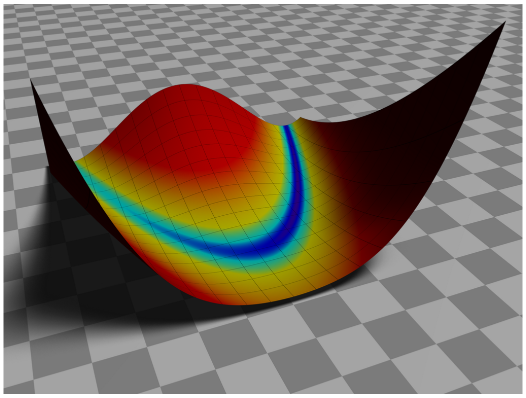
Karush-Kuhn-Tucker conditions

Lecture 15-16



ME EN 575
Andrew Ning
aning@byu.edu

Outline

Equality Constraints

Inequality Constraints

$$
\begin{aligned}
\text{minimize} \quad & f(x) \\
\text{with respect to} \quad & x \in \mathbb{R}^n \\
\text{subject to} \quad & lb < x < ub \\
& \hat{c}_j(x) = 0, \quad j = 1, \ldots, \hat{m} \\
& c_k(x) \leq 0, \quad k = 1, \ldots, m
\end{aligned}
$$

Please note slightly different notation than text.

Bound constraints are straightforward to deal with.

Inequality constraints are used more frequently in engineering, but we will start with equality constraints because the math is a bit easier.

Remember, we will use $\hat{c}$ for equality constraints and c for inequality constraints.

Equality Constraints

Motivating Problem:

$$\begin{aligned} \text{minimize} \quad & x_1 + x_2 \\ \text{subject to} \quad & x_1^2 + x_2^2 = 8 \end{aligned}$$

Discuss in a group of size 2-3

- Rewrite the constraint in our convention.
- Draw the direction of the function gradient: ∇f. What is the unconstrained optimum?
- Draw the direction(s) for the constraint gradients: ∇c.
- What is the optimal solution and where is it located?
- What do you notice about ∇f and ∇c at the optimum. Does that make sense?
- How would you define the optimality criteria mathematically?

∇f and ∇c are parallel:

$$\nabla f(x^*) = -\lambda \nabla c(x^*)$$

(minus sign just for convenience)

[picture]

A More Formal Motivation

Let's take a first-order Taylor's series expansion of our function.

$$f(x+s) \approx f(x) + \nabla f(x)^T s$$

If we were at a minimum point then

$$f(x+s) \geq f(x) \text{ for all feasible } s$$

Thus:

$$\nabla f(x)^T s \geq 0$$

$$\nabla f(x)^T s \geq 0$$

If unconstrained, all s are possible so this could only be true if $\nabla f(x) = 0$.

If constrained, let's assume there was a better point:

$$\nabla f(x)^T s < 0$$

The better point would also have to be feasible (first-order expansion):

$$\hat{c}(x+s) = \hat{c}(x) + \nabla \hat{c}(x)^T s = 0$$

Because $\hat{c}(x) = 0$ and $\hat{c}(x+s) = 0$ if feasible

$$\nabla \hat{c}(x)^T s = 0$$

We now have two conditions:

$$\nabla f(x)^T s < 0$$

$$\nabla \hat{c}(x)^T s = 0$$

This always has a solution unless ∇f and $\nabla \hat{c}$ are parallel.

In other words for this to be a minimum

$$\nabla f(x^*) = -\lambda \nabla c(x^*)$$

or:

$$\nabla f(x^*) + \lambda \nabla c(x^*) = 0$$

Define the Lagrangian

$$\mathcal{L}(x, \lambda) = f(x) + \lambda \hat{c}(x)$$

Instead of $\nabla g = 0$ we need $\nabla \mathcal{L} = 0$.

But $\mathcal{L}$ is a function of both x and λ so we need to set both partial derivatives equal to 0.

$$\nabla_x \mathcal{L}(x, \lambda) = \nabla f(x^*) + \lambda \nabla \hat{c}(x^*) = 0$$
$$\nabla_\lambda \mathcal{L}(x, \lambda) = \hat{c}(x) = 0$$

Notice, satisfaction of our equality constraint is an automatic outcome of this definition.

Extend to m constraints

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x_i} &= \frac{\partial f}{\partial x_i} - \sum_{j=1}^{\hat{m}} \hat{\lambda}_j \frac{\partial \hat{c}_j}{\partial x_i} = 0, \quad (i = 1, \ldots, n) \\
\frac{\partial \mathcal{L}}{\partial \hat{\lambda}_j} &= \hat{c}_j = 0, \quad (j = 1, \ldots, \hat{m}).
\end{aligned}
$$

Just like in unconstrained where $\nabla f = 0$ was necessary, but not sufficient we need a second order condition. What do you think it is?

$$w^T \nabla_{xx} \mathcal{L}(x, \lambda) w > 0 \text{ (positive definite)}$$

λ_j is called a Lagrange multiplier, and there is a separate one for each constraint.

Inequality Constraints

Recall, our optimality criteria so far:

$$\nabla \mathcal{L} = 0$$

$$\hat{c} = 0$$

$$c \leq 0$$

What else do we need to add to these conditions if adding inequality constraints?

Motivating Example

$$\begin{aligned} \text{minimize} \quad & x_1 + x_2 \\ \text{subject to} \quad & x_1^2 + x_2^2 \leq 8 \end{aligned}$$

Discuss again in your group

- Rewrite the constraint in our format.
- Assume the optimal point is in the interior. What is the sign of c? What does this imply about ∇f for optimality? Using our Lagrangian, what does that mean about λ?
- Now, imagine the optimal point is right on the boundary. What is the value of c? How ∇f related to ∇c and can you relate it to the Lagrangian? What does this imply about the sign of λ?

Either our constraint is inactive ($c < 0$) and $\lambda = 0$.

or

The constraint is active $c = 0$, and $\lambda > 0$.

We have discovered the complementarity condition

$$\lambda_i^* c_i(x^*) = 0$$

KKT Conditions

$$\begin{aligned} \nabla_x \mathcal{L}(x, \lambda) &= 0 \\ \hat{c}(x) &= 0 \\ c(x) &\leq 0 \\ \lambda &\geq 0 \\ \lambda c(x) &= 0 \end{aligned}$$

The formal proof is long, but consider the following:

We already know how to deal with equality constraints, so what if we changed our inequality constraints into equality constraints.

Instead of

$$c \leq 0$$

Let's change it to

$$c = -s^2$$

for some s (called a slack variable), or

$$c + s^2 = 0$$

Then our Lagrangian becomes:

$$\mathcal{L}(x, \hat{\lambda}, \lambda, s) = f(x) + \hat{\lambda}^T \hat{c}(x) + \lambda^T \left(c(x) + s^2 \right),$$

Now we just need to take a bunch of partial derivatives.

$$\nabla_x \mathcal{L} = 0 \Rightarrow \frac{\partial \mathcal{L}}{\partial x_i} =$$

$$\frac{\partial f}{\partial x_i} + \sum_{j=1}^{\hat{m}} \hat{\lambda}_j \frac{\partial \hat{c}_j}{\partial x_i} + \sum_{k=1}^{m} \lambda_k \frac{\partial c_k}{\partial x_i} = 0, \quad i = 1, \ldots, n$$

or group all c and $\hat{c}$ together for simplicity:

$$\frac{\partial f}{\partial x_i} + \sum_{j=1}^{m} \lambda_j \frac{\partial c_j}{\partial x_i} = 0, \quad i = 1, \ldots, n$$

$$\nabla_{\hat{\lambda}} \mathcal{L} = 0 \Rightarrow \frac{\partial \mathcal{L}}{\partial \hat{\lambda}_j} =$$

$$\hat{c}_j = 0, \quad j = 1, \ldots, \hat{m}$$

(recovered our equality constraints)

$$\nabla_\lambda \mathcal{L} = 0 \Rightarrow \frac{\partial \mathcal{L}}{\partial \lambda_k} =$$
$$c_k + s_k^2 = 0 \quad k = 1, \ldots, m$$

or

$$c_k \leq 0$$

(recovered inequality constraints)

$$\nabla_s \mathcal{L} = 0 \Rightarrow \frac{\partial \mathcal{L}}{\partial s_k} =$$
$$\lambda_k s_k = 0, \quad k = 1, \ldots, m$$

Still need to add $\lambda_k > 0$