Lecture 3: Optimality conditions for constrained optimization, cont’d

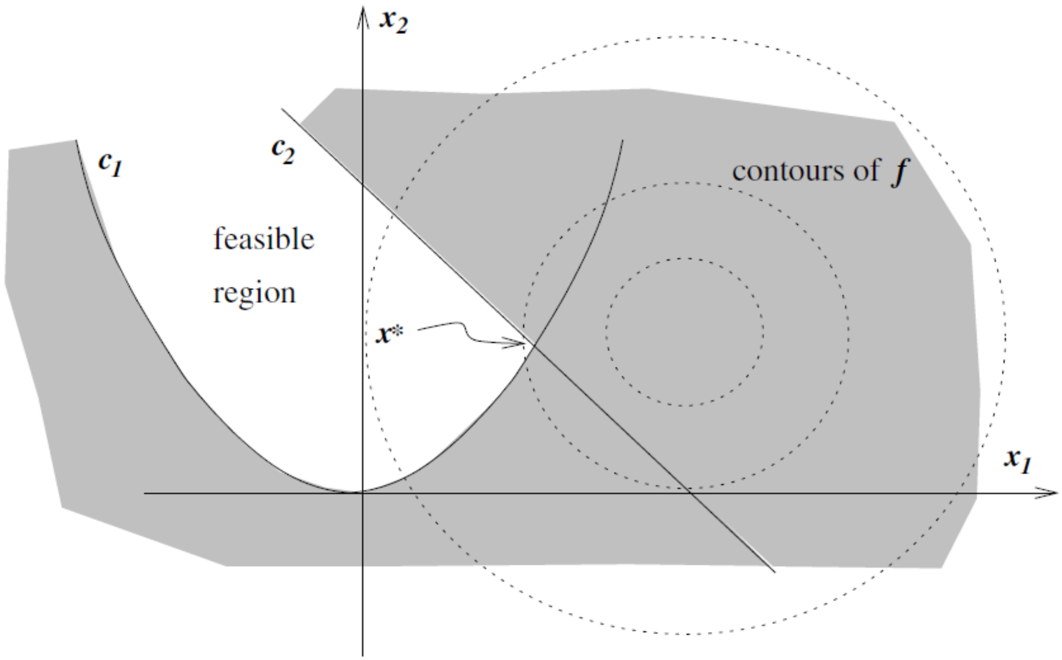
- KKT conditions
- 2^{nd} order conditions

Reference: Chapter 12.3, 12.5 (12.8, 12.9) in N&W

General optimization problem

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

- Example: $$\min\ (x_1 - 2)^2 + (x_2 - 1)^2 \quad \text{subject to} \quad \begin{aligned} x_1^2 - x_2 &\leq 0, \\ x_1 + x_2 &\leq 2. \end{aligned}$$



Gradient and Hessian

- 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 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*

Positive definiteness

A square, symmetric matrix A is *positive definite* if the following equivalent conditions hold:

- There is a positive scalar α such that
$$x^\top A x \geq \alpha x^\top x, \qquad \text{for all } x \in \mathbb{R}^n.$$
- $x^\top A x > 0, \qquad \text{for all } x \neq 0.$
- If all *eigenvalues* $\lambda_i > 0$.

We also write $A > 0$ when A is PD.

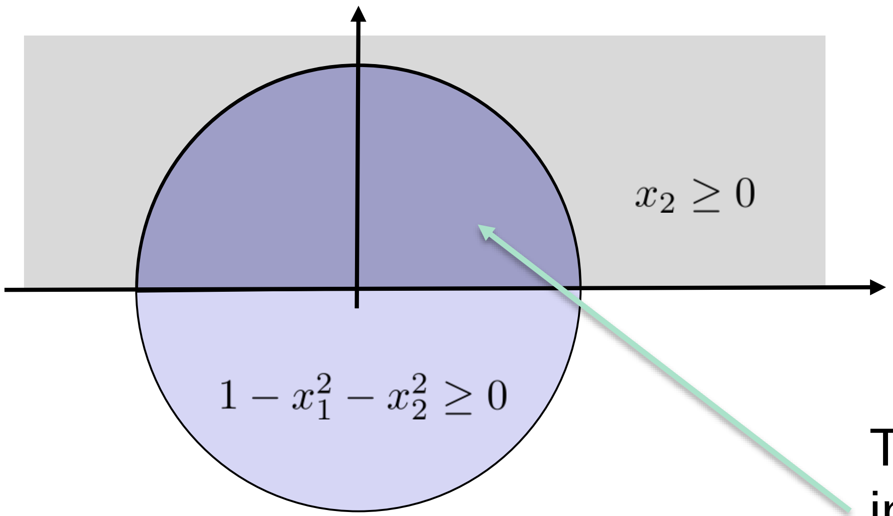
A square matrix A is *positive semidefinite* if

$$x^\top A x \geq 0, \qquad \text{for all } x \in \mathbb{R}^n$$

We also write $A \geq 0$ when A is PSD.

Feasible set

Feasible set: Collection of all points that satisfy all constraints:

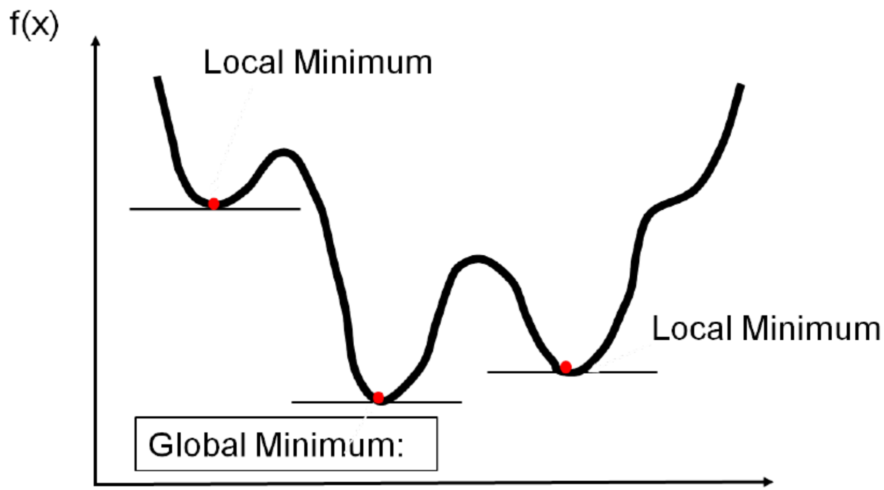


$$c_1(x) = x_2 \geq 0$$
$$c_2(x) = 1 - x_1^2 - x_2^2 \geq 0$$

The feasible set is the intersection of the grey and blue area

$$\text{Feasible set:} \qquad \Omega = \{x \in \mathbb{R}^n \mid c_i(x) = 0,\ i \in \mathcal{E};\ c_i(x) \geq 0,\ i \in \mathcal{I}\}$$

Local and global optima



$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{cases} \qquad \text{(P)}$$

A point x^* is a *global solution* to (P) if $x^* \in \Omega$ and $f(x) \geq f(x^*)$ for $x \in \Omega$.

A point x^* is a *local solution* to (P) if $x^* \in \Omega$ and there is a neighborhood $\mathcal{N}$ of x^* such that $f(x) \geq f(x^*)$ for $x \in \mathcal{N} \cap \Omega$.

Convex optimization problems: local solutions are global.

Unconstrained optimality conditions

$$\min_{x \in \mathbb{R}^n} f(x)$$

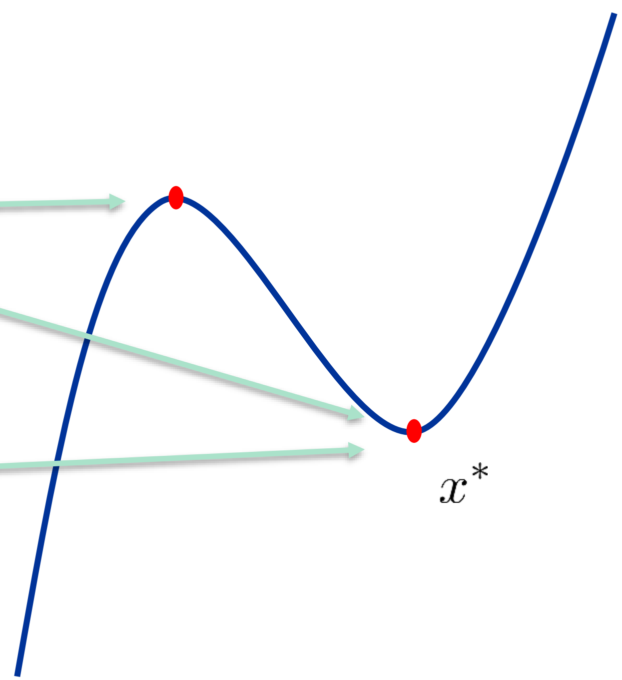
We want to test a point x^* for local optimality:

Necessary condition:

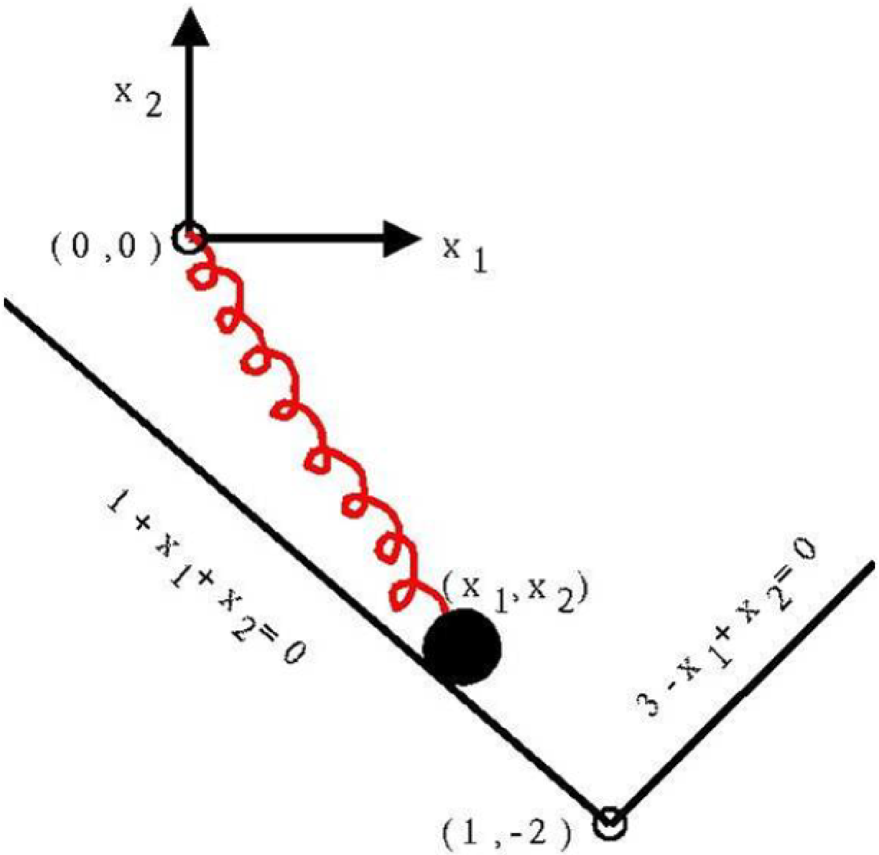
$\nabla f(x^*) = 0$ (stationarity)

Sufficient condition:

x^* stationary and $\nabla^2 f(x^*) > 0$



Simple example: Ball hanging on a string

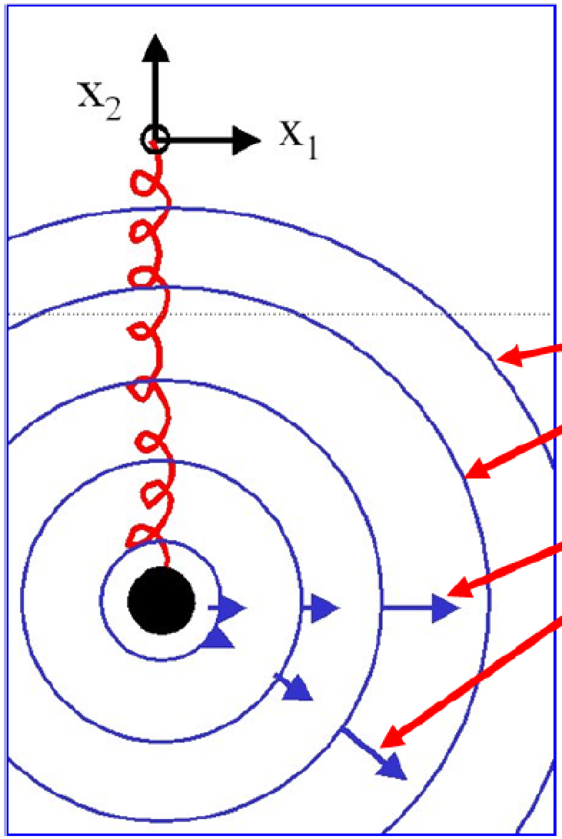


To find position at rest, minimize potential energy!

$$
\begin{aligned}
\text{minimize} \quad & \underbrace{x_1^2 + x_2^2}_{\text{spring}} + \underbrace{m\,x_2}_{\text{gravity}} \\
\text{subject to} \quad & 1 + x_1 + x_2 \geq 0 \\
& 3 - x_1 + x_2 \geq 0
\end{aligned}
$$

Ball on a spring without constraints

$$\boxed{\underset{x\in\mathbb{R}^2}{\text{minimize}} \quad x_1^2 + x_2^2 + m\,x_2}$$



contour lines of *f(x)*

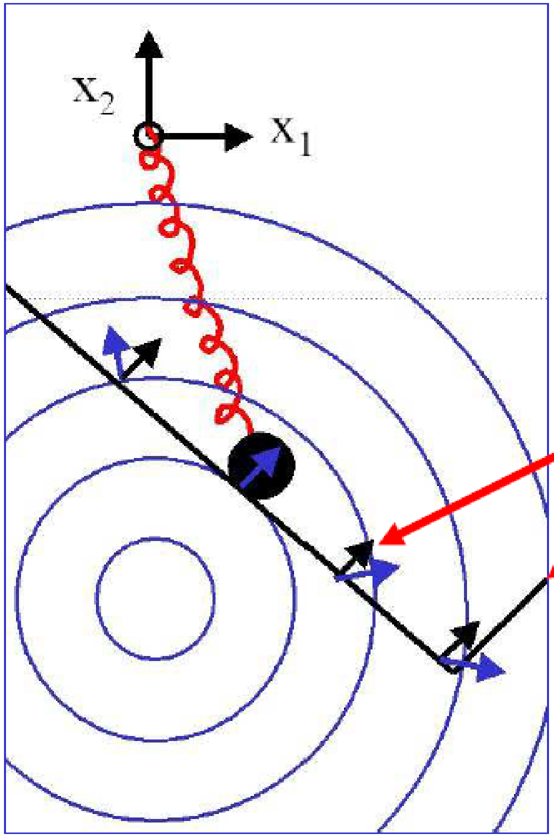
gradient vector

$$\nabla f(x) = (2x_1, 2x_2 + m)$$

unconstrained minimum:

$$0 = \nabla f(x^*) \Leftrightarrow (x_1^*, x_2^*) = (0, -\frac{m}{2})$$

Ball on a string with constraints



$$\min f(x)$$
$$h_1(x) := 1 + x_1 + x_2 \quad \geq \quad 0$$
$$h_2(x) := 3 - x_1 + x_2 \quad \geq \quad 0$$

gradient ∇h_1 of active constraint

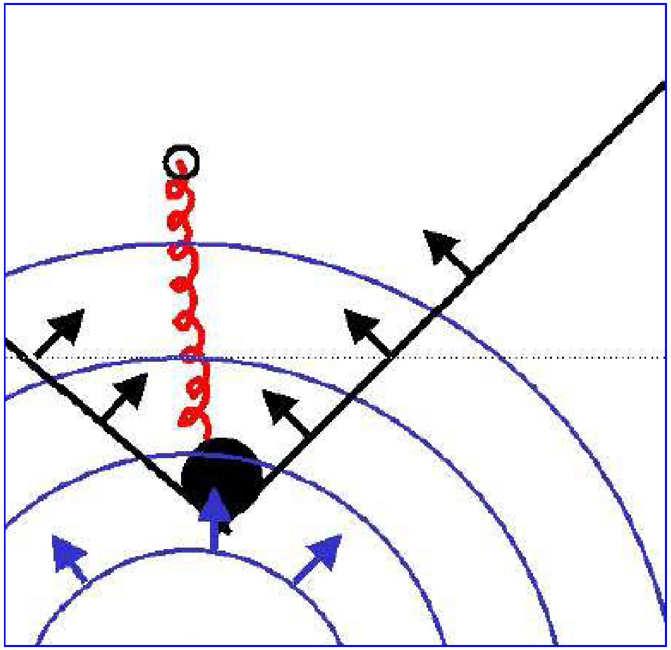
inactive constraint h_2

constrained minimum:

$$\boxed{\nabla f(x^*) = \mu_1 \nabla h_1(x^*)}$$

Lagrange multiplier

Ball on a string with two active constraints



$$\min f(x)$$
$$h_1(x) := 1 + x_1 + x_2 \quad \geq \quad 0$$
$$h_2(x) := 3 - x_1 + x_2 \quad \geq \quad 0$$

„equilibrium of forces“

$$\nabla f(x^*) = \mu_1 \nabla h_1(x^*) + \mu_2 \nabla h_2(x^*) \qquad \mu_1, \mu_2 \geq 0$$

„constraint forces“

The Lagrangian

For constrained functions, introduce modification of objective function:

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

- Multipliers for *equality* constrains may have both signs in a solution
- Multipliers for *inequality* constraints cannot be negative (cf. shadow prices)
- For (inequality) constraints that are *inactive*, multipliers are zero

Active set

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

$$\mathcal{A}(x) = \mathcal{E} \cup \left\{ i \in \mathcal{I} \middle| c_i(x) = 0 \right\}$$

KKT conditions (Theorem 12.1)

KKT-conditions (First-order necessary conditions): If x^* is a local solution and LICQ holds, then there exist λ^* such that

$$
\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, && && \text{(stationarity)} \\
c_i(x^*) &= 0, && \forall i \in \mathcal{E}, && \\
c_i(x^*) &\geq 0, && \forall i \in \mathcal{I}, && \text{(primal feasibility)} \\
\lambda_i^* &\geq 0, && \forall i \in \mathcal{I}, && \text{(dual feasibility)} \\
\lambda_i^* c_i(x^*) &= 0, && \forall i \in \mathcal{E} \cup \mathcal{I}. && \text{(complementarity condition/ complementary slackness)}
\end{aligned}
$$

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

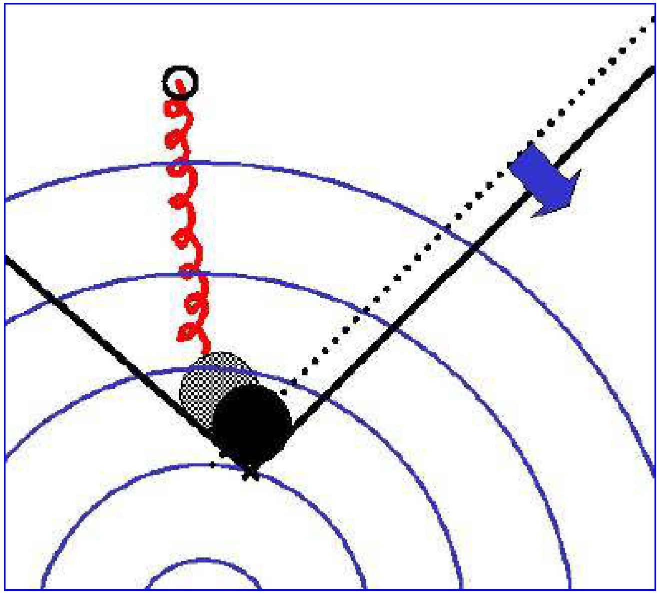
(*strict* complimentarity: Only one of them is zero)

Solvability

- KKT conditions can only be solved for very simple nonlinear programming problems
  - The main complexity is the complementarity conditions – that is, deciding which constraints are active or not

- What is then the use of the KKT conditions?
  - Algorithms for LP and QP are constructed by searching for points that fulfill the KKT conditions
    - LPs and (some) QPs are convex – a local solution is global
  - For nonlinear programming, we use KKT to check whether a certain point is a *candidate* solution
    - Remember, KKT are *necessary* conditions!

Multipliers are shadow prices



old constraint: $h(x) \geq 0$

new constraint: $h(x) + \varepsilon \geq 0$

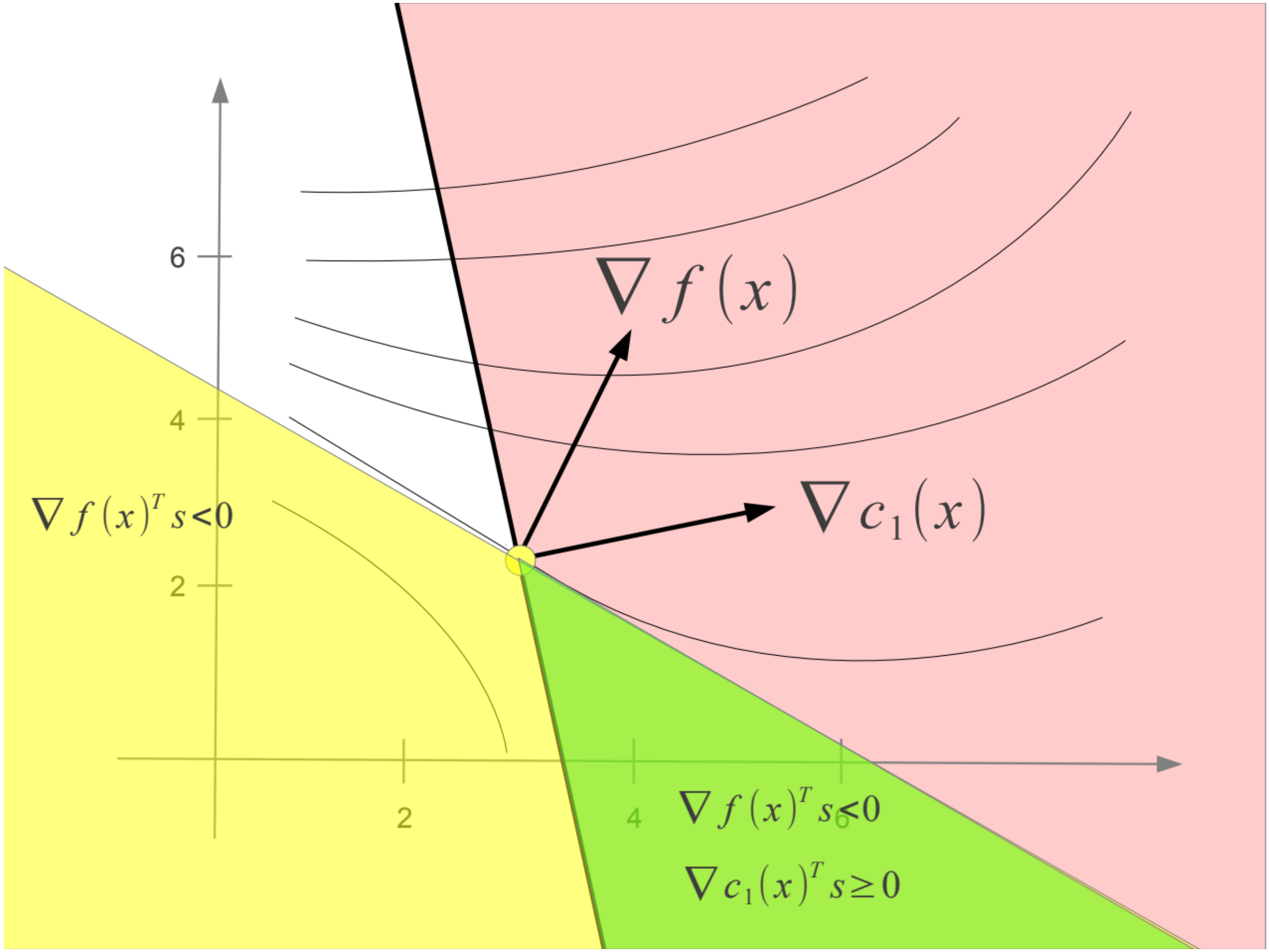
What happens if we relax a constraint?
Feasible set becomes bigger,
so new minimum $f(x_\varepsilon^*)$ becomes smaller.
How much would we gain?

$$\boxed{f(x_\varepsilon^*) \approx f(x^*) - \mu\varepsilon}$$

Multipliers show the hidden cost of constraints.

Constraint qualifications

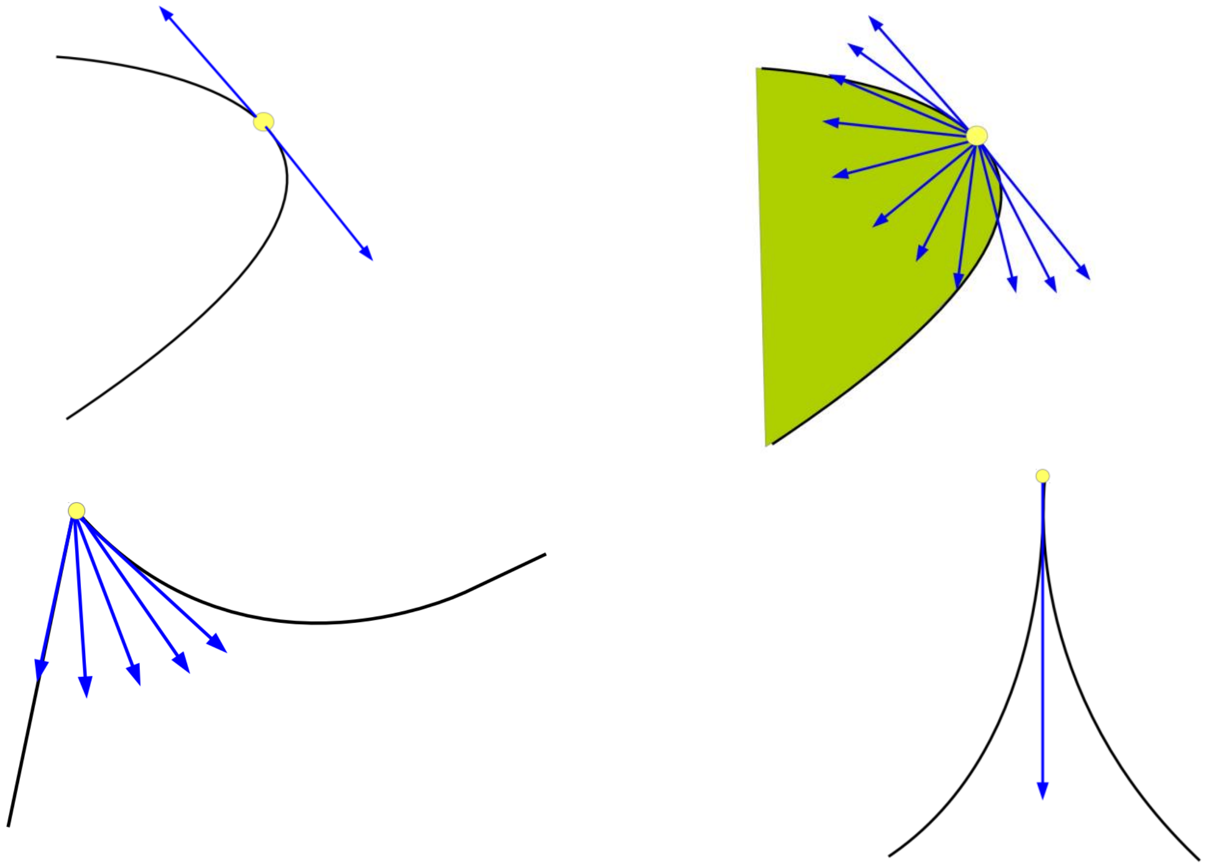
- Recall:



http://montes-de-oca.com/teaching/UD/S14-MATH529-10/Session13/handout.pdf

Tangent cone

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.



Set of (linearized) feasible directions

Given a feasible point x and the active constraint set $\mathcal{A}(x)$, the set of linearized feasible directions $\mathcal{F}(x)$ is

$$\mathcal{F}(x) = \left\{ d \;\middle|\; \begin{array}{ll} d^\top \nabla c_i(x) = 0, & \text{for all } i \in \mathcal{E}, \\ d^\top \nabla c_i(x) \geq 0, & \text{for all } i \in \mathcal{A}(x) \cap \mathcal{I} \end{array} \right\}$$

- Note 1: The definition of $T_\Omega(x)$ depends on the geometry of the feasible set Ω.
- Note 2: The definition of $\mathcal{F}(x)$ depends on the algebraic definition of the constraints

Constraint qualifications

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^*.

- In other words: Constraint qualifications ensure that the linearized feasible set $\mathcal{F}(x)$ and the tangent cone $T_\Omega(x)$ is similar (usually: identical).

- 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.

  - Other (weaker) constraint qualifications exists, that are easier to check than LICQ but implies LICQ (N&W 12.6, not syllabus)

- Note: LICQ implies uniqueness of Lagrange multipliers

Second order conditions, critical cone

- Say there are directions $w \in \mathcal{F}(x^*)$ that does not lead to an increase in the objective function, that is $w^\top \nabla f(x^*) = 0,\ w \neq 0$. How do we decide whether it is a minimum?
- Second-order conditions answer this by looking at the curvature in these directions

- Define the *critical cone*:

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

- These are the “undecided” directions, the directions that are not “frozen” by the active constraint

Second-order conditions

- Second-order necessary conditions (Theorem 12.5):

Suppose that x^* is a local solution and that the LICQ condition is satisfied. Let λ^* be the Lagrange multiplier vector for which the KKT conditions are satisfied. Then

$$w^\top \nabla^2_{xx}\mathcal{L}(x^*, \lambda^*)w \geq 0, \qquad \text{for all } w \in \mathcal{C}(x^*, \lambda^*)$$

- Second-order sufficient conditions (Theorem 12.6):

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

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

Then x^* is a strict local solution.

Example:

$$\min_{x \in \mathbb{R}^2} \quad f(x) = \frac{1}{2}\left((x_1 - 1)^2 + x_2^2\right)$$
$$\text{s.t.} \quad c_1(x) = -x_1 + \beta x_2^2 = 0$$

$$\mathcal{E} = \{1\}, \qquad \mathcal{I} = \emptyset.$$