

# Lecture 5:

## Duality and convex solvers

Optimization for data sciences



Rémy Sun  
[remy.sun@inria.fr](mailto:remy.sun@inria.fr)



# Course organization

# Course organization

- Introduction to optimization
  - A few problems of interest
  - Quick mathematical refresher
- Convex problems (Following Stephen Boyd)
  - Quick refresher on last week
  - Convex sets
  - Convex functions
  - Convex problems
  - Simplex algorithm for Linear Programming

# Course organization

- **Duality (for convex problems)**
  - Lagrangian and dual function
  - Dual problem
  - Qualification constraints
  - KKT conditions
- **Newton's Descent and Barrier methods for convex case**
  - Descent for the unconstrained problems
  - Equality constrained problems
  - Interior point methods

# Course organization

- (First order) descent methods for the general case
- Backpropagation
- Some more properties on stochastic gradient descent

- Reports on lab sessions
  - Labs on jupyter notebooks
    - Not every session
  - Explain the code done in the session
  - Summarize what is done in the practical
- Written Exam
  - Theoretical questions
  - We will do exercises in class

# Refresher on last weeks

# Convex problem

convex optimization problem:

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && Ax = b \end{aligned}$$

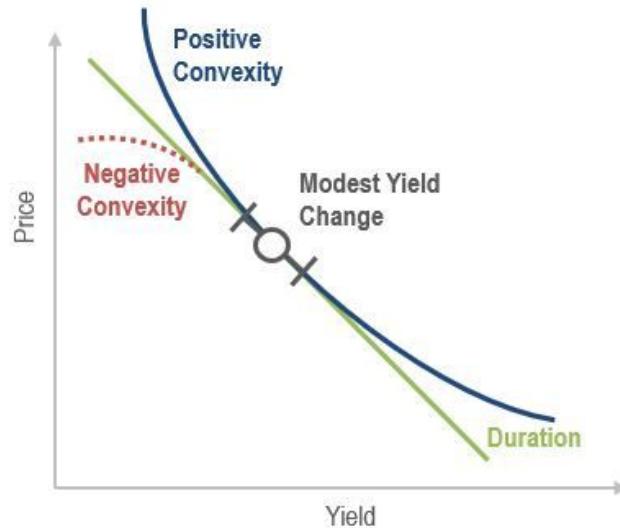
- ▶ variable  $x \in \mathbf{R}^n$
- ▶ equality constraints are linear
- ▶  $f_0, \dots, f_m$  are **convex**: for  $\theta \in [0, 1]$ ,

$$f_i(\theta x + (1 - \theta)y) \leq \theta f_i(x) + (1 - \theta)f_i(y)$$

i.e.,  $f_i$  have nonnegative (upward) curvature

# Easy problem

- ▶ classical view:
  - linear (zero curvature) is easy
  - nonlinear (nonzero curvature) is hard
- ▶ the classical view is **wrong**
- ▶ the correct view:
  - convex (nonnegative curvature) is easy
  - nonconvex (negative curvature) is hard



# Easy to solve!

- ▶ many different algorithms (that run on many platforms)
  - interior-point methods for up to 10000s of variables
  - first-order methods for larger problems
  - do not require initial point, babysitting, or tuning
- ▶ can develop and deploy quickly using modeling languages such as CVXPY
- ▶ solvers are reliable, so can be embedded
- ▶ code generation yields real-time solvers that execute in milliseconds (e.g., on Falcon 9 and Heavy for landing)

- **Convex sets (definition!)**
  - **Affine sets, norm balls, norm cones**
  - **Convex combination, convex hull and Convex cones**
  - **Hyperplanes, halfspaces and polyhedron**
  - Positive Semidefinite Cone
- **Showing a set is convex (with operations!)**
  - **Intersection**
  - **Affine mapping**
  - **Perspective and Linear fractional mappings**
- Proper cones and generalized inequalities
- Separating and supporting hyperplanes

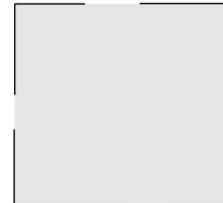
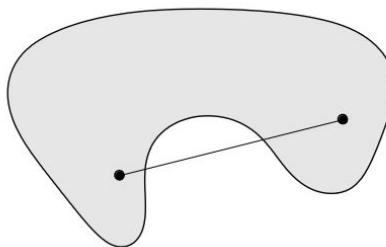
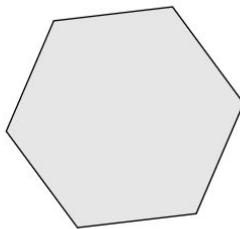
# Convex sets

**line segment** between  $x_1$  and  $x_2$ : all points of form  $x = \theta x_1 + (1 - \theta)x_2$ , with  $0 \leq \theta \leq 1$

**convex set**: contains line segment between any two points in the set

$$x_1, x_2 \in C, \quad 0 \leq \theta \leq 1 \quad \implies \quad \theta x_1 + (1 - \theta)x_2 \in C$$

**examples** (one convex, two nonconvex sets)



# Showing a set is convex

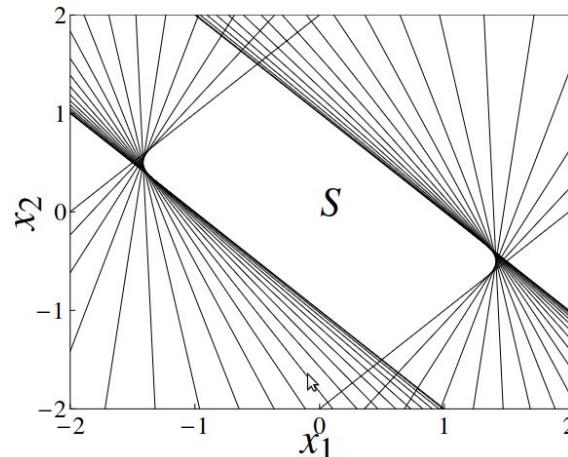
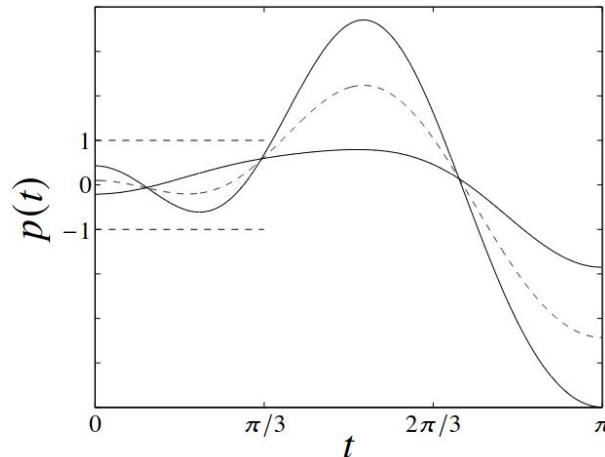
methods for establishing convexity of a set  $C$

1. apply definition: show  $x_1, x_2 \in C, 0 \leq \theta \leq 1 \implies \theta x_1 + (1 - \theta)x_2 \in C$ 
  - recommended only for **very simple** sets
2. use convex functions (next lecture)
3. show that  $C$  is obtained from simple convex sets (hyperplanes, halfspaces, norm balls, ...) by operations that preserve convexity
  - intersection
  - affine mapping
  - perspective mapping
  - linear-fractional mapping

you'll mostly use methods 2 and 3

# Showing a set is convex

- ▶ the intersection of (any number of) convex sets is convex
- ▶ **example:**
  - $S = \{x \in \mathbf{R}^m \mid |p(t)| \leq 1 \text{ for } |t| \leq \pi/3\}$ , with  $p(t) = x_1 \cos t + \cdots + x_m \cos mt$
  - write  $S = \bigcap_{|t| \leq \pi/3} \{x \mid |p(t)| \leq 1\}$ , i.e., an intersection of (convex) slabs
- ▶ picture for  $m = 2$ :



# Takeaway

$$x_1, x_2 \in C, \quad 0 \leq \theta \leq 1 \quad \implies \quad \theta x_1 + (1 - \theta)x_2 \in C$$

- Classic convex sets
  - Affine sets, hyperplanes, cones, balls, polyhedrons
- Convexity preserving operations
  - Intersection
  - Affine mapping
  - Perspective
  - Linear Fractional mapping

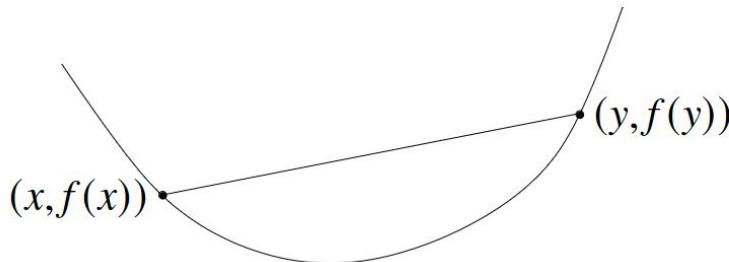
# Convex function overview

- **Convex functions (with definition!)**
  - **Examples of classic convex functions**
  - *Extended value function*
  - *Line restriction*
  - **First and second order conditions**
  - *Epigraph and sublevel sets*
- **Showing a function is convex with operations**
  - **Non-negative weighted sum and affine composition**
  - **Pointwise maximum**
  - **Composition rules**
  - *Partial minimization and perspective*
- *Conjugate function*

# Convex functions

- $f : \mathbf{R}^n \rightarrow \mathbf{R}$  is convex if  $\mathbf{dom}f$  is a convex set and for all  $x, y \in \mathbf{dom}f$ ,  $0 \leq \theta \leq 1$ ,

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$



- $f$  is concave if  $-f$  is convex
- $f$  is strictly convex if  $\mathbf{dom}f$  is convex and for  $x, y \in \mathbf{dom}f$ ,  $x \neq y$ ,  $0 < \theta < 1$ ,

$$f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y)$$

↳

# Showing a function is convex

methods for establishing convexity of a function  $f$

1. verify definition (often simplified by restricting to a line)
2. for twice differentiable functions, show  $\nabla^2 f(x) \geq 0$ 
  - recommended only for **very simple** functions
3. show that  $f$  is obtained from simple convex functions by operations that preserve convexity
  - nonnegative weighted sum
  - composition with affine function
  - pointwise maximum and supremum
  - composition
  - minimization
  - perspective

you'll mostly use methods 2 and 3



# Showing a function is convex

- ▶ **nonnegative multiple:**  $\alpha f$  is convex if  $f$  is convex,  $\alpha \geq 0$
  - ▶ **sum:**  $f_1 + f_2$  convex if  $f_1, f_2$  convex
  - ▶ **infinite sum:** if  $f_1, f_2, \dots$  are convex functions, infinite sum  $\sum_{i=1}^{\infty} f_i$  is convex
  - ▶ **integral:** if  $f(x, \alpha)$  is convex in  $x$  for each  $\alpha \in \mathcal{A}$ , then  $\int_{\alpha \in \mathcal{A}} f(x, \alpha) d\alpha$  is convex
- ↗
- ▶ there are analogous rules for concave functions

# Showing a function is convex

**(pre-)composition with affine function:**  $f(Ax + b)$  is convex if  $f$  is convex

## examples

- ▶ log barrier for linear inequalities

$$f(x) = - \sum_{i=1}^m \log(b_i - a_i^T x), \quad \text{dom } f = \{x \mid a_i^T x < b_i, i = 1, \dots, m\}$$

- ▶ norm approximation error:  $f(x) = \|Ax - b\|$  (any norm)

# Showing a function is convex

if  $f_1, \dots, f_m$  are convex, then  $f(x) = \max\{f_1(x), \dots, f_m(x)\}$  is convex

## examples

- ▶ piecewise-linear function:  $f(x) = \max_{i=1,\dots,m}(a_i^T x + b_i)$
- ▶ sum of  $r$  largest components of  $x \in \mathbf{R}^n$ :

$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

( $x_{[i]}$  is  $i$ th largest component of  $x$ )

□

proof:  $f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$

# Showing a function is convex

- ▶ composition of  $g : \mathbf{R}^n \rightarrow \mathbf{R}$  and  $h : \mathbf{R} \rightarrow \mathbf{R}$  is  $f(x) = h(g(x))$  (written as  $f = h \circ g$ )
- ▶ composition  $f$  is convex if
  - $g$  convex,  $h$  convex,  $\tilde{h}$  nondecreasing
  - or  $g$  concave,  $h$  convex,  $\tilde{h}$  nonincreasing(monotonicity must hold for extended-value extension  $\tilde{h}$ )
- ▶ proof (for  $n = 1$ , differentiable  $g, h$ )

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

## examples

- ▶  $f(x) = \exp g(x)$  is convex if  $g$  is convex
- ▶  $f(x) = 1/g(x)$  is convex if  $g$  is concave and positive

↳

# Takeaway

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

- Classic convex functions
  - Affine, exponential, norms, max, ...
- Convexity preserving operations
  - Non negative weighted sum, composition with affine
  - Pointwise maximum and supremum
  - Composition
  - Minimization
  - Perspective

# Standard form optimization problem

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

- ▶  $x \in \mathbf{R}^n$  is the optimization variable
- ▶  $f_0 : \mathbf{R}^n \rightarrow \mathbf{R}$  is the objective or cost function
- ▶  $f_i : \mathbf{R}^n \rightarrow \mathbf{R}, i = 1, \dots, m,$  are the inequality constraint functions
- ▶  $h_i : \mathbf{R}^n \rightarrow \mathbf{R}$  are the equality constraint functions



# Feasible and optimal points

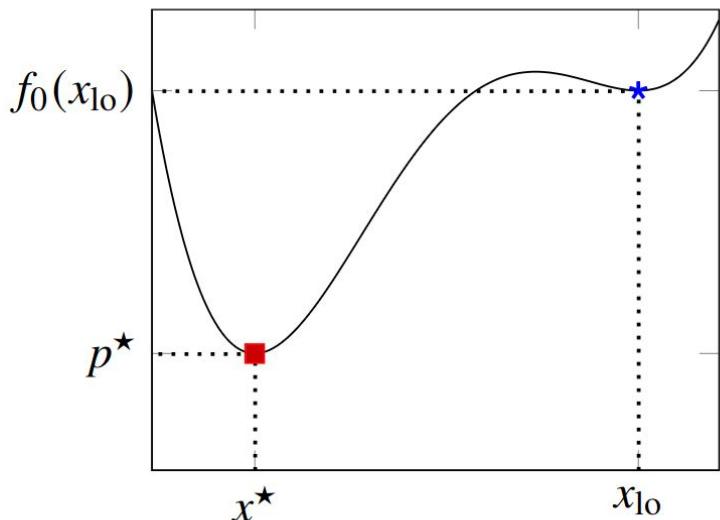
- ▶  $x \in \mathbf{R}^n$  is **feasible** if  $x \in \text{dom } f_0$  and it satisfies the constraints
- ▶ **optimal value** is  $p^\star = \inf\{f_0(x) \mid f_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p\}$
- ▶  $p^\star = \infty$  if problem is infeasible
- ▶  $p^\star = -\infty$  if problem is **unbounded below**
- ▶ a feasible  $x$  is **optimal** if  $f_0(x) = p^\star$
- ▶  $X_{\text{opt}}$  is the set of optimal points



# Optimal points

$x$  is **locally optimal** if there is an  $R > 0$  such that  $x$  is optimal for

$$\begin{array}{ll}\text{minimize (over } z) & f_0(z) \\ \text{subject to} & f_i(z) \leq 0, \quad i = 1, \dots, m, \quad h_i(z) = 0, \quad i = 1, \dots, p \\ & \|z - x\|_2 \leq R\end{array}$$



# Standard form convex problem

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && a_i^T x = b_i, \quad i = 1, \dots, p \end{aligned}$$

- ▶ objective and inequality constraints  $f_0, f_1, \dots, f_m$  are convex
  - ▶ equality constraints are affine, often written as  $Ax = b$
  - ▶ feasible and optimal sets of a convex optimization problem are convex
- ↗
- ▶ problem is **quasiconvex** if  $f_0$  is quasiconvex,  $f_1, \dots, f_m$  are convex,  $h_1, \dots, h_p$  are affine

# Optimum in a convex set

any locally optimal point of a convex problem is (globally) optimal

**proof:**

- ▶ suppose  $x$  is locally optimal, but there exists a feasible  $y$  with  $f_0(y) < f_0(x)$
- ▶  $x$  locally optimal means there is an  $R > 0$  such that

$$z \text{ feasible}, \quad \|z - x\|_2 \leq R \quad \implies \quad f_0(z) \geq f_0(x)$$

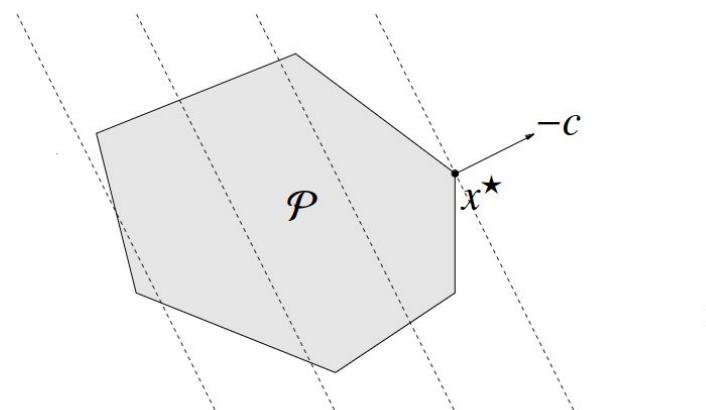
- ▶ consider  $z = \theta y + (1 - \theta)x$  with  $\theta = R/(2\|y - x\|_2)$
- ▶  $\|y - x\|_2 > R$ , so  $0 < \theta < 1/2$
- ▶  $z$  is a convex combination of two feasible points, hence also feasible
- ▶  $\|z - x\|_2 = R/2$  and  $f_0(z) \leq \theta f_0(y) + (1 - \theta)f_0(x) < f_0(x)$ , which contradicts our assumption that  $x$  is locally optimal



$$\begin{array}{ll}\text{minimize} & c^T x + d \\ \text{subject to} & Gx \leq h \\ & Ax = b\end{array}$$

**Often written as**  
**a maximization**

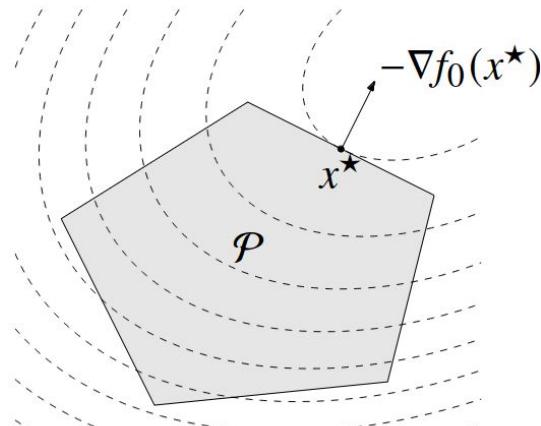
- ▶ convex problem with affine objective and constraint functions
- ▶ feasible set is a polyhedron



# Quadratic programming

$$\begin{array}{ll}\text{minimize} & (1/2)x^T Px + q^T x + r \\ \text{subject to} & Gx \leq h \\ & Ax = b\end{array}$$

- ▶  $P \in \mathbf{S}_+^n$ , so objective is convex quadratic
- ▶ minimize a convex quadratic function over a polyhedron



# Quadratically constrained Quadratic programming (QCQP)

$$\begin{aligned} & \text{minimize} && (1/2)x^T P_0 x + q_0^T x + r_0 \\ & \text{subject to} && (1/2)x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \dots, m \\ & && Ax = b \end{aligned}$$

- ▶  $P_i \in \mathbf{S}_+^n$ ; objective and constraints are convex quadratic
- ▶ if  $P_1, \dots, P_m \in \mathbf{S}_{++}^n$ , feasible region is intersection of  $m$  ellipsoids and an affine set

# Change of variable

- ▶  $\phi : \mathbf{R}^n \rightarrow \mathbf{R}^n$  is one-to-one with  $\phi(\mathbf{dom} \phi) \supseteq \mathcal{D}$
- ▶ consider (possibly non-convex) problem

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

- ▶ change variables to  $z$  with  $x = \phi(z)$
- ▶ can solve equivalent problem

$$\begin{aligned} & \text{minimize} && \tilde{f}_0(z) \\ & \text{subject to} && \tilde{f}_i(z) \leq 0, \quad i = 1, \dots, \overset{\mathbb{I}}{m} \\ & && \tilde{h}_i(z) = 0, \quad i = 1, \dots, p \end{aligned}$$

where  $\tilde{f}_i(z) = f_i(\phi(z))$  and  $\tilde{h}_i(z) = h_i(\phi(z))$

- ▶ recover original optimal point as  $x^\star = \phi(z^\star)$

# Transformation

suppose

- ▶  $\phi_0$  is monotone increasing
- ▶  $\psi_i(u) \leq 0$  if and only if  $u \leq 0$ ,  $i = 1, \dots, m$
- ▶  $\varphi_i(u) = 0$  if and only if  $u = 0$ ,  $i = 1, \dots, p$

standard form optimization problem is equivalent to

$$\begin{aligned} & \text{minimize} && \phi_0(f_0(x)) \\ & \text{subject to} && \psi_i(f_i(x)) \leq 0, \quad i = 1, \dots, m \\ & && \varphi_i(h_i(x)) = 0, \quad i = 1, \dots, p \end{aligned}$$



example: minimizing  $\|Ax - b\|$  is equivalent to minimizing  $\|Ax - b\|^2$

# Maximization and minimization

- ▶ suppose  $\phi_0$  is monotone decreasing
- ▶ the maximization problem

$$\begin{aligned} & \text{maximize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

is equivalent to the minimization problem

$$\begin{aligned} & \text{minimize} && \phi_0(f_0(x)) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

- ▶ **examples:**
  - $\phi_0(u) = -u$  transforms maximizing a concave function to minimizing a convex function
  - $\phi_0(u) = 1/u$  transforms maximizing a concave positive function to minimizing a convex function

# Eliminating equality constraints

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && Ax = b \end{aligned}$$

is equivalent to

$$\begin{aligned} & \text{minimize (over } z) && f_0(Fz + x_0) \\ & \text{subject to} && f_i(Fz + x_0) \leq 0, \quad i = 1, \dots, m \end{aligned}$$

where  $F$  and  $x_0$  are such that  $Ax = b \iff x = Fz + x_0$  for some  $z$

# Introducing equality constraints

$$\begin{aligned} & \text{minimize} && f_0(A_0x + b_0) \\ & \text{subject to} && f_i(A_i x + b_i) \leq 0, \quad i = 1, \dots, m \end{aligned}$$

is equivalent to

$$\begin{aligned} & \text{minimize (over } x, y_i\text{)} && f_0(y_0) \\ & \text{subject to} && f_i(y_i) \leq 0, \quad i = 1, \dots, m \\ & && y_i = A_i x + b_i, \quad i = 0, 1, \dots, m \end{aligned}$$



# Slack variables for linear equalities

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && a_i^T x \leq b_i, \quad i = 1, \dots, m \end{aligned}$$

is equivalent to

$$\begin{aligned} & \text{minimize (over } x, s) && f_0(x) \\ & \text{subject to} && a_i^T x + s_i = b_i, \quad i = 1, \dots, m \\ & && s_i \geq 0, \quad i = 1, \dots, m \end{aligned}$$

minimize  $f_0(x)$

subject to  $f_i(x) \leq 0, \quad i = 1, \dots, m$

$h_i(x) = 0, \quad i = 1, \dots, p$

- Convex  $f$  and linear  $h$ 
  - $X$  feasible: satisfies implicit and explicit constraints
- Quite a few classical convex problems(linear, quadratic, ...)
- Easy to change variables between equivalent problems

# Simplex tableau

$$\begin{array}{rcl}
 x_4 & = & 5 - 2x_1 - 3x_2 - x_3 \\
 x_5 & = & 11 - 4x_1 - x_2 - 2x_3 \\
 x_6 & = & 8 - 3x_1 - 4x_2 - 2x_3 \\
 \hline
 z & = & 5x_1 + 4x_2 + 3x_3.
 \end{array}$$

x1	x2	x3	x4	x5	x6	z	c
2	3	1	1	0	0	0	5
4	1	2	0	1	0	0	11
3	4	2	0	0	1	0	8
-5	-4	-3	0	0	0	1	0

# Simplex tableau

x1	x2	x3	x4	x5	x6	Z	C
2	3	1	1	0	0	0	5 (5/2)
4	1	2	0	1	0	0	11 (11/4)
3	4	2	0	0	1	0	8 (8/3)
-5	-4	-3	0	0	0	1	0

x1	x2	x3	x4	x5	x6	Z	C
1	1.5	0.5	0.5	0	0	0	2.5
4	1	2	0	1	0	0	11
3	4	2	0	0	1	0	8
-5	-4	-3	0	0	0	1	0

# Simplex tableau

x1	x2	x3	x4	x5	x6	Z	C
2	3	1	1	0	0	0	5 (5/2)
4	1	2	0	1	0	0	11 (11/4)
3	4	2	0	0	1	0	8 (8/3)
-5	-4	-3	0	0	0	1	0

x1	x2	x3	x4	x5	x6	Z	C
1	1.5	0.5	0.5	0	0	0	2.5
0	-5	0	-2	1	0	0	1
0	-0.5	0.5	-1.5	0	1	0	0.5
0	3.5	-0.5	2.5	0	0	1	12.5

# Simplex tableau

$$\begin{array}{rcl}
 x_1 & = & \frac{5}{2} - \frac{3}{2}x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4 \\
 x_5 & = & 1 + 5x_2 \\
 x_6 & = & \frac{1}{2} + \frac{1}{2}x_2 - \frac{1}{2}x_3 + \frac{3}{2}x_4 \\
 \hline
 z & = & \frac{25}{2} - \frac{7}{2}x_2 + \frac{1}{2}x_3 - \frac{5}{2}x_4.
 \end{array}$$

x1	x2	x3	x4	x5	x6	Z	C
1	1.5	0.5	0.5	0	0	0	2.5
0	-5	0	-2	1	0	0	1
0	-0.5	0.5	-1.5	0	1	0	0.5
0	3.5	-0.5	2.5	0	0	1	12.5

# Simplex tableau

x1	x2	x3	x4	x5	x6	Z	C
1	1.5	0.5	0.5	0	0	0	2.5 (5)
0	-5	0	-2	1	0	0	1 (inf)
0	-0.5	0.5	-1.5	0	1	0	0.5 (1)
0	3.5	-0.5	2.5	0	0	1	12.5

x1	x2	x3	x4	x5	x6	Z	C
1	1.5	0.5	0.5	0	0	0	2.5
0	-5	0	-2	1	0	0	1
0	-1	1	-3	0	2	0	1
0	3.5	-0.5	2.5	0	0	1	12.5

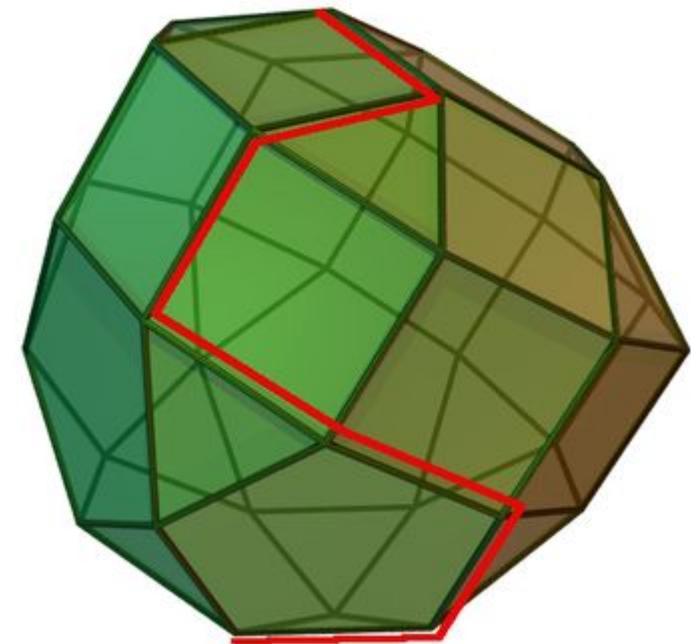
# Simplex tableau

$$\begin{array}{rcl}
 x_3 & = & 1 + x_2 + 3x_4 - 2x_6 \\
 x_1 & = & 2 - 2x_2 - 2x_4 + x_6 \\
 x_5 & = & 1 + 5x_2 + 2x_4 \\
 \hline
 z & = & 13 - 3x_2 - x_4 - x_6.
 \end{array}$$

x1	x2	x3	x4	x5	x6	Z	C
1	2	0	2	0	-1	0	2
0	-5	0	-2	1	0	0	1
0	-1	1	-3	0	2	0	1
0	3	0	1	0	1	1	13

# Algorithm

- Build tableau from canonical
- Check we have feasible solution
- Do a pivot step if negative coef
  - Pick column  $c$  w/ most negative coefficient
  - Pick row  $r$  w/ smallest ratio
  - Pivot! (set  $c$  to 1 in  $r$  and 0 in other rows)



# Unconstrained problem version

minimize       $f_0(x)$   
subject to     $f_i(x) \leq 0, \quad i = 1, \dots, m$   
                   $h_i(x) = 0, \quad i = 1, \dots, p,$

# Unconstrained problem version

minimize  $f_0(x)$   
subject to  $f_i(x) \leq 0, \quad i = 1, \dots, m$   
 $h_i(x) = 0, \quad i = 1, \dots, p,$

$$I_-(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0, \end{cases}$$

minimize  $f_0(x) + \sum_{i=1}^m I_-(f_i(x)) + \sum_{i=1}^p I_0(h_i(x)),$

# Relaxing the indicator function

$$I_-(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0, \end{cases}$$

minimize  $f_0(x) + \sum_{i=1}^m I_-(f_i(x)) + \sum_{i=1}^p I_0(h_i(x)),$

minimize  $L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x).$

# Lagrangian

- ▶ **standard form problem** (not necessarily convex)

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

variable  $x \in \mathbf{R}^n$ , domain  $\mathcal{D}$ , optimal value  $p^\star$

- ▶ **Lagrangian:**  $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \rightarrow \mathbf{R}$ , with  $\text{dom } L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$ ,

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

- weighted sum of objective and constraint functions
- $\lambda_i$  is **Lagrange multiplier** associated with  $f_i(x) \leq 0$
- $\nu_i$  is Lagrange multiplier associated with  $h_i(x) = 0$

# Lagrange dual

- ▶ **Lagrange dual function:**  $g : \mathbf{R}^m \times \mathbf{R}^p \rightarrow \mathbf{R}$ ,

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left( f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right)$$

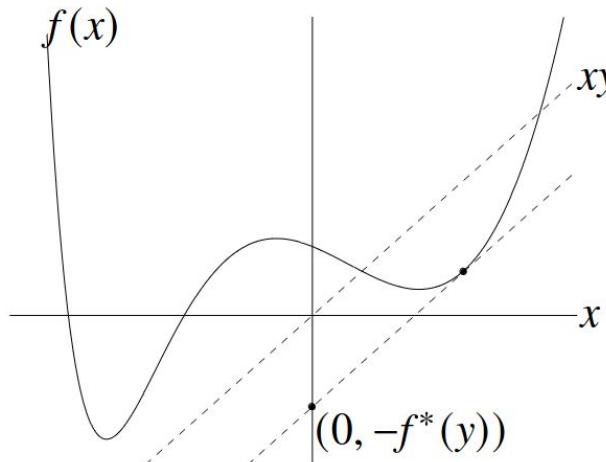
- ▶  $g$  is concave, can be  $-\infty$  for some  $\lambda, \nu$
- ▶ **lower bound property:** if  $\lambda \geq 0$ , then  $g(\lambda, \nu) \leq p^\star$
- ▶ proof: if  $\tilde{x}$  is feasible and  $\lambda \geq 0$ , then

$$f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)$$

minimizing over all feasible  $\tilde{x}$  gives  $p^\star \geq g(\lambda, \nu)$

# Conjugate

- ▶ the **conjugate** of a function  $f$  is  $f^*(y) = \sup_{x \in \text{dom } f} (y^T x - f(x))$



↳

- ▶  $f^*$  is convex (even if  $f$  is not)
- ▶ will be useful in chapter 5

# Lagrange dual and conjugate

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && Ax \leq b, \quad Cx = d \end{aligned}$$

- ▶ dual function

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathbf{dom} f_0} \left( f_0(x) + (A^T \lambda + C^T \nu)^T x - b^T \lambda - d^T \nu \right) \\ &= -f_0^*(-A^T \lambda - C^T \nu) - b^T \lambda - d^T \nu \end{aligned}$$

where  $f^*(y) = \sup_{x \in \mathbf{dom} f}(y^T x - f(x))$  is conjugate of  $f_0$

- ▶ simplifies derivation of dual if conjugate of  $f_0$  is known
- ▶ **example: entropy maximization**

$$f_0(x) = \sum_{i=1}^n x_i \log x_i, \quad f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

# Dual problem

## (Lagrange) **dual problem**

$$\begin{aligned} & \text{maximize} && g(\lambda, \nu) \\ & \text{subject to} && \lambda \geq 0 \end{aligned}$$

- ▶ finds best lower bound on  $p^*$ , obtained from Lagrange dual function
- ▶ a convex optimization problem, even if original **primal** problem is not
- ▶ dual optimal value denoted  $d^*$
- ▶  $\lambda, \nu$  are dual feasible if  $\lambda \geq 0, (\lambda, \nu) \in \mathbf{dom} g$
- ▶ often simplified by making implicit constraint  $(\lambda, \nu) \in \mathbf{dom} g$  explicit

# Weak and strong duality

**weak duality:**  $d^* \leq p^*$

- ▶ always holds (for convex and nonconvex problems)
- ▶ can be used to find nontrivial lower bounds for difficult problems, e.g., solving the SDP

$$\begin{aligned} & \text{maximize} && -\mathbf{1}^T v \\ & \text{subject to} && W + \mathbf{diag}(v) \succeq 0 \end{aligned}$$

gives a lower bound for the two-way partitioning problem on page 5.7

**strong duality:**  $d^* = p^*$

- ▶ does not hold in general
- ▶ (usually) holds for convex problems
- ▶ conditions that guarantee strong duality in convex problems are called **constraint qualifications**

# Slater's constraint qualification

strong duality holds for a convex problem

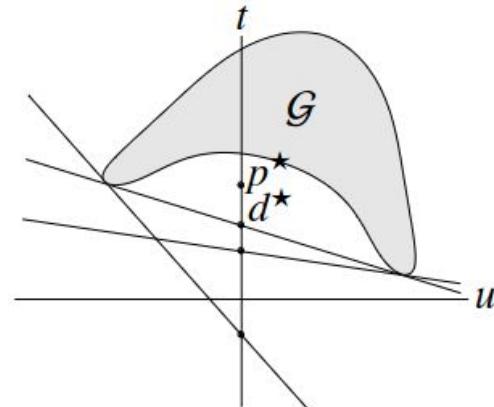
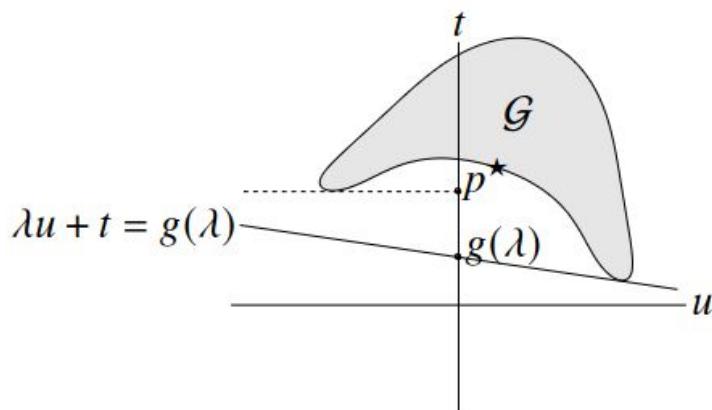
$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && Ax = b \end{aligned}$$

if it is **strictly feasible**, i.e., there is an  $x \in \mathbf{int} \mathcal{D}$  with  $f_i(x) < 0, i = 1, \dots, m, Ax = b$

- ▶ also guarantees that the dual optimum is attained (if  $p^* > -\infty$ )
- ▶ can be sharpened: e.g.,
  - can replace  $\mathbf{int} \mathcal{D}$  with  $\mathbf{relint} \mathcal{D}$  (interior relative to affine hull)
  - affine inequalities do not need to hold with strict inequality
- ▶ there are many other types of constraint qualifications

# Geometric interpretation

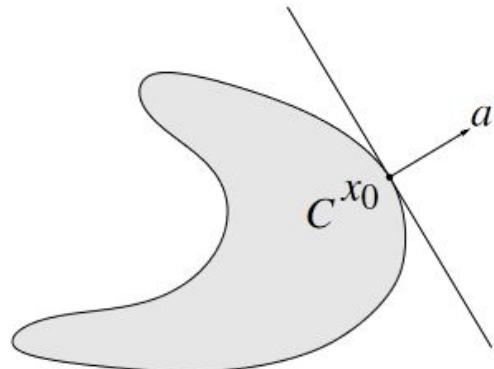
- ▶ for simplicity, consider problem with one constraint  $f_1(x) \leq 0$
- ▶  $\mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$  is set of achievable (constraint, objective) values
- ▶ **interpretation of dual function:**  $g(\lambda) = \inf_{(u,t) \in \mathcal{G}} (t + \lambda u)$



- ▶  $\lambda u + t = g(\lambda)$  is (non-vertical) supporting hyperplane to  $\mathcal{G}$
- ▶ hyperplane intersects  $t$ -axis at  $t = g(\lambda)$

# Supporting hyperplanes

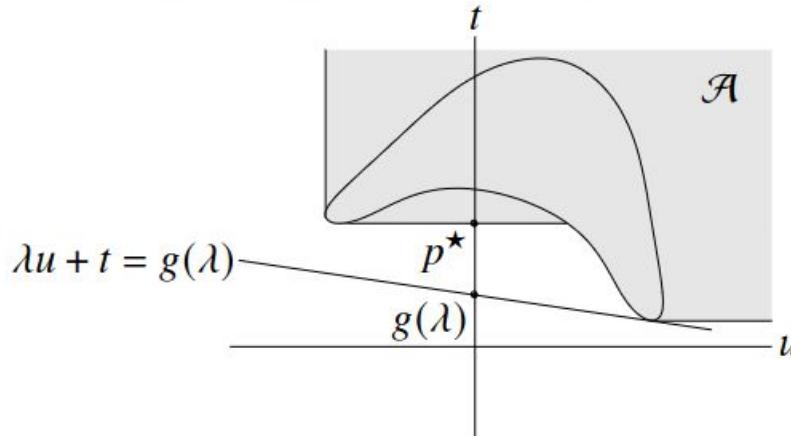
- ▶ suppose  $x_0$  is a boundary point of set  $C \subset \mathbf{R}^n$
- ▶ **supporting hyperplane** to  $C$  at  $x_0$  has form  $\{x \mid a^T x = a^T x_0\}$ , where  $a \neq 0$  and  $a^T x \leq a^T x_0$  for all  $x \in C$



- ▶ **supporting hyperplane theorem:** if  $C$  is convex, then there exists a supporting hyperplane at every boundary point of  $C$

# Geometric interpretation

- ▶ same with  $\mathcal{G}$  replaced with  $\mathcal{A} = \{(u, t) \mid f_1(x) \leq u, f_0(x) \leq t \text{ for some } x \in \mathcal{D}\}$



- ▶ strong duality holds if there is a non-vertical supporting hyperplane to  $\mathcal{A}$  at  $(0, p^*)$
- ▶ for convex problem,  $\mathcal{A}$  is convex, hence has supporting hyperplane at  $(0, p^*)$
- ▶ Slater's condition: if there exist  $(\tilde{u}, \tilde{t}) \in \mathcal{A}$  with  $\tilde{u} < 0$ , then supporting hyperplane at  $(0, p^*)$  must be non-vertical

# Complementary slackness

- ▶ assume strong duality holds,  $x^\star$  is primal optimal,  $(\lambda^\star, \nu^\star)$  is dual optimal

$$\begin{aligned} f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_x \left( f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\ &\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\ &\leq f_0(x^\star) \end{aligned}$$

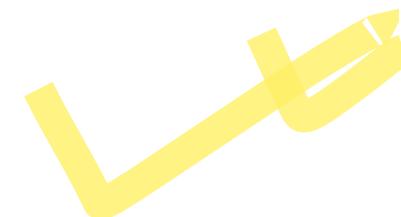
- ▶ hence, the two inequalities hold with equality
- ▶  $x^\star$  minimizes  $L(x, \lambda^\star, \nu^\star)$
- ▶  $\lambda_i^\star f_i(x^\star) = 0$  for  $i = 1, \dots, m$  (known as **complementary slackness**):

$$\lambda_i^\star > 0 \implies f_i(x^\star) = 0, \quad f_i(x^\star) < 0 \implies \lambda_i^\star = 0$$

# Karush-Kuhn-Tucker (KKT) conditions

the **KKT conditions** (for a problem with differentiable  $f_i, h_i$ ) are

1. primal constraints:  $f_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p$
2. dual constraints:  $\lambda \geq 0$
3. complementary slackness:  $\lambda_i f_i(x) = 0, i = 1, \dots, m$
4. gradient of Lagrangian with respect to  $x$  vanishes:



$$\nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + \sum_{i=1}^p \nu_i \nabla h_i(x) = 0$$

if strong duality holds and  $x, \lambda, \nu$  are optimal, they satisfy the KKT conditions

# KKT for convex problems

if  $\tilde{x}$ ,  $\tilde{\lambda}$ ,  $\tilde{v}$  satisfy KKT for a convex problem, then they are optimal:

- ▶ from complementary slackness:  $f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{v})$
- ▶ from 4th condition (and convexity):  $g(\tilde{\lambda}, \tilde{v}) = L(\tilde{x}, \tilde{\lambda}, \tilde{v})$

hence,  $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{v})$

if Slater's condition is satisfied, then

*x is optimal if and only if there exist  $\lambda, v$  that satisfy KKT conditions*

- ▶ recall that Slater implies strong duality, and dual optimum is attained
- ▶ generalizes optimality condition  $\nabla f_0(x) = 0$  for unconstrained problem

# Takeaway

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left( f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right)$$

- Mirror problem that is always convex!
- Gives a lower bound on solution (weak duality)
- Can give the exact solution
  - Under qualifications on constraints for convex problems
- KKT conditions can help reverse engineer a solution

# 1. Unconstrained problems

# General problem statement

- ▶ unconstrained minimization problem

minimize  $f(x)$

- ▶ we assume
  - $f$  convex, twice continuously differentiable (hence  $\text{dom } f$  open)
  - optimal value  $p^\star = \inf_x f(x)$  is attained at  $x^\star$  (not necessarily unique)
- ▶ optimality condition is  $\nabla f(x) = 0$
- ▶ minimizing  $f$  is the same as solving  $\nabla f(x) = 0$
- ▶ a set of  $n$  equations with  $n$  unknowns

# Closed form solution

- ▶ convex quadratic:  $f(x) = (1/2)x^T Px + q^T x + r, P \succeq 0$

- ▶ we can solve exactly via linear equations

$$\nabla f(x) = Px + q = 0$$

- ▶ much more on this special case later

- ▶ for most non-quadratic functions, we use **iterative methods**
- ▶ these produce a sequence of points  $x^{(k)} \in \text{dom} f$ ,  $k = 0, 1, \dots$
- ▶  $x^{(0)}$  is the **initial point** or **starting point**
- ▶  $x^{(k)}$  is the  $k$ th **iterate**
- ▶ we hope that the method **converges**, *i.e.*,

$$f(x^{(k)}) \rightarrow p^*, \quad \nabla f(x^{(k)}) \rightarrow 0$$



# Starting point and sublevel sets

- ▶ algorithms in this chapter require a starting point  $x^{(0)}$  such that
  - $x^{(0)} \in \text{dom } f$
  - sublevel set  $S = \{x \mid f(x) \leq f(x^{(0)})\}$  is closed
- ▶ 2nd condition is hard to verify, except when **all** sublevel sets are closed
  - equivalent to condition that  $\text{epi } f$  is closed
  - true if  $\text{dom } f = \mathbf{R}^n$
  - true if  $f(x) \rightarrow \infty$  as  $x \rightarrow \text{bd dom } f$
- ▶ examples of differentiable functions with closed sublevel sets:

$$f(x) = \log \left( \sum_{i=1}^m \exp(a_i^T x + b_i) \right), \quad f(x) = - \sum_{i=1}^m \log(b_i - a_i^T x)$$

# Descent methods

- descent methods generate iterates as

$$x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)}$$

with  $f(x^{(k+1)}) < f(x^{(k)})$  (hence the name)

- other notations:  $x^+ = x + t\Delta x$ ,  $x := x + t\Delta x$
- $\Delta x^{(k)}$  is the **step**, or **search direction**
- $t^{(k)} > 0$  is the **step size**, or **step length**
- from convexity,  $f(x^+) < f(x)$  implies  $\nabla f(x)^T \Delta x < 0$
- this means  $\Delta x$  is a **descent direction**

# Descent methods

---

**General descent method.**

**given** a starting point  $x \in \text{dom } f$ .

**repeat**

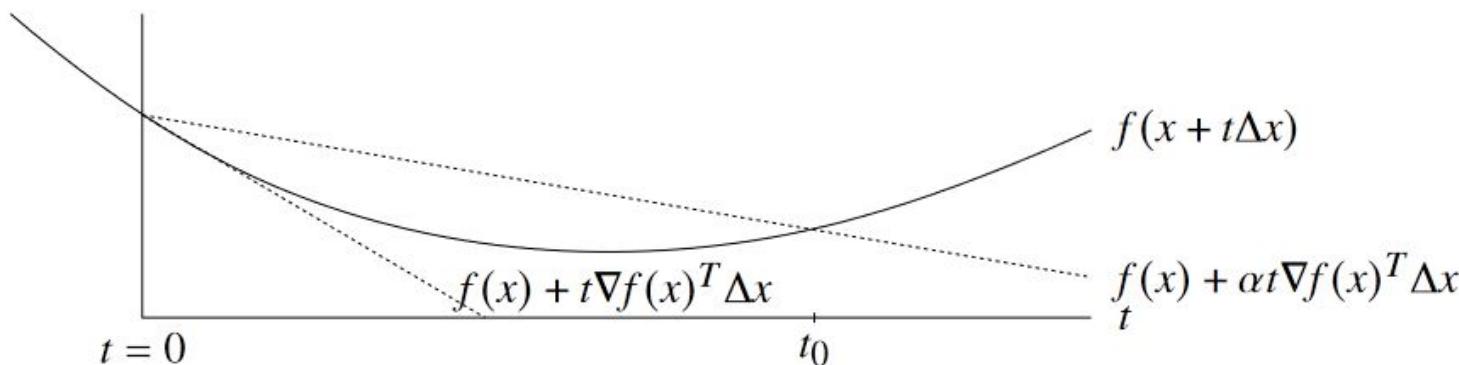
1. Determine a descent direction  $\Delta x$ .
2. **Line search.** Choose a step size  $t > 0$ .
3. **Update.**  $x := x + t\Delta x$ .

**until** stopping criterion is satisfied.

---

# Line search

- ▶ **exact line search:**  $t = \operatorname{argmin}_{t>0} f(x + t\Delta x)$
- ▶ **backtracking line search** (with parameters  $\alpha \in (0, 1/2)$ ,  $\beta \in (0, 1)$ )
  - starting at  $t = 1$ , repeat  $t := \beta t$  until  $f(x + t\Delta x) < f(x) + \alpha t \nabla f(x)^T \Delta x$
- ▶ graphical interpretation: reduce  $t$  (i.e., backtrack) until  $t \leq t_0$



# Gradient descent

- ▶ general descent method with  $\Delta x = -\nabla f(x)$

---

given a starting point  $x \in \text{dom } f$ .

repeat

1.  $\Delta x := -\nabla f(x)$ .

2. **Line search.** Choose step size  $t$  via exact or backtracking line search.

3. **Update.**  $x := x + t\Delta x$ .

until stopping criterion is satisfied.

---

- ▶ stopping criterion usually of the form  $\|\nabla f(x)\|_2 \leq \epsilon$
- ▶ convergence result: for strongly convex  $f$ ,

$$f(x^{(k)}) - p^\star \leq c^k(f(x^{(0)}) - p^\star)$$

$c \in (0, 1)$  depends on  $m$ ,  $x^{(0)}$ , line search type

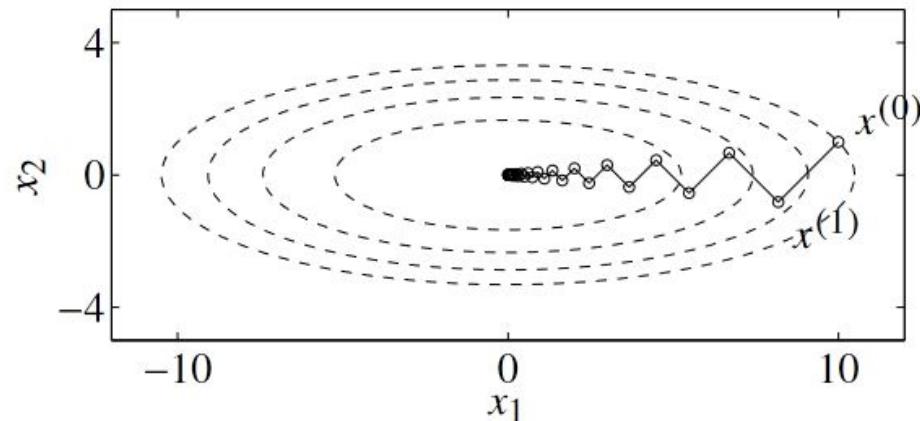
- ▶ very simple, but can be very slow

# Example of gradient descent

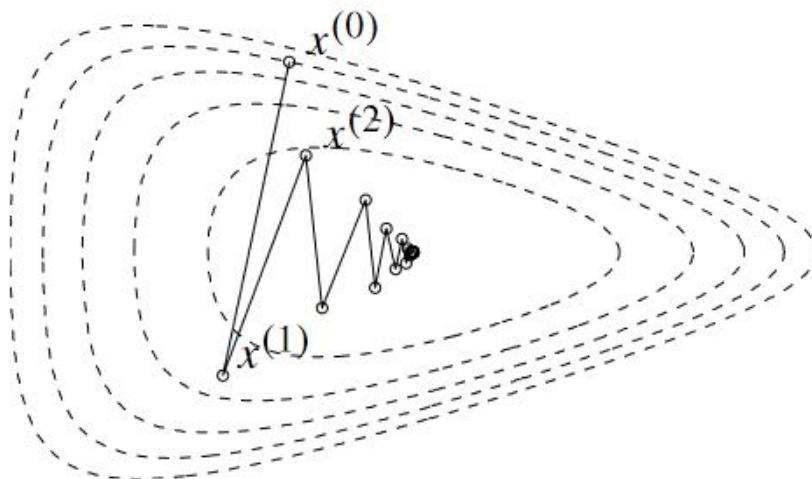
- ▶ take  $f(x) = (1/2)(x_1^2 + \gamma x_2^2)$ , with  $\gamma > 0$
- ▶ with exact line search, starting at  $x^{(0)} = (\gamma, 1)$ :

$$x_1^{(k)} = \gamma \left( \frac{\gamma - 1}{\gamma + 1} \right)^k, \quad x_2^{(k)} = \left( -\frac{\gamma - 1}{\gamma + 1} \right)^k$$

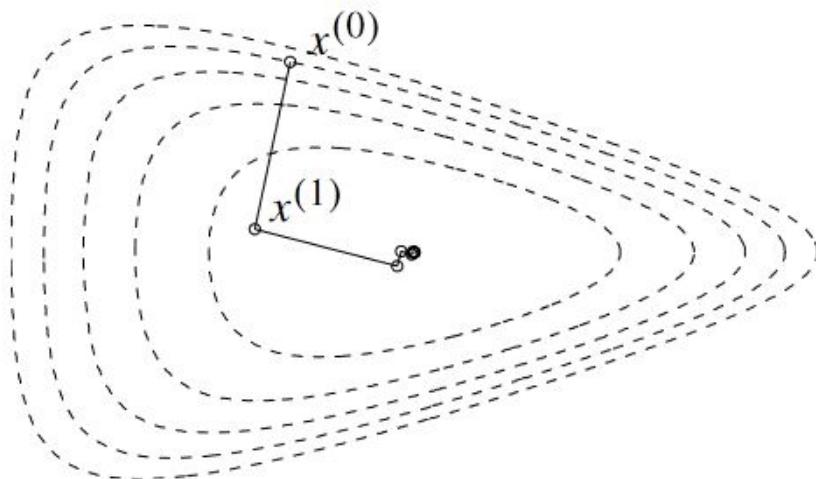
- very slow if  $\gamma \gg 1$  or  $\gamma \ll 1$
- example for  $\gamma = 10$  at right
- called zig-zagging



# Example of gradient descent

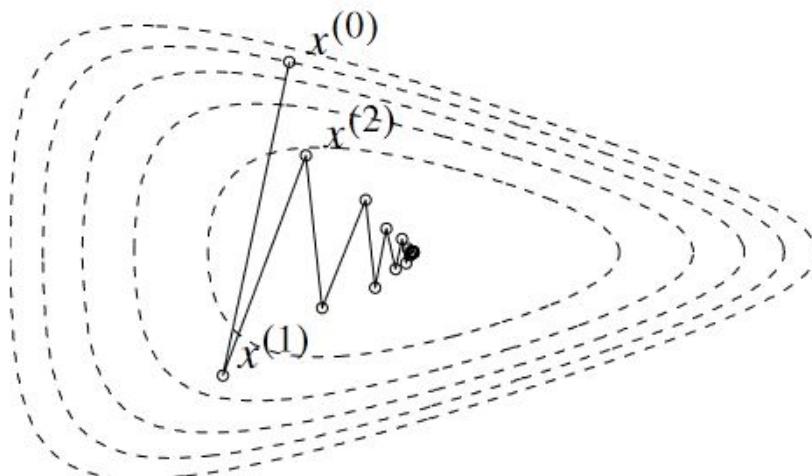


backtracking line search

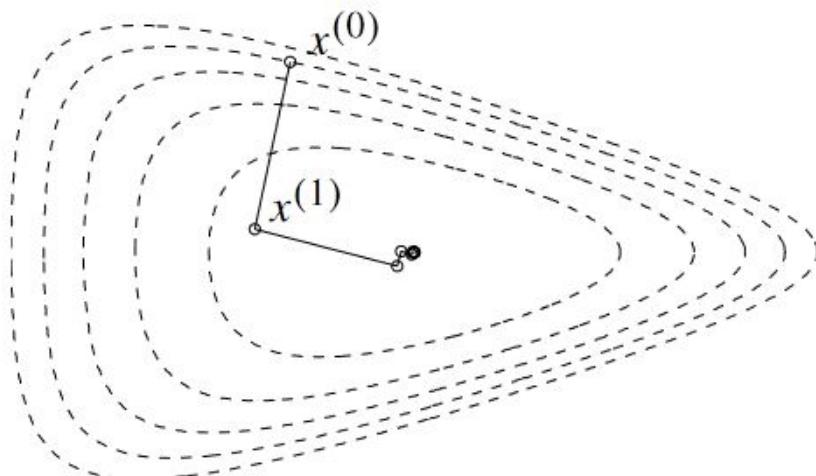


exact line search

# Example of gradient descent



backtracking line search



exact line search

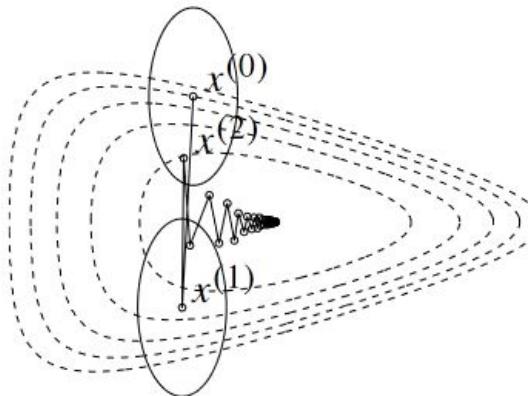
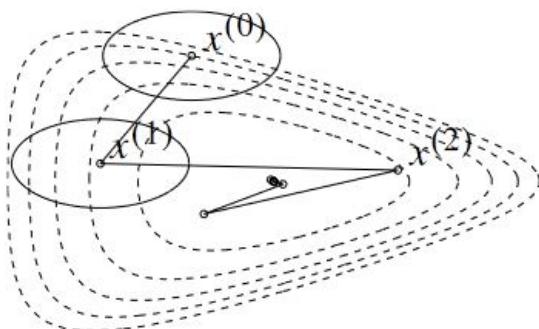
# Steepest descent

- ▶ **normalized steepest descent direction** (at  $x$ , for norm  $\|\cdot\|$ ):

$$\Delta x_{\text{nsd}} = \operatorname{argmin}\{\nabla f(x)^T v \mid \|v\| = 1\}$$

- ▶ interpretation: for small  $v$ ,  $f(x + v) \approx f(x) + \nabla f(x)^T v$ ;
- ▶ direction  $\Delta x_{\text{nsd}}$  is unit-norm step with most negative directional derivative
- ▶ **(unnormalized) steepest descent direction:**  $\Delta x_{\text{sd}} = \|\nabla f(x)\|_* \Delta x_{\text{nsd}}$
- ▶ satisfies  $\nabla f(x)^T \Delta x_{\text{sd}} = -\|\nabla f(x)\|_*^2$
- ▶ **steepest descent method**
  - general descent method with  $\Delta x = \Delta x_{\text{sd}}$
  - convergence properties similar to gradient descent

# Steepest descent norm

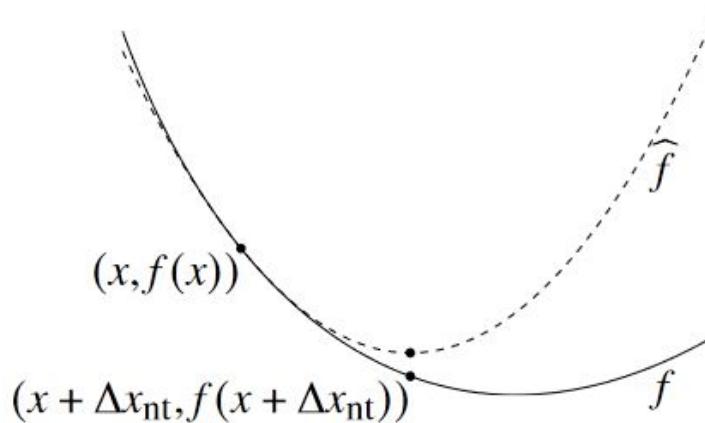


- ▶ steepest descent with backtracking line search for two quadratic norms
- ▶ ellipses show  $\{x \mid \|x - x^{(k)}\|_P = 1\}$
- ▶ interpretation of steepest descent with quadratic norm  $\|\cdot\|_P$ : gradient descent after change of variables  $\bar{x} = P^{1/2}x$
- ▶ shows choice of  $P$  has strong effect on speed of convergence

# The gold standard: Newton's method

- ▶ **Newton step** is  $\Delta x_{nt} = -\nabla^2 f(x)^{-1} \nabla f(x)$
- ▶ **interpretation:**  $x + \Delta x_{nt}$  minimizes second order approximation

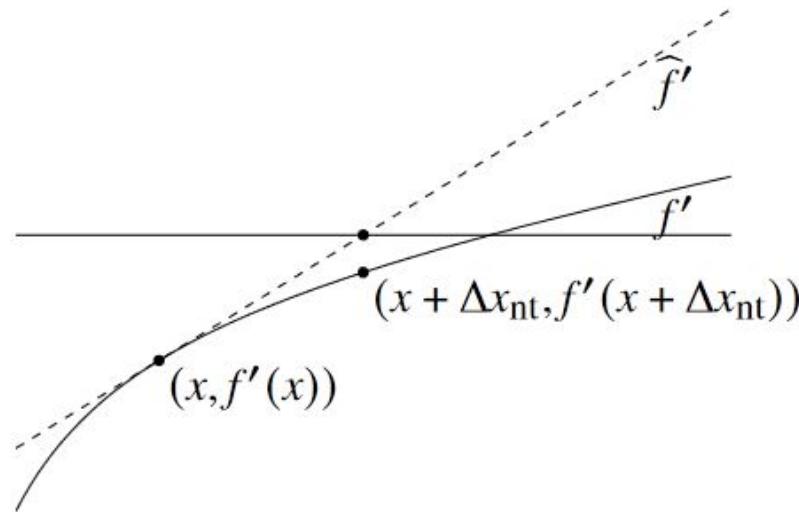
$$\hat{f}(x + v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v$$



# Linearized interpretation

- $x + \Delta x_{\text{nt}}$  solves linearized optimality condition

$$\nabla f(x + v) \approx \widehat{\nabla f}(x + v) = \nabla f(x) + \nabla^2 f(x)v = 0$$



# Hessian norm interpretation

- $\Delta x_{\text{nt}}$  is steepest descent direction at  $x$  in local Hessian norm  $\|u\|_{\nabla^2 f(x)} = (u^T \nabla^2 f(x) u)^{1/2}$



- dashed lines are contour lines of  $f$ ; ellipse is  $\{x + v \mid v^T \nabla^2 f(x) v = 1\}$
- arrow shows  $-\nabla f(x)$



# Newton decrement

- ▶ **Newton decrement** is  $\lambda(x) = (\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x))^{1/2}$
- ▶ a measure of the proximity of  $x$  to  $x^*$
- ▶ gives an estimate of  $f(x) - p^*$ , using quadratic approximation  $\widehat{f}$ :

$$f(x) - \inf_y \widehat{f}(y) = \frac{1}{2} \lambda(x)^2$$

- ▶ equal to the norm of the Newton step in the quadratic Hessian norm

$$\lambda(x) = (\Delta x_{\text{nt}}^T \nabla^2 f(x) \Delta x_{\text{nt}})^{1/2}$$

- ▶ directional derivative in the Newton direction:  $\nabla f(x)^T \Delta x_{\text{nt}} = -\lambda(x)^2$
- ▶ affine invariant (unlike  $\|\nabla f(x)\|_2$ )

# Algorithm

---

given a starting point  $x \in \text{dom } f$ , tolerance  $\epsilon > 0$ .

repeat

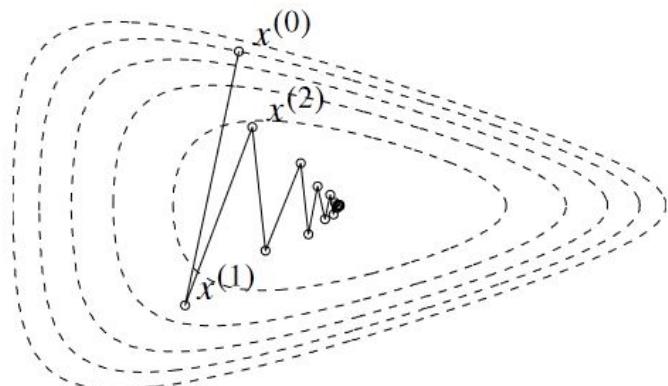
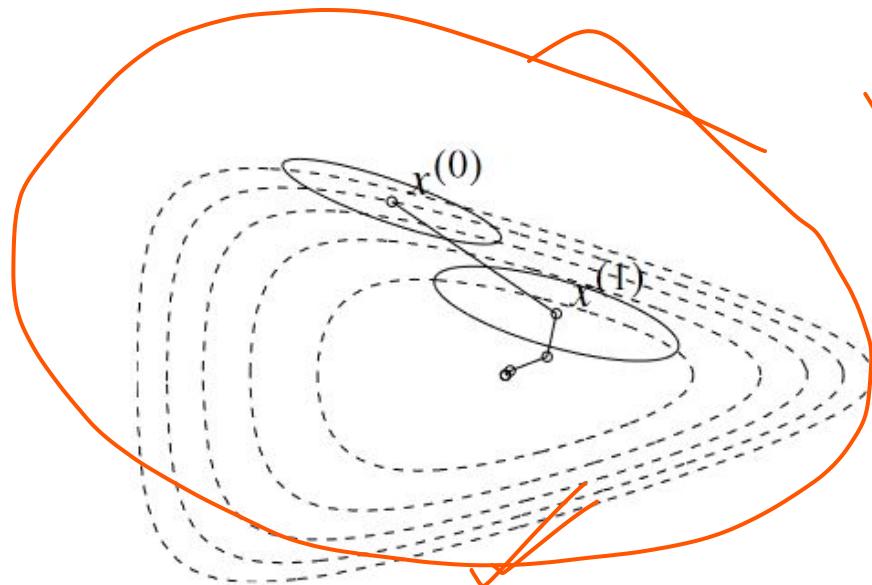
1. Compute the Newton step and decrement.

$$\Delta x_{\text{nt}} := -\nabla^2 f(x)^{-1} \nabla f(x); \quad \lambda^2 := \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x).$$

2. Stopping criterion. quit if  $\lambda^2/2 \leq \epsilon$ .
  3. Line search. Choose step size  $t$  by backtracking line search.
  4. Update.  $x := x + t \Delta x_{\text{nt}}$ .
- 

- ▶ **affine invariant**, i.e., independent of linear changes of coordinates
- ▶ Newton iterates for  $\tilde{f}(y) = f(Ty)$  with starting point  $y^{(0)} = T^{-1}x^{(0)}$  are  $y^{(k)} = T^{-1}x^{(k)}$

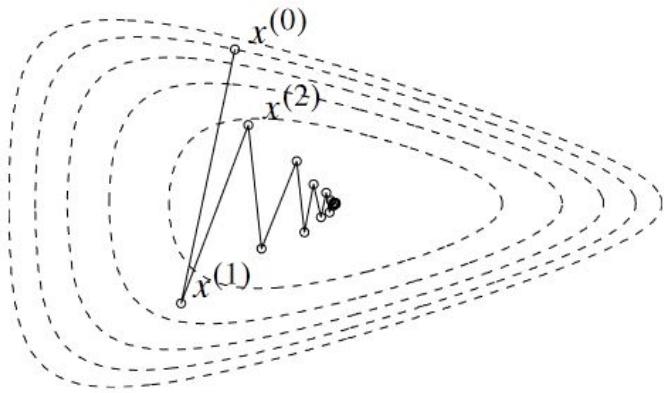
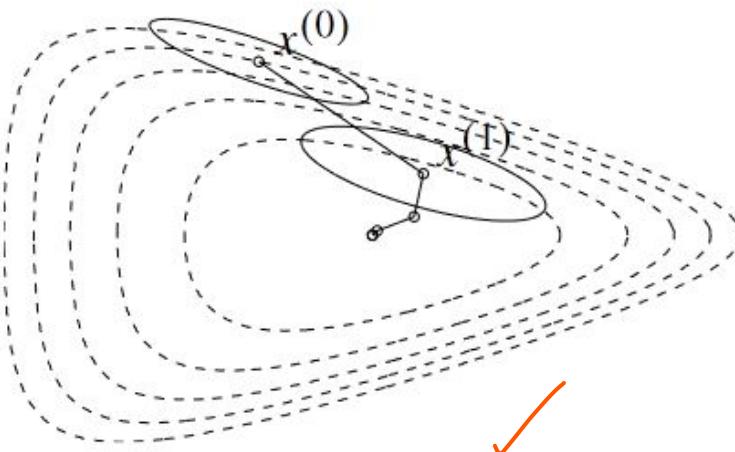
# Backtrack analysis



backtracking line search

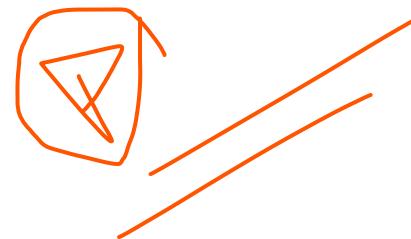
- ▶ backtracking parameters  $\alpha = 0.1, \beta = 0.7$
- ▶ converges in only 5 steps
- ▶ quadratic local convergence

# Backtrack analysis



backtracking line search

- ▶ backtracking parameters  $\alpha = 0.1, \beta = 0.7$
- ▶ converges in only 5 steps
- ▶ quadratic local convergence



## 2. Equality constrained problems



# Equality constrained problems

- ▶ equality constrained smooth minimization problem:

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && Ax = b \end{aligned}$$

- ▶ we assume
  - $f$  convex, twice continuously differentiable
  - $A \in \mathbf{R}^{p \times n}$  with  $\text{rank } A = p$
  - $p^*$  is finite and attained
- ▶ **optimality conditions:**  $x^*$  is optimal if and only if there exists a  $v^*$  such that

$$\nabla f(x^*) + A^T v^* = 0, \quad Ax^* = b$$

# KKT to the rescue

- ▶  $f(x) = (1/2)x^T Px + q^T x + r$ ,  $P \in \mathbf{S}_+^n$
- ▶  $\nabla f(x) = Px + q$
- ▶ optimality conditions are a **system of linear equations**

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ v^\star \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

- ▶ coefficient matrix is called KKT matrix
- ▶ KKT matrix is nonsingular if and only if

$$Ax = 0, \quad x \neq 0 \quad \implies \quad x^T Px > 0$$

- ▶ equivalent condition for nonsingularity:  $P + A^T A > 0$

- ▶ represent feasible set  $\{x \mid Ax = b\}$  as  $\{Fz + \hat{x} \mid z \in \mathbf{R}^{n-p}\}$ 
  - $\hat{x}$  is (any) **particular solution** of  $Ax = b$
  - range of  $F \in \mathbf{R}^{n \times (n-p)}$  is nullspace of  $A$  (**rank**  $F = n - p$  and  $AF = 0$ )
- ▶ **reduced or eliminated problem:** minimize  $f(Fz + \hat{x})$
- ▶ an unconstrained problem with variable  $z \in \mathbf{R}^{n-p}$
- ▶ from solution  $z^*$ , obtain  $x^*$  and  $v^*$  as

$$x^* = Fz^* + \hat{x}, \quad v^* = -(AA^T)^{-1}A\nabla f(x^*)$$

### 3. Inequality constrained problems (and interior point methods)

# Inequality constrained problems

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && Ax = b \end{aligned}$$


we assume

- ▶  $f_i$  convex, twice continuously differentiable
- ▶  $A \in \mathbf{R}^{p \times n}$  with  $\text{rank } A = p$
- ▶  $p^*$  is finite and attained
- ▶ problem is strictly feasible: there exists  $\tilde{x}$  with

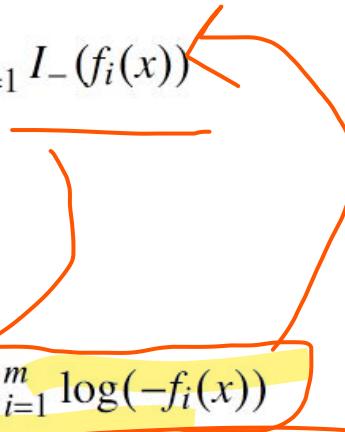
$$\tilde{x} \in \mathbf{dom} f_0, \quad f_i(\tilde{x}) < 0, \quad i = 1, \dots, m, \quad A\tilde{x} = b$$

hence, strong duality holds and dual optimum is attained

# Remember this?

- reformulation via **indicator function**:

$$\begin{aligned} & \text{minimize} && f_0(x) + \sum_{i=1}^m I_-(f_i(x)) \\ & \text{subject to} && Ax = b \end{aligned}$$



where  $I_-(u) = 0$  if  $u \leq 0$ ,  $I_-(u) = \infty$  otherwise

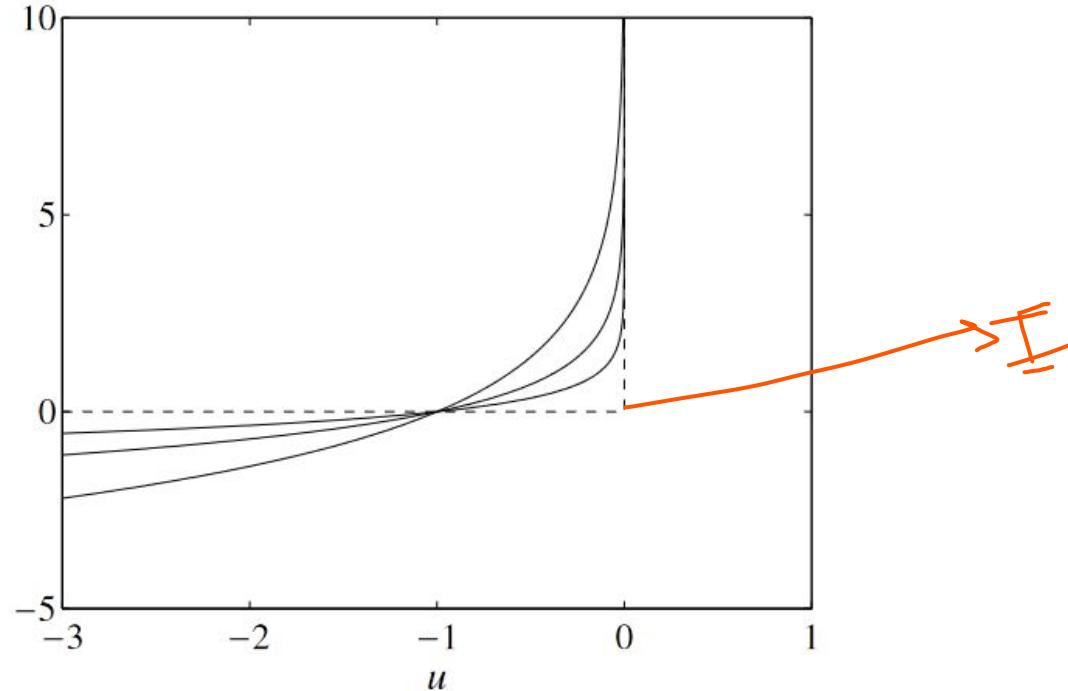
- **approximation via logarithmic barrier:**

$$\begin{aligned} & \text{minimize} && f_0(x) - (1/t) \sum_{i=1}^m \log(-f_i(x)) \\ & \text{subject to} && Ax = b \end{aligned}$$

- an equality constrained problem
- for  $t > 0$ ,  $-(1/t) \log(-u)$  is a smooth approximation of  $I_-$
- approximation improves as  $t \rightarrow \infty$

# Logarithmic barrier

►  $-(1/t) \log u$  for three values of  $t$ , and  $I_-(u)$



# Logarithmic barrier

- ▶ log barrier function for constraints  $f_1(x) \leq 0, \dots, f_m(x) \leq 0$

$$\phi(x) = - \sum_{i=1}^m \log(-f_i(x)), \quad \mathbf{dom} \phi = \{x \mid f_1(x) < 0, \dots, f_m(x) < 0\}$$

- ▶ convex (from composition rules)
- ▶ twice continuously differentiable, with derivatives

$$\nabla \phi(x) = \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x)$$

$$\nabla^2 \phi(x) = \sum_{i=1}^m \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla^2 f_i(x)$$

# Central path

- ▶ for  $t > 0$ , define  $x^*(t)$  as the solution of

$$\begin{array}{ll} \text{minimize} & tf_0(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

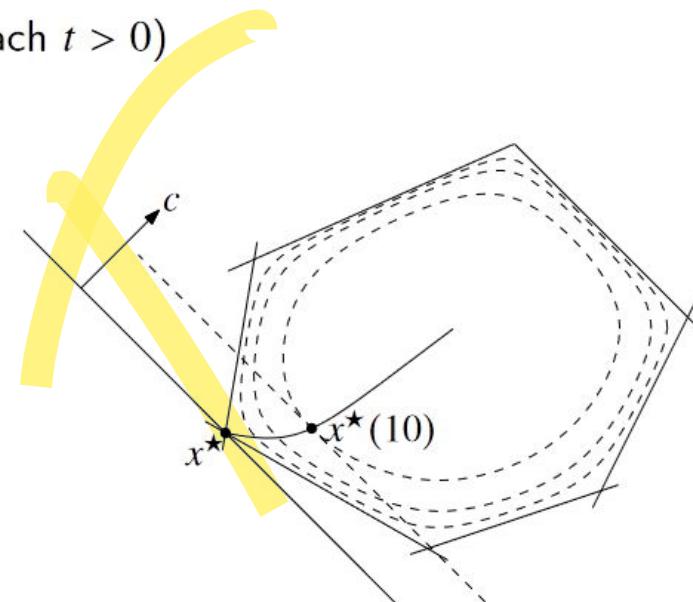
(for now, assume  $x^*(t)$  exists and is unique for each  $t > 0$ )

- ▶ central path is  $\{x^*(t) \mid t > 0\}$

**example:** central path for an LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \dots, 6 \end{array}$$

hyperplane  $c^T x = c^T x^*(t)$  is tangent to level curve of  $\phi$  through  $x^*(t)$



# Duality and the central path

- ▶  $x = x^\star(t)$  if there exists a  $w$  such that

$$t \nabla f_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x) + A^T w = 0, \quad Ax = b$$

- ▶ therefore,  $x^\star(t)$  minimizes the Lagrangian

$$L(x, \lambda^\star(t), v^\star(t)) = f_0(x) + \sum_{i=1}^m \lambda_i^\star(t) f_i(x) + v^\star(t)^T (Ax - b)$$

where we define  $\lambda_i^\star(t) = 1/(-tf_i(x^\star(t)))$  and  $v^\star(t) = w/t$

- ▶ this confirms the intuitive idea that  $f_0(x^\star(t)) \rightarrow p^\star$  if  $t \rightarrow \infty$ :

$$p^\star \geq g(\lambda^\star(t), v^\star(t)) = L(x^\star(t), \lambda^\star(t), v^\star(t)) = f_0(x^\star(t)) - m/t$$

# Approximated KKT

$x = x^\star(t)$ ,  $\lambda = \lambda^\star(t)$ ,  $\nu = \nu^\star(t)$  satisfy

1. primal constraints:  $f_i(x) \leq 0$ ,  $i = 1, \dots, m$ ,  $Ax = b$
2. dual constraints:  $\lambda \geq 0$
3. approximate complementary slackness:  $-\lambda_i f_i(x) = 1/t$ ,  $i = 1, \dots, m$
4. gradient of Lagrangian with respect to  $x$  vanishes:

$$\nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + A^T \nu = 0$$

difference with KKT is that condition 3 replaces  $\lambda_i f_i(x) = 0$

- ▶ **centering problem** (for problem with no equality constraints)

$$\text{minimize} \quad tf_0(x) - \sum_{i=1}^m \log(-f_i(x))$$

- ▶ **force field interpretation**

- $tf_0(x)$  is potential of force field  $F_0(x) = -t\nabla f_0(x)$

- $-\log(-f_i(x))$  is potential of force field  $F_i(x) = (1/f_i(x))\nabla f_i(x)$

- ▶ forces balance at  $x^\star(t)$ :

$$F_0(x^\star(t)) + \sum_{i=1}^m F_i(x^\star(t)) = 0$$

# Barrier method

given strictly feasible  $x$ ,  $t := t^{(0)} > 0$ ,  $\mu > 1$ , tolerance  $\epsilon > 0$ .

repeat

1. *Centering step.* Compute  $x^*(t)$  by minimizing  $tf_0 + \phi$ , subject to  $Ax = b$ .
2. *Update.*  $x := x^*(t)$ .
3. *Stopping criterion.* **quit** if  $m/t < \epsilon$ .
4. *Increase  $t$ .*  $t := \mu t$ .

- ▶ terminates with  $f_0(x) - p^* \leq \epsilon$  (stopping criterion follows from  $f_0(x^*(t)) - p^* \leq m/t$ )
- ▶ centering usually done using Newton's method, starting at current  $x$
- ▶ choice of  $\mu$  involves a trade-off: large  $\mu$  means fewer outer iterations, more inner (Newton) iterations; typical values:  $\mu = 10$  or  $20$
- ▶ several heuristics for choice of  $t^{(0)}$

# Phase I: feasibility

- ▶ barrier method needs strictly feasible starting point, i.e.,  $x$  with

$$f_i(x) < 0, \quad i = 1, \dots, m, \quad Ax = b$$

- ▶ (like the infeasible start Newton method, more sophisticated interior-point methods do not require a feasible starting point)
- ▶ **phase I** method forms an optimization problem that
  - is itself strictly feasible
  - finds a strictly feasible point for original problem, if one exists
  - certifies original problem as infeasible otherwise
- ▶ **phase II** uses barrier method starting from strictly feasible point found in phase I

# Phase I: feasibility

- ▶ introduce slack variable  $s$  in **phase I problem**

$$\begin{aligned} & \text{minimize (over } x, s) && s \\ & \text{subject to} && f_i(x) \leq s, \quad i = 1, \dots, m \\ & && Ax = b \end{aligned}$$

with optimal value  $\bar{p}^*$

- if  $\bar{p}^* < 0$ , original inequalities are strictly feasible
- if  $\bar{p}^* > 0$ , original inequalities are infeasible
- $\bar{p}^* = 0$  is an ambiguous case

- ▶ start phase I problem with
  - any  $\tilde{x}$  in problem domain with  $A\tilde{x} = b$
  - $s = 1 + \max_i f_i(\tilde{x})$

# Phase I: feasibility

- minimize **sum** of slacks, not max:

$$\begin{aligned} & \text{minimize} && \mathbf{1}^T s \\ & \text{subject to} && s \geq 0, \quad f_i(x) \leq s_i, \quad i = 1, \dots, m \\ & && Ax = b \end{aligned}$$

- will find a strictly feasible point if one exists
- for infeasible problems, produces a solution that satisfies many (but not all) inequalities
- can weight slacks to set **priorities** (in satisfying constraints)