multiple constraints, where there are non-binding constraints, 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $G : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be continuously differentiable functions, and let $\mathbf{b} \in \mathbb{R}^m$.

We want to characterize those vectors $\hat{\mathbf{x}} \in \mathbb{R}^n$ that satisfy

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

Maximize $f(\mathbf{x})$ subject to $\mathbf{x} \geq \mathbf{0}$ and $G(\mathbf{x}) \leq \mathbf{b}$,

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

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

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

(KT1) For $j = 1, \dots, n$: $\frac{\partial f}{\partial x_j} \leq \sum_{i=1}^m \lambda_i \frac{\partial G^i}{\partial x_j}$, with equality if $\hat{x}_j > 0$;

(KT2) For $i = 1, \dots, m$: $G^i(\hat{\mathbf{x}}) \leq b_i$, with equality if $\lambda_i > 0$,

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

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

$$(KT1) \quad \nabla f \leq \sum_{i=1}^m \lambda_i \nabla G^i \quad \text{and} \quad \hat{\mathbf{x}} \cdot (\nabla f - \sum_{i=1}^m \lambda_i \nabla G^i) = 0 ;$$

$$(KT2) \quad G(\hat{\mathbf{x}}) \leq \mathbf{b} \quad \text{and} \quad \lambda \cdot (\mathbf{b} - G(\hat{\mathbf{x}})) = 0 ,$$

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

The Kuhn-Tucker Theorems

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

Theorem 1: Assume that each G^i is quasi-convex and either (a) f is concave or (b) f is quasi-concave and $\nabla f \neq \mathbf{0}$ at $\hat{\mathbf{x}}$. If $\hat{\mathbf{x}}$ satisfies the Kuhn-Tucker Conditions then $\hat{\mathbf{x}}$ satisfies (*).

[Briefly, (KT) \Rightarrow (*).]

Theorem 2: Assume that f is quasi-concave and G satisfies one of the constraint qualifications (to be described shortly). If $\hat{\mathbf{x}}$ satisfies (*) then $\hat{\mathbf{x}}$ 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** $v : \mathbb{R}^m \rightarrow \mathbb{R}$ as follows:

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

Theorem 3: If (*) and (KT) are both satisfied at $\hat{\mathbf{x}}$, then $\lambda_i = \frac{\partial v}{\partial b_i}$ 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.

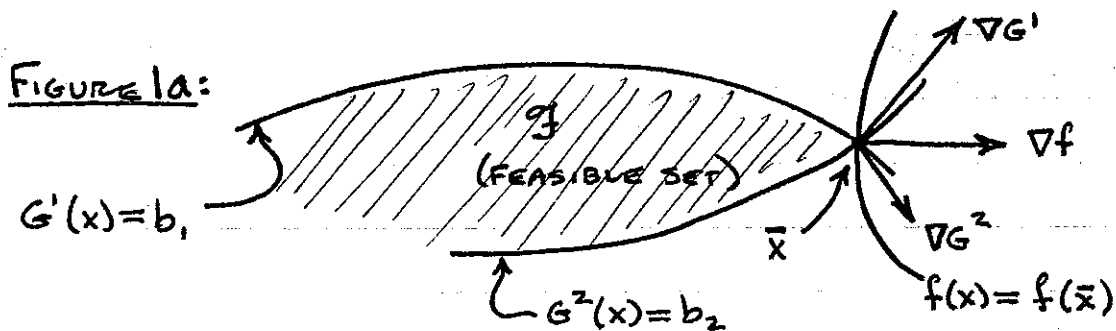
Note that second-order (curvature/convexity/concavity) conditions are required in order for the Kuhn-Tucker (first-order) conditions to be either necessary or sufficient for $\hat{\mathbf{x}}$ to be a solution to the nonlinear programming problem.

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

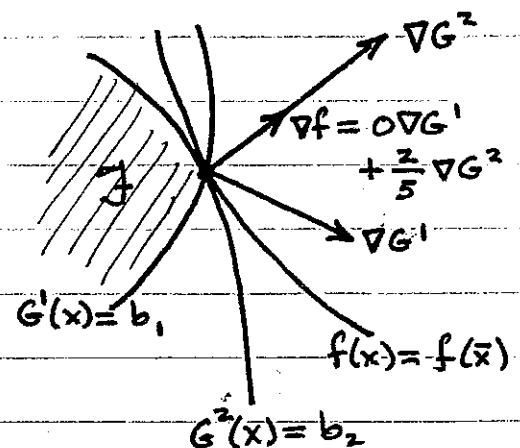
$$\nabla f = \sum_{i=1}^m \lambda_i \nabla G^i \text{ FOR SOME } \lambda_1, \dots, \lambda_m \geq 0;$$

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.

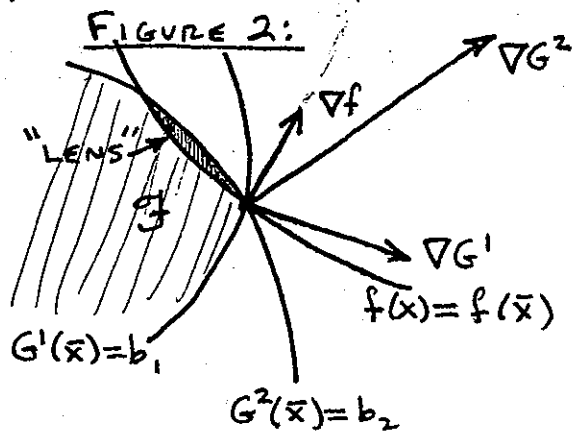


NOTICE THAT EACH $G^i(\bar{x}) = b_i$ IN THIS FIRST EXAMPLE, AND THAT EACH $\lambda_i > 0$. IT IS, HOWEVER, POSSIBLE TO HAVE $\lambda_i = 0$ FOR SOME i 'S, EVEN WHEN $G^i(\bar{x}) = b_i$; THIS WOULD STILL LEAVE ∇f IN THE CONE FORMED BY THE ∇G^i 'S, AS IN FIGURE 1b.

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 $G^1(x) \leq b_1$ AND $G^2(x) \leq b_2$) 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).

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^i 'S 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 $x_i = 0$ 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 $\nabla f \leq \sum \lambda_i \nabla G^i$. NOTICE IN FIGURE 4 THAT IT IS POSSIBLE TO HAVE $\nabla f \leq \lambda \nabla G$ (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:

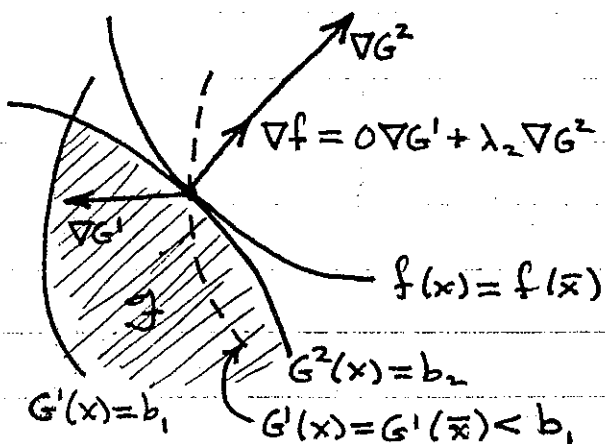
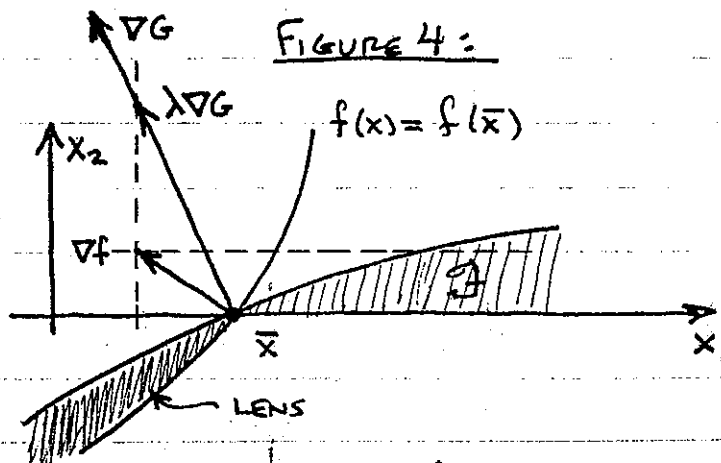
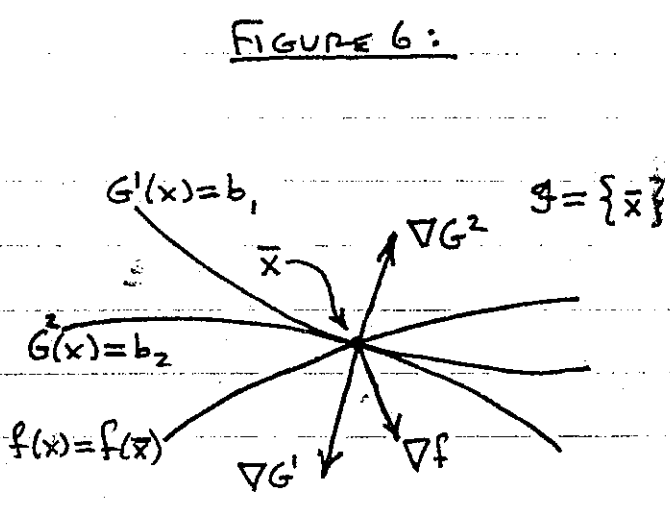
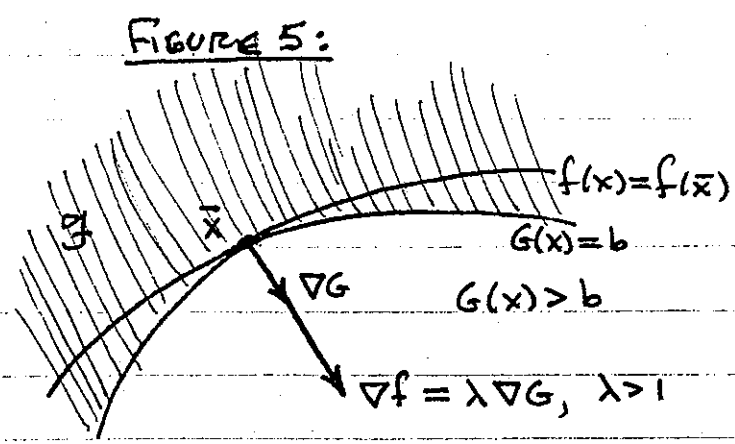


FIGURE 4:



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.

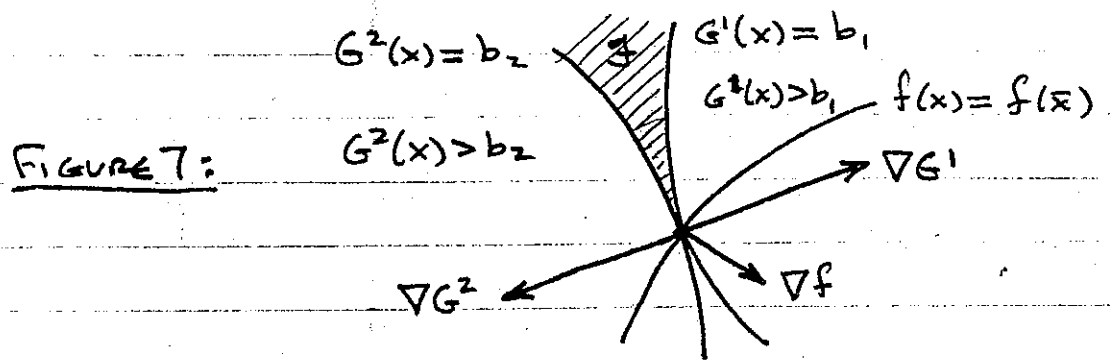
⑥ IF WE RECALL EXAMPLE 2, IT SEEMS AS IF (KT) MUST BE NECESSARY IF (*) IS TO HOLD: OTHERWISE, IT SEEMS, THERE WOULD BE A LENS PARTLY IN THE FEASIBLE SET. WITHOUT SECOND-ORDER CONDITIONS OF SOME KIND, HOWEVER, THIS IS NOT QUITE SO, AS FIGURE 6 DEMONSTRATES. HERE THE GRADIENTS ∇G^1 AND ∇G^2 ARE OPPOSITE 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.



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} ~~CONCAVE~~ OR ELSE QUASI-CONVEX ~~CONCAVE~~ 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.



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 STRAIGHT FORWARD CONSTRAINT QUALIFICATION:

CONSTRAINT QUALIFICATION: THE GRADIENTS $\nabla G^1(\bar{x}), \dots, \nabla G^m(\bar{x})$ ARE LINEARLY INDEPENDENT, I.E., THE JACOBIAN MATRIX $DG(\bar{x})$ HAS RANK m .