

Optimization: Convex Optimization

Convex Optimization

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

with f and g convex $\rightarrow$ **local minimum = global minimum**

Convex function:

For all $x, y \in \text{dom}(f)$ and $\lambda \in [0, 1]$:

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

Convex functions:

$$f(x) = ax \quad a \in \mathbb{R}$$

$$f(x) = x^{2n} \quad n \in \mathbb{N} \setminus \{0\}$$

$$f(x) = e^x$$

$$f(x) = h(\beta_0 + \beta^T x) \quad h \text{ is convex}$$

$$f(x) = \|x\| \quad \text{for any norm function}$$

Norm function $\| \cdot \|$ on $\mathbb{R}^n$:

- ① $\|u\| \geq 0$
- ② $\|u\| = 0 \Leftrightarrow u = 0$
- ③ $\|\alpha u\| = |\alpha| \|u\|$ for all $\alpha \in \mathbb{R}$
- ④ $\|u + v\| \leq \|u\| + \|v\|$ for all $u, v \in \mathbb{R}^n$

Operations that preserve convexity:

Positive-weighted sum: $f(x) = \sum \alpha_i f_i(x) \quad \alpha_i > 0$

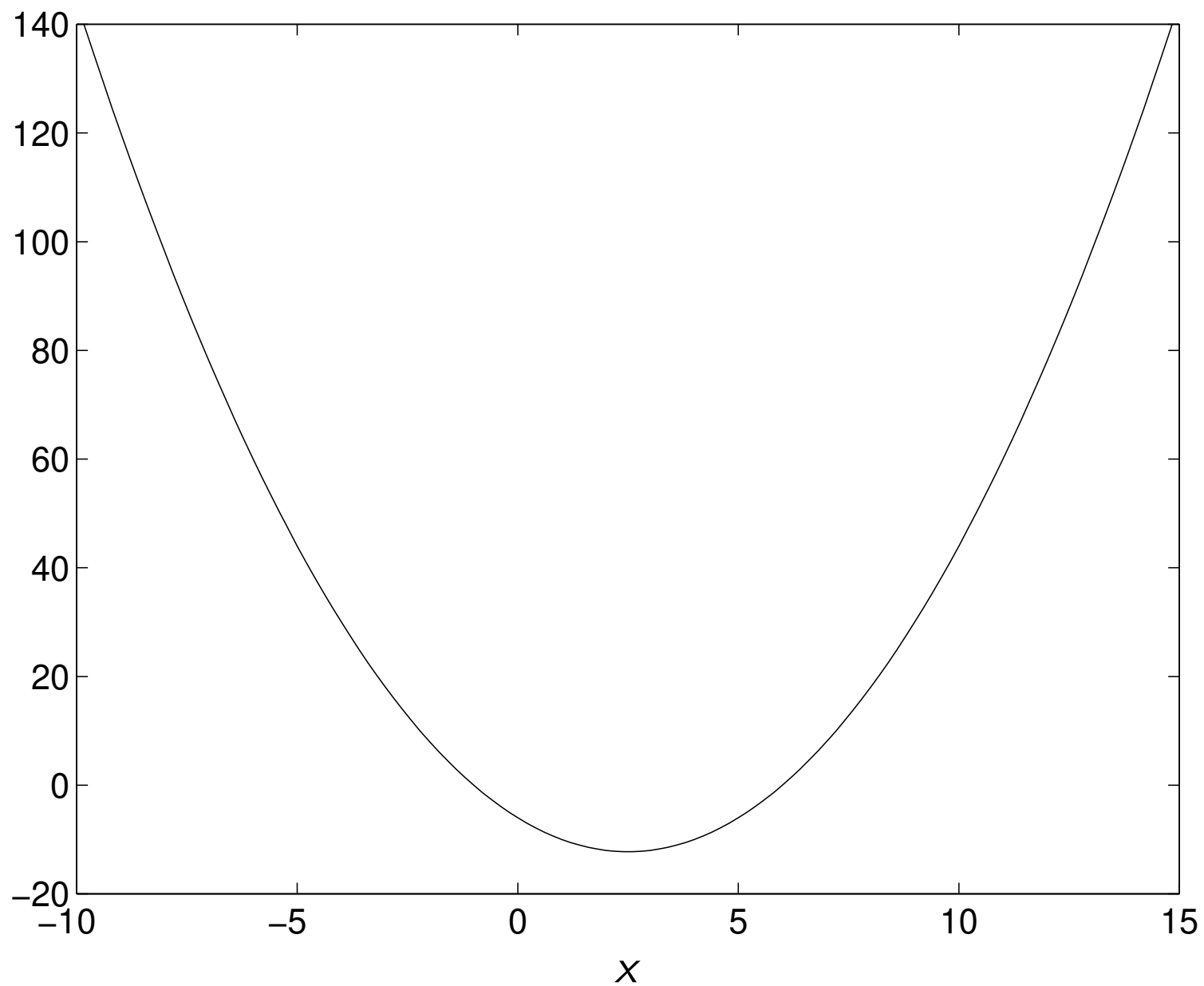
Pointwise maximum: $f(x) = \max f_i(x)$

Composition: $f(x) = h(g(x))$

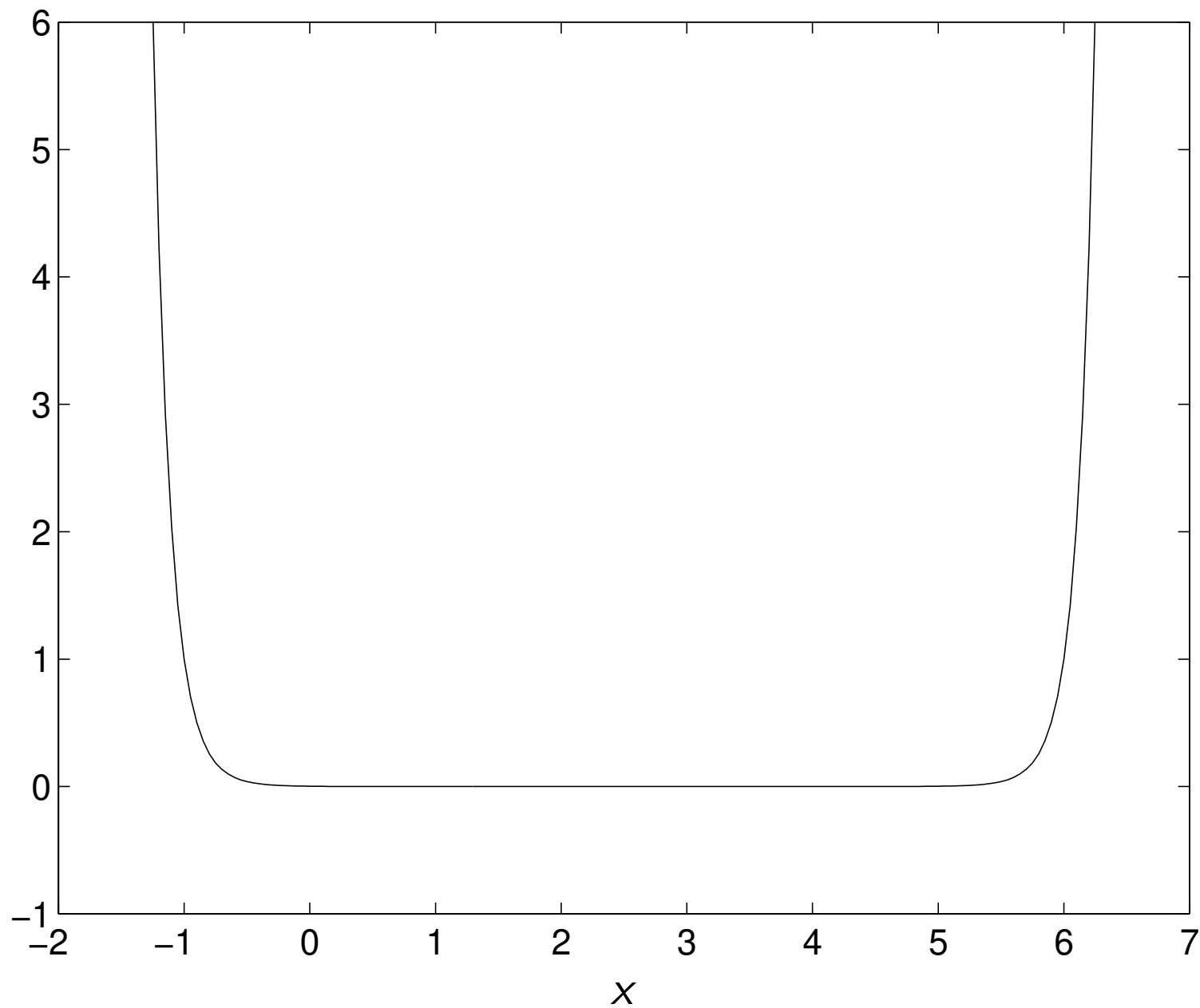
if g, h convex and $h \nearrow$

if $-g, h$ convex and $h \searrow$

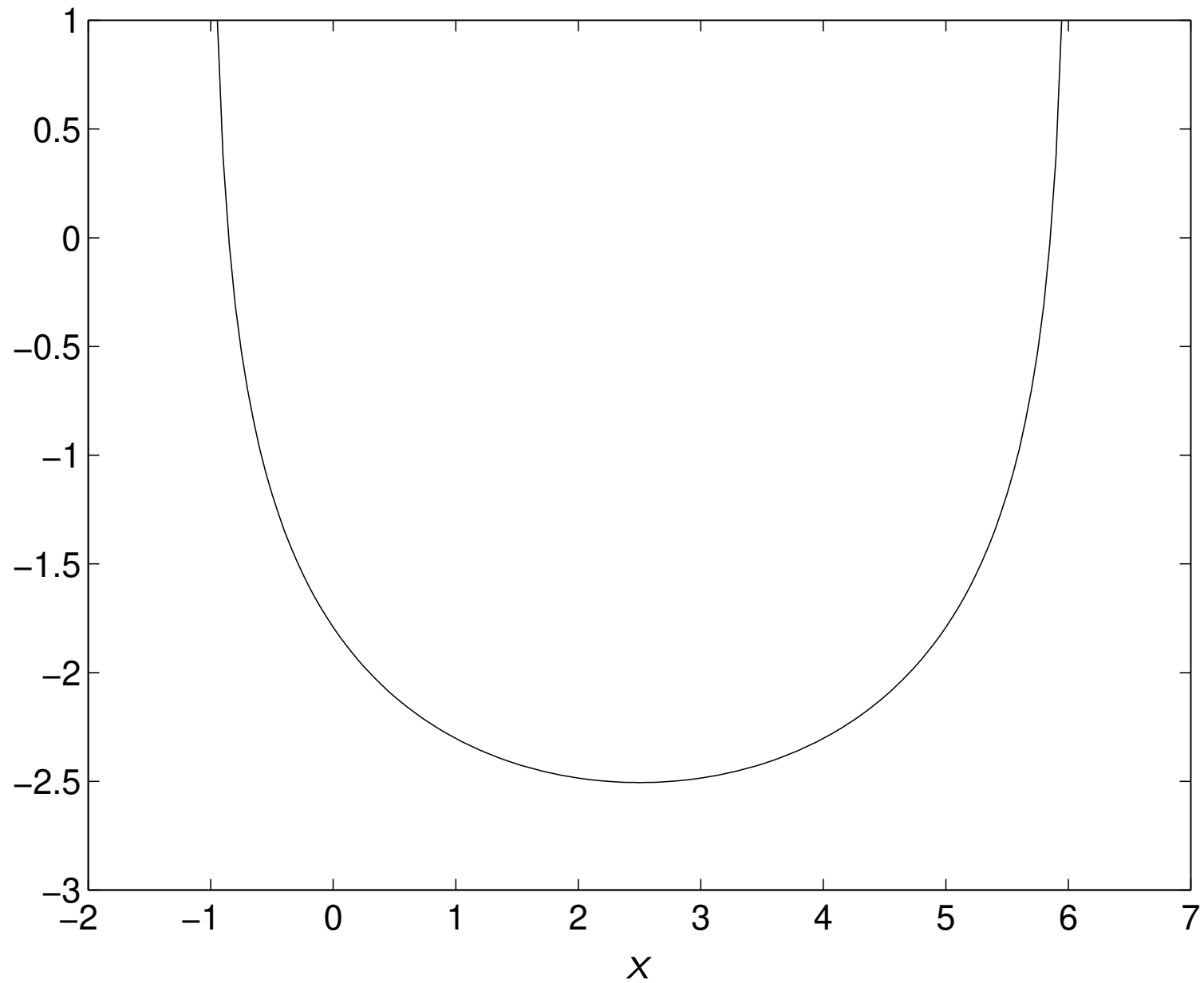
$$f(x) = -6 + (-5x) + x^2$$



$$f(x) = e^{-6+(-5x)+x^2}$$



$$f(x) = -\log(6 + 5x - x^2)$$

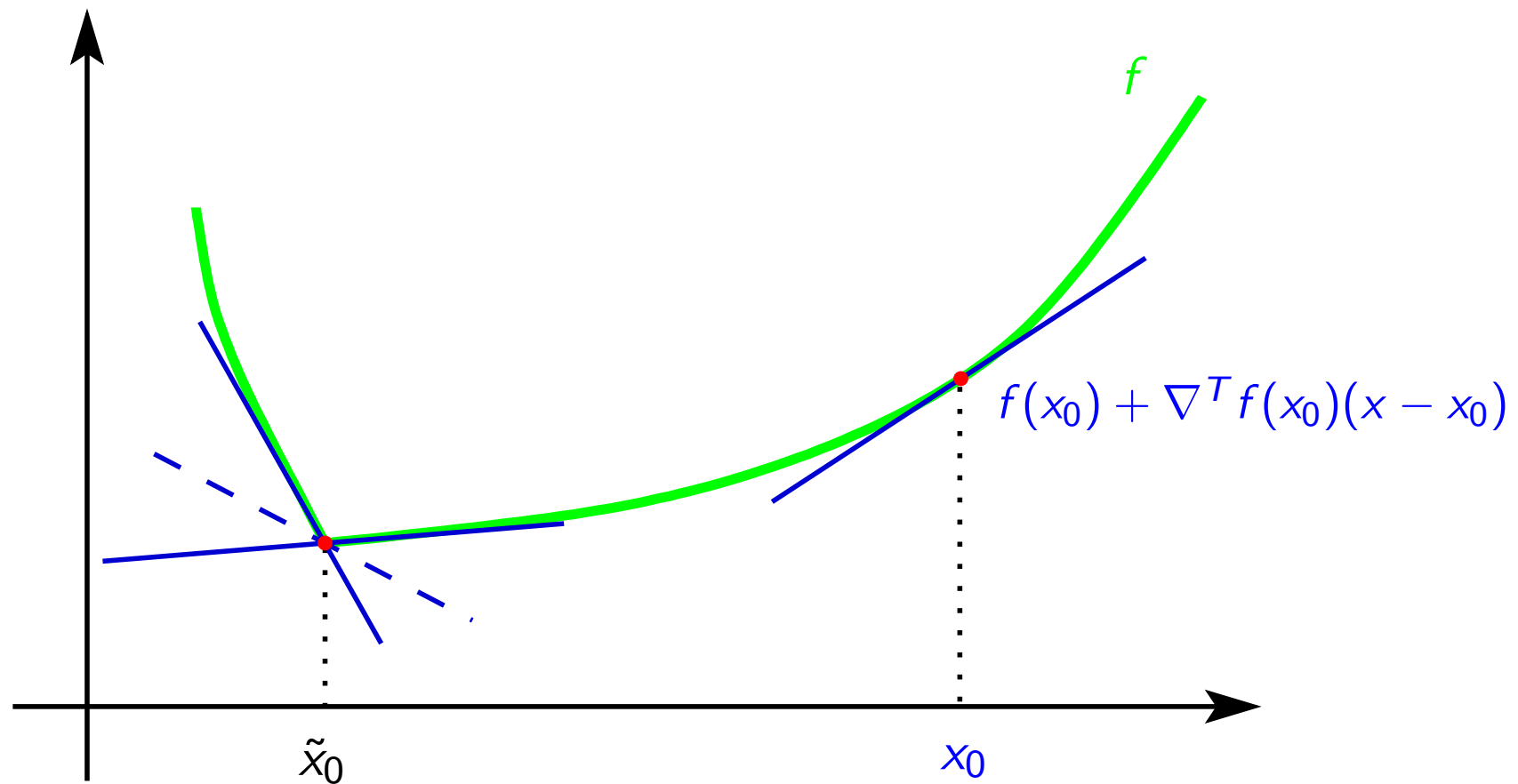


Test: Convex functions

Are the following functions convex? ($x, y \in \mathbb{R}$):

- $\cos(0.5x + y)$
- $\cosh(x^2 + y^2)$
- $|x + y - 5|$
- $-\log |x + y - 5|$

Subgradient of a **convex** function



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

Norms of affine functions

F_0 : scalar signal, F_1 : vector-valued signal

$$f_1(x) = \|F_0(n) + F_1^T(n)x\|_1 = \sum_n |F_0(n) + F_1^T(n)x|$$

$$f_2(x) = \|F_0(n) + F_1^T(n)x\|_2^2 = \sum_n (F_0(n) + F_1^T(n)x)^2$$

$$f_\infty(x) = \|F_0(n) + F_1^T(n)x\|_\infty = \max_n |F_0(n) + F_1^T(n)x|$$

Subgradients:

$$\nabla f_1(x) = \sum_n F_1(n) \operatorname{sign}(F_0(n) + F_1^T(n)x)$$

$$\nabla f_2(x) = \sum_n 2 F_1(n) (F_0(n) + F_1^T(n)x)$$

$$\nabla f_\infty(x) = F_1(n_{\max}) \operatorname{sign}(F_0(n_{\max}) + F_1^T(n_{\max})x)$$

$$\text{with } n_{\max} = \arg \max_n |F_0(n) + F_1^T(n)x|$$

Linear matrix inequalities

Linear matrix inequality

$$F(x) := F_0 + \sum_i F_i x_i > 0$$

where $F_i = F_i^T$

- $\{x \mid F(x) > 0\}$ is **convex** set
- Multiple LMIs $\Rightarrow$ one LMI

$$F^{(1)}(x) > 0, F^{(2)}(x) > 0, \dots, F^{(p)}(x) > 0$$

$$\Leftrightarrow \begin{bmatrix} F^{(1)}(x) & 0 & \dots & 0 \\ 0 & F^{(2)}(x) & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & F^{(p)}(x) \end{bmatrix} > 0$$

Linear matrix inequalities (continued)

Linear matrix inequality

$$F(x) := F_0 + \sum_i F_i x_i > 0$$

where $F_i = F_i^T$

- Schur complement:

$$\begin{bmatrix} Q(x) & S(x) \\ S^T(x) & R(x) \end{bmatrix} > 0$$

$$\Leftrightarrow R(x) > 0, \quad Q(x) - S(x)R^{-1}(x)S^T(x) > 0$$

$$\Leftrightarrow Q(x) > 0, \quad R(x) - S^T(x)Q^{-1}(x)S(x) > 0$$

Lyapunov theory

$$\frac{d}{dt}x(t) = Ax(t) \quad \text{is stable}$$

is equivalent to

$$A^T P + PA < 0, \quad P > 0$$

Stabilizing state feedback $u(t) = Kx(t)$:

$$\frac{d}{dt}x(t) = Ax(t) + Bu(t) = Ax(t) + BKx(t)$$

is equivalent to

$$(A^T + K^T B^T)P + P(A + BK) < 0, \quad P > 0$$

$$P^{-1}((A^T + K^T B^T)P + P(A + BK))P^{-1} < 0, \quad P > 0$$

Defining $X = P^{-1}$ and $Y = KP^{-1}$ yields

$$XA^T + Y^T B^T + AX + BY < 0, \quad X > 0$$

Quadratic objective:

$$\frac{d}{dt}x(t) = Ax(t) + Bu(t) = Ax(t) + BKx(t)$$

Given: $x(0) = x_0$. Find state feedback $u(t) = Kx(t)$

$$\min_K \underbrace{\int_0^\infty (x^T Qx + u^T Ru) dt}_{J(K)} = \min_K \int_0^\infty x^T (Q + K^T RK) x dt$$

Find $V(z) = z^T Pz$ with $P > 0$ such that

$$\frac{d}{dt}V(x) < -x^T (Q + K^T RK) x$$

Note: if $V(x(\infty)) = 0$, then $V(x(0)) > J(K)$

$$\begin{aligned} \frac{d}{dt}V(x) &= \frac{d}{dt}(x^T Px) = \frac{dx^T}{dt}Px + x^T P \frac{dx}{dt} \\ &= x^T ((A + BK)^T P + P(A + BK))x \end{aligned}$$

Quadratic objective (continued):

Hence, $(A + BK)^T P + P(A + BK) + Q + K^T R K < 0$

Setting $P = \gamma X^{-1}$ and $K = YX^{-1}$ yields

$$XA^T + Y^T B^T + AX + BY + \gamma^{-1} X Q X + \gamma^{-1} Y^T R Y < 0$$

Schur complement transformation:

$$\begin{bmatrix} -(XA^T + Y^T B^T + AX + BY) & XQ^{1/2} & Y^T R^{1/2} \\ Q^{1/2} X & \gamma I & 0 \\ R^{1/2} Y & 0 & \gamma I \end{bmatrix} > 0$$

$$\gamma > 0, \quad X > 0$$

For any solution X, Y, γ : $V(x_0) = x_0^T P x_0 = \gamma x_0^T X^{-1} x_0 > J(K)$

Extra constraint: $x_0^T X^{-1} x_0 < 1$ or equivalently

$$\begin{bmatrix} 1 & x_0^T \\ x_0 & X \end{bmatrix} > 0$$

$\Rightarrow \min_{X, Y, \gamma} \gamma$ subject to LMIs

Some common convex problems with LMIs

- **Entropy maximization:**

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & A(x) > 0 \end{array}$$

- **Determinant maximization:**

$$\begin{array}{ll} \text{minimize} & \log \det(A(x)^{-1}) \\ \text{subject to} & A(x) > 0, \quad B(x) > 0 \end{array}$$

- **Generalized eigenvalue problem:**

$$\begin{array}{ll} \text{minimize} & \lambda \\ \text{subject to} & \lambda B(x) - A(x) > 0, \quad B(x) > 0 \end{array}$$

Convex optimization algorithms

- Cutting-plane algorithm
- Ellipsoid algorithm
- Interior-point algorithm

Advantages

- Easy to implement
- Guaranteed and fast convergence
- Provide stopping criteria with “hard guarantee”:

$$|f(x^*) - f(x)| \leq \varepsilon_f$$

$$\|x^* - x\|_2 \leq \varepsilon_x \quad (\text{for ellipsoid})$$

where x^* is the **real** optimum

Cutting-plane algorithm

Given: k points $x_{(1)}, \dots, x_{(k)}$

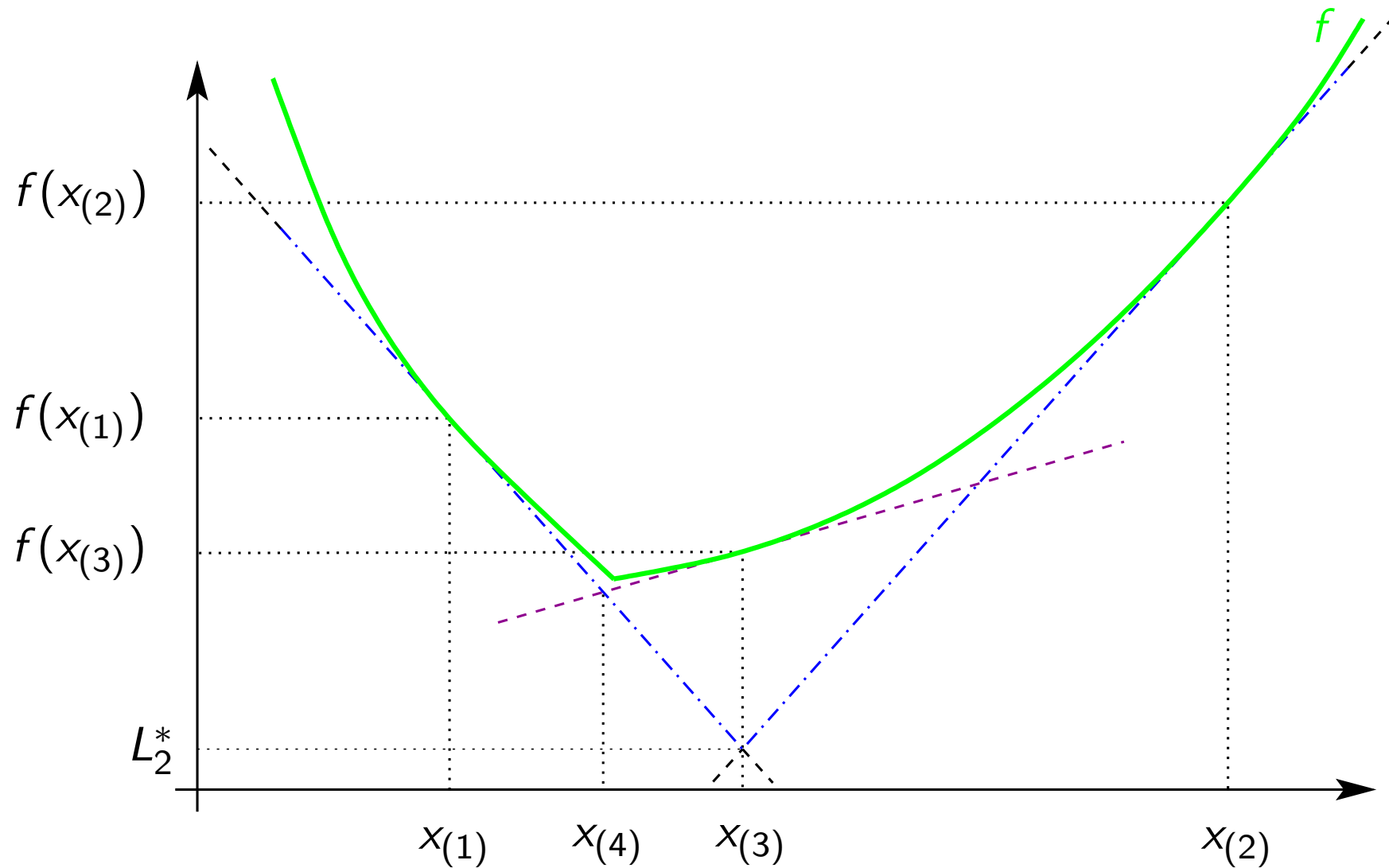
$$f(x) \geq f(x_{(i)}) + \nabla^T f(x_{(i)})(x - x_{(i)})$$

$$\Rightarrow f(x) \geq \max_{i=1,\dots,k} \left(f(x_{(i)}) + \nabla^T f(x_{(i)})(x - x_{(i)}) \right)$$

Hence,

$$f(x^*) \geq \min_x \max_{i=1,\dots,k} \left(f(x_{(i)}) + \nabla^T f(x_{(i)})(x - x_{(i)}) \right)$$

Cutting-plane algorithm (continued)



Cutting-plane algorithm (continued)

$$f(x^*) \geq \min_x \max_{i=1,\dots,k} \left(f(x_{(i)}) + \nabla^T f(x_{(i)})(x - x_{(i)}) \right)$$

Solve: $\min_{x, L_k} L_k$

$$\text{s.t. } f(x_{(i)}) + \nabla^T f(x_{(i)})(x - x_{(i)}) \leq L_k$$

= linear program $\rightarrow x_k^*, L_k^*$

Define: $U_k^* = \min_{i=1,\dots,k} f(x_{(i)})$

$$\Rightarrow L_k^* \leq f(x_k^*) \leq U_k^*$$

Define $x_{(k+1)} = x_k^*$ and repeat procedure

$$\Rightarrow L_k^* \leq L_{k+1}^* \leq f(x^*) \leq U_{k+1}^* \leq U_k^*$$

Iterate until $U_k^* - L_k^* \leq \varepsilon_f$

Cutting-plane: Handling constraints

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

Since g is convex:

$$g(x) \geq g(x_{(i)}) + \nabla^T g(x_{(i)})(x - x_{(i)})$$

So if $g(x) \leq 0$, then certainly $g(x_{(i)}) + \nabla^T g(x_{(i)})(x - x_{(i)}) \leq 0$, i.e.

$$\{g(x) \leq 0\} \subseteq \left\{g(x_{(i)}) + \nabla^T g(x_{(i)})(x - x_{(i)}) \leq 0\right\}$$

Add extra constraints to LP:

$$\nabla^T g(x_{(i)}) x \leq \nabla^T g(x_{(i)}) x_{(i)} - g(x_{(i)})$$

→ still LP!

Larger feasible region considered → resulting L_k^* is still lower bound for $f(x^*)$

Ellipsoid algorithm

One-dimensional case:

Suppose x^* in interval E_0 with center $x_{(0)}$:

$$E_0 = \{x \mid x_{(0)} - A_0 \leq x \leq x_{(0)} + A_0\}$$

Recall that

$$f(x) \geq f(x_{(0)}) + \nabla^T f(x_{(0)}) (x - x_{(0)})$$

Hence, x^* will be in the half-plane

$$H_0 = \{x \mid \nabla^T f(x_{(0)}) (x - x_{(0)}) \leq 0\}$$

Construct

$$\begin{aligned} E_1 &= H_0 \cap E_0 \\ &= \{x \mid x_{(1)} - A_1 \leq x \leq x_{(1)} + A_1\} \end{aligned}$$

$$\text{with } A_1 = \frac{A_0}{2}$$

$$x_{(1)} = x_{(0)} - A_1 \operatorname{sign}(\nabla f(x_{(0)}))$$

Ellipsoid algorithm

Multi-dimensional case:

Suppose x^* in ellipsoid E_0 with center $x_{(0)}$:

$$E_0 = \{x \mid (x - x_{(0)})^T A_0^{-1} (x - x_{(0)}) \leq 1\}$$

with $A_0 \in \mathbb{R}^{n \times n}$ is non-singular and positive definite.

Since

$$f(x) \geq f(x_{(0)}) + \nabla^T f(x_{(0)}) (x - x_{(0)})$$

x^* will be in half-plane

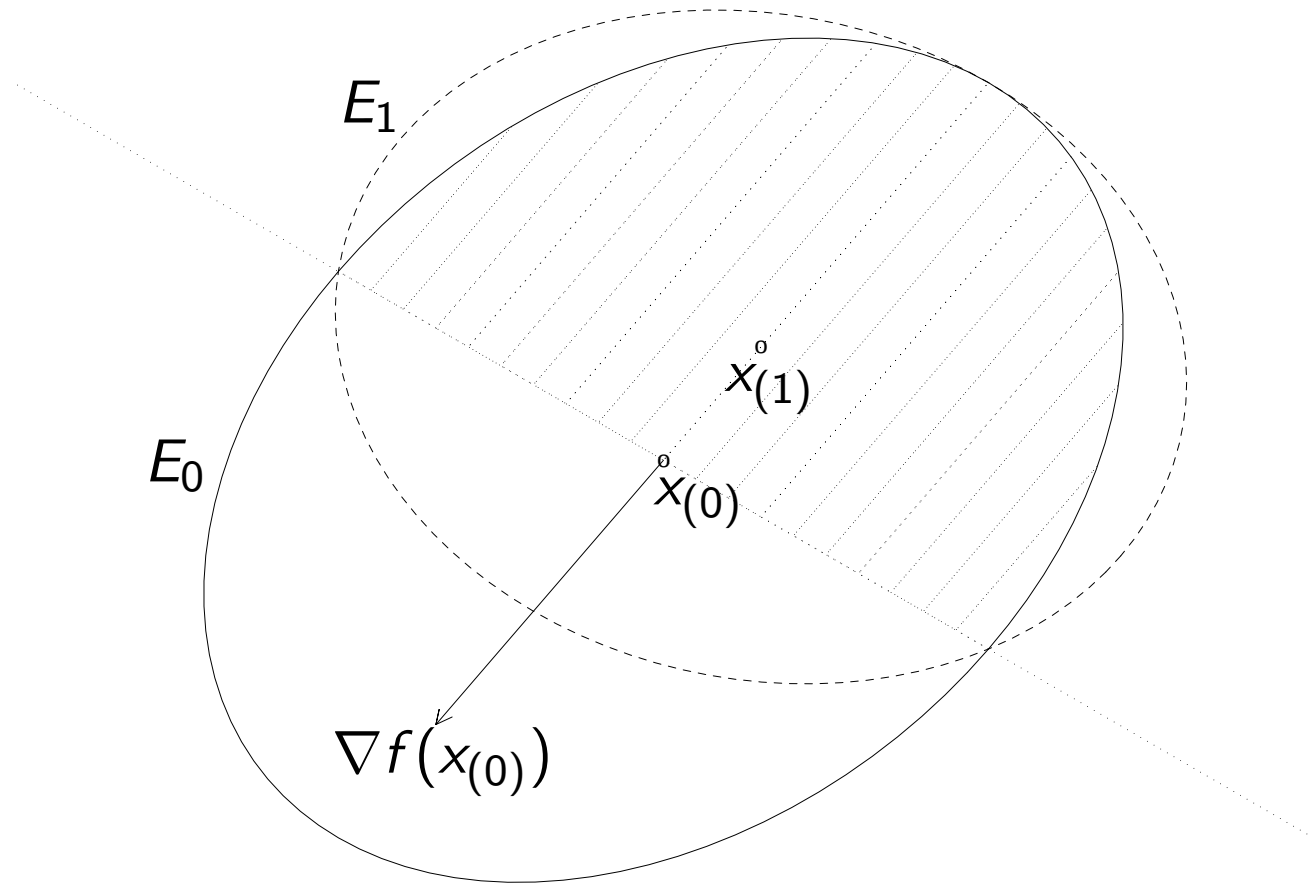
$$H_0 = \{x \mid \nabla^T f(x_{(0)}) (x - x_{(0)}) \leq 0\}$$

Construct new ellipsoid

$$E_1 = \{x \mid (x - x_{(1)})^T A_1^{-1} (x - x_{(1)}) \leq 1\}$$

such that $H_0 \cap E_0 \subseteq E_1$

Ellipsoid algorithm (continued)



$$f(x) \geq f(x_{(0)}) + \nabla^T f(x_{(0)}) (x - x_{(0)})$$

Ellipsoid algorithm (continued)

For $(k + 1)$ st ellipsoid E_{k+1} :

$$x_{(k+1)} = x_{(k)} - \frac{1}{n+1} \frac{A_k \nabla f(x_{(k)})}{\sqrt{\nabla^T f(x_{(k)}) A_k \nabla f(x_{(k)})}}$$

$$A_{k+1} = \frac{n^2}{n^2 - 1} \left(A_k - \frac{2}{n+1} \frac{A_k \nabla f(x_{(k)}) \nabla^T f(x_{(k)}) A_k^T}{\nabla^T f(x_{(k)}) A_k \nabla f(x_{(k)})} \right)$$

Properties:

$$\text{Vol}(E_k) \rightarrow 0 \text{ for } k \rightarrow \infty$$

$$f(x_{(k)}) - f(x^*) \leq \sqrt{\nabla^T f(x_{(k)}) A_k \nabla f(x_{(k)})}$$

Iterate until $\sqrt{\nabla^T f(x_{(k)}) A_k \nabla f(x_{(k)})} \leq \varepsilon_f$ and/or $\text{Vol}(E_k) \leq \varepsilon_x$

Ellipsoid algorithm: Handling constraints

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

g is convex $\Rightarrow g(x) \geq g(x_{(k)}) + \nabla^T g(x_{(k)}) (x - x_{(k)})$

So if $g(x_{(k)}) > 0$ then x^* will be in half-plane

$$H_k = \{x \mid \nabla^T g(x_{(k)}) (x - x_{(k)}) \leq 0\}$$

$\rightarrow$ replace ∇f by ∇g and use same formulas

Ellipsoid algorithm:

- If $x_{(k)}$ feasible $\rightarrow$ use ∇f in formulas for E_{k+1}
 $\rightarrow$ discard points that are not minimizers
“objective iteration”
- If $x_{(k)}$ not feasible $\rightarrow$ use ∇g in formulas for E_{k+1}
 $\rightarrow$ discard infeasible points
“constraint iteration”

Interior-point algorithm

$$f(x^*) = \min_x f(x) \quad \text{s.t.} \quad g(x) \leq 0$$

Strictly feasible set $\mathcal{G}$:

$$\mathcal{G} := \{ x \mid g_i(x) < 0, \ i = 1, \dots, m \}$$

Barrier function:

$$\phi(x) = \begin{cases} -\sum_{i=1}^m \log(-g_i(x)) & x \in \mathcal{G} \\ \infty & \text{otherwise} \end{cases}$$

→ convex

Optimization problem for $t \geq 0$: $\min_x \ t f(x) + \phi(x)$

Central path $x^*(t)$:

$$x^*(t) = \arg \min_x (t f(x) + \phi(x)) \quad \text{is always in } \mathcal{G}$$

$x^*(t)$ will converge to x^* for $t \rightarrow \infty$

Interior-point algorithm (continued)

We have $\Psi(x, t) = t f(x) + \phi(x)$. So

$$\nabla_x \Psi(x^*, t) = t \nabla f(x^*(t)) + \sum_{i=1}^m \frac{1}{-g_i(x^*(t))} \nabla g_i(x^*(t)) = 0$$

$$\Rightarrow \nabla f(x^*(t)) + \sum_{i=1}^m \mu_i^*(t) \nabla g_i(x^*(t)) = 0 \quad \text{with } \mu_i^*(t) := \frac{1}{-g_i(x^*(t)) t}$$

So $x^*(t)$ also minimizes $f(x) + \sum_i \mu_i^*(t) g_i(x)$ (*)

Moreover, we have: $g(x^*(t)) \leq 0$

$$\mu_i^*(t) \geq 0$$

$$\nabla f(x^*(t)) + \sum_{i=1}^m \mu_i^*(t) \nabla g_i(x^*(t)) = 0$$

$$\mu_i^*(t) g_i(x^*(t)) = -1/t$$

So Karush-Kuhn-Tucker conditions satisfied for $t \rightarrow \infty$

Interior-point algorithm (continued)

Dual function:

$$d(\mu^*(t)) = \min_x \left(f(x) + \sum_{i=1}^m \mu_i^*(t) g_i(x) \right)$$

Property: $d(\mu^*(t)) \leq f(x^*)$ for any $\mu^*(t) \geq 0$

Hence, $f(x^*) \geq d(\mu^*(t))$

$$\begin{aligned} &\geq \min_x \left(f(x) + \sum_{i=1}^m \mu_i^*(t) g_i(x) \right) \\ &\geq f(x^*(t)) + \sum_{i=1}^m \mu_i^*(t) g_i(x^*(t)) \quad (\text{by } (*)) \\ &\geq f(x^*(t)) - \frac{m}{t} \quad \text{since } \mu_i^*(t) = \frac{1}{-g_i(x^*(t)) t} \end{aligned}$$

This yields:

$$f(x^*(t)) \geq f(x^*) \geq f(x^*(t)) - \frac{m}{t}$$

Interior-point algorithm (continued)

$$f(x^*(t)) \geq f^* \geq f(x^*(t)) - \frac{m}{t}$$

Stopping criterion: $|f^* - f(x^*(t))| \leq \varepsilon_f$

Take $t = \frac{m}{\varepsilon_f} \rightarrow$ one unconstrained optimization problem

slow $\rightarrow$ gradually increase t

Sequential unconstrained minimization technique:

Given: $x \in \mathcal{G}$, $t > 0$ and tolerance ε_f

Step 1: Compute $x^*(t)$ starting from x :

$$x^*(t) = \arg \min_x (t f(x) + \phi(x))$$

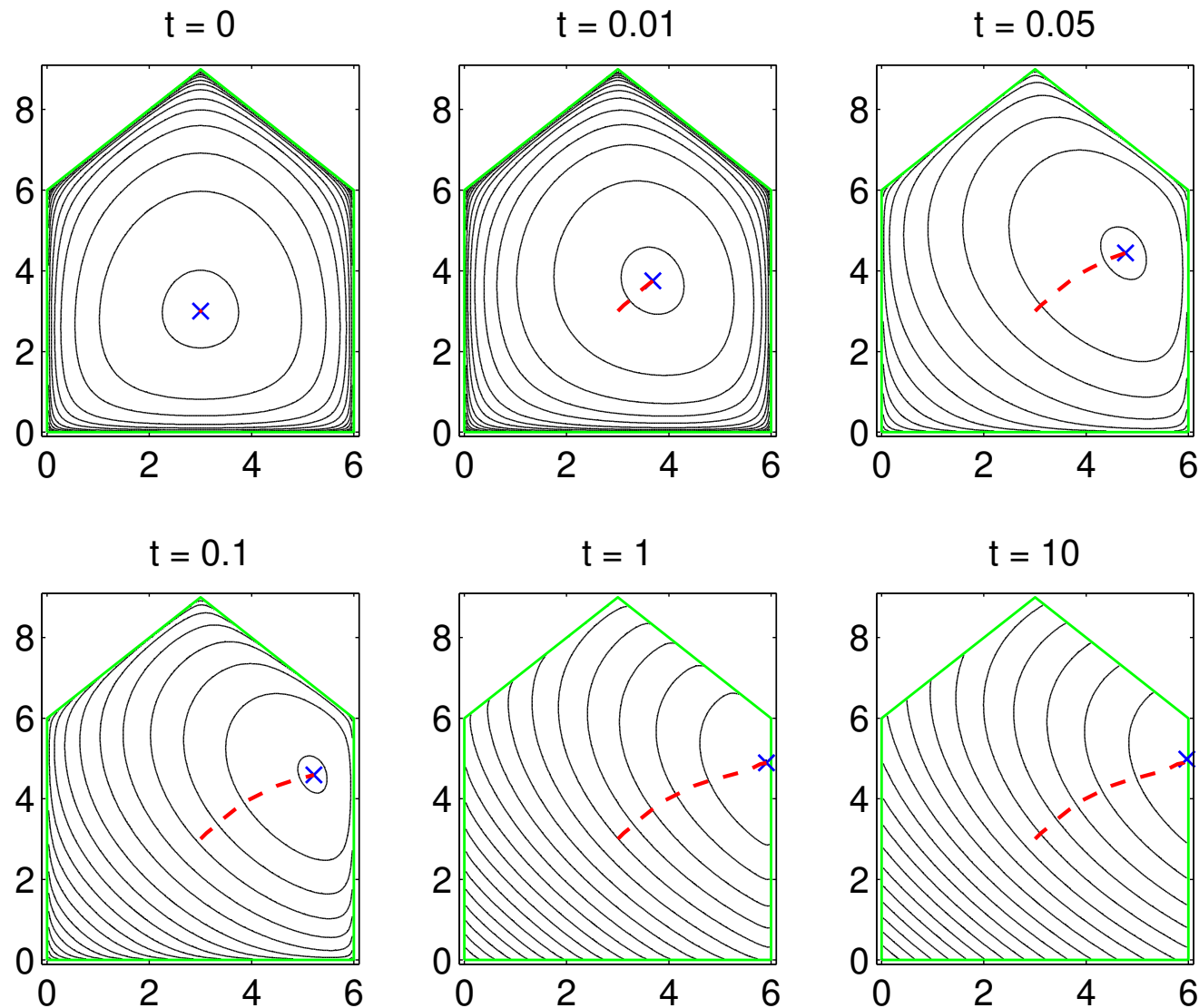
Step 2: Set $x = x^*(t)$

Step 3: If $\frac{m}{t} \leq \varepsilon_f$, return x and stop

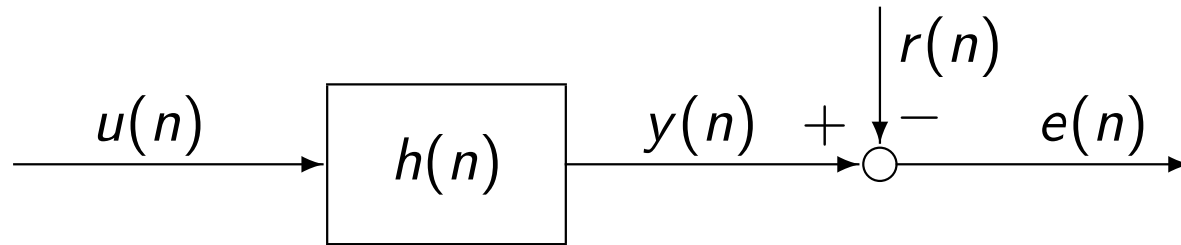
Step 4: Increase t and goto step 1

Interior-point algorithm: Example

$$f(x_1, x_2) = (x - x_0)^T C (x - x_0) \text{ with } C = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, x_0 = \begin{bmatrix} 9 \\ 4 \end{bmatrix}$$



Example from control theory



$$h(n) = \begin{cases} 2^{-n} & \text{for } 0 \leq n \leq 5, \quad n \in \mathbb{Z} \\ 0 & \text{else} \end{cases}$$

Reference signal: $r(n) = 1$

Input: $u(n) = 0$ for $n \in [6, 10]$

$$\text{Output: } y(n) = \sum_{m=0}^5 h(n-m) u(m) = F_1^T(n) x$$

$$x = \begin{bmatrix} u(0) & u(1) & \dots & u(5) \end{bmatrix}^T$$

$$F_1^T(n) = \begin{bmatrix} h(n) & h(n-1) & \dots & h(n-5) \end{bmatrix}$$

By defining $F_0(n) = -r(n)$, we obtain the error signal

$$e(n) = y(n) - r(n) = F_0(n) + F_1^T(n) x$$

Cost functions:

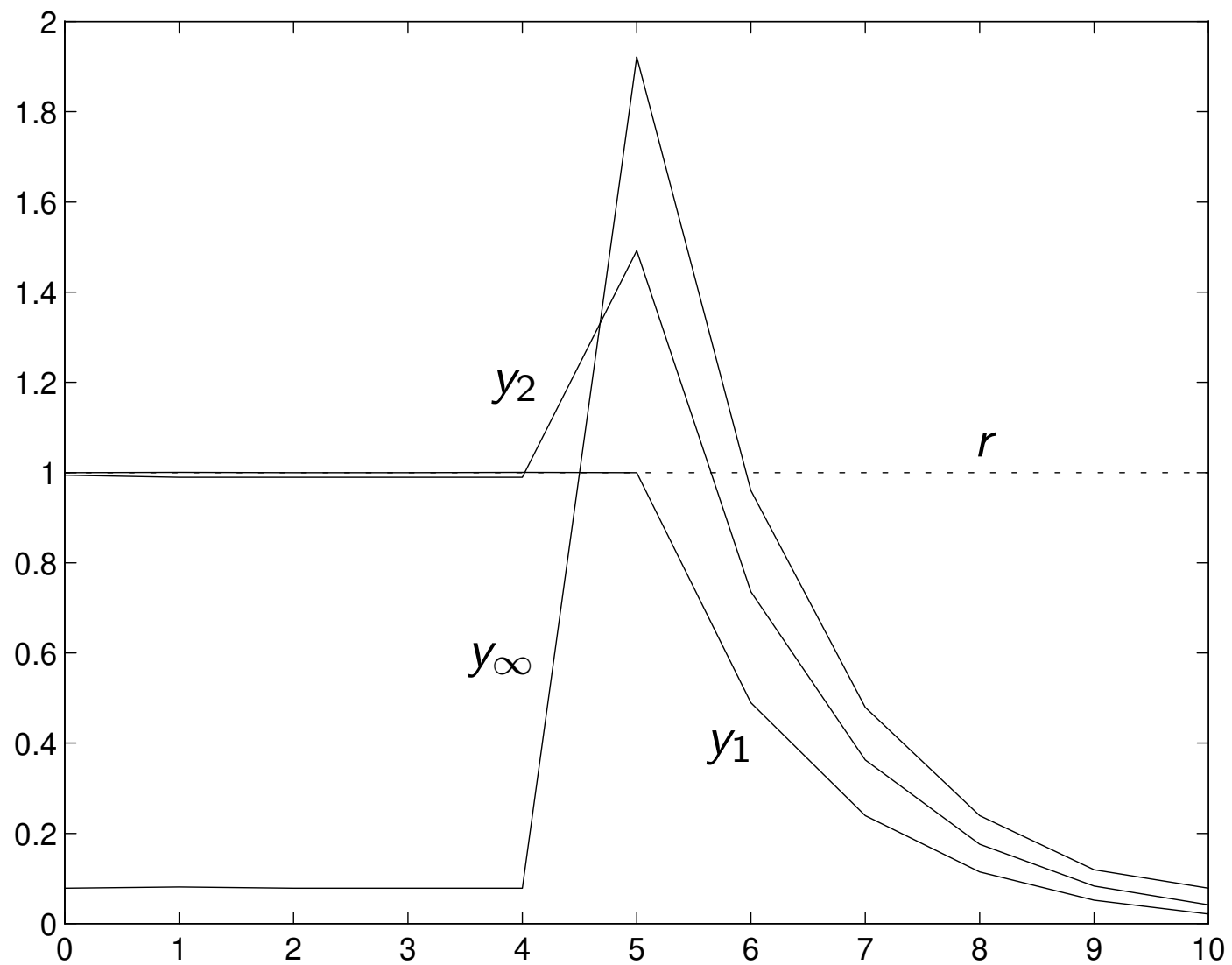
$$f_1(x) = \| e(n) \|_1$$

$$f_2(x) = \| e(n) \|_2^2$$

$$f_\infty(x) = \| e(n) \|_\infty$$

Optimization using cutting-plane algorithm yields:

$f(x)$	# iterations	$f_1(x^*)$	$f_2(x^*)$	$f_\infty(x^*)$
1-norm	12	4.083	3.480	0.979
2-norm	8	4.141	3.157	0.959
∞ -norm	7	8.650	7.567	0.922



Summary

- Convex functions + properties
- Linear matrix inequalities (LMIs)
- Convex optimization problem: Standard form

$$\begin{aligned} & \min_x f(x) \\ & \text{s.t. } g(x) \leq 0 \\ & \text{with } f \text{ and } g \text{ convex functions} \end{aligned}$$

- Algorithms for convex optimization:
 - ▶ cutting-plane
 - ▶ ellipsoid
 - ▶ interior-point
- Provide stopping criterion with hard guarantee: $|f(x^*) - f(x_k)| \leq \varepsilon_f$

Convex Functions - Recapitulation

Convex functions:

$$f(x) = a x \quad a \in \mathbb{R}$$

$$f(x) = x^{2n} \quad n \in \mathbb{N} \setminus \{0\}$$

$$f(x) = e^x$$

$$f(x) = h(\beta_0 + \beta^T x) \quad h \text{ is convex}$$

$$f(x) = \|x\| \quad \text{for any norm function}$$

Operations that preserve convexity:

Positive-weighted sum: $f(x) = \sum \alpha_i f_i(x) \quad \alpha_i > 0$

Pointwise maximum: $f(x) = \max f_i(x)$

Composition: $f(x) = h(g(x))$

if g, h convex and $h \nearrow$

if $-g, h$ convex and $h \searrow$