Unconstrained Minimization Algorithm

(1) Initialize $\boldsymbol{x}^0$ and ϵ, set $k := 0$.

(2) **while** $\|\boldsymbol{g}(\boldsymbol{x}^k)\| > \epsilon$

- (a) Find a descent direction $\boldsymbol{d}^k$ for f at $\boldsymbol{x}^k$
- (b) Find $\alpha^k (> 0)$ along $\boldsymbol{d}^k$ such that
  - (i) $f(\boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k) < f(\boldsymbol{x}^k)$
  - (ii) α^k satisfies Armijo-Wolfe conditions
- (c) $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k$
- (d) $k := k + 1$

endwhile

Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$, a stationary point of $f(\boldsymbol{x})$.

Does this algorithm converge?

Global Convergence Theorem

Global Convergence Theorem [*Zoutendijk*]

Consider the problem to minimize $f(\boldsymbol{x})$ over $\mathbb{R}^n$. Suppose f is bounded below in $\mathbb{R}^n$, $f \in \mathcal{C}^1$ and the gradient, $\nabla f (= \boldsymbol{g})$ is Lipschitz continuous. If at every iteration k of an optimization algorithm, a descent direction $\boldsymbol{d}^k$ is chosen such that $\cos^2 \theta_k > \delta(> 0)$ (where θ_k is the angle between $\boldsymbol{d}^k$ and $\boldsymbol{g}^k$) and α^k satisfies Armijo-Wolfe conditions, then the optimization algorithm either *terminates in a finite number of iterations* or

$$\lim_{k \to \infty} \|\boldsymbol{g}^k\| = 0.$$

Sufficient Decrease and Backtracking

- Armijo-Goldstein Conditions: Choose α^k such that

$$\phi_2(\alpha^k) \leq f(\boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k) \leq \phi_1(\alpha^k)$$

where $\phi_1(\alpha) = f(\boldsymbol{x}^k) + c_1 \alpha \boldsymbol{g}^{k^T} \boldsymbol{d}^k, \;\; c_1 \in (0,1)$ and $\phi_2(\alpha) = f(\boldsymbol{x}^k) + c_2 \alpha \boldsymbol{g}^{k^T} \boldsymbol{d}^k, \;\; c_2 \in (c_1, 1)$.

- Use of *backtracking* line search with Armijo's condition

Backtracking Line Search

---

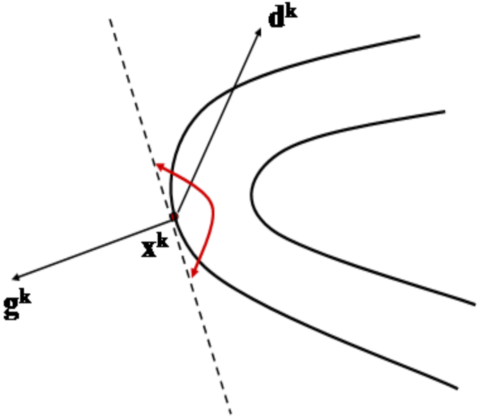
(1) Choose $\hat{\alpha}(> 0), \rho \in (0,1), c_1 \in (0,1)$. Set $\alpha = \hat{\alpha}$.

(2) **while** $f(\boldsymbol{x}^k + \alpha \boldsymbol{d}^k) > f(\boldsymbol{x}^k) + c_1 \alpha \boldsymbol{g}^{k^T} \boldsymbol{d}^k$

$\alpha := \rho\alpha$

endwhile

Output : $\alpha^k = \alpha$



- Descent direction set: $\{\boldsymbol{d} \in \mathbb{R}^n : \boldsymbol{g}^{k^T}\boldsymbol{d} < 0\}$ where $\boldsymbol{g}^k = \boldsymbol{g}(\boldsymbol{x}^k)$

Descent Directions

- Let $\boldsymbol{g}^k \neq \boldsymbol{0}$ and $\boldsymbol{d}^k = -\boldsymbol{A}^k\boldsymbol{g}^k$ where $\boldsymbol{A}^k$ is a symmetric matrix
- If $\boldsymbol{A}^k$ is positive definite,

$$\begin{aligned} {\boldsymbol{g}^k}^T\boldsymbol{d}^k &= -{\boldsymbol{g}^k}^T\boldsymbol{A}^k\boldsymbol{g}^k < 0 \\ &\Rightarrow \quad \boldsymbol{d}^k \text{ is a descent direction} \end{aligned}$$

- $\boldsymbol{d}^k = -\boldsymbol{A}^k\boldsymbol{g}^k$ is a *descent direction* if $\boldsymbol{A}^k$ is positive definite.
- Different optimization algorithms use different $\boldsymbol{A}^k$

How to find $\boldsymbol{d}^k$?

Consider the first order approximation to $f(x)$ about $\boldsymbol{x}^k$:

$$f(\boldsymbol{x}) \approx \hat{f}(\boldsymbol{x}) \triangleq f(\boldsymbol{x}^k) + \boldsymbol{g}^{k^T}(\boldsymbol{x} - \boldsymbol{x}^k) = f(\boldsymbol{x}^k) + \boldsymbol{g}^{k^T}\boldsymbol{d}$$

Maximum decrease in $\hat{f}(\boldsymbol{x})$ is possible by solving (**P1**):

$$\begin{array}{ll} \min_{\boldsymbol{d}} & \boldsymbol{g}^{k^T}\boldsymbol{d} \\ \text{s.t.} & \boldsymbol{d}^T\boldsymbol{d} = 1 \end{array}$$

Let θ_k be the angle between $\boldsymbol{g}^k$ and $\boldsymbol{d}$.

$$\begin{aligned} \boldsymbol{g}^{k^T}\boldsymbol{d} &= \|\boldsymbol{g}^k\|\|\boldsymbol{d}\|\cos\theta_k \\ &= \|\boldsymbol{g}^k\|\cos\theta_k \;\; (\because \boldsymbol{d}^T\boldsymbol{d} = 1) \end{aligned}$$

Therefore, the solution to the problem (**P1**) is $\boldsymbol{d}^k = -\boldsymbol{g}^k/\|\boldsymbol{g}^k\|$

Steepest Descent Method

- Uses the steepest descent direction, $\boldsymbol{d}^k = -\boldsymbol{g}^k$

Steepest Descent Algorithm

---

(1) Initialize $\boldsymbol{x}^0$ and ϵ, set $k := 0$.

(2) **while** $\|\boldsymbol{g}^k\| > \epsilon$

  (a) $\boldsymbol{d}^k = -\boldsymbol{g}^k$

  (b) Find $\alpha^k (> 0)$ along $\boldsymbol{d}^k$ such that

    (i) $f(\boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k) < f(\boldsymbol{x}^k)$

    (ii) α^k satisfies Armijo-Wolfe conditions

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

  (d) $k := k + 1$

endwhile

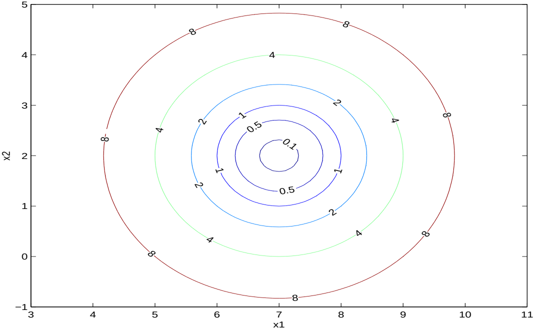
Output : $\boldsymbol{x}^* = \boldsymbol{x}^k$, a stationary point of $f(\boldsymbol{x})$.

- Exact or Backtracking line search can be used in step 2(b)

Example:

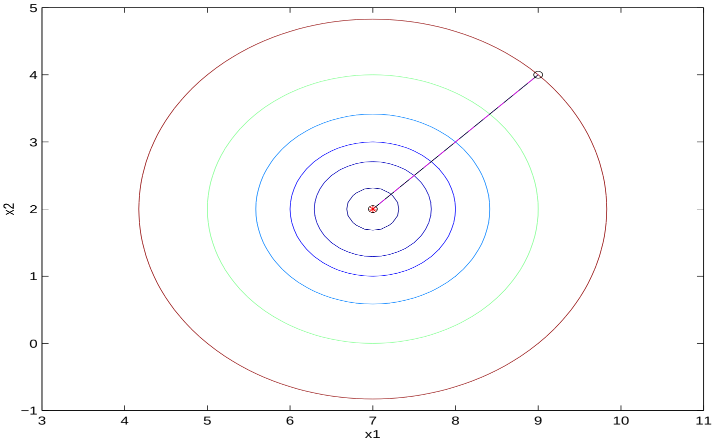
$$\min \quad f(\boldsymbol{x}) \triangleq (x_1 - 7)^2 + (x_2 - 2)^2$$

- $\boldsymbol{g}(\boldsymbol{x}) = \begin{pmatrix} 2(x_1 - 7) \\ 2(x_2 - 2) \end{pmatrix}$, $\boldsymbol{H}(\boldsymbol{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.
- $\boldsymbol{x}^* = \begin{pmatrix} 7 \\ 2 \end{pmatrix}$



Example:

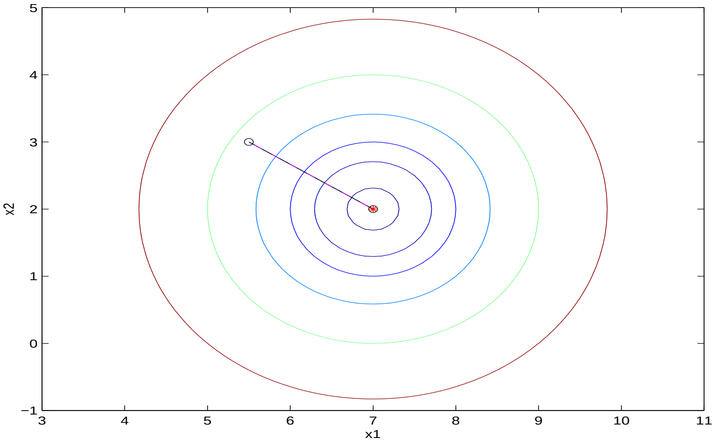
$$\min \quad f(\boldsymbol{x}) \triangleq (x_1 - 7)^2 + (x_2 - 2)^2$$



Behaviour of the steepest descent algorithm (with exact line search) applied to $f(\boldsymbol{x})$ using $\boldsymbol{x}^0 = (9, 4)^T$

Example:

$$\min \quad f(\boldsymbol{x}) \triangleq (x_1 - 7)^2 + (x_2 - 2)^2$$

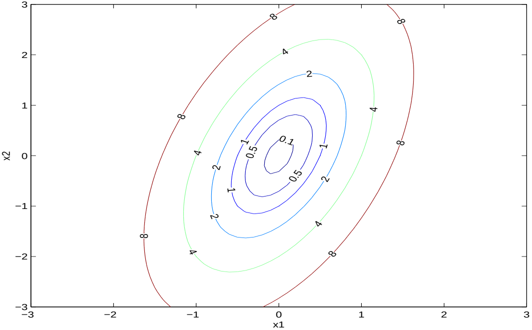


Behaviour of the steepest descent algorithm (with exact line search) applied to $f(\boldsymbol{x})$ using $\boldsymbol{x}^0 = (5.5, 3)^T$

Example:

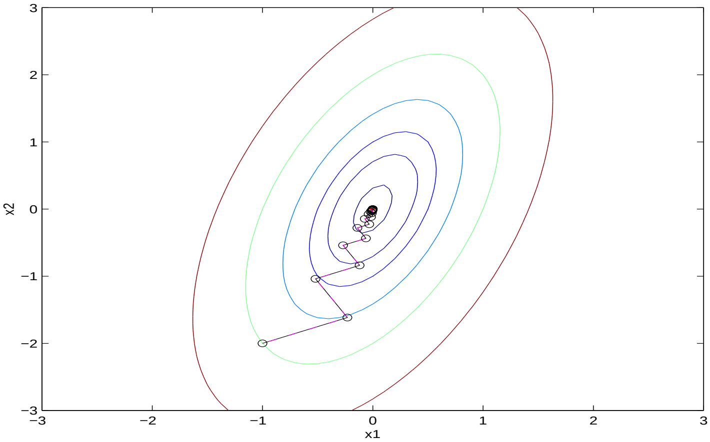
$$\min \ f(\boldsymbol{x}) \stackrel{\Delta}{=} 4x_1^2 + x_2^2 - 2x_1x_2$$

- $\boldsymbol{g}(\boldsymbol{x}) = \begin{pmatrix} 8x_1 - 2x_2 \\ 2x_2 - 2x_1 \end{pmatrix}$, $\boldsymbol{H}(\boldsymbol{x}) = \begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix}$.
- $\boldsymbol{x}^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$



Example:

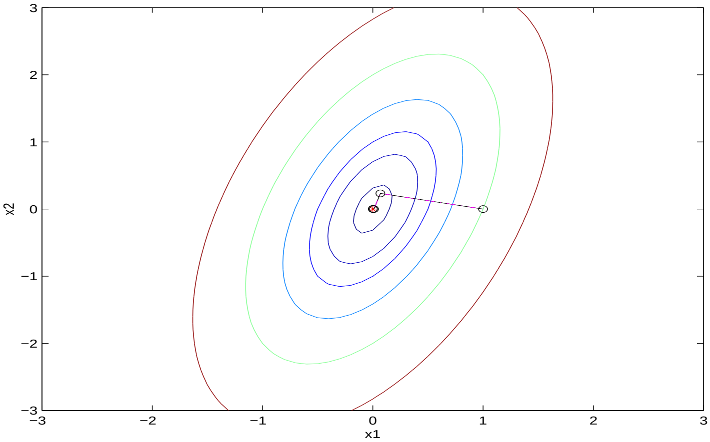
$$\min \quad f(\boldsymbol{x}) \stackrel{\Delta}{=} 4x_1^2 + x_2^2 - 2x_1x_2$$



Behaviour of the steepest descent algorithm (with exact line search) applied to $f(\boldsymbol{x})$ using $\boldsymbol{x}^0 = (-1, -2)^T$

Example:

$$\min \; f(\boldsymbol{x}) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$

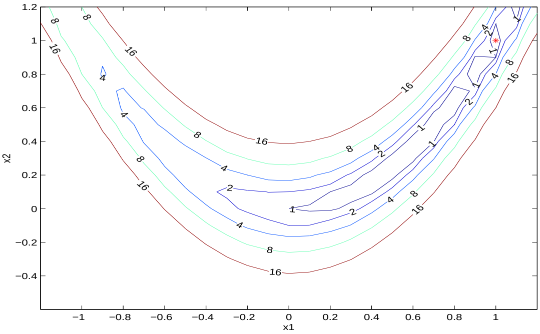


Behaviour of the steepest descent algorithm (with exact line search) applied to $f(\boldsymbol{x})$ using $\boldsymbol{x}^0 = (1,0)^T$

Example (Rosenbrock function):

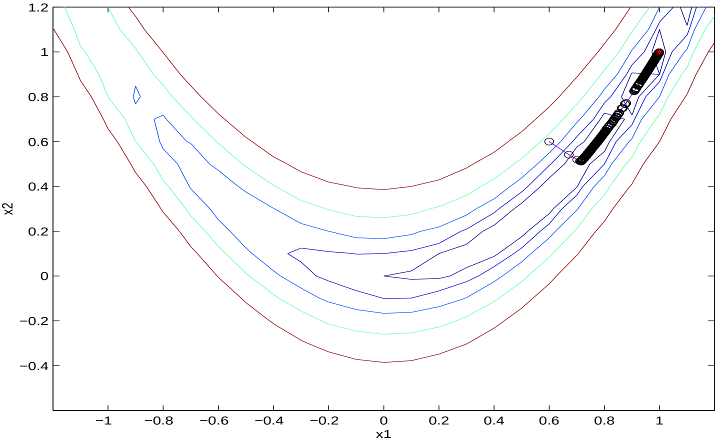
$$\min \quad f(\boldsymbol{x}) \triangleq 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

- $\boldsymbol{x}^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$



Example (Rosenbrock function):

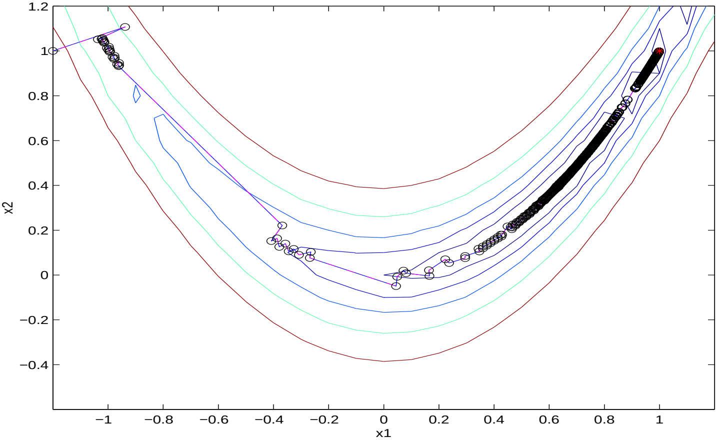
$$\min \;\; f(\boldsymbol{x}) \triangleq 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$



Behaviour of the steepest descent algorithm (with backtracking line search) applied to $f(\boldsymbol{x})$ using $\boldsymbol{x}^0 = (0.6, 0.6)^T$

Example (Rosenbrock function):

$$\min \quad f(\boldsymbol{x}) \triangleq 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$



Behaviour of the steepest descent algorithm (with backtracking line search) applied to $f(\boldsymbol{x})$ using $\boldsymbol{x}^0 = (-1.2, 1)^T$