4. Sequential convex programming

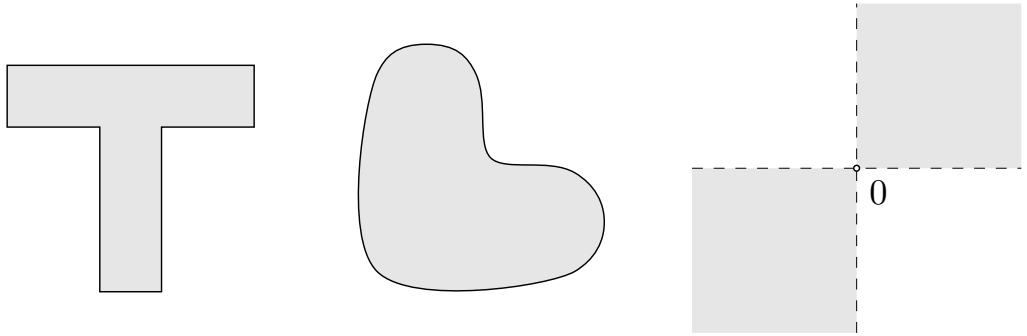
- biconvex sets and functions
- functional operations preserving biconvexity
- biconvex optimization problems
- alternate convex search
- difference-of-convex functions
- first-order majorization
- difference-of-convex programming
- convex-concave procedure
- sequential convex approximation
- Taylor expansions and particle methods

Biconvex sets

biconvex set $B \subseteq \mathbf{R}^n \times \mathbf{R}^k$ is convex when restricted to either variable:

- $B_{\tilde{y}} = \{x \in \mathbf{R}^n \mid (x, \tilde{y}) \in B\}$ convex for all $\tilde{y} \in \mathbf{R}^k$
- $B_{\tilde{x}} = \{y \in \mathbf{R}^k \mid (\tilde{x}, y) \in B\}$ convex for all $\tilde{x} \in \mathbf{R}^n$

examples:



- solution set of a system of *bilinear equations* is biconvex:

$$B = \{(x, y) \in \mathbf{R}^n \times \mathbf{R}^k \mid (Ax + b)^T (Cy + d) = 0\}$$

- intersection of (potentially infinite) biconvex sets is biconvex

Biconvex functions

$f : \mathbf{R}^n \times \mathbf{R}^k \to \mathbf{R}$ is **biconvex function** if

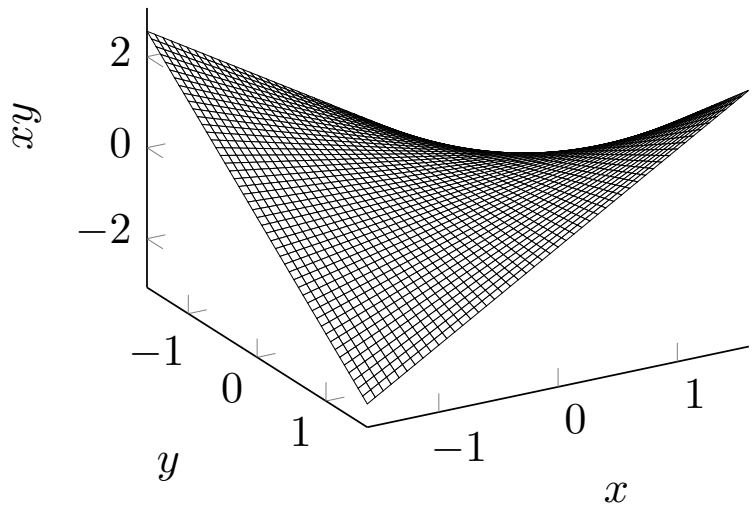
- $\mathbf{dom}\, f$ is a biconvex set
- $f_{\tilde{y}} : \mathbf{R}^n \to \mathbf{R}$ given by $x \mapsto f(x, \tilde{y})$ is convex in x
- $f_{\tilde{x}} : \mathbf{R}^k \to \mathbf{R}$ given by $y \mapsto f(\tilde{x}, y)$ is convex in y

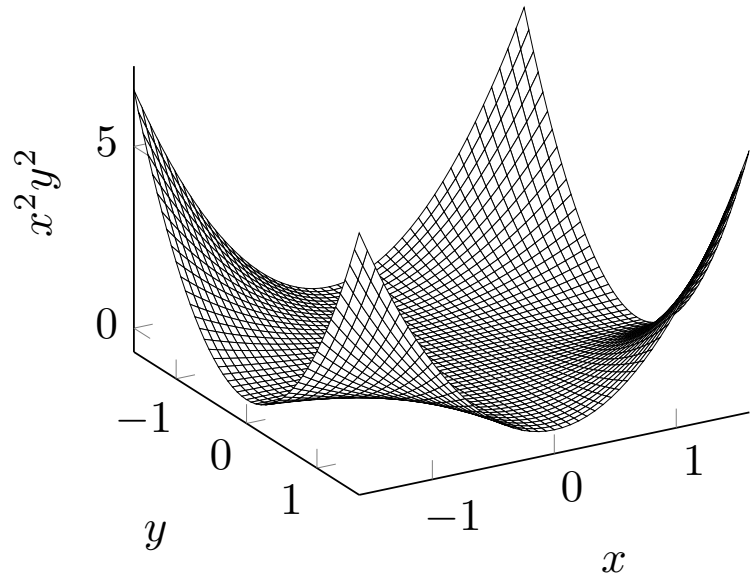
i.e., f is convex in each variable when the other is fixed

biconcave, **bilinear**, and **biaffine** functions are defined similarly

examples:

- $f(x, y) = x^T y$ is bilinear (hence biaffine)
- $f(x, y) = (Ax + b)^T (Cy + d)$ is biaffine (hence biconvex and biconcave)
- $f(X, Y) = XY - A$ is biaffine (on $\mathbf{R}^{n \times m} \times \mathbf{R}^{m \times k}$)
- $f(x, y) = x^2 y^2$ is biconvex





Sublevel sets

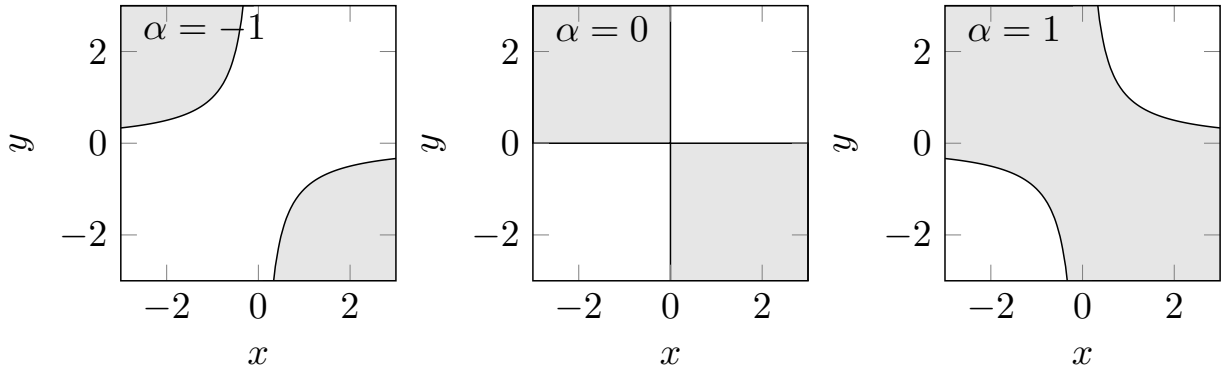
α-sublevel set of biconvex function $f : \mathbf{R}^n \times \mathbf{R}^k \to \mathbf{R}$:

$$C_\alpha = \{(x, y) \in \mathbf{dom}\, f \mid f(x, y) \leq \alpha\}$$

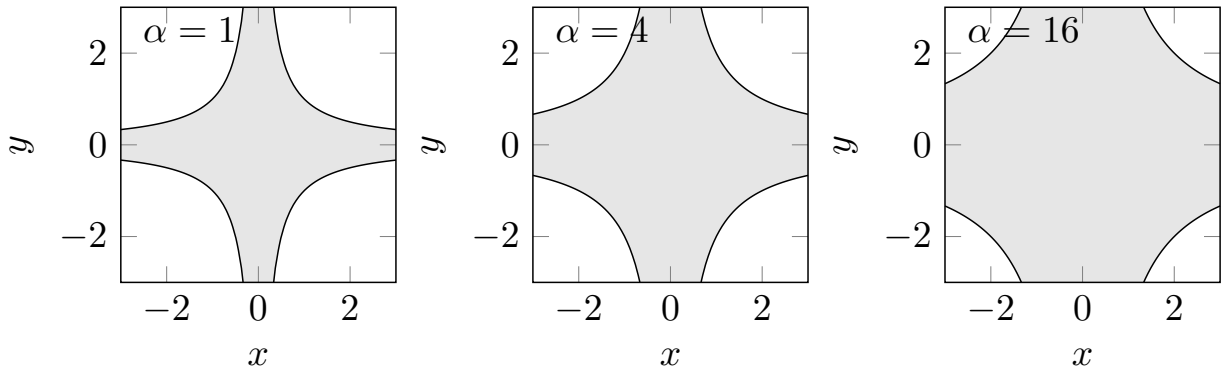
is biconvex

examples:

- $f(x, y) = xy$:



- $f(x, y) = x^2y^2$:



Functional operations preserving biconvexity

nonnegative weighted sum of biconvex functions is biconvex

pointwise maximum and supremum of biconvex functions is biconvex

biaffine composition:

$$h\colon \mathbf{R} \to \mathbf{R} \text{ convex} \implies h((Ax+b)^T(Cy+d)) \text{ biconvex}$$

general composition:

$$\begin{array}{l} h\colon \mathbf{R} \to \mathbf{R} \text{ convex nondecreasing} \\ g\colon \mathbf{R}^n \times \mathbf{R}^k \to \mathbf{R} \text{ biconvex} \end{array} \implies h(g(x,y)) \text{ biconvex}$$

- can be generalized to multivariate h

example: matrix factorization cost

$$f(X,Y) = \|XY - A\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{k}\left((XY)_{ij} - A_{ij}\right)^2$$

- $A \in \mathbf{R}^{n\times k}$ is a given matrix

- composition of convex function $h(Z) = \|Z\|_F^2$ with biaffine function $g(X,Y) = XY - A$

Biconvex optimization problems

standard form biconvex optimization problem:

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

- $f_i\colon \mathbf{R}^n \times \mathbf{R}^k \to \mathbf{R}$ biconvex for $i = 0, 1, \ldots, m$
- $h_i\colon \mathbf{R}^n \times \mathbf{R}^k \to \mathbf{R}$ biaffine for $i = 1, \ldots, p$
- not convex in the most general case

feasible set:

$$\mathcal{D} = \left\{ (x,y) \in \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i \;\middle|\; \begin{array}{l} f_i(x,y) \leq 0,\ i=1,\ldots,m \\ h_i(x,y) = 0,\ i=1,\ldots,p \end{array} \right\}$$

is the intersection of biconvex sets and hence biconvex

Optimal value and points

feasible point $(x^\star, y^\star) \in \mathcal{D}$ is **partially optimal** if

- $f_0(x^\star, y^\star) \leq f_0(x, y^\star)$ for all $x \in \mathcal{D}_{y^\star} = \{x \mid (x, y^\star) \in \mathcal{D}\}$
- $f_0(x^\star, y^\star) \leq f_0(x^\star, y)$ for all $y \in \mathcal{D}_{x^\star} = \{y \mid (x^\star, y) \in \mathcal{D}\}$

$f(x^\star, y^\star)$ called **partially optimal value**

- every stationary point of differentiable f_0 over $\mathcal{D}$ is partially optimal, and vice versa
- not necessarily a local optimum (and of course global optimum)
- turns out to be quite useful in practice
- no stronger optimality conditions exist for general biconvex problems

Alternate convex search (ACS)

iterate between solving convex subproblems in x and y:

Alternate convex search.

given feasible starting point $(x^{(0)}, y^{(0)})$.
$k := 0$.
repeat

1. $x^{(k+1)} := \underset{x \in \mathbf{R}^n}{\operatorname{argmin}} \left\{ f_0(x, y^{(k)}) \;\middle|\; \begin{array}{l} f_i(x, y^{(k)}) \leq 0, \; i = 1, \ldots, m \\ h_i(x, y^{(k)}) = 0, \; i = 1, \ldots, p \end{array} \right\}.$

2. $y^{(k+1)} := \underset{y \in \mathbf{R}^k}{\operatorname{argmin}} \left\{ f_0(x^{(k+1)}, y) \;\middle|\; \begin{array}{l} f_i(x^{(k+1)}, y) \leq 0, \; i = 1, \ldots, m \\ h_i(x^{(k+1)}, y) = 0, \; i = 1, \ldots, p \end{array} \right\}.$

3. $k := k + 1$.

until stopping criteria are satisfied.

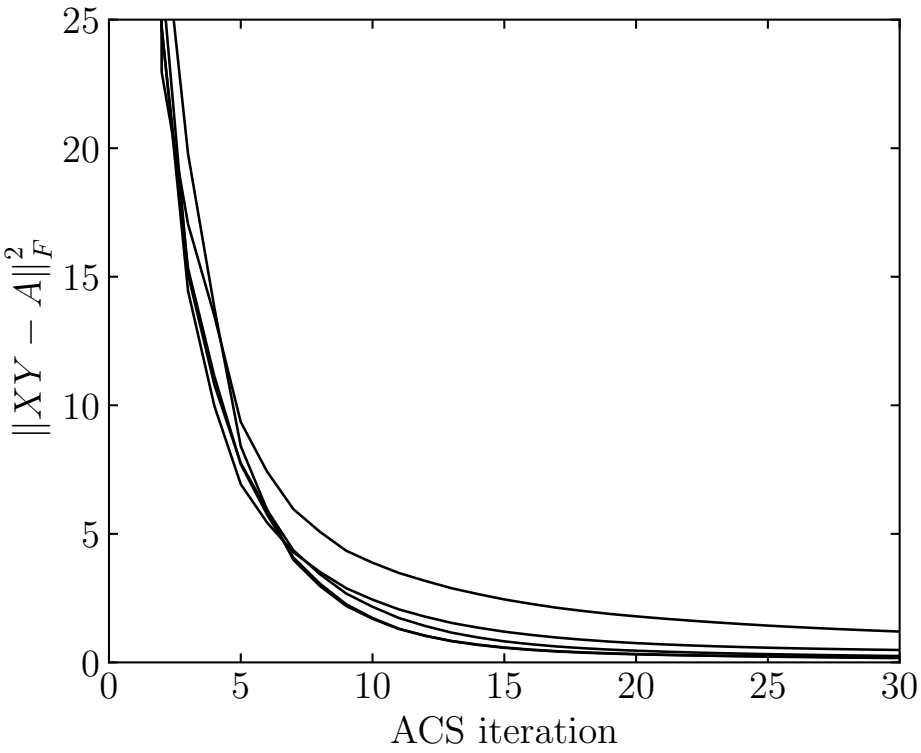
- objective value is nonincreasing
- converges to stationary point (hence partially optimal)

example: nonnegative matrix factorization

$$\begin{array}{ll} \text{minimize} & \|XY - A\|_F^2 \\ \text{subject to} & X_{ij} \geq 0, \quad i = 1, \ldots, m, \quad j = 1, \ldots, k \\ & Y_{ij} \geq 0, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n \end{array}$$

- variables $X \in \mathbf{R}^{m \times k}$ and $Y \in \mathbf{R}^{k \times n}$, $A \in \mathbf{R}^{m \times n}$ is given matrix

for $m = n = 50$, $k = 5$,
multiple random initializations:



Difference-of-convex functions

difference-of-convex (DC) function $h\colon \mathbf{R}^n \to \mathbf{R}$ has the form

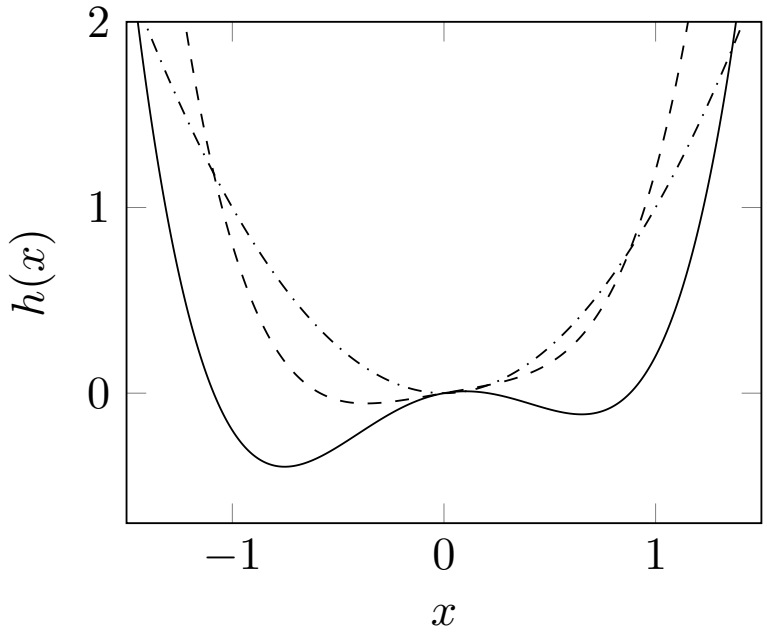
$$h(x) = f(x) - g(x), \quad f,\, g \text{ convex}$$

- any convex f is DC with $f(x) = f(x) - 0$
- any concave h is DC with $h(x) = 0 - (-h(x))$

examples:

- $h(x) = x^4 + (1/5)x - x^2$ is DC with

  $$f(x) = x^4 + (1/5)x, \quad g(x) = x^2$$



- nonconvex quadratic function

$$h(x) = x^T P x + q^T x + r, \quad P \in \mathbf{S}^n,\ q \in \mathbf{R}^n,\ r \in \mathbf{R}$$

  is expressed in DC form as

$$h(x) = x^T P_+ x + q^T x + r - x^T P_- x$$

  where $P_+ \succeq 0$, $P_- \succeq 0$, and $P = P_+ - P_-$
  (obtained from eigenvalue decomposition)

First-order majorization

for DC function $h(x) = f(x) - g(x)$ and differentiable g:

- 1st-order Taylor expansion of g at $\tilde{x}$ is affine function $\hat{g}_{\tilde{x}} : \mathbf{R}^n \to \mathbf{R}$:

$$\hat{g}_{\tilde{x}}(x) = g(\tilde{x}) + \nabla g(\tilde{x})^T (x - \tilde{x}) \leq g(x)$$

  which is a global underestimator of g

- define $\hat{h}_{\tilde{x}} : \mathbf{R}^n \to \mathbf{R}$ as

$$\hat{h}_{\tilde{x}}(x) = f(x) - \hat{g}_{\tilde{x}}(x) = f(x) - g(\tilde{x}) - \nabla g(\tilde{x})^T (x - \tilde{x})$$

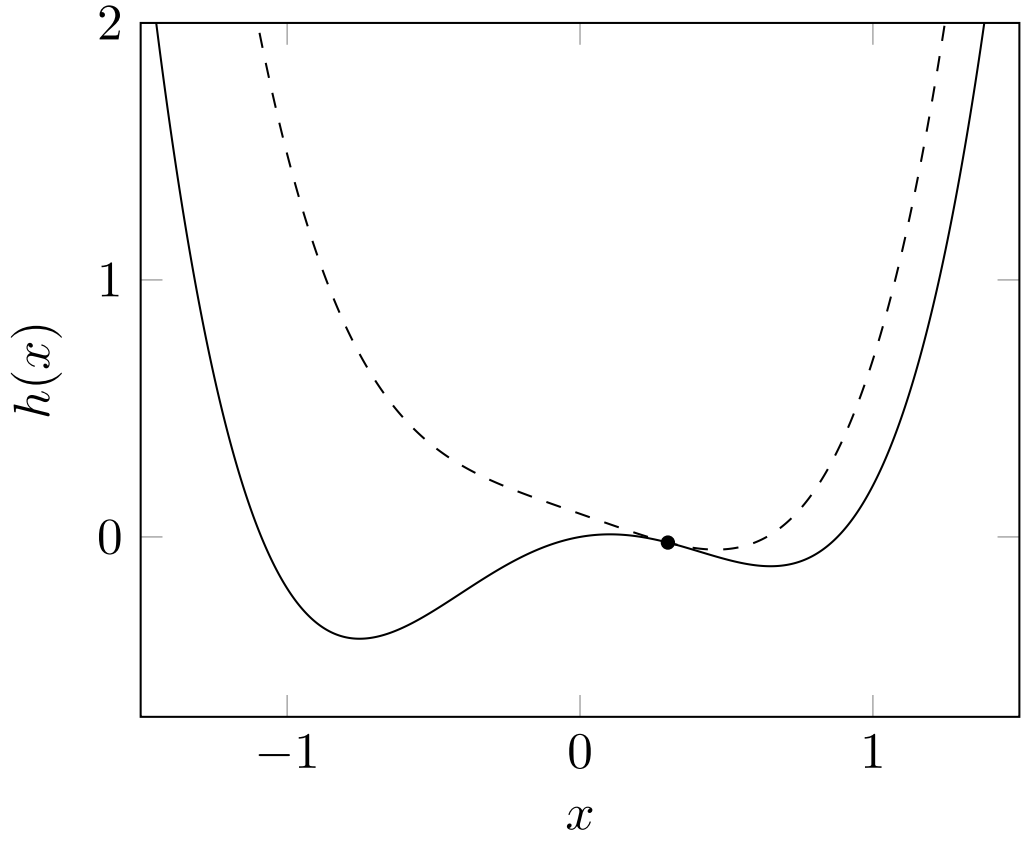
  – $\hat{h}_{\tilde{x}}(x) \geq f(x) - g(x) = h(x) \implies \hat{h}_{\tilde{x}}$ is a global *majorizer* of h

  – $\hat{h}_{\tilde{x}}(\tilde{x}) = h(\tilde{x}) \implies$ majorizer is tight at $\tilde{x}$

example: 1st-order majorization of DC function $h(x) = x^4 + (1/5)x - x^2$

$$\begin{aligned}\hat{h}_{\tilde{x}}(x) &= x^4 + (1/5)x - (\tilde{x}^2 + 2\tilde{x}(x - \tilde{x})) \\ &= x^4 + (1/5)x - 2\tilde{x}x + \tilde{x}^2\end{aligned}$$

at $\tilde{x} = 0.3$:



Difference-of-convex programming

difference-of-convex optimization problem has the form

$$\begin{array}{ll} \text{minimize} & f_0(x) - g_0(x) \\ \text{subject to} & f_i(x) - g_i(x) \leq 0, \quad i = 1, \ldots, m \end{array}$$

- f_i, g_i convex for $i = 0, 1, \ldots, m$
- nonconvex in general, reduces to convex program if g_i are all affine
- includes equality constraints, *e.g.*,

$$p_i(x) = q_i(x), \quad i = 1, \ldots, k$$

with convex p_i, q_i as

$$p_i(x) - q_i(x) \leq 0, \quad q_i(x) - p_i(x) \leq 0, \quad i = 1, \ldots, k$$

example: boolean least squares problem

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_2^2 \\ \text{subject to} & x_i \in \{0,1\}, \quad i=1,\ldots,n \end{array}$$

is expresses as a DC program as

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_2^2 \\ \text{subject to} & x_i^2 - x_i \leq 0, \quad i=1,\ldots,n \\ & x_i - x_i^2 \leq 0, \quad i=1,\ldots,n, \end{array}$$

since for $i=1,\ldots,n$:

$$\begin{array}{lcll} x_i \in \{0,1\} & \iff & x_i(1-x_i)=0 & \\ & \iff & x_i(1-x_i) \leq 0 & \text{and} \quad -x_i(1-x_i) \leq 0 \end{array}$$

Convex-concave procedure (CCP)

iteratively solve convex approximations of DC program:

Convex-concave procedure.

given a feasible starting point $x^{(0)}$.
$k := 0$.
repeat
1. *Convexify.*

$$\hat{g}_{i,x^{(k)}}(x) := g_i(x^{(k)}) + \nabla g_i(x^{(k)})^T(x - x^{(k)}), \quad i = 0, 1, \ldots, m$$

2. *Solve the convex approximation.*

$$\begin{array}{lll} x^{(k+1)} := \mathrm{argmin}_x & f_0(x) - \hat{g}_{0,x^{(k)}}(x) & \\ \text{subject to} & f_i(x) - \hat{g}_{i,x^{(k)}}(x) \leq 0, & i = 1, \ldots, m \end{array}$$

3. $k := k + 1$.

until stopping criteria are satisfied.

---

interpretation: iteratively find a point

- minimizing a global upper bound of the DC objective, since

$$f_0(x) - g_0(x) \leq f_0(x) - \hat{g}_{0,x^{(k)}}(x), \quad k = 0, 1, \ldots$$

- over a convex restriction of the original feasible set, since

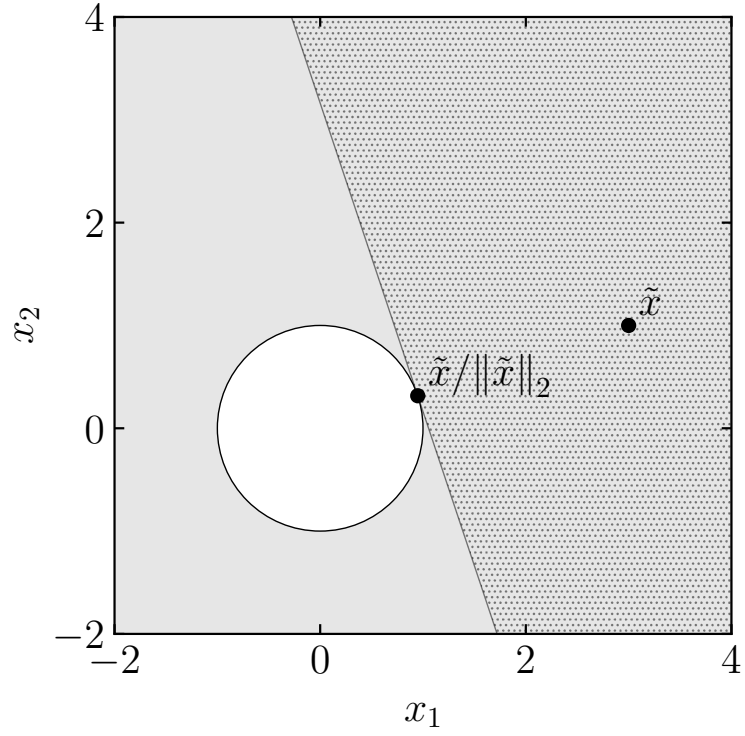
$$f_i(x) - g_i(x) \leq f_i(x) - \hat{g}_{i,x^{(k)}}(x) \leq 0, \quad k = 0, 1, \ldots$$

example:

- shaded region: $\|x\|_2 \geq 1$
- dotted region: approximation

$$\frac{\tilde{x}^T}{\|\tilde{x}\|_2} x \geq 1$$

  in each CCP iteration at $\tilde{x}$



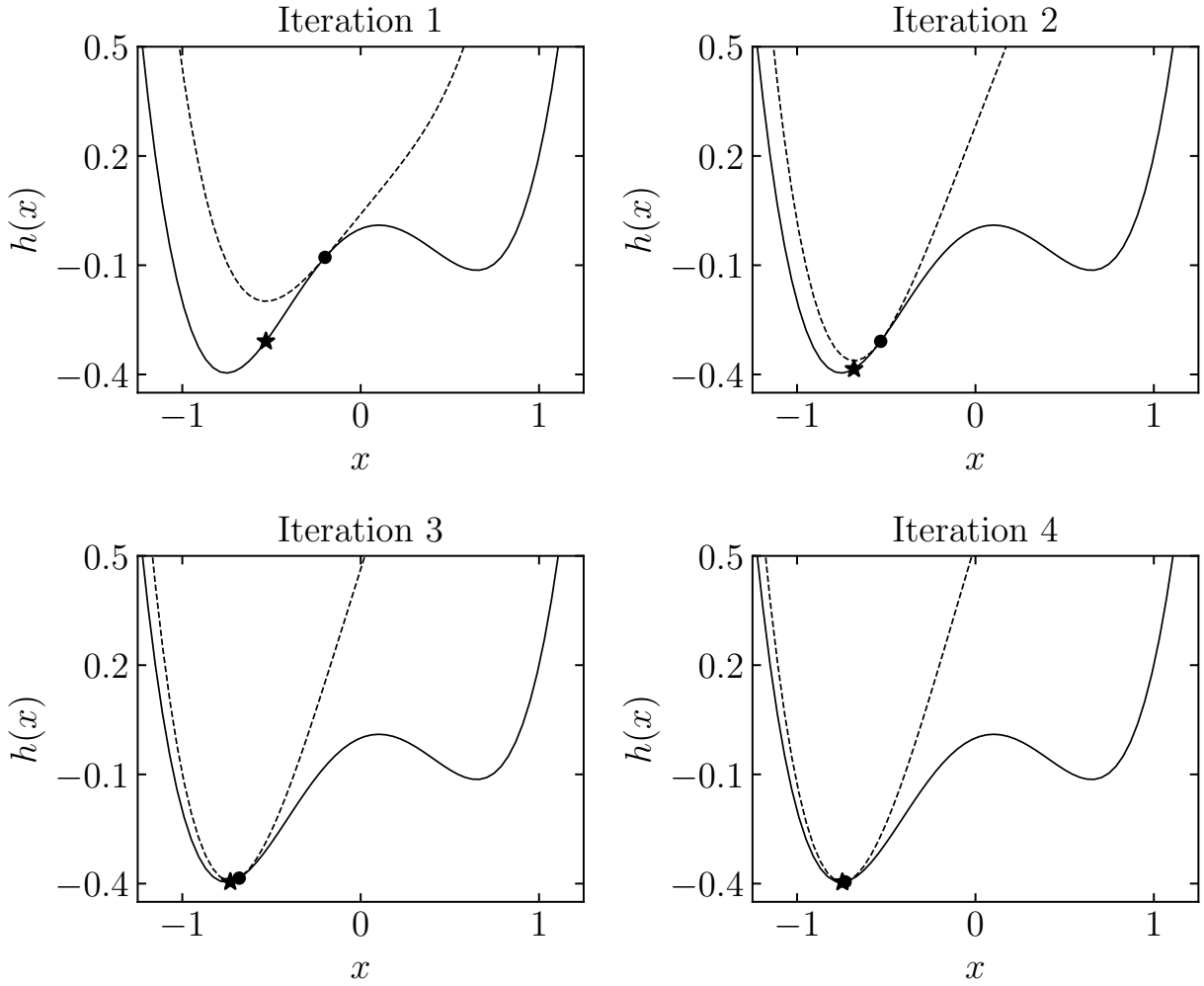
Minimizing a DC function

unconstrained DC program

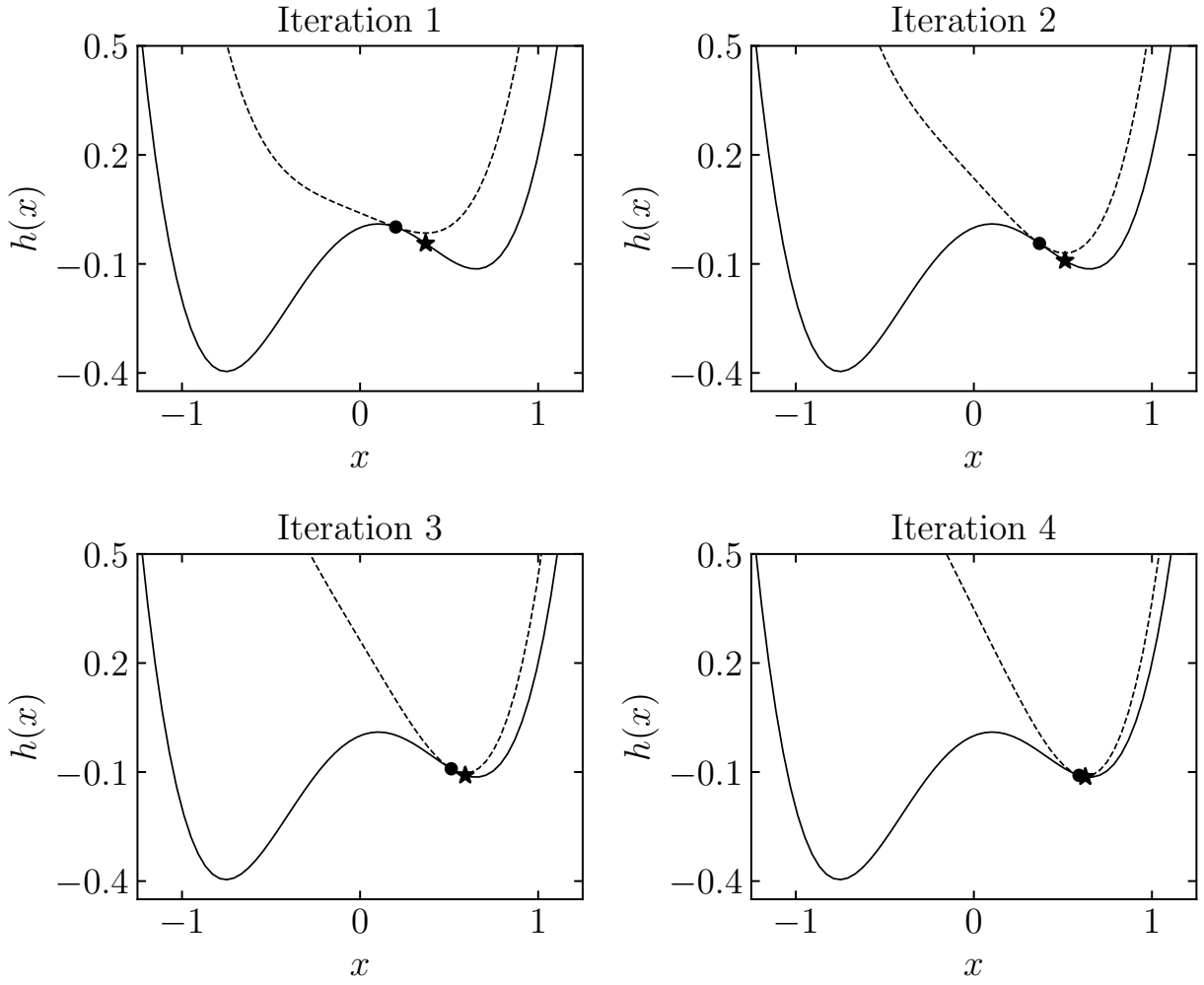
minimize $x^4 + (1/5)x - x^2$

CCP iteration at $\tilde{x}$ solves

minimize $x^4 + (1/5)x - 2\tilde{x}x + \tilde{x}^2$



CCP is initialization-dependent:



Closest point outside unit ball

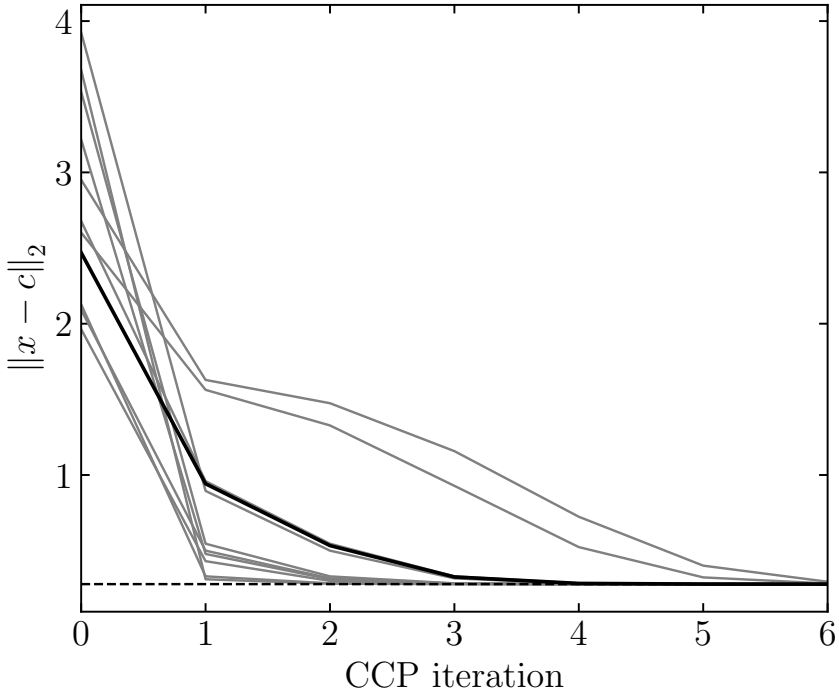
find $x \in \mathbf{R}^n$ closest to given point $c \in \mathbf{R}^n$ outside the unit ball:

$$\begin{array}{ll} \text{minimize} & \|x - c\|_2 \\ \text{subject to} & \|x\|_2 \geq 1 \end{array}$$

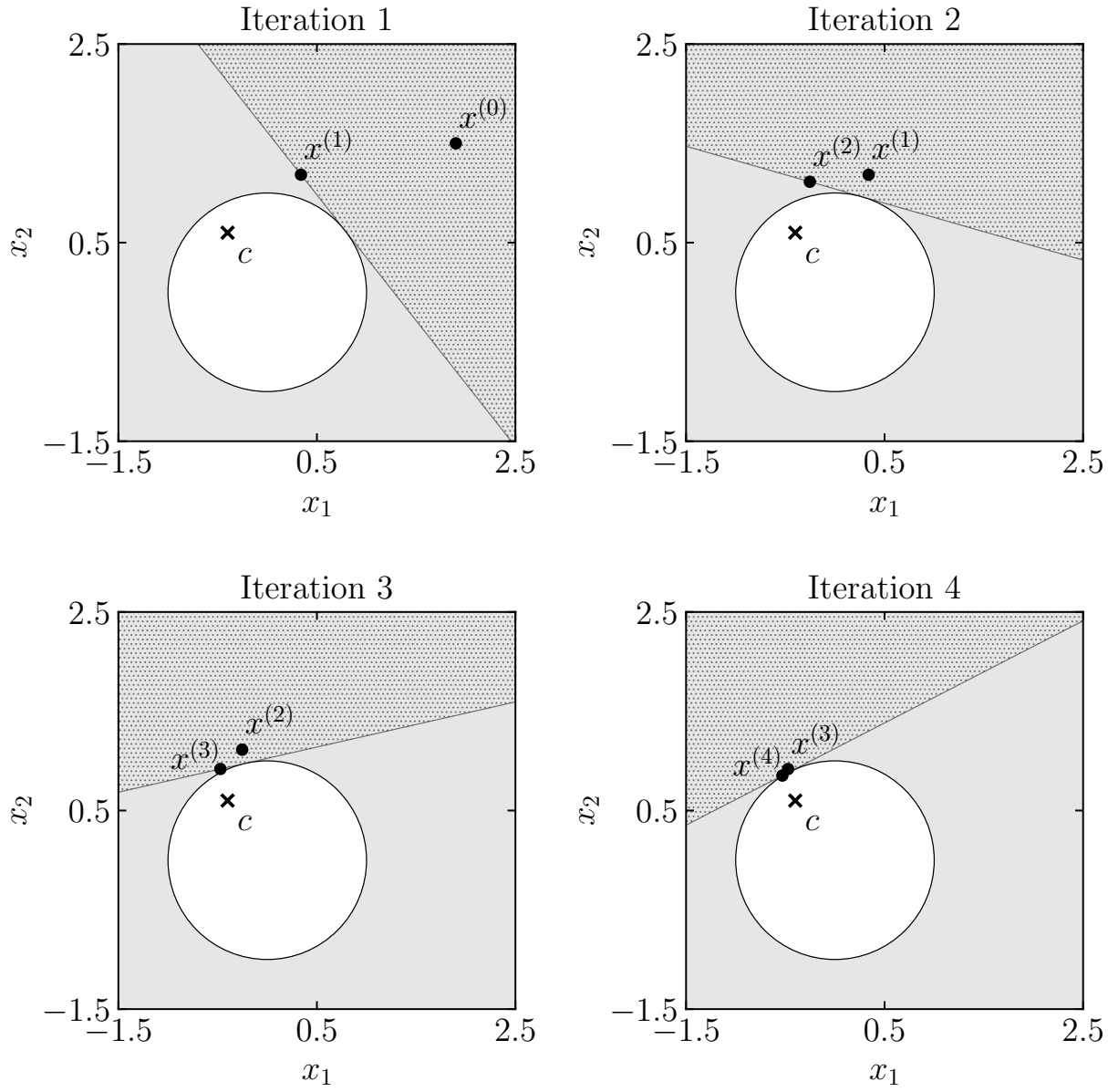
CCP iteration at $\tilde{x}$ solves

$$\begin{array}{ll} \text{minimize} & \|x - c\|_2 \\ \text{subject to} & (1/\|\tilde{x}\|_2)\tilde{x}^T x \geq 1 \end{array}$$

right shows example for $n = 2$,
10 random initializations:



results of initial 4 iterations:



Sequential convex approximation

given general nonlinear program with variable $x \in \mathbf{R}^n$

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

- $f_i \colon \mathbf{R}^n \to \mathbf{R}$ not necessarily convex for $i = 0, 1, \ldots, m$
- $h_i \colon \mathbf{R}^n \to \mathbf{R}$ not necessarily affine for $i = 1, \ldots, p$

basic idea: at current point $\tilde{x} \in \mathbf{R}^n$ iteratively form

- convex approximation $\hat{f}_{i,\tilde{x}}$ of f_i for $i = 0, 1, \ldots, m$
- affine approximation $\hat{h}_{i,\tilde{x}}$ of h_i for $i = 1, \ldots, p$
- over some *trust region* $\mathcal{T}_{\tilde{x}} \subseteq \mathbf{R}^n$ around $\tilde{x}$

so we end up solving a sequence of convex approximations

$$\begin{array}{ll} \text{minimize} & \hat{f}_{0,\tilde{x}}(x) \\ \text{subject to} & \hat{f}_{i,\tilde{x}}(x) \leq 0, \quad i = 1, \ldots, m \\ & \hat{h}_{i,\tilde{x}}(x) = 0, \quad i = 1, \ldots, p \\ & x \in \mathcal{T}_{\tilde{x}} \end{array}$$

trust region $\mathcal{T}_{\tilde{x}}$:

- Euclidean ball with center $\tilde{x}$ and radius $\rho > 0$:

$$\mathcal{T}_{\tilde{x}} = \{x \in \mathbf{R}^n \mid \|x - \tilde{x}\|_2 \leq \rho\}$$

- box centered at $\tilde{x}$ bounded by $\rho \in \mathbf{R}^n_{++}$:

$$\mathcal{T}_{\tilde{x}} = \{x \in \mathbf{R}^n \mid |x_i - \tilde{x}_i| \leq \rho_i,\ i = 1, \ldots, n\}$$

  - $\rho_i = \infty$ if x_i appears only in convex f_i or affine h_i

interpretation: the next point should not be too far from the current point $\tilde{x}$, so that the approximations are valid

Taylor expansions

for twice differentiable $f : \mathbf{R}^n \to \mathbf{R}$

affine approximation at $\tilde{x}$:

$$\hat{f}_{\tilde{x}}(x) = f(\tilde{x}) + \nabla f(\tilde{x})^T (x - \tilde{x})$$

convex approximation at $\tilde{x}$:

$$\hat{f}_{\tilde{x}}(x) = f(\tilde{x}) + \nabla f(\tilde{x})^T (x - \tilde{x}) + (1/2)(x - \tilde{x})^T H_+ (x - \tilde{x})$$

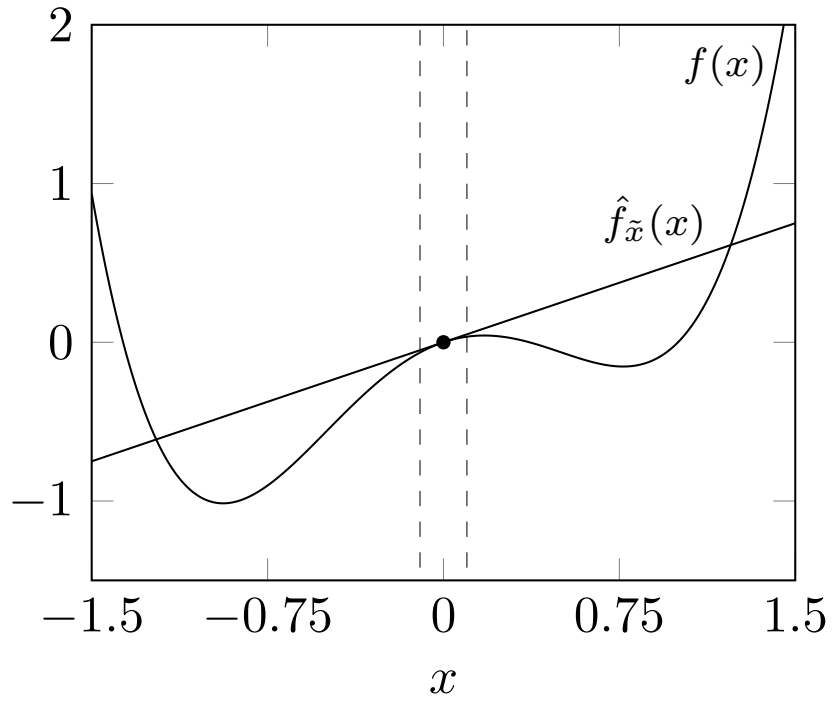
- $H_+ \succeq 0$ is positive semidefinite part of Hessian $\nabla^2 f(\tilde{x})$, *i.e.*,

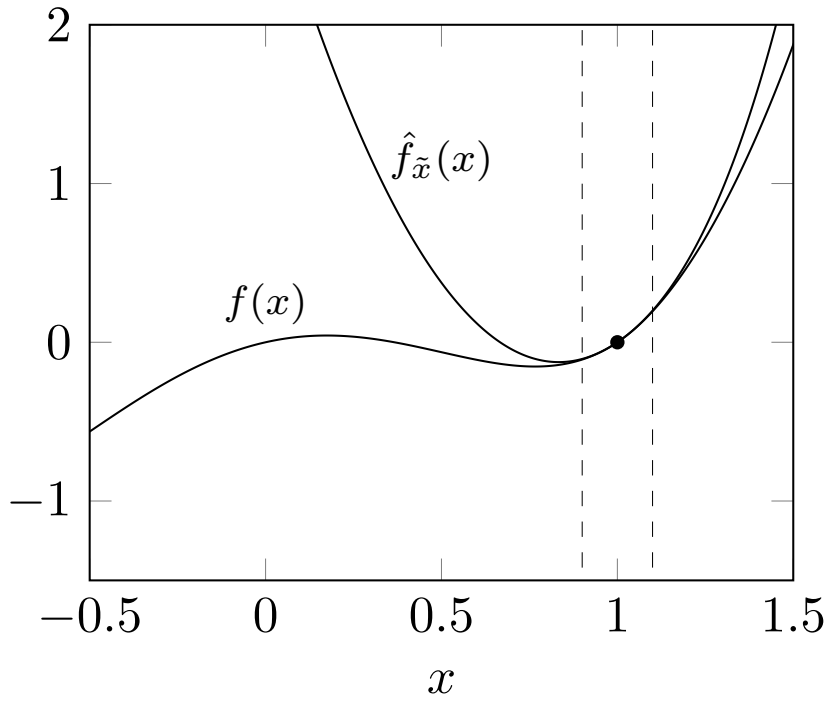
$$H_+ = Q \Lambda_+ Q^T$$

  where $\nabla^2 f(\tilde{x}) = Q \Lambda Q^T$,

$$\Lambda_+ = \mathbf{diag}(\max\{\lambda_1, 0\}, \ldots, \max\{\lambda_n, 0\})$$

examples:





Taylor expansions are **local** approximations, *i.e.*,

- if x is close to $\tilde{x}$, then approximation is very accurate
- approximation invariant to selection of $\mathcal{T}_{\tilde{x}}$
- but large $\mathcal{T}_{\tilde{x}}$ may lead to very poor approximation (in the whole $\mathcal{T}_{\tilde{x}}$)

Particle methods

given target function $f : \mathbf{R}^n \to \mathbf{R}$ and trust region $\mathcal{T}_{\tilde{x}}$:

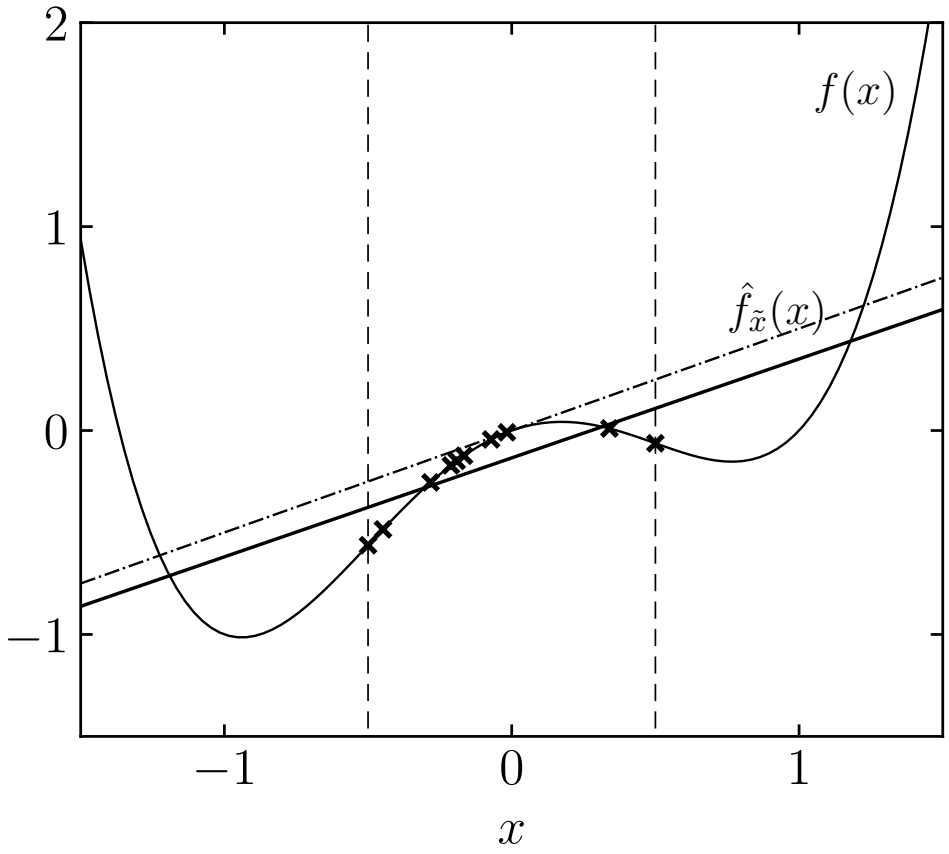
- sample K points $z_1, \ldots, z_K \in \mathcal{T}_{\tilde{x}}$, evaluate $y_i = f(z_i)$ for $i = 1, \ldots, K$
- so we have *particles* (z_i, y_i), $i = 1, \ldots, K$
- fit affine or convex function $\hat{f}_{\tilde{x}}$ to the particles (via, $e.g.$, least squares)

- allow approximation of nondifferentiable functions
- provide **'regional'** approximations, roughly speaking,
  - approximation is valid in the whole $\mathcal{T}_{\tilde{x}}$
  - accuracy not so sensitive to the size of $\mathcal{T}_{\tilde{x}}$ (compared to local approximations)
- approximation depends on $\mathcal{T}_{\tilde{x}}$ and the particles (z_i, y_i)
- needs extraordinary sampling and function evaluation

examples:

- affine approximation $\hat{f}_{\tilde{x}}(x) = a^T(x - \tilde{x}) + b$ fitted to particles (z_i, y_i):

$$\text{minimize} \quad \sum_{i=1}^{K} \left(a^T(z_i - \tilde{x}) + b - y_i\right)^2$$

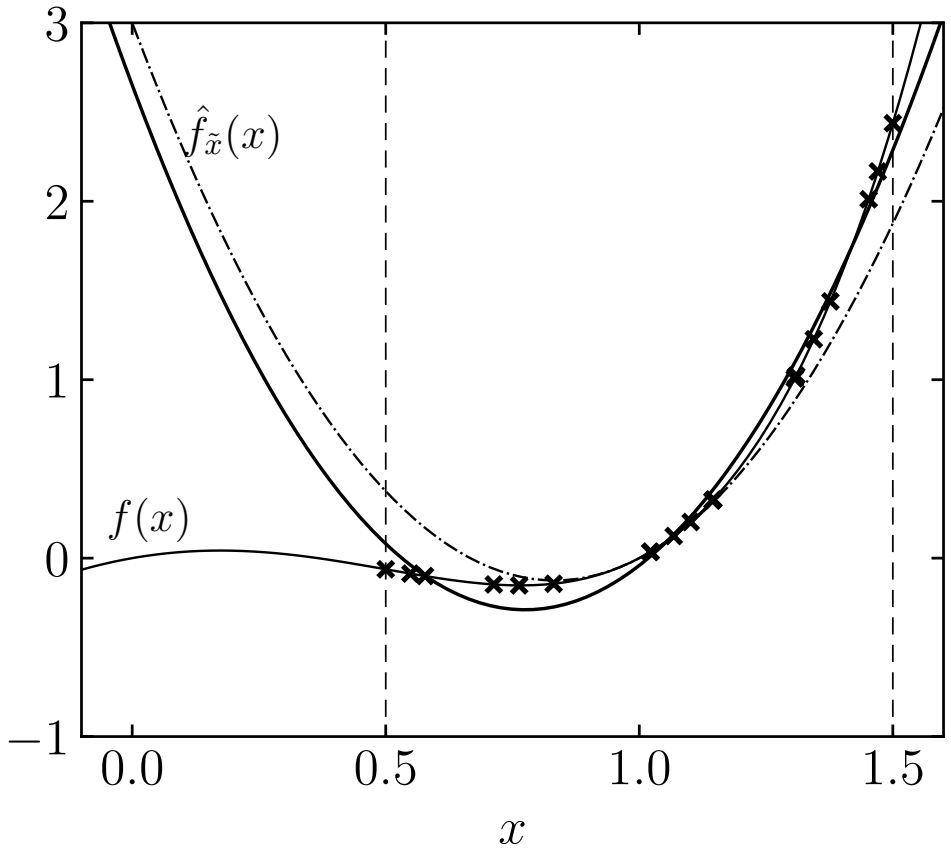


dashdotted line shows Taylor expansion at $\tilde{x}$

- fit convex quadratic $\hat{f}_{\tilde{x}}(x) = (1/2)(x-\tilde{x})^T P(x-\tilde{x}) + q^T(x-\tilde{x}) + r$:

  minimize $\sum_{i=1}^K \left((1/2)(z_i-\tilde{x})^T P(z_i-\tilde{x}) + q^T(z_i-\tilde{x}) + r - y_i\right)^2$
  subject to $P \succeq 0$

  (least squares problem with variables $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, $r \in \mathbf{R}$)



Fit convex envelope to particles

given particles $(z_i, y_i = f(z_i))$, $i = 1, \ldots, K$, fit $h\colon \mathbf{R}^n \to \mathbf{R}$,

$$h(x) = \max\{h_i + g_i^T(x - z_i) \mid i = 1, \ldots, K\}$$

with parameters $h_i \in \mathbf{R}$, $g_i \in \mathbf{R}^n$ for $i = 1, \ldots, K$

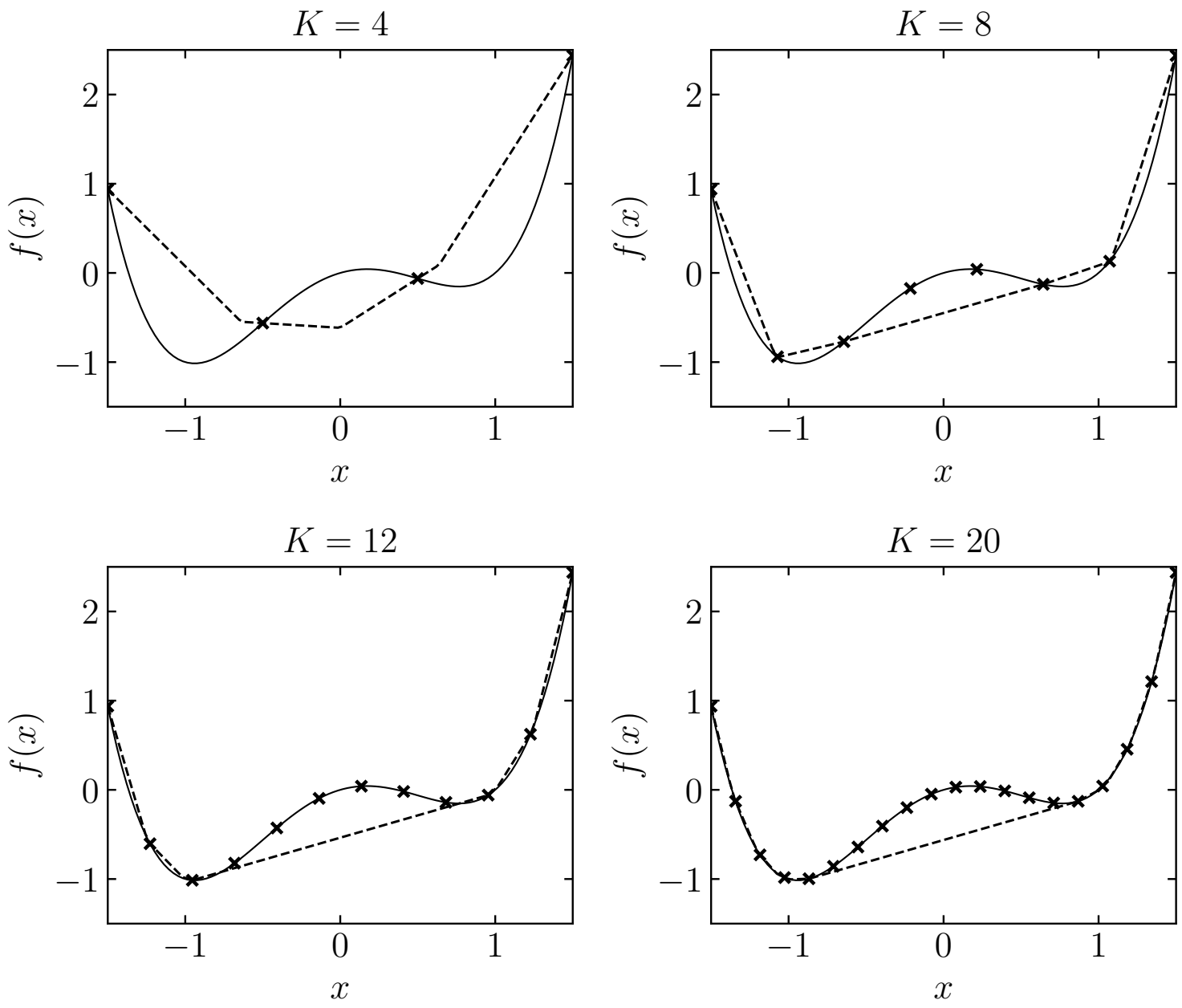
- h is always convex, since it is the pointwise maximum of affine functions
- require h lower bounds f at the particles, for $i = 1, \ldots, K$:

$$\begin{aligned} & y_i \geq h(z_i) = \max\{h_j + g_j^T(z_i - z_j) \mid j = 1, \ldots, K\} \\ \iff \quad & y_i \geq h_j + g_j^T(z_i - z_j), \quad j = 1, \ldots, K \end{aligned}$$

- maximize $\sum_{i=1}^K h_i$ make h as tight as possible (since $h(z_i) \geq h_i$)

put together, we have the LP

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^K h_i \\ \text{subject to} & y_i \geq h_j + g_j^T(z_i - z_j), \quad i,\ j = 1, \ldots, K \end{array}$$


$K = 4$
$K = 8$
$K = 12$
$K = 20$
$f(x)$
x
2
1
0
−1
−1
0
1

Example nonconvex QP

for the nonconvex QP with $P \in \mathbf{S}^n$ indefinite

$$\begin{array}{ll} \text{minimize} & f(x) = (1/2)x^T P x + q^T x \\ \text{subject to} & \|x\|_\infty \leq 1 \end{array}$$

iteratively solve convex approximations

$$\begin{array}{ll} \text{minimize} & f(\tilde{x}) + (P\tilde{x} + q)^T (x - \tilde{x}) + (1/2)(x - \tilde{x})^T P_+ (x - \tilde{x}) \\ \text{subject to} & \|x\|_\infty \leq 1 \\ & \|x - \tilde{x}\|_\infty \leq \rho \end{array}$$

- convex objective from 2nd-order Taylor approximation at $\tilde{x}$
- $P_+ \succeq 0$ is positive semidefinite part of P
- $\rho > 0$ is fixed trust region radius

example results for $n = 20$ from 10 random initializations:

