Optimization — Introduction

Optimization deals with how to do things in the best possible manner:

- Design of multi-criteria controllers
- Clustering in fuzzy modeling
- Trajectory planning of robots
- Scheduling in process industry
- Estimation of system parameters
- Simulation of continuous time systems on digital computers
- Design of predictive controllers with input-saturation

Related courses:

- SC42025: Filtering & identification
- SC42125: Model predictive control
- SC42101: Networked and distributed control systems
- EE4530: Applied convex optimization
- WI4227-14: Discrete optimization
- WI4410: Advanced discrete optimization

Overview

Three subproblems:

Formulation (other courses):
: Translation of engineering demands and requirements into a mathematically well-defined optimization problem

Optimization procedure:
: Choice of right algorithm
: Various optimization techniques
: Various computer platforms

Initialization & approximation (other courses):
: Choice of initial values for parameters
: Approximation of problem by more simple one

Teaching goals

- Insight into basic operation of optimization algorithms
- Optimization problem $\rightarrow$ most efficient and best suited optimization algorithm
- Reduce complexity of optimization problem using simplifications and/or reformulations

Contents

Optimization Techniques

Mathematical framework

$$\min_x f(x)$$
$$\text{s.t. } h(x) = 0$$
$$g(x) \leqslant 0$$

- f : objective function
- x : parameter vector
- $h(x) = 0$: equality constraints
- $g(x) \leqslant 0$: inequality constraints

$f(x)$ is a scalar
$g(x)$ and $h(x)$ may be vectors

- **Unconstrained optimization:**

$$f(x^*) = \min_x f(x)$$

where

$$x^* = \arg\min_x f(x)$$

- **Constrained optimization:**

$$f(x^*) = \min_x f(x)$$

$$h(x^*) = 0$$

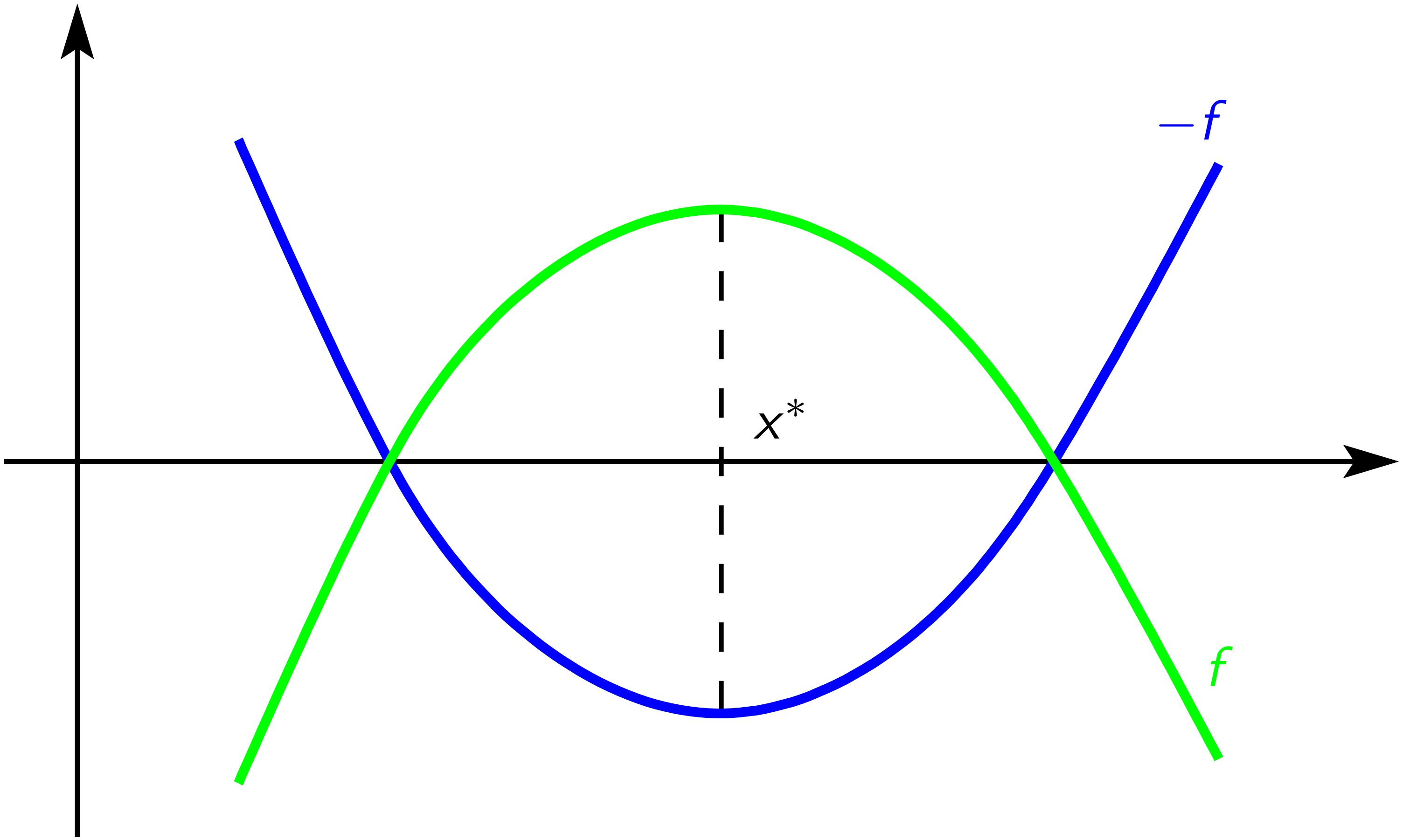
$$g(x^*) \leqslant 0$$

where

$$x^* = \arg\left(\min_x f(x) \text{ s.t. } h(x) = 0,\ g(x) \leqslant 0\right)$$

Maximization = Minimization

$$\max_x \; f(x) = -\min_x \; \big(-f(x)\big)$$



Classes of optimization problems

- Linear programming

$$\min_x c^T x \ , \quad Ax = b \ , \quad x \geqslant 0$$

$$\min_x c^T x \ , \quad Ax \leqslant b \ , \quad x \geqslant 0$$

- Quadratic programming

$$\min_x \tfrac{1}{2} x^T H x + c^T x \ , \quad Ax = b \ , \quad x \geqslant 0$$

$$\min_x \tfrac{1}{2} x^T H x + c^T x \ , \quad Ax \leqslant b \ , \quad x \geqslant 0$$

- Convex optimization

$$\min_x f(x) \ , \quad g(x) \leqslant 0 \qquad \text{where } f \text{ and } g \text{ are convex}$$

- Nonlinear optimization

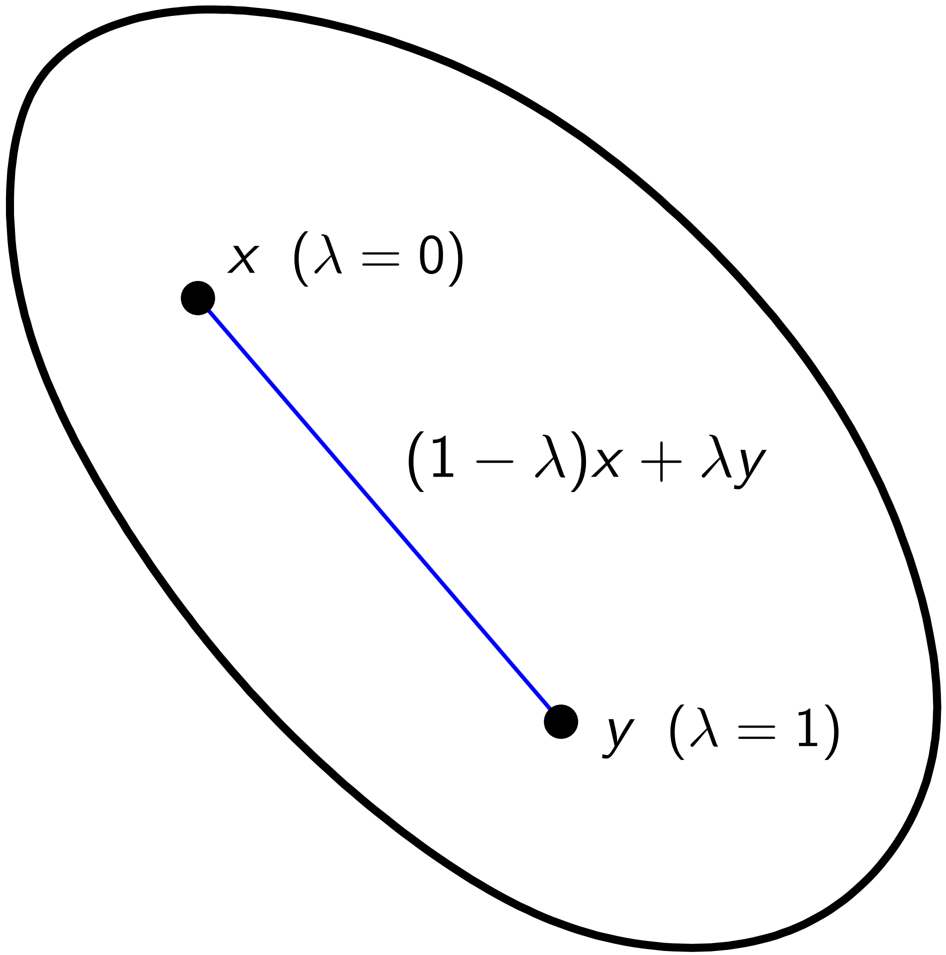
$$\min_x f(x) \ , \quad h(x) = 0 \ , \quad g(x) \leqslant 0$$

where f, h, and g are non-convex and nonlinear

Convex set

Set $\mathcal{C}$ in $\mathbb{R}^n$ is convex if for all $x, y \in \mathcal{C}$, and for all $\lambda \in [0,1]$:

$$(1-\lambda)x + \lambda y \in \mathcal{C}$$



Unimodal function

A function f is unimodal if

a) The domain $\mathrm{dom}(f)$ is a convex set.

b) $\exists\, x^* \in \mathrm{dom}(f)$ such

$$f(x^*) \leqslant f(x) \ \ \forall x \in \mathrm{dom}(f)$$

c) For all $x_0 \in \mathrm{dom}(f)$

there is a trajectory $x(\lambda) \in \mathrm{dom}(f)$

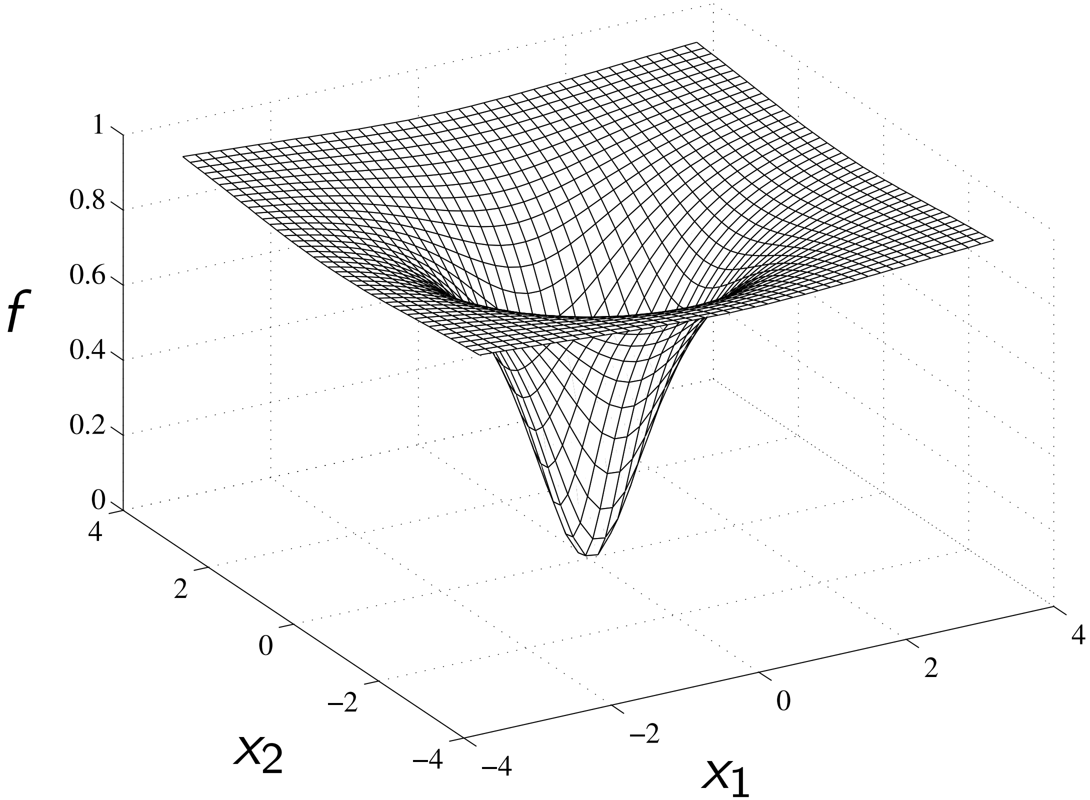
with $x(0) = x_0$ and $x(1) = x^*$

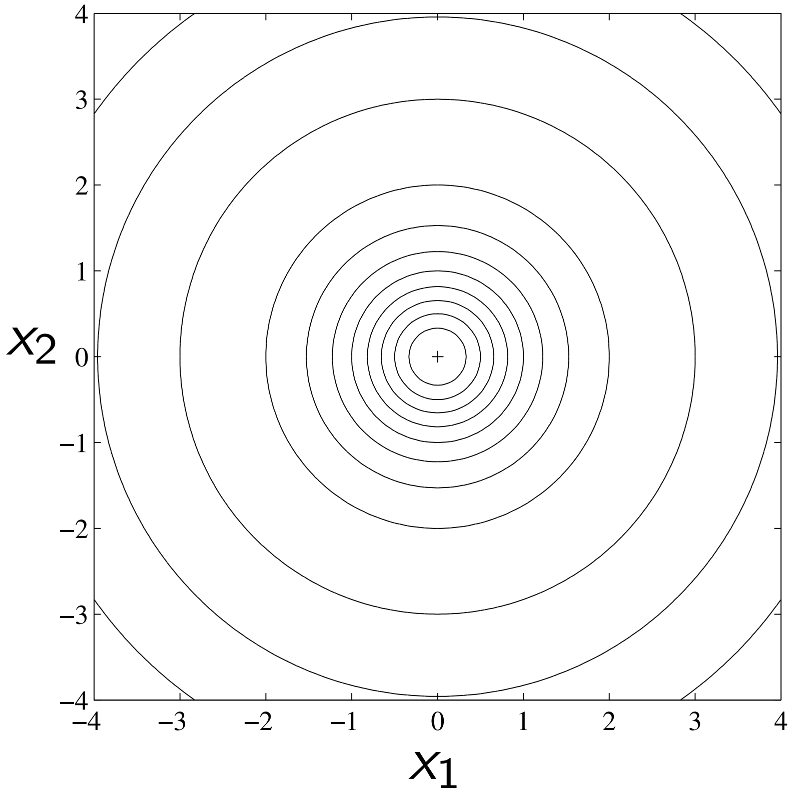
such that

$$f\Big(x(\lambda)\Big) \leqslant f(x_0) \quad \forall \lambda \in [0, 1]$$

Inverted Mexican hat

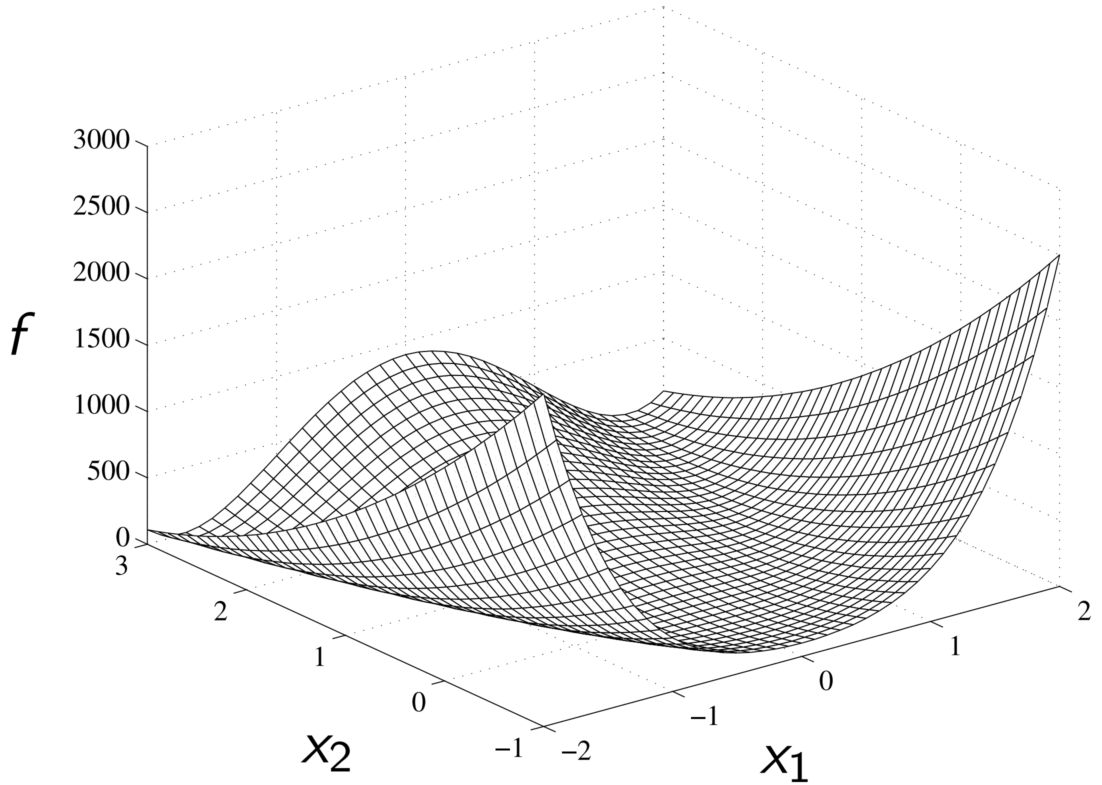
$$f(x) = \frac{x^T x}{1 + x^T x} \qquad x \in \mathbb{R}^2$$

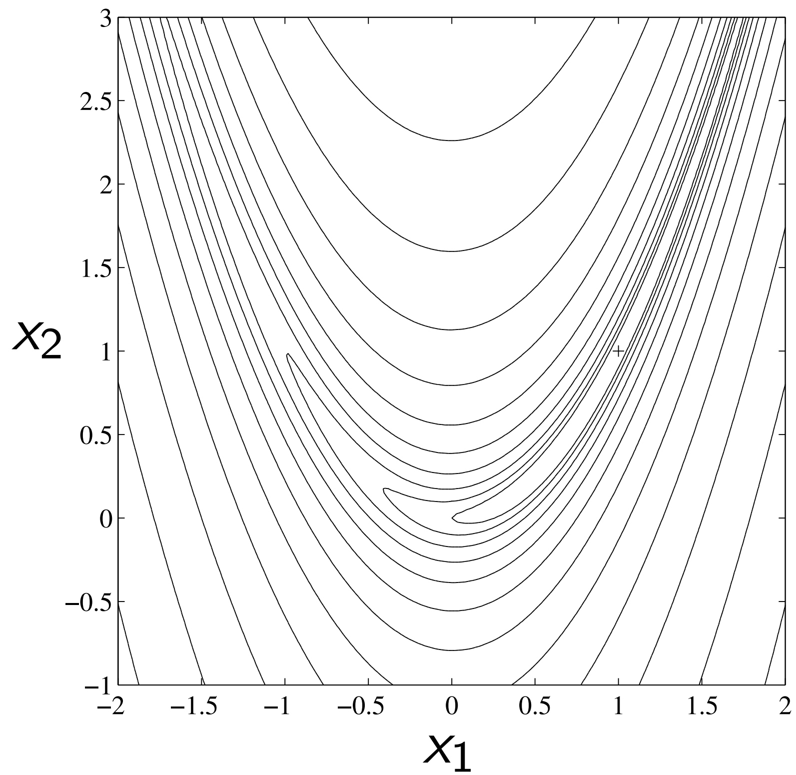




Rosenbrock function

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$





Quasiconvex function

A function f is quasiconvex if

a) Domain $\text{dom}(f)$ is a convex set

b) For all $x, y \in \text{dom}(f)$

and $0 \leqslant \lambda \leqslant 1$

there holds

$$f\Big((1-\lambda)x + \lambda y\Big) \leqslant \max\left(f(x), f(y)\right)$$

Quasiconvex function

Alternative definition:

A function f is quasiconvex if the sublevel set

$$\mathcal{L}(\alpha) = \{\, x \in \mathrm{dom}(f) : \ f(x) \leqslant \alpha \,\}$$

is convex for every real number α

Convex function

A function f is convex if

a) Domain $\mathrm{dom}(f)$ is a convex set.

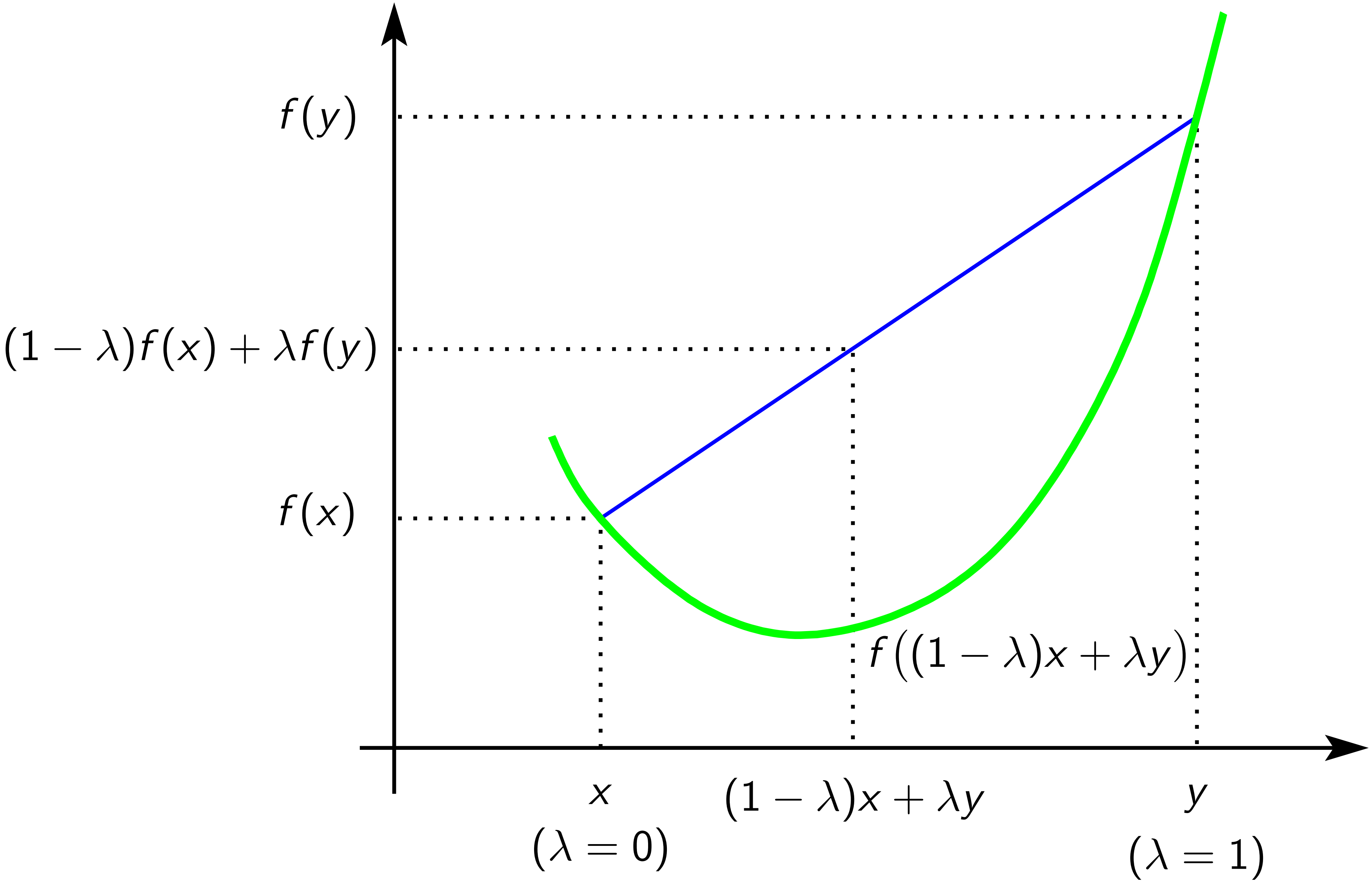
b) For all $x, y \in \mathrm{dom}(f)$

   and $0 \leqslant \lambda \leqslant 1$

   there holds

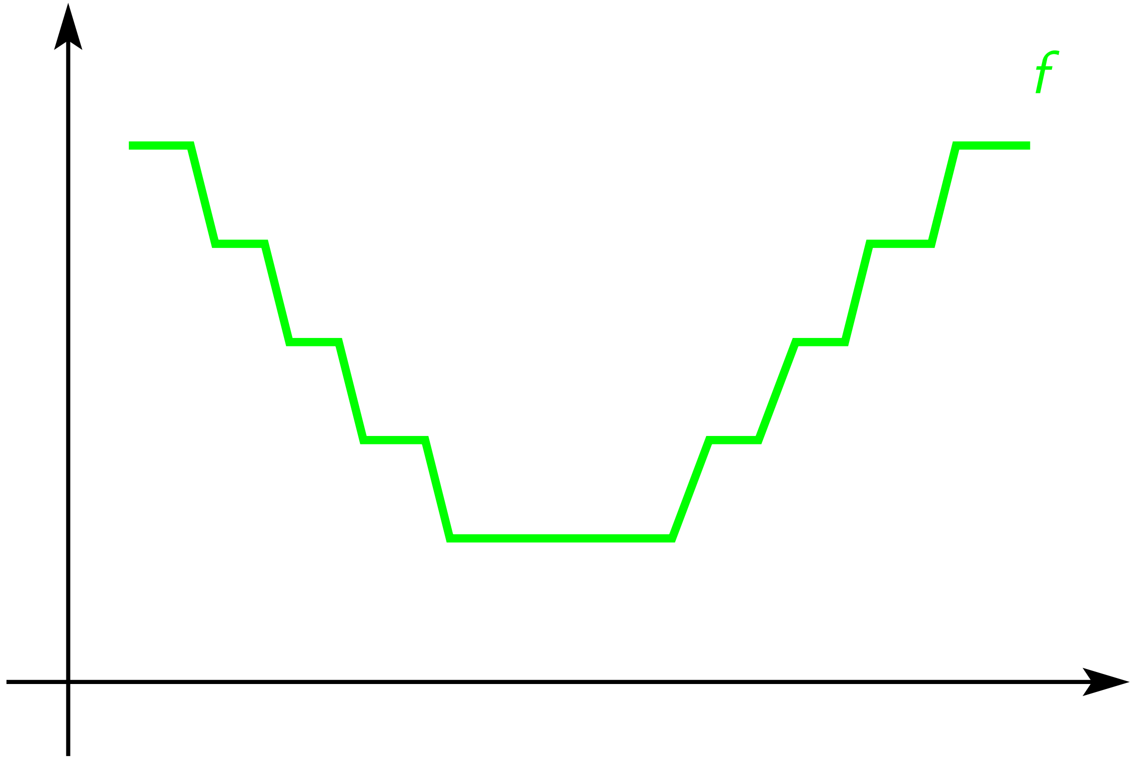
$$f\Big((1-\lambda)x + \lambda y\Big) \leqslant (1-\lambda)f(x) + \lambda f(y)$$

Convex function



Test: Unimodal, quasiconvex, convex

Given: Function f with graph



Question: f is (check all that apply)

- ☐ unimodal
- ☐ quasiconvex
- ☐ convex

Gradient and Hessian

Gradient of f: $$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Hessian of f: $$H(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Jacobian

x: vector
h: vector-valued

$$\nabla h(x) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_2}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_1} \\ \frac{\partial h_1}{\partial x_2} & \frac{\partial h_2}{\partial x_2} & \cdots & \frac{\partial h_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_1}{\partial x_n} & \frac{\partial h_2}{\partial x_n} & \cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix}$$

Graphical interpretation of gradient

- Directional derivative of function f in x_0 in direction of unit vector β:

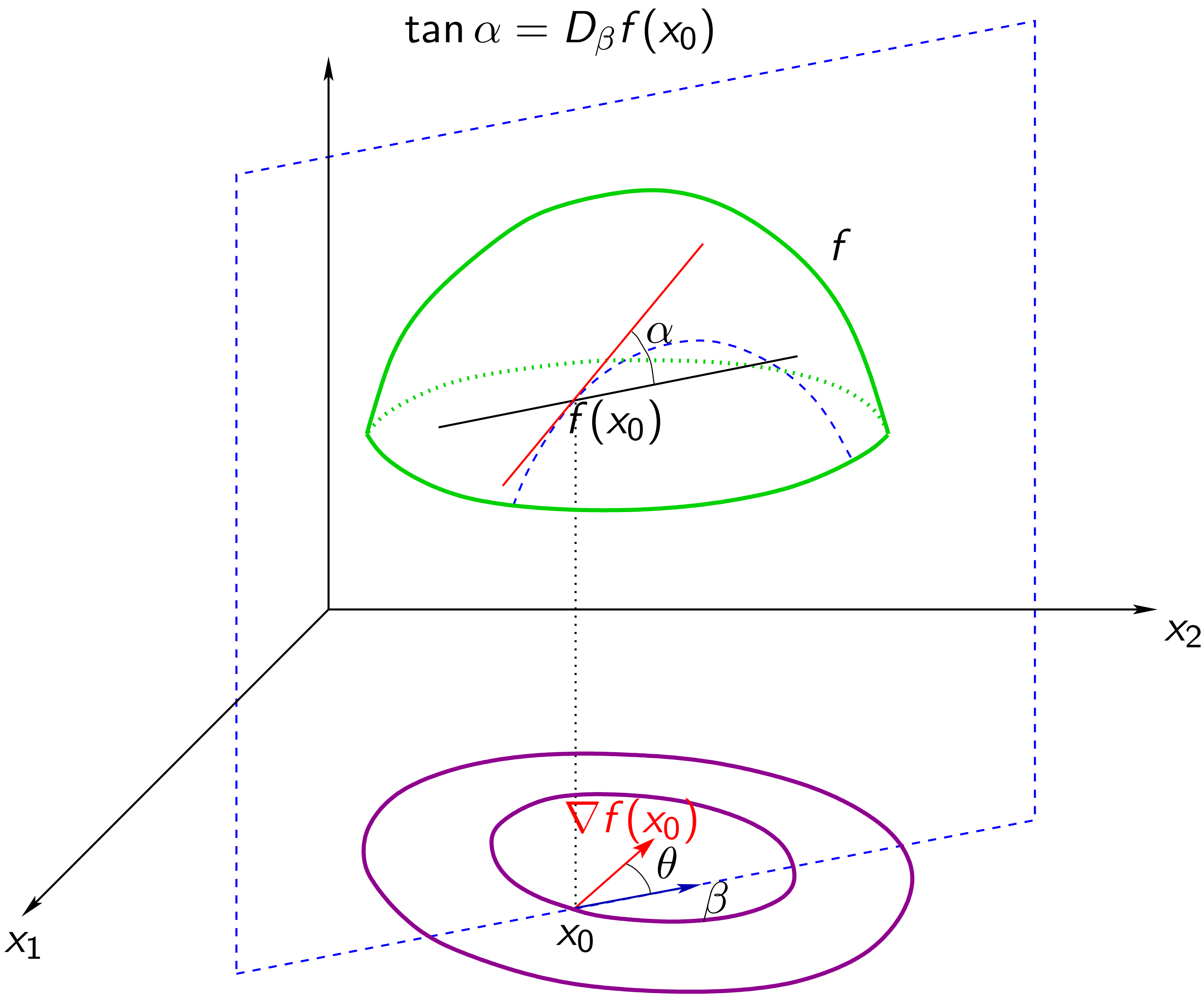
$$D_\beta f(x_0) = \nabla^T f(x_0) \cdot \beta = \|\nabla f(x_0)\|_2 \cos\theta$$

  with θ angle between $\nabla f(x_0)$ and β

- $D_\beta f(x_0)$ is maximal if $\nabla f(x_0)$ and β are parallel
  $\rightarrow$ function values exhibit largest increase in direction of $\nabla f(x_0)$
  $\rightarrow$ function values exhibit largest decrease in direction of $-\nabla f(x_0)$

- $-\nabla f(x_0)$ is called *steepest descent direction*

- $D_\beta f(x_0)$ is equal to 0 (i.e., function values f do not change) if $\nabla f(x_0) \perp \beta$
  $\rightarrow$ $\nabla f(x_0)$ is perpendicular to contour line through x_0


$\tan\alpha = D_\beta f(x_0)$
f
α
$f(x_0)$
x_2
x_1
$\nabla f(x_0)$
θ
β
x_0

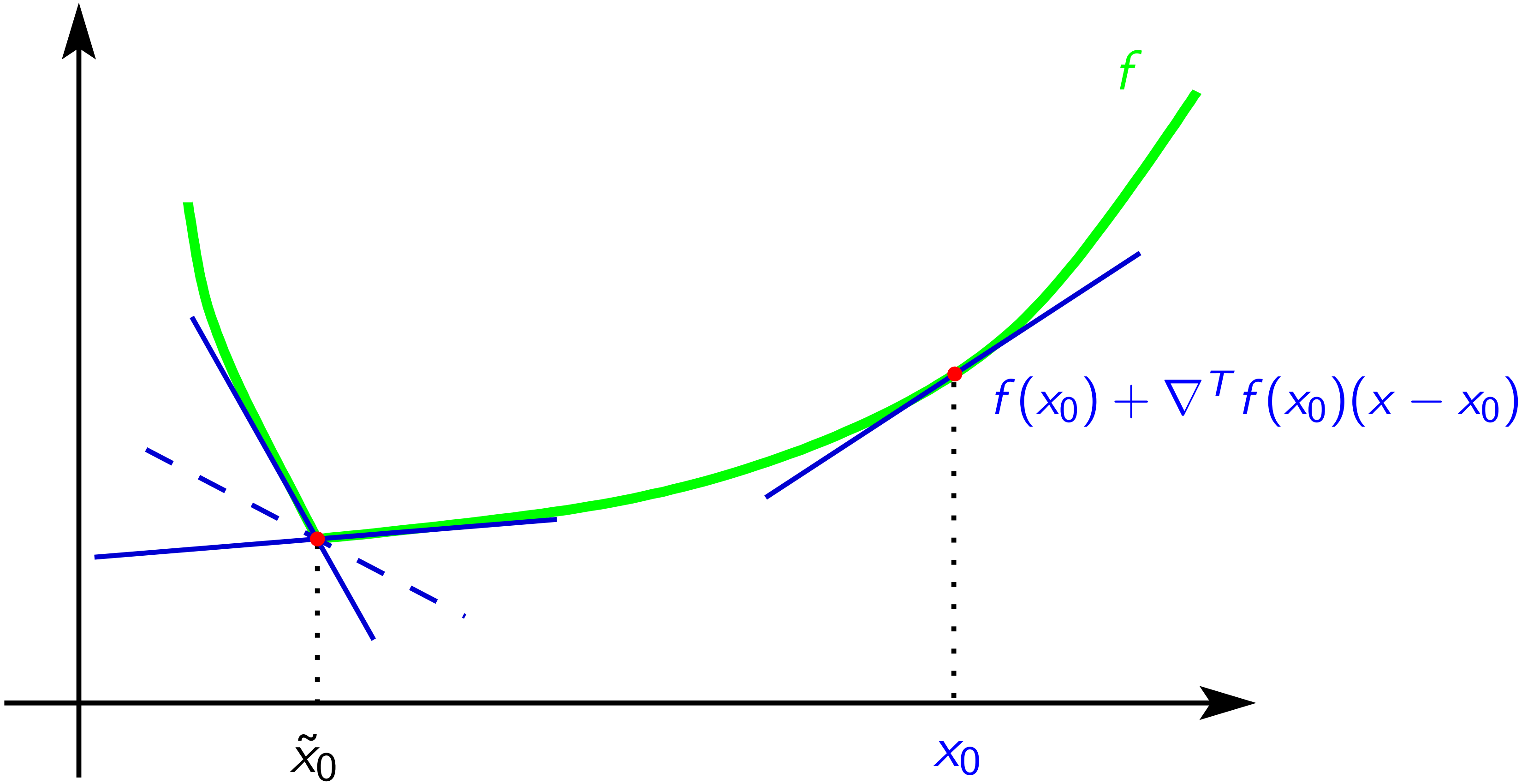
Subgradient

Let f be a *convex* function.

$\nabla f(x_0)$ is a subgradient of f in x_0 if

$$f(x) \geqslant f(x_0) + \nabla^T f(x_0)\,(x - x_0)$$

for all $x \in \mathbb{R}^n$



Positive definite matrices

Let $A \in \mathbb{R}^{n \times n}$ be symmetric

A is positive definite ($A > 0$) if $x^T A x > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$

A is positive semi-definite ($A \geqslant 0$) if $x^T A x \geqslant 0$ for all $x \in \mathbb{R}^n$

Property

- $A > 0$ if all its leading principal minors are positive or if all its eigenvalues are positive
- $A \geqslant 0$ if all its principal minors are nonnegative or if all its eigenvalues are nonnegative

Note:

- principal minor: determinant of submatrix A_{JJ} consisting of rows and columns in J
- leading principal minor: determinant of submatrix A_{JJ} with $J = \{1, 2, \ldots, k\}$, $k \leqslant n$

Classes of optimization problems

- Linear programming

$$\min_x \; c^T x \;,\; Ax = b \;,\; x \geqslant 0$$

$$\min_x \; c^T x \;,\; Ax \leqslant b \;,\; x \geqslant 0$$

- Quadratic programming

$$\min_x \; \tfrac{1}{2}x^T Hx + c^T x \;,\; Ax = b \;,\; x \geqslant 0$$

$$\min_x \; \tfrac{1}{2}x^T Hx + c^T x \;,\; Ax \leqslant b \;,\; x \geqslant 0$$

- Convex optimization

$$\min_x \; f(x) \;,\; g(x) \leqslant 0 \quad \text{where } f \text{ and } g \text{ are convex}$$

- Nonlinear optimization

$$\min_x \; f(x) \;,\; h(x) = 0 \;,\; g(x) \leqslant 0$$

where f, h, and g are non-convex and nonlinear

Necessary conditions for extremum → learn by heart!

- Unconstrained optimization problem: $\min_x f(x)$

  Zero-gradient condition: $\nabla f(x) = 0$

- Equality constrained optimization problem: $\min_x f(x)$ s.t. $h(x) = 0$

  Lagrange conditions:

$$\begin{aligned} \nabla f(x) + \nabla h(x)\,\lambda &= 0 \\ h(x) &= 0 \end{aligned}$$

- Inequality constrained optimization problem: $\min_x f(x)$ s.t. $g(x) \leqslant 0$, $h(x) = 0$

  Karush-Kuhn-Tucker conditions:

$$\begin{aligned} \nabla f(x) + \nabla g(x)\,\mu + \nabla h(x)\,\lambda &= 0 \\ \mu^T g(x) &= 0 \\ \mu &\geqslant 0 \\ h(x) &= 0 \\ g(x) &\leqslant 0 \end{aligned}$$

Necessary and sufficient conditions for extremum

- Unconstrained optimization problem:

  $$\min_x f(x)$$

  $\nabla f(x) = 0$ and $H(x) > 0 \rightarrow$ local minimum

  $\nabla f(x) = 0$ and $H(x) < 0 \rightarrow$ local maximum

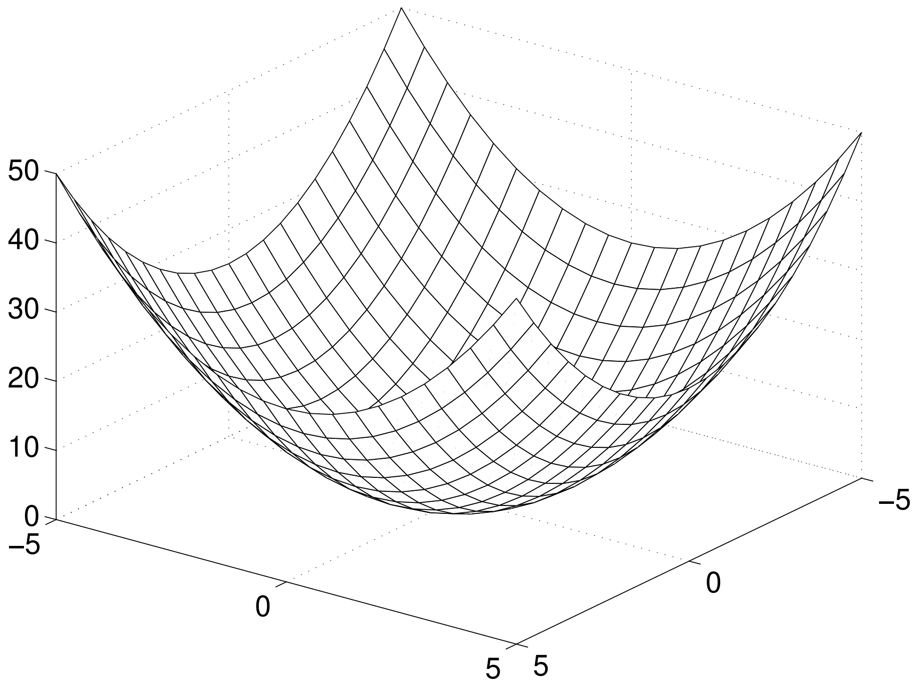
  $\nabla f(x) = 0$ and $H(x)$ indefinite $\rightarrow$ saddle point

- Convex optimization problem:

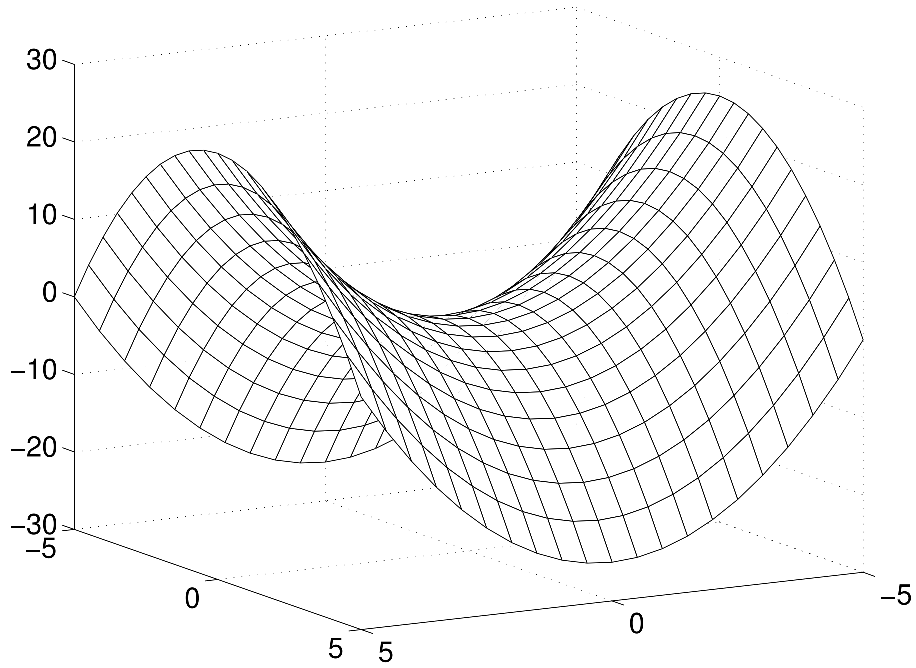
  $$\min_x f(x)$$
  $$\text{s.t. } g(x) \leqslant 0$$

  Karush-Kuhn-Tucker conditions are
  necessary *and* sufficient
  for *global* optimum

Unconstrained optimization



local mininum



saddle point

Stopping criteria

- Linear and Quadratic programming: Finite number of steps
- Convex optimization: $|f(x_k) - f(x^*)| \leqslant \varepsilon_f$, $g(x_k) \leqslant \varepsilon_g$, and for ellipsoid: $\|x_k - x^*\|_2 \leqslant \varepsilon_x$
- Unconstrained nonlinear optimization: $\|\nabla f(x_k)\|_2 \leqslant \varepsilon_\nabla$
- Constrained nonlinear optimization:

$$\begin{aligned} \| \, \nabla f(x_k) + \nabla g(x_k) \, \mu + \nabla h(x_k) \, \lambda \, \|_2 &\leqslant \varepsilon_{KT1} \\ | \, \mu^T \, g(x_k) \, | &\leqslant \varepsilon_{KT2} \\ \mu &\geqslant -\varepsilon_{KT3} \\ \| \, h(x_k) \, \|_2 &\leqslant \varepsilon_{KT4} \\ g(x_k) &\leqslant \varepsilon_{KT5} \end{aligned}$$

- Maximum number of steps
- Heuristic stopping criteria (*last* resort):
  $\|x_{k+1} - x_k\|_2 \leqslant \varepsilon_x$ or $|f(x_{k+1}) - f(x_k)| \leqslant \varepsilon_f$

Summary

- Standard form of optimization problem:
  $\min\limits_{x} f(x)$ s.t. $h(x) = 0, \; g(x) \leqslant 0$
- Classes of optimization problems: linear, quadratic, convex, nonlinear
- Convex sets & functions
- Gradient, subgradient, and Hessian
- Conditions for extremum
- Stopping criteria

Test: Gradient

Given: Level lines of *unimodal* function f with minimum x^*, a point x_0, and vectors v_1, v_2, v_3, v_4, v_5, one of which is equal to $\nabla f(x_0)$.

Question: Which vector v_i is equal to $\nabla f(x_0)$?