Basics of Unconstrained Optimization

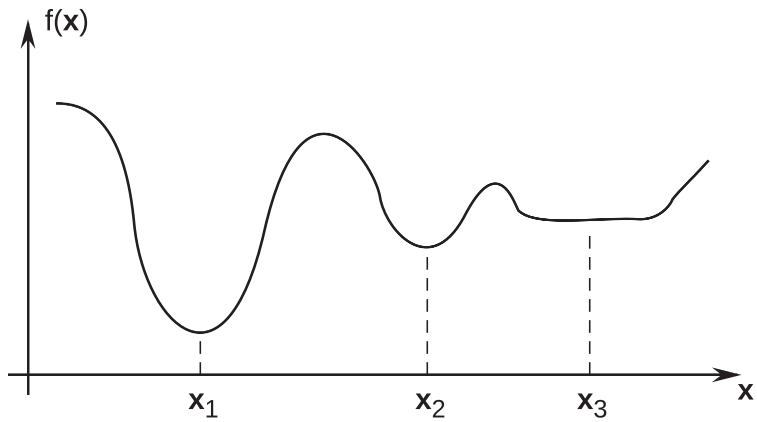
Introduction (§6.1)

- Optimization problem:

$$\begin{array}{rl} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \Omega \end{array}$$

- Solution to the problem: a *minimizer*, $\boldsymbol{x}^*$.
- Idea: $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega \setminus \{\boldsymbol{x}^*\}$.
- Ω: constraint set or feasible set.

Types of minimizers



Several ways to classify minimizer $\boldsymbol{x}^*$:

- Global minimizer: $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega \setminus \{\boldsymbol{x}^*\}$.
- Local minimizer: there exists $\varepsilon > 0$ such that $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega \setminus \{\boldsymbol{x}^*\}$ and $\|\boldsymbol{x} - \boldsymbol{x}^*\| < \varepsilon$.
- Also, *strict* global minimizer and local minimizer.
- Ideal solution: global minimizer.
- Often have to be satisfied with local minimizer.

Existence of minimizers

- Theorem of Weierstrass: If f is continuous and Ω is closed and bounded, then a global minimizer exists.

Conditions: Necessary and Sufficient

- We seek conditions that characterize minimizers.
- Two types of conditions: necessary and sufficient.
- Necessary condition: If $\boldsymbol{x}^*$ is a minimizer, then $\boldsymbol{x}^*$ satisfies this particular condition.
- Sufficient condition: If $\boldsymbol{x}^*$ satisfies this particular condition, then $\boldsymbol{x}^*$ is a minimizer.

- A necessary condition limits the set of candidates for minimizers.
- A sufficient condition guarantees that a point is a minimizer.
- We consider conditions that are based on gradients and Hessians. These conditions apply to *local minimzers*.

Conditions for local minimizers (§6.2)

- Consider the totally unconstrained problem:

$$\begin{array}{rl} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^n. \end{array}$$

- Assume $f \in \mathcal{C}^1$.
- Theorem: If $\boldsymbol{x}^*$ is a local minimizer, then

$$\nabla f(\boldsymbol{x}^*) = \mathbf{0}.$$

- For $n = 1$, this theorem should be very familiar ("slope=0").
- First order necessary condition (FONC).

Idea of proof of theorem (by contraposition):

- Suppose $\nabla f(\boldsymbol{x}^*) \neq \mathbf{0}$.
- Since $-\nabla f(\boldsymbol{x}^*)$ points in the direction of decreasing f, there will be some points close to $\boldsymbol{x}^*$ that have smaller f value.
- Specifically, consider $\boldsymbol{x}_\alpha = \boldsymbol{x}^* - \alpha \nabla f(\boldsymbol{x}^*)$, $\alpha > 0$. From an equation before (using Taylor's formula),

$$f(\boldsymbol{x}_\alpha) = f(\boldsymbol{x}^*) - \alpha \|\nabla f(\boldsymbol{x}^*)\|^2 + o(\alpha).$$

- For sufficiently small α, we have $f(\boldsymbol{x}_\alpha) < f(\boldsymbol{x}^*)$.
- Hence, $\boldsymbol{x}^*$ is not a local minimizer.

Example: quadratic case

- Let f be a quadratic:
$$f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c,$$
- We have $\nabla f(\boldsymbol{x}) = 2\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}$.
- If $\boldsymbol{x}^*$ is a local minimizer, then $2\boldsymbol{Q}\boldsymbol{x}^* + \boldsymbol{b} = \boldsymbol{0}$.
- If $\boldsymbol{Q} > 0$, then
$$\boldsymbol{x}^* = -\frac{1}{2}\boldsymbol{Q}^{-1}\boldsymbol{b}.$$
- In general, $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ is necessary but not sufficient for local minimizer.

- General constrained problem:
$$\begin{array}{rl} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \Omega \end{array}$$
- If $\boldsymbol{x}^*$ is a local minimizer and an *interior* point of Ω, then same result as previously. Why?
- Need to consider the boundary case.
- $\boldsymbol{x}^*$ is a *boundary point* of Ω if inside any ball around $\boldsymbol{x}^*$, there are points inside Ω and points outside Ω.

Directional derivatives

- Given $f : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{x}_0 \in \mathbb{R}^n$.
- Consider a vector $\boldsymbol{d} \in \mathbb{R}^n$.
- Construct the function $\phi : \mathbb{R} \to \mathbb{R}$ by
$$\phi(\alpha) = f(\boldsymbol{x}_0 + \alpha\boldsymbol{d}).$$
- The directional derivative of f at $\boldsymbol{x}_0$ in the direction $\boldsymbol{d}$ is
$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}_0) = \lim_{\alpha\to 0} \frac{f(\boldsymbol{x}_0 + \alpha\boldsymbol{d}) - f(\boldsymbol{x}_0)}{\alpha} = \phi'(0).$$

- By the chain rule,
$$\phi'(0) = \nabla f(\boldsymbol{x}_0)^T \boldsymbol{d} = \boldsymbol{d}^T \nabla f(\boldsymbol{x}_0)$$

- Hence,
$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}_0) = \boldsymbol{d}^T \nabla f(\boldsymbol{x}_0)$$

- Note: The directional derivative of f at $\boldsymbol{x}_0$, in a direction tangent to the level set, is 0.

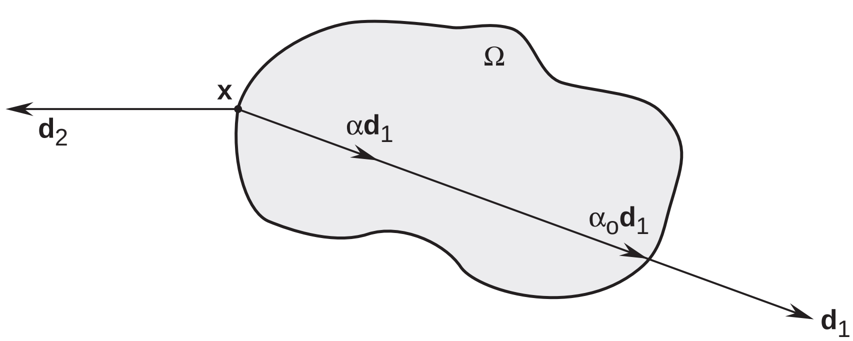
Example:

- Define $f : \mathbb{R}^3 \to \mathbb{R}$ by $f(\boldsymbol{x}) = x_1 x_2 x_3$.
- Direction $\boldsymbol{d} = [1, 1, \sqrt{2}]^T$.
- We have

$$\begin{aligned}\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) &= \boldsymbol{d}^T \nabla f(\boldsymbol{x}) = \left[1, 1, \sqrt{2}\right] \begin{bmatrix} x_2 x_3 \\ x_1 x_3 \\ x_1 x_2 \end{bmatrix} \\ &= x_2 x_3 + x_1 x_3 + \sqrt{2} x_1 x_2.\end{aligned}$$

Feasible directions

- Given: a nonzero vector $\boldsymbol{d} \in \mathbb{R}^n$.
- $\boldsymbol{d}$ is a *feasible direction* at $\boldsymbol{x} \in \Omega$ if there exists $\alpha_0 > 0$ such that $\boldsymbol{x} + \alpha \boldsymbol{d} \in \Omega$ for all $\alpha \in [0, \alpha_0]$.
- At an interior point, all directions are feasible.
- At a boundary point, only some directions are feasible. These are directions that "point into" the set.

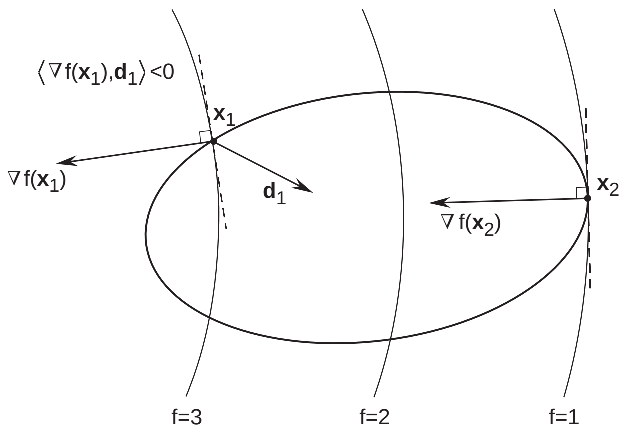


- Theorem: If $\boldsymbol{x}^*$ is a local minimizer, then

$$\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*) \geq \mathbf{0}$$

  for all feasible directions $\boldsymbol{d}$.

- First order necessary condition (FONC), general case.



- Interpretation: if $\boldsymbol{x}^*$ is a local minimizer, then the directional derivative of f in any feasible direction must be ≥ 0 (because the function must be increasing in that direction).

- In the interior case, the general FONC reduces to $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.

- Read proof in book (it is similar to what we gave earlier).

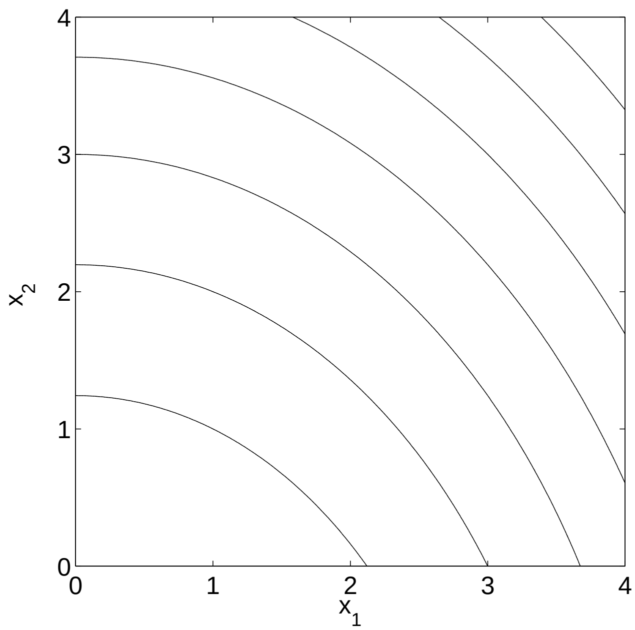
Summary: First order necessary condition

- Interior case: $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$

- General case: $\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*) \geq \mathbf{0}$ for all feasible directions $\boldsymbol{d}$.

- Example (6.3)

$$\begin{array}{ll} \text{minimize} & x_1^2 + 0.5x_2^2 + 3x_2 + 4.5 \\ \text{subject to} & x_1, x_2 \geq 0. \end{array}$$



Second order conditions (§6.2)

- The FONC is only necessary, but not sufficient (in general).
- Can we say more?
- Yes, if we use second derivatives also.
- Example:

$$f(x) = ax^2 + bx + c$$

  $x^* = -b/2a$ satisfies FONC.

$$x^* \text{ minimizer} \qquad \Rightarrow \qquad a \geq 0.$$

- Consider the totally unconstrained problem:

$$\begin{array}{rl} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^n. \end{array}$$

- Assume $f \in \mathcal{C}^2$.
- Theorem: If $\boldsymbol{x}^*$ is a local minimizer, then

$$\boldsymbol{F}(\boldsymbol{x}^*) \geq 0.$$

- Second order necessary condition (SONC).

- Think about quadratic case.

Idea of proof of theorem (by contraposition):

- Suppose $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ but for some $\boldsymbol{d}$, we have $\boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} < 0$.
- Consider $\boldsymbol{x}_\alpha = \boldsymbol{x}^* + \alpha \boldsymbol{d}$, $\alpha > 0$. Using Taylor's formula,

$$f(\boldsymbol{x}_\alpha) = f(\boldsymbol{x}^*) + \frac{1}{2}\alpha^2 \boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} + o(\alpha^2).$$

- For sufficiently small α, we have $f(\boldsymbol{x}_\alpha) < f(\boldsymbol{x}^*)$.
- Hence, $\boldsymbol{x}^*$ is not a local minimizer.

- General constrained problem:

$$\begin{array}{rl} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \Omega \end{array}$$

- If $\boldsymbol{x}^*$ is a local minimizer and an interior point of Ω, then same result as previously.
- Directional second derivative in direction $\boldsymbol{d}$: $\boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d}$.
- For boundary case, positivity of 2nd derivative unnecessary if $\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*) > 0$. (Why?)

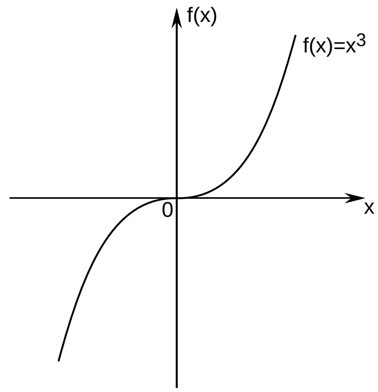
- Theorem: Suppose $\boldsymbol{x}^*$ is a local minimizer, and $\boldsymbol{d}$ a feasible direction. If $\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*) = \mathbf{0}$, then $\boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0$.
- Second order necessary condition (SONC), general case.
- Read proof in book (it is similar to what we gave earlier).

Summary: Second order necessary condition

- Interior case: $\boldsymbol{F}(\boldsymbol{x}^*) \geq \mathbf{0}$
- General case: If $\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*) = \mathbf{0}$ for a feasible direction $\boldsymbol{d}$, then $\boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0$.

- SONC is necessary, but not sufficient (in general).

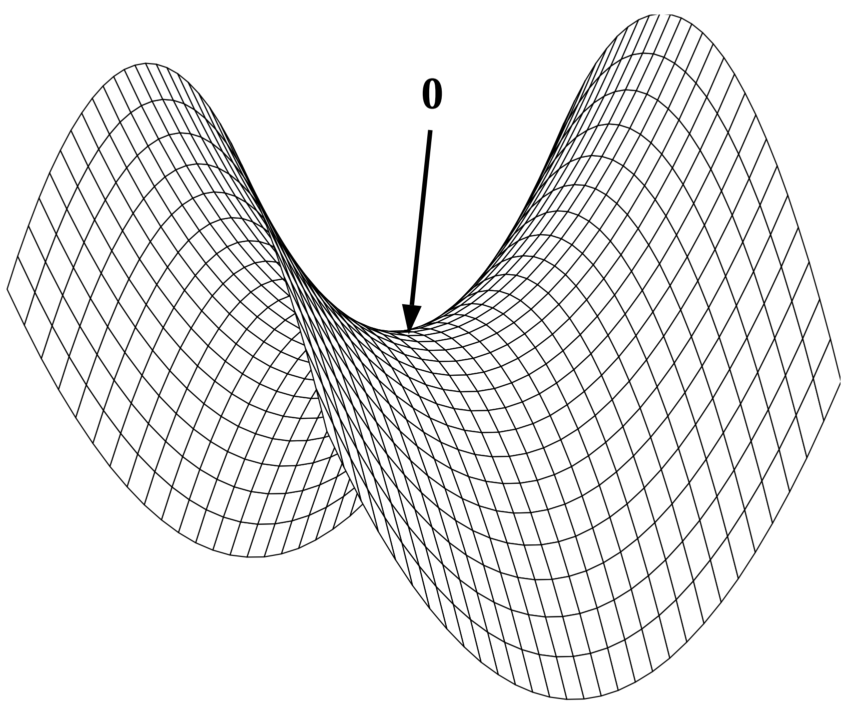
- Example (6.4): $f(x) = x^3$, $x \in \mathbb{R}$.



- $f'(0) = 0$, $f''(0) = 0$.
- 0 is not a minimizer.

Example (6.5)

- Consider $f(\boldsymbol{x}) = x_1^2 - x_2^2$.
- $\nabla f(\boldsymbol{x}) = [2x_1, -2x_2]^T$, $\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$.
- FONC is satisfied at $\boldsymbol{x}^* = \boldsymbol{0}$.
- SONC not satisfied at $\boldsymbol{x}^* = \boldsymbol{0}$. For if $\boldsymbol{d} = [0, 1]^T$, then $\boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{0}) \boldsymbol{d} < 0$.



What about a sufficient condition?

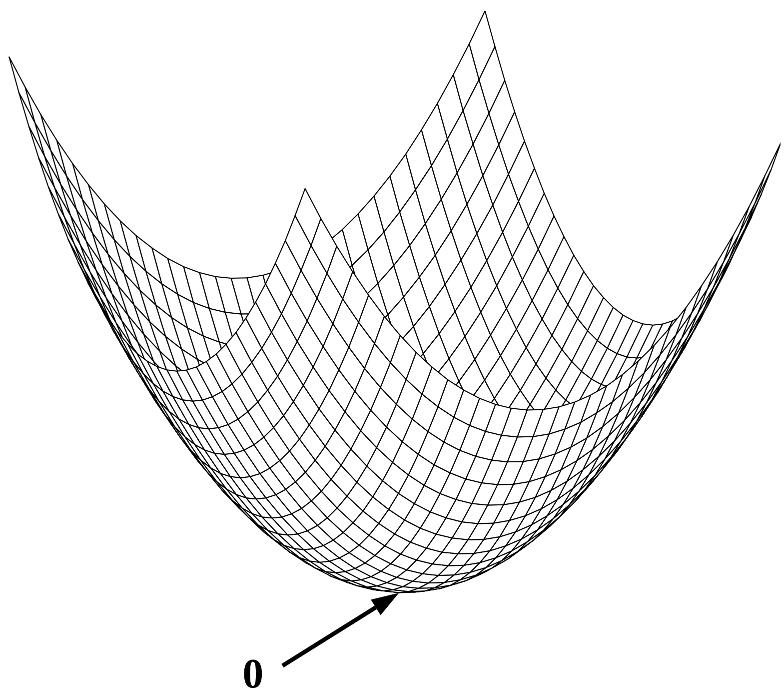
- Suppose $f \in \mathcal{C}^2$, $\boldsymbol{x}^*$ an *interior* point.
- Theorem: Suppose that $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ and $\boldsymbol{F}(\boldsymbol{x}^*) > 0$. Then, $\boldsymbol{x}^*$ is a strict local minimizer.
- Second order sufficient condition (SOSC).

Example (6.5)

- Consider $f(\boldsymbol{x}) = x_1^2 - x_2^2$.
- $\nabla f(\boldsymbol{x}) = [2x_1, -2x_2]^T$, $\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$.
- No point satisfies the SOSC.

Example (6.6)

- Consider $f(\boldsymbol{x}) = x_1^2 + x_2^2$.
- $\nabla f(\boldsymbol{x}) = [2x_1, 2x_2]^T$, $\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$.
- $\boldsymbol{x}^* = \boldsymbol{0}$ satisfies the SOSC.



Sketch of proof of SOSC:

- Rayleigh's inequality: $\boldsymbol{F}(\boldsymbol{x}^*) > 0$ implies f is locally bounded below by a positive definite quadratic function at $\boldsymbol{x}^*$.
- For bounding quadratic function, $\boldsymbol{x}^*$ strict local minimizer.
- Hence, $\boldsymbol{x}^*$ must be a strict local minimizer for f.

A note about quadratics

- Consider a quadratic
$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b} + c,$$
where $\boldsymbol{Q} > 0$.

- We can write f as
$$f(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^T\boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{x}^*) + \left(c - \frac{1}{2}\boldsymbol{x}^{*T}\boldsymbol{Q}\boldsymbol{x}^*\right),$$
where $\boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{b}$.

- Hence, $\boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{b}$ is the unique global minimizer.

How do we use optimality conditions?

- To "solve" for minimizers.

Example: unconstrained case

- First, find points satisfying $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ (FONC).
- Then, among these, check for $\boldsymbol{F}(\boldsymbol{x}^*) \geq 0$ (SONC).
- Then, among those remaining, check for $\boldsymbol{F}(\boldsymbol{x}^*) > 0$ (SOSC).
- These are guaranteed to be strict local minimizers.

Alternatively:

- Among those points satisfying the FONC, evaluate objective function values one by one.
- Pick the point with smallest objective function value.
- When does this give us a global minimizer?

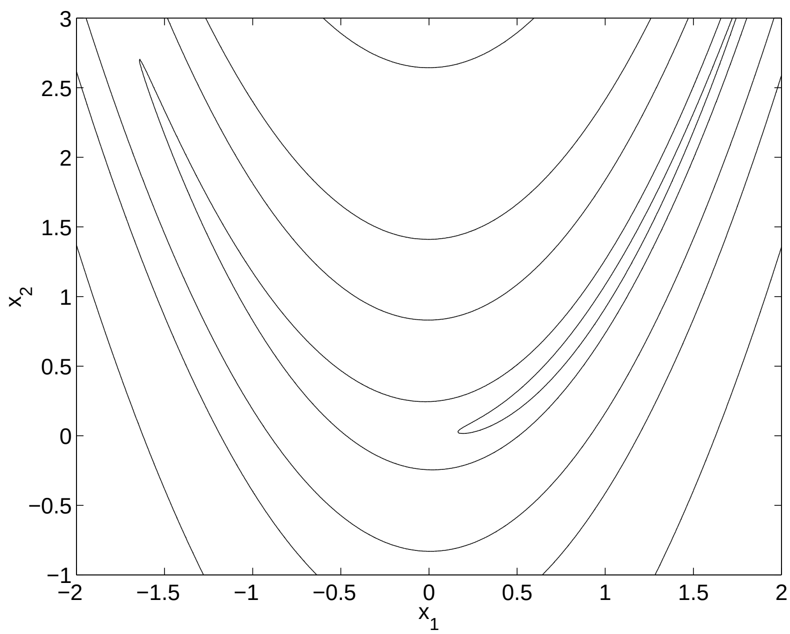
- Potential problem: Solving $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ may be very difficult.
- Numerical algorithms for solving $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ may be as complex as algorithms for solving the original optimization problem!
- Often, we directly apply numerical algorithms to the optimization problem.
- Optimality conditions remain useful in *development* and *analysis* of optimization techniques.

Optimization algorithms

- Basic form:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}$$

- $\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots$: iterates
- α_k: step size (positive)
- $\boldsymbol{d}^{(k)}$: search direction



- Typical choice for α_k: (greedy strategy)

$$\alpha_k = \arg\min_{\alpha \geq 0} f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}).$$

  "Line search"

- Choice of $\boldsymbol{d}^{(k)}$:
  - Tradeoff between fast convergence and complexity.
  - Want to guarantee descent property.
- Algorithm has *descent* property if $f(\boldsymbol{x}^{(k+1)}) < f(\boldsymbol{x}^{(k)})$ whenever $\nabla f(\boldsymbol{x}^{(k)}) \neq \mathbf{0}$.

