Lecture Notes on Machine Learning: The Karush-Kuhn-Tucker Conditions (Part 2)

Technical Report · August 2019

2 authors:


Christian Bauckhage
University of Bonn


Tiansi Dong
University of Bonn

Lecture Notes on Machine Learning
The Karush-Kuhn-Tucker Conditions (Part 2)

Christian Bauckhage and Tiansi Dong
B-IT, University of Bonn

This note demonstrates how to solve a (simple) inequality constrained optimization problem. In passing, we introduce the notions of active and inactive constraints as well as of quadratic programming.

Introduction

In our previous note within this mini series on optimization,[1] we introduced the Karush-Kuhn-Tucker conditions that come into play when solving *inequality constrained* problems. Here, we consider a rather didactic example of such a problem and walk through the process of solving it. This should help us to develop a more practical understanding of the KKT conditions and their significance.

[1] C. Bauckhage and D. Speicher. Lecture Notes on Machine Learning: The Karush-Kuhn-Tucker Conditions (Part 1). B-IT, University of Bonn, 2019b

An Inequality Constrained Problem . . .

To keep things simple and intuitive, we next look at a constrained convex minimization problem over $\mathbb{R}^2$. Moreover, there are no no equality constraints and only two inequality constraints. That is, we consider the following kind of problem

$$\begin{aligned} \boldsymbol{x}^* = \underset{\boldsymbol{x} \in \mathbb{R}^2}{\operatorname{argmin}} \; & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h_1(\boldsymbol{x}) \leq 0 \\ & h_2(\boldsymbol{x}) \leq 0. \end{aligned} \tag{1}$$

To be more specific, suppose that the objective function is given by

$$f(\boldsymbol{x}) = \left(x_1 - 7\right)^2 + \left(x_2 - 3\right)^2 \tag{2}$$

and that the two inequality constraint functions are

$$h_1(\boldsymbol{x}) = x_1 + 3\,x_2 - 18 \tag{3}$$

$$h_2(\boldsymbol{x}) = 2\,x_1 + x_2 - 7. \tag{4}$$

Figure 1 provides an illustration of this problem. It shows contour lines of the objective $f(x)$ as well as the two half-spaces $\mathcal{HS}_1$ and $\mathcal{HS}_2$ defined by the constraints $h_1(x) \leq 0$ and $h_2(x) \leq 0$. Just from looking at this figure, we realize that the solution we are after must be found somewhere to lower left of the picture because this is where both half-spaces intersect. Any feasible minimum of our objective must reside in the region $\mathcal{HS}_1 \cap \mathcal{HS}_2$. But where exactly?

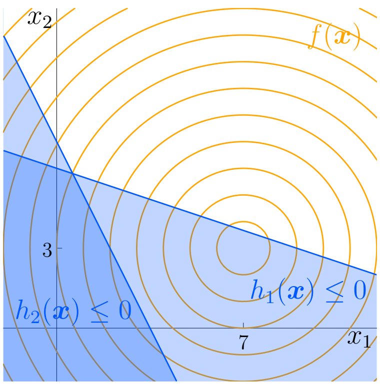

Figure 1: Contour plot of the objective function $f(x)$ in (2) and visualizations of the two half-spaces defined by the inequality constraint functions $h_1(x)$ and $h_2(x)$ in (3) and (4), respectively.

. . . and Its Solution

From our previous discussion, we deduce that the Lagrangian for our problem involves two Lagrange multipliers and is given by

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \mu_1\, h_1(\boldsymbol{x}) + \mu_2\, h_2(\boldsymbol{x}). \tag{5}$$

We also know that the first KKT condition (stationary) requires us to solve $\nabla f(\boldsymbol{x}) + \mu_1 \nabla h_1(\boldsymbol{x}) + \mu_2 \nabla h_2(\boldsymbol{x}) = \mathbf{0}$. Written in coordinates, this becomes

$$\begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} + \mu_1 \begin{bmatrix} \frac{\partial h_1}{\partial x_1} \\ \frac{\partial h_1}{\partial x_2} \end{bmatrix} + \mu_2 \begin{bmatrix} \frac{\partial h_2}{\partial x_1} \\ \frac{\partial h_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{6}$$

and, upon deriving (2)–(4), leads to

$$2\,x_1 + 0\,x_2 + 1\,\mu_1 + 2\,\mu_2 = 14 \tag{7}$$
$$0\,x_1 + 2\,x_2 + 3\,\mu_1 + 1\,\mu_2 = 6. \tag{8}$$

Next, we assume that both inequality constraints of our problem are ***active***. That is, we pretend they hold with equality which hence provides us with

$$1\,x_1 + 3\,x_2 = 18 \tag{9}$$
$$2\,x_1 + 1\,x_2 = 7. \tag{10}$$

This way, we obtain 4 equations in 4 unknowns (x_1, x_2, μ_1, μ_2). As our equations are linear, we may write them in matrix-vector form

$$\begin{bmatrix} 2 & 0 & 1 & 2 \\ 0 & 2 & 3 & 1 \\ 1 & 3 & 0 & 0 \\ 2 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 14 \\ 6 \\ 18 \\ 7 \end{bmatrix}. \tag{11}$$

Left multiplying both sides of this equation by the inverse of the matrix then leads to the solution

$$\begin{bmatrix} x_1^* \\ x_2^* \\ \mu_1^* \\ \mu_2^* \end{bmatrix} = \begin{bmatrix} 0.6 \\ 5.8 \\ -4.8 \\ 8.8 \end{bmatrix}. \tag{12}$$

This, however, violates the third KKT condition (dual feasibility) which demands that Lagrange multipliers for inequality constraints must be non-negative.

We therefore ***deactivate*** the first constraint and simply declare that $\mu_1 = 0$. This is equivalent to removing one of the unknowns and thus one of the equations from our problem and leads to a modified or reduced problem in only 3 unknowns (x_1, x_2, μ_2), namely

$$\begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & 1 \\ 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 14 \\ 6 \\ 7 \end{bmatrix}. \tag{13}$$

Solving for the variables in this new problem results in

$$\begin{bmatrix} x_1^* \\ x_2^* \\ \mu_2^* \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} \tag{14}$$

which satisfies all KKT conditions and therefore constitutes a feasible solution. The point $\boldsymbol{x}^* = [x_1^*\ x_2^*]^\intercal$ contained in the vector on the left hand side is shown in Fig. 2; as expected, it lies in the intersection of the two constraint half-spaces.

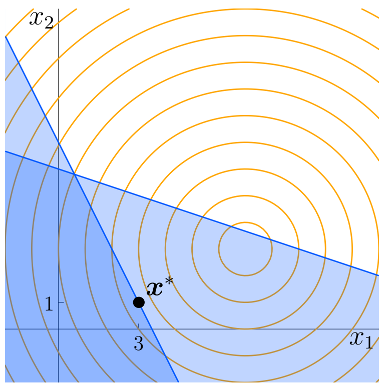

Figure 2: Solution $\boldsymbol{x}^*$ to the constrained problem specified in (1)–(4).

Discussion

Looking at what we just did, there are several aspects that merit further elaboration.

FIRST OF ALL, what did we just do? Well, we went through the steps of a *two stage algorithm* that apparently solves convex minimization problems under inequality constraints.

Our particular problem led to a Lagrangian with 4 parameters (x_1, x_2, μ_1, μ_2). To determine their optimal values, we used the first KKT condition to obtain 2 equations in the 4 unknowns. Assuming that both constraints of our problem were ***active*** led to 2 additional equations. Since all these equations happened to be linear, we had a solvable algebraic problem at our hands.

However, its solution violated the third KKT condition. This meant that our initial guess of both constrained being active was incorrect. In the second stage of the algorithm, we therefore eliminated the ***inactive*** constraint, i.e. fixed its Lagrange multiplier to zero, and solved the resulting reduced system of 3 equations in 3 unknowns. This produced a feasible solution $\boldsymbol{x}^*$ to our problem.

active and **inactive constraints**

Note that an inequality constraint j is called ***active*** (or binding) if $h_j(\boldsymbol{x}^*) = 0$ and ***inactive*** if $h_j(\boldsymbol{x}^*) < 0$.

Geometrically, this is to say that, if a constraint $1 \leq j \leq q$ is active, then $\boldsymbol{x}^*$ resides on the surface of the region defined by $h_j(\boldsymbol{x}) \leq 0$; if it is inactive, then $\boldsymbol{x}^*$ resides inside of this region. For our example, we can clearly see this in Fig. 2: $\boldsymbol{x}^*$ lies on the boundary of the region defined by $h_2(\boldsymbol{x}) \leq 0$ and well inside the region defined by $h_1(\boldsymbol{x}) \leq 0$.

The crucial observation is that inactive constraints do not impact the solution of a constrained optimization problem and may therefore be ignored. This, too, can be seen in Fig. 2. It shows that it is indeed sufficient to only work with the constraint $h_2(\boldsymbol{x}) \leq 0$ in order to determine the solution to our problem.

Whether or not a constraint is active is determined by checking KKT 3 after the first stage of the above algorithm; the ignoring of inactive constraints then happens in the second stage.

SECOND OF ALL, if we define $\boldsymbol{q} = [7\ 3]^\intercal \in \mathbb{R}^2$, we realize that the objective function in (2) can also be written as

$$f(\boldsymbol{x}) = \left\|\boldsymbol{x} - \boldsymbol{q}\right\|^2. \tag{15}$$

But this is to say that we are dealing with another instance of a "closest point" problem. We already looked at some of these in an earlier note.[2] Back then, we were interested in the closest point within the intersection of two hyperplanes; here, we are interested in the closest point in the intersection of two half-spaces.

[2] C. Bauckhage and T. Donge. Lecture Notes on Machine Learning: Lagrange Multipliers (Part 2). B-IT, University of Bonn, 2019

While the former led to an equality constrained minimization problem, we just saw that the latter leads to an inequality constrained minimization problem.

Third of all, we note that the two inequality constraint functions $h_1(\boldsymbol{x})$ and $h_2(\boldsymbol{x})$ in (3) and (4), respectively, are linear in $\boldsymbol{x}$. Hence, if we introduce

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{b} = \begin{bmatrix} 18 \\ 7 \end{bmatrix}, \tag{16}$$

we can express (1)–(4) much more succinctly, namely as

$$\begin{aligned} \boldsymbol{x}^* = \underset{\boldsymbol{x} \in \mathbb{R}^2}{\operatorname{argmin}} \ & \left\| \boldsymbol{x} - \boldsymbol{q} \right\|^2 \\ \text{s.t.} \quad & \boldsymbol{A}\,\boldsymbol{x} \preceq \boldsymbol{b}. \end{aligned} \tag{17}$$

Earlier,[3] we said that constrained optimization problems of this type are rather commonplace; our example in this note therefore corroborates this claim.

[3] C. Bauckhage and D. Speicher. Lecture Notes on Machine Learning: Equality and Inequality Constraints. B-IT, University of Bonn, 2019a

Indeed, the problem formulation in (17) makes it explicit that we just dealt with a *linearly constrained quadratic optimization problem* or ***quadratic programming problem*** for short.

quadratic programming problem

In machine learning, quadratic programming problems arise all over the place. In their most general form, they are given by[4]

$$\begin{aligned} \boldsymbol{x}^* = \underset{\boldsymbol{x} \in \mathbb{R}^m}{\operatorname{argmin}} \ & \boldsymbol{x}^\intercal \boldsymbol{P}\,\boldsymbol{x} + \boldsymbol{x}^\intercal \boldsymbol{p} \\ \text{s.t.} \quad & \boldsymbol{A}\,\boldsymbol{x} \preceq \boldsymbol{b} \\ & \boldsymbol{C}\,\boldsymbol{x} = \boldsymbol{d}. \end{aligned} \tag{18}$$

[4] Observe that, for the objective function of our problem, we have

$$\begin{aligned} \left\| \boldsymbol{x} - \boldsymbol{q} \right\|^2 &= \boldsymbol{x}^\intercal \boldsymbol{x} - 2\boldsymbol{x}^\intercal \boldsymbol{q} + \boldsymbol{q}^\intercal \boldsymbol{q} \\ &\propto \boldsymbol{x}^\intercal \boldsymbol{P}\,\boldsymbol{x} + \boldsymbol{x}^\intercal \boldsymbol{p} \end{aligned}$$

where $\boldsymbol{P} = \boldsymbol{I}$ and $\boldsymbol{p} = -2\,\boldsymbol{q}$.

While this may look daunting, our practical examples up to this point demonstrate that solving a quadratic program might be a bit involved but, in the end, is actually easy.

Summary and Outlook

In this note, we showed how to work with the KKT conditions to solve a simple inequality constrained optimization problem.

In particular, we went through a two stage procedure. In the first stage, we used KKT 1 and equations for active inequality constraints to set up expressions for the variables we wanted to solve for. Using KKT 3, we then identified inactive constraints and eliminated them in the second stage of the procedure to obtain the sought after solution.

We also pointed out that the example we considered in this note is an instance of a quadratic programming problem. Further examples of this kind of problem will be discussed later on.

Acknowledgments

This material was prepared within project P3ML which is funded by the Ministry of Education and Research of Germany (BMBF) under grant number 01/S17064. The authors gratefully acknowledge this support.

References

C. Bauckhage and T. Donge. Lecture Notes on Machine Learning: Lagrange Multipliers (Part 2). B-IT, University of Bonn, 2019.

C. Bauckhage and D. Speicher. Lecture Notes on Machine Learning: Equality and Inequality Constraints. B-IT, University of Bonn, 2019a.

C. Bauckhage and D. Speicher. Lecture Notes on Machine Learning: The Karush-Kuhn-Tucker Conditions (Part 1). B-IT, University of Bonn, 2019b.