Simple Optimization Problems

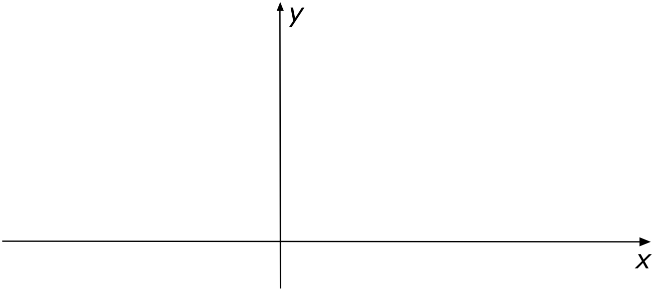
Antonio Frangioni

Department of Computer Science
University of Pisa
https://www.di.unipi.it/~frangio
mailto:frangio@di.unipi.it

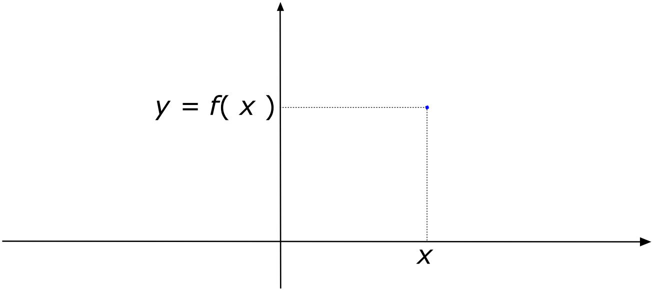
Computational Mathematics for Learning and Data Analysis
Master in Computer Science – University of Pisa

A.Y. 2024/25

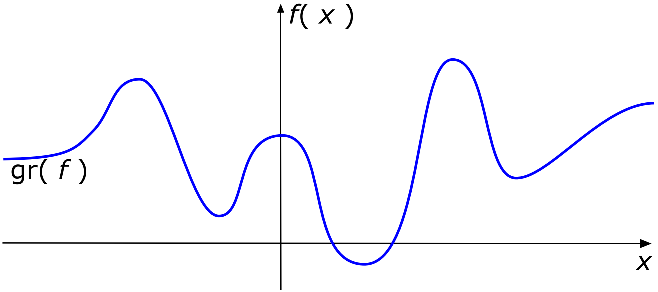
Outline



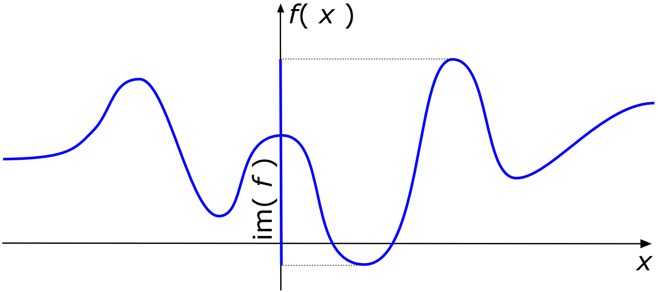
- Let's start simple: x input space, y output space, both just $\mathbb{R}$



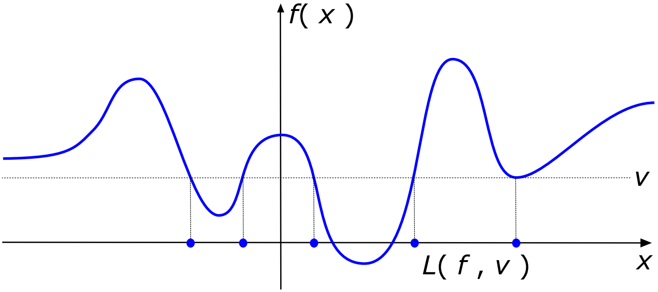
- Let's start simple: x input space, y output space, both just $\mathbb{R}$
- Function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = y =$ output with input x



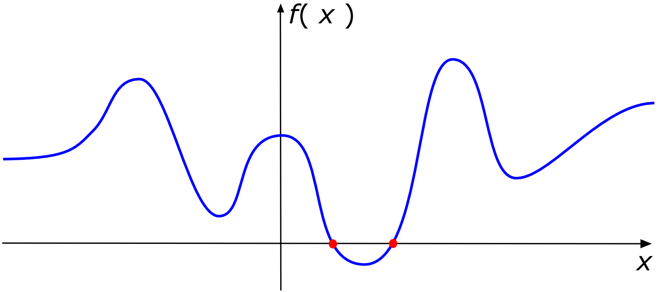
- Let's start simple: x input space, y output space, both just $\mathbb{R}$
- Function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = y =$ output with input x
- Graph of f: $\mathrm{gr}(f) = \{ (f(x), x) : x \in \mathbb{R} \} \subset \mathbb{R}^2$



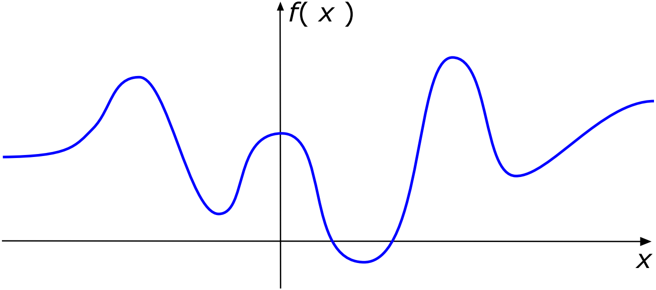
- Let's start simple: x input space, y output space, both just $\mathbb{R}$
- Function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = y =$ output with input x
- Graph of f: $\mathrm{gr}(f) = \{ (f(x), x) \ : \ x \in \mathbb{R} \} \subset \mathbb{R}^2$
- Image of f: $\mathrm{im}(f) = \{ y \ : \ \exists x \in \mathbb{R} \text{ s.t. } y = f(x) \} \subset \mathbb{R}$
  i.e., projection of gr(f) on output space (a.k.a. co-domain)



- Let's start simple: x input space, y output space, both just $\mathbb{R}$
- Function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = y =$ output with input x
- Graph of f: $\mathrm{gr}(f) = \{ (f(x), x) \ : \ x \in \mathbb{R} \} \subset \mathbb{R}^2$
- Image of f: $\mathrm{im}(f) = \{ y \ : \ \exists x \in \mathbb{R} \text{ s.t. } y = f(x) \} \subset \mathbb{R}$
  i.e., projection of gr(f) on output space (a.k.a. co-domain)
- Level set at value v: $L(f, v) = \{ x \in \mathbb{R} \ : \ f(x) = v \} \subset \mathbb{R}$



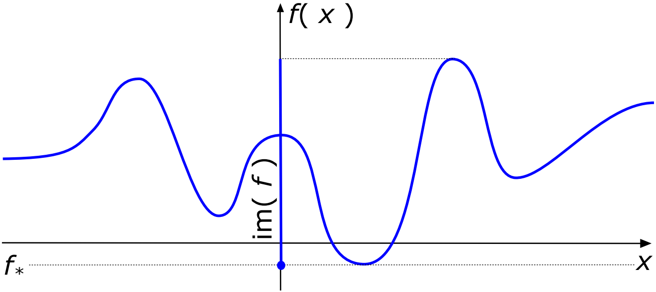
- Let's start simple: x input space, y output space, both just $\mathbb{R}$
- Function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = y =$ output with input x
- Graph of f: $\mathrm{gr}(f) = \{ (f(x), x) \ : \ x \in \mathbb{R} \} \subset \mathbb{R}^2$
- Image of f: $\mathrm{im}(f) = \{ y \ : \ \exists x \in \mathbb{R} \text{ s.t. } y = f(x) \} \subset \mathbb{R}$
  i.e., projection of $\mathrm{gr}(f)$ on output space (a.k.a. co-domain)
- Level set at value v: $L(f, v) = \{ x \in \mathbb{R} \ : \ f(x) = v \} \subset \mathbb{R}$
  (roots of $f = L(f, 0) =$ level set at value 0)



- f objective (function) of (univariate, unconstrained) optimization problem

  $(P) \quad f_* = \min\{\, f(x) \,:\, x \in \mathbb{R} \,\}$

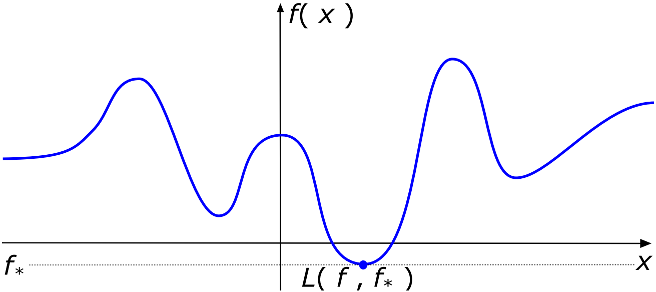
- $f_* = \nu(P)$ optimal value (unique if $\exists$, which it may not)



- f objective (function) of (univariate, unconstrained) optimization problem

$$(P) \quad f_* = \min\{\, f(x) \,:\, x \in \mathbb{R} \,\}$$

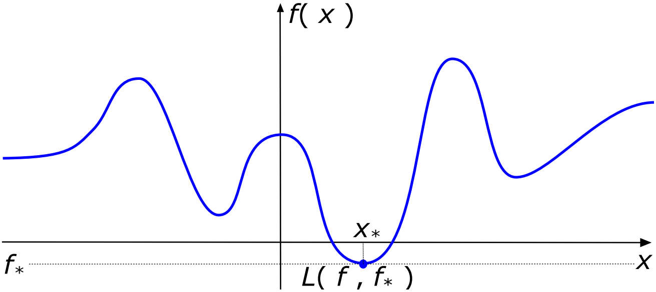
- $f_* = \nu(P)$ optimal value (unique if $\exists$, which it may not)
- f_* = smaller element of im(f)



- f objective (function) of (univariate, unconstrained) optimization problem

  $(P) \quad f_* = \min\{ f(x) : x \in \mathbb{R} \}$

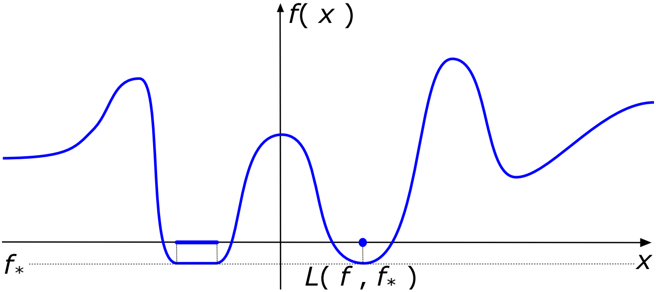
- $f_* = \nu(P)$ optimal value (unique if $\exists$, which it may not)
- f_* = smaller element of $\mathrm{im}(f)$ = smaller v s.t. $L(f, v) \neq \emptyset$



- f objective (function) of (univariate, unconstrained) optimization problem

$$(P) \quad f_* = \min\{\, f(\,x\,) \,:\, x \in \mathbb{R}\,\}$$

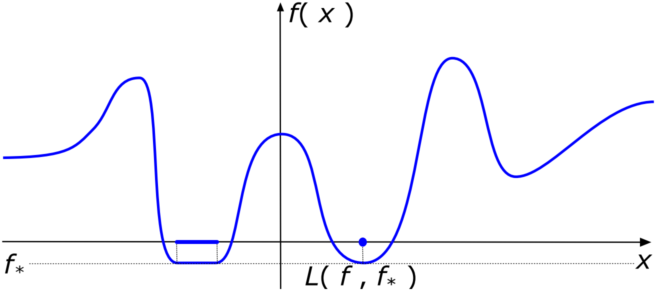
- $f_* = \nu(\,P\,)$ optimal value (unique if $\exists$, which it may not)
- f_* = smaller element of $\mathrm{im}(\,f\,)$ = smaller v s.t. $L(\,f\,,\,v\,) \neq \emptyset$
- In fact, the problem is $\quad (P) \quad x_* \in \mathrm{argmin}\,\{\, f(\,x\,) \,:\, x \in \mathbb{R}\,\}$
- x_* s.t. $f_* = f(\,x_*\,) \leq f(\,x\,) \;\; \forall x \in \mathbb{R}$ optimal solution (if $\exists$, which it may not)



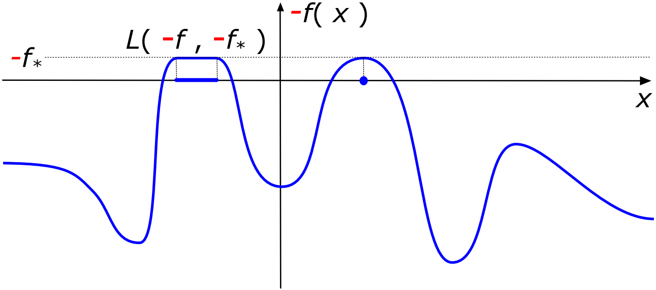
- f objective (function) of (univariate, unconstrained) optimization problem

  $$(P) \quad f_* = \min\{ f(x) : x \in \mathbb{R} \}$$

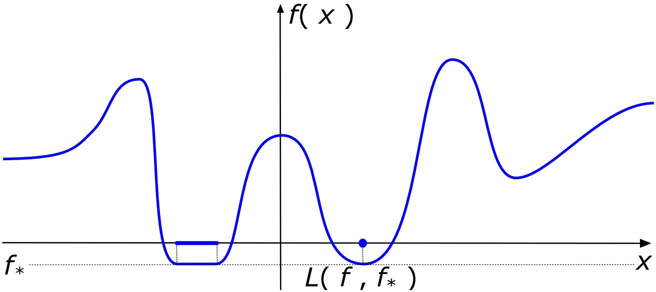
- $f_* = \nu(P)$ optimal value (unique if $\exists$, which it may not)
- f_* = smaller element of $\mathrm{im}(f)$ = smaller v s.t. $L(f, v) \neq \emptyset$
- In fact, the problem is $\quad (P) \quad x_* \in \mathrm{argmin}\{ f(x) : x \in \mathbb{R} \}$
- x_* s.t. $f_* = f(x_*) \leq f(x) \;\; \forall x \in \mathbb{R}$ optimal solution (if $\exists$, which it may not)
- x_* may not be unique: $\exists\, x' \neq x_* \in L(f, f_*) = X_*$ set of optimal solutions, but we don't care (mostly): all optimal solutions equivalent "in the eyes of f"



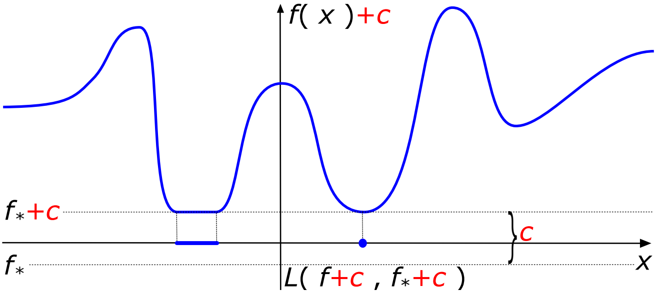
- Sometimes changing f changes f_* "in a simple way" but does not change X_*: the corresponding problem is equivalent, a reformulation of (P)
- "min" w.l.o.g.: $\min\{\, f(x) \,:\, x \in \mathbb{R} \,\} =$



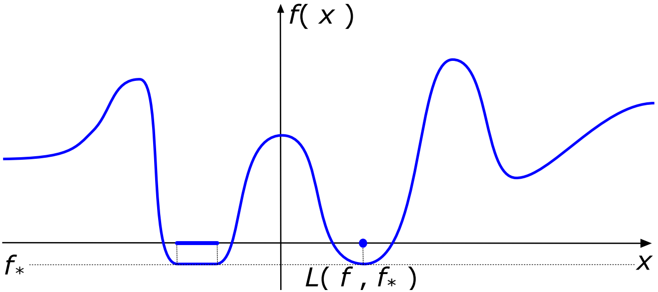
- Sometimes changing f changes f_* "in a simple way" but does not change X_*: the corresponding problem is equivalent, a reformulation of (P)
- "min" w.l.o.g.: $\min\{\, f(x) \,:\, x \in \mathbb{R}\,\} = -\max\{\, -f(x) \,:\, x \in \mathbb{R}\,\}$
  i.e., $\operatorname{argmin}\{\, f(x) \,:\, x \in \mathbb{R}\,\} = \operatorname{argmax}\{\, -f(x) \,:\, x \in \mathbb{R}\,\}$
  (but $\min\{\, f(x)\,\} \neq \max\{\, f(x)\,\}$, often rather different problems)



- Sometimes changing f changes f_* "in a simple way" but does not change X_*: the corresponding problem is equivalent, a reformulation of (P)
- "min" w.l.o.g.: $\min\{\, f(x) \,:\, x \in \mathbb{R} \,\} = -\max\{\, -f(x) \,:\, x \in \mathbb{R} \,\}$
  i.e., $\operatorname{argmin}\{\, f(x) \,:\, x \in \mathbb{R} \,\} = \operatorname{argmax}\{\, -f(x) \,:\, x \in \mathbb{R} \,\}$
  (but $\min\{\, f(x) \,\} \neq \max\{\, f(x) \,\}$, often rather different problems)
- Analogously, $\min\{\, f(x) + c \,:\, x \in \mathbb{R} \,\} =$



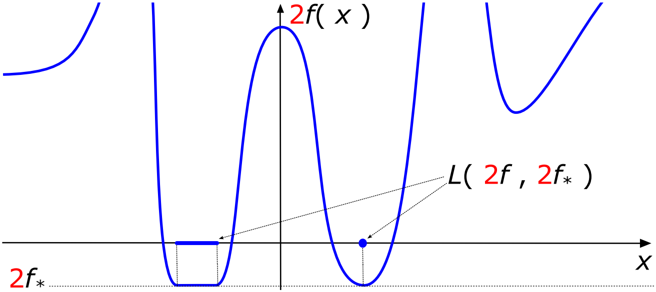
- Sometimes changing f changes f_* "in a simple way" but does not change X_*: the corresponding problem is equivalent, a reformulation of (P)
- "min" w.l.o.g.: $\min\{\, f(x) \,:\, x \in \mathbb{R} \,\} = -\max\{\, -f(x) \,:\, x \in \mathbb{R} \,\}$
  i.e., $\operatorname{argmin}\,\{\, f(x) \,:\, x \in \mathbb{R} \,\} = \operatorname{argmax}\,\{\, -f(x) \,:\, x \in \mathbb{R} \,\}$
  (but $\min\{\, f(x) \,\} \neq \max\{\, f(x) \,\}$, often rather different problems)
- Analogously, $\min\{\, f(x) + c \,:\, x \in \mathbb{R} \,\} = c + \min\{\, f(x) \,:\, x \in \mathbb{R} \,\}$
  i.e., $\operatorname{argmin}\,\{\, f(x) + c \,:\, x \in \mathbb{R} \,\} = \operatorname{argmin}\,\{\, f(x) \,:\, x \in \mathbb{R} \,\} = X_*$



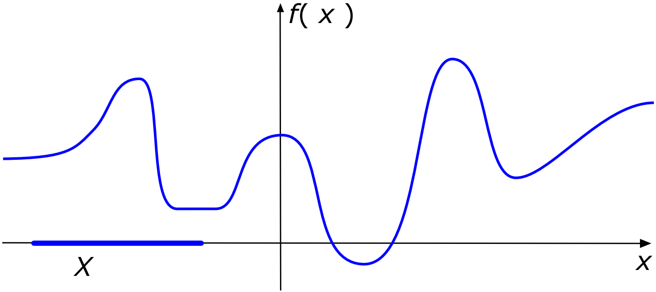
- Sometimes changing f changes f_* "in a simple way" but does not change X_*: the corresponding problem is equivalent, a reformulation of (P)
- "min" w.l.o.g.: $\min\{\, f(\, x\,) \;:\; x \in \mathbb{R} \,\} = -\max\{\, -f(\, x\,) \;:\; x \in \mathbb{R} \,\}$

  i.e., $\operatorname{argmin}\{\, f(\, x\,) \;:\; x \in \mathbb{R} \,\} = \operatorname{argmax}\{\, -f(\, x\,) \;:\; x \in \mathbb{R} \,\}$

  (but $\min\{\, f(\, x\,)\,\} \neq \max\{\, f(\, x\,)\,\}$, often rather different problems)
- Analogously, $\min\{\, cf(\, x\,) \;:\; x \in \mathbb{R} \,\} =$

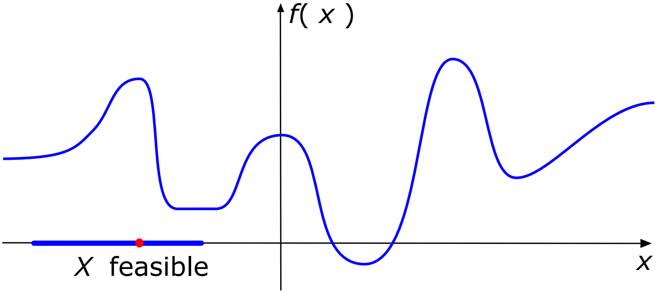


- Sometimes changing f changes f_* "in a simple way" but does not change X_*: the corresponding problem is equivalent, a reformulation of (P)
- "min" w.l.o.g.: $\min\{ f(x) : x \in \mathbb{R} \} = - \max\{ -f(x) : x \in \mathbb{R} \}$
  i.e., $\text{argmin}\ \{ f(x) : x \in \mathbb{R} \} = \text{argmax}\ \{ -f(x) : x \in \mathbb{R} \}$
  (but $\min\{ f(x) \} \neq \max\{ f(x) \}$, often rather different problems)
- Analogously, $\min\{ cf(x) : x \in \mathbb{R} \} = c \min\{ f(x) : x \in \mathbb{R} \}$ (if $c > 0$)
  i.e., $\text{argmin}\ \{ cf(x) : x \in \mathbb{R} \} = \text{argmin}\ \{ f(x) : x \in \mathbb{R} \} = X_*$



- More general: feasible region any set X ($\subseteq \mathbb{R}$), objective $f : X \to \mathbb{R}$

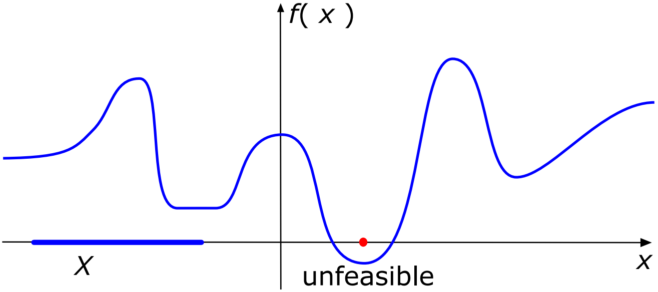
$$(P) \quad f_* = \min\{\, f(x) \;:\; x \in X \,\}$$ constrained optimization problem



- More general: feasible region any set X ($\subseteq \mathbb{R}$), objective $f : X \to \mathbb{R}$

$$(P) \quad f_* = \min\{\, f(x) \;:\; x \in X \,\}$$ constrained optimization problem

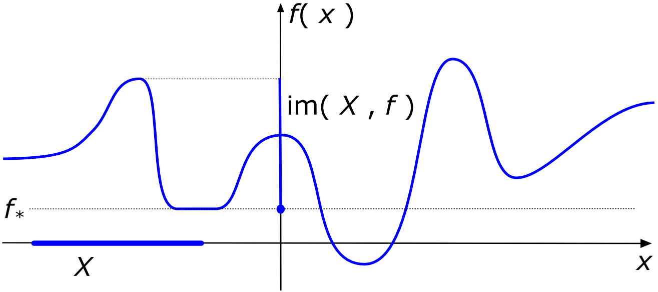
- $x \in X$ feasible solution;



- More general: feasible region any set X ($\subseteq \mathbb{R}$), objective $f : X \to \mathbb{R}$

$$(P) \quad f_* = \min\{\, f(x) \;:\; x \in X \,\}$$ constrained optimization problem

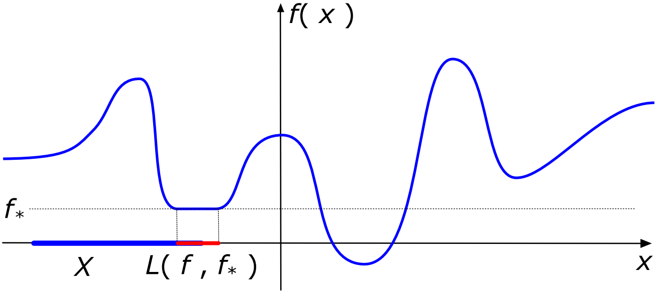
- $x \in X$ feasible solution; $x \in \mathbb{R} \setminus X$ unfeasible solution



- More general: feasible region any set X ($\subseteq \mathbb{R}$), objective $f : X \rightarrow \mathbb{R}$

$$(P) \quad f_* = \min\{ f(x) \; : \; x \in X \} \quad \text{constrained optimization problem}$$

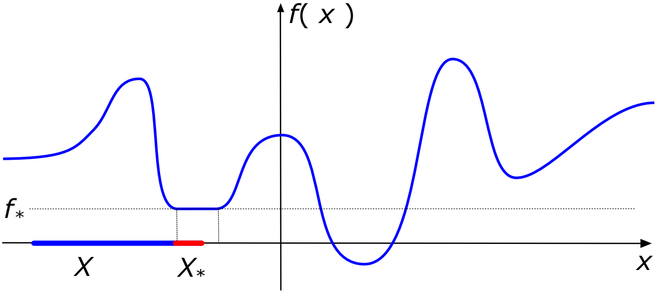
- $x \in X$ feasible solution; $x \in \mathbb{R} \setminus X$ unfeasible solution
- $f_* = \nu(P) = \min(im(X, f)) =$ smaller element of image of X through f



- More general: feasible region any set X ($\subseteq \mathbb{R}$), objective $f : X \to \mathbb{R}$

  $(P) \quad f_* = \min\{\, f(x) \;:\; x \in X \,\}$ constrained optimization problem

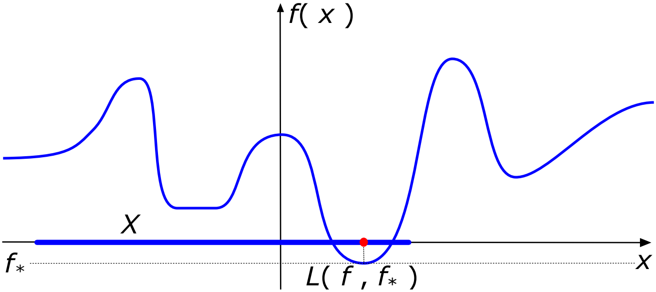
- $x \in X$ feasible solution; $x \in \mathbb{R} \setminus X$ unfeasible solution
- $f_* = \nu(P) = \min(\, im(X, f)\,) =$ smaller element of image of X through f
- $X_* = L(f, f_*)$



- More general: feasible region any set X ($\subseteq \mathbb{R}$), objective $f : X \to \mathbb{R}$

  $$(P) \quad f_* = \min\{\, f(\, x\,) \;:\; x \in X \,\}$$ constrained optimization problem

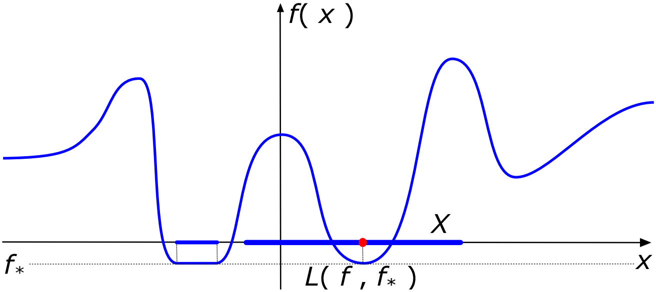
- $x \in X$ feasible solution; $x \in \mathbb{R} \setminus X$ unfeasible solution
- $f_* = \nu(\, P\,) = \min(\, im(\, X\,,\, f\,)\,)$ = smaller element of image of X through f
- $X_* = L(\, f\,,\, f_*) \cap X$: set of best feasible solutions



- More general: feasible region any set X ($\subseteq \mathbb{R}$), objective $f : X \to \mathbb{R}$

  $$(P) \quad f_* = \min\{ f(x) \; : \; x \in X \}$$ constrained optimization problem

- $x \in X$ feasible solution; $x \in \mathbb{R} \setminus X$ unfeasible solution
- $f_* = \nu(P) = \min(\, im(X, f)\,) =$ smaller element of image of X through f
- $X_* = L(f, f_*) \cap X$: set of best feasible solutions
- X can be "useless" (X_* same)

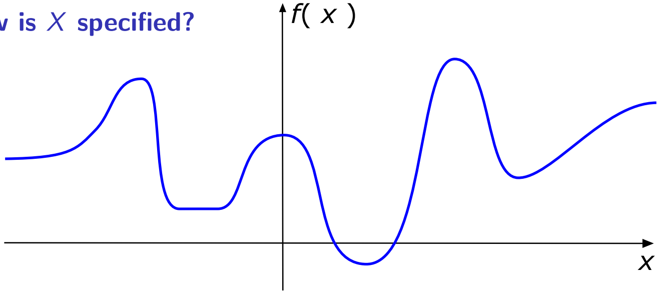


- More general: feasible region any set X ($\subseteq \mathbb{R}$), objective $f : X \to \mathbb{R}$

$$(P) \quad f_* = \min\{\, f(\,x\,) \ : \ x \in X \,\}$$ constrained optimization problem

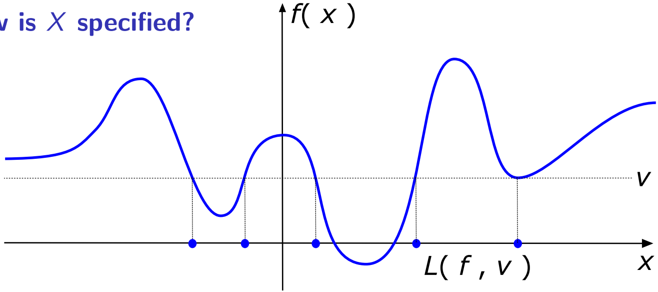
- $x \in X$ feasible solution; $x \in \mathbb{R} \setminus X$ unfeasible solution
- $f_* = \nu(\,P\,) = \min(\,im(\,X\,,\,f\,)\,) =$ smaller element of image of X through f
- $X_* = L(\,f\,,\,f_*) \cap X$: set of best feasible solutions
- X can be "useless" (X_* same) or partly so (f_* same) $\Longrightarrow$ makes sense to study the unconstrained case $X = \mathbb{R}$ first

Anyhow, how is X specified?



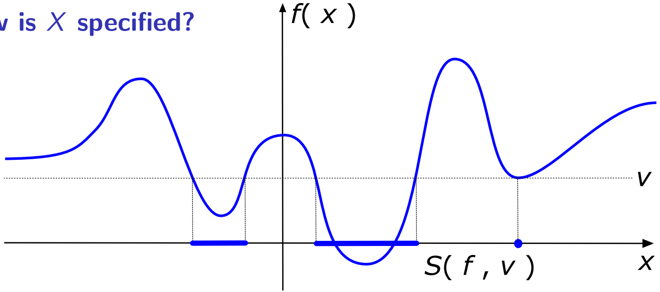
- The "abstract constraint $x \in X$" need be specified somehow
- Often useful to represent a set via (more than) one function(s)

Anyhow, how is X specified?



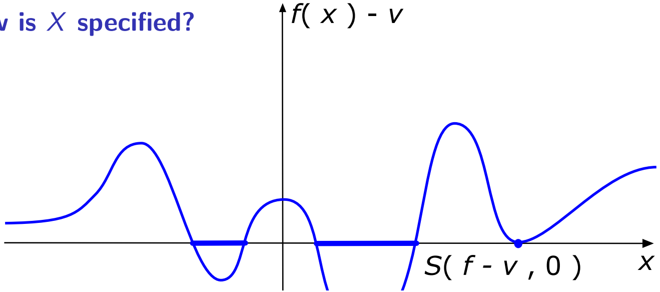
- The "abstract constraint $x \in X$" need be specified somehow
- Often useful to represent a set via (more than) one function(s)
- Standard ways: equality constraint $g(x) = v \equiv X =$ level set $L(g, v)$,

Anyhow, how is X specified?



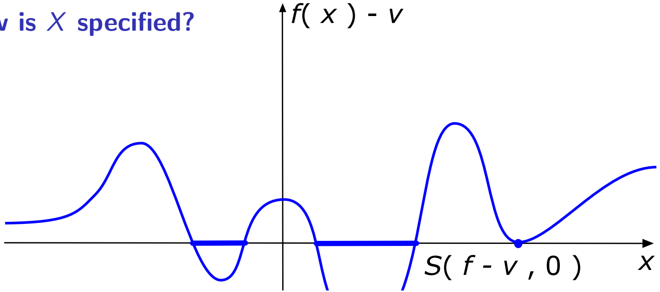
- The "abstract constraint $x \in X$" need be specified somehow
- Often useful to represent a set via (more than) one function(s)
- Standard ways: equality constraint $g(x) = v \;\equiv\; X =$ level set $L(g, v)$, inequality constraint $g(x) \leq v \;\equiv\;$ sublevel set $S(g, v) = \{ x : g(x) \leq v \}$

Anyhow, how is X specified?



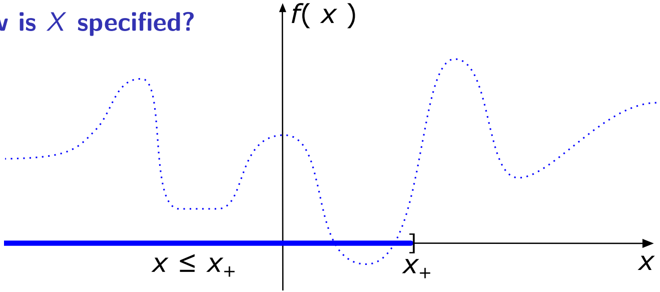
- The "abstract constraint $x \in X$" need be specified somehow
- Often useful to represent a set via (more than) one function(s)
- Standard ways: equality constraint $g(x) = v \equiv X =$ level set $L(g, v)$, inequality constraint $g(x) \leq v \equiv$ sublevel set $S(g, v) = \{ x : g(x) \leq v \}$
- For convenience "v hidden in f" $\Longrightarrow f(x) \leq 0 \ , \ f(x) = 0$
- What if one rather wants $g(x) \geq 0$? Simply $-g(x) \leq 0$

Anyhow, how is X specified?



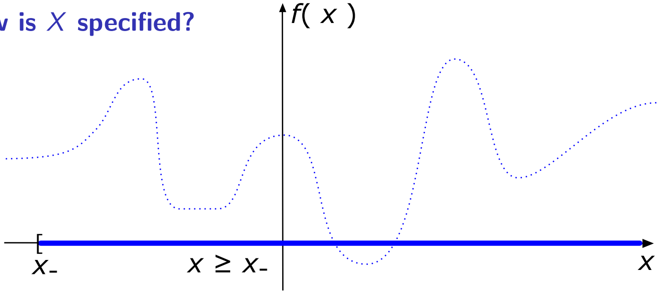
- The "abstract constraint $x \in X$" need be specified somehow
- Often useful to represent a set via (more than) one function(s)
- Standard ways: equality constraint $g(x) = v \;\equiv\; X =$ level set $L(g, v)$, inequality constraint $g(x) \leq v \;\equiv\;$ sublevel set $S(g, v) = \{ x : g(x) \leq v \}$
- For convenience "v hidden in f" $\Longrightarrow\; f(x) \leq 0 \;,\; f(x) = 0$
- What if one rather wants $g(x) \geq 0$? Simply $-g(x) \leq 0$
- Usually multiple constraints: "$g_1(x) \leq 0 \;,\; g_2(x) \leq 0$" $\equiv$ logical conjunction ("first condition and second condition") $\equiv$ intersection of (sub)level sets

Anyhow, how is X specified?



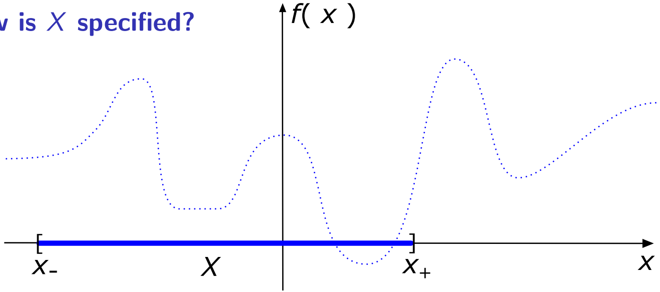
- The "abstract constraint $x \in X$" need be specified somehow
- Often useful to represent a set via (more than) one function(s)
- Standard ways: equality constraint $g(x) = v \;\equiv\; X =$ level set $L(g, v)$, inequality constraint $g(x) \leq v \;\equiv\;$ sublevel set $S(g, v) = \{ x : g(x) \leq v \}$
- For convenience "v hidden in f" $\Longrightarrow\; f(x) \leq 0 \;,\; f(x) = 0$
- What if one rather wants $g(x) \geq 0$? Simply $-g(x) \leq 0$
- Usually multiple constraints: "$g_1(x) \leq 0 \;,\; g_2(x) \leq 0$" $\equiv$ logical conjunction ("first condition and second condition") $\equiv$ intersection of (sub)level sets
- Simple and common: bounds $x \leq x_+$ (up)

Anyhow, how is X specified?



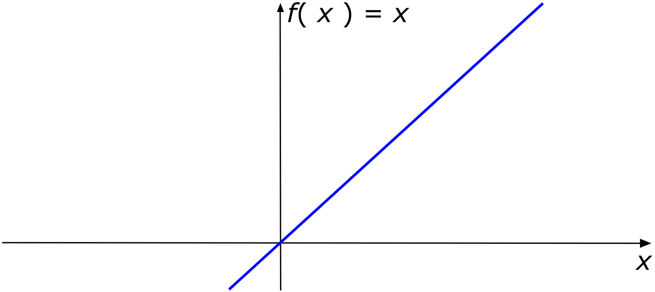
- The "abstract constraint $x \in X$" need be specified somehow
- Often useful to represent a set via (more than) one function(s)
- Standard ways: equality constraint $g(x) = v \equiv X =$ level set $L(g, v)$, inequality constraint $g(x) \leq v \equiv$ sublevel set $S(g, v) = \{ x : g(x) \leq v \}$
- For convenience "v hidden in f" $\Longrightarrow f(x) \leq 0$, $f(x) = 0$
- What if one rather wants $g(x) \geq 0$? Simply $-g(x) \leq 0$
- Usually multiple constraints: "$g_1(x) \leq 0$, $g_2(x) \leq 0$" $\equiv$ logical conjunction ("first condition and second condition") $\equiv$ intersection of (sub)level sets
- Simple and common: bounds $x \leq x_+$ (up) / $x \geq x_-$ (dn),

Anyhow, how is X specified?

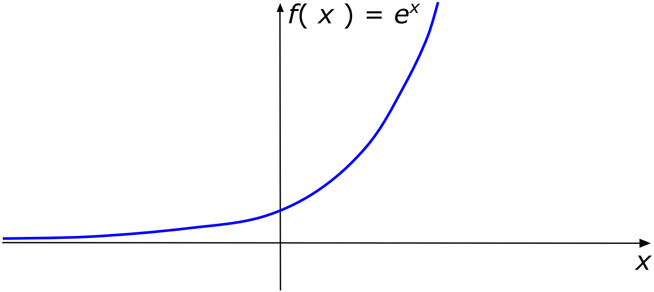


- The "abstract constraint $x \in X$" need be specified somehow
- Often useful to represent a set via (more than) one function(s)
- Standard ways: equality constraint $g(x) = v \;\equiv\; X =$ level set $L(g, v)$, inequality constraint $g(x) \leq v \;\equiv\;$ sublevel set $S(g, v) = \{ x : g(x) \leq v \}$
- For convenience "v hidden in f" $\Longrightarrow\; f(x) \leq 0 \;,\; f(x) = 0$
- What if one rather wants $g(x) \geq 0$? Simply $-g(x) \leq 0$
- Usually multiple constraints: "$g_1(x) \leq 0 \;,\; g_2(x) \leq 0$" $\equiv$ logical conjunction ("first condition and second condition") $\equiv$ intersection of (sub)level sets
- Simple and common: bounds $x \leq x_+$ (up) / $x \geq x_-$ (dn), boxes $x_- \leq x \leq x_+$

Outline

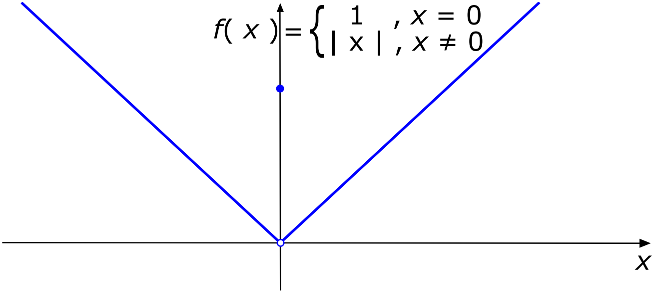


- f has no minimum, (P) unbounded (below): $f_* = \nu(\,P\,) = -\infty$
- Just a convenient shorthand for $\forall\, t \in \mathbb{R}\ \exists\, x \in \mathbb{R}$ s.t. $f(\,x\,) \leq t$
  i.e., "there is no (finite) lower bound on $\mathrm{im}(\,f\,)$"
- Solving (P) actually (at least) two different things:
  - finding x_* and proving it is optimal (how??)
  - constructively proving f unbounded below (how??)
- Hardly ever happens in learning since $\mathcal{L}(\,w\,) \geq 0$
- Nontrivial and important in optimization (tied with duality, nonemptiness, ...)



- im(f) is bounded below but has no minimum
- Either "naturally"

What if $f_* \exists$ but $x_* \nexists$?



- im(f) is bounded below but has no minimum
- Either "naturally" or "forcibly"
- $\inf\{ f(x) : x \in \mathbb{R} \} \exists$, but $\min\{ f(x) : x \in \mathbb{R} \} \nexists$
- Arguably $f_* = \inf\{ f(x) : x \in \mathbb{R} \}$, but $\nexists x_*$ s.t. $f_* = f(x_*)$
- im(f) is open, does not contain its boundary (will see)

- $\mathbb{R}$ totally ordered $\Longrightarrow \forall\, x, y \in \mathbb{R}$, at least one among $x \leq y$, $y \leq x$ holds
- $S \subseteq \mathbb{R}$, $\underline{s} = \inf S \iff \underline{s} \leq s \;\; \forall s \in S \;\; \wedge \;\; \forall t > \underline{s} \, \exists\, s \in S$ s.t. $s \leq t$
  $\bar{s} = \sup S \iff \bar{s} \geq s \;\; \forall s \in S \;\; \wedge \;\; \forall t < \bar{s} \, \exists\, s \in S$ s.t. $s \geq t$
- $\underline{s} \in S \Longrightarrow \underline{s} = \min S$, $\bar{s} \in S \Longrightarrow \bar{s} = \max S$
- Issues: i) $\inf S / \sup S$ may not $\exists$ in $\mathbb{R}$, ii) $\inf S / \sup S$ may not $\in S$
- Should write "$\inf\{ f(x) \ldots$", but we want (approximately) optimal solutions
- Set of extended reals: $\overline{\mathbb{R}} = \{ -\infty \} \cup \mathbb{R} \cup \{ +\infty \}$ (usually just $\mathbb{R}$)
- For all $S \subseteq \mathbb{R}$, $\exists$ $\sup / \inf S \in \overline{\mathbb{R}}$
- $\inf S = -\infty \iff \forall\, t \in \mathbb{R} \, \exists\, s \in S$ s.t. $s \leq t$
  $\sup S = +\infty \iff \forall\, t \in \mathbb{R} \, \exists\, s \in S$ s.t. $s \geq t$
  just a convenient shorthand for "there is no (finite) inf / sup"
- $\inf \emptyset = \infty$, $\sup \emptyset = -\infty$

- Several ways to ensure this never happens (hypotheses on f, X)
- On computers "$x \in \mathbb{R}$" typically is "$x \in \mathbb{Q}$" with up to 16 digits precision $\Longrightarrow$ approximation errors unavoidable anyway
- Exact algebraic computation may be possible (if f is algebraic, which it may be not) but anyway usually too slow
- In fact learning going the opposite way (`float`, `half`, `FP8`, ...)
- Anyway, finding the exact x_* impossible in general [4, p. 408]
- For any fixed $\varepsilon > 0$, plenty of ε-approximate solutions (ε-optima):

  $x_\varepsilon \in \mathbb{R}$ s.t. $f_* \leq f(x_\varepsilon) \leq f_* + \varepsilon$

  "as close to the optimal solution (value) as you want"
- Cost of solution algorithms typically depends on ε (sometimes very badly)
- And ε can't really become very small anyway (see above)

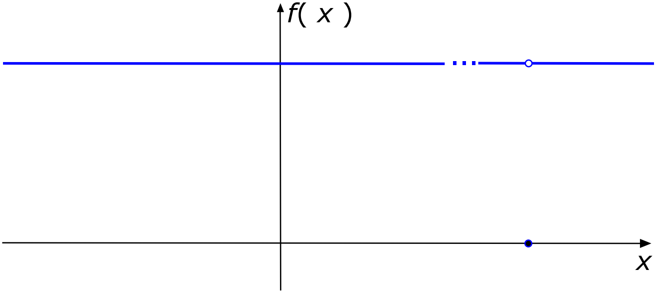
- Absolute gap: $A(x) = f(x) - f_* \ (\geq 0)$
- Relative gap: $R(x) = (f(x) - f_*) / |f_*| = A(x) / |f_*| \ (\geq 0)$
- Why $R(x)$? Because $\forall \alpha > 0 \quad (P) \equiv (P_\alpha) \ \min\{ \alpha f(x) : x \in \mathbb{R} \}$
  $\nu(P_\alpha) = \alpha f_* = \alpha \nu(P) \implies$ same $R(x)$ (scale invariant), different $A(x)$

Exercise: $R(x)$ ill-defined if $f_* = 0$, propose solutions & justify them (change f_*)

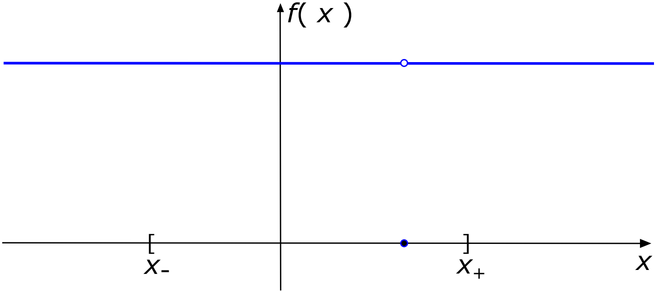
- Absolute gap: $A(x) = f(x) - f_* \ (\geq 0)$
- Relative gap: $R(x) = (f(x) - f_*) / |f_*| = A(x) / |f_*| \ (\geq 0)$
- Why $R(x)$? Because $\forall \alpha > 0 \quad (P) \equiv (P_\alpha) \ \min\{ \alpha f(x) : x \in \mathbb{R} \}$
  $\nu(P_\alpha) = \alpha f_* = \alpha \nu(P) \implies$ same $R(x)$ (scale invariant), different $A(x)$

Exercise: $R(x)$ ill-defined if $f_* = 0$, propose solutions & justify them (change f_*)

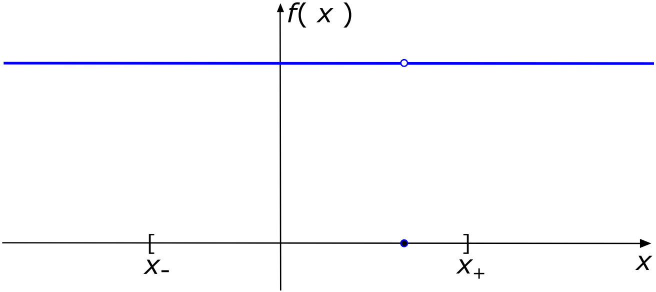
- (Approximately) solve (P): fix ε, find x s.t. either $A(x) \leq \varepsilon$ or $R(x) \leq \varepsilon$
- Issue: computing $A(x)$ or $R(x)$ requires f_* which is typically unkown
- Could argue this is “the issue” in optimization: compute (an estimate of) f_*
- Sometimes $\approx$ known in learning ($f_* \approx 0$ in NN, but not in SVM)
- A real issue only if global optimum x_* needed, hence not always



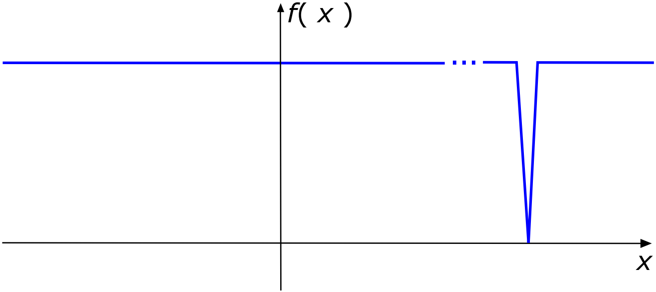
- Impossible because isolated minima can be anywhere [4, p. 408]



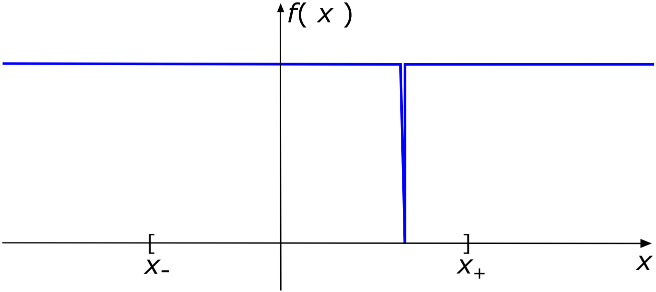
- Impossible because isolated minima can be anywhere [4, p. 408]
- Does it help restricting to $x \in X = [\, x_- \,,\, x_+ \,]$ $(-\infty < x_- < x_+ < +\infty)$?
- No: still uncountably many points to try



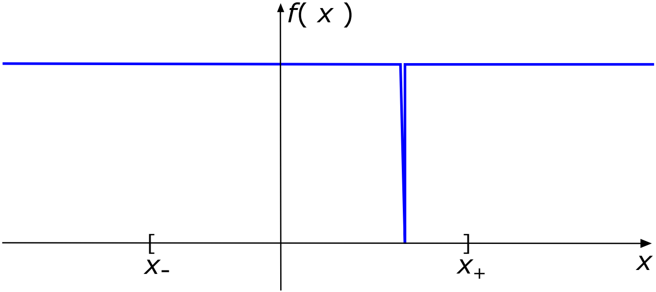
- Impossible because isolated minima can be anywhere [4, p. 408]
- Does it help restricting to $x \in X = [\, x_- \,,\, x_+ \,]$ $(-\infty < x_- < x_+ < +\infty)$?
- No: still uncountably many points to try
- Is it because f "jumps"?



- Impossible because isolated minima can be anywhere [4, p. 408]
- Does it help restricting to $x \in X = [\, x_- \, , \, x_+ \,]$ $(-\infty < x_- < x_+ < +\infty)$?
- No: still uncountably many points to try
- Is it because f "jumps"? No, f can have isolated $\downarrow$ spikes anywhere

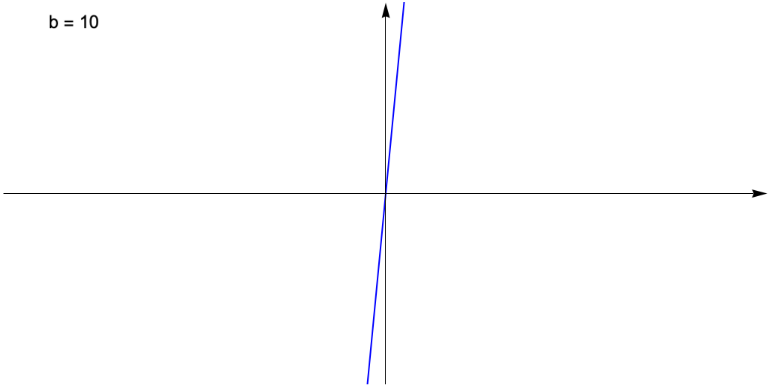


- Impossible because isolated minima can be anywhere [4, p. 408]
- Does it help restricting to $x \in X = [\, x_- \,,\, x_+ \,]$ $(-\infty < x_- < x_+ < +\infty)$?
- No: still uncountably many points to try
- Is it because f "jumps"? No, f can have isolated $\downarrow$ spikes anywhere
  ...even on $X = [\, x_- \,,\, x_+ \,]$ as spikes can be arbitrarily narrow

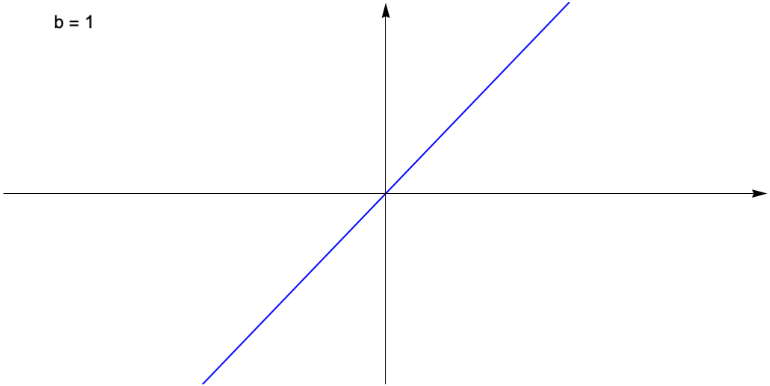


- Impossible because isolated minima can be anywhere [4, p. 408]
- Does it help restricting to $x \in X = [\, x_- \, , \, x_+ \,]$ $(-\infty < x_- < x_+ < +\infty)$?
- No: still uncountably many points to try
- Is it because f "jumps"? No, f can have isolated $\downarrow$ spikes anywhere
  ...even on $X = [\, x_- \, , \, x_+ \,]$ as spikes can be arbitrarily narrow
- To make (even approximate) optimization even possible, f must be "nice"
- Let's start with the nicest possible ones where optimization is $(\approx)$ trivial

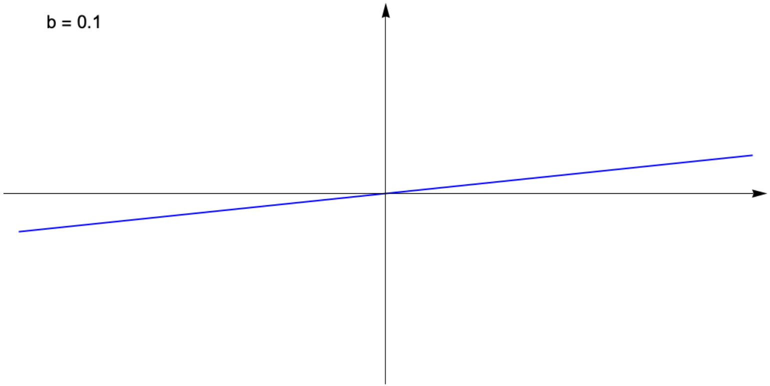
Outline



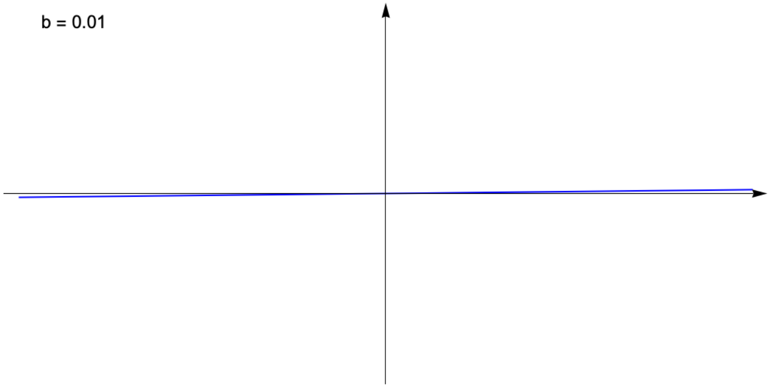
- The simplest possible function: $f(x) = bx$ (linear), fixed $b \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $b > 0 \equiv$ increasing: $x > z \Longrightarrow f(x) > f(z)$



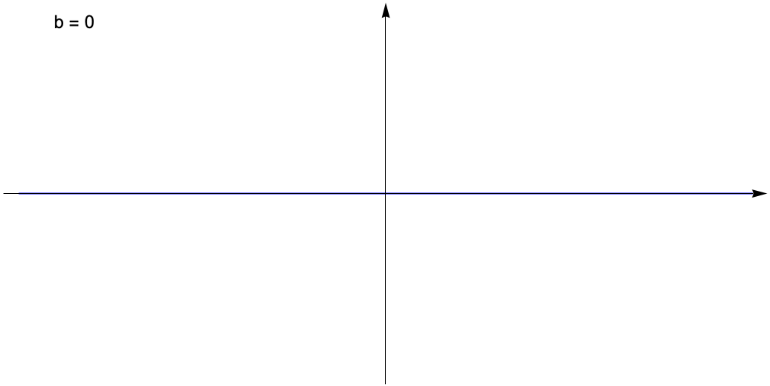
- The simplest possible function: $f(x) = bx$ (linear), fixed $b \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $b > 0 \equiv$ increasing: $x > z \Longrightarrow f(x) > f(z)$



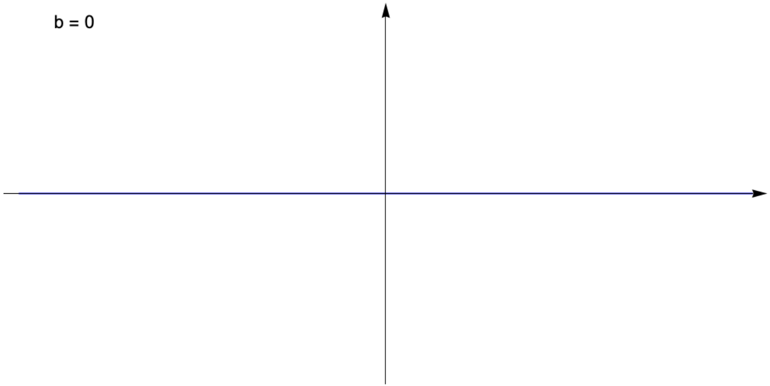
- The simplest possible function: $f(x) = bx$ (linear), fixed $b \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $b > 0 \equiv$ increasing: $x > z \Longrightarrow f(x) > f(z)$



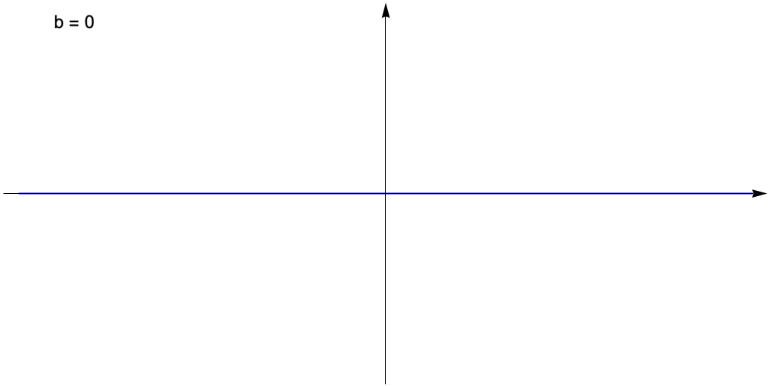
- The simplest possible function: $f(x) = bx$ (linear), fixed $b \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $b > 0 \equiv$ increasing: $x > z \Longrightarrow f(x) > f(z)$



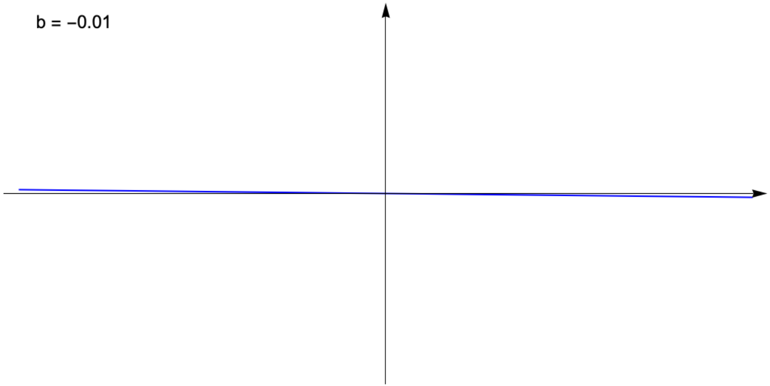
- The simplest possible function: $f(x) = bx$ (linear), fixed $b \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $b = 0 \equiv$ nondecreasing: $x > z \Longrightarrow f(x) \geq f(z)$ and



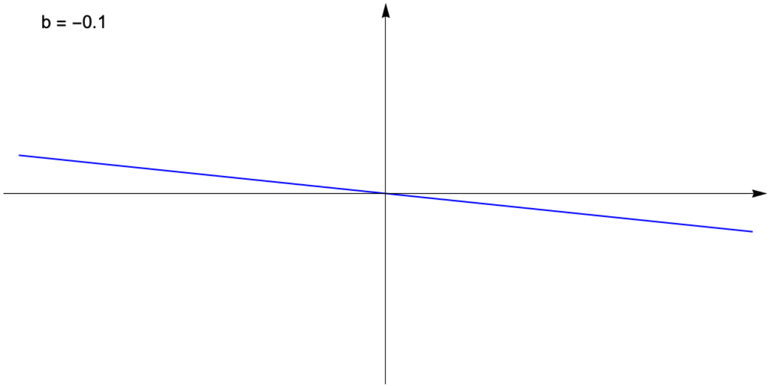
- The simplest possible function: $f(x) = bx$ (linear), fixed $b \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $b = 0 \equiv$ nonincreasing: $x > z \Longrightarrow f(x) \leq f(z)$



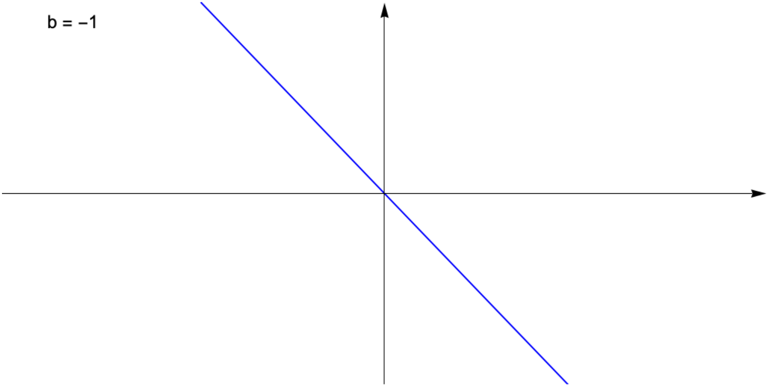
- The simplest possible function: $f(x) = bx$ (linear), fixed $b \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $b = 0 \equiv$ constant: $x > z \Longrightarrow f(x) = f(z)$



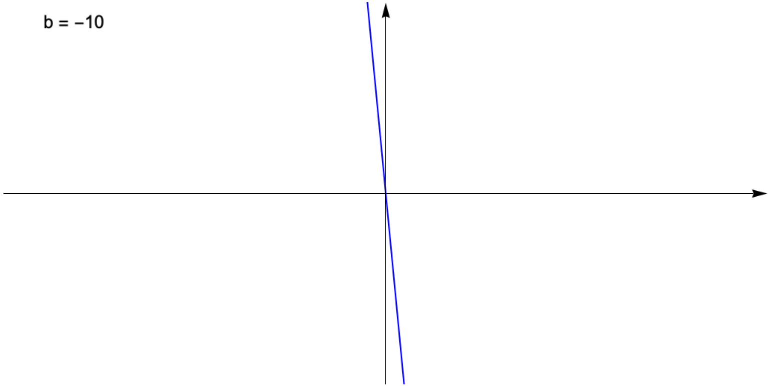
- The simplest possible function: $f(x) = bx$ (linear), fixed $b \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $b < 0 \equiv$ decreasing: $x > z \Longrightarrow f(x) < f(z)$



- The simplest possible function: $f(x) = bx$ (linear), fixed $b \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $b < 0 \equiv$ decreasing: $x > z \Longrightarrow f(x) < f(z)$

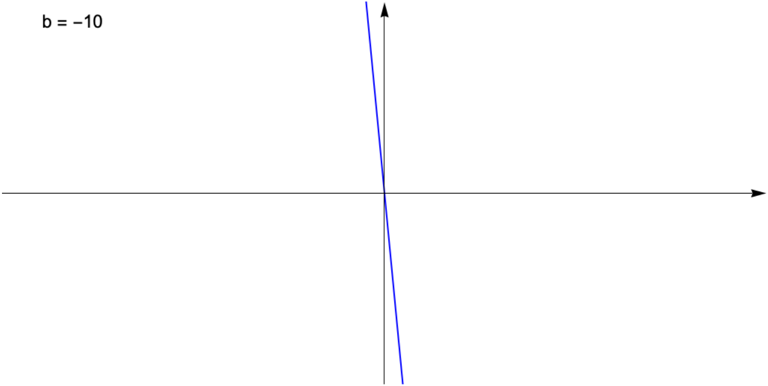


- The simplest possible function: $f(x) = bx$ (linear), fixed $b \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $b < 0 \equiv$ decreasing: $x > z \Longrightarrow f(x) < f(z)$



- The simplest possible function: $f(x) = bx$ (linear), fixed $b \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $b < 0 \equiv$ decreasing: $x > z \Longrightarrow f(x) < f(z)$

Exercise: Formally prove the stated properties



- The simplest possible function: $f(x) = bx$ (linear), fixed $b \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $b < 0 \equiv$ decreasing: $x > z \Longrightarrow f(x) < f(z)$

Exercise: Formally prove the stated properties

- $b =$ linear coefficient $=$ slope: the larger $|b|$, the steeper the line

- Too easy: $\min = -\infty$, $\max = +\infty$ unless $b = 0 \implies \min = \max = 0$
- More interesting: box-constrained optimization

  $$(P) \;\; \min\{ f(x) \, : \, x \in [x_- , x_+] \}$$

  with $-\infty \leq x_- \leq x_+ \leq +\infty \; \equiv \; X$ possibly (half-)infinite interval
- Constraints often useful, (finite) box constraints (very simple) especially so
- $b > 0 \implies$ argmin $= x_-$, $\min = f(x_-)$, argmax $= x_+$, $\max = f(x_+)$
- "Works" even if $x_- = -\infty$ and/or $x_+ = +\infty$, as $b \cdot (\pm\infty) = \pm\infty$

- Too easy: $\min = -\infty$, $\max = +\infty$ unless $b = 0 \implies \min = \max = 0$

- More interesting: box-constrained optimization

$$(P) \quad \min\{\, f(\,x\,) \,:\, x \in [\, x_- \,,\, x_+ \,] \,\}$$

  with $-\infty \leq x_- \leq x_+ \leq +\infty \;\equiv\; X$ possibly (half-)infinite interval

- Constraints often useful, (finite) box constraints (very simple) especially so

- $b > 0 \implies$ argmin $= x_-$, min $= f(\,x_-\,)$, argmax $= x_+$, max $= f(\,x_+\,)$

- "Works" even if $x_- = -\infty$ and/or $x_+ = +\infty$, as $b \cdot (\pm\infty) = \pm\infty$

Exercise: Formally prove the result, state & prove cases $b < 0$ and $b = 0$

- Too easy: $\min = -\infty$, $\max = +\infty$ unless $b = 0 \implies \min = \max = 0$
- More interesting: box-constrained optimization

  $$(P) \quad \min\{\, f(x) \,:\, x \in [\, x_- \,,\, x_+ \,] \,\}$$

  with $-\infty \leq x_- \leq x_+ \leq +\infty \;\equiv\; X$ possibly (half-)infinite interval
- Constraints often useful, (finite) box constraints (very simple) especially so
- $b > 0 \implies \operatorname{argmin} = x_-$, $\min = f(x_-)$, $\operatorname{argmax} = x_+$, $\max = f(x_+)$
- "Works" even if $x_- = -\infty$ and/or $x_+ = +\infty$, as $b \cdot (\pm\infty) = \pm\infty$

Exercise: Formally prove the result, state & prove cases $b < 0$ and $b = 0$

- Closed formula $O(1)$, don't get used to it
- Yet solving simple problems the basis of solving complicated ones

- Could have used $X = (\, x_- \,,\, x_+ \,) = \{\, x \in \mathbb{R} \; : \; x_- < x < x_+ \,\}$ (open interval)?

- Could have used $X = (\, x_- \,,\, x_+ \,) = \{\, x \in \mathbb{R} \,:\, x_- < x < x_+ \,\}$ (open interval)?
- Bad idea: again, inf $\exists$ but min $\nexists$, finite f_* but $x_* \nexists$

- Could have used $X = (\, x_- \,,\, x_+ \,) = \{\, x \in \mathbb{R} \;:\; x_- < x < x_+ \,\}$ (open interval)?

- Bad idea: again, inf $\exists$ but min $\nexists$, finite f_* but $x_* \nexists$

- Would it make sense for applications? Hardly. x a physical quantity
  $\Longrightarrow$ cannot be chosen/measured to ∞ precision
  (Plank scale, Heisenberg's Uncertainty Principle, ...)

- Could have used $X = (\, x_- \,,\, x_+ \,) = \{\, x \in \mathbb{R} \;:\; x_- < x < x_+ \,\}$ (open interval)?

- Bad idea: again, inf $\exists$ but min $\nexists$, finite f_* but $x_* \nexists$

- Would it make sense for applications? Hardly. x a physical quantity
  $\Longrightarrow$ cannot be chosen/measured to ∞ precision
  (Plank scale, Heisenberg's Uncertainty Principle, . . .)

- It is a problem for algorithms? In theory yes, in practice hardly:
  again, plenty of ε-optimal solutions however chosen $\varepsilon > 0$

- Could have used $X = (\, x_- \,,\, x_+ \,) = \{\, x \in \mathbb{R} \,:\, x_- < x < x_+ \,\}$ (open interval)?

- Bad idea: again, inf $\exists$ but min $\nexists$, finite f_* but $x_* \nexists$

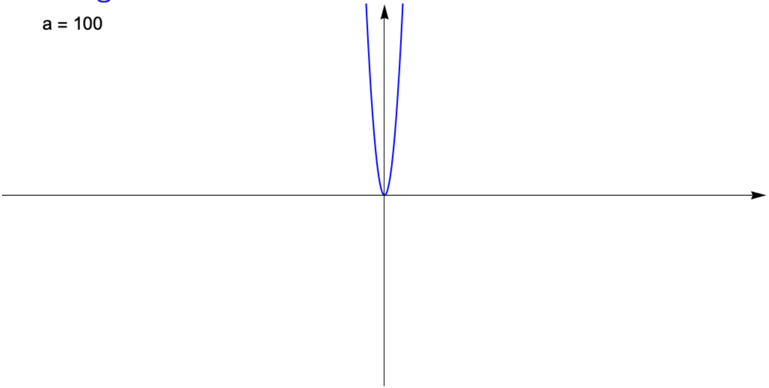
- Would it make sense for applications? Hardly. x a physical quantity
  $\Longrightarrow$ cannot be chosen/measured to ∞ precision
  (Plank scale, Heisenberg's Uncertainty Principle, ...)

- It is a problem for algorithms? In theory yes, in practice hardly:
  again, plenty of ε-optimal solutions however chosen $\varepsilon > 0$

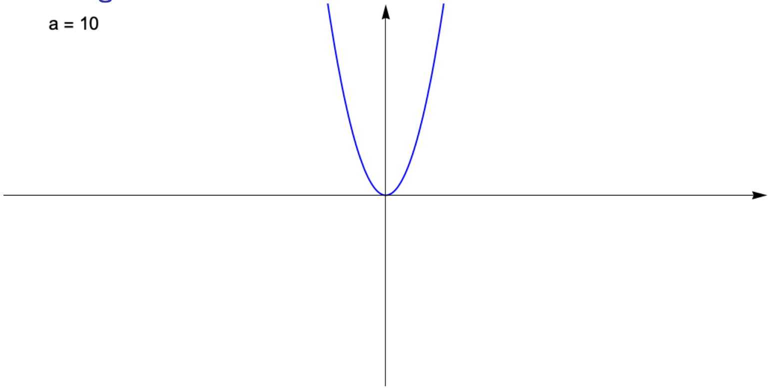
- Does it make any sense at all? Hardly: if x_-, x_+ "can't be touched", use
  $X = [\, x_- + \varepsilon_- \,,\, x_+ - \varepsilon_+]$ for appropriately chosen $\varepsilon_\pm$

- All in all? Just use closed intervals and be done with it

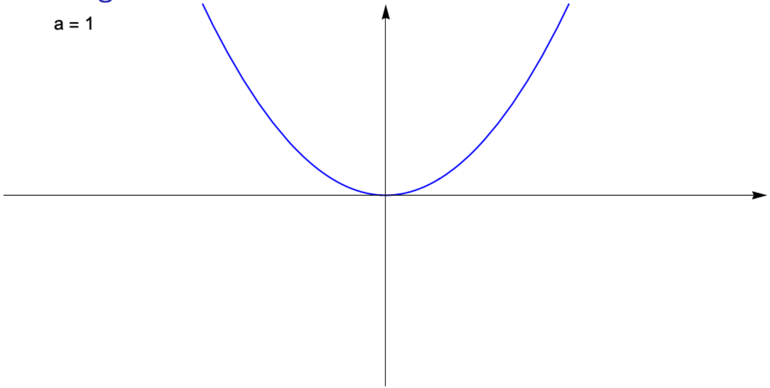
- Will generalise to "just use closed sets and be done with it"



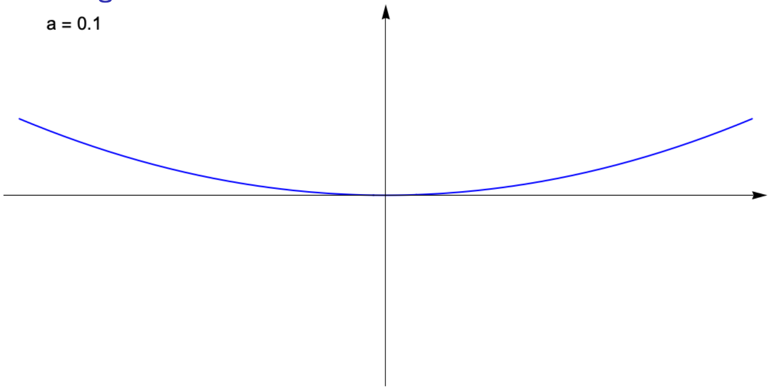
- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a > 0 \equiv$ decreasing for $x \leq 0$, increasing for $x \geq 0$



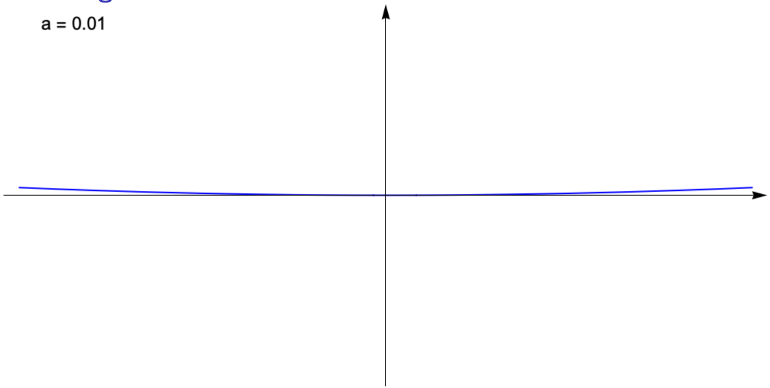
- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a > 0 \equiv$ decreasing for $x \leq 0$, increasing for $x \geq 0$



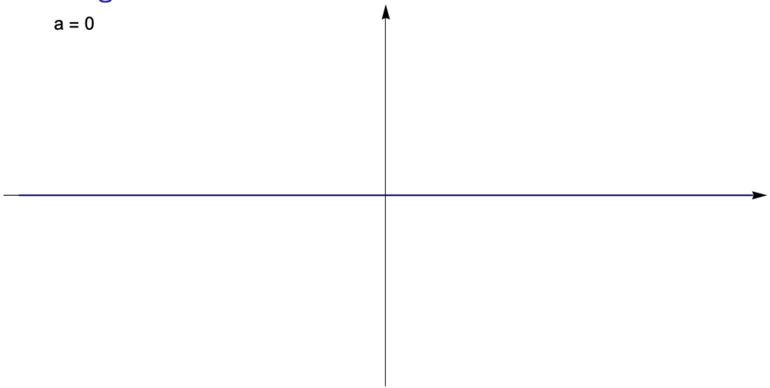
- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a > 0 \equiv$ decreasing for $x \leq 0$, increasing for $x \geq 0$



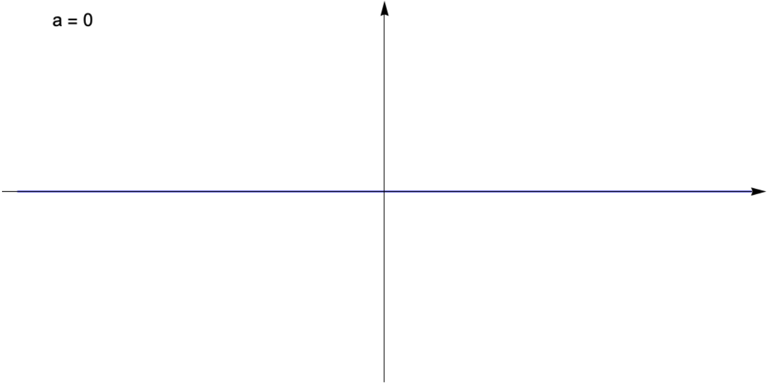
- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a > 0 \equiv$ decreasing for $x \leq 0$, increasing for $x \geq 0$



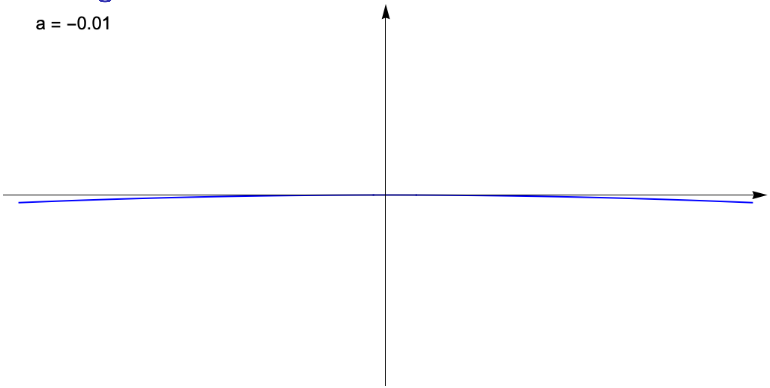
- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a > 0 \;\equiv\;$ decreasing for $x \leq 0$, increasing for $x \geq 0$



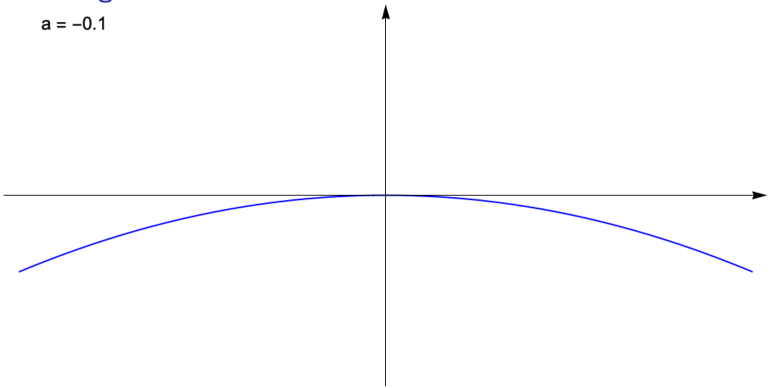
- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a = 0 \equiv$ nonincreasing for $x \leq 0$, nondecreasing for $x \geq 0$ and



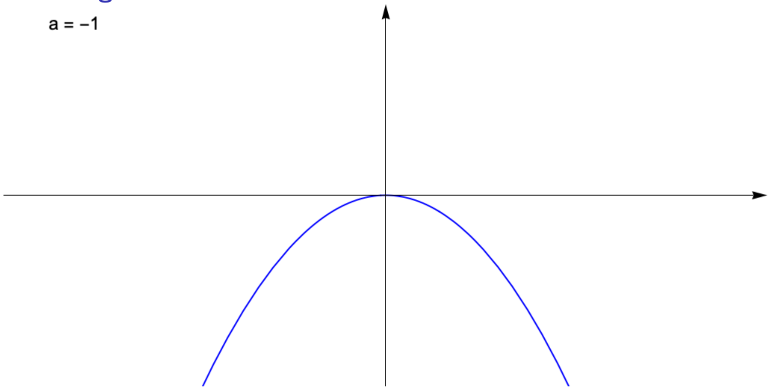
- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a = 0 \equiv$ nondecreasing for $x \leq 0$, nonincreasing for $x \geq 0$ (constant)



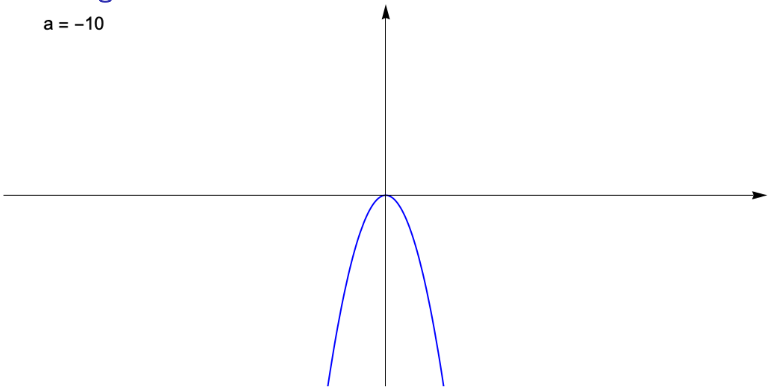
- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a < 0 \equiv$ increasing for $x \leq 0$, decreasing for $x \geq 0$



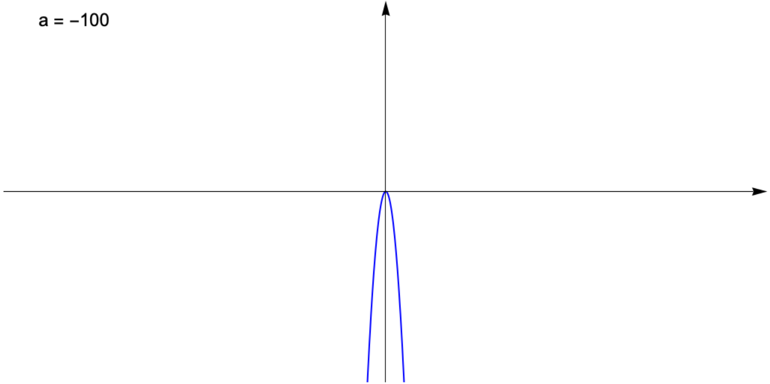
- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a < 0 \equiv$ increasing for $x \leq 0$, decreasing for $x \geq 0$



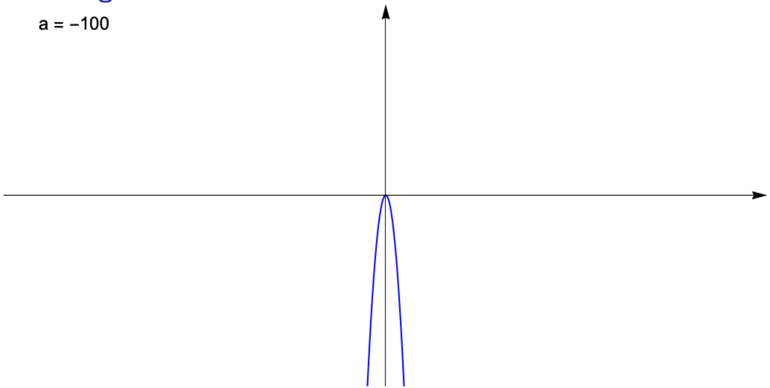
- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a < 0 \equiv$ increasing for $x \leq 0$, decreasing for $x \geq 0$



- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a < 0 \equiv$ increasing for $x \leq 0$, decreasing for $x \geq 0$

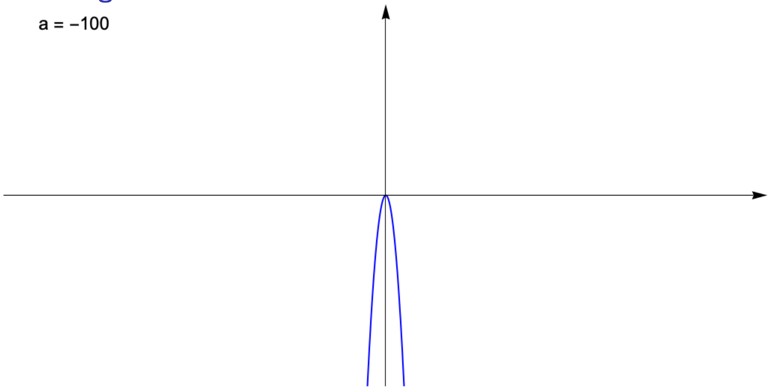


- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a < 0 \equiv$ increasing for $x \leq 0$, decreasing for $x \geq 0$



- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a < 0 \equiv$ increasing for $x \leq 0$, decreasing for $x \geq 0$

Exercise: Formally prove the stated properties



- Next simplest function: $f(x) = ax^2$ (homogeneous quadratic), fixed $a \in \mathbb{R}$
- As many different functions as real numbers (bijection)
- $a < 0 \equiv$ increasing for $x \leq 0$, decreasing for $x \geq 0$

Exercise: Formally prove the stated properties

- a = quadratic coefficient = curvature: the larger $|a|$, the steeper the parabola

- Clearly depends (and symmetric) on sign of a:
  - $a > 0 \implies$ min = argmin = 0, max = $+\infty$, argmax = $\pm\infty$
  - $a < 0 \implies$ max = argmax = 0, min = $-\infty$, argmin = $\pm\infty$
- Box-constrained optimization on (closed) $X = [\, x_- \,,\, x_+ \,]$ more interesting
- $a > 0 \implies$ three cases
  - $x_+ < 0 \implies$ argmin = x_+, argmax = x_-
  - $x_- > 0 \implies$ argmin = x_-, argmax = x_+
  - $x_- \leq 0 \leq x_+ \implies$ argmin = 0, argmax = $\text{argmax}\{\, f(\, x_- \,) \,,\, f(\, x_+ \,) \,\}$
- "Works" even if $x_- = -\infty$ and/or $x_+ = +\infty$, as $a \cdot (\pm\infty)^2 = +\infty$

- Clearly depends (and symmetric) on sign of a:
  - $a > 0 \implies$ min = argmin = 0, max = $+\infty$, argmax = $\pm\infty$
  - $a < 0 \implies$ max = argmax = 0, min = $-\infty$, argmin = $\pm\infty$
- Box-constrained optimization on (closed) $X = [\, x_- \,,\, x_+ \,]$ more interesting
- $a > 0 \implies$ three cases
  - $x_+ < 0 \implies$ argmin = x_+, argmax = x_-
  - $x_- > 0 \implies$ argmin = x_-, argmax = x_+
  - $x_- \leq 0 \leq x_+ \implies$ argmin = 0, argmax = argmax$\{\, f(\, x_- \,) \,,\, f(\, x_+ \,) \,\}$
- "Works" even if $x_- = -\infty$ and/or $x_+ = +\infty$, as $a \cdot (\pm\infty)^2 = +\infty$

Exercise: Formally prove the result, state & prove cases $a < 0$ and $a = 0$

- Clearly depends (and symmetric) on sign of a:
  - $a > 0 \implies$ min = argmin = 0, max = $+\infty$, argmax = $\pm\infty$
  - $a < 0 \implies$ max = argmax = 0, min = $-\infty$, argmin = $\pm\infty$
- Box-constrained optimization on (closed) $X = [\, x_- \,,\, x_+ \,]$ more interesting
- $a > 0 \implies$ three cases
  - $x_+ < 0 \implies$ argmin = x_+, argmax = x_-
  - $x_- > 0 \implies$ argmin = x_-, argmax = x_+
  - $x_- \leq 0 \leq x_+ \implies$ argmin = 0, argmax = $\text{argmax}\{ f(x_-) \,,\, f(x_+) \}$
- "Works" even if $x_- = -\infty$ and/or $x_+ = +\infty$, as $a \cdot (\pm\infty)^2 = +\infty$

Exercise: Formally prove the result, state & prove cases $a < 0$ and $a = 0$

- Again closed formula $O(1)$, don't get used to it
- $\max\{ f(x) \}$ and $\min\{ f(x) \}$ somewhat $\neq$ (cf. last case), will see much more

- Next step: $f(x) = ax^2 + bx$ (non-homogeneous quadratic), fixed $(a, b) \in \mathbb{R}^2$
- As many different functions as pairs of real numbers (bijection)
- Basically, a homogeneous quadratic $+$ a linear
- However, $\min\{ ax^2 + bx \} \neq \min\{ ax^2 \} + \min\{ bx \}$
- 0 clearly always a root, but in general not the argmin / argmax
- Powerful general concept: if $f(x)$ is "too complicated", make it "simpler"
- Sometimes this can be done by changing the space of variables (reformulation)
- In this case: change the input space so that it becomes homogeneous
- Clearly only needed if both $a \neq 0$ and $b \neq 0$

- Fundamental trick: $\bar{x} = -b/2a$ (because I say so), $z = x - \bar{x} \equiv x = z + \bar{x}$
- The z-space is the x-space where the origin is moved to $\bar{x}$
- Just algebra: $f(x) = a(z+\bar{x})^2 + b(z+\bar{x}) = az^2 + 2az\bar{x} + a\bar{x}^2 + bz + b\bar{x}$
  $= az^2 + (2a\bar{x} + b)z + [\, a\bar{x}^2 + b\bar{x} \,] = az^2 + f(\bar{x}) = g(z) \quad [\, 2a\bar{x} + b = 0 \,]$
- Translated by $\bar{x}$ horizontally (and by $f(\bar{x})$ vertically), $f(x)$ is homogeneous
- Its argmin / argmax (depending on sign of a) is $z = 0 \equiv x = \bar{x}$
- Then, just Optimizing a quadratic homogeneous function for $g(z)$
- Yet again, closed formula $O(1)$, don't get used to it

Exercise: Flesh out the details: describe all cases in terms of f and x

Exercise: Discuss the position of $\bar{x}$ and the roots of f depending on a, b

Outline

- Next crucial step: $f : \mathbb{R}^n \to \mathbb{R}$, i.e., $f(x_1, x_2, \ldots, x_n) = f(x)$ with $x = [x_i]_{i=1}^n = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^n$
- n can be smallish (2, 3, 100), largish (10^4, 10^5) or heinously large (10^9, 10^{11})
- All previous stuff (f_*, X_*, constraints, ...) immediately generalises

- Next crucial step: $f : \mathbb{R}^n \to \mathbb{R}$, i.e., $f(x_1, x_2, \dots, x_n) = f(x)$ with $x = [x_i]_{i=1}^n = [x_1, x_2, \dots, x_n] \in \mathbb{R}^n$
- n can be smallish (2, 3, 100), largish (10^4, 10^5) or heinously large (10^9, 10^{11})
- All previous stuff (f_*, X_*, constraints, ...) immediately generalises
- $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \dots \mathbb{R}$, Cartesian product of $\mathbb{R}$ n times $\Longrightarrow$ "exponentially larger than $\mathbb{R}$" $\equiv$ finding stuff a lot harder

- Next crucial step: $f : \mathbb{R}^n \rightarrow \mathbb{R}$, i.e., $f(\, x_1 \,,\, x_2 \,, \ldots ,\, x_n \,) = f(\, x \,)$ with $x = [\, x_i \,]_{i=1}^{n} = [\, x_1 \,,\, x_2 \,, \ldots ,\, x_n \,] \in \mathbb{R}^n$
- n can be smallish (2, 3, 100), largish (10^4, 10^5) or heinously large (10^9, 10^{11})
- All previous stuff (f_*, X_*, constraints, ...) immediately generalises
- $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \ldots \mathbb{R}$, Cartesian product of $\mathbb{R}$ n times $\Longrightarrow$ "exponentially larger than $\mathbb{R}$" $\equiv$ finding stuff a lot harder

> "Space is big. Really big. You just won't believe how vastly, hugely, mind-bogglingly big it is." [15]

- Next crucial step: $f : \mathbb{R}^n \to \mathbb{R}$, i.e., $f(x_1, x_2, \ldots, x_n) = f(x)$ with $x = [x_i]_{i=1}^n = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^n$
- n can be smallish (2, 3, 100), largish (10^4, 10^5) or heinously large (10^9, 10^{11})
- All previous stuff (f_*, X_*, constraints, ...) immediately generalises
- $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \ldots \mathbb{R}$, Cartesian product of $\mathbb{R}$ n times $\Longrightarrow$ "exponentially larger than $\mathbb{R}$" $\equiv$ finding stuff a lot harder

> "The vector space $\mathbb{R}^n$ is big. Really big. You just won't believe how vastly, hugely, mind-bogglingly big it is." [15]

- Next crucial step: $f : \mathbb{R}^n \to \mathbb{R}$, i.e., $f(x_1, x_2, \dots, x_n) = f(x)$ with $x = [x_i]_{i=1}^n = [x_1, x_2, \dots, x_n] \in \mathbb{R}^n$
- n can be smallish (2, 3, 100), largish (10^4, 10^5) or heinously large (10^9, 10^{11})
- All previous stuff (f_*, X_*, constraints, ...) immediately generalises
- $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \dots \mathbb{R}$, Cartesian product of $\mathbb{R}$ n times $\Longrightarrow$ "exponentially larger than $\mathbb{R}$" $\equiv$ finding stuff a lot harder

  > "The vector space $\mathbb{R}^n$ is big. Really big. You just won't believe how vastly, hugely, mind-bogglingly big it is." [15]

- Assume we can even luckily restrict to a "small" $x \in X \subset \mathbb{R}^n$: a "box" (hyperrectangle) $X = \{ x \in \mathbb{R}^n : x_- \leq x \leq x_+ \}$, $x_\pm \in \mathbb{R}^n$ (with $x_- \leq x_+$)
- Assume $x_- = 0$, $x_+ = u = [1, \dots, 1]$ and we can only look to integer values: still have 2^n points to look at (binary hypercube), grows too fast with n

- Next crucial step: $f : \mathbb{R}^n \to \mathbb{R}$, i.e., $f(x_1, x_2, \ldots, x_n) = f(x)$ with $x = [x_i]_{i=1}^n = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^n$
- n can be smallish (2, 3, 100), largish (10^4, 10^5) or heinously large (10^9, 10^{11})
- All previous stuff (f_*, X_*, constraints, …) immediately generalises
- $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \ldots \mathbb{R}$, Cartesian product of $\mathbb{R}$ n times $\Longrightarrow$ "exponentially larger than $\mathbb{R}$" $\equiv$ finding stuff a lot harder

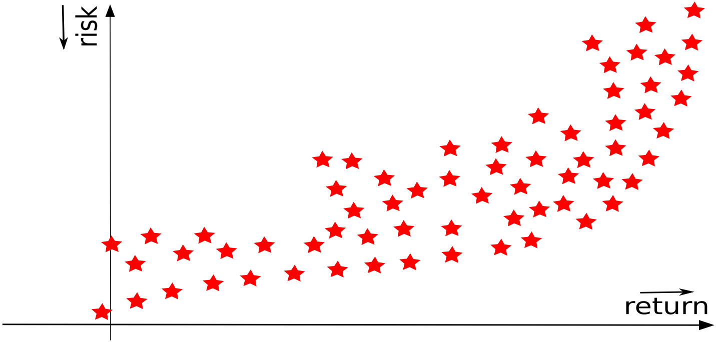
  > "The vector space $\mathbb{R}^n$ is big. Really big. You just won't believe how vastly, hugely, mind-bogglingly big it is." [15]

- Assume we can even luckily restrict to a "small" $x \in X \subset \mathbb{R}^n$: a "box" (hyperrectangle) $X = \{ x \in \mathbb{R}^n : x_- \leq x \leq x_+ \}$, $x_\pm \in \mathbb{R}^n$ (with $x_- \leq x_+$)
- Assume $x_- = 0$, $x_+ = u = [1, \ldots, 1]$ and we can only look to integer values: still have 2^n points to look at (binary hypercube), grows too fast with n
- Even picturing things is more complex and requires appropriate tools

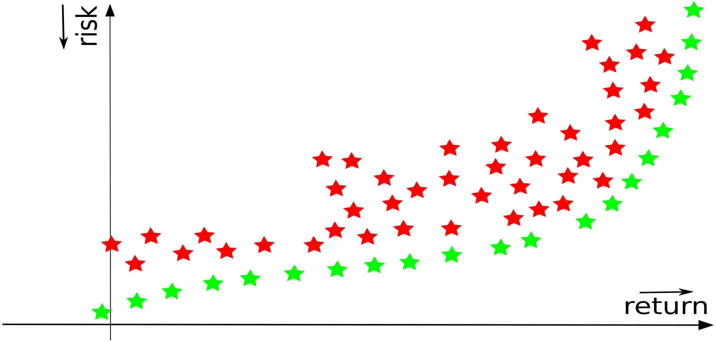
- Already "$f : X \to \mathbb{R}$" a rather strong assumption:
  can "express all the value of any $x \in X$ with a single number" $\Longrightarrow$
  given x' and x'' I can always tell which one I like best ($\mathbb{R}$ has total order)
- Often there would be more than one objective:

$$(P) \qquad \min \{ [f_1(x), f_2(x), \ldots] \ : \ x \in X \}$$

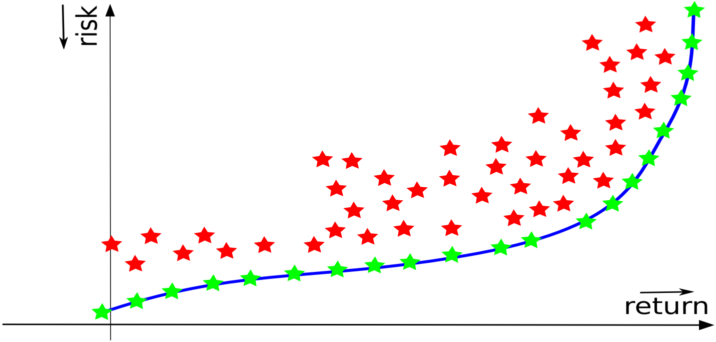
  with f_1, f_2, ... contrasting and/or with incomparable units (apples vs. oranges)
  - car cost vs. flashiness vs. km/l vs. # seats vs. trunk space ...
  - loss function $\mathcal{L}(w)$ vs. regularity $R(w)$ in ML
  - ...
- Vector-valued (a.k.a. multi-objective) optimization: $f : X \to \mathbb{R}^k$ with $k > 1$
- Textbook example: portfolio selection problem
  - X = set of financial instruments portfolios available to buy
  - $f_1(x)$ = expected return of portfolio x (€)
  - $f_2(x)$ = risk of portfolio x not achieving the expected return (%, CVAR, ...)



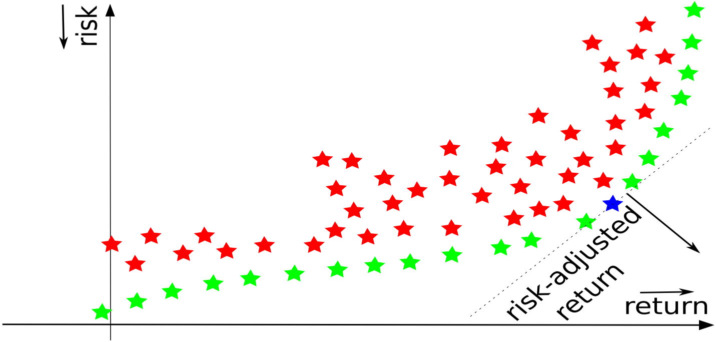
- $\mathbb{R}^k$ with $k > 1$ has no total order $\Longrightarrow$ no "best" solution, only



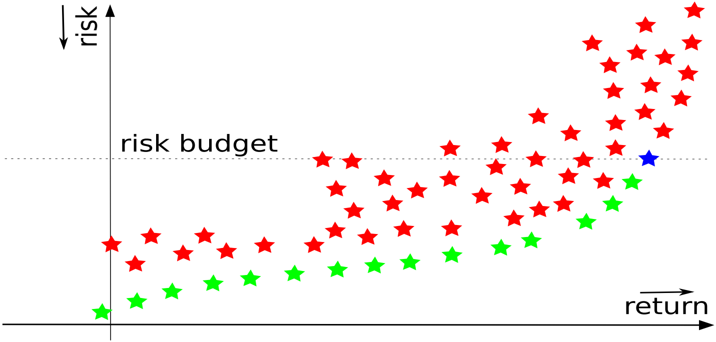
- $\mathbb{R}^k$ with $k > 1$ has no total order $\Longrightarrow$
  no "best" solution, only non-dominated ones on the



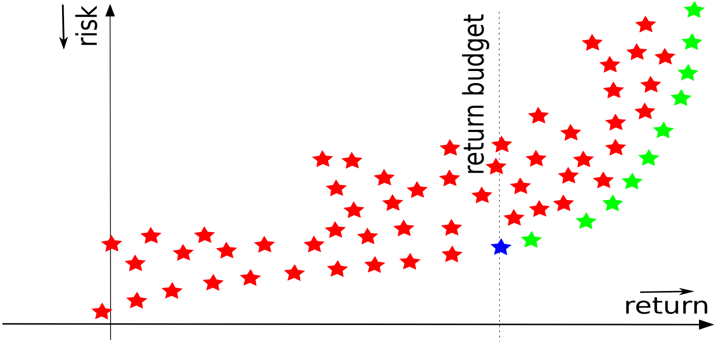
- $\mathbb{R}^k$ with $k > 1$ has no total order $\Longrightarrow$
  no "best" solution, only non-dominated ones on the Pareto frontier
- Two practical solutions:



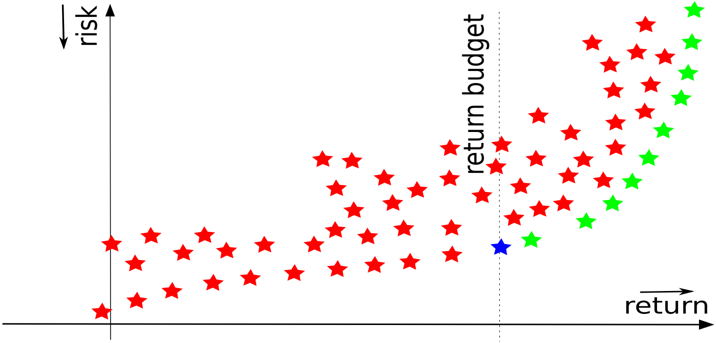
- $\mathbb{R}^k$ with $k > 1$ has no total order $\Longrightarrow$
  no "best" solution, only non-dominated ones on the Pareto frontier
- Two practical solutions: maximize risk-adjusted return,
  a.k.a. scalarization $\min \{ f_1(x) + \alpha f_2(x) \; : \; x \in X \}$ (which α??)



- $\mathbb{R}^k$ with $k > 1$ has no total order $\Longrightarrow$ no "best" solution, only non-dominated ones on the Pareto frontier
- Two practical solutions: maximize return with budget on maximum risk, a.k.a. budgeting $\min \{ f_1(x) : f_2(x) \leq \beta_2 , x \in X \}$ (which β_2??)



- $\mathbb{R}^k$ with $k > 1$ has no total order $\Longrightarrow$
  no "best" solution, only non-dominated ones on the Pareto frontier
- Two practical solutions: minimize risk with budget on minimum return,
  a.k.a. budgeting $\min \{ f_2(x) : f_1(x) \geq \beta_1 , x \in X \}$ (which β_1??)



- $\mathbb{R}^k$ with $k > 1$ has no total order $\Longrightarrow$
  no "best" solution, only non-dominated ones on the Pareto frontier
- Two practical solutions: minimize risk with budget on minimum return,
  a.k.a. budgeting $\min \{ f_2(x) \; : \; f_1(x) \geq \beta_1 \; , \; x \in X \}$ (which β_1??)
- All a bit fuzzy, but it's the nature of the beast
- We always assume this done if necessary at modelling stage
  (regularization, grid search used to divine hyperparameters α, β_1, β_2)

- (Euclidean) scalar product of $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$:

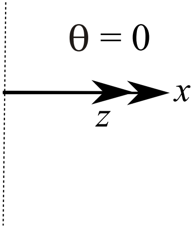
  $$\langle x , z \rangle = \sum_{i=1}^{n} x_i z_i = x_1 z_1 + \cdots + x_n z_n$$

- (Euclidean) norm: $\| x \| := \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\langle x , x \rangle}$ (induced by $\langle \cdot , \cdot \rangle$)

- (Euclidean) scalar product of $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$:

  $\langle x , z \rangle = \sum_{i=1}^{n} x_i z_i = x_1 z_1 + \cdots + x_n z_n$

- (Euclidean) norm: $\| x \| := \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\langle x , x \rangle}$ (induced by $\langle \cdot , \cdot \rangle$)



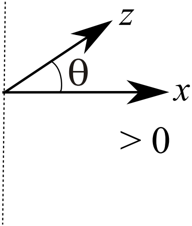
- Geometric interpretation: $\langle x , z \rangle = \| x \| \cdot \| z \| \cdot \cos(\theta)$

  $\langle x , z \rangle > 0 \;\equiv\;$ "x and z point in the same direction"

- (Euclidean) scalar product of $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$:

  $\langle x , z \rangle = \sum_{i=1}^{n} x_i z_i = x_1 z_1 + \cdots + x_n z_n$

- (Euclidean) norm: $\| x \| := \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\langle x , x \rangle}$ (induced by $\langle \cdot , \cdot \rangle$)



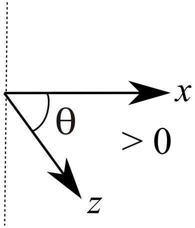
- Geometric interpretation: $\langle x , z \rangle = \| x \| \cdot \| z \| \cdot \cos(\theta)$

  $\langle x , z \rangle > 0 \; \equiv$ "x and z point in the same direction"

- (Euclidean) scalar product of $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$:

  $\langle x , z \rangle = \sum_{i=1}^{n} x_i z_i = x_1 z_1 + \cdots + x_n z_n$

- (Euclidean) norm: $\| x \| := \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\langle x , x \rangle}$ (induced by $\langle \cdot , \cdot \rangle$)



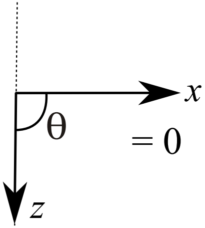
- Geometric interpretation: $\langle x , z \rangle = \| x \| \cdot \| z \| \cdot \cos(\theta)$

  $\langle x , z \rangle > 0 \equiv$ "x and z point in the same direction"

- (Euclidean) scalar product of $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$:

  $\langle x , z \rangle = \sum_{i=1}^{n} x_i z_i = x_1 z_1 + \cdots + x_n z_n$

- (Euclidean) norm: $\| x \| := \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\langle x , x \rangle}$ (induced by $\langle \cdot , \cdot \rangle$)



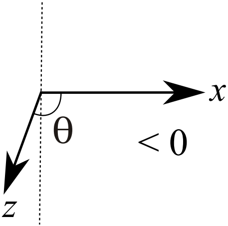
- Geometric interpretation: $\langle x , z \rangle = \| x \| \cdot \| z \| \cdot \cos(\theta)$

  $\langle x , z \rangle = 0 \equiv x \perp z$ (orthogonal)

- (Euclidean) scalar product of $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$:

  $\langle x , z \rangle = \sum_{i=1}^n x_i z_i = x_1 z_1 + \cdots + x_n z_n$

- (Euclidean) norm: $\| x \| := \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\langle x , x \rangle}$ (induced by $\langle \cdot , \cdot \rangle$)



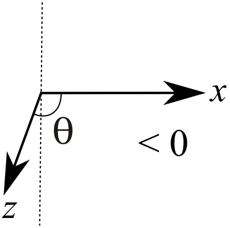
- Geometric interpretation: $\langle x , z \rangle = \| x \| \cdot \| z \| \cdot \cos(\theta)$

  $\langle x , z \rangle < 0 \; \equiv$ "x and z point in the opposite direction"

- (Euclidean) scalar product of $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$:

  $\langle\, x\,,\, z\,\rangle = \sum_{i=1}^n x_i z_i = x_1 z_1 + \cdots + x_n z_n$

- (Euclidean) norm: $\|\, x\,\| := \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\langle\, x\,,\, x\,\rangle}$ (induced by $\langle\, \cdot\,,\, \cdot\,\rangle$)



- Geometric interpretation: $\langle\, x\,,\, z\,\rangle = \|\, x\,\| \cdot \|\, z\,\| \cdot \cos(\,\theta\,)$

  $\langle\, x\,,\, z\,\rangle < 0 \;\equiv\;$ "x and z point in the opposite direction"

- Cauchy-Schwarz inequality: $|\,\langle\, x\,,\, z\,\rangle\,| \leq \|\, x\,\|\|\, z\,\|\;\; \forall x\,,\, z$

- (Euclidean) distance between x and z = norm of x when z is the origin:

  $d(\,x\,,\, z\,) := \|\, x - z\,\| = \sqrt{(\,x_1 - z_1\,)^2 + \cdots + (\,x_n - z_n\,)^2}$

- Ball, center $x \in \mathbb{R}^n$, radius $r > 0$: $\mathcal{B}(\,x\,,\, r\,) = \{\, z \in \mathbb{R}^n \,:\, \|\, z - x\,\| \leq r\,\}$

- $\mathbb{R}^n \in$ vector space $\equiv$ closed under sum and scalar multiplication

$$x + z = [\, x_1 + z_1 \,, \,\dots\,, \, x_n + z_n \,] \;\;,\;\; \alpha x = [\, \alpha x_1 \,, \,\dots\,, \, \alpha x_n \,]$$

- Finite-dimensional vector space: $\{\, u^i \,\}_{i=1}^n$ finite base s.t. $\forall\, x \in \mathbb{R}^n \;\exists\, \alpha_1, \dots, \alpha_n$ s.t. $x = \alpha_1 u^1 + \ldots + \alpha_n u^n$ (canonical base: $u^i_i = 1$, $u^i_h = 0$ for $h \neq i$, $\alpha_i = x_i$)

- Not all vector spaces are finite-dimensional (function spaces, ...)

- Properties $\equiv$ definition of scalar product:
  1. $\langle\, x \,,\, z \,\rangle = \langle\, z \,,\, x \,\rangle \quad \forall x \,,\, z \in \mathbb{R}^n$ (symmetry)
  2. $\langle\, x \,,\, x \,\rangle \geq 0 \quad \forall x \in \mathbb{R}^n \;\;,\;\; \langle\, x \,,\, x \,\rangle = 0 \iff x = 0$
  3. $\langle\, \alpha x \,,\, z \,\rangle = \alpha \langle\, x \,,\, z \,\rangle \quad \forall x \in \mathbb{R}^n \,,\, \alpha \in \mathbb{R}$
  4. $\langle\, x + w \,,\, z \,\rangle = \langle\, x \,,\, z \,\rangle + \langle\, w \,,\, z \,\rangle \quad \forall x \,,\, w \,,\, z \in \mathbb{R}^n$

- $\exists$ other scalar products that make sense in other spaces (matrices, integrable functions, random variables, ...)

- Not just theoretical stuff (cf. kernel in SVM)

Mathematically speaking: Norm, distance [14][1, A.1.2][6, p. 600]

- Properties $\equiv$ definition of norm:
  1. $\| x \| \geq 0 \quad \forall x \in \mathbb{R}^n \quad , \quad \| x \| = 0 \iff x = 0$
  2. $\| \alpha x \| = | \alpha | \| x \| \quad \forall x \in \mathbb{R}^n , \; \alpha \in \mathbb{R}$
  3. $\| x + z \| \leq \| x \| + \| z \| \quad \forall x , z \in \mathbb{R}^n$ (triangle inequality)

- $\| x + z \|^2 = \| x \|^2 + \| z \|^2 + 2\langle x , z \rangle$ (only Euclidean norm)

- $2\| x \|^2 + 2\| z \|^2 = \| x + z \|^2 + \| x - z \|^2$ (Parallelogram Law)

- Properties $\equiv$ definition of distance:
  1. $d(x , z) \geq 0 \quad \forall x , z \in \mathbb{R}^n \quad , \quad d(x , z) = 0 \iff x = z$
  2. $d(\alpha x , 0) = |\alpha| d(x , 0) \quad \forall x \in \mathbb{R}^n , \; \alpha \in \mathbb{R}$
  3. $d(x , w) \leq d(x , z) + d(z , w) \quad \forall x , w , z \in \mathbb{R}^n$ (triangle inequality)

- $\| \cdot \|$ defines $\mathcal{B}(\cdot , \cdot) \equiv$ the topology of the vector space: what is next to what (will be useful later on)

- $\mathrm{gr}(f) \in \mathbb{R}^{n+1}$, impossible if $n > 3$ ($n = 3$ hard already)
- $L(f, \cdot) \in \mathbb{R}^{n}$, impossible if $n > 4$ ($n = 4$ hard already)

- $\mathrm{gr}(\,f\,) \in \mathbb{R}^{n+1}$, impossible if $n > 3$ ($n = 3$ hard already)
- $L(\,f\,,\,\cdot\,) \in \mathbb{R}^{n}$, impossible if $n > 4$ ($n = 4$ hard already)
- General n, $f : \mathbb{R}^n \to \mathbb{R}$, $x \in \mathbb{R}^n$ (origin), $d \in \mathbb{R}^n$ (direction):
  $\varphi_{x,d}(\,\alpha\,) = f(\,x + \alpha d\,) : \mathbb{R} \to \mathbb{R}$ tomography of f from x along d
- $\mathrm{gr}(\,\varphi_{x,d}\,)$ can always be pictured, but too many of them: which x, d?
- $\|\,d\,\|$ only changes the scale: $\varphi_{x,\beta d}(\,\alpha\,) = \varphi_{x,d}(\,\beta\alpha\,)$ (**check**) $\Longrightarrow$ often (but not always) convenient to use normalised direction ($\|\,d\,\| = 1$)
- Simplest case: restriction along i-th coordinate ($\|\,u^i\,\| = 1$)
  $f^i_x(\,\alpha\,) = f(\,x_1\,,\,\ldots\,,\,x_{i-1}\,,\,\alpha\,,\,x_{i+1}\,,\,\ldots\,,\,x_n\,) \;\equiv\; \varphi_{[\,x_1\,,\ldots,\,x_{i-1}\,,\,0\,,\,x_{i+1}\,,\ldots,\,x_n\,],u^i}(\,\alpha\,)$
- For small n can "look at all d"
- Otherwise, find the specific d that "shows what you want to see"
- When x and d clear from context (will happen a lot), just $\varphi(\,\alpha\,)$

- Linear function: $f(x) = \langle b, x \rangle = \sum_{i=1}^{n} b_i x_i$, fixed $b \in \mathbb{R}^n$
- Linear $\equiv$ i. $f(\gamma x) = \gamma f(x)$, ii. $f(x + z) = f(x) + f(z) \quad \forall x, \gamma, z$

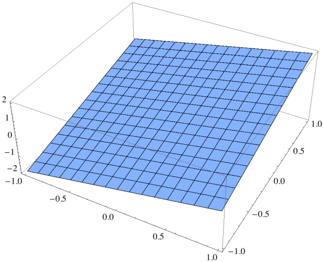
Exercise: Linear $\implies$ i) + ii) trivial, prove $\impliedby$; extends to affine $(\ldots + c)$?

- $\langle b, x \rangle = \sum_{i=1}^{n} [f_i(x_i) = b_i x_i]$, sum of n univariate linear functions

- Linear function: $f(x) = \langle b, x \rangle = \sum_{i=1}^{n} b_i x_i$, fixed $b \in \mathbb{R}^n$
- Linear $\equiv$ i. $f(\gamma x) = \gamma f(x)$, ii. $f(x + z) = f(x) + f(z) \quad \forall x, \gamma, z$

Exercise: Linear $\implies$ i) + ii) trivial, prove $\impliedby$; extends to affine $(\ldots + c)$?

- $\langle b, x \rangle = \sum_{i=1}^{n} [f_i(x_i) = b_i x_i]$, sum of n univariate linear functions

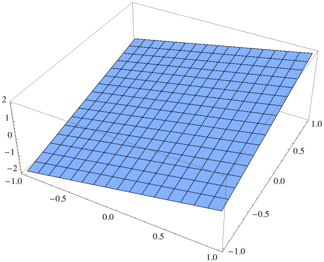


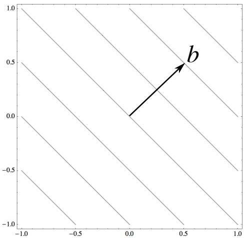
- $\mathrm{gr}(f) =$ hyperplane in $\mathbb{R}^{n+1}$ (plane in $\mathbb{R}^3$)

- Linear function: $f(x) = \langle b, x \rangle = \sum_{i=1}^{n} b_i x_i$, fixed $b \in \mathbb{R}^n$
- Linear $\equiv$ i. $f(\gamma x) = \gamma f(x)$, ii. $f(x+z) = f(x) + f(z) \quad \forall x, \gamma, z$

Exercise: Linear $\implies$ i) + ii) trivial, prove $\impliedby$; extends to affine $(\ldots + c)$?

- $\langle b, x \rangle = \sum_{i=1}^{n} [\, f_i(x_i) = b_i x_i \,]$, sum of n univariate linear functions



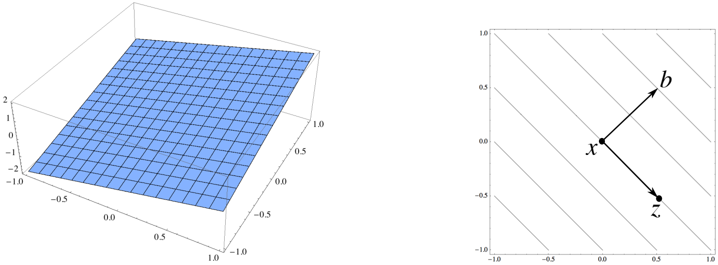


- $\mathrm{gr}(f)$ = hyperplane in $\mathbb{R}^{n+1}$ (plane in $\mathbb{R}^3$)
- Level sets are parallel hyperplanes in $\mathbb{R}^n$ (lines in $\mathbb{R}^2$) $\perp b$:

- Linear function: $f(x) = \langle b, x \rangle = \sum_{i=1}^{n} b_i x_i$, fixed $b \in \mathbb{R}^n$
- Linear $\equiv$ i. $f(\gamma x) = \gamma f(x)$, ii. $f(x+z) = f(x) + f(z) \quad \forall x, \gamma, z$

Exercise: Linear $\Longrightarrow$ i) + ii) trivial, prove $\Longleftarrow$; extends to affine $(\ldots + c)$?

- $\langle b, x \rangle = \sum_{i=1}^{n} [\, f_i(x_i) = b_i x_i \,]$, sum of n univariate linear functions

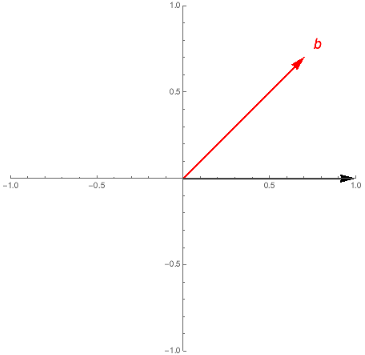


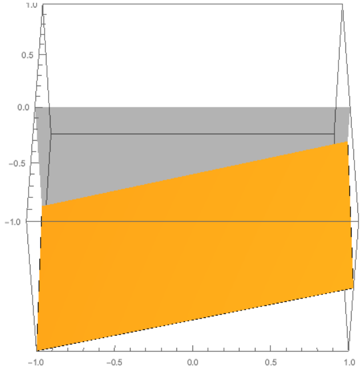
- $\mathrm{gr}(f)$ = hyperplane in $\mathbb{R}^{n+1}$ (plane in $\mathbb{R}^3$)
- Level sets are parallel hyperplanes in $\mathbb{R}^n$ (lines in $\mathbb{R}^2$) $\perp b$:

  $f(x) = f(z) \equiv \langle b, x \rangle = \langle b, z \rangle \equiv \langle b, z - x \rangle = 0 \equiv b \perp z - x$

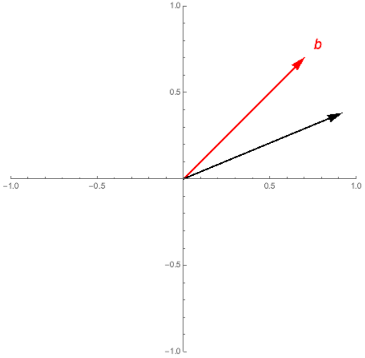
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha\langle b, d \rangle = \alpha\| b \| \cos(\theta)$

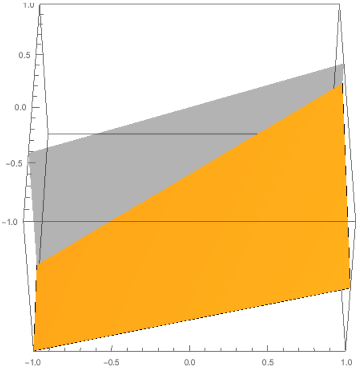
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$



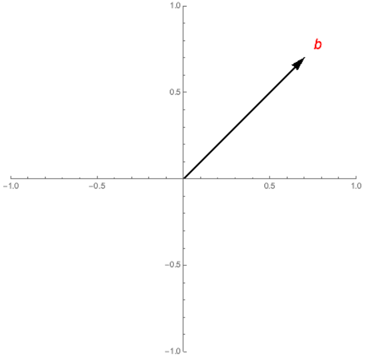


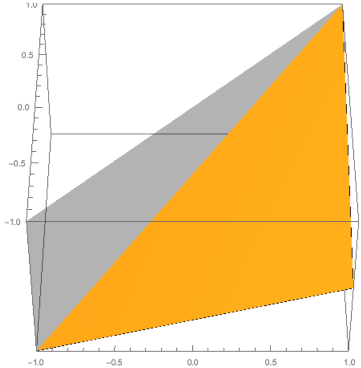
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$



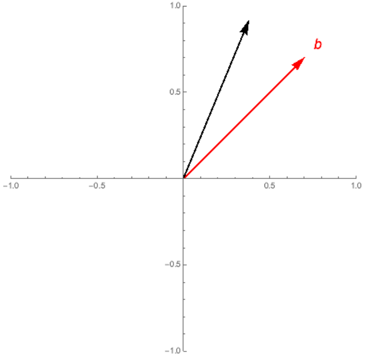


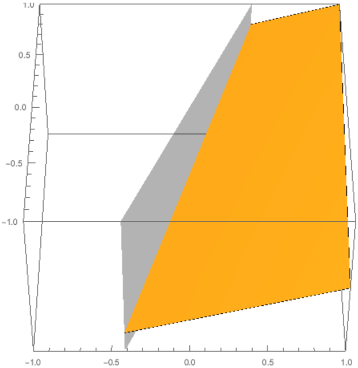
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$



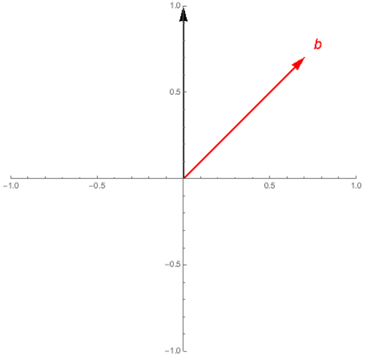


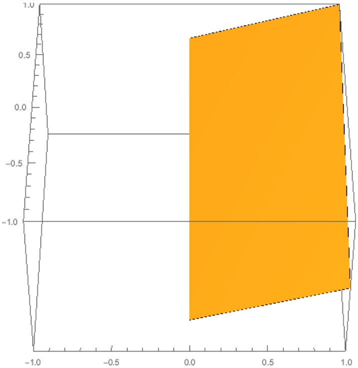
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$



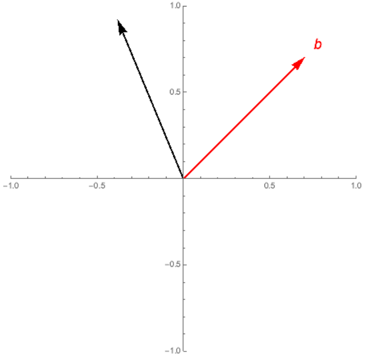


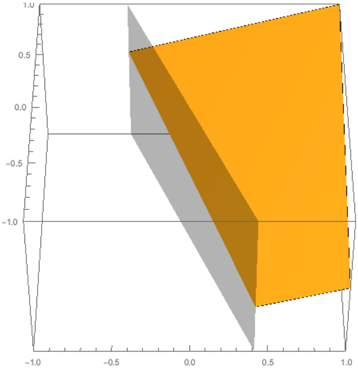
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$



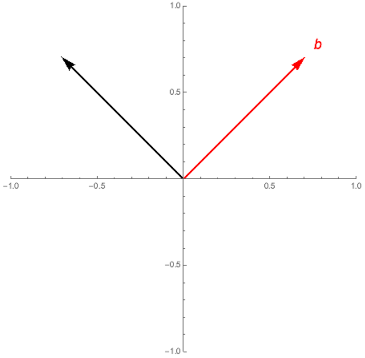


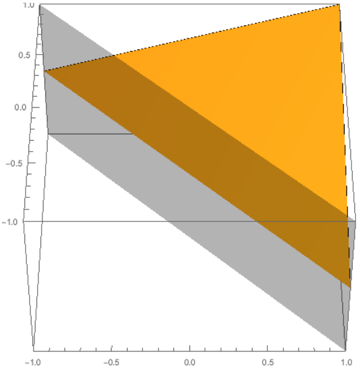
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$



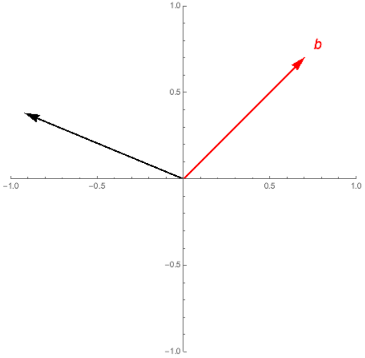


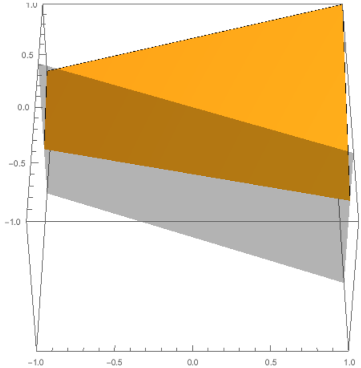
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$



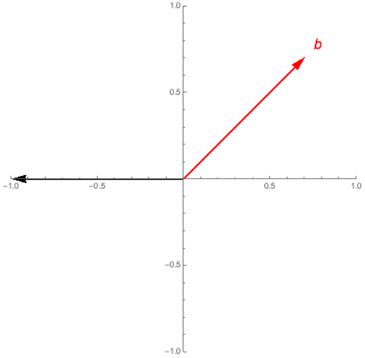


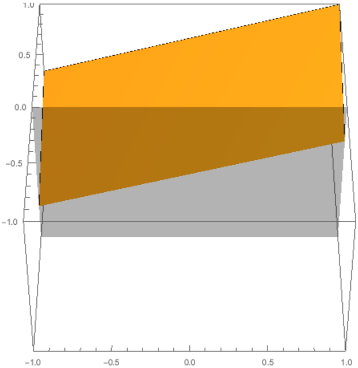
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$



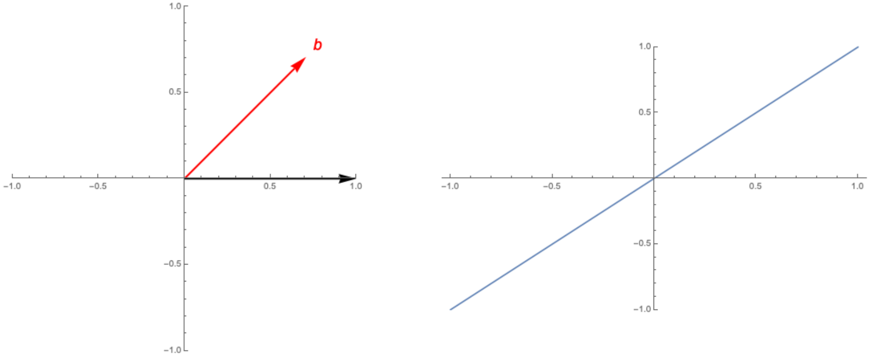


- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$



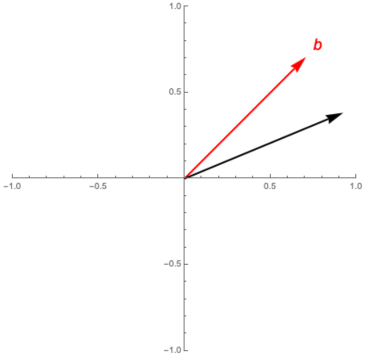


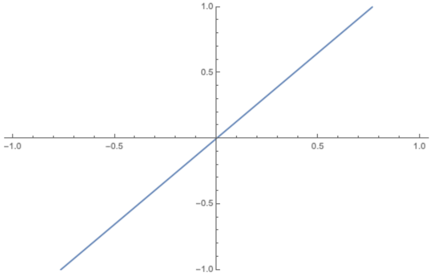
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$



- Increasing if "b same direction as d",

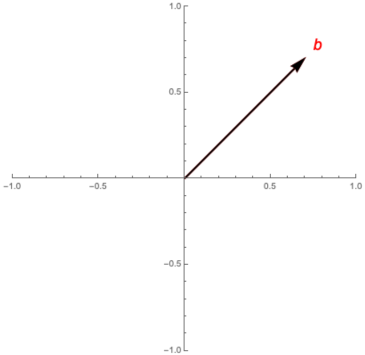
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$

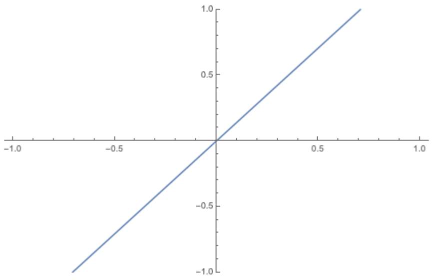




- Increasing if "*b* same direction as *d*", "more collinear" $\implies$ steeper

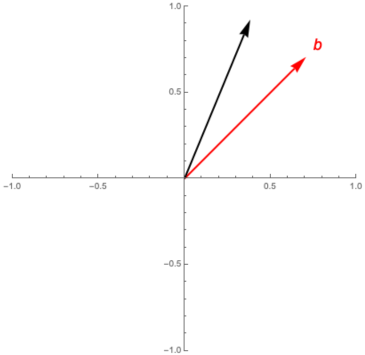
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha\langle b, d \rangle = \alpha\| b \| \cos(\theta)$

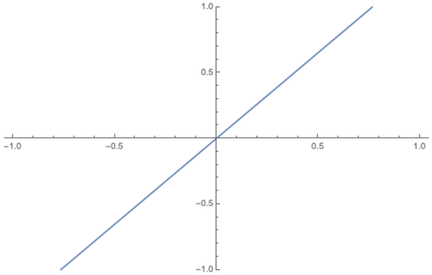




- Increasing if "b same direction as d", collinear $\Longrightarrow$ steepest

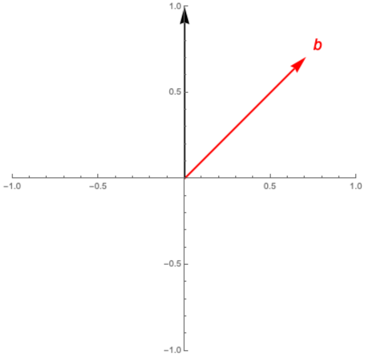
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$

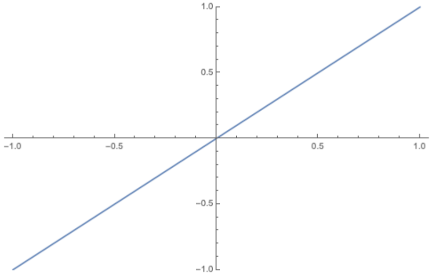




- Increasing if "b same direction as d", "less collinear" $\implies$ less steep

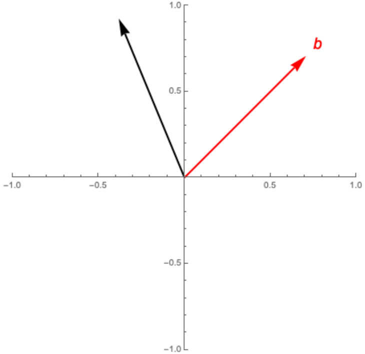
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha\langle b, d \rangle = \alpha\| b \| \cos(\theta)$

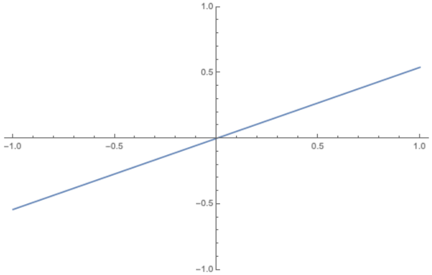




- Increasing if "*b* same direction as *d*", "less collinear" $\implies$ less steep

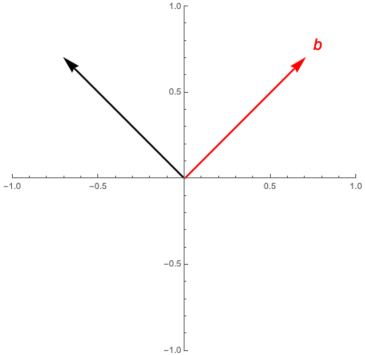
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$

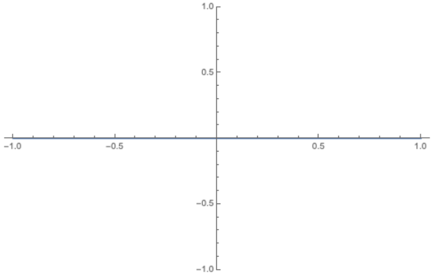




- Increasing if "b same direction as d", "less collinear" $\Longrightarrow$ less steep

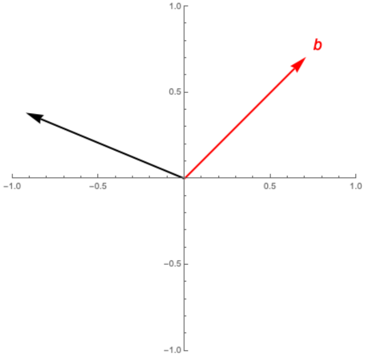
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$

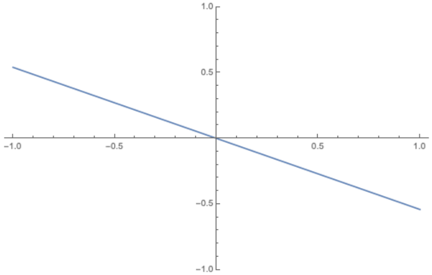




- "Flat" if $d \perp b$

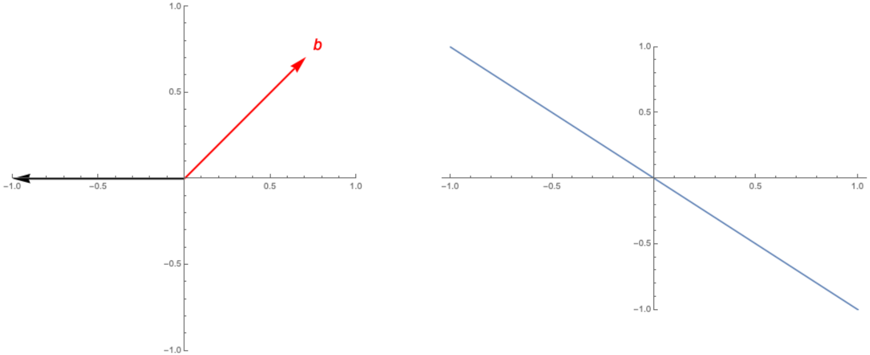
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$





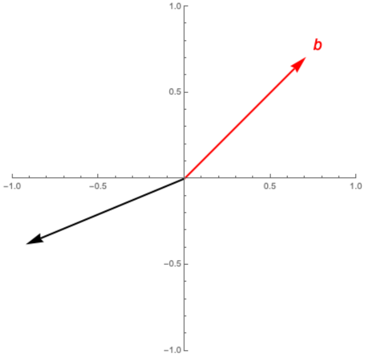
- Decreasing if "*b* opposite direction as *d*",

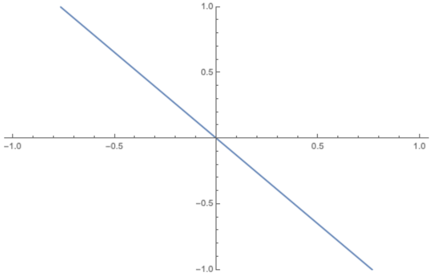
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$



- Decreasing if "b opposite direction as d", "more collinear" $\implies$ steeper

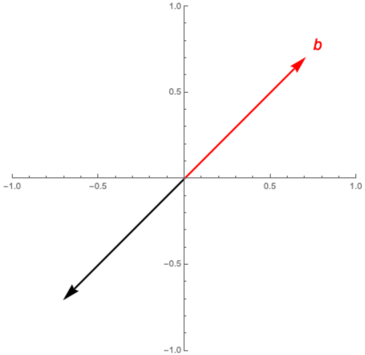
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha\langle b, d \rangle = \alpha\| b \| \cos(\theta)$

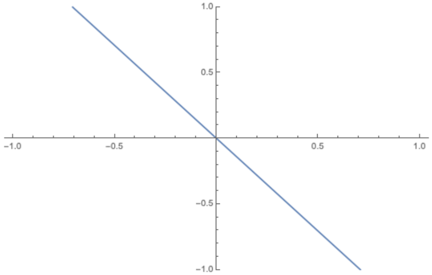




- Decreasing if "b opposite direction as d", "more collinear" $\Longrightarrow$ steeper

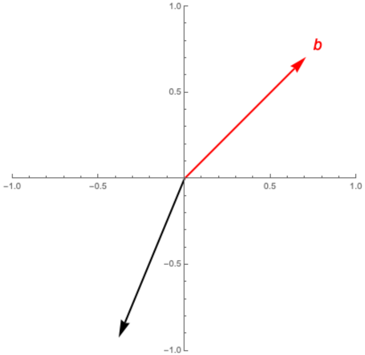
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$

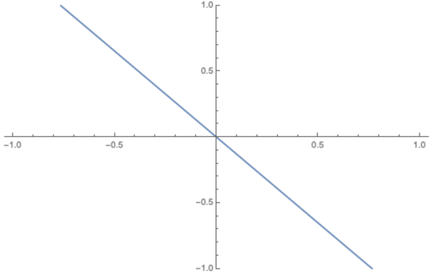




- Decreasing if "b opposite direction as d", collinear $\Longrightarrow$ steepest (negative)

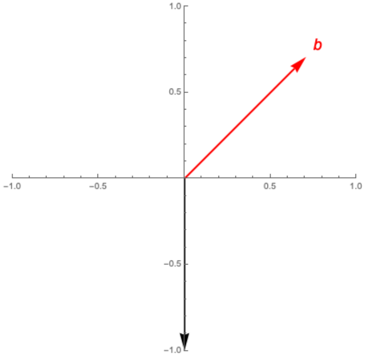
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$

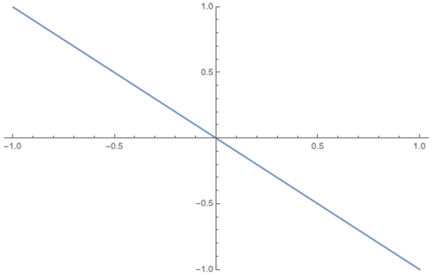




- Decreasing if "*b* opposite direction as *d*",

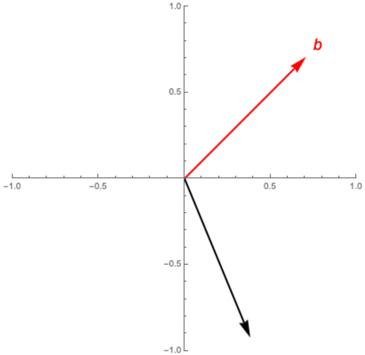
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$

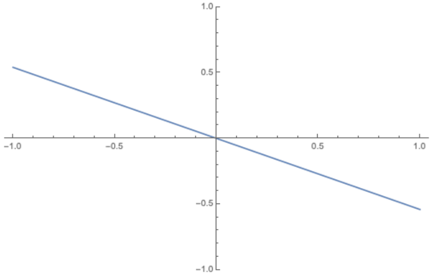




- Decreasing if "*b* opposite direction as *d*", "less collinear" $\Longrightarrow$ less steep

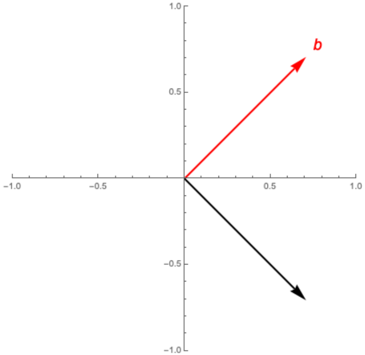
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$

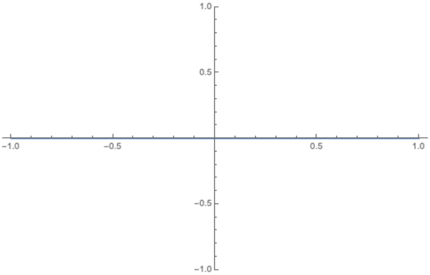




- Decreasing if "b opposite direction as d", "less collinear" $\Longrightarrow$ less steep

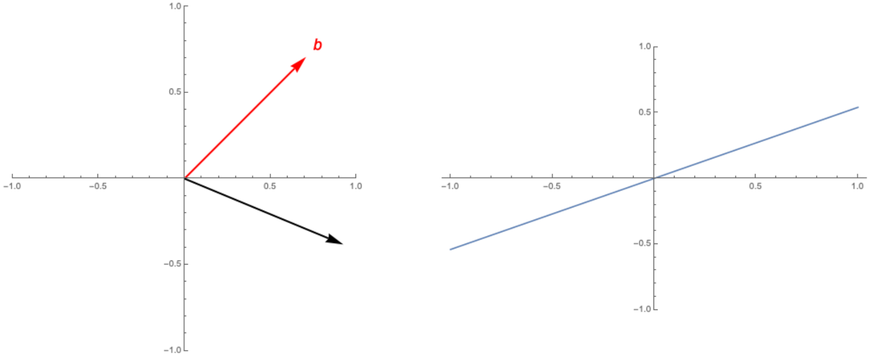
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$





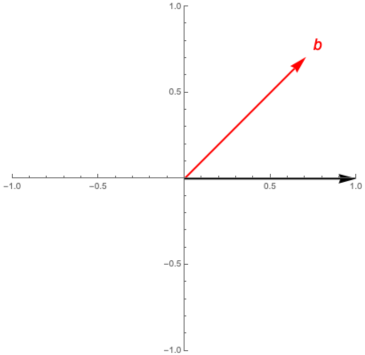
- "Flat" if $d \perp b$

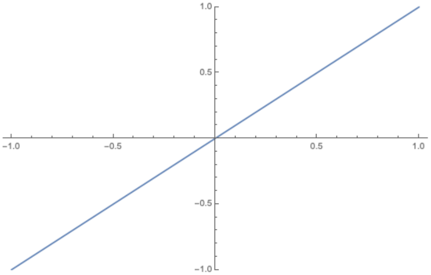
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha \langle b, d \rangle = \alpha \| b \| \cos(\theta)$



- Increasing if "b in the same direction as d"

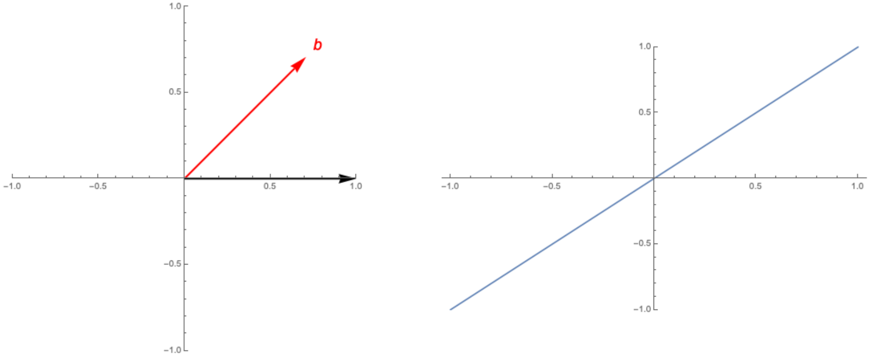
- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha\langle b, d \rangle = \alpha\| b \| \cos(\theta)$





- Increasing if "*b* in the same direction as *d*"

- $f(x) = \langle b, x \rangle$, $x = 0$, $\| d \| = 1$: $\varphi(\alpha) = \alpha\langle b, d \rangle = \alpha\| b \| \cos(\theta)$



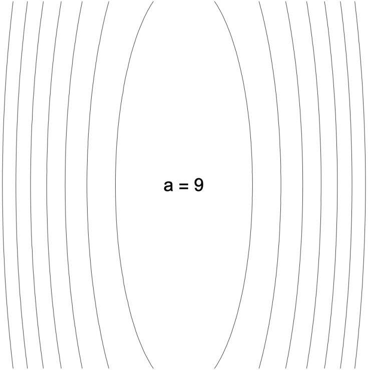
- Increasing if "b in the same direction as d"
- $f_* = \min\{ f(x) \} = -\infty$ except if $b = 0$, in which case $f_* = 0$ (same for max)
- $\min\{ f(x) : x \in X \}$, X hyperrectangle, ‣ Optimizing a linear function (same for max)
  n independent problems, as nothing links x_i and x_j for $i \neq i$
- n closed formulæ $O(1)$ each, almost the last time

- Separable (non-homogeneous) quadratic function:

  $f(x) = \sum_{i=1}^{n} [\, f_i(x_i) = a_i x_i^2 + b_i x_i \,]$, fixed $(a, b) \in \mathbb{R}^{2n}$

  = sum of n univariate quadratic (non-homogeneous) functions

- $f(x) = \|x\|^2 = \sum_{i=1}^{n} x_i^2$ an important special case



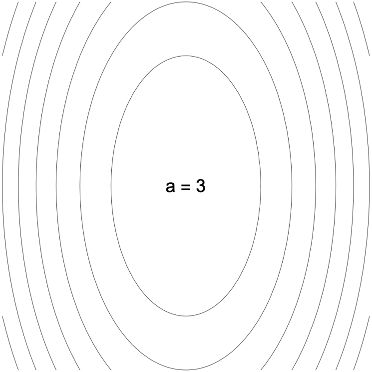
- $f(x_1, x_2) = a x_1^2 + x_2^2 \; [+0x_1 + 0x_2]$
- Contour plots for different values of a

- Separable (non-homogeneous) quadratic function:

  $f(x) = \sum_{i=1}^{n} [\, f_i(x_i) = a_i x_i^2 + b_i x_i \,]$, fixed $(a, b) \in \mathbb{R}^{2n}$

  = sum of n univariate quadratic (non-homogeneous) functions

- $f(x) = \| x \|^2 = \sum_{i=1}^{n} x_i^2$ an important special case



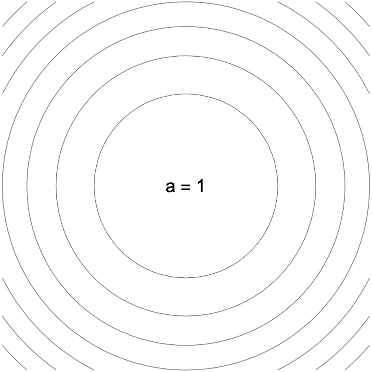
- $f(x_1, x_2) = a x_1^2 + x_2^2 \; [+0x_1 + 0x_2]$
- Contour plots for different values of a

- Separable (non-homogeneous) quadratic function:

  $f(x) = \sum_{i=1}^{n} [\, f_i(x_i) = a_i x_i^2 + b_i x_i \,]$, fixed $(a, b) \in \mathbb{R}^{2n}$

  $=$ sum of n univariate quadratic (non-homogeneous) functions

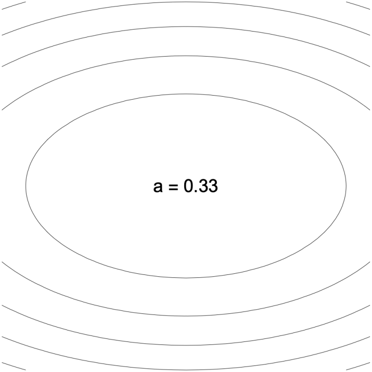
- $f(x) = \| x \|^2 = \sum_{i=1}^{n} x_i^2$ an important special case



- $f(x_1, x_2) = a x_1^2 + x_2^2 \; [+0x_1 + 0x_2]$
- Contour plots for different values of a
- For $a = 1$, perfect circles

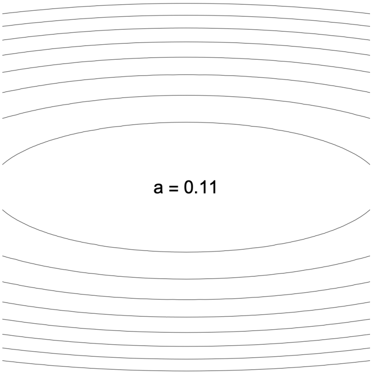
- Separable (non-homogeneous) quadratic function:
  $f(x) = \sum_{i=1}^{n} [\, f_i(x_i) = a_i x_i^2 + b_i x_i \,]$, fixed $(a, b) \in \mathbb{R}^{2n}$
  = sum of n univariate quadratic (non-homogeneous) functions

- $f(x) = \| x \|^2 = \sum_{i=1}^{n} x_i^2$ an important special case



- $f(x_1, x_2) = a x_1^2 + x_2^2$ $[+0x_1 + 0x_2]$
- Contour plots for different values of a
- For $a = 1$, perfect circles
- Larger / smaller a, more $\updownarrow$ / $\leftrightarrow$ elongated

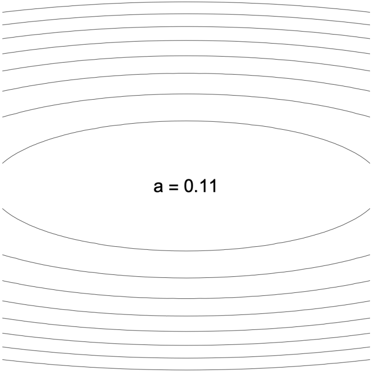
- Separable (non-homogeneous) quadratic function:
  $f(x) = \sum_{i=1}^{n} [\, f_i(x_i) = a_i x_i^2 + b_i x_i \,]$, fixed $(a, b) \in \mathbb{R}^{2n}$
  $=$ sum of n univariate quadratic (non-homogeneous) functions
- $f(x) = \| x \|^2 = \sum_{i=1}^{n} x_i^2$ an important special case



- $f(x_1, x_2) = a x_1^2 + x_2^2 \; [+0x_1 + 0x_2]$
- Contour plots for different values of a
- For $a = 1$, perfect circles
- Larger / smaller a, more $\updownarrow$ / $\leftrightarrow$ elongated
- Could be non-homogeneous,
  $[\, 0, 0 \,] \to [\, -b_1/2a_1, -b_2/2a_2 \,]$
- $O(n)$ ▸ Optimizing a quadratic non-homogeneous function,
  this is the last time

- Separable (non-homogeneous) quadratic function:
  $f(x) = \sum_{i=1}^{n} [\, f_i(x_i) = a_i x_i^2 + b_i x_i \,]$, fixed $(a, b) \in \mathbb{R}^{2n}$
  = sum of n univariate quadratic (non-homogeneous) functions

- $f(x) = \|x\|^2 = \sum_{i=1}^{n} x_i^2$ an important special case



- $f(x_1, x_2) = a x_1^2 + x_2^2 \; [+0x_1 + 0x_2]$
- Contour plots for different values of a
- For $a = 1$, perfect circles
- Larger / smaller a, more $\updownarrow$ / $\leftrightarrow$ elongated
- Could be non-homogeneous,
  $[\, 0, 0 \,] \rightarrow [\, -b_1/2a_1, -b_2/2a_2 \,]$
- $O(n)$ ▸ Optimizing a quadratic non-homogeneous function,
  this is the last time

- Not a general quadratic function, coming right next

- Nonseparable homogeneous quadratic function: fixed $Q \in \mathbb{R}^{n \times n}$ (n $Q_i \in \mathbb{R}^n$)

$$f(x) = \frac{1}{2} x^T Q x = \frac{1}{2} \left[\sum_{i=1}^n Q_{ii} x_i^2 + \sum_{i=1}^n \sum_{j=1,\, j \neq i}^n Q_{ij} x_i x_j \right]$$

- Not linear: $f(x+z) = \frac{1}{2}(x+z)^T Q (x+z) = f(x) + f(z) + z^T Q x$
- W.l.o.g. Q symmetric:

$$x^T Q x = [\,(x^T Q x) + (x^T Q x)^T\,] / 2 = x^T [\,(Q + Q^T) / 2\,] x$$

- f symmetric: $f(x) = f(-x) \implies$ "centred in $x = 0$"
- Tomography: $\varphi(\alpha) = f(\alpha d) = \frac{1}{2} \alpha^2 (d^T Q d) \implies$

  homogeneous quadratic univariate, sign and steepness depend on $d^T Q d$

- Need to know about signs of $d^T Q d$ when d changes: (multi)linear algebra
- Crucial stuff: spectral decomposition, eigenvalues, eigenvectors of Q

- $Q \in \mathbb{R}^{n \times n}$, $v \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$ s.t. $Qv = \lambda v$: v eigenvector of Q, λ eigenvalue
- v eigenvector $\equiv Qv = \lambda v \equiv Q(-v) = \lambda(-v) \equiv -v$ eigenvector
- Q symmetric $\implies$ has n distinct eigenvectors $H_1, H_2, \ldots, H_n$ and n (not necessarily distinct) corresponding real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$
- Eigenvectors can always be taken orthonormal: $H_i \perp H_j$ for $i \neq j$, $\| H_i \| = 1$ $\implies$ linearly independent (**check**) $\implies$ a(n orthonormal) basis of $\mathbb{R}^n$
- Spectral decomposition: $H = [\, H_1 , \ldots, H_n \,] \in \mathbb{R}^{n \times n}$, $\Lambda = \text{diag}(\, \lambda_1 , \ldots, \lambda_n \,)$

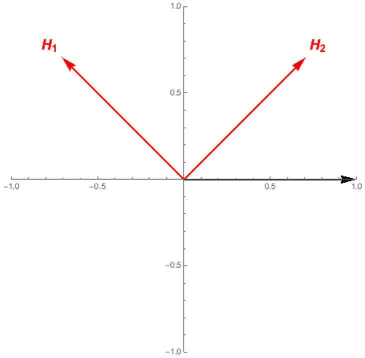
  $$Q = H \Lambda H^T = \lambda_1 H_1 H_1^T + \ldots + \lambda_n H_n H_n^T \qquad (\textbf{check})$$
- Notation: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ ($\lambda_1 = \max$, $\lambda_n = \min$)
- Variational characterization of eigenvalues:

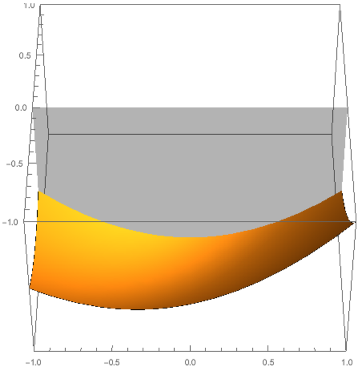
  $$\lambda_1 = \max\{\, d^T Q d \,/\, d^T d \;:\; d \neq 0 \,\} = \max\{\, d^T Q d \;:\; \| d \| = 1 \,\}$$
  $$\lambda_n = \min\{\, d^T Q d \,/\, d^T d \;:\; d \neq 0 \,\} = \min\{\, d^T Q d \;:\; \| d \| = 1 \,\}$$
- $Q \succ 0$ = positive definite if $\lambda_i > 0 \; \forall \, i \equiv \lambda_n > 0 \equiv d^T Q d > 0 \; \forall \, d \neq 0$

  $Q \succeq 0$ = positive semi-definite if $\lambda_i \geq 0 \; \forall \, i \equiv \lambda_n \geq 0 \equiv d^T Q d \geq 0 \; \forall \, d \neq 0$

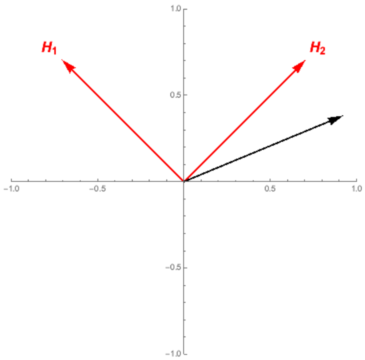
  negative definite ($\prec$), semi-definite ($\preceq$), indefinite ($\not\succeq\!\!\not\preceq$) obvious

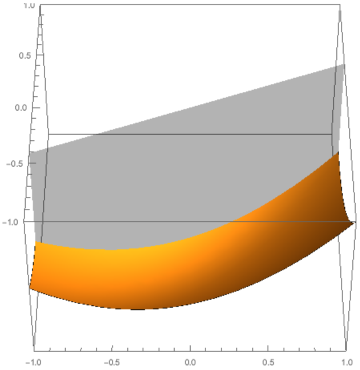
- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$



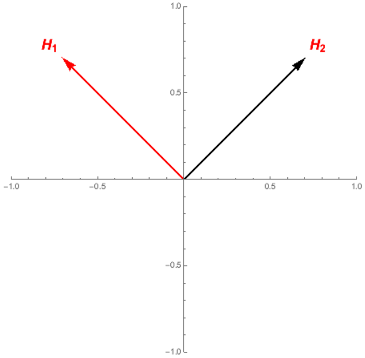


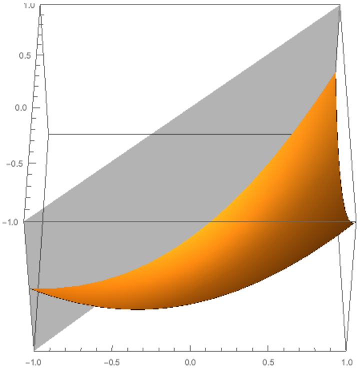
- Fundamental relation: $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$



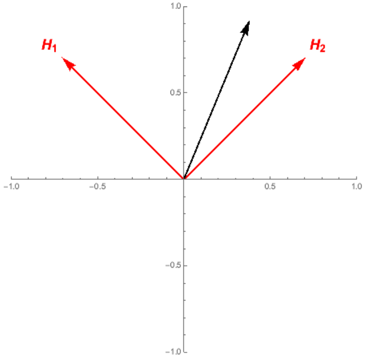


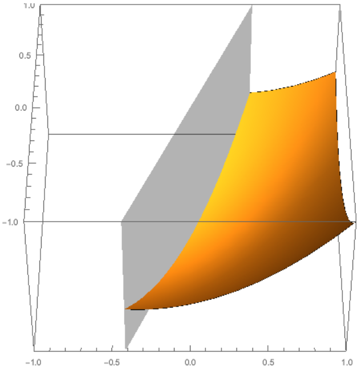
- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ **(check)**
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$



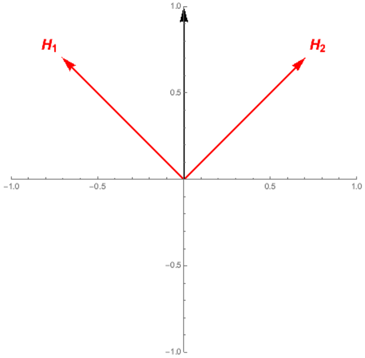


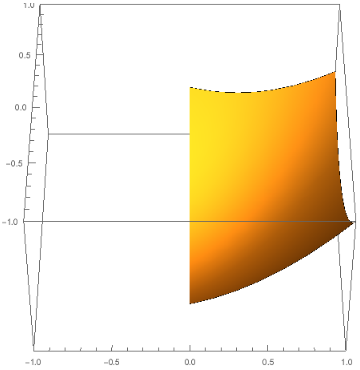
- Fundamental relation: $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$



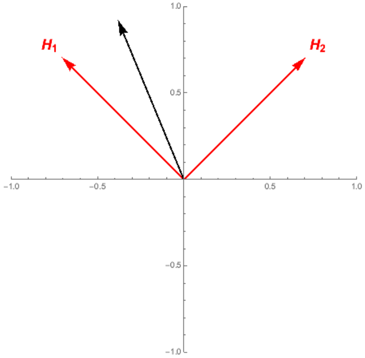


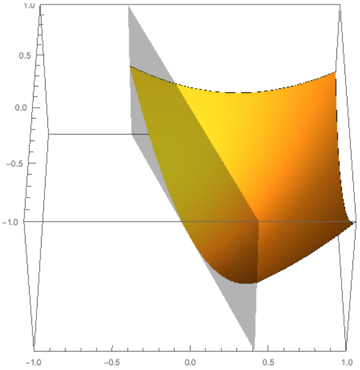
- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$



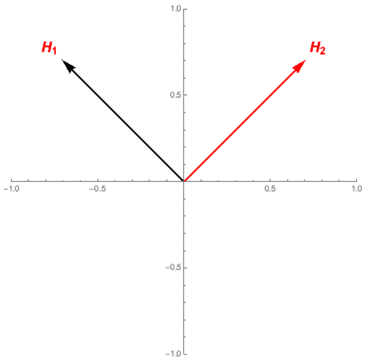


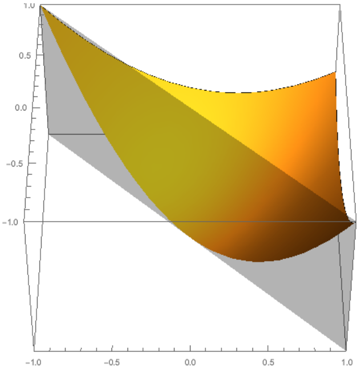
- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$



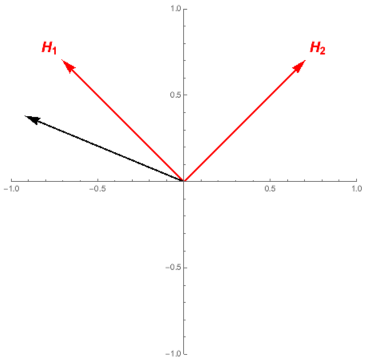


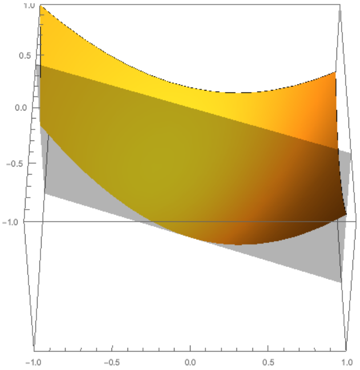
- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$



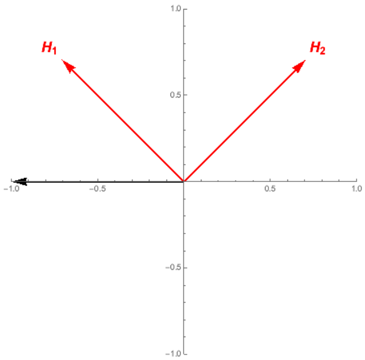


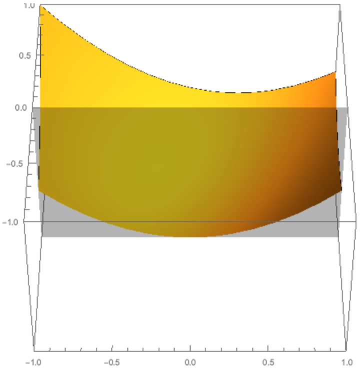
- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$



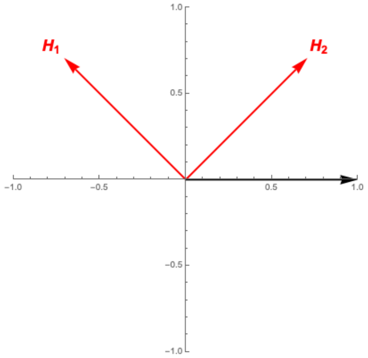


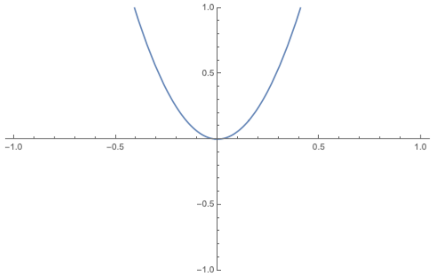
- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$





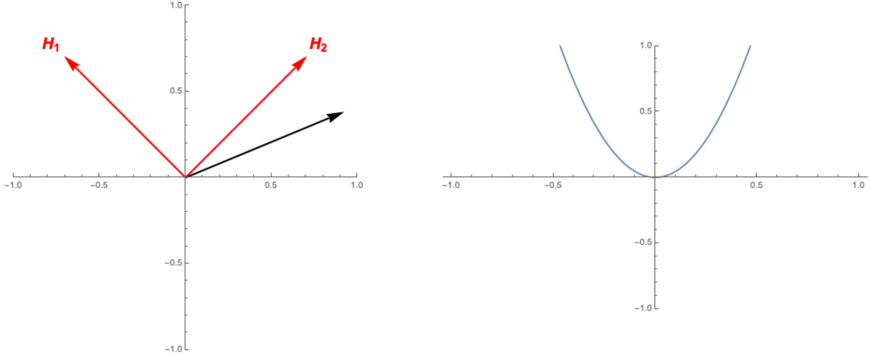
- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$





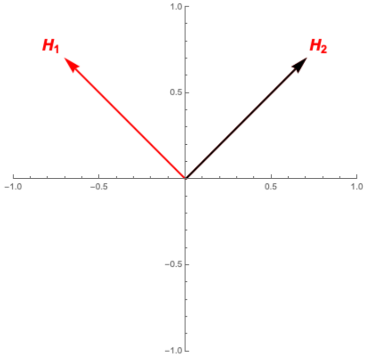
- $d^T Q d > 0 \; \forall \, d$, steepness change with d
-

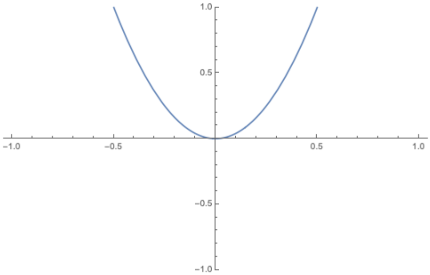
- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$



- $d^T Q d > 0 \; \forall \, d$, steepness change with d
-

- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

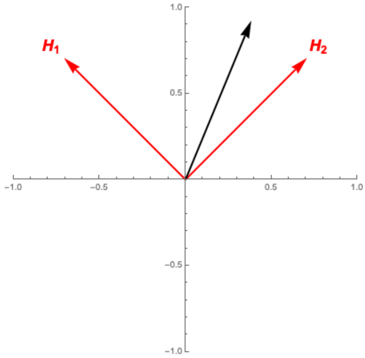


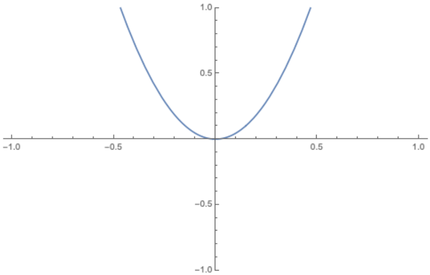


- $d^T Qd > 0 \; \forall \, d$, steepness change with d
- least steep along H_2 ($\lambda_2 = 4$)

- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)

- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

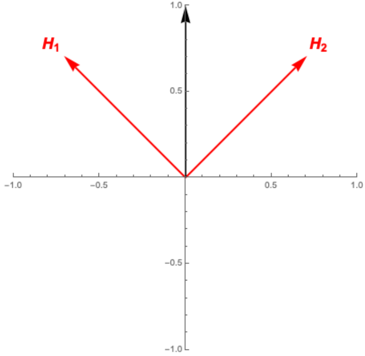


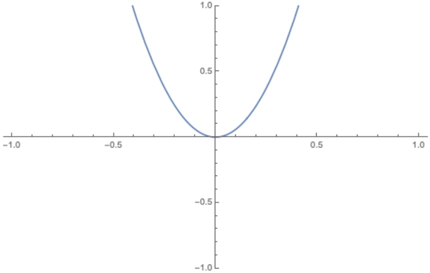


- $d^T Q d > 0 \; \forall \, d$, steepness change with d

-

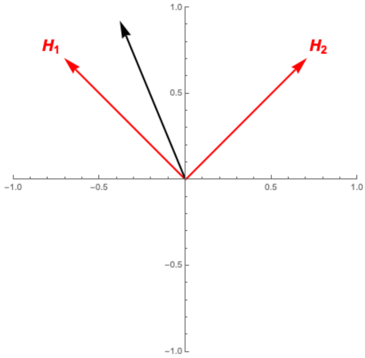
- Fundamental relation: $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

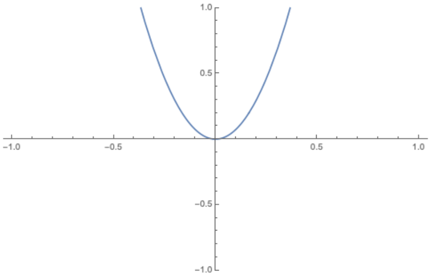




- $d^T Q d > 0 \; \forall \, d$, steepness change with d
-

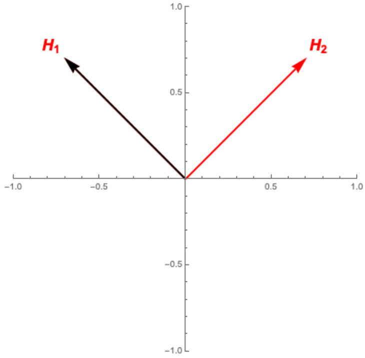
- Fundamental relation: $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

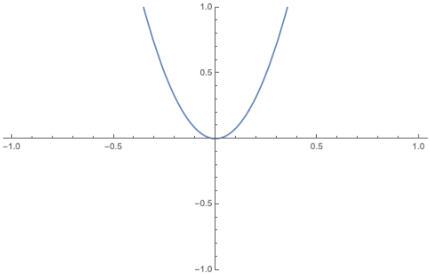




- $d^T Q d > 0 \; \forall \, d$, steepness change with d
-

- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

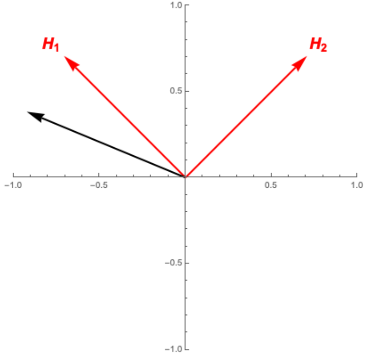


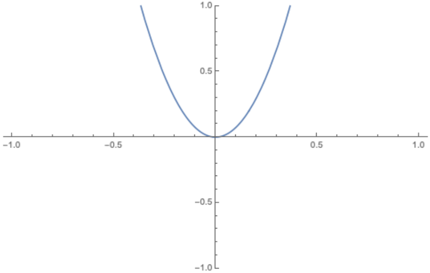


- $d^T Q d > 0 \; \forall \, d$, steepness change with d
- steepest along H_1 ($\lambda_1 = 8$)

- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)

- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

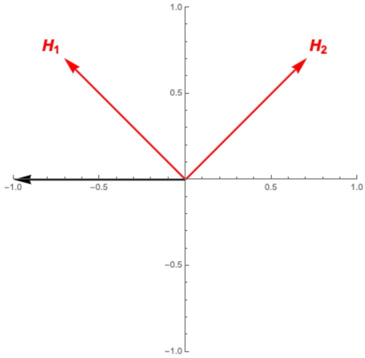


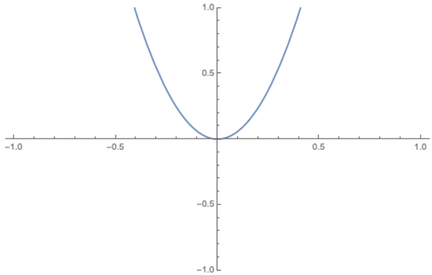


- $d^T Q d > 0 \; \forall \, d$, steepness change with d

- intermediate steepness "in between"

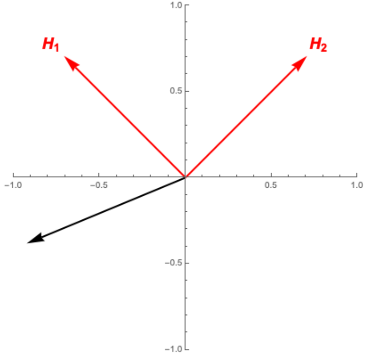
- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

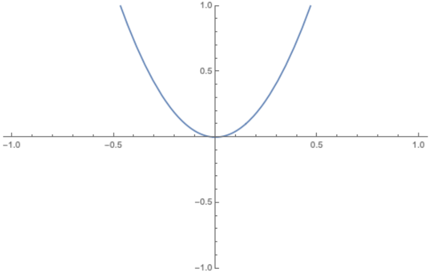




- $d^T Qd > 0 \; \forall \, d$, steepness change with d
- intermediate steepness "in between"

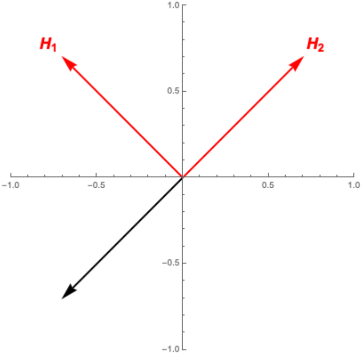
- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

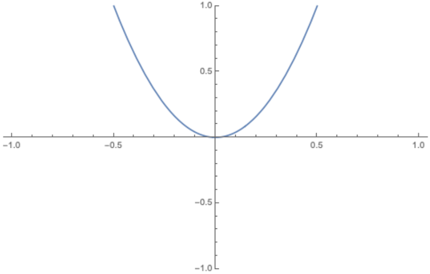




- $d^T Q d > 0 \ \forall\, d$, steepness change with d
- intermediate steepness "in between"

- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

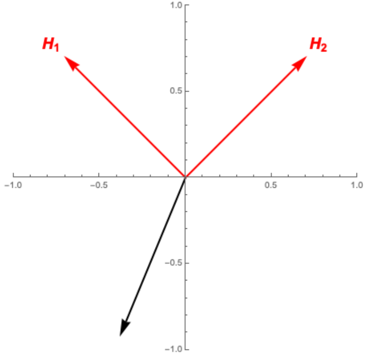


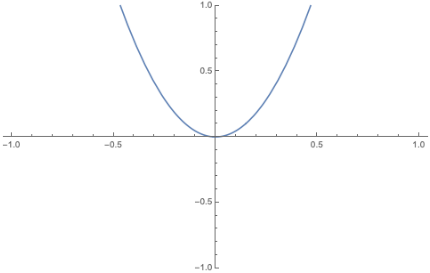


- $d^T Q d > 0 \; \forall \, d$, steepness change with d
- least steep along $-H_2$ ($\lambda_2 = 4$)

- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)

- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

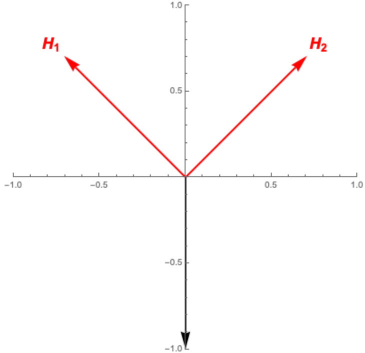


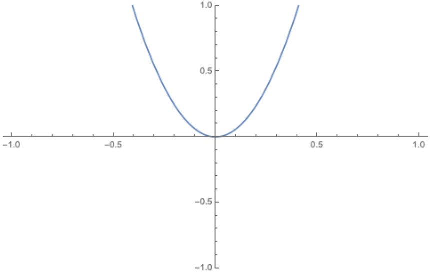


- $d^T Q d > 0 \; \forall \, d$, steepness change with d

- intermediate steepness "in between"

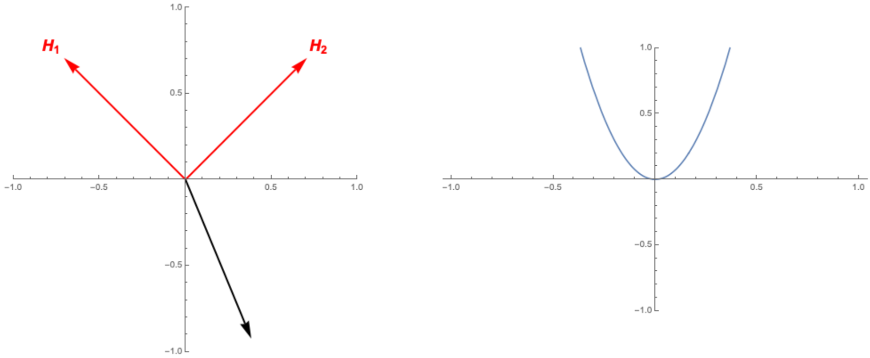
- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$





- $d^T Q d > 0 \; \forall \, d$, steepness change with d
- intermediate steepness "in between"

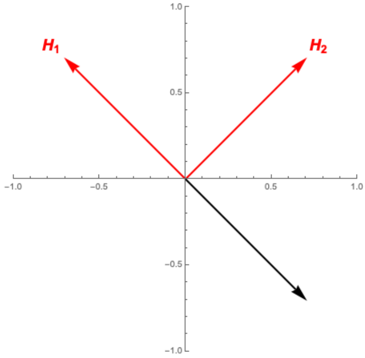
- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

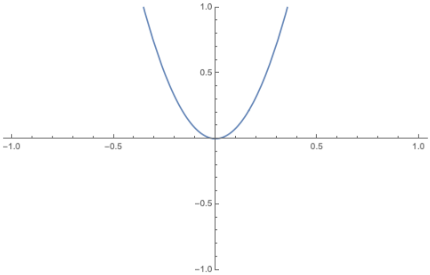


- $d^T Q d > 0 \; \forall \, d$, steepness change with d
- intermediate steepness "in between"

- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)

- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

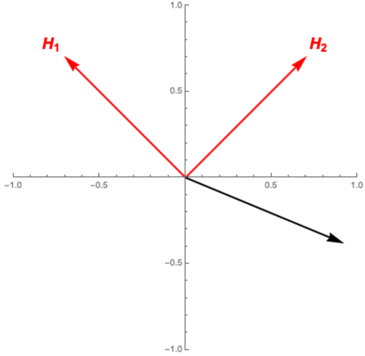


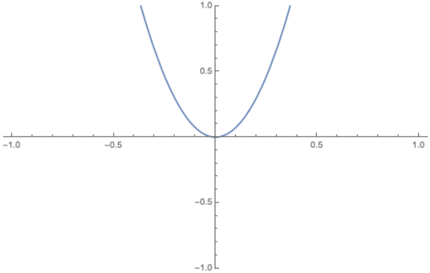


- $d^T Q d > 0 \; \forall \, d$, steepness change with d

- steepest along $-H_1$ ($\lambda_1 = 8$)

- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

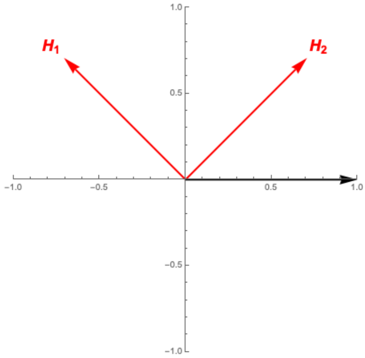


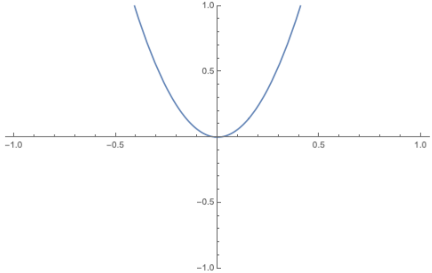


- $d^T Q d > 0 \; \forall \, d$, steepness change with d
- intermediate steepness "in between"

- Fundamental relation: $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$ (**check**)

- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$

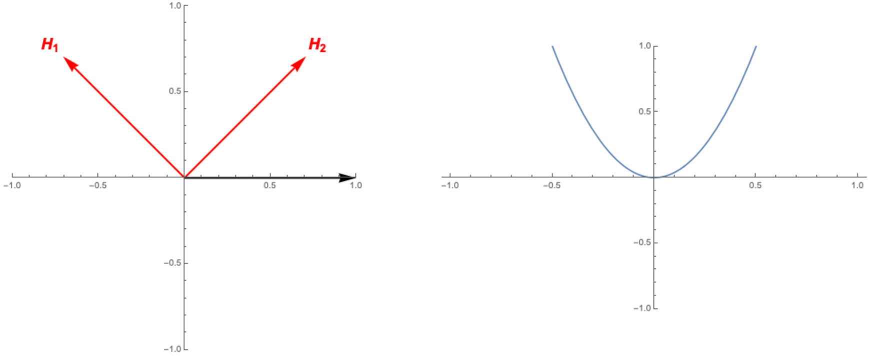




- $d^T Q d > 0 \; \forall \, d$, steepness change with d

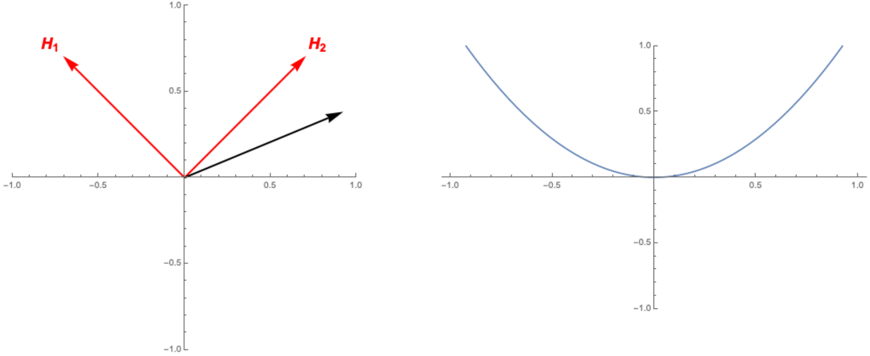
- intermediate steepness "in between"

- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$



- $d^T Q d \geq 0 \; \forall \, d$, but $\exists \, d$ s.t. $d^T Q d = 0$
-

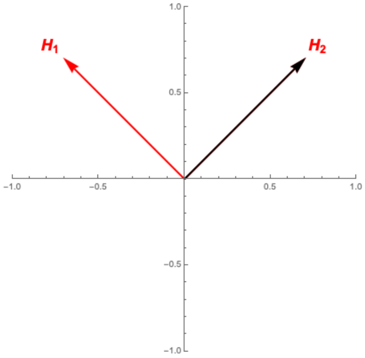
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$

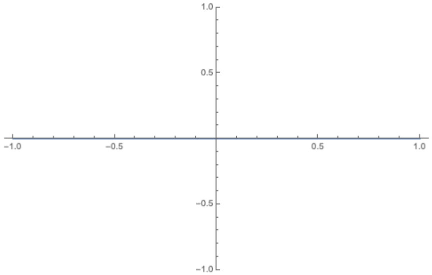


- $d^T Q d \geq 0 \; \forall \, d$, but $\exists \, d$ s.t. $d^T Q d = 0$
-

- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$

- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$

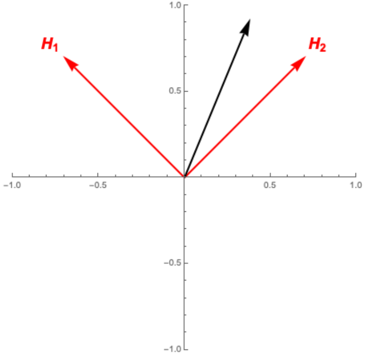


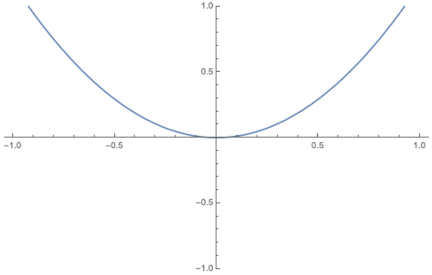


- $d^T Q d \geq 0 \; \forall \, d$, but $\exists \, d$ s.t. $d^T Q d = 0$

- completely flat along H_2 ($\lambda_2 = 0$)

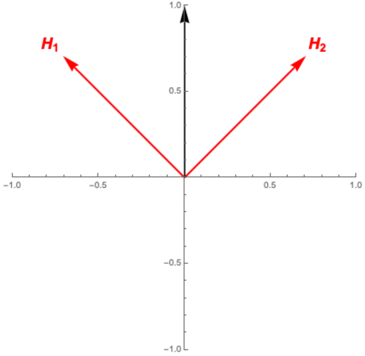
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$

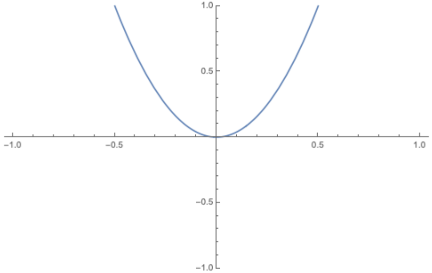




- $d^T Q d \geq 0 \; \forall \, d$, but $\exists \, d$ s.t. $d^T Q d = 0$
-

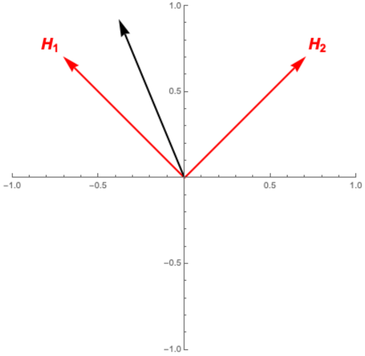
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$

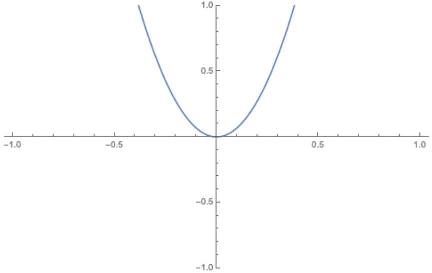




- $d^T Q d \geq 0 \; \forall \, d$, but $\exists \, d$ s.t. $d^T Q d = 0$
-

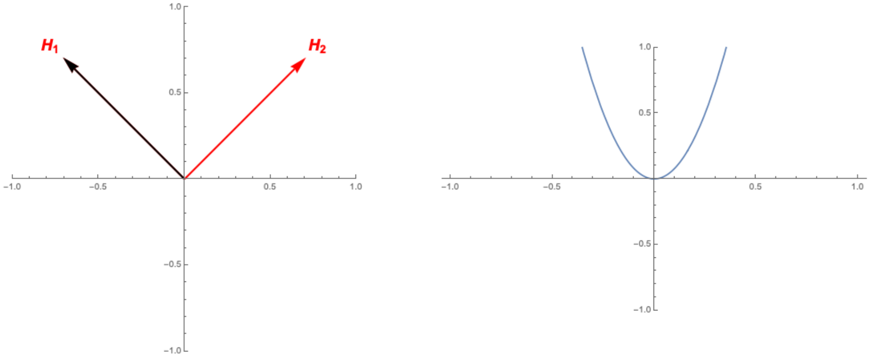
- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$





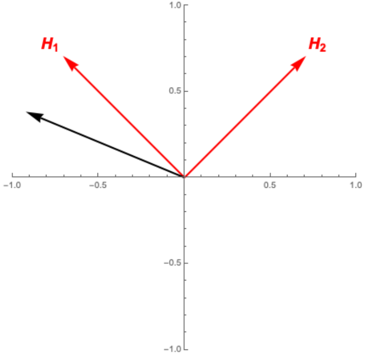
- $d^T Q d \geq 0 \;\forall\, d$, but $\exists\, d$ s.t. $d^T Q d = 0$
-

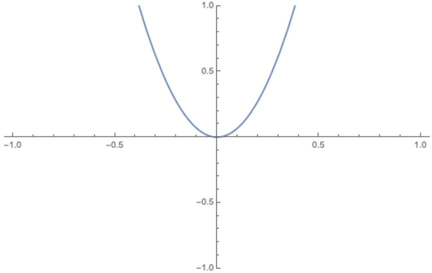
- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$



- $d^T Q d \geq 0 \;\forall\, d$, but $\exists\, d$ s.t. $d^T Q d = 0$
- steepest along H_1 ($\lambda_1 = 8$)

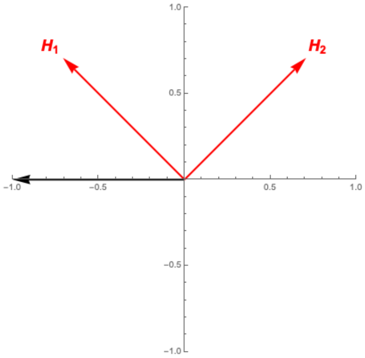
- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$

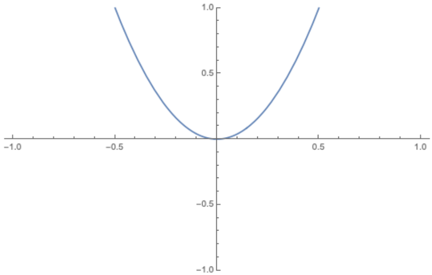




- $d^T Qd \geq 0 \;\forall\, d$, but $\exists\, d$ s.t. $d^T Qd = 0$
- intermediate steepness “in between”

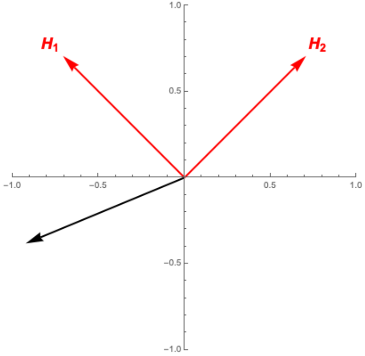
- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$

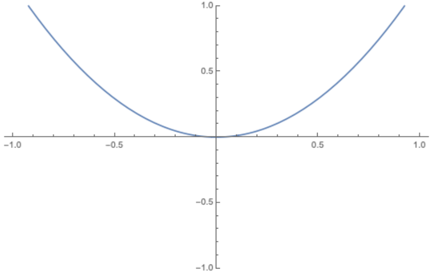




- $d^T Q d \geq 0 \; \forall \, d$, but $\exists \, d$ s.t. $d^T Q d = 0$
- intermediate steepness "in between"

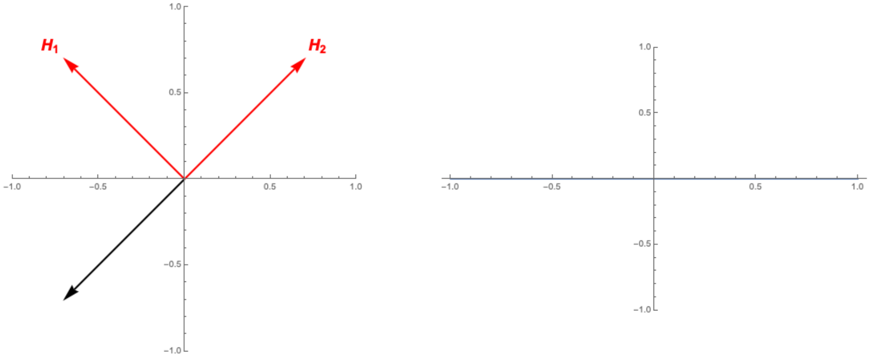
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$





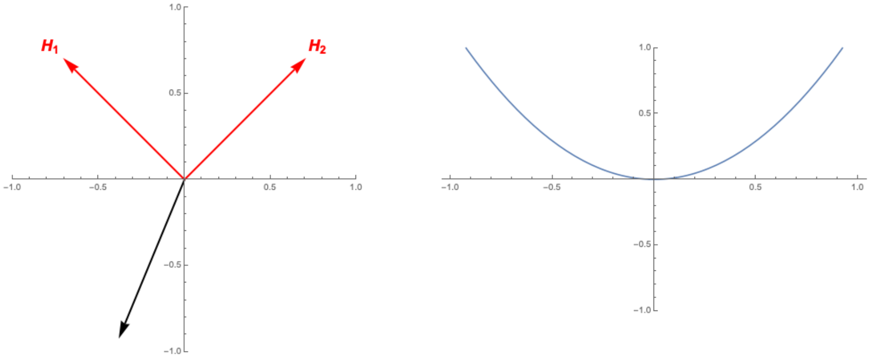
- $d^T Q d \geq 0 \; \forall \, d$, but $\exists \, d$ s.t. $d^T Q d = 0$
- intermediate steepness “in between”

- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$



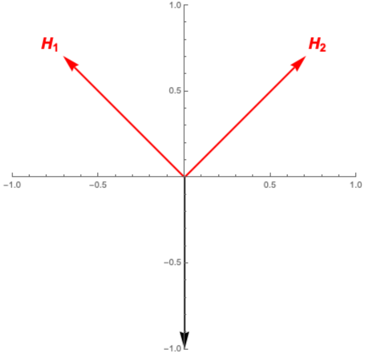
- $d^T Q d \geq 0 \; \forall \, d$, but $\exists \, d$ s.t. $d^T Q d = 0$
- completely flat along $-H_2$ ($\lambda_2 = 0$)

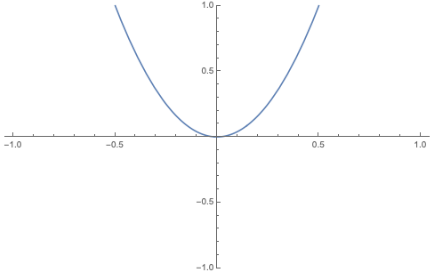
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$



- $d^T Q d \geq 0 \; \forall \, d$, but $\exists \, d$ s.t. $d^T Q d = 0$
- intermediate steepness “in between”

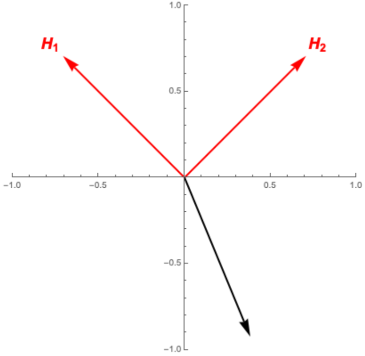
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$

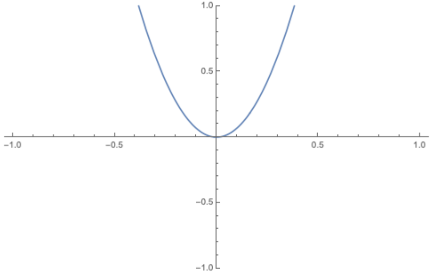




- $d^T Qd \geq 0 \; \forall \, d$, but $\exists \, d$ s.t. $d^T Qd = 0$
- intermediate steepness "in between"

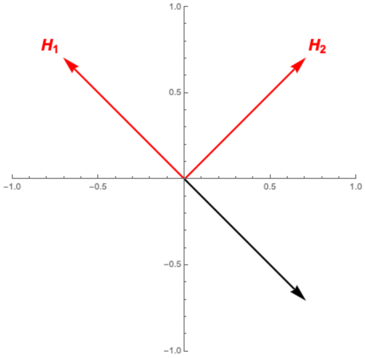
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$

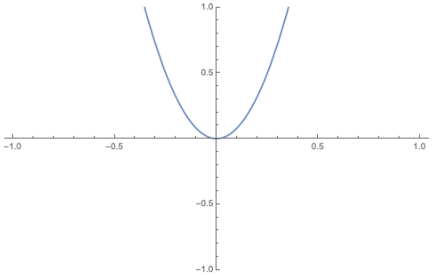




- $d^T Q d \geq 0 \;\forall\, d$, but $\exists\, d$ s.t. $d^T Q d = 0$
- intermediate steepness "in between"

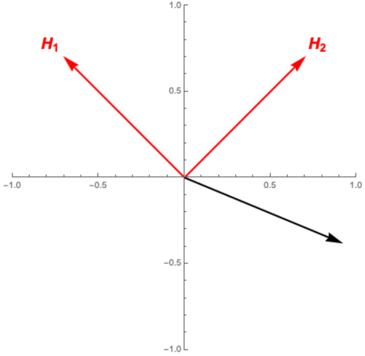
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$

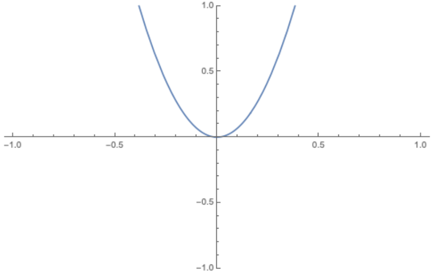




- $d^T Q d \geq 0 \;\forall\, d$, but $\exists\, d$ s.t. $d^T Q d = 0$
- steepest along $-H_1$ $(\lambda_1 = 8)$

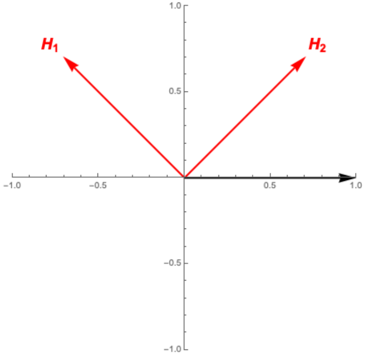
- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$

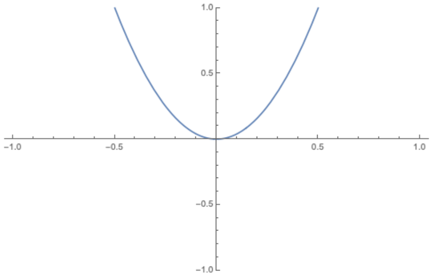




- $d^T Qd \geq 0 \; \forall \, d$, but $\exists \, d$ s.t. $d^T Qd = 0$
- intermediate steepness “in between”

- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$

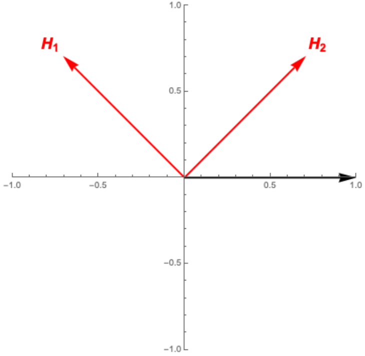


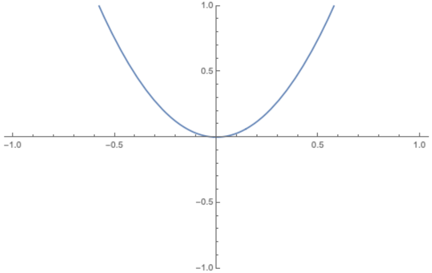


- $d^T Q d \geq 0 \; \forall \, d$, but $\exists \, d$ s.t. $d^T Q d = 0$
- intermediate steepness "in between"

- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$

- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \not\succeq 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

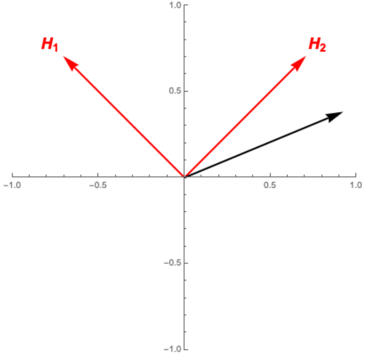


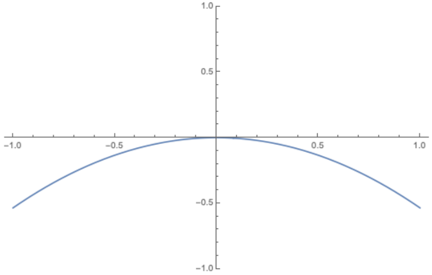


- $d^T Q d$ can be both > 0 and < 0

-

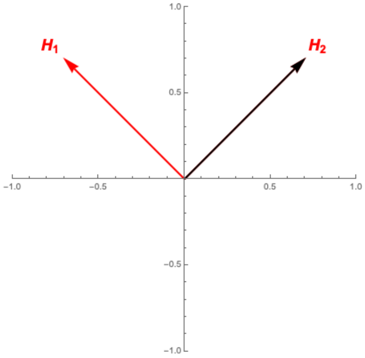
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

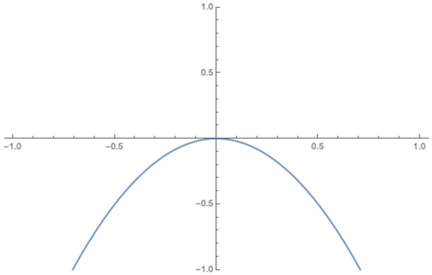




- $d^T Qd$ can be both > 0 and < 0
-

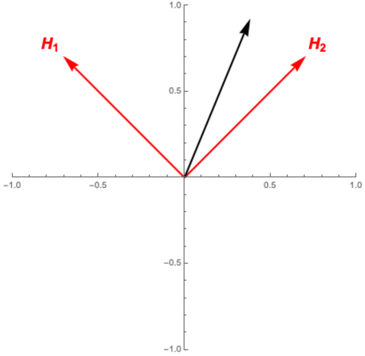
- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

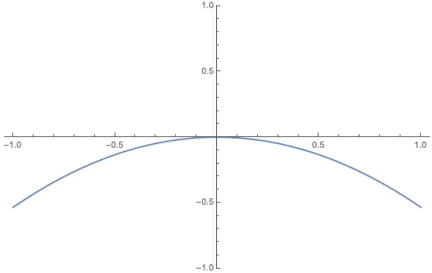




- $d^T Qd$ can be both > 0 and < 0
- steepest negative along H_2 ($\lambda_2 = -2$)

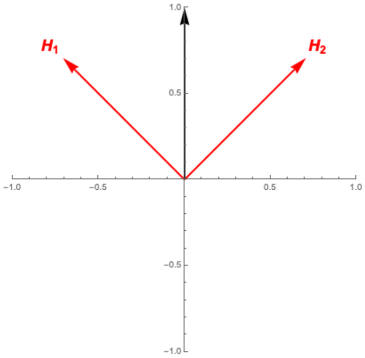
- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

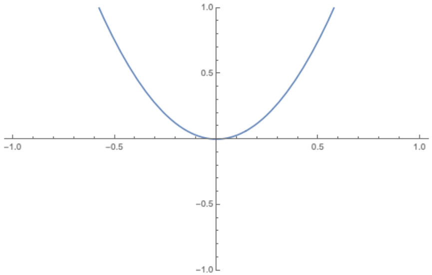




- $d^T Q d$ can be both > 0 and < 0
-

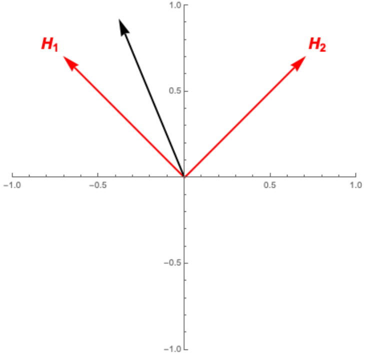
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

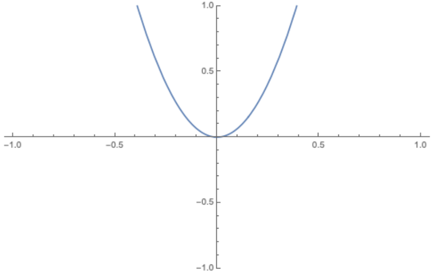




- $d^T Q d$ can be both > 0 and < 0
-

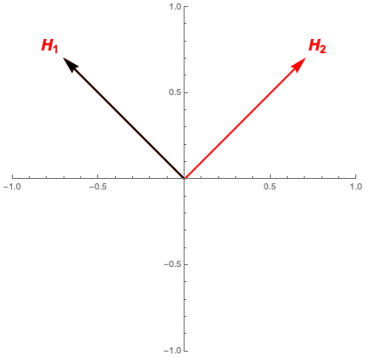
- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

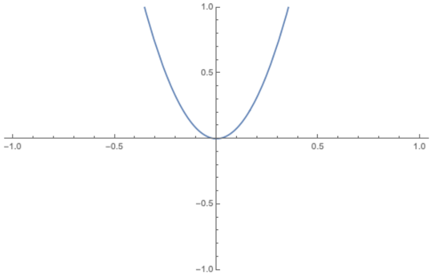




- $d^T Qd$ can be both > 0 and < 0
-

- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

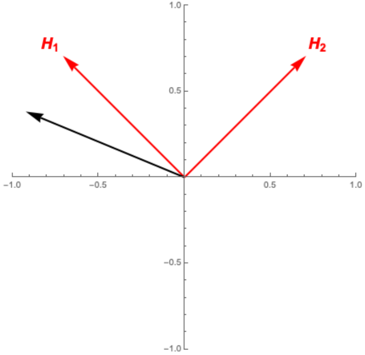


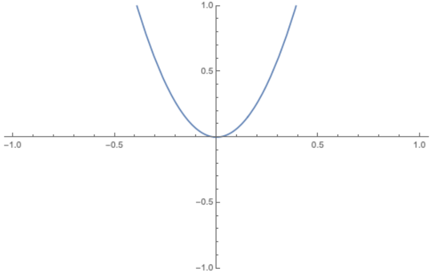


- $d^T Q d$ can be both > 0 and < 0
- steepest positive along H_1 ($\lambda_1 = 8$)

- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$

- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \qquad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \qquad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

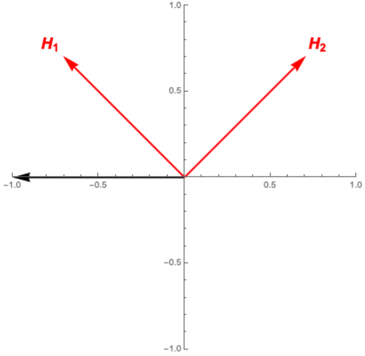


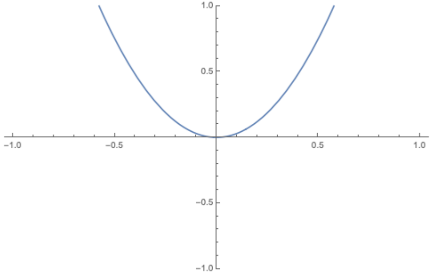


- $d^T Q d$ can be both > 0 and < 0

- intermediate steepness (positive or negative) “in between”

- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

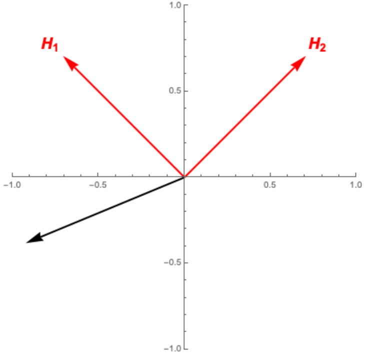


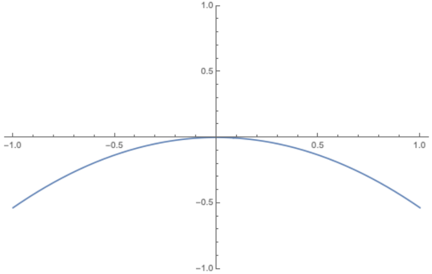


- $d^T Qd$ can be both > 0 and < 0
- intermediate steepness (positive or negative) “in between”

- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$

- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \dfrac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

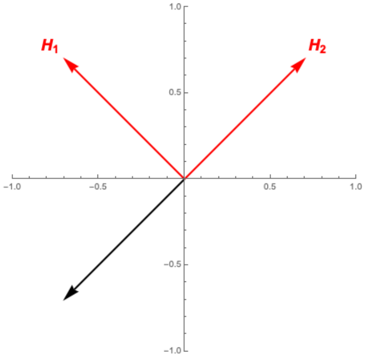


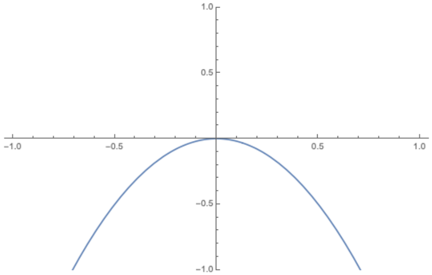


- $d^T Q d$ can be both > 0 and < 0

- intermediate steepness (positive or negative) “in between”

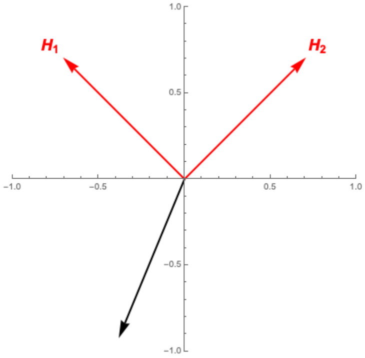
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

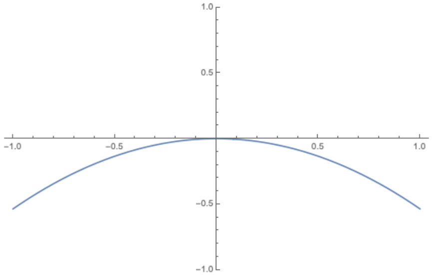




- $d^T Q d$ can be both > 0 and < 0
- steepest negative along $-H_2$ ($\lambda_2 = -2$)

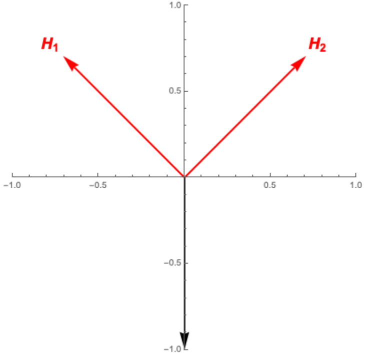
- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

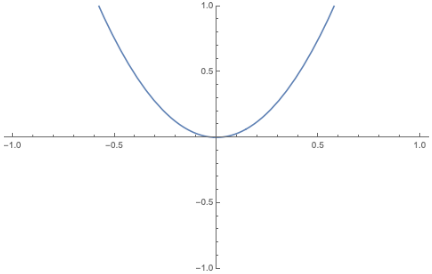




- $d^T Q d$ can be both > 0 and < 0
- intermediate steepness (positive or negative) “in between”

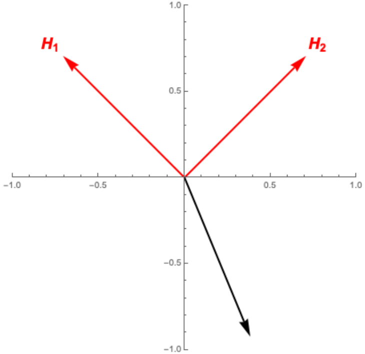
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

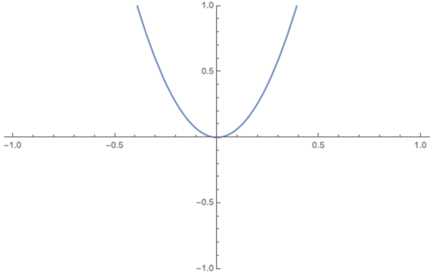




- $d^T Q d$ can be both > 0 and < 0
- intermediate steepness (positive or negative) "in between"

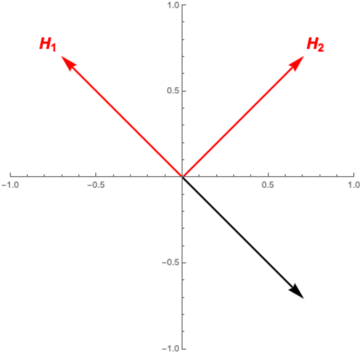
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

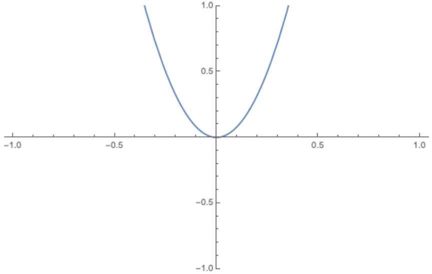




- $d^T Qd$ can be both > 0 and < 0
- intermediate steepness (positive or negative) "in between"

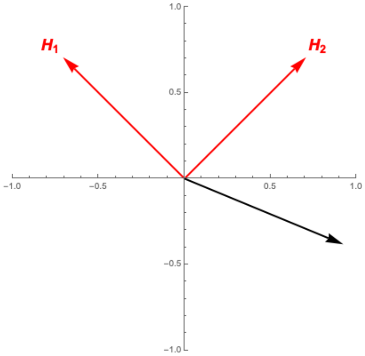
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

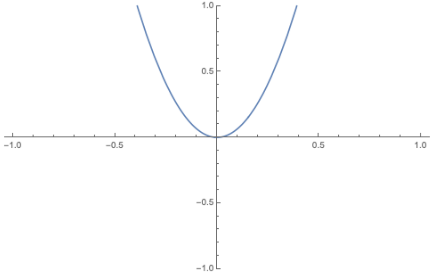




- d^TQd can be both > 0 and < 0
- steepest positive along $-H_1$ ($\lambda_1 = 8$)

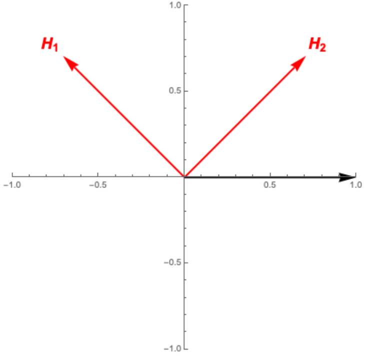
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

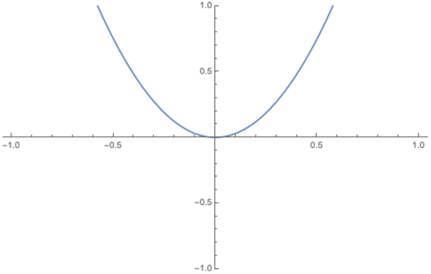




- $d^T Q d$ can be both > 0 and < 0
- intermediate steepness (positive or negative) “in between”

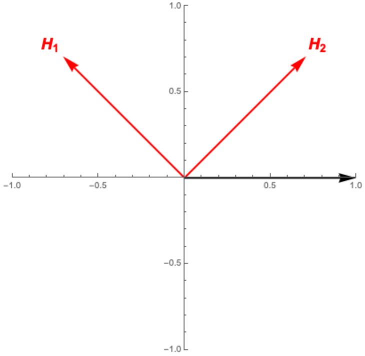
- Recall $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \asymp 0 \quad H = \frac{\sqrt{2}}{2}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$

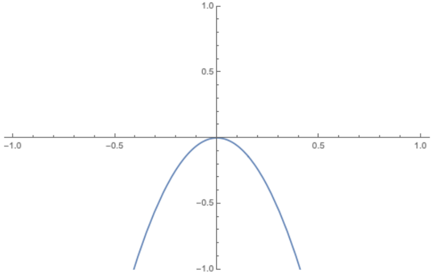




- $d^T Q d$ can be both > 0 and < 0
- intermediate steepness (positive or negative) “in between”

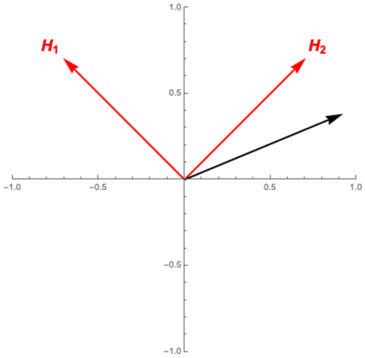
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

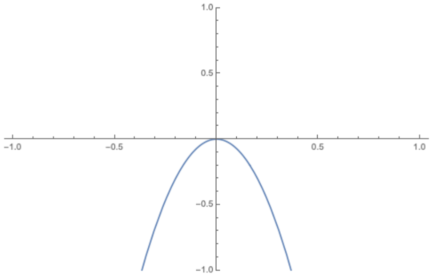




- $d^T Q d < 0 \; \forall \, d$, steepness change with d
-

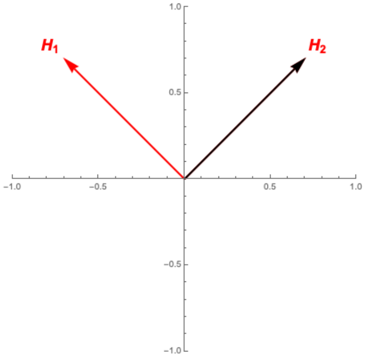
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

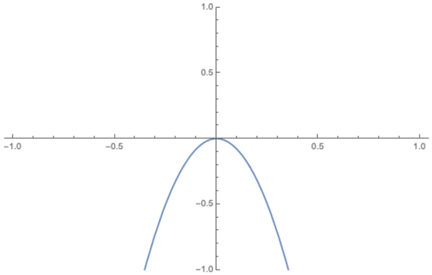




- $d^T Q d < 0 \; \forall \, d$, steepness change with d
-

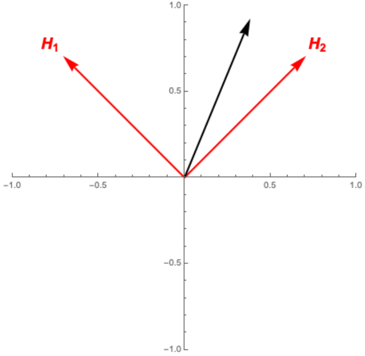
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

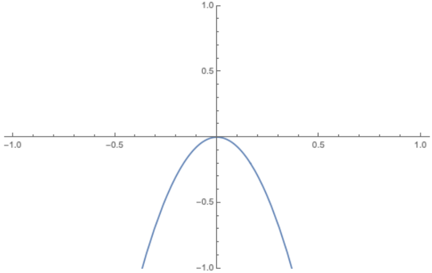




- $d^T Q d < 0 \; \forall \, d$, steepness change with d
- steepest negative along H_2 ($\lambda_2 = -8$)

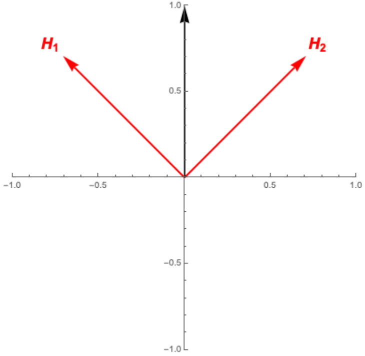
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

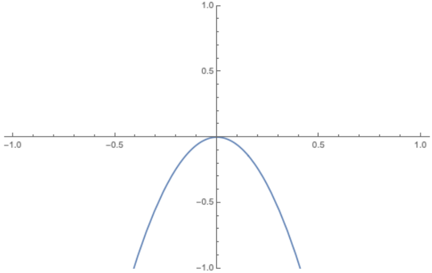




- $d^T Q d < 0 \; \forall \, d$, steepness change with d
-

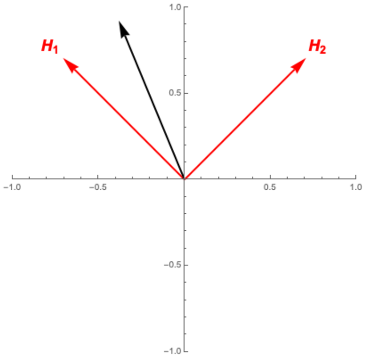
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

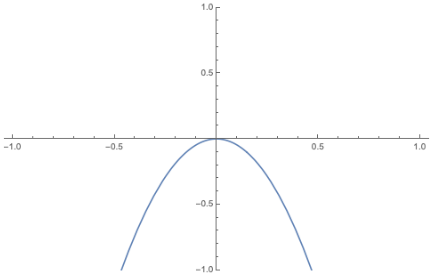




- $d^T Q d < 0 \; \forall \, d$, steepness change with d
-

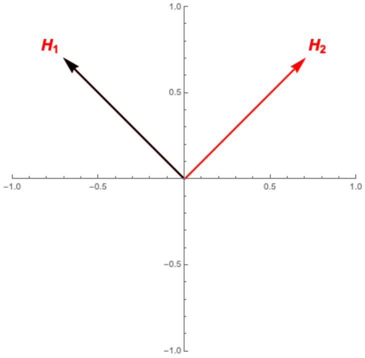
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

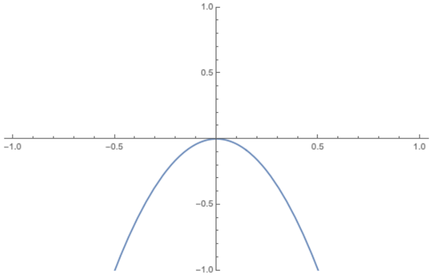




- $d^T Q d < 0 \; \forall \, d$, steepness change with d
-

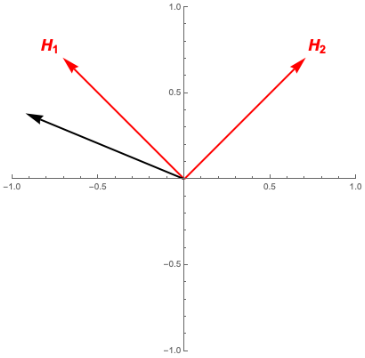
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

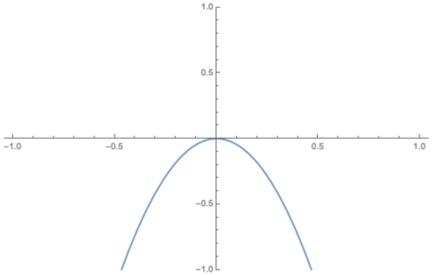




- $d^T Q d < 0 \; \forall \, d$, steepness change with d
- least steep negative along H_1 ($\lambda_1 = -4$)

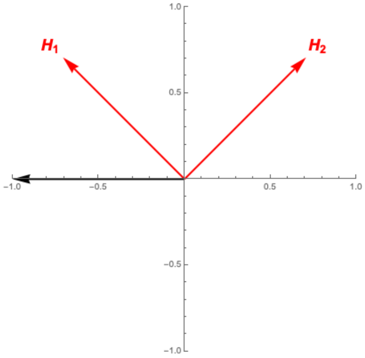
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

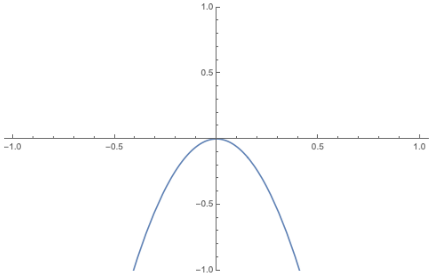




- $d^T Q d < 0 \; \forall \, d$, steepness change with d
- intermediate steepness (negative) "in between"

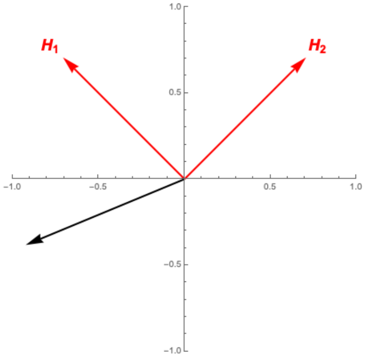
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

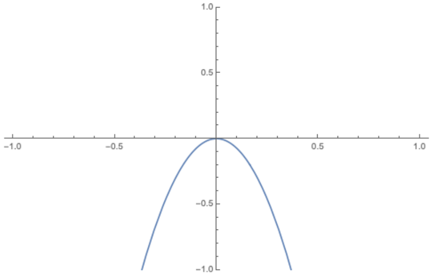




- $d^T Q d < 0 \; \forall \, d$, steepness change with d
- intermediate steepness (negative) "in between"

- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

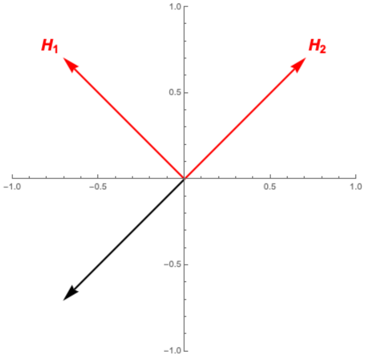


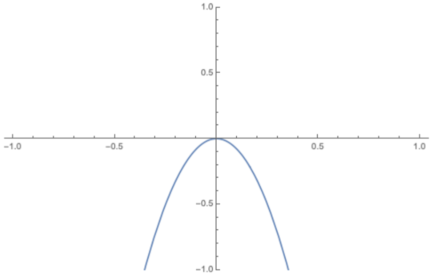


- $d^T Q d < 0 \; \forall \, d$, steepness change with d
- intermediate steepness (negative) "in between"

- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$

- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

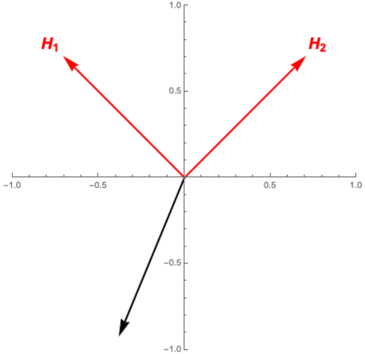


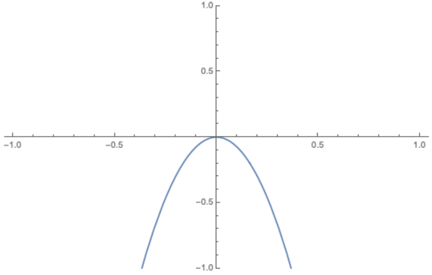


- $d^T Q d < 0 \; \forall \, d$, steepness change with d

- steepest negative along $-H_2$ ($\lambda_2 = -8$)

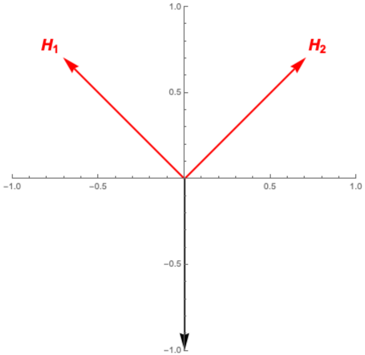
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

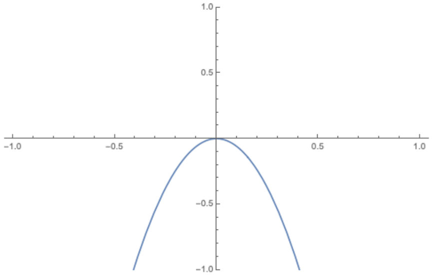




- $d^T Q d < 0 \; \forall \, d$, steepness change with d
- intermediate steepness (negative) “in between”

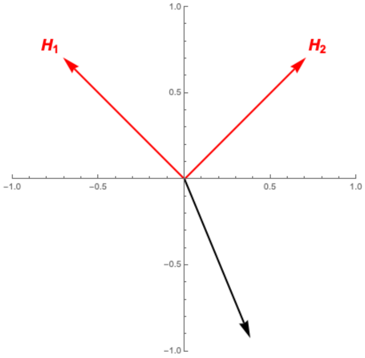
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

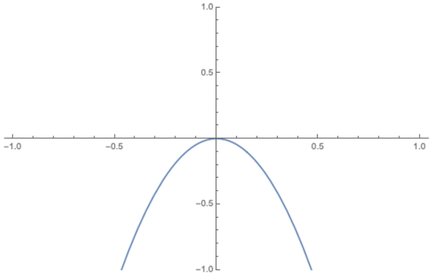




- $d^T Q d < 0 \; \forall \, d$, steepness change with d
- intermediate steepness (negative) “in between”

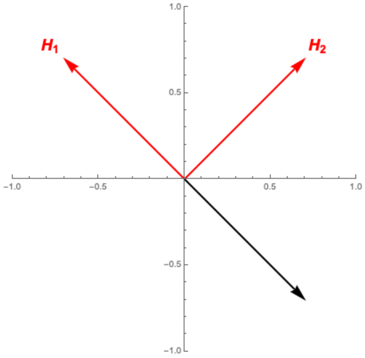
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

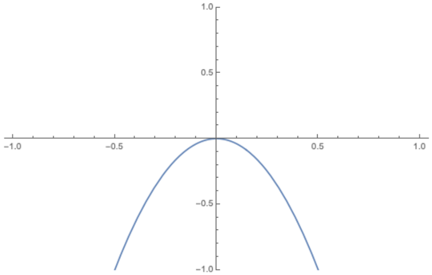




- $d^T Q d < 0 \; \forall \, d$, steepness change with d
- intermediate steepness (negative) "in between"

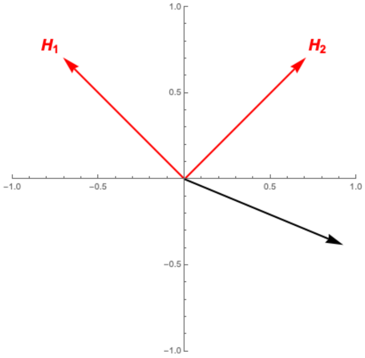
- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

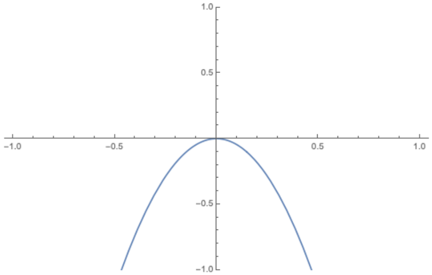




- $d^T Q d < 0 \; \forall \, d$, steepness change with d
- least steep negative along $-H_1$ ($\lambda_1 = -4$)

- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$

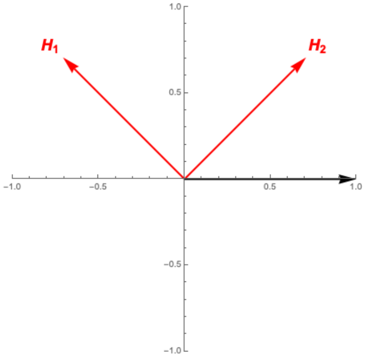


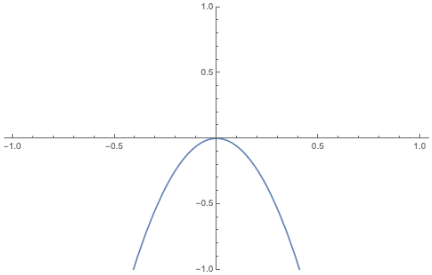


- $d^T Q d < 0 \; \forall \, d$, steepness change with d
- intermediate steepness (negative) “in between”

- Recall $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$

- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0 \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad \lambda = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$



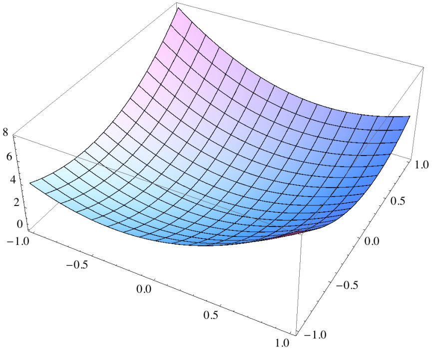


- $d^T Q d < 0 \; \forall \, d$, steepness change with d

- intermediate steepness (negative) "in between"

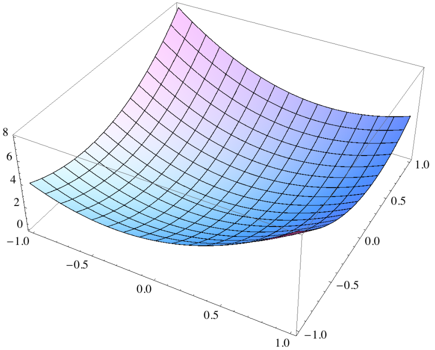
- All level sets centred in $x = 0$ by symmetry

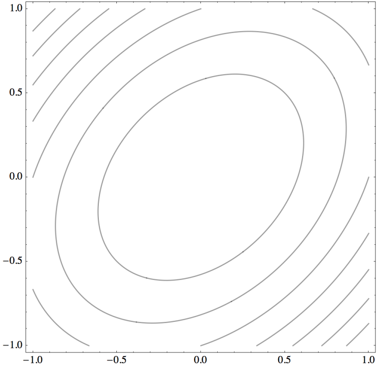
- All level sets centred in $x = 0$ by symmetry



- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0$ graph is a (convex) paraboloid

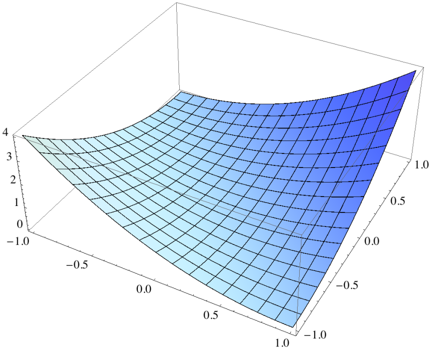
- All level sets centred in $x = 0$ by symmetry





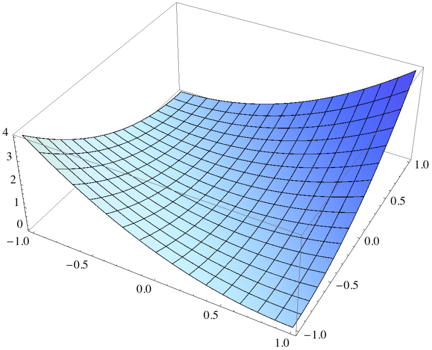
- $Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \succ 0$ graph is a (convex) paraboloid
  level sets are ellipsoids

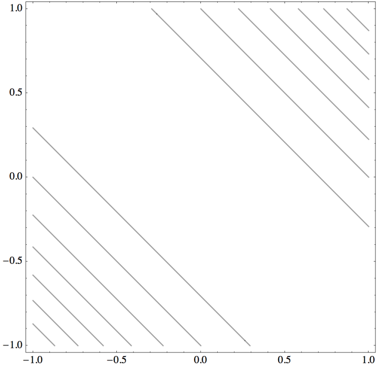
- All level sets centred in $x = 0$ by symmetry



- $Q = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \succeq 0$ graph is a degenerate paraboloid

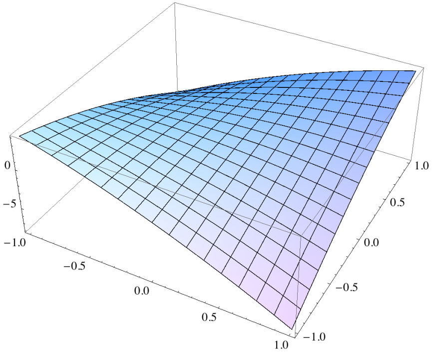
- All level sets centred in $x = 0$ by symmetry





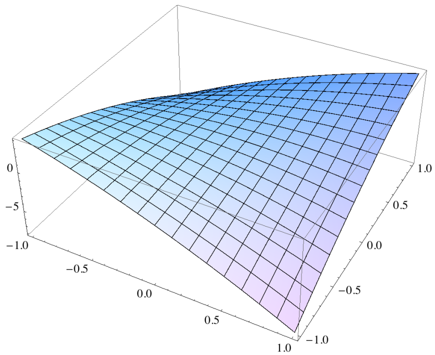
- $Q = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \succeq 0$ graph is a degenerate paraboloid
  level sets are degenerate ellipsoids

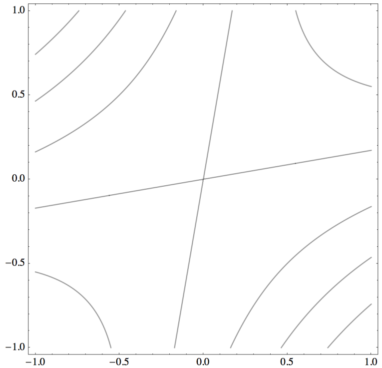
- All level sets centred in $x = 0$ by symmetry



- $Q = \begin{bmatrix} 3 & -5 \\ -5 & -3 \end{bmatrix} \asymp 0$ graph saddle-shaped (0 is a saddle point)

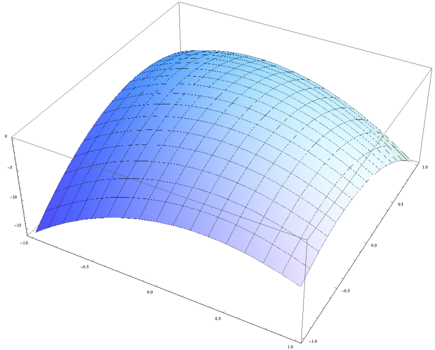
- All level sets centred in $x = 0$ by symmetry





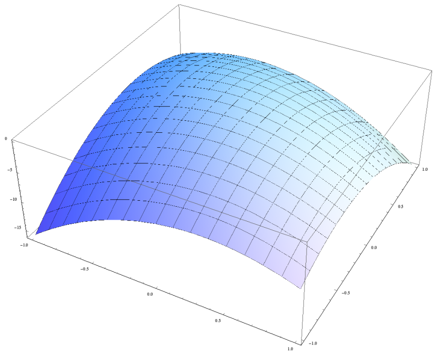
- $Q = \begin{bmatrix} 3 & -5 \\ -5 & -3 \end{bmatrix} \asymp 0$ graph saddle-shaped (0 is a saddle point)
  level sets are hyperboloids

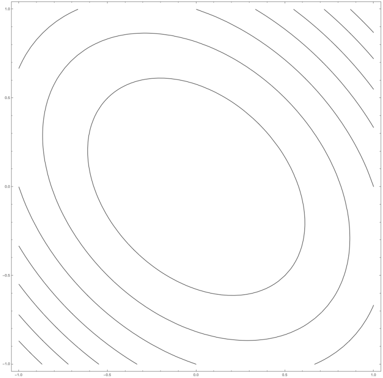
- All level sets centred in $x = 0$ by symmetry



- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0$ graph a (concave, i.e., "upside-down") paraboloid

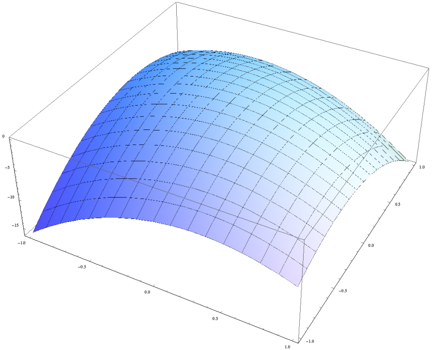
- All level sets centred in $x = 0$ by symmetry

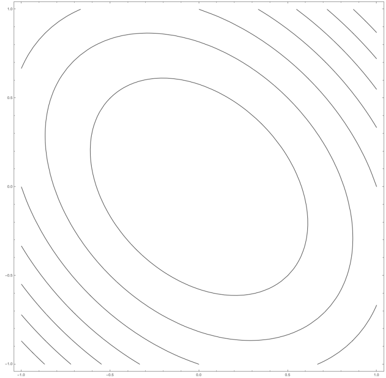




- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0$ graph a (concave, i.e., "upside-down") paraboloid level sets are ellipsoids again

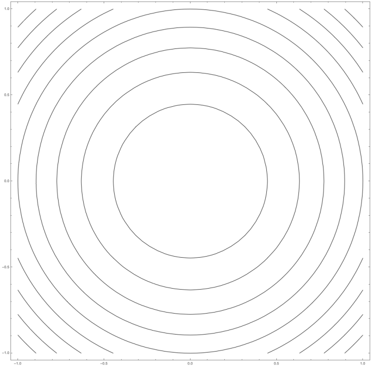
- All level sets centred in $x = 0$ by symmetry





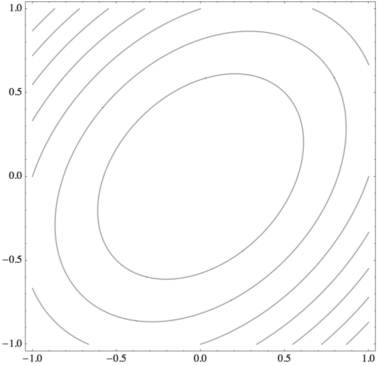
- $Q = \begin{bmatrix} -6 & -2 \\ -2 & -6 \end{bmatrix} \prec 0$ graph a (concave, i.e., "upside-down") paraboloid
  level sets are ellipsoids again

- Level sets can be precisely described in terms of H_i, λ_i



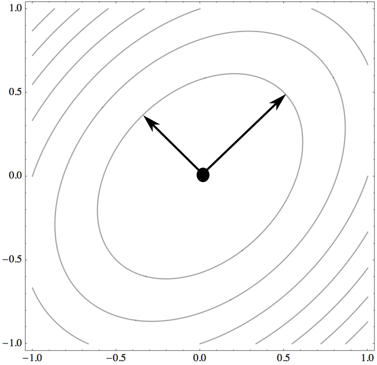
- $\|x\|_2^2 \equiv Q = H = \Lambda = I$: perfect circles

$$Q = H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad , \quad \lambda = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$



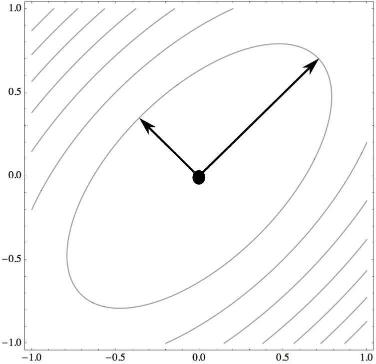
- Recall again $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $L(f, 1) \cap H_i \equiv \varphi_{H_i}(\alpha) = 1 \implies \lambda_i > 0$
- $\varphi_{H_i}(\alpha) = 1 \equiv \alpha = \sqrt{1/\lambda_i} \implies$

$$Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \quad , \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad , \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$



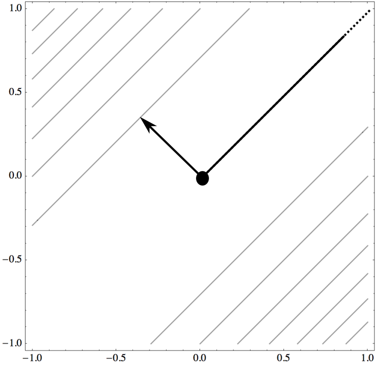
- Recall again $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $L(f, 1) \cap H_i \;\equiv\; \varphi_{H_i}(\alpha) = 1 \implies \lambda_i > 0$
- $\varphi_{H_i}(\alpha) = 1 \;\equiv\; \alpha = \sqrt{1/\lambda_i} \implies$
  $H_i \perp$ axes of $L(f, 1)$, length $\sqrt{1/\lambda_i}$

$$Q = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \quad , \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad , \quad \lambda = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$



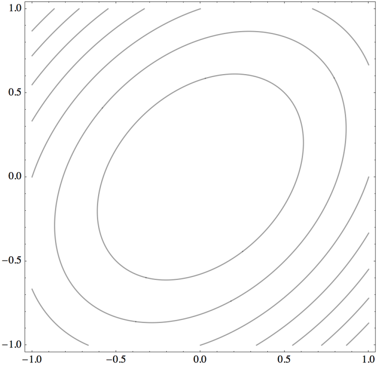
- Recall again $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $L(f, 1) \cap H_i \equiv \varphi_{H_i}(\alpha) = 1 \implies \lambda_i > 0$
- $\varphi_{H_i}(\alpha) = 1 \equiv \alpha = \sqrt{1/\lambda_i} \implies$
  $H_i \perp$ axes of $L(f, 1)$, length $\sqrt{1/\lambda_i}$
- $\lambda_i \searrow \equiv$ axis $\nearrow$,

$$Q = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix} \quad , \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad , \quad \lambda = \begin{bmatrix} 8 \\ 2 \end{bmatrix}$$



- Recall again $\varphi_{H_i}(\,\alpha\,) = \alpha^2 \lambda_i$
- $L(\,f\,,\,1\,) \cap H_i \;\equiv\; \varphi_{H_i}(\,\alpha\,) = 1 \;\Longrightarrow\; \lambda_i > 0$
- $\varphi_{H_i}(\,\alpha\,) = 1 \;\equiv\; \alpha = \sqrt{1/\lambda_i} \;\Longrightarrow$
  $H_i \perp$ axes of $L(\,f\,,\,1\,)$, length $\sqrt{1/\lambda_i}$
- $\lambda_i \searrow \equiv$ axis $\nearrow$, $\lambda_i = 0 \;\Longrightarrow\;$ "axis $\rightarrow \infty$"
- $\lambda_i < 0 \;\Longrightarrow\;$ same with $\varphi_{H_i}(\,\alpha\,) = \; -1$

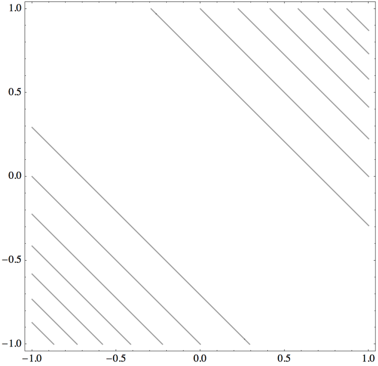
$$Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \quad , \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad , \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$$



- Recall again $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $L(f, 1) \cap H_i \equiv \varphi_{H_i}(\alpha) = 1 \implies \lambda_i > 0$
- $\varphi_{H_i}(\alpha) = 1 \equiv \alpha = \sqrt{1/\lambda_i} \implies$
  $H_i \perp$ axes of $L(f, 1)$, length $\sqrt{1/\lambda_i}$
- $\lambda_i \searrow \equiv$ axis $\nearrow$, $\lambda_i = 0 \implies$ "axis $\to \infty$"
- $\lambda_i < 0 \implies$ same with $\varphi_{H_i}(\alpha) = -1$

$$Q = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \quad , \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad , \quad \lambda = \begin{bmatrix} 8 \\ 0 \end{bmatrix}$$

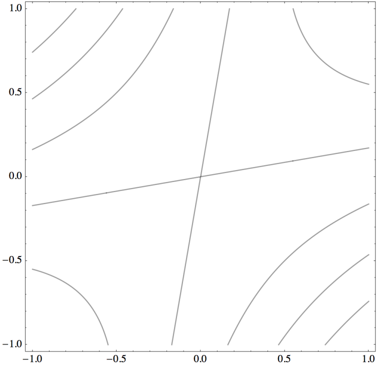
- All λ_i have the same sign: $f(x)$ either ≥ 0 or $\leq 0 \implies$ ellipsoids



- Recall again $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $L(f, 1) \cap H_i \equiv \varphi_{H_i}(\alpha) = 1 \implies \lambda_i > 0$
- $\varphi_{H_i}(\alpha) = 1 \equiv \alpha = \sqrt{1/\lambda_i} \implies$
  $H_i \perp$ axes of $L(f, 1)$, length $\sqrt{1/\lambda_i}$
- $\lambda_i \searrow \equiv$ axis $\nearrow$, $\lambda_i = 0 \implies$ "axis $\to \infty$"
- $\lambda_i < 0 \implies$ same with $\varphi_{H_i}(\alpha) = -1$

$$Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \quad , \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad , \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$$

- All λ_i have the same sign: $f(x)$ either ≥ 0 or $\leq 0 \implies$ ellipsoids
- Some $\lambda_i = 0 \implies$ "degenerate" ellipsoids (∞ axis)



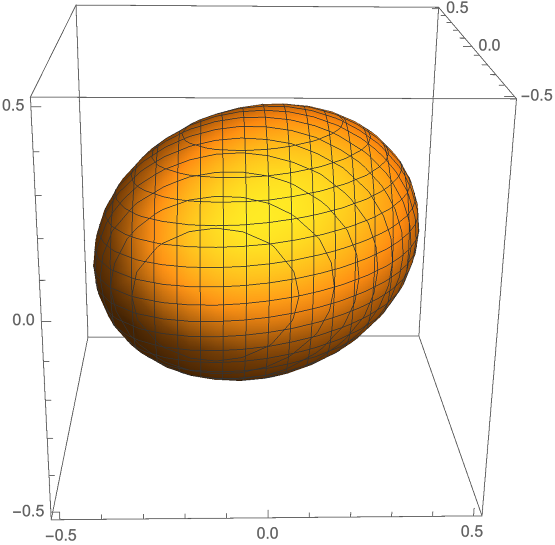
- Recall again $\varphi_{H_i}(\alpha) = \alpha^2 \lambda_i$
- $L(f, 1) \cap H_i \equiv \varphi_{H_i}(\alpha) = 1 \implies \lambda_i > 0$
- $\varphi_{H_i}(\alpha) = 1 \equiv \alpha = \sqrt{1/\lambda_i} \implies$
  $H_i \perp$ axes of $L(f, 1)$, length $\sqrt{1/\lambda_i}$
- $\lambda_i \searrow \equiv$ axis $\nearrow$, $\lambda_i = 0 \implies$ "axis $\to \infty$"
- $\lambda_i < 0 \implies$ same with $\varphi_{H_i}(\alpha) = -1$

$$Q = \begin{bmatrix} 3 & -5 \\ -5 & 3 \end{bmatrix} \quad , \quad H = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \quad , \quad \lambda = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$$

- All λ_i have the same sign: $f(x)$ either ≥ 0 or $\leq 0 \implies$ ellipsoids
- Some $\lambda_i = 0 \implies$ "degenerate" ellipsoids (∞ axis)
- $\lambda_i > 0$ and $\lambda_j < 0$: $\exists\, \alpha_i, \alpha_j$ s.t. $\varphi_{H_i}(\alpha_i) + \varphi_{H_j}(\alpha_j) = 0 \implies$ hyperboloids

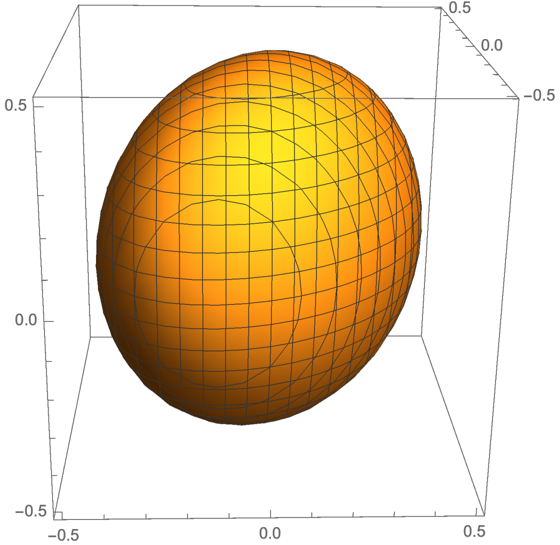
$$Q = \begin{bmatrix} 6 & -2 & 0 \\ -2 & 6 & 0 \\ 0 & 0 & 8 \end{bmatrix} \quad , \quad H = \begin{bmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad , \quad \lambda = \begin{bmatrix} 8 \\ 4 \\ 8 \end{bmatrix}$$

$L(f, 1)$



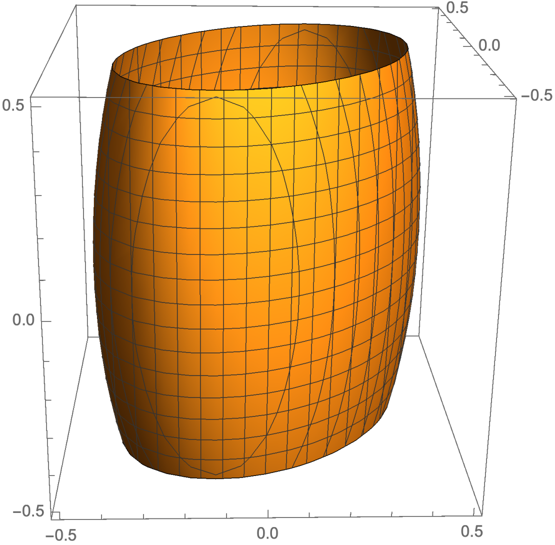
$$Q = \begin{bmatrix} 6 & -2 & 0 \\ -2 & 6 & 0 \\ 0 & 0 & 4 \end{bmatrix} , \quad H = \begin{bmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} , \quad \lambda = \begin{bmatrix} 8 \\ 4 \\ 4 \end{bmatrix}$$

$L(f , 1)$



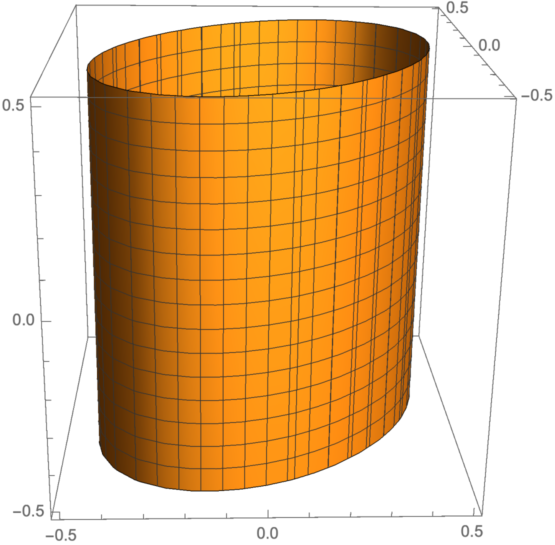
$$Q = \begin{bmatrix} 6 & -2 & 0 \\ -2 & 6 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad H = \begin{bmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \lambda = \begin{bmatrix} 8 \\ 4 \\ 1 \end{bmatrix}$$

$L(f, 1)$



$$Q = \begin{bmatrix} 6 & -2 & 0 \\ -2 & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad H = \begin{bmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \lambda = \begin{bmatrix} 8 \\ 4 \\ 0 \end{bmatrix}$$

$L(f, 1)$



- Clearly depends sign of eigenvalues of Q $\equiv$ definiteness:
  - $Q \succeq 0 \land Q \preceq 0 \equiv \lambda_1 = \lambda_n = 0 \equiv Q = 0 \implies \text{min} = \text{max} = 0$ (constant)
  - $Q \succeq 0 \implies \text{min} = 0, \text{argmin} = 0, \text{max} = +\infty$
  - $Q \preceq 0 \implies \text{max} = 0, \text{argmax} = 0, \text{min} = -\infty$
  - $Q \asymp 0 \implies \text{max} = +\infty, \text{min} = -\infty$

  analogous to univariate case, but "many more ways to be > 0 / < 0"

Exercise: Formally prove all the unboundedness results

- Clearly depends sign of eigenvalues of $Q \equiv$ definiteness:
  - $Q \succeq 0 \wedge Q \preceq 0 \equiv \lambda_1 = \lambda_n = 0 \equiv Q = 0 \Longrightarrow \min = \max = 0$ (constant)
  - $Q \succeq 0 \Longrightarrow \min = 0, \operatorname{argmin} = 0, \max = +\infty$
  - $Q \preceq 0 \Longrightarrow \max = 0, \operatorname{argmax} = 0, \min = -\infty$
  - $Q \asymp 0 \Longrightarrow \max = +\infty, \min = -\infty$

  analogous to univariate case, but "many more ways to be > 0 / < 0"

Exercise: Formally prove all the unboundedness results

- Box-constrained optimization on (closed) hyperrectangle X

- Clearly depends sign of eigenvalues of $Q \equiv$ definiteness:
  - $Q \succeq 0 \wedge Q \preceq 0 \equiv \lambda_1 = \lambda_n = 0 \equiv Q = 0 \implies \min = \max = 0$ (constant)
  - $Q \succeq 0 \implies \min = 0$, $\mathrm{argmin} = 0$, $\max = +\infty$
  - $Q \preceq 0 \implies \max = 0$, $\mathrm{argmax} = 0$, $\min = -\infty$
  - $Q \asymp 0 \implies \max = +\infty$, $\min = -\infty$

  analogous to univariate case, but "many more ways to be > 0 / < 0"

Exercise: Formally prove all the unboundedness results

- Box-constrained optimization on (closed) hyperrectangle X absolutely not analogous to the univariate case:
  - $\mathcal{NP}$-hard in most cases [3]
  - min with $Q \succeq 0$ and max with $Q \preceq 0$ polynomial but nontrivial (will see)
- $\mathcal{NP}$-hardness due to $\mathbb{R}^n$ "big" (X has 2^n vertices), issue also in $\mathcal{P}$ case
- $\max\{ f(x) \}$ and $\min\{ f(x) \}$ very very different

- $f(x) = \frac{1}{2}x^T Qx + \langle q, x \rangle$: a homogeneous quadratic plus a linear
- $q \neq 0$ but Q nonsingular $\equiv \lambda_i \neq 0 \, \forall i$ (regardless of the sign)
- Then $f(x) = g(z) = \frac{1}{2}z^T Qz + f(\bar{x})$ for $z = x - \bar{x}$ and $\bar{x} = -Q^{-1}q$

Exercise: Prove the result, but it should look familiar

▸ Optimizing a quadratic non-homogeneous function

- $f(x) = \frac{1}{2}x^T Qx + \langle q, x \rangle$: a homogeneous quadratic plus a linear
- $q \neq 0$ but Q nonsingular $\equiv \lambda_i \neq 0 \, \forall i$ (regardless of the sign)
- Then $f(x) = g(z) = \frac{1}{2}z^T Qz + f(\bar{x})$ for $z = x - \bar{x}$ and $\bar{x} = -Q^{-1}q$

Exercise: Prove the result, but it should look familiar

▸ Optimizing a quadratic non-homogeneous function

- $\bar{x} (\neq 0)$ centre of the level sets: repeat ▸ Optimizing a homogeneous quadratic multivariate function for $g(z)$, translate the results back in x-space
- Box-constrained case remains hard / nontrivial
- Analogous to univariate case, but many more ways for (pieces of) Q to be 0 and therefore the result not be applicable
- More complicated analysis needed, coming right next

- $Q \in \mathbb{R}^{n \times n}$, eigenvalue decomposition $(\, H \,,\, \Lambda \,)$, $I = \{\, 1 \,,\, 2 \,,\, \dots \,,\, n \,\}$
- $I^0 = \{\, i \in I \;;\; \lambda_i = 0 \,\}$, $I^+ = I \setminus I^0$, nonempty $(k = |\, I^0 \,| > 0,\; h = |\, I^+ \,| > 0)$
- $\ker(\, Q \,) = \{\, v \in \mathbb{R}^n \;:\; \exists \eta \in \mathbb{R}^k \text{ s.t. } v = \sum_{i \in I^0} \eta_i H_i \}$
- $Qv = 0 \;\forall v \in \ker(\, Q \,)\; [\supset \{\, 0 \,\}]$ **(check)**
- $\mathrm{im}(\, Q \,) = \{\, w \in \mathbb{R}^n \;:\; \exists \mu \in \mathbb{R}^h \text{ s.t. } w = \sum_{i \in I^+} \mu_i H_i \}$:
- $\forall\, w \in \mathrm{im}(\, Q \,)\; \exists x \in \mathbb{R}^n$ s.t. $Qx = w$, $\mathrm{im}(\, Q \,) = \mathrm{im}(\, -Q \,)$

Exercise: Prove the result (recall $Q = \lambda_1 H_1 H_1^T + \ldots + \lambda_n H_n H_n^T$, use [16])

- $\mathbb{R}^n = \mathrm{im}(\, Q \,) + ker(\, Q \,)$, $\mathrm{im}(\, Q \,) \perp ker(\, Q \,)$ (H is a hortonormal base of $\mathbb{R}^n$)
- $q = q^+ + q^0$, $q^+ \perp q^0$, with $q^0 \in ker(\, Q \,) \;\equiv\; Qq^0 = 0$, and
  $q^+ \in \mathrm{im}(\, Q \,) = \mathrm{im}(\, -Q \,) \;\equiv\; \exists\, \bar{x}$ s.t. $(-Q)\bar{x} = q^+$
- Then $f(\, x \,) = g(\, z \,) = \frac{1}{2} z^T Q z + q^0 z + f(\, \bar{x} \,)$ for $z = x - \bar{x}$

Exercise: Prove the result, but it should look very very familiar

▸ Optimizing non-homogeneous nonsingular quadratic functions

- f is "truly quadratic" on $\mathrm{im}(\,Q\,)$ but actually linear on $\ker(\,Q\,)$
- No surprise: $v \in \ker(\,Q\,) \implies f(\,v\,) = qv$
- Assume $Q \succeq 0$: f has minimum $\iff q^0 = 0 \;\equiv\; Q\bar{x} = -q$ has solution $\equiv\; q \in \mathrm{im}(\,Q\,)$
- $\bar{x}$ is not unique, in fact ∞-ly many of them: "all are centres"
- $\bar{x}$ solution $\implies \bar{x} + v$ solution $\forall\, v \in \ker(\,Q\,)$, all have the same objective value $\equiv$ they are all and only the minima of f

Exercise: Prove the result

Exercise: Discuss the cases $Q \preceq 0$ and $Q \asymp 0$

- $q_0 \neq 0 \;\equiv\; q \notin \mathrm{im}(\,Q\,) \implies \min = -\infty$, $\max = +\infty$
- Box-constrained version $\mathcal{P}$ (but nontrivial) if $Q \succeq 0$ / $Q \preceq 0$, hard otherwise

- f is "truly quadratic" on $\mathrm{im}(\,Q\,)$ but actually linear on $\ker(\,Q\,)$
- No surprise: $v \in \ker(\,Q\,) \implies f(\,v\,) = qv$
- Assume $Q \succeq 0$: f has minimum $\iff q^0 = 0 \;\equiv\; Q\bar{x} = -q$ has solution $\equiv\; q \in \mathrm{im}(\,Q\,)$
- $\bar{x}$ is not unique, in fact ∞-ly many of them: "all are centres"
- $\bar{x}$ solution $\implies \bar{x} + v$ solution $\forall\, v \in \ker(\,Q\,)$, all have the same objective value $\equiv$ they are all and only the minima of f

Exercise: Prove the result

Exercise: Discuss the cases $Q \preceq 0$ and $Q \asymp 0$

- $q_0 \neq 0 \;\equiv\; q \notin \mathrm{im}(\,Q\,) \implies \min = -\infty$, $\max = +\infty$
- Box-constrained version $\mathcal{P}$ (but nontrivial) if $Q \succeq 0$ / $Q \preceq 0$, hard otherwise
- All in all: solving system $Q\bar{x} = -q$ (or proving no solutions) required

Outline

- If one is lucky, optimising a quadratic function $\equiv$ solving $Q\bar{x} = -q$
- Linear system $O(\, n^3 \,)$ at worst, so doable for $n \approx 100$

- If one is lucky, optimising a quadratic function $\equiv$ solving $Q\bar{x} = -q$
- Linear system $O(\, n^3\,)$ at worst, so maybe doable for $n \approx 10000$

- If one is lucky, optimising a quadratic function $\equiv$ solving $Q\bar{x} = -q$
- Linear system $O(\, n^3\,)$ at worst, so not doable for $n \approx 10^{9+}$ (no memory)

- If one is lucky, optimising a quadratic function $\equiv$ solving $Q\bar{x} = -q$
- Linear system $O(\, n^3\,)$ at worst, so not doable for $n \approx 10^{9+}$ (no memory)
- Iterative procedures: start from initial guess x^0, some process $x^i \rightsquigarrow x^{i+1}$ $\Longrightarrow$ a sequence $\{\, x^i \,\}$ that should "go towards an optimal solution"
- The natural way: $\{\, f^i = f(\, x^i \,) \,\}$ sequence of values "go towards f_*"
- Typically we can't get f_* in finite time ($\exists\, i \;\; v_i = f_*$), but we can "get as close as we want": there in the limit
- Recall: (infinite) sequence $\{\, v_i \,\} = \{\, v_1 \,,\, v_2 \,,\, \ldots \}$,

$$\{\, v_i \,\} \to v \;\;\equiv\;\; \lim_{i\to\infty} v_i = v \;\;\equiv\;\; \forall \varepsilon > 0 \,\exists\, h \text{ s.t. } |\, v_i - v \,| \leq \varepsilon \;\forall i \geq h$$

$$\lim_{i\to\infty} v_i = +\infty \quad\Longleftrightarrow\quad \forall M > 0 \,\exists\, h \text{ s.t. } v_i \geq M \;\forall i \geq h$$

- If one is lucky, optimising a quadratic function $\equiv$ solving $Q\bar{x} = -q$
- Linear system $O(\, n^3 \,)$ at worst, so not doable for $n \approx 10^{9+}$ (no memory)
- Iterative procedures: start from initial guess x^0, some process $x^i \rightsquigarrow x^{i+1}$ $\Longrightarrow$ a sequence $\{\, x^i \,\}$ that should "go towards an optimal solution"
- The natural way: $\{\, f^i = f(\, x^i \,) \,\}$ sequence of values "go towards f_*"
- Typically we can't get f_* in finite time ($\exists\, i \;\; v_i = f_*$), but we can "get as close as we want": there in the limit
- Recall: (infinite) sequence $\{\, v_i \,\} = \{\, v_1 \,,\, v_2 \,,\, \ldots \}$,

  $\{\, v_i \,\} \to v \;\;\equiv\;\; \lim_{i \to \infty} v_i = v \;\;\equiv\;\; \forall \varepsilon > 0 \,\exists\, h \text{ s.t. } |\, v_i - v \,| \leq \varepsilon \;\forall i \geq h$

  $\lim_{i \to \infty} v_i = -\infty \quad \Longleftrightarrow \quad \forall M > 0 \,\exists\, h \text{ s.t. } v_i \leq -M \;\forall i \geq h$

- If one is lucky, optimising a quadratic function $\equiv$ solving $Q\bar{x} = -q$
- Linear system $O(\, n^3 \,)$ at worst, so not doable for $n \approx 10^{9+}$ (no memory)
- Iterative procedures: start from initial guess x^0, some process $x^i \rightsquigarrow x^{i+1}$ $\Longrightarrow$ a sequence $\{\, x^i \,\}$ that should "go towards an optimal solution"
- The natural way: $\{\, f^i = f(\, x^i \,) \,\}$ sequence of values "go towards f_*"
- Typically we can't get f_* in finite time ($\exists\, i \;\; v_i = f_*$), but we can "get as close as we want": there in the limit
- Recall: (infinite) sequence $\{\, v_i \,\} = \{\, v_1 \,,\, v_2 \,,\, \ldots \}$,

  $\{\, v_i \,\} \to v \;\; \equiv \;\; \lim_{i \to \infty} v_i = v \;\; \equiv \;\; \forall \varepsilon > 0 \,\exists\, h \text{ s.t. } |\, v_i - v \,| \leq \varepsilon \;\forall i \geq h$

  $\lim_{i \to \infty} v_i = -\infty \quad \Longleftrightarrow \quad \forall M > 0 \,\exists\, h \text{ s.t. } v_i \leq -M \;\forall i \geq h$
- $\{\, x^i \,\}$ s.t. $\{\, f^i \,\} \to f_*$ a minimizing sequence
- note that $\{\, f^i \,\} \to -\infty \implies f_* = -\infty \implies$ minimizing sequence
- A sequence may not have limit: are we "not converging"?
- Any monotone sequence has a limit (monotone algorithms are good)

- We generally assume minimization, but maximization is equivalent
- Given x^i, necessarily compute $g^i = Qx^i + q$: if $g^i = 0$ then stop
- "$g^i = 0$" not doable in floating point arithmetic $\Longrightarrow \| g^i \| \leq \varepsilon$ (which ε?)
- Idea: if $\| g^i \| > [\varepsilon >] 0$, produce a x^{i+1} "better" than x^i
- How? Consider the tomography $\varphi_{x^i,-g^i}(\alpha) = f(x^i - \alpha g^i) - f(x^i)$

$$= \tfrac{1}{2}(x^i - \alpha g^i)^T Q(x^i - \alpha g^i) + q(x^i - \alpha g^i) - f(x^i)$$

$$= \tfrac{1}{2}\alpha^2 (g^i)^T Q g^i - \alpha[(g^i)^T Q x^i + q g^i] = \underbrace{\tfrac{1}{2}\alpha^2 (g^i)^T Q g^i}_{\text{positive}} \underbrace{- \alpha \| g^i \|^2}_{\text{negative}}$$

- For some $\alpha > 0$, $\varphi_{x^i,-g^i}(\alpha) < 0 \Longrightarrow f(x^i - \alpha g^i) < f(x^i)$

Exercise: Check all the above (recall ▸ Optimizing a quadratic non-homogeneous function)

- The same information (called gradient, we'll see why) saying "you can't stop" is at the same time saying "you can get a better solution than x^i over there"
- This immediately suggests a (monotone, $f^{i+1} < f^i$) algorithm

- In fact it is easy to minimize $\varphi_{x^i, -g^i}(\alpha)$ (▸ Optimizing a quadratic non-homogeneous function)
  $\alpha^i = \| g^i \|^2 / ((g^i)^T Q g^i)$ $[1 / \lambda_1 \leq \alpha \leq 1 / \lambda_n$ (**check**)]
- Computing g^i and the optimal value of α is $O(n^2) \Longrightarrow$
  n "large" $\Longrightarrow$ "we can do may iterations before hitting $O(n^3)$"

```
procedure x = SDQ ( Q , q , x , ε )
  do forever
    g ← Qx + q;
    if( ‖ g ‖ ≤ ε ) then break;
    α ← stepsize(); x ← x − αg;
```

- stepsize() { **return**($\| g \|^2 / (g^T Q g)$); }, others possible

Exercise: something can go wrong with that formula ↑: what does it mean? Improve the pseudo-code to take that occurrence into account.

Exercise: what happens if $Q \not\succeq 0$? Does the (improved) code need be fixed?

Exercise: Discuss how to change the code to solve $\max\{ f(x) \}$ instead

Exercise: Rewrite the code with one product with Q per iteration

- It is very simple, but does it work? And is it efficient?

- Optimal stepsize $\implies g^{i+1} \perp g^i$ (**check**)
- "Homogeneous form of the error": $A(x) = \frac{1}{2}(x - x_*)^T Q(x - x_*)$ (**check**)
- The above for $x = x^{i+1}$, $Q \succ 0$ and some algebra [5, Lm. 8.6.1] gives
$$A(x^{i+1}) = \left(1 - \frac{\| g^i \|^4}{((g^i)^T Q g^i)((g^i)^T Q^{-1} g^i)}\right) A(x^i) \quad (\textbf{check})[\text{tedious}]$$
- Easy to derive an estimate using $\kappa = \lambda_1 / \lambda_n$ [≥ 1] condition number of Q
$$\frac{\|x\|^4}{(x^T Q x)(x^T Q^{-1} x)} \geq \frac{\lambda_n}{\lambda_1} = \frac{1}{\kappa} \quad (\textbf{check}) \implies A(x^{i+1}) \leq \left(1 - \frac{1}{\kappa}\right) A(x^i)$$
- This means the algorithm converges: $A(x^i) \leq r^i A(x^0)$ (**check**) with $r \leq (\kappa - 1)/\kappa < 1 \implies A(x^i) \to 0$ exponentially fast as $i \to \infty$
- Kantorovich inequality [5, 8.6.(34)] gives better estimate
$$\frac{\|x\|^4}{(x^T Q x)(x^T Q^{-1} x)} \geq \frac{4\lambda_1 \lambda_n}{(\lambda_1 + \lambda_n)^2} \implies r \leq \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^2 = \left(\frac{\kappa - 1}{\kappa + 1}\right)^2$$
- Let's see it in practice

- Crucial sequences: $\{\, x^i \,\}$ / $\{\, d^i = \| \, x^i - x_* \, \| \,\}$ [iterates / distance from x_*]
  $\{\, f^i = f(\, x^i \,) \,\}$ / $\{\, a^i = A(\, x^i \,) \,\}$ / $\{\, r^i = R(\, x^i \,) \,\}$ [f-values / A/R gaps]

- Complexity as a function of prescribed accuracy ε:
  max number of iterations k such that d^i / a^i / $r^i \leq \varepsilon \; \forall i \geq k$

- General formula: $v^k \leq r^k v^1 \leq \varepsilon$ for $k \geq [\, 1 \,/\, \log(\, 1 \,/\, r \,) \,] \log(\, v^1 \,/\, \varepsilon \,)$ (**check**)

- $r \approx 1 \implies k \in O(\, [\, r \,/\, (\, 1 - r \,) \,] \log(\, v^1 \,/\, \varepsilon \,) \,)$ (**check**)

- Good news: dimension independent (n not there) $\implies$ very-large-scale

- $O(\, \log(\, 1 \,/\, \varepsilon \,) \,)$ (good), but the constant $\uparrow \infty$ as $r \to 1$ (bad)

- $v^1 = f(\, x^1 \,) - f_*$: starting closer to f_* helps (would be strange if not)

- "$\| \, x^i - x_* \, \| \leq \varepsilon$" and "$f(\, x^i \,) - f_* \leq \varepsilon$" not the same ($\varepsilon$):
  $a^i = \frac{1}{2}(x^i - x_*)^T Q (x^i - x_*) \leq \varepsilon \implies \lambda_n \| \, x^i - x_* \, \|^2 \leq \varepsilon \implies$
  $d^i = \| \, x^i - x_* \, \| \leq \sqrt{\varepsilon \,/\, \lambda_n}$

Exercise: Cook up the other direction ($d^i \leq \varepsilon \implies \ldots$)

- Converge: $\{\, f^i \,\} \to f_* \approx\equiv \{\, a^i \,\} \to 0 \equiv \{\, r^i \,\} \to 0 \Longleftarrow \{\, d^i \,\} \to 0$ ($\not\Longrightarrow$)

Exercise: Discuss why $\{\, f^i \,\} \to f_*$ is only $\approx\equiv$ to $\{\, a^i \,\} \to 0$ and why the $\not\Longrightarrow$

- But how rapidly does it ("in the tail")? Rate/order of convergence

$$\lim_{i\to\infty} \left[\frac{f^{i+1} - f_*}{(\, f^i - f_* \,)^p} = \frac{a^{i+1}}{(\, a^i \,)^p} \approx \frac{r^{i+1}}{(\, r^i \,)^p} \right] = r \qquad \left[\begin{array}{l} x^p \to 0 \text{ faster than} \\ x \to 0 \text{ when } p > 1 \end{array} \right. \qquad \textbf{(check)}$$

- $p = 1$, $r = 1$ $\equiv$ sublinear: important examples

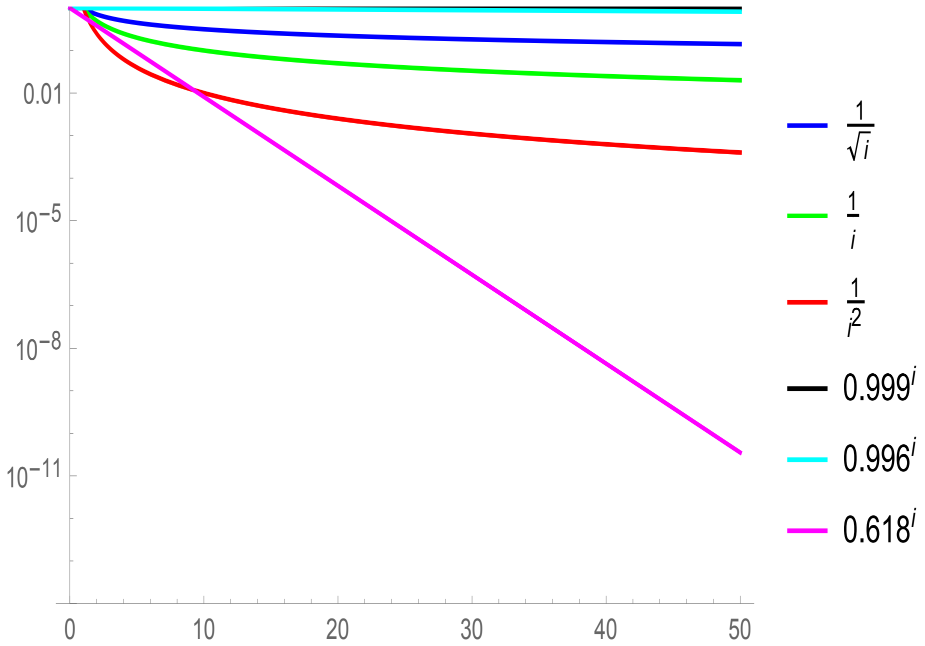
error	$O(\, 1 \,/\, i \,)$	$O(\, 1 \,/\, i^2 \,)$	$O(\, 1 \,/\, \sqrt{i} \,)$
i	$O(\, 1 \,/\, \varepsilon \,)$ (bad)	$O(\, 1 \,/\, \sqrt{\varepsilon} \,)$ (a bit better)	$O(\, 1 \,/\, \varepsilon^2 \,)$ (horrible)

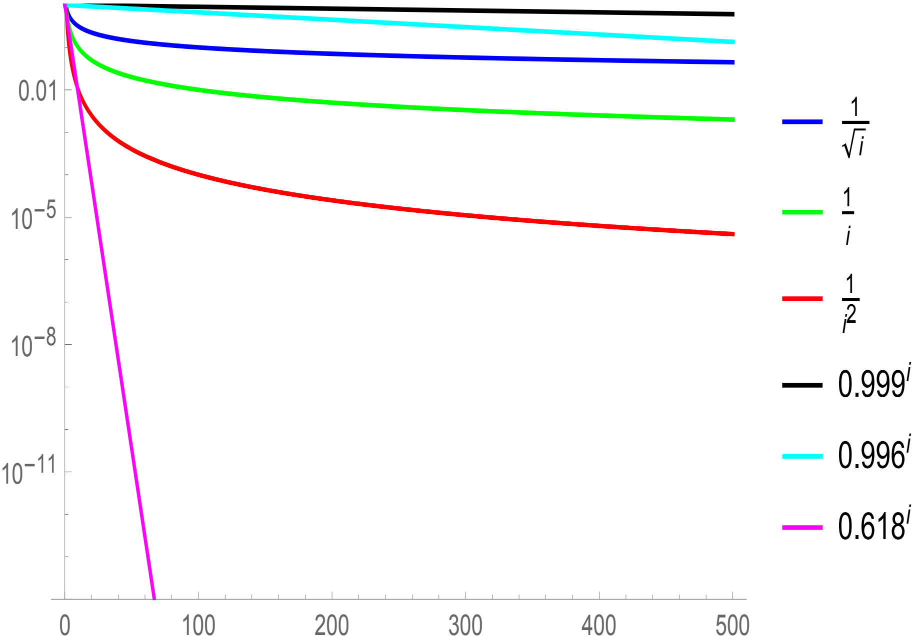
- $p = 1$, $r < 1$ $\equiv$ linear: $r^i \Longrightarrow i \in O(\, \log(\, 1/\varepsilon \,)\,)$ (good unless $r \approx 1$)

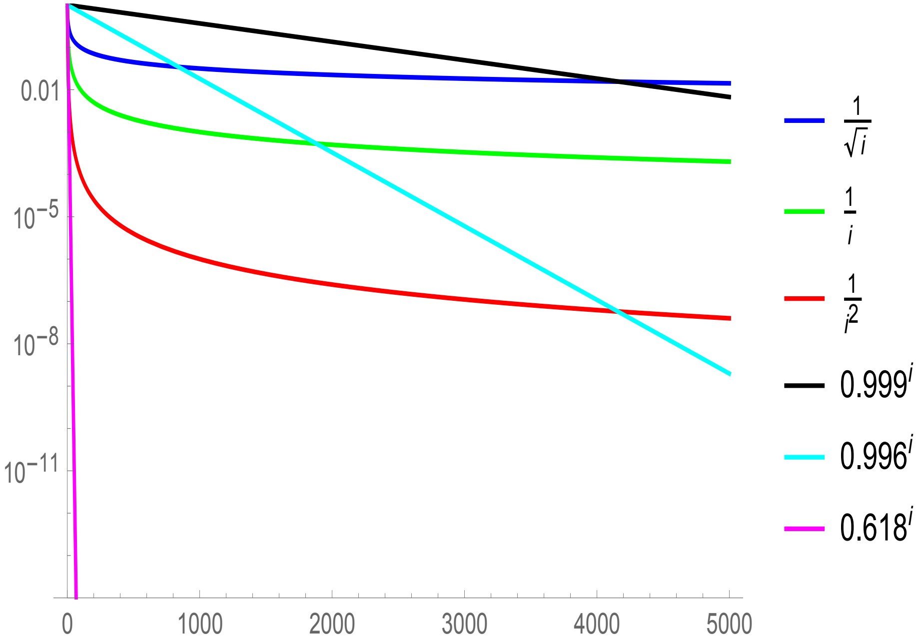
- $p = 2$, $r > 0$ $\equiv$ quadratic (!!!): $\approx 1 \,/\, 2^{2^i} \Longrightarrow i \in O(\, \log(\, \log(\, 1/\varepsilon \,)\,)\,)$
  in practice $O(\, 1 \,)$ (correct digits double at each iteration)

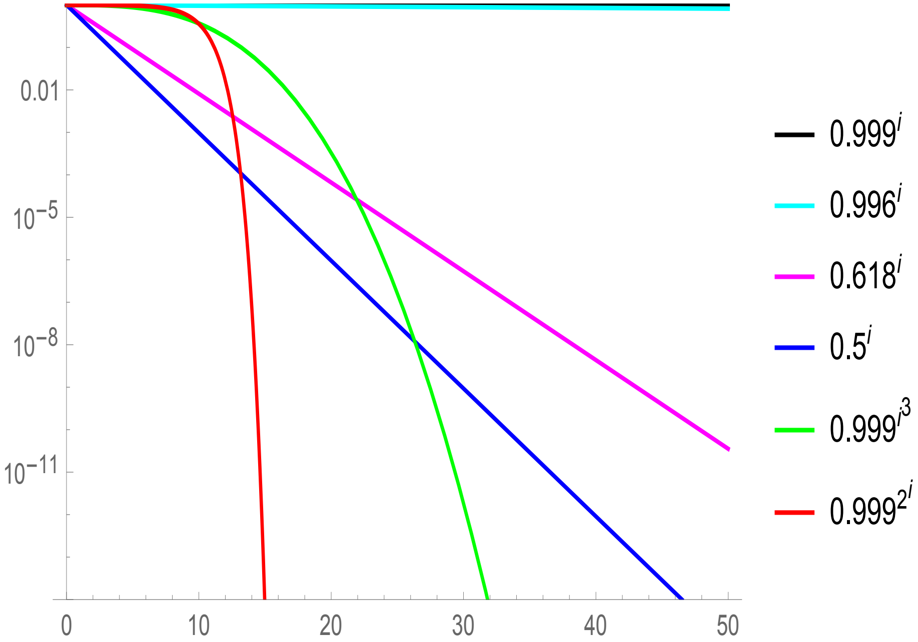
- $p \in (\, 1 \,,\, 2 \,)$ $\equiv$ $p = 1$, $r = 0$ $\equiv$ superlinear (!): "something in the middle"

- $p = 2$ the best you can reasonably hope for: possible but not easy


0.01
10^{-5}
10^{-8}
10^{-11}
0
10
20
30
40
50
$\frac{1}{\sqrt{i}}$
$\frac{1}{i}$
$\frac{1}{i^2}$
0.999^i
0.996^i
0.618^i


0.01
10^{-5}
10^{-8}
10^{-11}
0
100
200
300
400
500
$\frac{1}{\sqrt{i}}$
$\frac{1}{i}$
$\frac{1}{i^2}$
0.999^i
0.996^i
0.618^i


0.01
10−5
10−8
10−11
0
1000
2000
3000
4000
5000
1/√i
1/i
1/i²
0.999^i
0.996^i
0.618^i


0.01
10^{-5}
10^{-8}
10^{-11}
0
10
20
30
40
50
0.999^i
0.996^i
0.618^i
0.5^i
0.999^{i^3}
0.999^{2^i}

- The stopping criterion one would want: $A(\,x^i\,) \leq \varepsilon \;/\; R(\,x^i\,) \leq \varepsilon$
- Issue: f_* typically unknown, cannot be used on-line
- $\|\,g^i\,\|$ "proxy" of $A(\,x^i\,)$: hopefully $\|\,g^i\,\|$ "small" $\;\Longrightarrow\;$ $A(\,x^i\,)$ "small"
  but exact relationship nontrivial $\Longrightarrow$ choosing ε non obvious
- $\|\,g^i\,\| = Q(\,x^i - x_*\,) \Longrightarrow \|\,g^i\,\| \leq \lambda_1 \|\,x^i - x_*\,\| \ldots (??)$ wrong inequality:
  $\|\,g^i\,\| \leq \varepsilon \;\not\Longrightarrow\; \|\,x^i - x_*\,\|$ "small"
- $a^i = \frac{1}{2}(x^i - x_*)^T Q(x^i - x_*) = \frac{1}{2}\langle\, x^i - x_*\,,\, g^i\,\rangle \leq \frac{1}{2}\|\,g^i\,\|\|\,x^i - x_*\,\|$;
  if we knew $\delta \geq \|\,x^i - x_*\,\|$, which we don't, then $\|\,g^i\,\| \leq 2\varepsilon\,/\,\delta \;\Longrightarrow\; a^i \leq \varepsilon$
- If we knew $\lambda_n > 0$, which we don't, $\|\,g^i\,\| \leq \sqrt{2\lambda_n\varepsilon} \;\Longrightarrow\; a^i \leq \varepsilon$ (**check**)
- All in all, exact control on final $a^i \;/\; r^i$ not obvious (not always really needed)

- Convergence fast if $\lambda_1 \approx \lambda_n$ (one iteration for $\| x \|^2$), rather slow if $\lambda_1 \gg \lambda_n$:
  $\kappa = \lambda_1 / \lambda_n \to \infty$ (Q ill conditioned) $\Longrightarrow$ $r \to 1$ $\Longrightarrow$ slow in practice

- $g^{i+1} \perp g^i$ + level sets very elongated $\Longrightarrow$ lots of "zig-zags" $\Longrightarrow$ slow

- Ex.: $\kappa = 1000 \Longrightarrow r \approx 0.996 \Longrightarrow r / (1 - r) \approx 250$
  $f(x^1) - f_* = 1$, $\varepsilon = 10^{-6} \Longrightarrow k \geq 3450$ for $n = 2$

- Convergence fast if $\lambda_1 \approx \lambda_n$ (one iteration for $\| x \|^2$), rather slow if $\lambda_1 \gg \lambda_n$:
  $\kappa = \lambda_1 / \lambda_n \to \infty$ (Q ill conditioned) $\Longrightarrow$ $r \to 1$ $\Longrightarrow$ slow in practice

- $g^{i+1} \perp g^i$ + level sets very elongated $\Longrightarrow$ lots of "zig-zags" $\Longrightarrow$ slow

- Ex.: $\kappa = 1000 \Longrightarrow r \approx 0.996 \Longrightarrow r / (1 - r) \approx 250$
  $f(x^1) - f_* = 1$, $\varepsilon = 10^{-6} \Longrightarrow k \geq 3450$ for $n = 2$ ... but also for $n = 10^8$

- Note: with coarser formula $r = 0.999 \equiv r / (1 - r) \approx 1000 \Longrightarrow k \geq 13800$

- In other words: $0.996^{10} \approx 0.96071$ $\qquad$ $0.999^{10} \approx 0.99004$

- Convergence fast if $\lambda_1 \approx \lambda_n$ (one iteration for $\| x \|^2$), rather slow if $\lambda_1 \gg \lambda_n$: $\kappa = \lambda_1 / \lambda_n \to \infty$ (Q ill conditioned) $\implies r \to 1 \implies$ slow in practice

- $g^{i+1} \perp g^i$ + level sets very elongated $\implies$ lots of "zig-zags" $\implies$ slow

- Ex.: $\kappa = 1000 \implies r \approx 0.996 \implies r / (1 - r) \approx 250$
  $f(x^1) - f_* = 1$, $\varepsilon = 10^{-6} \implies k \geq 3450$ for $n = 2$ ... but also for $n = 10^8$

- Note: with coarser formula $r = 0.999 \equiv r / (1 - r) \approx 1000 \implies k \geq 13800$

- In other words: $0.996^{100} \approx 0.66978$ $\qquad$ $0.999^{100} \approx 0.90479$

- Convergence fast if $\lambda_1 \approx \lambda_n$ (one iteration for $\| x \|^2$), rather slow if $\lambda_1 \gg \lambda_n$:
  $\kappa = \lambda_1 / \lambda_n \to \infty$ (Q ill conditioned) $\Longrightarrow$ $r \to 1$ $\Longrightarrow$ slow in practice
- $g^{i+1} \perp g^i$ + level sets very elongated $\Longrightarrow$ lots of "zig-zags" $\Longrightarrow$ slow
- Ex.: $\kappa = 1000 \Longrightarrow r \approx 0.996 \Longrightarrow r / (1 - r) \approx 250$
  $f(x^1) - f_* = 1$, $\varepsilon = 10^{-6} \Longrightarrow k \geq 3450$ for $n = 2$ ... but also for $n = 10^8$
- Note: with coarser formula $r = 0.999 \equiv r / (1 - r) \approx 1000 \Longrightarrow k \geq 13800$
- In other words: $0.996^{1000} \approx 0.01816$ $\qquad 0.999^{1000} \approx 0.36769$

- Convergence fast if $\lambda_1 \approx \lambda_n$ (one iteration for $\| x \|^2$), rather slow if $\lambda_1 \gg \lambda_n$:
  $\kappa = \lambda_1 / \lambda_n \to \infty$ (Q ill conditioned) $\Longrightarrow$ $r \to 1$ $\Longrightarrow$ slow in practice
- $g^{i+1} \perp g^i$ + level sets very elongated $\Longrightarrow$ lots of "zig-zags" $\Longrightarrow$ slow
- Ex.: $\kappa = 1000 \Longrightarrow r \approx 0.996 \Longrightarrow r / (1 - r) \approx 250$
  $f(x^1) - f_* = 1$, $\varepsilon = 10^{-6} \Longrightarrow k \geq 3450$ for $n = 2$ ... but also for $n = 10^8$
- Note: with coarser formula $r = 0.999 \equiv r / (1 - r) \approx 1000 \Longrightarrow k \geq 13800$
- In other words: $0.996^{2000} \approx 0.00033$ $\quad$ $0.999^{2000} \approx 0.13520$

- Convergence fast if $\lambda_1 \approx \lambda_n$ (one iteration for $\|x\|^2$), rather slow if $\lambda_1 \gg \lambda_n$: $\kappa = \lambda_1 / \lambda_n \to \infty$ (Q ill conditioned) $\implies r \to 1 \implies$ slow in practice
- $g^{i+1} \perp g^i$ + level sets very elongated $\implies$ lots of "zig-zags" $\implies$ slow
- Ex.: $\kappa = 1000 \implies r \approx 0.996 \implies r / (1 - r) \approx 250$
  $f(x^1) - f_* = 1$, $\varepsilon = 10^{-6} \implies k \geq 3450$ for $n = 2$ ... but also for $n = 10^8$
- Note: with coarser formula $r = 0.999 \equiv r / (1 - r) \approx 1000 \implies k \geq 13800$
- In other words: $0.996^{2000} \approx 0.00033$ $\qquad$ $0.999^{2000} \approx 0.13520$
- More bad news, "hidden dependency":
  λ_1 and λ_n may depend on n, κ may grow as $n \to \infty$
- More bad news: the behaviour in practice is close to the bound
- Even more bad news: $\lambda_n = 0 \equiv \kappa = \infty$ happens

What if $\lambda_n = 0$?

- $\lambda_n = 0 \implies$ not converging?

What if $\lambda_n = 0$?

- $\lambda_n = 0 \implies$ not converging? No, just can't prove it this way
- In fact we can prove convergence (in a more general setting) [2, Theorem 3.3]: $\alpha = 1 / \lambda_1 \implies f(x^i) - f_* \leq 2\lambda_1 \| x^1 - x_* \|^2 / (i - 1)$

- $\lambda_n = 0 \implies$ not converging? No, just can't prove it this way
- In fact we can prove convergence (in a more general setting) [2, Theorem 3.3]:
  $\alpha = 1 / \lambda_1 \implies f(x^i) - f_* \leq 2\lambda_1 \| x^1 - x_* \|^2 / (i - 1)$
- Is it good news? Only partly. Because complexity is $k \geq 2\lambda_1 d^1 / \varepsilon$
- $O(1/\varepsilon)$ vs. $O(\log(1/\varepsilon))$: sublinear convergence, exponentially slower
- One further digit of accuracy $\equiv$ 10 times more iterations $\implies$ typically unfeasible to get more than 1e−3 / 1e−4 accuracy
- The result cannot be improved (in general, will see)
- Is it bad? Rather. Can it be worse? Yes (in general, will see)

- $\lambda_n = 0 \implies$ not converging? No, just can't prove it this way
- In fact we can prove convergence (in a more general setting) [2, Theorem 3.3]:
  $\alpha = 1 / \lambda_1 \implies f(x^i) - f_* \leq 2\lambda_1 \| x^1 - x_* \|^2 / (i - 1)$
- Is it good news? Only partly. Because complexity is $k \geq 2\lambda_1 d^1 / \varepsilon$
- $O(1/\varepsilon)$ vs. $O(\log(1/\varepsilon))$: sublinear convergence, exponentially slower
- One further digit of accuracy $\equiv$ 10 times more iterations $\implies$ typically unfeasible to get more than 1e−3 / 1e−4 accuracy
- The result cannot be improved (in general, will see)
- Is it bad? Rather. Can it be worse? Yes (in general, will see)
- If $\lambda_n > 0$, can we do better than $O(\log(1/\varepsilon))$? Yes – @Federico
- Fundamental idea, will see more than once: changing the space

Outline

- Optimization problems are difficult
- Clever strategy: start simple, then use what you learnt to go more complex

Wrap up

- Optimization problems are difficult
- Clever strategy: start simple, then use what you learnt to go more complex
- Simple problems provide intuition for the solution of more complex ones
- Solving a complex problem may entail solving a sequence of simpler ones
- Usual concept: if XYZ complex, use “simpler” ABC $\approx$ XYZ – a model

- Optimization problems are difficult
- Clever strategy: start simple, then use what you learnt to go more complex
- Simple problems provide intuition for the solution of more complex ones
- Solving a complex problem may entail solving a sequence of simpler ones
- Usual concept: if XYZ complex, use "simpler" ABC $\approx$ XYZ – a model
- Linear functions "too simple": optimising (on simple constraints) always easy
- Quadratic functions already a different story: few really simple cases, often polynomial but not with low exponent, up to $\mathcal{NP}$-hard
- Solving (simple) optimization problems requires linear algebra, and vice-versa

- Optimization problems are difficult
- Clever strategy: start simple, then use what you learnt to go more complex
- Simple problems provide intuition for the solution of more complex ones
- Solving a complex problem may entail solving a sequence of simpler ones
- Usual concept: if XYZ complex, use "simpler" ABC $\approx$ XYZ – a model
- Linear functions "too simple": optimising (on simple constraints) always easy
- Quadratic functions already a different story: few really simple cases, often polynomial but not with low exponent, up to $\mathcal{NP}$-hard
- Solving (simple) optimization problems requires linear algebra, and vice-versa
- We now know all we need about simple problems, time to step up the game
- Will keep following an incremental approach: next step is more complicated functions but only one variable

[1] S. Boyd, L. Vandenberghe *Convex Optimization*,
https://web.stanford.edu/~boyd/cvxbook
Cambridge University Press, 2008

[2] S. Bubeck *Convex Optimization: Algorithms and Complexity*,
arXiv:1405.4980v2, https://arxiv.org/abs/1405.4980, 2015

[3] E. de Klerk "The complexity of optimizing over a simplex, hypercube or sphere: a short survey" *Central European Journal of Operations Research* 16: 111–125, 2008 https://link.springer.com/content/pdf/10.1007/s10100-007-0052-9.pdf

[4] P. Hansen, B. Jaumard "Lipschitz Optimization" in *Handbook of Global Optimization – Nonconvex optimization and its applications*, R. Horst and P.M. Pardalos (Eds.), Chapter 8, 407–494, Springer, 1995

[5] D.G. Luenberger, Y. Ye *Linear and Nonlinear Programming*, Springer International Series in Operations Research & Management Science, 2008

[6] J. Nocedal, S.J. Wright, *Numerical Optimization – second edition*, Springer Series in Operations Research and Financial Engineering, 2006

[7] Wikipedia – Cubic equation
https://en.wikipedia.org/wiki/Cubic_equation

[8] Wikipedia – Determinant
https://en.wikipedia.org/wiki/Determinant

[9] Wikipedia – Laplace expansion
https://en.wikipedia.org/wiki/Laplace_expansion

[10] Wikipedia – Eigenvalue Algorithm
https://en.wikipedia.org/wiki/Eigenvalue_algorithm

[11] Wikipedia – Eigenvalues and Eigenvectors
https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors

[12] Wikipedia – Islands of Space
https://en.wikipedia.org/wiki/Islands_of_Space

[13] Wikipedia – Matrix Norm
https://en.wikipedia.org/wiki/Matrix_norm

[14] Wikipedia – Norm
https://en.wikipedia.org/wiki/Norm_(mathematics)

[15] Wikipedia – The Hitchhiker's Guide to the Galaxy
https://en.wikipedia.org/wiki/The_Hitchhiker's_Guide_to_the_Galaxy

[16] Wikipedia – Underdetermined System
https://en.wikipedia.org/wiki/Underdetermined_system

Outline

- Use $\max\{\, | f_* | \,,\, 1 \,\}$ instead; this corresponds to $\min\{\, f(\,x\,) + 1 \,\}$ [**back**]

- $b > 0$ and $x - z > 0 \implies b(x - z) > 0 \equiv bx > bz$;
  the others are analogous (or simpler) [**back**]

- If $x_+ = +\infty$, obviously $x_* = +\infty = x_+$
  If $x_+ < +\infty$, since $f(\,x\,)$ is increasing, $f(\,x\,) < f(\,x_+\,)\ \forall x < x_+$
  The treatment of x_- is analogous.
  If $b < 0$, the role of x_+ and x_- reverses ($x_+ = \text{argmin}$, $x_- = \text{argmax}$)
  If $b = 0$, every point in X is an optimal solution [**back**]

- $x > z$, $a > 0$ and $x > 0 \implies ax^2 > axz > az^2$. Since $f(\,x\,)$ is symmetric ($ax^2 = a(-x)^2$), increasing for $x > 0 \equiv$ deceasing for $x < 0$. When $a < 0$ the sign of the inequalities in inverted (the function is reflected upon the x axis). The case $a = 0$ is trivial [**back**]

- $f(x)$ has a minimum in 0, is decreasing for $x < 0$ and increasing for $x > 0$. If $x_- > 0$ then $f(x)$ is increasing along all X, hence x_- is the minimum and x_+ the maximum. The argument is symmetric if $x_+ < 0$. Obviously, if $0 \in X$ then it is the minimum; for the maximization, since the function is increasing when x moves away from 0 in both directions, the maximum has to be one of the two extremes but we don't know which until we test. The rest is too trivial [**back**]

- No, this is both too trivial and didactic [**back**]

- $f(x) = (ax + b)x$, hence the roots are $x = 0$ and $x = x_p = -b / a$. Clearly, $\bar{x} = -b / 2a$ is always in the middle of the interval defined by the roots. If a and b have the same sign then $x_p < \bar{x} < 0$, otherwise $x_p > \bar{x} > 0$ [**back**]

- $\varphi_{x,(\beta d)}(\alpha) = f(x + \alpha(\beta d)) = f(x + (\alpha\beta)d) = \varphi_{x,d}(\alpha\beta)$ [**back**]

- We assume that i. and ii. hold for f and we want to show that $f(x) = \langle b, x \rangle$ for some $b \in \mathbb{R}^n$. Let u_i, $i = 1, \dots, n$, the i-th vector of the canonical base of $\mathbb{R}^n$ (having 1 in the i-th position and 0 otherwise), and $b_i = f(u_i)$. For any $x \in \mathbb{R}^n$, $x = \sum_{i=1}^n x_i u_i$, hence $f(x) = f(\sum_{i=1}^n x_i u_i) = \sum_{i=1}^n f(x_i u_i)$ (using ii. recursively n times) $= \sum_{i=1}^n x_i f(u_i)$ (using i. on each individual term) $= \sum_{i=1}^n b_i x_i$ (using the definition of b_i) $= \langle b, x \rangle$ (using the definition of scalar product). The results clearly breaks in the affine case ($c \neq 0$): $f(x) = x + 1 \implies f(2x) = 2x + 1 \neq 2(x + 1) = 2f(x)$ **[back]**

- By contradiction, $\exists\, \gamma \in \mathbb{R}^n \setminus \{0\}$ s.t. $H\gamma = 0 \implies$ $0 = \| H\gamma \|^2 = \gamma^T [H^T H] \gamma = \| \gamma \|^2 > 0$ $[\gamma \neq 0]$ ↯ **[back]**

- This is based on a general result: for $[\, A^1\,,\, A^2\,,\, \ldots\,,\, A^n\,] = A \in \mathbb{R}^{m \times n}$ (not necessarily square) written by columns, $AA^T = M \in \mathbb{R}^{m \times m}$ (symmetric, prove it using $[\, AB\,]^T = B^T A^T$) can be written as the sum of the n rank-one matrices corresponding to the columns, i.e., $M = \sum_{i=1}^{n} [\, D^i = A^i (A^i)^T\,]$. In fact, the h-th row of A is $A_h = [\, A_h^1\,,\, A_h^2\,,\, \ldots\,,\, A_h^n\,]$ and the k-th column of A^T is the k-th row of A, thus $M_{hk} = \langle\, A_h\,,\, A_k\,\rangle = \sum_{i=1}^{n} A_h^i A_k^i$. But $D_{hk}^i = A_h^i A_h^i$, hence $M_{hk} = \sum_{i=1}^{n} D_{hk}^i$ for all h and k
  To complete the result, for $\Lambda = \mathrm{diag}(\, [\, \lambda_1\,,\, \lambda_2\,,\, \ldots\,,\, \lambda_n\,]\,) \in \mathbb{R}^{n \times n}$, $L = A\Lambda =$ $= [\, \lambda_1 A^1\,,\, \lambda_2 A^2\,,\, \ldots\,,\, \lambda_n A^n\,]$. In fact, the h-th row $A_h = [\, A_h^1\,,\, A_h^2\,,\, \ldots\,,\, A_h^n\,]$ and the k-th column of Λ, i.e., $\lambda_k u_k$ (u_k being the k-th vector of the canonical base) give $L_{hk} = \langle\, A^h\,,\, \lambda_k u_k\,\rangle = \lambda_k A_h^k$ **[back]**

- $\varphi_{H_i}(\,\alpha\,) = (\alpha H_i)^T Q(\alpha H_i) = \alpha^2 [\, H_i^T(\, \lambda_i H_i\,)\,] = \lambda_i \alpha^2$ **[back]**

- $\lambda_n < 0 \implies \varphi_{H_n}(\,\alpha\,)\,[= \lambda_n \alpha^2]$ unbounded below $\implies f(\,x\,)$ unbounded below
  $\lambda_1 > 0 \implies \varphi_{H_1}(\,\alpha\,)$ unbounded above $\implies f(\,x\,)$ unbounded above **[back]**

- $x = z + \bar{x} \implies \frac{1}{2}x^T Qx + qx = \frac{1}{2}(z + \bar{x})^T Q(z + \bar{x}) + q(z + \bar{x}) =$
  $\frac{1}{2}z^T Qz + z^T(Q\bar{x} + q) + [\frac{1}{2}\bar{x}^T Q\bar{x} + q\bar{x}] = \frac{1}{2}z^T Qz + f(\bar{x})$
  as $Q\bar{x} + q = Q(-Q^{-1}q) + q = -q + q = 0$ **[back]**

- $Qv = Q[\sum_{i \in Z} \eta_i H_i] = \sum_{i \in Z} \eta_i QH_i = \sum_{i \in Z} \eta_i \lambda_i H_i = 0$ **[back]**

- $Q = H\Lambda H^T = \sum_{i=1}^{n} \lambda_i H_i H_i^T = \sum_{i \in Z} \lambda_i H_i H_i^T [= 0] + \sum_{i \in N} \lambda_i H_i H_i^T$
  We want to prove $\exists\, x$ s.t. $(\sum_{i \in N} \lambda_i H_i H_i^T)x = \sum_{i \in N} \mu_i H_i = w$
  True if $\lambda_i H_i^T x = \mu_i \;\; i \in N \;\; \equiv \;\; H_i^T x = \gamma_i = \mu_i / \lambda_i \;\; i \in N$,
  a linear system of $k \leq n$ equations in n variables (likely underdetermined)
  All H_i linearly independent, $H_N = [H_i]_{i \in N} \in \mathbb{R}^{n \times k} \implies rank(H_N) = k$
  $\implies [H_N^T, \gamma] \in \mathbb{R}^{k \times n+1}$ has rank k (rank $\leq$ number of rows) $\implies$
  by [16] the system has a solution x (∞-ly many if $k < n$) **[back]**

- $\frac{1}{2}x^T Qx + qx = \frac{1}{2}(z + \bar{x})^T Q(z + \bar{x}) + q(z + \bar{x}) =$
  $\frac{1}{2}z^T Qz + z^T(Q\bar{x} + q^+ + q^0) + f(\bar{x}) = \frac{1}{2}z^T Qz + q^0 z + f(\bar{x})$ **[back]**

- We know that $f(\,z\,) = z^T Qz + f(\,barx\,)$, with $z = x - \bar{x}$. For $x \in \bar{x} + v$, with $v \in \ker(\,Q\,)$, $z = x - \bar{x} = \bar{x} + v - \bar{x} = v$. Hence $f(\,z\,) = f(\,\bar{x}\,)$. On the other hand, $f(\,z\,) \geq f(\,\bar{x}\,)$ for all z since $Q \succeq 0$, thus any such point is a minimum. Any point $x \in \bar{x} + v$ with $v \notin \ker(\,Q\,)$ has $f(\,x\,) = v^T Qv + f(\,\bar{x}\,) > f(\,\bar{x}\,)$ since $v^T Qv > 0$ [**back**]

- No, this is both too trivial and didactic [**back**]

- $\varphi(\,\alpha\,) = a\alpha^2 + b\alpha$ quadratic non-homogeneous with $a = (g^i)^T Qg^i \geq 0$ and $b = -\|\,g^i\,\|^2 < 0$. If $a > 0$, then $\varphi(\,\bar{\alpha}\,) < \varphi(\,0\,) = f(\,x^i\,)\;\forall\,\bar{\alpha} \in (\,0\,,\,-b/a\,)$; in particular, $\bar{\alpha} = \|\,g^i\,\|^2 \,/\, (2(g^i)^T Qg^i)$ is the minimum of φ. If $a = 0$ then φ is decreasing and $\varphi(\,\bar{\alpha}\,) < \varphi(\,0\,) = f(\,x^i\,)\;\forall\bar{\alpha} > 0$ [**back**]

- The variational characterization of the eigenvalues implies that $\lambda_1 \geq d^T Qd \,/\, \|\,d\,\|^2 \geq \lambda_n$ for all $d \neq 0$; this immediately gives $1 \,/\, \lambda_1 \leq \|\,d\,\|^2 \,/\, d^T Qd \leq 1 \,/\, \lambda_n$ for all d, and therefore in particular $d = g^i$ (knowing that $g^i \neq 0$ otherwise the algorithm would have stopped) [**back**]

- The issue clearly is $g^T Qg = 0$ (very small), which means that $\varphi_{x,-g}$ is (almost) linear, and therefore f is unbounded below. One should therefore add a line
  if($g^T Qg \leq \delta$) **then break**;
  for a "very small" δ, but also add a proper way for the algorithm to signal that the returned x is not optimal, e.g., by also returning a "status code" [**back**]

- Having added the extra check above, the code just works: if $g^T Qg < 0$ then $(-)g$ is direction where φ has negative curvature, which still implies f is unbounded below. Note that this is not guaranteed to happen [**back**]

- Because $a < 0$, the step α will be negative, which basically means one is going in direction g rather than $-g$. The algorithm remains the same, except that the extra check above has to become $g^T Qg \geq -\delta$ [**back**]

- Assuming the gradient is computed in the "natural way" as $g = Q * x + q$ before the algorithm starts (i.e., with x the initial guess x^0), both quantities depending from matrix-vector products can be recovered by computing the vector $v = Q * g$. In fact, $a = g^T Qg = \langle g , v \rangle$. Then, with $x' = x - \alpha g$ one has $g' = Qx' + q = Q(x - \alpha g) + q = (Qx + q) - \alpha Qg = g - \alpha v$. Hence, the gradient at the next iteration can be computed in $O(n)$ out of that of the previous iteration and the vector v. As for the objective function, $1/2x^T Qx + \langle q , x \rangle = 1/2(x^T Qx + 2\langle q , x \rangle) = 1/2x^T(Qx + q + q) = 1/2\langle q + g , x \rangle$, i.e., it can be computed in $O(n)$ once g is known [**back**]

- $g^i = Q(x^i - x_*) = Qx^i + q$, $\alpha^i = \| g^i \|^2 / [(g^i)^T Qg^i]$
  $g^{i+1} = Qx^{i+1} + q = Q(x^i - \alpha^i g^i) + q = (I - \alpha^i Q)g^i \implies$
  $\langle g^{i+1} , g^i \rangle = \| g^i \|^2 - \alpha^i [(g^i)^T Qg^i] = 0$ [**back**]

- All arguments boil down to the crucial $Qx^* + q = 0$. This first of all gives that $f(\, x^* \,) = \frac{1}{2}(x^*)^T Qx^* + \langle\, x^* \,,\, q \,\rangle = (x^*)^T Qx^* + \langle\, x^* \,,\, q \,\rangle - \frac{1}{2}(x^*)^T Qx^* = (x^*)^T(Qx^* + q) - \frac{1}{2}(x^*)^T Qx^* = -\frac{1}{2}(x^*)^T Qx^*$. Then, $\frac{1}{2}(x - x^*)^T Q(x - x^*) = \frac{1}{2}x^T Qx + \frac{1}{2}(x^*)^T Qx^* - x^T(Qx^*) = \frac{1}{2}x^T Qx - \langle\, x \,,\, q \,\rangle + \frac{1}{2}(x^*)^T Qx^* = f(\, x \,) - f(\, x^* \,)$ (in the penultimate step we have used $Qx^* = -q$) **[back]**

- Just induction: obvious for $i = 0$, if it holds for $i - 1$ then $A(\, x^i \,) \leq rA(\, x^{i-1} \,) \leq r(\, r^{i-1} A(\, x^0 \,)\,)$ **[back]**

- Q nonsingular $\implies x^i - x_* = Q^{-1} g^i \implies$
  $a^i = \frac{1}{2}(x^i - x_*)^T Q(x^i - x_*) = \frac{1}{2}(g^i)^T Q^{-1} g^i \implies$
  $a^{i+1} = \frac{1}{2}(x^{i+1} - x_*)^T Q(x^{i+1} - x_*) = \frac{1}{2}(x^i - \alpha^i g^i - x_*)^T g^{i+1} = \frac{1}{2}(x^i - x_*)^T g^{i+1}$
  [using $\langle\, g^{i+1} \,,\, g^i \,\rangle = 0$] $= \frac{1}{2}(x^i - x_*)^T Q(x^i - \alpha^i g^i - x_*)$
  $= \frac{1}{2}(x^i - x_*)^T Q(x^i - x_*) - \frac{1}{2}\alpha^i(x^i - x_*)^T Qg^i = a^i - \frac{1}{2}\alpha^i \|\, g^i \,\|^2$
  [using $Q(x^i - x_*) = g^i$] $= a^i - \frac{1}{2}\|\, g^i \,\|^4 \,/\, (g^i)^T Qg^i$

$$= a^i - \frac{\| g^i \|^4 \qquad \frac{1}{2}(g^i)^T Q^{-1} g^i}{((g^i)^T Q g^i)((g^i)^T Q^{-1} g^i)} = a^i \left(1 - \frac{\| g^i \|^4}{((g^i)^T Q g^i)((g^i)^T Q^{-1} g^i)} \right)$$ **[back]**

- Recall $1 / \lambda_n \geq \ldots \geq 1 / \lambda_1 > 0$ eigenvalues of Q^{-1}; from the usual $\lambda_n \| x \|^2 \leq x^T Q x \leq \lambda_1 \| x \|^2$ (applied to Q^{-1} as well) one has $\| g \|^2 / g^T Q g \geq 1 / \lambda_1$ and $\| g \|^2 / g^T Q^{-1} g \geq 1 / [1 / \lambda_n]$ **[back]**

- $r^k v_1 \leq \varepsilon \equiv r^k \leq \varepsilon / v^1 \equiv \log(r^k) \leq \log(\varepsilon / v^1)$ (log monotone) $\equiv$ $k \log(r) \leq \log(\varepsilon / v^1)$ (property of log); since $r < 1$, $\log(r) < 0$, giving $k \geq \log(\varepsilon / v^1) / \log(r) = [- \log(\varepsilon / v^1)] / [- \log(r)] =$ $\log(v^1 / \varepsilon) / \log(1/r) = \log(v^1 / \varepsilon)[1 / \log(1 / r)]$ **[back]**

- This requires a bit of elementary calculus. The derivative of $\ln(x)$ is $1 / x$. The first-order Taylor approximation is $f(x + \delta) \approx f(x) + f'(x)\delta$ for $\delta \approx 0$. Applied to $\ln(\cdot)$ with $x = 1$ gives $\ln(1 + \delta) \approx \delta$, whence $1 / \ln(1 / r) =$ $= 1 / \ln(1 + (1 - r) / r) = r / (1 - r)$. But $\log_a(x) = \log_b(x) / \log_b(a)$, hence $\ln(x) = \log_e(x) = \log_{10}(x) / \log_{10}(e) \approx \log(x) / 0.43 \approx 2.3 \log(x)$, i.e., $\ln(x) \in O(\log(x))$ **[back]**

- $\lambda_1 \| x^i - x_* \|^2 \geq (x_i - x_*)^T Q(x_i - x_*) = 2a^i \equiv \| x^i - x_* \| \geq \sqrt{2a^i / \lambda_1}$, hence $d^i \leq \varepsilon \implies a^i \leq \lambda_1 \varepsilon^2 / 2$ **[back]**

- $a^i = \frac{1}{2}(x^i - x_*)^T Q(x^i - x_*) = \frac{1}{2}\langle g^i , x^i - x_* \rangle \leq \frac{1}{2}\| g^i \| \| x^i - x_* \|$. On the other hand, $\| g^i \|^2 = (x^i - x_*)^T Q^T Q(x^i - x_*) \geq \lambda_n^2 \| x^i - x_* \|^2$ (recall λ_n^2 eigenvalue of Q^2, clearly the smallest), i.e., $\| g^i \| \geq \lambda_n \| x^i - x_* \|$. Hence, $\| g^i \| \leq \sqrt{2\lambda_n \varepsilon} \implies \varepsilon \geq \frac{1}{2\lambda_n}\| g^i \|^2 \geq \frac{1}{2}\| g^i \| \| x^i - x_* \| \geq a^i$ **[back]**

- If $f_* = -\infty$, $f_i \to -\infty$ is OK (minimising sequence) but $a^i = a^{i+1} = \infty$ and therefore their ratio is not well-defined. Since f is continuous, $\{ d^i \} \to 0 \implies \{ a^i \} \to 0$, but the converse need not happen in general: say, $\{ x^{2i} \} \to x'_*$ and $\{ x^{2i+1} \} \to x''_*$ with $x'_* \neq x''_*$ optimal solutions **[back]**

- Simply, $\lim_{x \to 0} x^p / x = \lim_{x \to 0} x^{p-1} = 0$: the numerator goes to 0 faster than the denominator **[back]**