6. **Nonlinear Programming Problems and the Karush-Kuhn Tucker conditions**

6.1. Existence and Uniqueness

6.2. Necessary conditions for a solution to an NPP

6.3. Role of the Constraint Qualification

6.4. Demonstration that KKT conditions are necessary

---

6. **NONLINEAR PROGRAMMING PROBLEMS AND THE KARUSH-KUHN TUCKER CONDITIONS**

Going to look at the technique for solving the general nonlinear programming problem. We did this graphically at the beginning of the year, but we now need to do it formally. See why the calculus conditions do what they are supposed to do.

The general nonlinear programming problem (NPP) is the following:

\[
\text{maximize } f(x) \text{ subject to } g(x) \leq b,
\]

where \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) and \( g: \mathbb{R}^n \rightarrow \mathbb{R}^m \)

Terminology:
• $f$ is called the *objective function*;

• $g$ is a vector of $m$ *constraint functions*, and, of course, $b \in \mathbb{R}^m$. That is, the individual constraints are stacked together to form a vector-valued function.

• The set of $x$ such that $g(x) \leq b$ is called the *feasible set* or *constraint set* for the problem.

For the remainder of the course, unless otherwise notified, we will assume that both the objective function and the constraint functions are continuously differentiable. Indeed, we will in fact assume that they are as many times continuously differentiable as we could ever need.

Emphasize that this setup is *completely* general, i.e., covers every problem you are ever likely to encounter.

• can handle constraints of the form $g(x) \geq b$;

• can handle constraints of the form $x \geq 0$;

• can even handle constraints of the form $g(x) = b$.

For example, given $u : \mathbb{R}^2 \to \mathbb{R}$, consider the problem

$$\text{maximize } u(x) \text{ subject to } p \cdot x \leq y, x \geq 0;$$

What’s $g$: in this case, $g$ is a linear function, defined as follows:

$$g^0(x) = p \cdot x$$

$$g^1(x) = -x_1$$

$$g^2(x) = -x_2$$
so that the problem can be written as

$$\text{maximize } u(x) \text{ subject to } Gx \leq b,$$

where

$$G = \begin{pmatrix} p_1 & p_2 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{and } b = \begin{pmatrix} y \\ 0 \\ 0 \end{pmatrix}.$$

While there are many advantages to having a single, general version of the KT conditions, the generality comes at a (small) cost. When it comes to actually computing the solution to an NPP problem (as opposed to just understanding what’s going on), it’s very convenient to treat equality constraints differently from inequality constraints. I explain what you need to do on page 10. You should make sure to refer to this discussion before you start on the NPP problem set.

6.1. Existence and Uniqueness

The first step is: under what conditions does the NPP have a solution at all? Under what conditions will the solution be unique? We’ll answer this question by considering when a function defined on an arbitrary set \(A\) attains a maximum and when this maximum is unique. Then we’ll apply these results to the NPP, letting \(A\) denote the constraint set for the problem. Recall the following theorem, also known as the extremum value theorem.

**Theorem (Weierstrass):** If \(f : A \rightarrow \mathbb{R}\) is continuous and \(A\) is compact, nonempty, then \(f\) attains a maximum on \(A\).

- relationship between \(A\) and the constraint sets of the NPP: \(A\) is the intersection of the lower contour sets defined by the NPP, i.e., \(g_j(\cdot) \leq b^j\), for all \(j\).
• role of compact, i.e., closed and bounded, and nonempty; not guaranteed in the specification
  of the NPP.
  – example of why existence of a maximum may fail if $A$ is not closed: try to maximize
    $f(x) = 1/x$ on the interval $(0, 1]$.
  – example of why existence of a maximum may fail if $A$ is not bounded try to maximize
    $f(x) = x$ on $\mathbb{R}_+$.

• when you set up the problem, have to make sure the constraint set you end up with is
  nonempty and compact, otherwise you may not get a solution.

Theorem: If $f : A \rightarrow \mathbb{R}$ is continuous and $A$ is compact, nonempty and convex, and $f$ is strictly
quasi-concave then $f$ attains a unique maximum on $A$.

• example of why uniqueness may fail if $f$ is not quasi-concave: let $A$ be a circle and construct
  a function $f$ which has a level set that has two points of tangency with the ball.
• example of why uniqueness may fail if $A$ is not convex: let $A$ be a ball with a bite out of
  it, and have a quasi-concave level set that touches the edge of the constraint surface, where
  the surface has been bitten, at two distinct points.

Theorem: If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous and strictly quasi-concave and $g^j : \mathbb{R}^n \rightarrow \mathbb{R}$ is quasi-convex
for each $j$, then if $f$ attains a local maximum on $A = \{x \in \mathbb{R}^n : \forall j, g^j(x) \leq b_j\}$, this max is the
unique global maximum of $f$ on $A$.

Proof:

• recall that the upper contour sets of a quasi-concave function are convex sets.
• similarly, the lower contour sets of a quasi-convex function are convex sets.
• the condition that \( g^j(x) \leq b_j \) is the condition that \( x \) lie in the lower contour set of the \( j \)'th constraint identified by \( b_j \).

• finally, the intersection of convex sets is convex.

• now apply previous theorem about uniqueness given conditions on \( A \).

Thus, quasi-convexity of \( g^j \)'s guarantees convexity of \( A \). However, it doesn’t imply COMPACTNESS, hence existence isn’t guaranteed.

Take stock: haven’t told you how to solve one of these problems, only when it will have a solution and when that solution will be unique.

6.2. **Necessary conditions for a solution to an NPP**

In this topic, we’ll give two theorems, one is necessary and one is sufficient for \( \bar{x} \) to be a maximum. For now, we’re focusing on necessary conditions.

Recall the mantra: (Except for a bizarre exception) a necessary condition for \( x \) to solve a constrained maximization problem is that the gradient vector of the objective function at \( x \) belongs to the nonnegative cone defined by the gradient vectors of the constraints that are satisfied with equality at \( x \).

The bizarre exception is called the *constraint qualification*. Worry about it later.

Let \( \lambda^T \) denotes the transpose of \( \lambda \), i.e., if \( \lambda \) is assumed to be a column vector, then \( \lambda^T \) will be a row vector.
Theorem (Kuhn-Tucker necessary conditions): If $\bar{x}$ solves the maximization problem and the constraint qualification holds at $\bar{x}$ then there exists a vector $\bar{\lambda} \in \mathbb{R}_+^m$ such that

$$\nabla f(\bar{x})^T = \bar{\lambda}^T Jg(\bar{x})$$

Moreover, $\bar{\lambda}$ has the property that $\bar{\lambda}_j = 0$, for each $j$ such that $g^j(\bar{x}) < b_j$.

Why didn’t I write $Jg(\bar{x})\bar{\lambda}$?

- Recall that if $Ax = b$, then $b$ is a weighted linear combination of the columns of $A$.
- Imagine if you wrote $x'A = b$, then what would $b$ be? Ans.: it is a weighted linear combination of the rows of $A$.
- That’s what we want here: in words, what the theorem says is: if $\bar{x}$ solves the maximization problem and the constraint qualification holds then the gradient of $f$ at $\bar{x}$ is a nonnegative weighted linear combination of the gradients of the constraints that are satisfied with equality.

Needless to say, the above conditions aren’t sufficient for a solution. You need conditions that guarantee you a *global* max on the constraint set.

- draw constraint set that isn’t from a quasi-convex function; linear level sets for the objective function; show that the KKT conditions are satisfied at a local tangency, but that there’s a point further away that gives a higher value for the objective function. Note here that a local maximum on the constraint set is *not* a solution. By a solution we mean a *global* max on the constraint set.

---

1 Actually, this is not the Kuhn-Tucker theorem but the Karush-Kuhn-Tucker theorem. Karush was a graduate student who was partly responsible for the result. Somehow, his name has been forgotten. A regrettable thing for graduate students.
- draw constraint set from a quasi-convex function but objective isn’t quasi-concave. Exactly the same thing. Because of a “kink” you can get a local max on the constraint set that isn’t a solution to the problem.

6.3. Role of the Constraint Qualification

The right way to think about the constraint qualification is that it is sufficient to ensure that the KT conditions are necessary for a solution. Some people have wondered the following: “might the KT conditions lead me to believe that I’ve found a solution to the NPP, but then I find that because the constraint qualification is violated, and so candidate solution turns out not to be a solution?”

The answer to this question is a resounding NO. If the constraint qualification is violated, there may be solutions to the NPP that don’t satisfy the KT conditions. On the other hand, the KT conditions may lead indeed me to believe that I’ve found a solution to the NPP, which for reasons other than the constraint qualification failing, is not in fact a solution.

If you say that finding a point that satisfies the KT conditions is a “positive” and not finding such a point is a “negative”, then we can talk about “false positives” (in the sense of a positive result of a test, even though the thing you’re testing for (say pregnancy) hasn’t happened) and “false negatives” (the test says you aren’t pregnant when you are).

In KT theory, we’ve seen both kinds of “falseness”: there are several ways that we can get “false positives” in the sense that the KT conditions are satisfied at some point which does not solve the NPP. For example, the gradient may vanish at a non-solution point, or the second-order requirements for a solution may be violated. The only way to get a “false negative” is for the CQ to be violated (e.g., Varian’s “Hershey Kiss” example (Fig. 1 below)). Thus, provided the CQ is not violated, there will be no “false negatives” and the maximum (if it exists) must satisfy KT.
We'll now show that in the absence of the caveat about the constraint qualification, the above theorem would be false.

- the constraint qualification relates to the mysterious “bizarre exception” that is added as a caveat to the mantra.

- In Fig. 1 below, the gradient vectors for the constraints satisfied with equality are not linearly independent, and don’t span the space. Can’t write the objective function as a linear combination of them. Doesn’t mean that we don’t have a maximum at the tip of the constraint set.

- A sufficient condition for the CQ to be satisfied is that the gradients of the constraints satisfied with equality at $\bar{x}$ form a linearly independent set.

- Another way of stating the above is as follows: let $M(x)$ denote what’s left of $Jg(x)$ (i.e., the Jacobian matrix of $g$ evaluated at $x$), after you’ve eliminated all rows of this matrix corresponding to constraints satisfied with strict inequality at $x$; a sufficient condition for the CQ to be satisfied at $x$ is that the matrix $M(x)$ has full rank.

- If this condition isn’t satisfied then you could have a maximum at $\bar{x}$ but wouldn’t know about it by looking at the KKT conditions.

- The right way to think about what goes wrong is that when the CQ is violated, the information conveyed by the gradient vectors doesn’t accurately describe the constraint set.

- Saying this more precisely, when we use the KT conditions to solve an NPP, what we are actually doing is solving the linearized version of the problem. By this I mean, we replace the boundary of the constraint set, as defined by the level sets of the constraint functions, by the piecewise linear surface defined by the tangent planes to the level sets. In order for the KT conditions to work, it had better be the case that in a neighborhood of the solution you are considering, this “linear shape” looks, locally, pretty much like the shape of the true nonlinear constraint set for the problem.
• In the case of our example, the tangent lines to the two level sets, at the tip of the pointy cone, are parallel to each other, so that the “linearized problem” is a line that goes on for ever.

  – Given the objective function that we’ve drawn, the linearized problem has no solution, you just go on and on forever along the line.

  – the constraint set for the true problem doesn’t look anything like a line, so that the linearized problem is completely misleading

  – observe that if you “bumped” the original problem a little bit, so that the tangent planes at the tip were not parallel to each other, then the close relationship between the original and the linearized problems would be restored.

  – what determines whether the linearized problem accurately represents the original problem? Answer is apparent from comparing the relatinoship between the gradient vectors to the tangent planes for the original, pathological problem to this relationship for the bumped, well-behaved problem: in the former case, the two gradient vectors are colinear.

  – More generally, the linearized problem will locally accurately represent the original nonlinear problem at a point if the gradient vectors of the constraints that are satisfied with equality at that point form a linear independent set.

6.4. **Demonstration that KKT conditions are necessary**

Rather than prove the result generally, we are just going to look at special cases.

**The case of one inequality constraint:** With one constraint, the picture is easy: the condition is that \( \nabla f(\bar{x}) = \lambda \nabla g(\bar{x}) \), for some nonnegative scalar \( \lambda \). Also, with one constraint the CQ is vacuous.
Suppose you don’t have the above condition satisfied. That is, suppose you don’t have nonnegative colinearity: $\nabla f(\bar{x})$ is not a nonnegative scalar multiple of $\nabla g(\bar{x})$. Then you can find a vector $\mathbf{d}x$ such that $\nabla g(\bar{x}) \cdot \mathbf{d}x < 0$ and $\nabla f(\bar{x}) \cdot \mathbf{d}x > 0$. (Easy to see this diagrammatically, but it involves some linear algebra work to show that this must be true generally. We’ll just assert it for now.)

But this means that $\mathbf{x} + \mathbf{d}x$ satisfies the inequality constraint and (by Taylor’s theorem) increases $f$ provided $\mathbf{d}x$ is sufficiently small.

The case of one equality constraint, $g(\mathbf{x}) = b$: In this case, let $g^1 = g$ and $g^2 = -g$. The KT conditions say that

$$\nabla f(\bar{x}) = \lambda_1 \nabla g^1(\bar{x}) + \nabla \lambda_2 g^2(\bar{x}), \text{ for some pair } (\lambda_1, \lambda_2) \geq 0. \quad (1)$$

For the purposes of computation, however, it is convenient to recognize that the above condition is equivalent to: $\nabla f(\bar{x}) = \lambda \nabla g(\bar{x})$, for some arbitrary scalar $\lambda$, i.e., (i.e., $\lambda$ in this case is not restricted to be nonnegative). To see why the two conditions are equivalent note that
\(\lambda_1 \nabla g^1(\bar{x}) + \lambda_2 \nabla g^2(\bar{x}) = (\lambda_1 - \lambda_2)\nabla g(\bar{x})\)

(2) At most one of the two \(\lambda\)'s can be non-zero.

(a) if \((\lambda_1 - \lambda_2) > 0\) then the \(g^1\) constraint cannot be binding, therefore \(\lambda_2\) must be zero.

(b) if \((\lambda_1 - \lambda_2) < 0\) then the \(g^2\) constraint cannot be binding, therefore \(\lambda_1\) must be zero.

(c) if \((\lambda_1 - \lambda_2) = 0\) then \(\nabla f\) must be zero, in which case neither constraint can be binding, so that \(\lambda_1 = \lambda_2 = 0\).

Because of this (i.e., because at most one \(\lambda\) can be non-zero), we can collapse the two \(\lambda\)'s into one, as follows:

\[
\lambda = \begin{cases} 
\lambda_1 & \text{if } g^1 \text{ is binding} \\
-\lambda_2 & \text{if } g^2 \text{ is binding} \\
0 & \text{if neither are binding}
\end{cases}
\]

and observe that (1) is now equivalent to

\[
\nabla f(\bar{x}) = \lambda \nabla g(\bar{x}), \quad \lambda \in \mathbb{R} \quad (1')
\]

The two constraint case: With two constraints the condition says that at the solution the gradient vector \(\nabla f(\bar{x})\), can be written as a nonnegative linear combination of the gradients of the constraints that are satisfied with equality

What does this mean geometrically:

\bullet means that the gradient vector of the objective function points into the cone defined by the gradients of the constraints that are satisfied with equality. Go over what it means for a vector to be inside the positive cone defined by two others. Show geometrically that if a vector \(x\) is inside the cone, you can take positive scalar multiples of the vectors that define the cone, and reconstruct \(x\). Otherwise, you can’t reconstruct \(x\) with positive coefficients.
Figure 2. Graphical Illustration of the KKT conditions: two constraints

- to see why must this condition hold geometrically, we draw a picture (Fig. 2) that has only the gradients drawn in (none of the constraint sets or indifference curves will be drawn in) and argue from the vectors alone why the above condition is necessary.

- Suppose the above condition is violated, so that $\nabla f(\bar{x})$ lies outside the cone defined by the gradients of the constraints that are satisfied with equality. Clearly, we can then draw a line (the horizontal dotted line in Fig. 2) such that the gradient vectors of all of the constraints lie on one side of the line, and the gradient vector of the objective function lies on the other side. (Again this is obvious geometrically, but requires some work to prove rigorously.)
• We can now draw a vector $\mathbf{d}\mathbf{x}$ that makes an acute angle with $\nabla f(\mathbf{x})$ but an obtuse angle with all of the constraint vectors. We can *always* do this: make $\mathbf{d}\mathbf{x}$ arbitrarily close to the line perpendicular to the first dotted line, which ensures that $\mathbf{d}\mathbf{x}$ will make an obtuse angle with the constraint gradients, but must make an acute angle with $\nabla f(\mathbf{x})$ since both vectors are trapped within the quadrant defined by the dotted lines.

• Observe that it makes an obtuse angle with all of the constraint vectors.

• Reason from there:
  – add $\mathbf{d}\mathbf{x}$ to $\mathbf{x}$
  – observe that $f(\mathbf{x} + \mathbf{d}\mathbf{x}) - f(\mathbf{x}) \approx \nabla f(\mathbf{x}) \cdot \mathbf{d}\mathbf{x} > 0$;
  – i.e., moving in this direction increases the objective function.
  – similarly, for all $j$ such that the $j$'th constraint is satisfied with equality, observe that $g^j(\mathbf{x} + \mathbf{d}\mathbf{x}) - g^j(\mathbf{x}) \approx \nabla g^j(\mathbf{x}) \cdot \mathbf{d}\mathbf{x} < 0$, i.e., reduces the value of all of the constraints that are satisfied with equality.
  – for $j$'s that aren’t, if $\mathbf{d}\mathbf{x}$ is sufficiently small, then you can increase the value of $g^j(\cdot)$ a little bit and still be less than $b^j$.
  – in other words, you can move a little in the direction of $\mathbf{d}\mathbf{x}$ and stay within the constraint set, yet increase the value of $f$.
  – This establishes that $\mathbf{x}$ couldn’t have been a maximum.