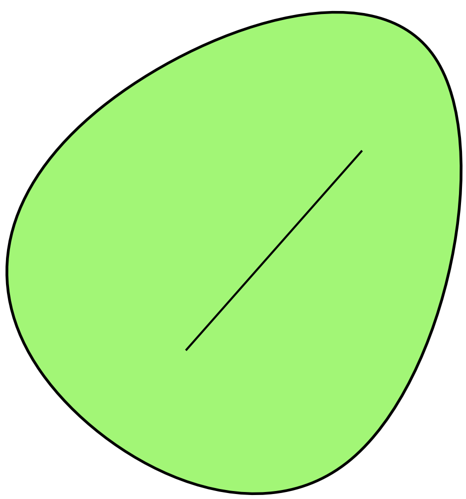
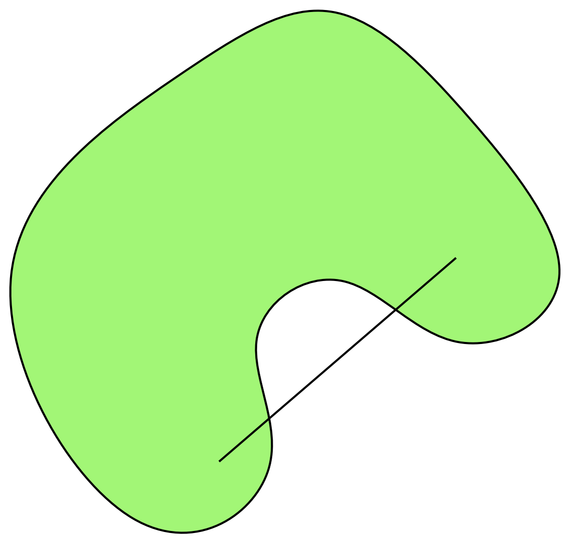
Lecture 2: Optimality Conditions for Constrained Optimization

January 10, 2025

Lecture 1 introduced the concept of conditions and how they can be used to verify candidate solutions to a problem. It also mentioned conditions for unconstrained optimization. In this lecture we look at sufficient and necessary conditions for constrained optimization.
From last time

- Problem (P): $\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \qquad s.t. \qquad \begin{cases} c_i(\mathbf{x}) = 0,\ i \in \mathcal{E} \\ c_i(\mathbf{x}) \geq 0,\ i \in \mathcal{I} \end{cases}$
- Feasible set: $\Omega x \in \mathbb{R}^n | c_i(\mathbf{x}) = 0,\ i \in \mathcal{E}; c_i(\mathbf{c}) \geq 0,\ i \in \mathcal{I}$
- A vector $\mathbf{x}^*$ is a *global solution* to (P) if $\mathbf{x}^* \in \Omega$ and $f(\mathbf{x}) \geq f(\mathbf{x}^*)$
- A vector $\mathbf{x}^*$ is a *local solution* to (P) if $\mathbf{x}^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of $\mathbf{x}^*$ such that $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for $x \in \mathcal{N} \cap \Omega$
- A vector $\mathbf{x}^*$ is a *strict local solution* to (P) if $\mathbf{x}^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of $\mathbf{x}^*$ such that $f(\mathbf{x}) > f(\mathbf{x}^*)$ for $\mathbf{x} \in \mathcal{N} \cap \Omega$ with $\mathbf{x} \neq \mathbf{x}^*$.





The figure shows a convex set and a non-convex set. Any line between points in the first set will not leave the set. But some points will cause the line to leave the set in the second drawing.

Proof: Convex problems - Any local solution is global

Proof by contradiction: Assume that we have

- Convex optimization problem.
- $\mathbf{x}^*$ is a local solution, but not a global solution.

This means that there exist $\mathbf{x}'in \in \Omega,\ \mathbf{x}' \neq \mathbf{x}^*$, s.t. $f(x') < f(\mathbf{x}^*)$

Define $z = \alpha\mathbf{x}^* + (1-\alpha)\mathbf{x}',\ \alpha \in [0,1)$

$$\begin{aligned}
f(z) &= f(\alpha\mathbf{x}^*) + (1-\alpha)\mathbf{x}' \\
&\leq \alpha f(\mathbf{x}^*) + (1-\alpha)f(\mathbf{x}') \\
&= f(\mathbf{x}^*) - (1-\alpha)f(\mathbf{x}^*) + (1-\alpha)f(\mathbf{x}') \\
&= f(\mathbf{x}^*) + (1-\alpha)(f(\mathbf{x}') - f(\mathbf{x}^*)) \\
&\leq f(\mathbf{x}^*)
\end{aligned}$$

This implies that $\mathbf{x}^*$ cannot be a local solution, which is a contradiction.

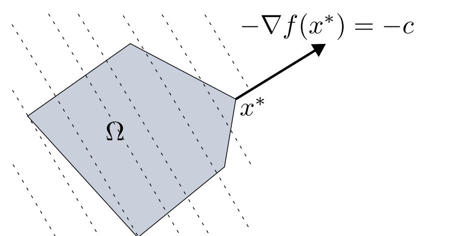
Contours and directions - Given $f(\mathbf{x}) = x_1^2 + x_2^2$ the gradient is $\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$. For any point, it gives a vector in the direction of the steepest increase. The above above gradient always points away from the origin. This means that the origin is the lowest point.

Given a direction d: If $\nabla f(\mathbf{x})^\top d < 0$, then d is a descent direction in $\mathbf{x}$. We can use this dot product to check a direction.

Types of Constrained Optimization Problems :

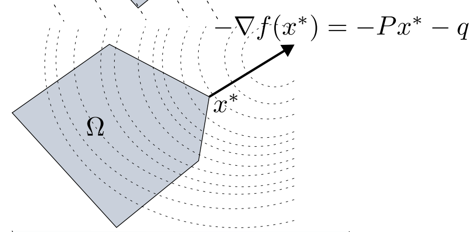
- Linear programming
  - Convex problem
  - Feasible set polyhedron

$$\begin{aligned} \min \quad & c^\top x \\ \text{subject to} \quad & Ax \leq b \\ & Cx = d \end{aligned}$$



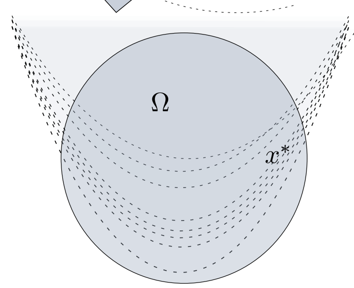
- Quadratic programming
  - Convex problem if $P \geq 0$
  - Feasible set polyhedron

$$\begin{aligned} \min \quad & \frac{1}{2}x^\top Px + q^\top x \\ \text{subject to} \quad & Ax \leq b \\ & Cx = d \end{aligned}$$



- Nonlinear programming
  - Not convex in general

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } \begin{cases} c_i(x) = 0,, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{cases}.$$

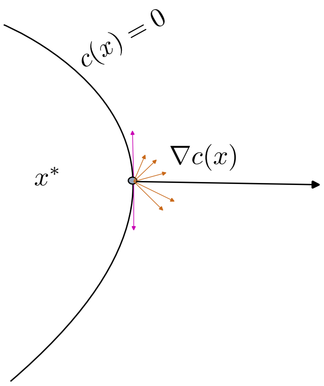


(This took way to long...)

Necessary conditions for Unconstrained Optimization $\min_{x \in \mathbb{R}^n} f(x)$

Theorem 1. *(First-Order Necessary Conditions). If $\mathbf{x}^*$ is a local minimizer and f is continuously differentiable in an open neighborhood of $\mathbf{x}^*$, then $\nabla f(\mathbf{x}^*) = 0$.*

Feasible directions with respect to constraints.



$\nabla c(x')^\top d = 0$: d is a feasible direction for $cf(x) = 0$ in $x^{'}$. Indicated by the purple lines (vertical).

$\nabla c(x^{'})^\top d \geq 0$: d is a feasible direction for $c(x) \geq 0$, in $x^{'}$

Observation : In a local solution, there cannot be feasible downhill descent directions. Conditions that guarantee that a point has no feasible descent direction, are necessary conditions for optimality.

KKT Conditions will do just that.

To simplify notation, introduce the Lagrangian:

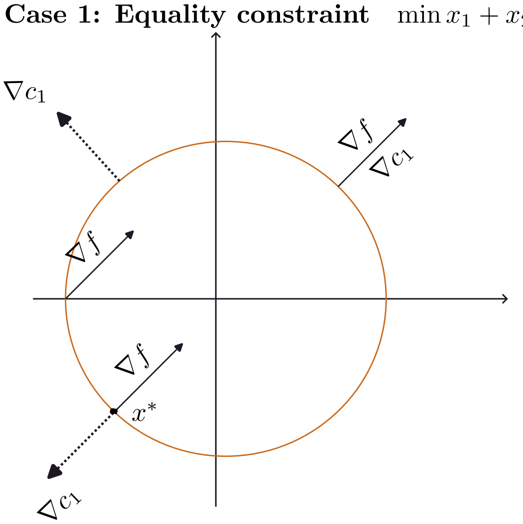
$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i \in \mathcal{E} \cap \mathcal{I}} \lambda_i c_i(x).$$

The necessary conditions for $\mathbf{x}^*$ to be a local solution:

$$\begin{aligned} \nabla_x \mathcal{L}(x^* \lambda^*) &= 0 & \\ c_i(x^*) &= 0 & \forall i \in \mathcal{E} \\ c_i(x^*) &\geq 0 & \forall i \in \mathcal{I} \\ \lambda_i^* &\geq 0 & \forall i \in \mathcal{I} \\ \lambda_i^* c_i(x^*) &= 0 & \forall i \in \mathcal{E} \cap \mathcal{I} \end{aligned}$$

These are hard to prove, so no need to care about that. Instead we look at some examples to understand why they must be like this.

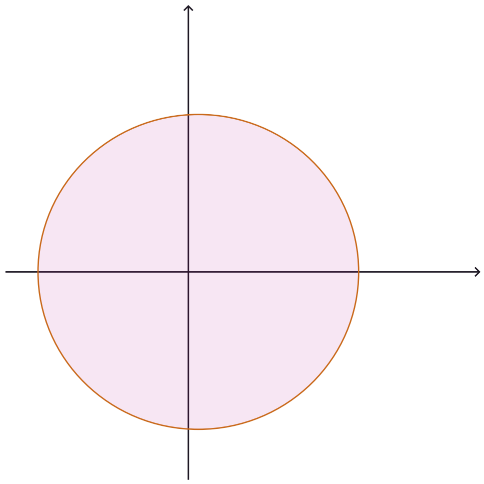
Case 1: Equality constraint $\min x_1 + x_2$ s.t. $x_1^2 + x_2^2 - 2 = 0$.



$\nabla f = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\nabla c_1 = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$. We can see that the gradient of the objective function is always pointing up to the right. The gradient of the constraint is always pointing out of the circle.

We see that at the point x^* there are no feasible descent directions. There are no $\lambda > 0$ such that $\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*)$
It is sometimes possible to use the KKT conditions directly to find candidate solutions directly. For harder problems this is not possible.

Case II: Inequality constraints $\min x_1 + x_2$ s.t. $2 - x_1^2 - x_2^2 \geq 0$.



$\nabla f = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\nabla c_1(x) = \begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix}$. Here there are two different types of feasible points: inside of the circle, or on the border. If the optimal point is inside the circle, the lagrangian multiplier becomes 0, and we are essentially solving and unconstrained optimization problem.

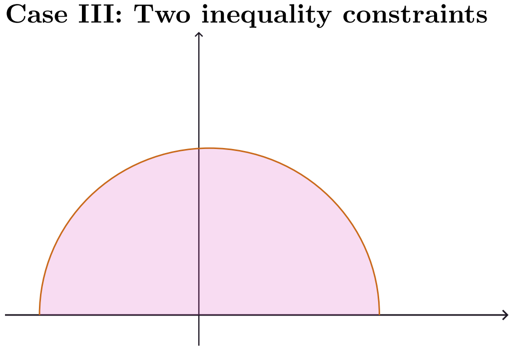
Active set $\mathcal{A}(x)$ at any feasible point x consists of the equality constraints indices form $\mathcal{E}$ together with the indices of the inequalty constraints i for which $c_i(x) = 0$. That is,

$$\mathcal{A}(x) = \mathcal{E} \cap i \in \mathcal{I} | c_i(x) = 0.$$

Set of Linearized Feasible Directions $\mathcal{F}$: given a feasible point x and the active constraint set $\mathcal{A}(x)$, the set of linearized feasible directions $\mathcal{F}(x)$ is

$$\mathcal{F}(x) = \begin{cases} d \quad | & \begin{matrix} d^\top \nabla c_i(x) = 0, & \forall i \in \mathcal{E} \\ d^\top \nabla c_i(x) \geq 0, & \forall i \in \mathcal{A} \cup \mathcal{I} \end{matrix} \end{cases}.$$

Case III: Two inequality constraints $\min x_1 + x_2$ s.t. $2 - x_1^2 - x_2^2 >= \geq 0, \; x_2 \geq 0$



The active set is empty for points inside the region. The active set is 1 for points on the top part of the circle. In the corners of the x-axis, the active set is 1, 2, since both constraints are active.

LICQ - Linear Independence Constraint Qualification.
Given the point x and the active set $\mathcal{A}(x)$, we say that the linear Independence constraint Qualification (LICQQ) holds if the set of active constraint gradients $\nabla c_i(x), \; i \in \mathcal{A}(x)$ is linearly independent.