KKT Examples

Stanley B. Gershwin*
Massachusetts Institute of Technology

The purpose of this note is to supplement the slides that describe the Karush-Kuhn-Tucker conditions. Neither these notes nor the slides are a complete description of these conditions; they are only intended to provide some intuition about how the conditions are sometimes used and what they mean.

The KKT conditions are usually not solved directly in the analysis of practical large nonlinear programming problems by software packages. Iterative successive approximation methods are most often used. The results, however they are obtained, must satisfy these conditions.

Example 0

No constraints:

$$\min J = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

The solution is $x_1 = x_2 = x_3 = x_4 = 0$ and $J = 0$.

Example 1

One equality constraint:

$$\min J = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

subject to

$$x_1 + x_2 + x_3 + x_4 = 1 \tag{1}$$

Solution: Adjoin the constraint

$$\min \bar{J} = x_1^2 + x_2^2 + x_3^2 + x_4^2 + \lambda(1 - x_1 - x_2 - x_3 - x_4)$$

subject to 1

$$\boxed{x_1 + x_2 + x_3 + x_4 = 1}$$

In this context, λ is called a *Lagrange multiplier.* The KKT conditions reduce, in this case, to setting $\partial \bar{J}/\partial x$ to zero:

2

$$\boxed{\frac{\partial \bar{J}}{\partial x} = \begin{pmatrix} 2x_1 - \lambda \\ 2x_2 - \lambda \\ 2x_3 - \lambda \\ 2x_4 - \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}} \tag{2}$$

Therefore

$$x_1 = x_2 = x_3 = x_4 = \frac{\lambda}{2}$$

so

$$x_1 + x_2 + x_3 + x_4 = 4\left(\frac{\lambda}{2}\right) = 1 \quad \text{or} \quad \lambda = \frac{1}{2}$$

so

$$x_1 = x_2 = x_3 = x_4 = \frac{1}{4} \text{ and } J = \frac{1}{4}$$

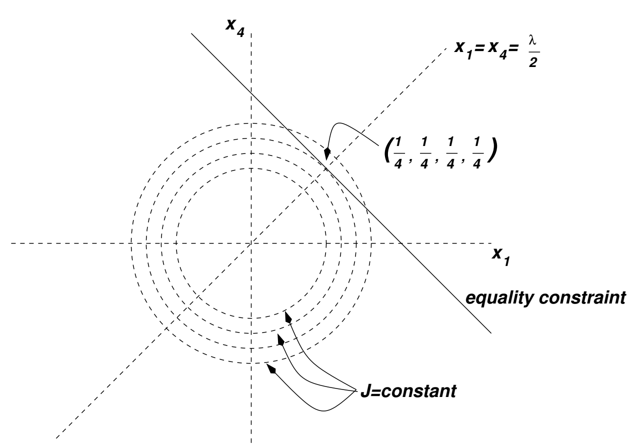


Figure 1: Example 1, represented in two dimensions

Comments

- *Why* should we adjoin the constraints? The answer to this question has two parts:

  - ⋆ First, it does no harm. There is no difference between J and $\bar{J}$ for any set of x_i that satisfies the problem since the set of x_i that satisfies the problem must satisfy (1).

    Furthermore, it works. This may not feel like a very satisfying answer, but this is also why we plug $x = Ce^{\lambda t}$ into a linear differential equation with constant coefficients. We could have done a zillion other things that would have done no harm, but they would not have done any good either. There is not much point in studying them; we only study what works.

  - ⋆ In this case, it provides a relationship that we can use among the components of x (equation (2)). In general, it replaces the minimization requirement with a set of equations and inequalities, and there are just enough of them to determine a unique solution (when the problem has a unique solution).

- The solution is illustrated in Figure 1. We are seeking the smallest 4-dimensional sphere that intersects with the equality constraint (a 3-dimensional plane in 4-dimensional space). Equation (2) essentially tells us that the solution point is on a line that intersects with that plane.

- If we asked for the maximum rather than the minimum of J, the same necessary conditions would have applied, and we would have gotten the same answer following the steps shown here. However, that answer would be ***wrong***. After all, if it is a minimum, it cannot also be a maximum unless the function is a constant, which it certainly is not.

  We know that we have found a minimum because of the second order conditions: the second derivative matrix $\partial^2 J/\partial x^2$ is positive definite.

- It is a coincidence that for both Example 0 and Example 1 $J = x_i, i = 1, ..., 4$ at the optimum. What is *not* a coincidence is that J for Example 1 is greater than J for Example 0. If you change an optimization problem by adding a constraint, you make the optimum worse; or, at best, you leave it unchanged.

Example 2

One equality constraint and one inequality constraint:

$$\min J = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

subject to

$$x_1 + x_2 + x_3 + x_4 = 1 \tag{3}$$

$$x_4 \leq A \tag{4}$$

in which A is a parameter that we will play with.

Figures 2 and 3 illustrate two possible versions of this problem, depending on the value of A. (The shaded regions are the forbidden values of x, the places where $x_4 > A$.)

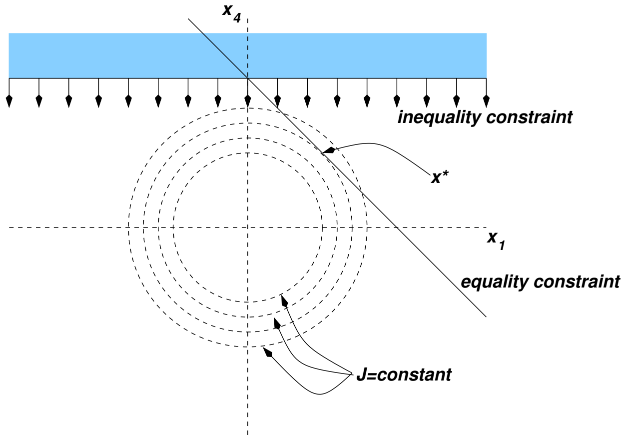


Figure 2: Example 2, A large

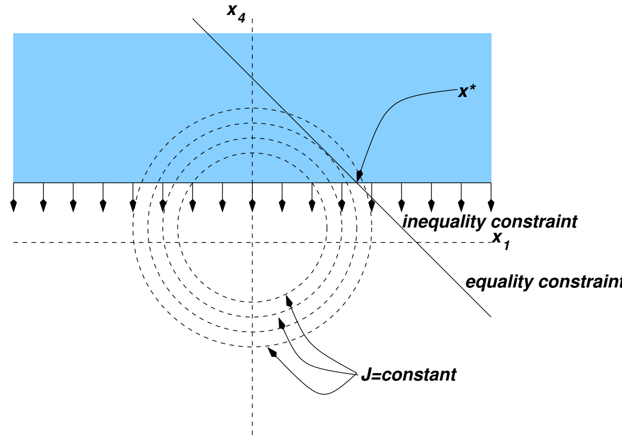


Figure 3: Example 2, A small

Digression: The inequality constraint requires a new Lagrange multiplier. To understand it, let us temporarily ignore the equality constraint and consider the following scalar problem, in which J and g are arbitrary functions that are differentiable, whose derivatives are continuous, and where J has a minimum:

$$\min_x J(x) \tag{5}$$

subject to

$$g(x) \leq 0$$

There are two possibilities: the solution $x^\star$ satisfies $g(x^\star) < 0$ (ie, where the solution is strictly in the interior of the inequality condition) or it satisfies $g(x^\star) = 0$ (where the solution is on the boundary of the interior of the inequality condition). Figures 4 and 5 illustrate these two possibilities.

Possibility 1, the interior solution: The necessary condition is that $x^\star$ satisfies

$$\frac{dJ}{dx}(x^\star) = 0 \tag{6}$$

That is, the solution is the same as that of the problem without the inequality constraint.

Possibility 2, the boundary solution: Let $x = x^\star + \delta x$. For $x^\star$ to be the optimal solution to (5), $\delta x = 0$ must be the solution to

$$\min_{\delta x} J(x^\star + \delta x) \tag{7}$$

subject to

$$g(x^\star + \delta x) \leq 0$$

This is certainly true if we restrict our attention to small δx. In that case, we can expand J and g as first order Taylor approximations and (7) becomes, approximately

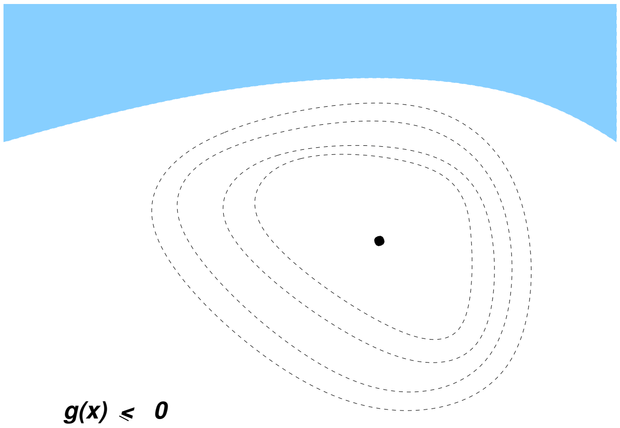


Figure 4: Case 1, Interior solution

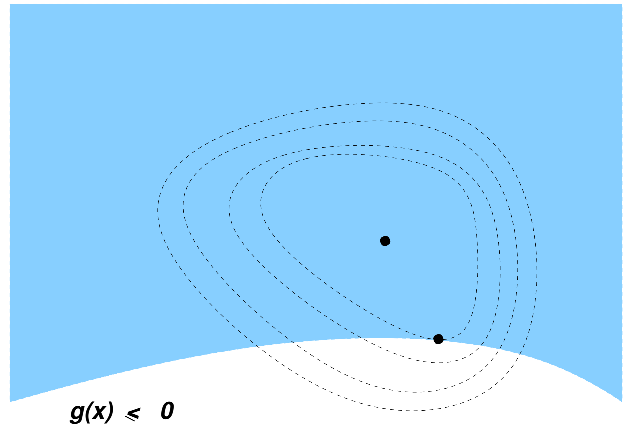


Figure 5: Case 2, Boundary solution

$$\min_{\delta x} J(x^\star) + \frac{dJ}{dx}(x^\star)\delta x$$

subject to

$$g(x^\star) + \frac{dg}{dx}(x^\star)\delta x \leq 0$$

Since $J(x^\star)$ is independent of δx and $g(x^\star) = 0$, this can be written

$$\min_{\delta x} \frac{dJ}{dx}(x^\star)\delta x \tag{8}$$

subject to

$$\frac{dg}{dx}(x^\star)\delta x \leq 0$$

We now seek conditions on $\dfrac{dJ}{dx}(x^\star)$ and $\dfrac{dg}{dx}(x^\star)$ such that the solution to (8) is $\delta x = 0$. Note that (8) is a linear programming problem.

To be really exhaustive about it, we must now consider the four cases in the table below[1]. Case 1, for example, reduces to

$$\min_{\delta x} \delta x$$

subject to

$$\delta x \leq 0$$

[1]Actually, there are nine cases, since we could also have $dJ/dx = 0$ or $dg/dx = 0$. However, if $dJ/dx = 0$ at $x = x^*$, then we really have a situation like the interior solution. That is, the constraint is still ineffective since we would have the same x^* even if we ignored the $g(x) \leq 0$ condition. If $dg/dx = 0$ at $x = x^*$, then the constraint qualification is violated. This is discussed below.

since the magnitudes do not matter — only the signs matter. The solution is clearly $\delta x = -\infty$. Similarly, Case 2 becomes

$$\min_{\delta x} \delta x$$

subject to

$$\delta x \geq 0$$

and it is clear that the solution is $\delta x = 0$. The other cases are similar.

Case	$\dfrac{dJ}{dx}(x^\star)$	$\dfrac{dg}{dx}(x^\star)$	Solution
1	> 0	> 0	$\delta x = -\infty$
2	> 0	< 0	$\delta x = 0$
3	< 0	> 0	$\delta x = 0$
4	< 0	< 0	$\delta x = \infty$

Therefore, the only cases in which the solution is $\delta x = 0$ are Cases 2 and 3, the cases in which the signs of $\dfrac{dJ}{dx}(x^\star)$ and $\dfrac{dg}{dx}(x^\star)$ are opposite. Therefore, there is some ***positive*** number μ such that

$$\frac{dJ}{dx}(x^\star) = -\mu \frac{dg}{dx}(x^\star)$$

or

$$\frac{dJ}{dx}(x^\star) + \mu \frac{dg}{dx}(x^\star) = 0$$

Equation (6) is implied by this if we require $\mu = 0$ when $g(x^\star) < 0$.

Solution: For the problem of Example 2, define

$$\bar{J} = x_1^2 + x_2^2 + x_3^2 + x_4^2 + \lambda(1 - x_1 - x_2 - x_3 - x_4) + \mu(x_4 - A)$$

Then, the KKT conditions are

$$\frac{\partial \bar{J}}{\partial x} = 0 \tag{9}$$

$$x_1 + x_2 + x_3 + x_4 = 1 \tag{10}$$

$$x_4 \le A \tag{11}$$

$$\mu \ge 0 \tag{12}$$

$$\mu(x_4 - A) = 0 \tag{13}$$

From (9):

$$\frac{\partial \bar{J}}{\partial x} = \begin{pmatrix} 2x_1 - \lambda \\ 2x_2 - \lambda \\ 2x_3 - \lambda \\ 2x_4 - \lambda + \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Therefore

$$x_1 = x_2 = x_3 = \frac{\lambda}{2},\ x_4 = \frac{\lambda - \mu}{2}$$

so, from (10),

$$x_1 + x_2 + x_3 + x_4 = 4\left(\frac{\lambda}{2}\right) - \frac{\mu}{2} = 1$$

or

$$4\lambda - \mu = 2 \text{ or } \lambda = \frac{2+\mu}{4}$$

Therefore

1

$$x_1 = x_2 = x_3 = \frac{2+\mu}{8} = \frac{1}{4} + \frac{\mu}{8},\ x_4 = \frac{2+\mu}{8} - \frac{\mu}{2} = \frac{1}{4} - \frac{3\mu}{8} \tag{14}$$

From (11),

$$\frac{1}{4} - \frac{3\mu}{8} \le A$$

or

2

$$\frac{3\mu}{8} \ge \frac{1}{4} - A \tag{15}$$

Case 1: $A > 1/4$ This is the interior case illustrated in Figures 2 and 4. Since $1/4 - A \le 0$, (15) implies that (12) is automatically satisfied. From (14)

$$x_1 = x_2 = x_3 \ge \frac{1}{4}; \quad x_4 = 1 - (x_1 + x_2 + x_3) \le \frac{1}{4}$$

But we have more from (14): (13) implies that $\mu = 0$. Therefore

$$x_1 = x_2 = x_3 = x_4 = \frac{1}{4}$$

which is consistent with the answer to Example 1 and common sense. Example 1 says that this is optimal; if we also require that x_4 is less than A and A is greater than $1/4$, we haven't changed anything. Note that the optimal J is again $1/4$.

Case 2: $A = 1/4$ This behaves like Case 1. The unconstrained optimum lies on the boundary. Therefore, if we ignored the inequality constraint, we would get the same x^*.

Case 3: $A < 1/4$ If x_4 were strictly less than A, then (13) would require that $\mu = 0$. But then (14) would imply $x = 1/4$, which violates (11).

Therefore $x_4 = A$ and

$$x_1 = x_2 = x_3 = \frac{1}{3}(1 - A)$$

Also

$$\begin{aligned} J &= 3\left(\frac{1}{9}(1-A)^2\right) + A^2 = \frac{1}{3}(1-A)^2 + A^2 \\ &= \frac{1}{3}\left(1 - 2A + 4A^2\right) \end{aligned}$$

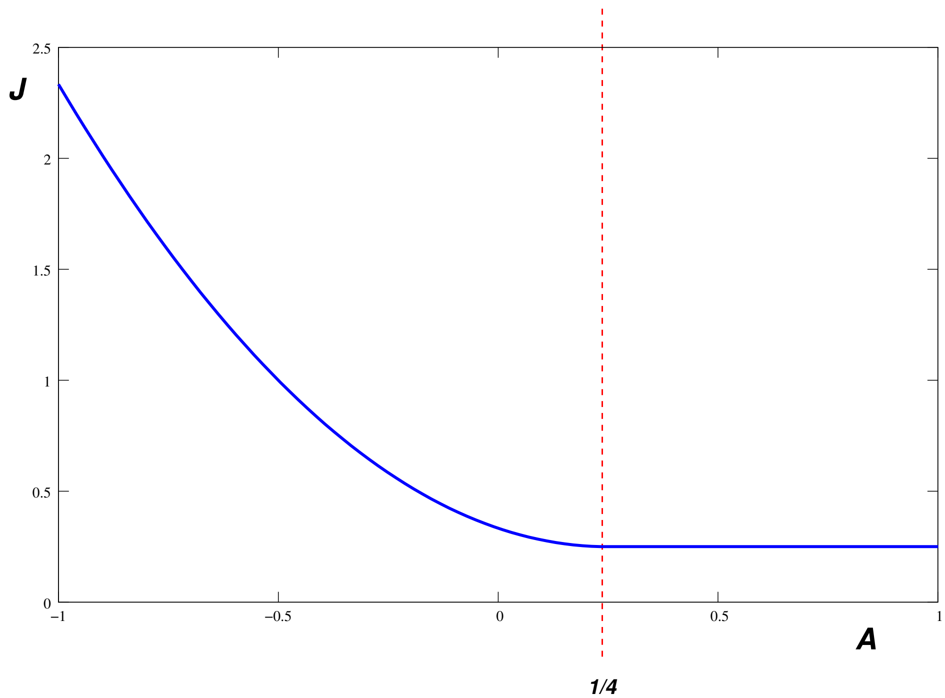
and that $J \geq 1/4$ and that $J = 1/4$ when $A = 1/4$.

Comments

- The comments after Example 1 hold here.
-

$$J = \begin{cases} \dfrac{1}{4} & \text{if } A \geq \dfrac{1}{4} \\[2ex] \dfrac{1}{3}\left(1 - 2A + 4A^2\right) & \text{otherwise} \end{cases}$$

The graph of J as a function of A is:



- The extension of the KKTs to more than one equation and/or more than one inequality constraint is straightforward, but there is one more condition to be applied in general, the so-called *constraint qualification*. This says that the gradients of the equations and of the effective inequality constraints must be linearly independent at the solution $x^\star$.

- The informal discussion of the KKT conditions was modeled on that of Bryson and Ho (1975). There are plenty of other references, but this discussion is especially intuitive.

References

Bryson, A. E. and Y.-C. Ho (1975). *Applied Optimal Control: Optimization, Estimation, and Control*. Hemisphere Publishing Corporation.