Optimization: Convex Optimization

Convex Optimization

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

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

<u>Convex function</u>:

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

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

Convex functions:

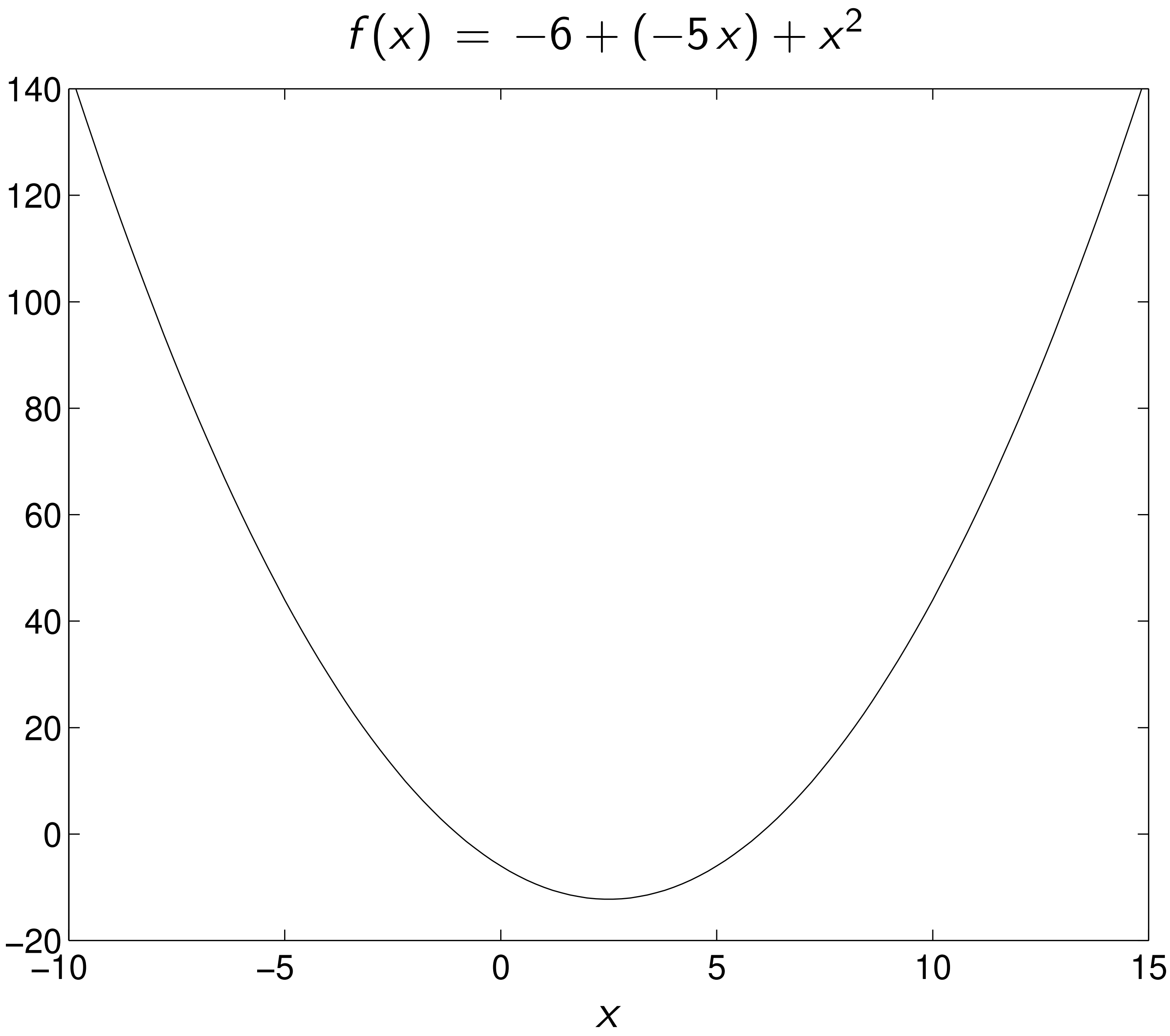
$$\begin{aligned}
f(x) &= a\,x && a \in \mathbb{R} \\
f(x) &= x^{2n} && n \in \mathbb{N} \setminus \{0\} \\
f(x) &= e^x && \\
f(x) &= h(\beta_0 + \beta^T x) && h \text{ is convex} \\
f(x) &= \|x\| && \text{for any norm function}
\end{aligned}$$

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

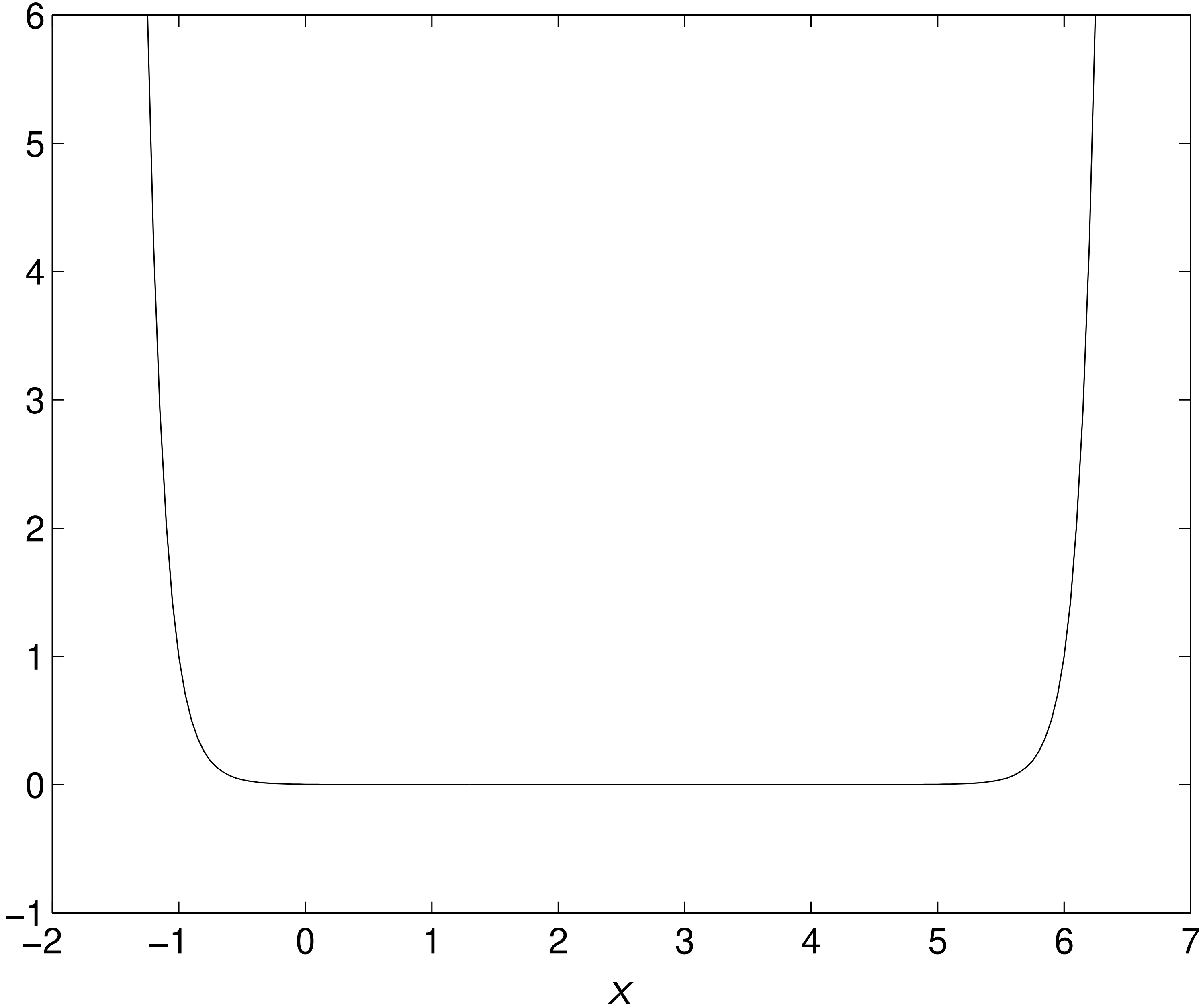
1. $\|u\| \geqslant 0$
2. $\|u\| = 0 \;\Leftrightarrow\; u = 0$
3. $\|\alpha\, u\| = |\alpha|\|u\| \quad$ for all $\alpha \in \mathbb{R}$
4. $\|u + v\| \leqslant \|u\| + \|v\| \quad$ for all $u, v \in \mathbb{R}^n$

Operations that preserve convexity:

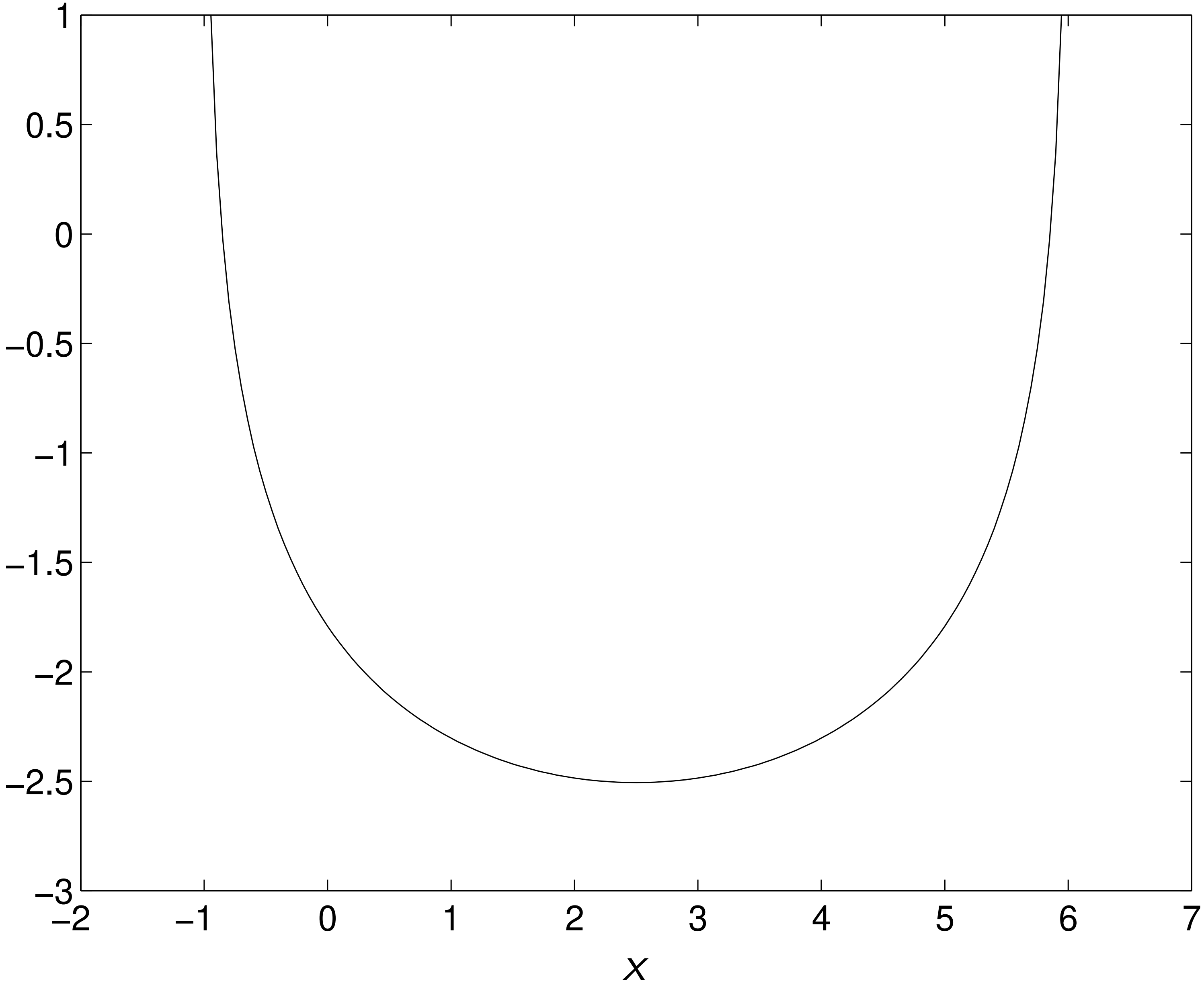
Positive-weighted sum:	$f(x) = \sum \alpha_i f_i(x) \quad \alpha_i > 0$
Pointwise maximum:	$f(x) = \max f_i(x)$
Composition:	$f(x) = h(g(x))$
	if g, h convex and $h \nearrow$
	if $-g, h$ convex and $h \searrow$


$f(x) = -6 + (-5x) + x^2$
140
120
100
80
60
40
20
0
–20
–10
–5
0
5
10
15
x

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



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

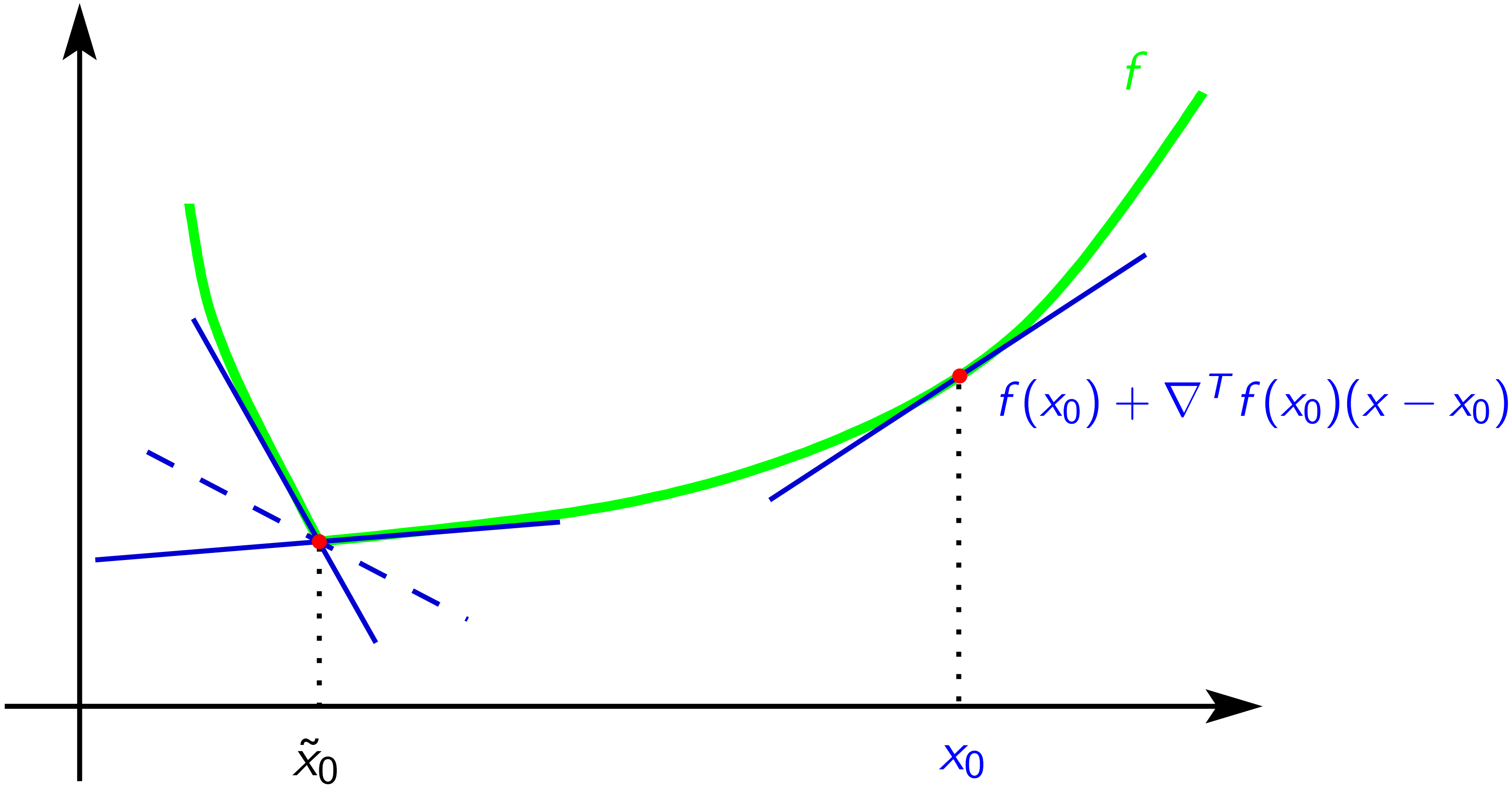


Test: Convex functions

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



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

Norms of affine functions

F_0: scalar signal, F_1: vector-valued signal

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

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

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

Subgradients:

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

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

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

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

Linear matrix inequalities

Linear matrix inequality

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

where $F_i = F_i^T$

- $\{x \,|\, F(x) > 0\}$ is convex set
- Multiple LMIs $\;\Rightarrow\;$ one LMI

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

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

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:

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

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

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

Lyapunov theory

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

is equivalent to

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

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

$$\frac{d}{dt}x(t) = A\,x(t) + B\,u(t) = A\,x(t) + B\,Kx(t)$$

is equivalent to

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

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

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

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

Quadratic objective:

$$\frac{d}{dt}x(t) = A\,x(t) + B\,u(t) = A\,x(t) + B\,Kx(t)$$

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

$$\min_K \underbrace{\int_0^\infty (x^T Q x + u^T R u)\,dt}_{J(K)} = \quad \min_K \int_0^\infty x^T (Q + K^T R K) x\,dt$$

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

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

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

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

Quadratic objective (continued):

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

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

$$XA^T + Y^T B^T + AX + BY + \gamma^{-1} XQX + \gamma^{-1} Y^T RY < 0$$

Schur complement transformation:

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

$$\gamma > 0\,, \;\; X > 0$$

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

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

$\Rightarrow \quad \min\limits_{X,Y,\gamma} \; \gamma \quad$ subject to LMIs

Some common convex problems with LMIs

- **Entropy maximization**:

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

- **Determinant maximization**:

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

- **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":

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

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

where x^* is the real optimum

Cutting-plane algorithm

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

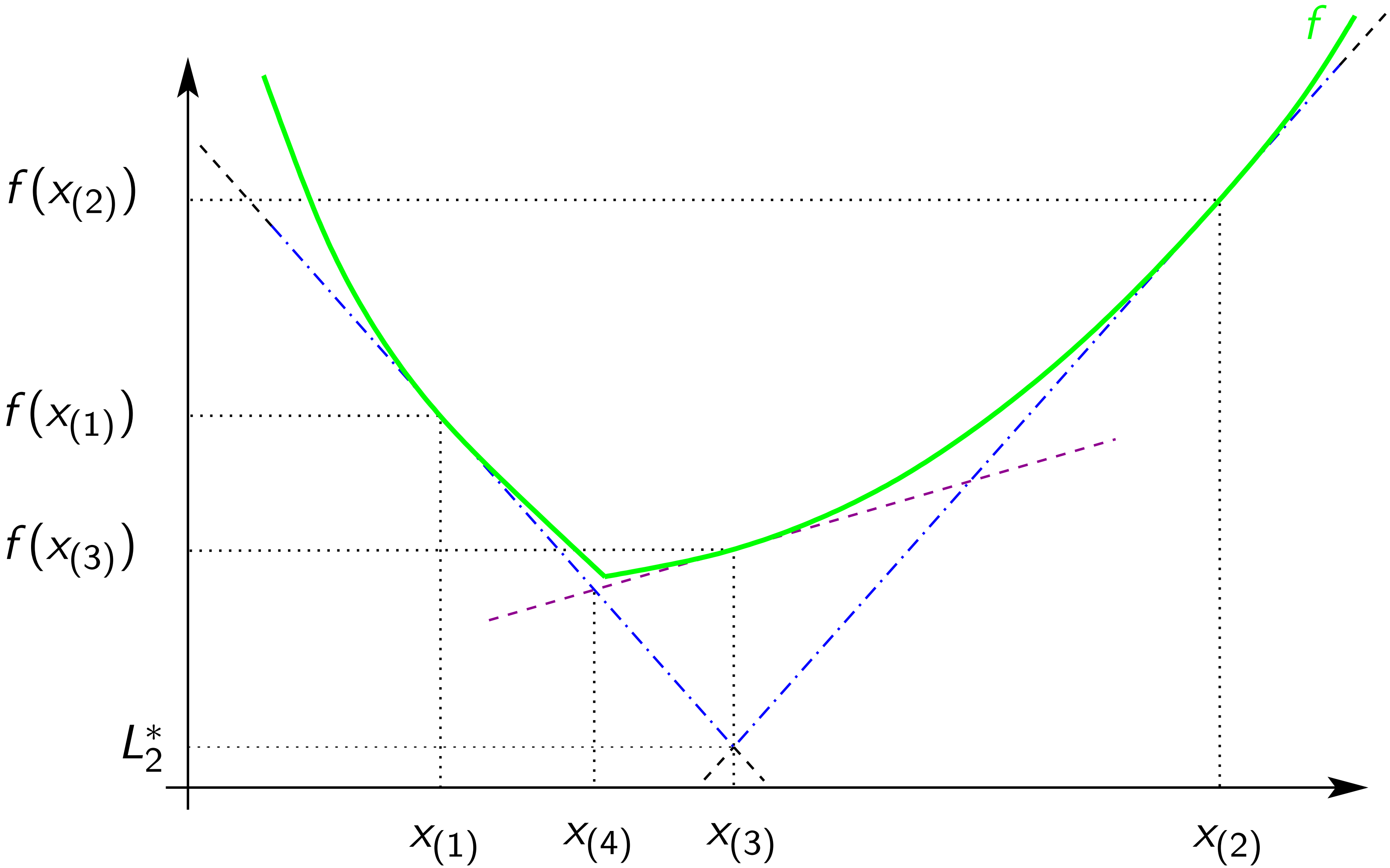
$$f(x) \geqslant f(x_{(i)}) + \nabla^T f(x_{(i)})(x - x_{(i)})$$

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

Hence,

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

Cutting-plane algorithm (continued)



Cutting-plane algorithm (continued)

$$f(x^*) \geqslant \min_{x} \max_{i=1,\ldots,k} \Big(f(x_{(i)}) + \nabla^T f(x_{(i)})(x - x_{(i)}) \Big)$$

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

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

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

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

$$\Rightarrow \quad L_k^* \leqslant f(x_k^*) \leqslant U_k^*$$

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

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

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

Cutting-plane: Handling constraints

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

Since g is convex:

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

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

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

Add extra constraints to LP:

$$\nabla^T g(x_{(i)}) \, x \; \leqslant \; \nabla^T g(x_{(i)}) \, x_{(i)} - g(x_{(i)})$$

$\rightarrow$ still LP!

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

Ellipsoid algorithm

One-dimensional case:

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

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

Recall that

$$f(x) \; \geqslant \; f(x_{(0)}) + \nabla^T f(x_{(0)})\,(x - x_{(0)})$$

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

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

Construct

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

with $\quad A_1 = \dfrac{A_0}{2}$

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

Ellipsoid algorithm

Multi-dimensional case:

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

$$E_0 = \{x \,|\, (x - x_{(0)})^T A_0^{-1} (x - x_{(0)}) \leqslant 1\}$$

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

$$f(x) \;\geqslant\; f(x_{(0)}) + \nabla^T f(x_{(0)})\,(x - x_{(0)})$$

x^* will be in half-plane

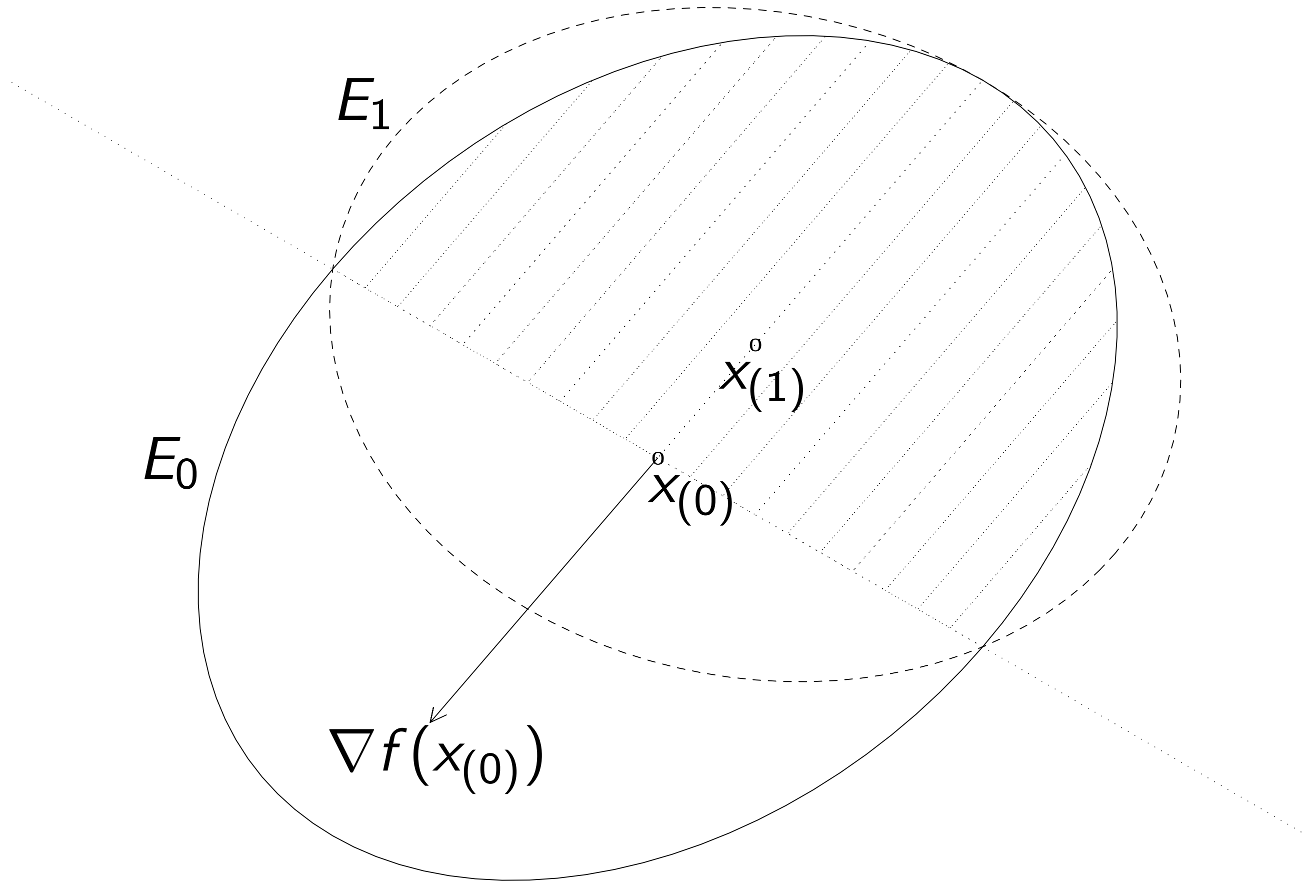
$$H_0 \;=\; \{\, x \,|\, \nabla^T f(x_{(0)})\,(x - x_{(0)}) \;\leqslant\; 0 \,\}$$

Construct new ellipsoid

$$E_1 \;=\; \{x \,|\, (x - x_{(1)})^T A_1^{-1} (x - x_{(1)}) \leqslant 1\}$$

such that $H_0 \cap E_0 \subseteq E_1$

Ellipsoid algorithm (continued)



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

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:

$$\mathrm{Vol}(E_k) \to 0 \quad \text{for} \quad k \to \infty$$

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

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

Ellipsoid algorithm: Handling constraints

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

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

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

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

$\rightarrow$ 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

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

Strictly feasible set $\mathcal{G}$:

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

Barrier function:

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

$\rightarrow$ convex

Optimization problem for $t \geqslant 0$: $\min_x \; t\, f(x) + \phi(x)$

Central path $x^*(t)$:

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

$x^*(t)$ 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:

$$\begin{aligned} g(x^*(t)) &\leqslant 0 \\ \mu_i^*(t) &\geqslant 0 \\ \nabla f(x^*(t)) + \sum_{i=1}^{m} \mu_i^*(t) \nabla g_i(x^*(t)) &= 0 \\ \mu_i^*(t) g_i(x^*(t)) &= -1/t \end{aligned}$$

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

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: $d(\mu^*(t)) \leqslant f(x^*)$ for any $\mu^*(t) \geqslant 0$

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

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

This yields:

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

Interior-point algorithm (continued)

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

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

Take $t = \dfrac{m}{\varepsilon_f} \rightarrow$ 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:

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

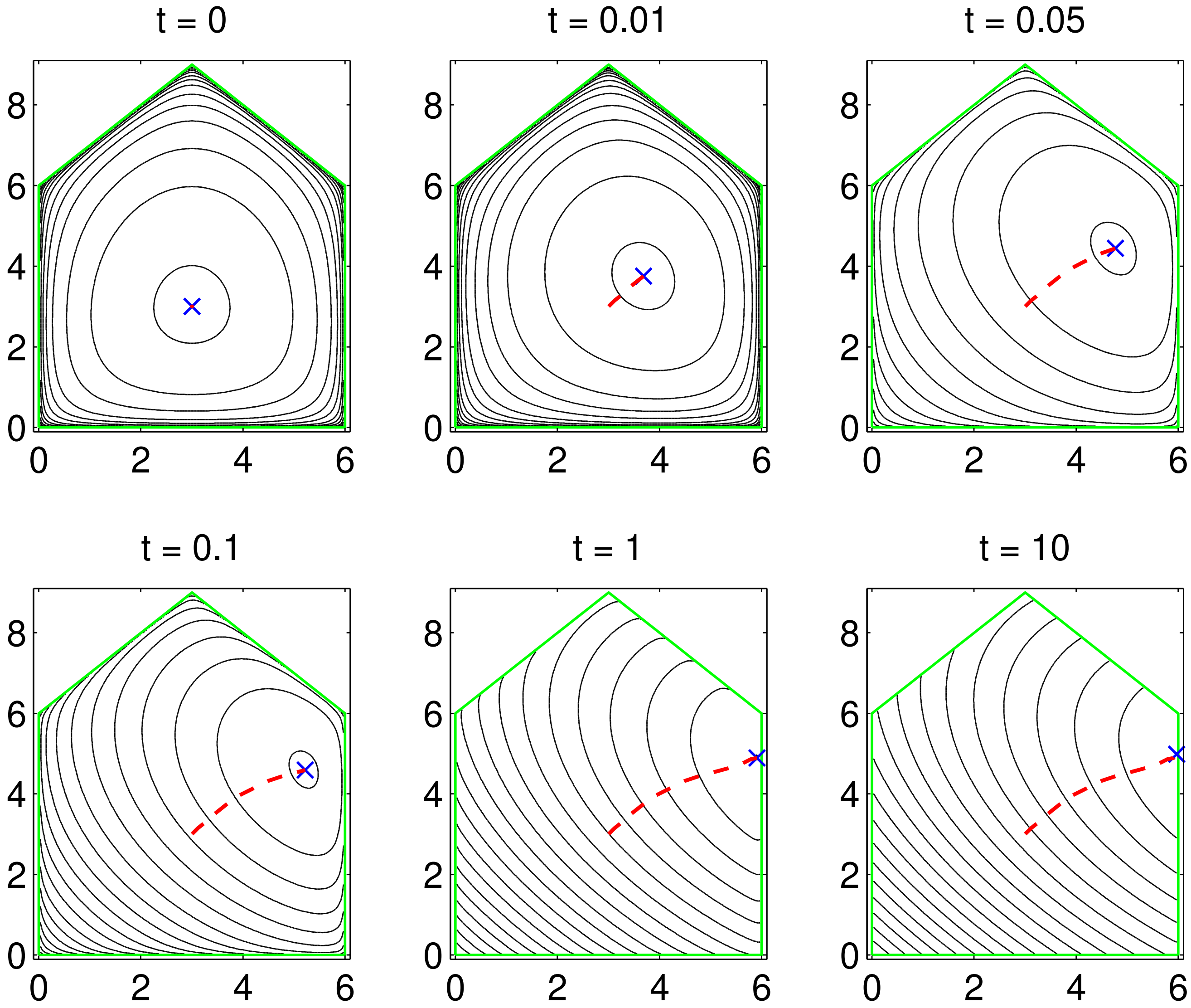
Step 2: Set $x = x^*(t)$

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

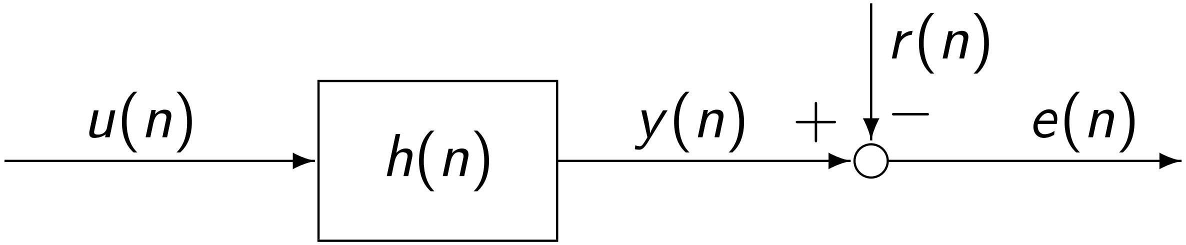
Step 4: Increase t and goto step 1

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



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

Reference signal: $r(n) = 1$

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

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

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

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

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

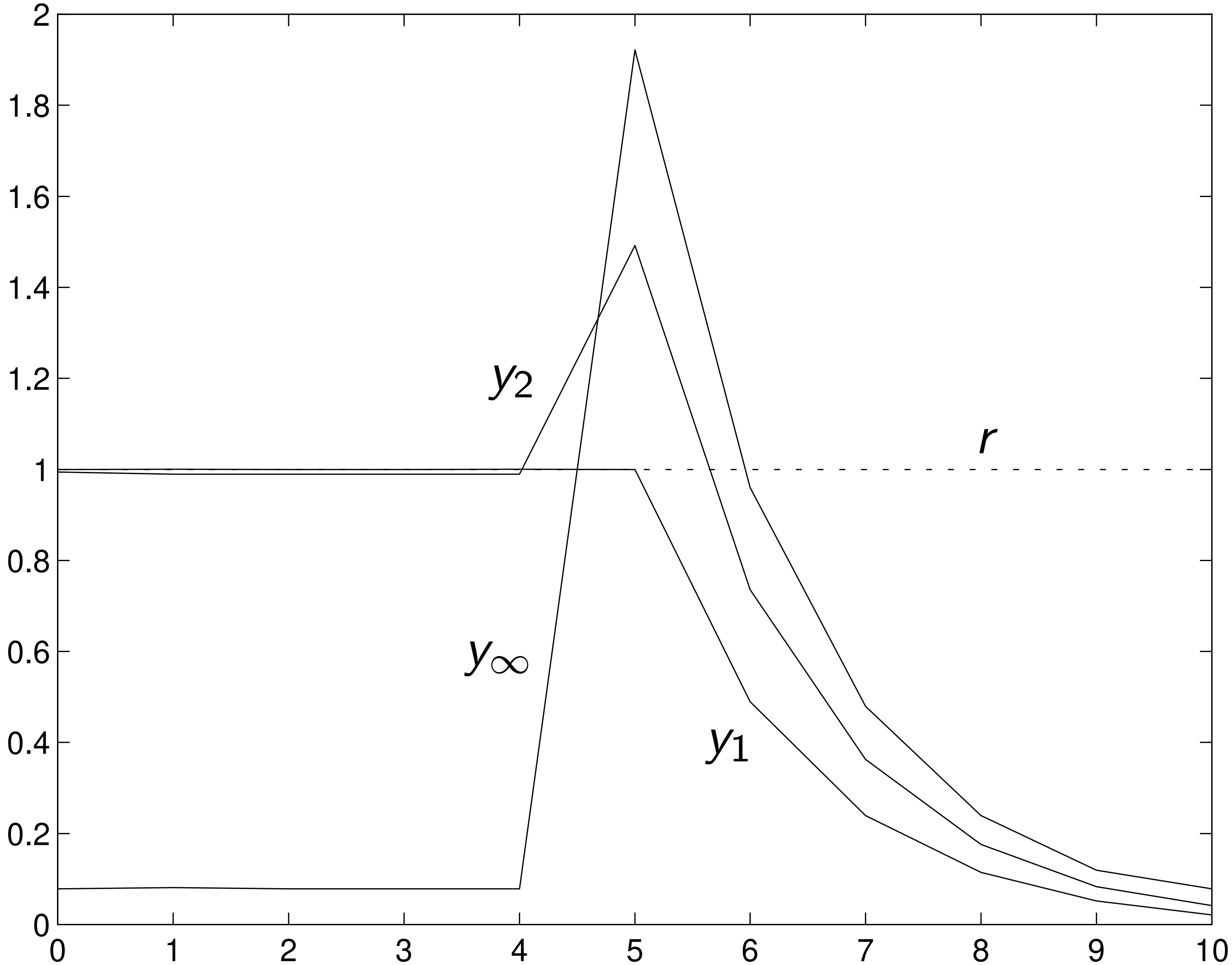
$$e(n) = y(n) - r(n) = F_0(n) + F_1^T(n)\, x$$

Cost functions:

$$\begin{aligned} f_1(x) &= \| \, e(n) \, \|_1 \\ f_2(x) &= \| \, e(n) \, \|_2^2 \\ f_\infty(x) &= \| \, e(n) \, \|_\infty \end{aligned}$$

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


y_2
r
y_∞
y_1
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
1
2
3
4
5
6
7
8
9
10

Summary

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

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

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

Convex Functions - Recapitulation

Convex functions:

$$
\begin{aligned}
f(x) &= a\,x && a \in \mathbb{R} \\
f(x) &= x^{2n} && n \in \mathbb{N} \setminus \{0\} \\
f(x) &= e^{x} && \\
f(x) &= h(\beta_0 + \beta^T x) && h \text{ is convex} \\
f(x) &= \|x\| && \text{for any norm function}
\end{aligned}
$$

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 $h \searrow$