Lecture Notes on Machine Learning: The Karush-Kuhn-Tucker Conditions (Part 1)

Technical Report · August 2019

2 authors:

Christian Bauckhage
University of Bonn

Daniel Speicher
University of Bonn

Lecture Notes on Machine Learning
The Karush-Kuhn-Tucker Conditions (Part 1)

Christian Bauckhage and Daniel Speicher
B-IT, University of Bonn

Having studied how the method of Lagrange multipliers allows us to solve equality constrained optimization problems, we next look at the more general case of inequality constrained optimization problems. In short, these, too, can be dealt within the Lagrangian framework but require us to consider the Karush-Kuhn-Tucker conditions which we discuss in note.

Introduction

So far, we reviewed the notion of equality- and inequality constrained optimization problems[1] and studied the method of Lagrange multipliers as a tool especially for solving equality constrained problems.[2] Next, we extend our discussion to optimization problems involving inequality constraints

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

[2] C. Bauckhage and D. Speicher. Lecture Notes on Machine Learning: Lagrange Multipliers (Part 3). B-IT, University of Bonn, 2019b

$$\begin{aligned} \boldsymbol{x}^* = \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{R}^m} \; & f(\boldsymbol{x}) & \qquad \boldsymbol{x}^* = \operatorname*{argmax}_{\boldsymbol{x}\in\mathbb{R}^m} \; & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) = 0 & \text{s.t.} \quad & g_i(\boldsymbol{x}) = 0 \\ & h_j(\boldsymbol{x}) \leq 0 & & h_j(\boldsymbol{x}) \leq 0. \end{aligned} \tag{1}$$

Throughout, we assume that the objective function f, the $1 \leq i \leq p$ equality constraint functions g_i, as well as the $1 \leq j \leq q$ inequality constraint functions h_j are continuously differentiable mappings from $\mathbb{R}^m$ to $\mathbb{R}$.

A KEY OBSERVATION regarding the problems in (1) is that we can solve them within the Lagrangian framework, too.[3] That is, we may again consider Lagrange functions and determine their stationary points in order to obtain feasible solutions.

[3] We will not prove this claim but take it as a given. Strictly speaking, the problem under consideration must satisfy some mild regularity conditions for the claim to be true. We ignore these and simply remark that they usually hold in the context of machine learning.

However, whereas the simpler case of merely equality constrained problems only required Lagrange multipliers λ_i for the g_i, we now need to introduce further multipliers μ_j for the inequality constraints h_j. In other words, in the more general setting, we need to work with Lagrangians $\mathcal{L} : \mathbb{R}^{m+p+q} \to \mathbb{R}$.

To be more precise, for constrained minimization problems as on the left in (1), we consider Lagrangians

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{i=1}^{p} \lambda_i \, g_i(\boldsymbol{x}) + \sum_{j=1}^{q} \mu_j \, h_j(\boldsymbol{x}) \tag{2}$$

and for constrained maximization problems as on the right in (1), the Lagrangians amount to

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) - \sum_{i=1}^{p} \lambda_i \, g_i(\boldsymbol{x}) - \sum_{j=1}^{q} \mu_j \, h_j(\boldsymbol{x}). \tag{3}$$

STATIONARY POINTS $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ of these extended Lagrangians can again be determined by solving $\nabla \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0}$. However, since inequality constrained problems are (slightly) more involved than mere equality constrained ones, it should not be surprising that a stationary point now has to satisfy more involved conditions in order for $\boldsymbol{x}^*$ to be a feasible solution to the original problem.

The Karush-Kuhn-Tucker Conditions

The fact that the method of Lagrange multipliers also applies to problems with inequality constraints rests on the following assertion: If $\boldsymbol{x}^* \in \mathbb{R}^m$ feasibly solves the minimization- or maximization problem in (1) whose Lagrangians are given by (2) and (3), respectively, then there exist vectors of Lagrange multipliers $\boldsymbol{\lambda}^* \in \mathbb{R}^p$ and $\boldsymbol{\mu}^* \in \mathbb{R}^q$ such that the ***Karush-Kuhn-Tucker conditions*** are met.

Karush-Kuhn-Tucker or simply **KKT conditions**

The Karush-Kuhn-Tucker conditions are necessary conditions and of utmost practical importance far beyond machine learning. Next, we list them by their commonly agreed upon names.

KKT 1: Stationarity

A feasible solution of the constrained minimization- or maximization problems in (1), must obey

$$\nabla f(\boldsymbol{x}^*) + \sum_i \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_j \mu_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0} \tag{4}$$

or

$$\nabla f(\boldsymbol{x}^*) - \sum_i \lambda_i^* \nabla g_i(\boldsymbol{x}^*) - \sum_j \mu_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0}, \tag{5}$$

respectively. This obviously generalizes the gradient conditions for equality constrained problems we discussed earlier.[4] Note, however, that we emphasize the signs of the Lagrange multipliers. Contrary to mere equality constrained problems where they do not matter, here they do.

[4] C. Bauckhage and D. Speicher. Lecture Notes on Machine Learning: Lagrange Multipliers (Part 3). B-IT, University of Bonn, 2019b

KKT 2: Primal Feasibility

Any feasible solution to an inequality constrained problem must of course comply with the constraints, hence

$$g_i(\boldsymbol{x}^*) = 0 \quad \forall\, 1 \leq i \leq p \tag{6}$$

$$h_j(\boldsymbol{x}^*) \leq 0 \quad \forall\, 1 \leq j \leq q. \tag{7}$$

KKT 3: Dual Feasibility

At a feasible solution, the Lagrange multipliers of all the inequality constraints must be non-negative, that is

$$\mu_j^* \geq 0 \quad \forall\, 1 \leq j \leq q. \tag{8}$$

KKT 4: Complementary Slackness

At a feasible solution, either each inequality constraint must evaluate to zero, or the corresponding Lagrange multiplier must be zero. This condition is usually expressed as

$$\mu_j^* \, h_j(\boldsymbol{x}^*) = 0 \quad \forall \, 1 \leq j \leq q. \tag{9}$$

AT THIS POINT, all of this may seem abstract and hardly to make sense. In our next note, we will therefore look at a simple inequality constrained problem and show how the KKT conditions come into play when solving it. Later on, we will see that many machine learning problems are of either of the forms in (1) and there will be ample opportunities for us to work with the KKT conditions.

Summary and Outlook

In this note, we claimed (without proof) that the method of Lagrange multipliers generalizes to solving *inequality constrained* optimization problems. However, this generalization requires us to consider the Karush-Kuhn-Tucker conditions which we listed in this note.

In other words, if we use Lagrange multipliers to solve inequality constrained problems, we must verify the conditions in (4)/(5)–(9) to rest assured that we obtain feasible solutions.

It is no exaggeration to say that the KKT conditions are of fundamental importance to our civilization. Airplanes fly, bridges stand, and portfolios can be managed because humankind knows how to solve constrained optimization problems. In machine learning, too, constrained optimization problems occur all over the place and we will study various practical examples later on.

Notes and Further Reading

The Karush-Kuhn-Tucker conditions are named after the mathematicians William Karush (∗1917, †1997), Harold W. Kuhn (∗1925, †2014), and Albert W. Tucker (∗1905, †1995). Karush first formalized them in his 1939 master's thesis[5] but this was widely realized only long after Kuhn and Tucker independently published their Kuhn-Tucker theorem in 1951.[6]

For a more formal introduction to the KKT conditions, readers may consult the book by Boyd and Vandenberghe.[7] At present, it is *the* standard textbook on optimization and freely available on Boyd's homepage.

[5] W. Karush. Minima of Functions of Several Variables with Inequalities as Side Constraints. Master's thesis, University of Chicago, 1939

[6] H.W. Kuhn and A.W. Tucker. Nonlinear Programming. In *Proc. 2nd Berkley Symposium on Mathematical Statistics and Probability*, 1951

[7] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004

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 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: Lagrange Multipliers (Part 3). B-IT, University of Bonn, 2019b.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

W. Karush. Minima of Functions of Several Variables with Inequalities as Side Constraints. Master's thesis, University of Chicago, 1939.

H.W. Kuhn and A.W. Tucker. Nonlinear Programming. In *Proc. 2nd Berkley Symposium on Mathematical Statistics and Probability*, 1951.