# Lecture 1: Problems and solutions. Optimality conditions for unconstrained optimization

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

# Problems and solutions

$$\text{minimize } f(x) \quad \text{subject to } x \in \Omega \subseteq \mathbb{R}^n. \qquad (\dagger)$$

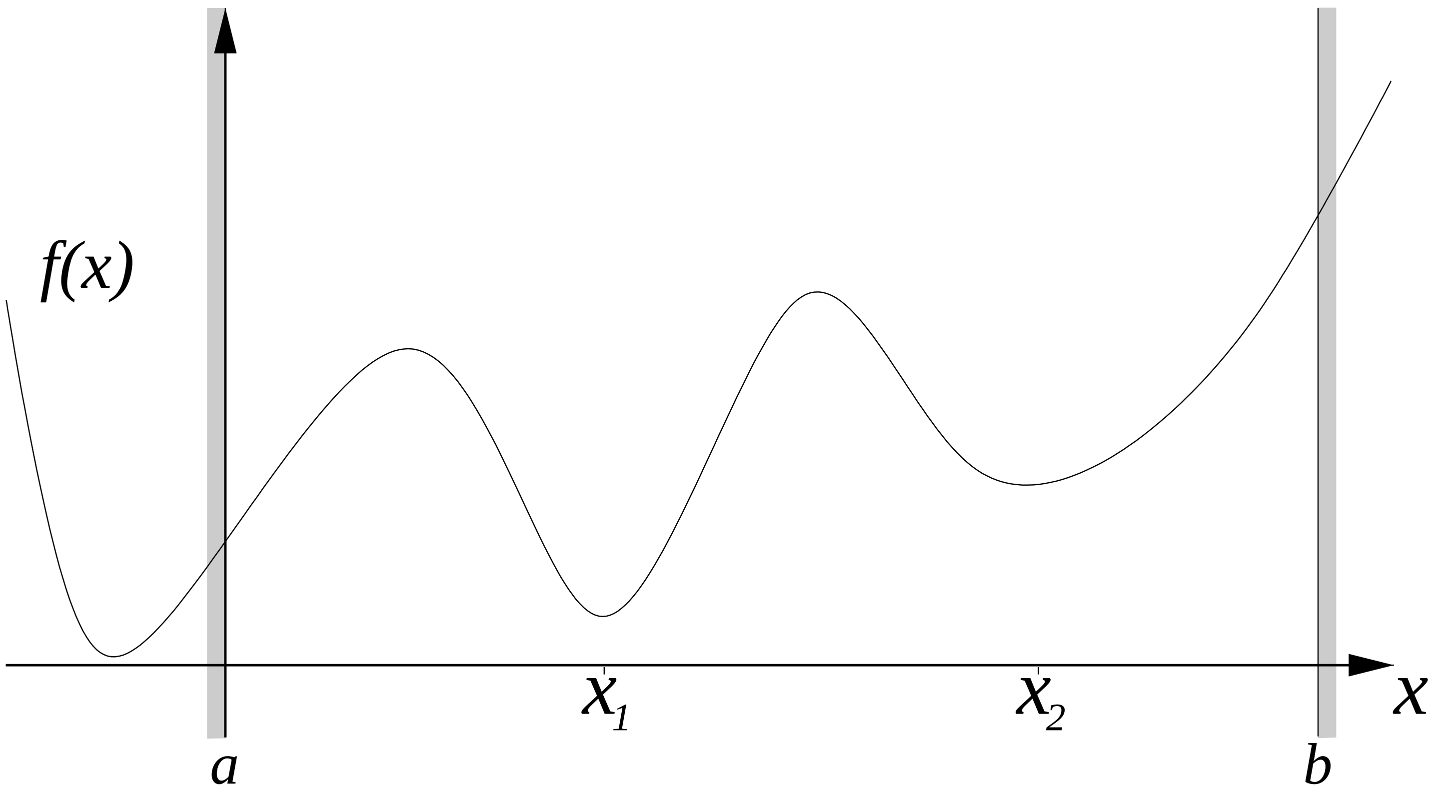
- $f : \Omega \to \mathbb{R}$ is (sufficiently) smooth ($f \in \mathcal{C}^i(\Omega)$, $i \in \{1, 2\}$).
- $f$ objective; $x$ variables; $\Omega$ feasible set (determined by finitely many constraints).
- $n$ may be large.
- minimizing $-f(x) \equiv -$ maximizing $f(x)$. Wlog, minimize.

$x^*$ global minimizer of $f$ over $\Omega \iff f(x) \geq f(x^*)$, $\forall x \in \Omega$.

$x^*$ local minimizer of $f$ over $\Omega \iff$ there exists $\mathcal{N}(x^*, \delta)$ such that $f(x) \geq f(x^*)$, for all $x \in \Omega \cap \mathcal{N}(x^*, \delta)$,
where $\mathcal{N}(x^*, \delta) := \{x \in \mathbb{R}^n : \|x - x^*\| \leq \delta\}$ and $\|\cdot\|$ is the Euclidean norm.

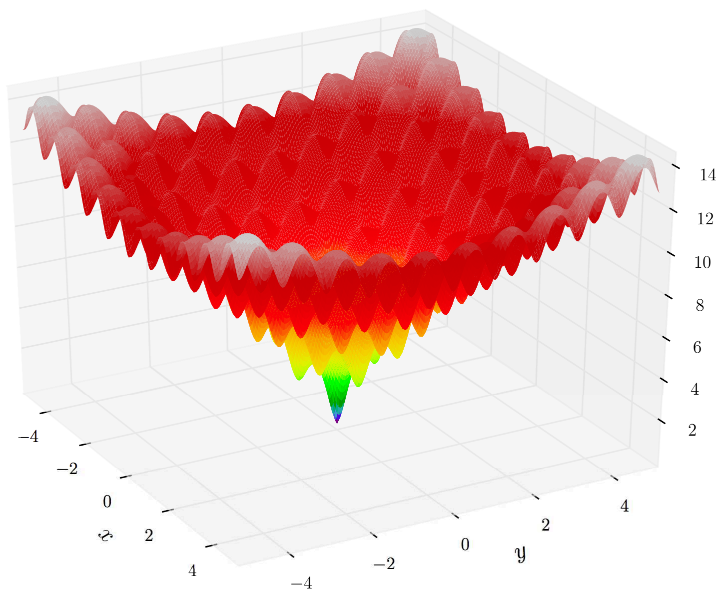
# Example problem in one dimension

$$\textbf{Example :}\quad \min f(x) \quad \text{subject to} \quad a \le x \le b.$$

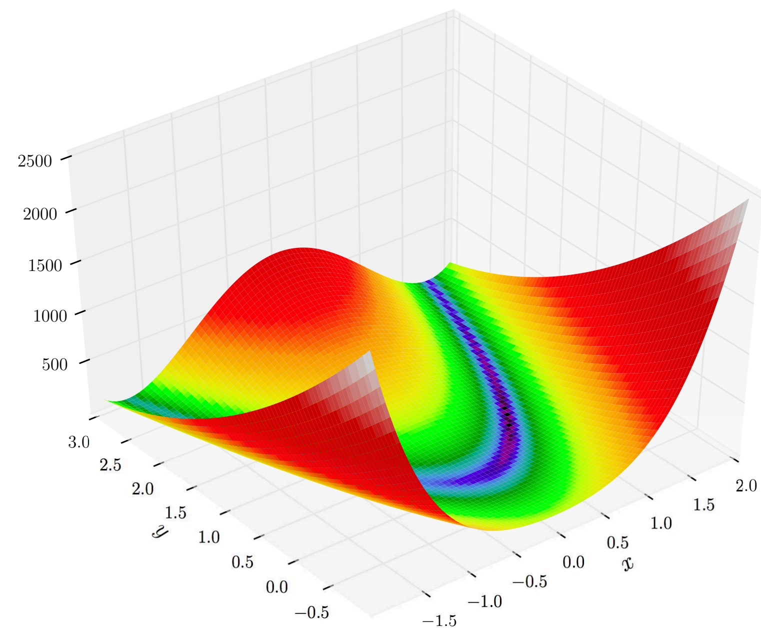



- The feasible region $\Omega$ is the interval $[a, b]$.
- The point $x_1$ is the global minimizer; $x_2$ is a local (non-global) minimizer; $x = a$ is a constrained local minimizer.

# Example problems in two dimensions




Ackeley's test function




Rosenbrock's test function

[see Wikipedia]

# Main classes of continuous optimization problems

Linear (Quadratic) programming: linear (quadratic) objective and linear constraints in the variables

$$\min_{x \in \mathbb{R}^n} c^T x \left( + \frac{1}{2} x^T H x \right) \text{ subject to } a_i^T x = b_i, i \in E; \; a_i^T x \geq b_i, i \in I,$$

where $c$, $a_i \in \mathbb{R}^n$ for all $i$ and $H$ is $n \times n$ symmetric matrix; $E$ and $I$ are finite index sets.

Unconstrained (Constrained) nonlinear programming

$$\min_{x \in \mathbb{R}^n} f(x) \;\; (\text{subject to } c_i(x) = 0, i \in E; \; c_i(x) \geq 0, i \in I)$$

where $f, c_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ are (smooth, possibly nonlinear) functions for all $i$; $E$ and $I$ are finite index sets.
Most real-life problems are nonlinear, often large-scale !

# Example: an OR application [Gould'06]

Optimization of a high-pressure gas network

pressures $p = (p_i,\ i)$; flows $q = (q_j,\ j)$; demands $d = (d_k,\ k)$; compressors. Maximize net flow s.t. the constraints:

$$\begin{cases} Aq - d = 0 \\ A^T p^2 + K q^{2.8359} = 0 \\ A_2^T q + z \cdot c(p, q) = 0 \\ p_{\min} \le p \le p_{\max} \\ q_{\min} \le q \le q_{\max} \end{cases}$$


- $A,\ A_2 \in \{\pm 1, 0\}$; $z \in \{0, 1\}$
- 200 nodes and pipes, 26 machines: 400 variables;
- variable demand, $(p, d)$ 10mins.

$\longrightarrow$ 58,000 vars; real-time.

# Example: an inverse problem application[MetOffice]

Data assimilation for weather forecasting

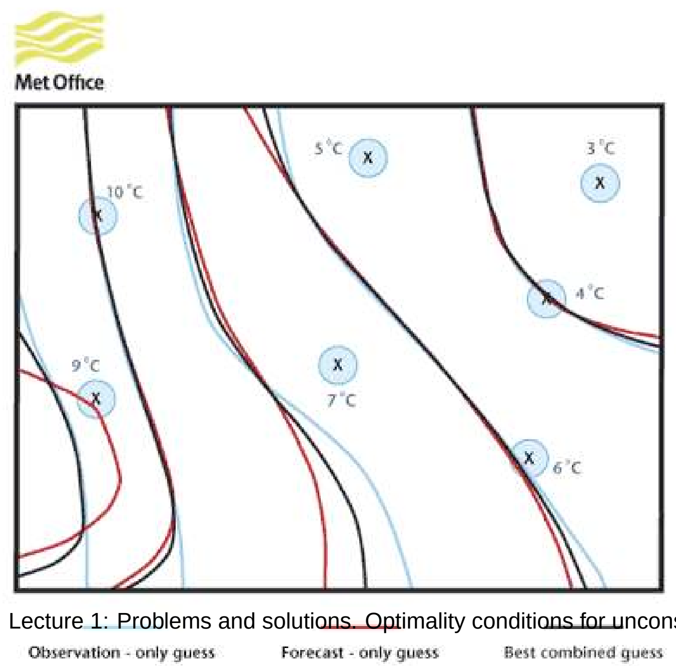
- best estimate of the current state of the atmosphere

$\longrightarrow$ find initial conditions $x_0$ for the numerical forecast by solving the (ill-posed) nonlinear inverse problem

$$\min_{x_0} \sum_{i=0}^{m} (H_i[x_i] - y_i)^T R_i^{-1} (H[x_i] - y_i),$$

$x_i = S(t_i, t_0, x_0)$, $S$ solution operator of the discrete nonlinear model; $H_i$ maps $x_i$ to observations $y_i$, $R_i$ error covariance matrix of the observations at $t_i$

$x_0$ of size $10^7 - 10^8$;
observations $m \approx$ 250,000.



# Optimality conditions for unconstrained problems

== algebraic characterizations of solutions $\longrightarrow$ suitable for computations.

- provide a way to guarantee that a candidate point is optimal (sufficient conditions)
- indicate when a point is not optimal (necessary conditions)

$$\text{minimize } f(x) \quad \text{subject to} \quad x \in \mathbb{R}^n. \qquad \text{(UP)}$$

First-order necessary conditions: $f \in \mathcal{C}^1(\mathbb{R}^n)$;

$x^*$ a local minimizer of $f \implies \nabla f(x^*) = 0$.

$\nabla f(x) = 0 \iff x$ stationary point of $f$.

# Optimality conditions for unconstrained problems...

Lemma 1. Let $f \in \mathcal{C}^1$, $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ with $s \neq 0$. Then $\nabla f(x)^T s < 0 \implies f(x + \alpha s) < f(x), \ \forall \alpha > 0$ sufficiently small.

Proof. $f \in \mathcal{C}^1 \implies \exists \overline{\alpha} > 0$ such that

$$\nabla f(x + \alpha s)^T s < 0, \ \forall \alpha \in [0, \overline{\alpha}]. \qquad (\lozenge)$$

Taylor's/Mean value theorem:

$f(x + \alpha s) = f(x) + \alpha \nabla f(x + \tilde{\alpha} s)^T s$, for some $\tilde{\alpha} \in (0, \alpha)$.

$(\lozenge) \implies f(x + \alpha s) < f(x)$, $\forall \alpha \in [0, \overline{\alpha}]$. $\square$

- $s$ descent direction for $f$ at $x$ if $\nabla f(x)^T s < 0$.

Proof of 1st order necessary conditions. assume $\nabla f(x^*) \neq 0$.
$s := -\nabla f(x^*)$ is a descent direction for $f$ at $x = x^*$:
$\nabla f(x^*)^T(-\nabla f(x^*)) = -\nabla f(x^*)^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$
since $\nabla f(x^*) \neq 0$ and $\|a\| \geq 0$ with equality iff $a = 0$.
Thus, by Lemma 1, $x^*$ is not a local minimizer of $f$. $\square$

# Optimality conditions for unconstrained problems...

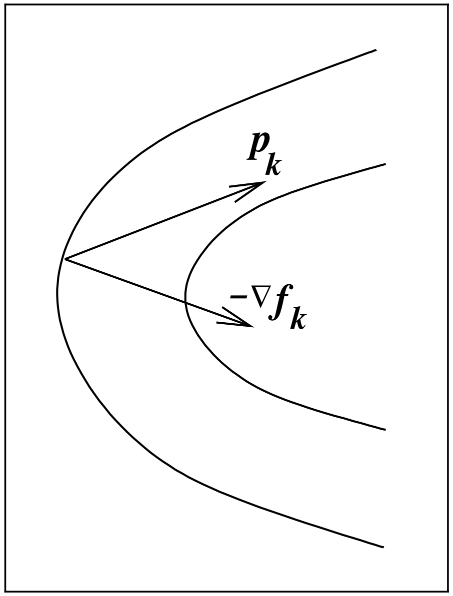
- $-\nabla f(x)$ is a descent direction for $f$ at $x$ whenever $\nabla f(x) \neq 0$.
- $s$ descent direction for $f$ at $x$ if $\nabla f(x)^T s < 0$, which is

equivalent to

$$\cos\langle -\nabla f(x), s\rangle = \frac{(-\nabla f(x))^T s}{\|\nabla f(x)\| \cdot \|s\|} = \frac{|\nabla f(x)^T s|}{\|\nabla f(x)\| \cdot \|s\|} > 0,$$

and so:

$\langle -\nabla f(x), s\rangle \in [0, \pi/2)$.




A descent direction $p_k$.

# Summary of first-order conditions. A look ahead

$$\text{minimize } f(x) \quad \text{subject to} \quad x \in \mathbb{R}^n. \qquad \text{(UP)}$$

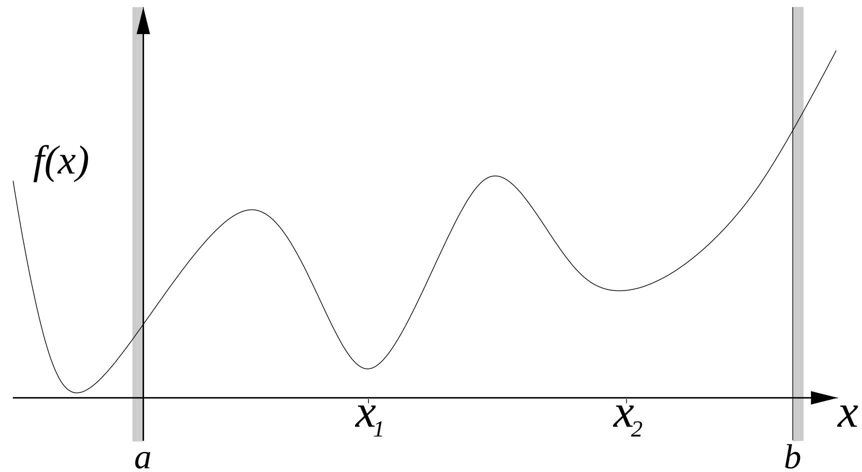
First-order necessary optimality conditions: $f \in \mathcal{C}^1(\mathbb{R}^n)$;

$x^*$ a local minimizer of $f \implies \nabla f(x^*) = 0$.

$$\tilde{x} = \arg\max_{x \in \mathbb{R}^n} f(x)$$

$$\Downarrow$$

$$\nabla f(\tilde{x}) = 0.$$




- Look at higher-order derivatives to distinguish between minimizers and maximizers.

. . . except for convex functions.

# Optimality conditions for convex problems

- $f$ convex $\iff f(x+\alpha(y-x)) \le f(x) + \alpha(f(y)-f(x))$,
  for all $x,\ y \in \mathbb{R}^n$, $\alpha \in [0,1]$.
- $\iff \nabla^2 f(x) \succeq 0$ (positive semidefinite), for all $x \in \mathbb{R}^n$, i.e.,
  - $s^T \nabla^2 f(x^*) s \ge 0$, $\forall\, s \in \mathbb{R}^n$; equivalently,
  - eigenvalues $\lambda_i(\nabla^2 f(x^*)) \ge 0$, $\forall i \in \{1, \ldots, n\}$.

If $f$ convex, then [Pb Sheet 1]

$$x^* \text{ local minimizer} \implies x^* \text{ global minimizer.}$$

$$x^* \text{ stationary point} \implies x^* \text{ global minimizer.}$$

Quadratic functions: $q(x) := g^T x + \frac{1}{2} x^T H x$.

$\nabla^2 q(x) = H$, for all $x$; if $H$ is positive semidefinite, then $q$ convex; any stationary point $x^*$ is a global minimizer of $q$.

# Second-order optimality conditions (nonconvex fcts.)

Second-order necessary conditions: $f \in \mathcal{C}^2(\mathbb{R}^n)$;

$x^*$ local minimizer of $f \implies \nabla^2 f(x^*) \succeq 0$ (positive semidefinite),

namely, $s^T \nabla^2 f(x^*) s \geq 0$, for all $s \in \mathbb{R}^n$. [local convexity]

Example: $f(x) := x^3$, $x^* = 0$ not a local minimizer but $f'(0) = f''(0) = 0$.

Second-order sufficient conditions: $f \in \mathcal{C}^2(\mathbb{R}^n)$;

$\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) \succ 0$ (positive definite), namely,

$$s^T \nabla^2 f(x^*) s > 0, \text{ for all } s \neq 0.$$

$\implies x^*$ (strict) local minimizer of $f$.

Example: $f(x) := x^4$, $x^* = 0$ is a (strict) local minimizer but $f''(0) = 0$.

# Proof of second-order conditions

Let $x$ and $s \neq 0$ in $\mathbb{R}^n$ be fixed. Let $\Phi : [0, \infty) \longrightarrow \mathbb{R}$ where $\Phi(\alpha) := f(x + \alpha s)$ with $f \in \mathcal{C}^2(\mathbb{R}^n)$. Then (univariate) Taylor's/Mean-value theorem gives for any $\alpha > 0$ that

$$\Phi(\alpha) = \Phi(0) + \alpha\Phi'(0) + \frac{\alpha^2}{2}\Phi''(\tilde{\alpha}) \text{ for some } \tilde{\alpha} \in (0, \alpha),$$

or equivalently, from def. of $\Phi$ and differentiation/chain rule:[Pb Sheet 1]

$$f(x + \alpha s) = f(x) + \alpha s^T \nabla f(x) + \frac{\alpha^2}{2} s^T \nabla^2 f(x + \tilde{\alpha} s) s \qquad (\Diamond)$$

for some $\tilde{\alpha} \in (0, \alpha)$.

Proof of second order necessary conditions. Assume there exists $s \in \mathbb{R}^n$ with $s^T \nabla^2 f(x^*) s < 0$. Then $s \neq 0$ and $f \in \mathcal{C}^2$ imply $s^T \nabla^2 f(x^* + \alpha s) s < 0$ for all $\alpha \in [0, \overline{\alpha}]$. Employing this and $\nabla f(x^*) = 0$ in ($\Diamond$) with $x = x^*$ gives that for each $\alpha \in (0, \overline{\alpha})$, there exists $\tilde{\alpha} \in (0, \alpha)$ such that

$$f(x^* + \alpha s) = f(x^*) + \frac{\alpha^2}{2} s^T \nabla^2 f(x^* + \tilde{\alpha} s) s < f(x^*).$$

We have reached a contradiction with $x^*$ being a local minimizer.□

# Proof of second-order conditions ...

Recall (from previous slide) that for $x \in \mathbb{R}^n$, $s \neq 0$ and $\alpha > 0$,

$$f(x+\alpha s) = f(x) + \alpha s^T \nabla f(x) + \frac{\alpha^2}{2} s^T \nabla^2 f(x + \tilde{\alpha} s) s \qquad (\Diamond)$$

for some $\tilde{\alpha} \in (0, \alpha)$.

Proof of second order sufficient conditions. $f \in \mathcal{C}^2$ and $\nabla^2 f(x^*) \succ 0$ imply $\nabla^2 f(x^* + s) \succ 0$ for all $x^* + s \in \mathcal{N}(x^*, \delta)$ some neighbourhood of $x^*$. For any such $s$ with $\|s\| \leq \delta$, $(\Diamond)$ with $\alpha = 1$ and $x = x^*$, gives, for some $\tilde{\alpha} \in (0, 1)$,

$$f(x^* + s) = f(x^*) + \frac{1}{2} s^T \nabla^2 f(x^* + \tilde{\alpha} s) s \geq f(x^*)$$

where we also used $\nabla f(x^*) = 0$ in the first equality and $\nabla^2 f(x^* + \tilde{\alpha} s) \succ 0$ in the second inequality (note that $\|x^* + \tilde{\alpha} s - x^*\| \leq \delta$ since $\|s\| \leq \delta$ and $\tilde{\alpha} \in (0, 1)$; thus $x^* + \tilde{\alpha} s \in \mathcal{N}(x^*, \delta)$ which ensures that $\nabla^2 f(x^* + \tilde{\alpha} s) \succ 0$.) $\Box$

# Stationary points of quadratic functions

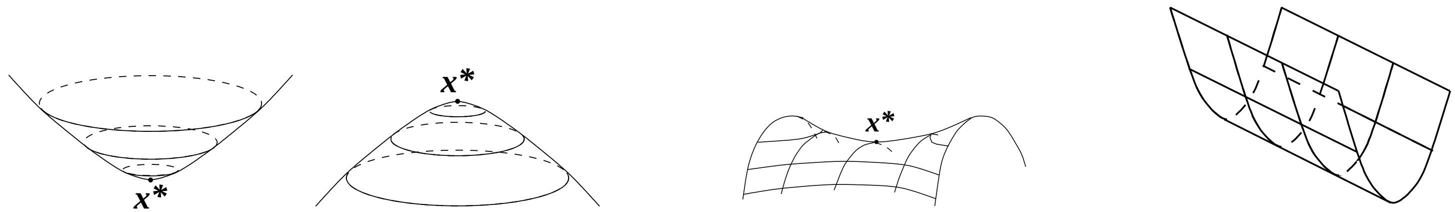
- $H \in \mathbb{R}^{n\times n}$ symmetric, $g \in \mathbb{R}^n$: $q(x) := g^T x + \frac{1}{2} x^T H x$.

$\nabla q(x^*) = 0 \iff Hx^* + g = 0$: linear system.

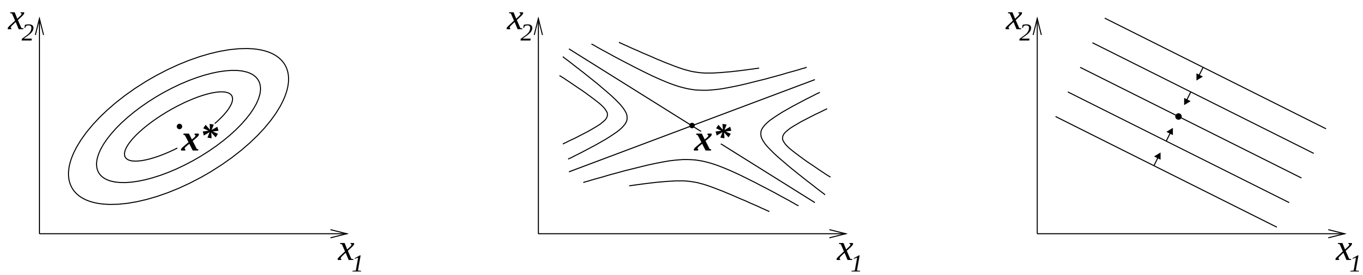
- $H$ nonsingular: $x^* = -H^{-1} g$ unique stationary point.
  - $H$ positive definite $\Longrightarrow$ $x^*$ minimizer (a), e)).
  - $H$ negative definite $\Longrightarrow$ $x^*$ maximizer (b), e)).
  - $H$ indefinite $\Longrightarrow$ $x^*$ saddle point (c), f)).
- $H$ singular and $g + Hx = 0$ consistent:
  - $H$ positive semidefinite $\Longrightarrow$ infinitely many global minimizers (d), g)).
  - Similarly $H$ negative semidefinite or indefinite.

General $f$: approximately locally quadratic around $x^*$ stationary.

# Stationary points of quadratic functions...




(a) Minimum (b) Maximum (c) Saddle (d) Semidefinite




(e) Maximum or minimum (f) Saddle (g) Semidefinite