Lecture 3: Optimality conditions, constraint qualifications, 2nd order optimality conditions

Sondre Pedersen

January 16, 2025

Some basics:

The gradient (or first derivative) of a function $f(x)$ of several variables is defined as

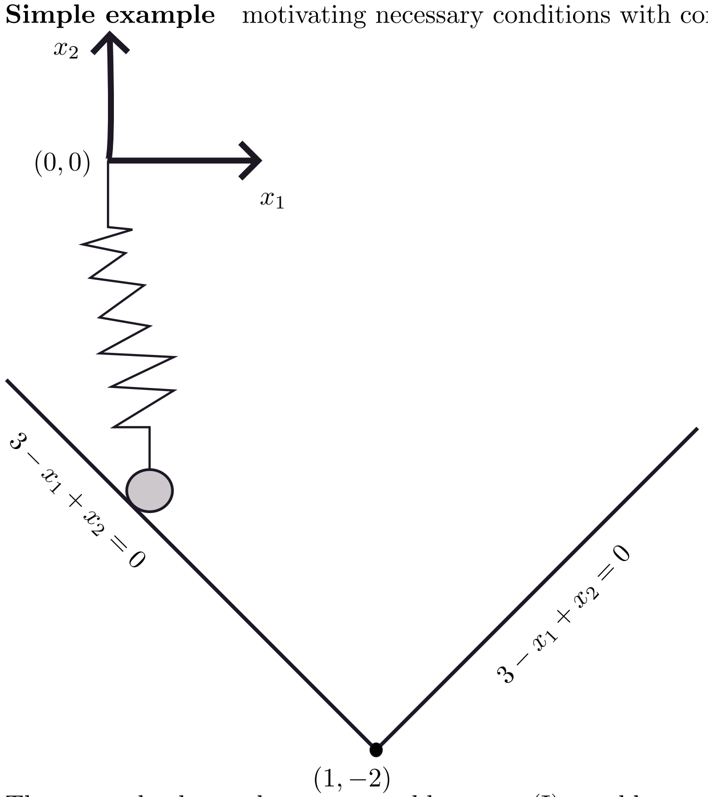
$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}^\top .$$

The matrix of the second partial derivatives of $f(x)$ is known as the *Hessian*, and is defined as

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} .$$

We will frequently use $\nabla^2_{xx}\mathcal{L}(x^*, \lambda^*)$, the *Hessian of the Lagrangian*.
For unconstrained optimization $\nabla f(x^*) = 0$ (gradient) is a necessary condition, and $\nabla^2 f(x^*) > 0$ (positive Hessian) is a sufficient condition.

Simple example motivating necessary conditions with constraints: Ball and spring.

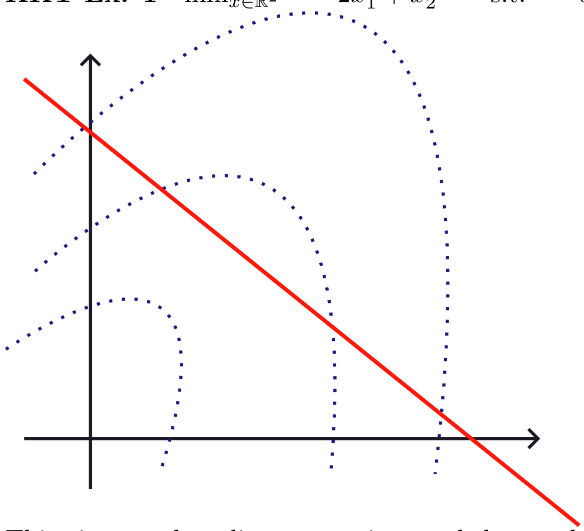


To find position at rest, minimize potential energy.

$$\begin{aligned} \min_{x \in \mathbb{R}^2} \quad & x_1^2 + x_2^2 + m x_2 \\ \text{subject to} \quad & c_1(x) = 1 + x_1 + x_2 \geq 0 \\ & c_2(x) 3 - x_1 + x_2 \geq 0 \end{aligned}$$

The example shows that most problems we (I) would normally solve with forces and mechanics can be formulated as an optimization problem. This is why optimization is a versatile tool.
We get a constrained minimum at "equilibrium of forces": $\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*) + \lambda_2^* \nabla c_2(x^*),\ \lambda_1^*, \lambda_2^* \geq 0$.

KKT Ex. 1 $\min_{x\in\mathbb{R}^2} \quad 2x_1^2 + x_2^2 \quad \text{s.t.} \quad c_1(x) = x_1 + x_2 - 1 = 0.$



$$
\mathcal{L}(x,\lambda) = 2x_2^2 + x_2^2 - \lambda_1(x_1 - x_2 - 1)
$$

KKT:

$$
\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0 \\
c_1(x^*) &= 0 \\
\frac{\partial \mathcal{L}}{\partial x_1} &= 4x_1^* - \lambda_1^* = 0 \\
\frac{\partial \mathcal{L}}{\partial x_2} &= 2x_2^* - \lambda_1^* = 0 \\
x_1^* + x_2^* - 1 &= 0
\end{aligned}
$$

This gives us three linear equations and three unknown. So in this case we can directly find the points that satisfy the KKT conditions. The result is

$$
x_1^* = \frac{1}{3}, \qquad x_2^* = \frac{2}{3}, \qquad \lambda_1^* = \frac{4}{3}.
$$

Solvability of KKT conditions - The above procedure is only possible for simple problems. The main challenge are the complementary conditions - that is, deciding which inequality constraints are active or not. Problems with only equality constraints are a little easier.

What is the point then? Algorithms for LP and QP are constructed by searching for points that fulfill the KKT conditions. LPs and some QPs are convex, making KKT necessary and sufficient. For nonlinear programming, KKT is used to check if a point is at least a candidate solution. This is because they are only *necessary* but not sufficient.

Multipliers also known as "Shadow proces'. When we relax a constraint, the feasible set becomes larger. In some cases, a new minimum becomes available too. The shadow price tells us how much we would gain by expanding the constraint.

$$
f(x_{\mathcal{E}}^*) \approx f(x^*) - \lambda_{\mathcal{E}}.
$$

Multipliers are shadow prices

Consider $\min_x f(x)$ s.t. $c(x) \geq 0$.

KKT stationary: $\nabla f(x^*) = \lambda^* \nabla_c(x^*)$. Assume constraint is active, $c(x^*) = 0$.

Relax constraint: $\min_x f(x)$ s.t. $c(x) \geq -\epsilon$, $\epsilon > 0$ This gives a new solution $x^*(\epsilon)$, with $c(x^*(\epsilon)) = -\epsilon$

$$
\begin{aligned}
f(x^*(\epsilon)) - f(x^*) &\approx \nabla f(x^*)^\top (x^*(\epsilon) - x^*) \\
&= \lambda^* \nabla c(x^*)^\top (x^*(\epsilon) - x^*) \\
&\approx \lambda^* (c(x^*(\epsilon)) - c(x^*)) \\
&= -\lambda^* \epsilon
\end{aligned}
$$

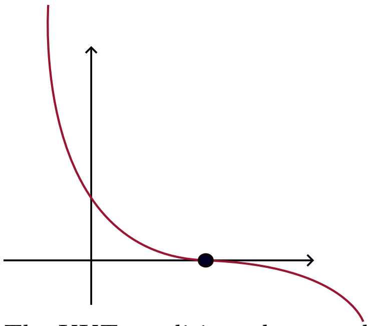
Geometric description of feasible directions Recall - A possible solution is a point where there are no directions that are both feasible and descent directions (directions should only be interpreted in a geometric sense).

The tangent cone to a set Ω at a point $x \in \Omega$, denoted by $T_\Omega(x)$, consists of the limits of all (secant) rays which originate at x and pass through a sequence of points $p_i \in \Omega - x$ which converges to x.
This geometric interpretation must be equivalent to the algebraic method used by KKT. To ensure this, the constraints must satisfy some qualifications. These rule out special cases where the optimal solution does not fulfill the KKT conditions.

A constraint qualification is an assumption that ensures similarity of the constraint set Ω and its linearized approximation, in a neighborhood of a point x^*. Informally: make sure that $\mathcal{F}(x^*)$ and $T_\Omega(x^*)$ are the same.

The most used constraint qualification is LICQ: *Given the point x and the active set $\mathcal{A}(x)$, we say that the linear Independence constraint qualification (LICQ) holds if the set of active constraint gradients $\nabla c_i(x),\ i \in \mathcal{A}(x)$ is linearly independent.*

LICQ Ex. $\min_{x\in\mathbb{R}^2} -x_1$ s.t. $\begin{cases} c_1(x) = (1-x_1)^3 - x_2 \geq 0 \\ c_2(x) = x_1 \geq 0 \\ c_3(x) = x_2 \geq 0 \end{cases}$



$x^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\mathcal{A}(x^*) = 1, 3.$

$\nabla c_1(x^*) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$, $\nabla c_3(x^*) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

The gradients are not independent, so LICQ does not hold.

The KKT conditions does not hold for the optimal point.

Slater's condition is not mentioned in the book, but is still important. *For a convex constrained optimization problem where Slater's condition is fulfulled, the KKT conditions are necessary and sufficient.*

Slater's condition is a constraint qualification, similar to LICQ. Inequality constraints are strictly feasible. LICQ implies Slater's condition, and Slater's condition implies that strong duality hold. Slater's condition is fulfilled for linear constraints.

2^{nd} **Order Conditions: Critical Cone** - After having found a point x^* that fulfills KKT conditions. Now, all all feasible directions $w \in \mathcal{F}(x^*)$ for first order approximations will either (i) lead to an increase in the objective function, or (ii) not change the value.

In case (ii), how do we know if the objective function increases or decreases? The second order conditions answers this by looking at the curvature.

The undecided directions are given by the *critical cone*:

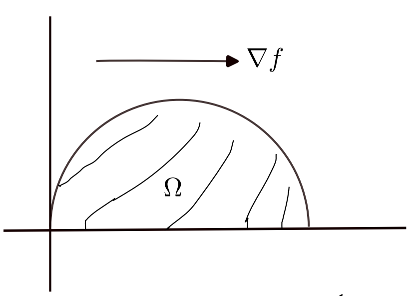
$$w \in \mathcal{C}(x^*, \lambda^*) \implies \begin{cases} \nabla c_i(x^*)^\top w = 0 & \forall i \in \mathcal{E} \\ \nabla c_i(x^*)^\top w = 0 & \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \text{ with } \lambda_i^* > 0 \\ \nabla c_i(x^*)^\top w \geq 0 & \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \end{cases}.$$

To see why this is the critical cone:

- Note that $w \in \mathcal{C}(x^*, \lambda^*)^\top w = 0 \qquad \forall i \in \mathcal{E} \cap \mathcal{I}$
- KKT stationary condition: $\nabla f(x^*) = \sum_{i\in\mathcal{E}\cap\mathcal{I}} \lambda_i^* \nabla c(x^*)$
- Therefore: $w \in \mathcal{C}(c^*, \lambda^*) \implies w^\top \nabla f(x^*) = \sum_{i\in\mathcal{E}\cap\mathcal{I}} \lambda_i^* w^\top c_i(x^*) = 0$

So we need to investigate the critical cone, a subset of the linearized set of feasible directions.

Critical cone Ex. $\min_{x\in\mathbb{R}^2} x_1$ s.t. $\begin{cases} c_1(x) = x_2 >= 0 \\ c_2(x) = -(x_1-1)^2 - x_2^2 + 1 \geq 0 \end{cases}$



$x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\mathcal{A}(x^*) = \{1, 2\}$

KKT fulfilled with $\lambda_1^* = 0,\ \lambda_2^* = 0.5$

$\nabla c_1(x^*) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\nabla c_2(x^*) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

$\mathcal{F}(x^*) = \{d | d \geq 0\}$ (any positive direction is legal)

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \in \mathcal{C}(x^*, \lambda^*) \implies \begin{cases} \nabla c_2(x^*)^\top w = 0 \implies w_1 = 0 \\ \nabla c_1(x^*)^\top w \geq 0 \implies w_2 \geq 0 \end{cases}$$

$\mathcal{C}(c^*, \lambda^*) = \{w = \begin{bmatrix} 0 \\ w_2 \end{bmatrix},\ w_2 \geq 0\}$ is the critical cone. It is only one direction, which we have to check for 2^{nd} order condition.

Theorem 1. *(Second-order sufficient conditions) Suppose that for some feasible point $x^* \in \mathbb{R}^n$ there is a Lagrange multiplier vector λ^* such that the KKT conditions are satisfied. Suppose also that*

$$w^\top \nabla^2_{xx}\mathcal{L}(x^*, \lambda^*)w > 0 \qquad \forall w \in \mathcal{C}(c^*, \lambda^*),\ w \neq 0.$$

Then x^* is a strict local solution.

Continuing the previous example:

$$\begin{aligned}
\mathcal{L}(x, \lambda) &= x_1 - \lambda_1 x_2 - \lambda_2(-(x_1 - 1)^2 - x_2^2 + 1) \\
\nabla_x \mathcal{L}(x, \lambda) &= \begin{bmatrix} 1 + 2\lambda_2(x_1 - 1) \\ -\lambda_1 + 2\lambda_2 x_2 \end{bmatrix} \\
\nabla^2_{xx}\mathcal{L}(x, \lambda) &= \begin{bmatrix} 2\lambda_2 & 0 \\ 0 & 2\lambda_2 \end{bmatrix} \\
\nabla^2_{xx}\mathcal{L}(x^*, \lambda^*) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\end{aligned}$$

This is a positive definite matrix. We need to check it against the direction in the critical cone.

$$w \in \mathcal{C}(x^*, \lambda^*) : w^\top \nabla^2_{xx}\mathcal{L}(x^*, \lambda^*)w = \begin{bmatrix} 0 \\ w_2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & w_2 \end{bmatrix} > w_2^2 > 0.$$

It would have been enough to just look at the Hessian, because no matter what you multiply it with (on both sides) it would be positive.