7 - Solution methods for unconstrained optimization problems

Mauro Passacantando

Department of Computer Science, University of Pisa
mauro.passacantando@unipi.it

Optimization Methods and Game Theory
Master of Science in Artificial Intelligence and Data Engineering
University of Pisa – A.Y. 2020/21

Gradient method

Consider an unconstrained problem: $\min_{x \in \mathbb{R}^n} f(x)$.

Current point x^k, search direction $d^k = -\nabla f(x^k)$ (steepest descent direction)

Gradient method

Choose $x^0 \in \mathbb{R}^n$, set $k = 0$
while $\nabla f(x^k) \neq 0$ **do**
    [search direction] $d^k = -\nabla f(x^k)$
    [step size] compute an optimal solution t_k of the problem: $\min_{t>0} f(x^k + t\, d^k)$
    $x^{k+1} = x^k + t_k\, d^k$, $k = k + 1$
end

Example. $f(x) = x_1^2 + 2\, x_2^2 - 3\, x_1 - 2\, x_2$, starting point $x^0 = (2, 1)$.

$\nabla f(x^0) = (1, 2)$, $d^0 = (-1, -2)$, $f(x^0 + td^0) = 9\, t^2 - 5\, t - 2$, $t_0 = 5/18$,

$$x^1 = (2, 1) - \frac{5}{18}(1, 2) = \left(\frac{31}{18}, \frac{4}{9}\right).$$

Gradient method - step size

If $f(x) = \frac{1}{2}x^{\mathsf{T}}Qx + c^{\mathsf{T}}x$, with Q positive definite matrix, then

$$f(x^k + td^k) = \frac{1}{2}\,(d^k)^{\mathsf{T}}Q\,d^k\,t^2 + (g^k)^{\mathsf{T}}d^k\,t + f(x^k),$$

where $g^k = \nabla f(x^k) = Q\,x^k + c$. Thus the step size is equal to

$$t_k = -\frac{(g^k)^{\mathsf{T}}d^k}{(d^k)^{\mathsf{T}}Q\,d^k}.$$

Gradient method - convergence

Proposition

- $(d^k)^\mathsf{T} d^{k+1} = 0$ for any iteration k.
- If $\{x^k\}$ converges to x^*, then $\nabla f(x^*) = 0$, i.e. x^* is a stationary point of f.

Theorem

If f is coercive, then for any starting point x^0 the generated sequence $\{x^k\}$ is bounded and any of its cluster points is a stationary point of f.

Corollary

If f is coercive and convex, then for any starting point x^0 the generated sequence $\{x^k\}$ is bounded and any of its cluster points is a global minimum of f.

Corollary

If f is strongly convex, then for any starting point x^0 the generated sequence $\{x^k\}$ converges to the global minimum of f.

Gradient method - convergence rate

Two subsequent directions are orthogonal: $(d^k)^\mathsf{T} d^{k+1} = 0 \;\rightarrow\;$ the generated sequence has a zig-zag behaviour.

Theorem (Error bound)

If $f(x) = \frac{1}{2}\, x^\mathsf{T} Q\, x + c^\mathsf{T} x$, with Q positive definite matrix, and x^* is the global minimum of f, then the sequence $\{x^k\}$ satisfies the following inequality:

$$\|x^{k+1} - x^*\|_Q \leq \left(\frac{\dfrac{\lambda_n}{\lambda_1} - 1}{\dfrac{\lambda_n}{\lambda_1} + 1} \right) \|x^k - x^*\|_Q, \quad \forall\, k \geq 0, \quad \text{(linear convergence)}$$

where $\|x\|_Q = \sqrt{x^\mathsf{T} Q\, x}$ and $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of Q.

Remark. If λ_n/λ_1 (condition number of Q) is $>> 1$, then the ratio $\left(\frac{\frac{\lambda_n}{\lambda_1} - 1}{\frac{\lambda_n}{\lambda_1} + 1} \right) \simeq 1$ and the convergence may be slow.

Gradient method - convergence rate

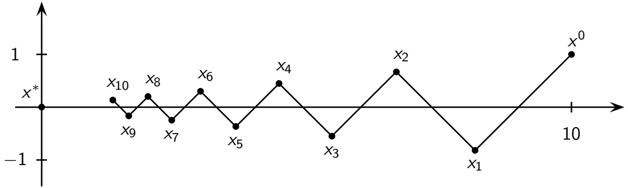
Example. $f(x) = x_1^2 + 10\,x_2^2$, global minimum is $x^* = (0,0)$.

If the starting point is $x^0 = (10,1)$, then the generated sequence is:

$$x^k = \left(10\,\left(\frac{9}{11}\right)^k, \left(-\frac{9}{11}\right)^k\right), \qquad \forall\ k \geq 0,$$

hence

$$\|x^{k+1} - x^*\| = \frac{9}{11}\,\|x^k - x^*\| \qquad \forall\ k \geq 0.$$



Gradient method - exercise

Exercise 7.1. Implement in MATLAB the gradient method for solving the problem

$$\begin{cases} \min \ \frac{1}{2} x^{\mathsf{T}} Q x + c^{\mathsf{T}} x \\ x \in \mathbb{R}^n \end{cases}$$

where Q is a positive definite matrix.

Solve the problem

$$\begin{cases} \min \ 3x_1^2 + 3x_2^2 + 3x_3^2 + 3x_4^2 - 4x_1x_3 - 4x_2x_4 + x_1 - x_2 + 2x_3 - 3x_4 \\ x \in \mathbb{R}^4 \end{cases}$$

starting from the point $(0, 0, 0, 0)$. [Use $\|\nabla f(x)\| < 10^{-6}$ as stopping criterion.]

Gradient method - Armijo inexact line search

When f is not a quadratic function, the exact line search may be computationally expensive.

Gradient method with the Armijo inexact line search

Set $\alpha, \gamma \in (0,1)$ and $\bar{t} > 0$. Choose $x^0 \in \mathbb{R}^n$, set $k = 0$.

while $\nabla f(x^k) \neq 0$ **do**
  $d^k = -\nabla f(x^k)$
  $t_k = \bar{t}$
  while $f(x^k + t_k\, d^k) > f(x^k) + \alpha\, t_k\, (d^k)^{\mathsf{T}} \nabla f(x^k)$ **do**
    $t_k = \gamma t_k$
  end
  $x^{k+1} = x^k + t_k d$, $k = k+1$
end

Theorem

If f is coercive, then for any starting point x^0 the generated sequence $\{x^k\}$ is bounded and any of its cluster points is a stationary point of f.

Gradient method - Armijo inexact line search

Example. Let $f(x_1, x_2) = x_1^4 + x_1^2 + x_2^2$. Set $\alpha = 10^{-4}$, $\gamma = 0.5$, $\bar{t} = 1$, choose $x^0 = (1, 1)$.
$d^0 = -\nabla f(x^0) = (-6, -2)$.
Line search. If $t_0 = 1$ then

$$f(x^0 + t_0 d^0) = 651 > f(x^0) + \alpha\, t_0 (d^0)^{\mathsf{T}} \nabla f(x^0) = 2.996,$$

if $t_0 = 0.5$ then

$$f(x^0 + t_0 d^0) = 20 > f(x^0) + \alpha\, t_0 (d^0)^{\mathsf{T}} \nabla f(x^0) = 2.998,$$

if $t_0 = 0.25$ then

$$f(x^0 + t_0 d^0) = 0.5625 < f(x^0) + \alpha\, t_0 (d^0)^{\mathsf{T}} \nabla f(x^0) = 2.999$$

hence the step size is $t_0 = 0.25$ and the new iterate is

$$x^1 = x^0 + t_0\, d^0 = (1, 1) + \frac{1}{4}(-6, -2) = \left(-\frac{1}{2}, \frac{1}{2}\right).$$

Gradient method - Armijo inexact line search

Exercise 7.2. Solve the problem

$$\begin{cases} \min\ 2\,x_1^4 + 3\,x_2^4 + 2\,x_1^2 + 4\,x_2^2 + x_1\,x_2 - 3\,x_1 - 2\,x_2 \\ x \in \mathbb{R}^2 \end{cases}$$

by means of the gradient method with the Armijo inexact line search setting $\alpha = 0.1$, $\gamma = 0.9$, $\bar{t} = 1$ and starting from the point $(0,0)$.
[Use $\|\nabla f(x)\| < 10^{-3}$ as stopping criterion.]

Conjugate gradient method

The search direction involves the gradient computed at the current iteration and the direction computed at the previous iteration.

First, consider the quadratic case:

$$f(x) = \frac{1}{2}\, x^{\mathsf{T}} Q\, x + c^{\mathsf{T}} x,$$

where Q is positive definite. Set $g = \nabla f(x) = Qx + c$.

At iteration k, the search direction is

$$d^k = \begin{cases} -g^0 & \text{if } k = 0, \\ -g^k + \beta_k\, d^{k-1} & \text{if } k \geq 1, \end{cases}$$

where β_k is such that d^k and d^{k-1} are conjugate with respect to Q, i.e.,

$$(d^k)^{\mathsf{T}} Q\, d^{k-1} = 0.$$

Conjugate gradient method

- It easy to compute β_k:

$$\beta_k = \frac{(g^k)^{\mathsf{T}} Q\, d^{k-1}}{(d^{k-1})^{\mathsf{T}} Q\, d^{k-1}}$$

- If we perform exact line search, then d^k is a descent direction
- The step size given by exact line search is $t_k = -\dfrac{(g^k)^{\mathsf{T}} d^k}{(d^k)^{\mathsf{T}} Q\, d^k}$

Conjugate gradient method for quadratic functions

Choose $x^0 \in \mathbb{R}^n$, set $g^0 = Q\, x^0 + c$, $k := 0$
while $g^k \neq 0$ **do**
  if $k = 0$ **then** $d^k = -g^k$
  else $\beta_k = \dfrac{(g^k)^{\mathsf{T}} Q\, d^{k-1}}{(d^{k-1})^{\mathsf{T}} Q\, d^{k-1}}, \quad d^k = -g^k + \beta_k\, d^{k-1}$
  end
  $t_k = -\dfrac{(g^k)^{\mathsf{T}} d^k}{(d^k)^{\mathsf{T}} Q\, d^k}$
  $x^{k+1} = x^k + t_k\, d^k$, $g^{k+1} = Q\, x^{k+1} + c$, $k = k + 1$
end

Conjugate gradient method

Example. Consider $f(x) = x_1^2 + 10\,x_2^2$, starting point $x^0 = (10, 1)$.

$k = 0$: $g^0 = (20, 20)$, $d^0 = -g^0 = (-20, -20)$,
$t_0 = -((g^0)^{\mathsf{T}} d^0)/((d^0)^{\mathsf{T}} Q\, d^0) = 1/11$, hence $x^1 = x^0 + t_0\, d^0 = (90/11, -9/11)$

$k = 1$: $g^1 = (180/11, -180/11)$, $\beta_1 = ((g^1)^{\mathsf{T}} Q\, d^0)/((d^0)^{\mathsf{T}} Q\, d^0) = 81/121$,
$d^1 = -g^1 + \beta_1\, d^0 = (-3600/121, 360/121)$,
$t_1 = -((g^1)^{\mathsf{T}} d^1)/((d^1)^{\mathsf{T}} Q\, d^1) = 11/40$, hence $x^2 = x^1 + t_1\, d^1 = (0, 0)$ which is the global minimum of f.

Conjugate gradient method - convergence

Proposition

- An alternative formula for the step size is $t_k = \dfrac{\|g^k\|^2}{(d^k)^\mathsf{T} Q\, d^k}$
- An alternative formula for β_k is $\beta_k = \dfrac{\|g^k\|^2}{\|g^{k-1}\|^2}$
- If we did not find the global minimum after k iterations, then the gradients $\{g^0, g^1, \dots, g^k\}$ are orthogonal
- If we did not find the global minimum after k iterations, then the directions $\{d^0, d^1, \dots, d^k\}$ are conjugate w.r.t. Q and x^k is the minimum of f on $x^0 + \mathrm{Span}(d^0, d^1, \dots, d^k)$

Theorem (Convergence)

- The CG method finds the global minimum in at most n iterations.
- If Q has r distinct eigenvalues, then CG method finds the global minimum in at most r iterations.

Conjugate gradient method - convergence rate

Theorem (Error bound)

If $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of Q, then the following bounds hold:

- $$\|x^k - x^*\|_Q \leq 2 \left(\frac{\sqrt{\dfrac{\lambda_n}{\lambda_1}} - 1}{\sqrt{\dfrac{\lambda_n}{\lambda_1}} + 1} \right)^k \|x^0 - x^*\|_Q, \qquad \forall\ k \geq 0,$$

- $$\|x^k - x^*\|_Q \leq \left(\frac{\lambda_{n-k+1} - \lambda_1}{\lambda_{n-k+1} + \lambda_1} \right) \|x^0 - x^*\|_Q, \qquad \forall\ k \geq 0.$$

Conjugate gradient method - exercise

Exercise 7.3. Implement in MATLAB the conjugate gradient method for solving the problem

$$\begin{cases} \min \ \frac{1}{2} x^{\mathsf{T}} Q x + c^{\mathsf{T}} x \\ x \in \mathbb{R}^n \end{cases}$$

where Q is a positive definite matrix.
Solve the problem

$$\begin{cases} \min \ 3x_1^2 + 3x_2^2 + 3x_3^2 + 3x_4^2 - 4x_1x_3 - 4x_2x_4 + x_1 - x_2 + 2x_3 - 3x_4 \\ x \in \mathbb{R}^4 \end{cases}$$

starting from the point $(0, 0, 0, 0)$. [Use $\|\nabla f(x)\| < 10^{-6}$ as stopping criterion.]

Conjugate gradient method - nonlinear functions

Conjugate gradient method for general nonlinear functions

Choose $x^0 \in \mathbb{R}^n$, set $k := 0$

while $\nabla f(x^k) \neq 0$ **do**

  if $k = 0$ **then** $d^k = -\nabla f(x^k)$

  else $\beta_k = \dfrac{\|\nabla f(x^k)\|^2}{\|\nabla f(x^{k-1})\|^2}, \quad d^k = -\nabla f(x^k) + \beta_k\, d^{k-1}$

  end

  Compute the step size t_k

  $x^{k+1} = x^k + t_k\, d^k$, $k = k + 1$

end

Conjugate gradient method - nonlinear functions

Proposition

- If t_k is computed by exact line search, then d^k is a descent direction.
- If t_k satisfies the following conditions:

$$\begin{cases} f(x^k + t_k\, d^k) \leq f(x^k) + \alpha\, t_k\, \nabla f(x^k)^\mathsf{T} d^k, \\ |\nabla f(x^k + t_k\, d^k)^\mathsf{T} d^k| \leq -\beta\, \nabla f(x^k)^\mathsf{T} d^k, \end{cases} \tag{1}$$

with $0 < \alpha < \beta < 1/2$, then d^k is a descent direction.

Theorem

If f is coercive, then the conjugate gradient method, where t_k satisfies conditions (1), generates a sequence $\{x^k\}$ such that

$$\liminf_{k\to\infty} \|\nabla f(x^k)\| = 0.$$

Newton method – basic version

We want to find a stationary point $\nabla f(x) = 0$.

At iteration k, make a linear approximation of $\nabla f(x)$ at x^k, i.e.

$$\nabla f(x) \simeq \nabla f(x^k) + \nabla^2 f(x^k)(x - x^k),$$

the new iterate x^{k+1} is the solution of the linear system

$$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0.$$

Note that x^{k+1} is a stationary point of the quadratic approximation of f at x^k:

$$f(x) \simeq f(x^k) + (x - x^k)^\mathsf{T} \nabla f(x^k) + \frac{1}{2}\,(x - x^k)^\mathsf{T} \nabla^2 f(x^k)\,(x - x^k).$$

Newton method – basic version

Newton method (basic version)

Choose $x^0 \in \mathbb{R}^n$, set $k = 0$
while $\nabla f(x^k) \neq 0$ **do**
    Solve the linear system $\nabla^2 f(x^k) d^k = -\nabla f(x^k)$
    $x^{k+1} = x^k + d^k$, $k = k + 1$
end

Theorem (Convergence)

If x^* is a local minimum of f and $\nabla^2 f(x^*)$ is positive definite, then there exists $\delta > 0$ such that for any $x^0 \in B(x^*, \delta)$ the sequence $\{x^k\}$ converges to x^* and

$$\|x^{k+1} - x^*\| \leq C\, \|x^k - x^*\|^2 \qquad \forall\ k > \bar{k}, \qquad \text{(quadratic convergence)}$$

for some $C > 0$ and $\bar{k} > 0$.

Newton method – basic version

Example. $f(x) = 2\,x_1^4 + 3\,x_2^4 + 2\,x_1^2 + 4\,x_2^2 + x_1\,x_2 - 3\,x_1 - 2\,x_2$ is strongly convex because

$$\nabla^2 f(x) = \begin{pmatrix} 24\,x_1^2 + 4 & 1 \\ 1 & 36\,x_2^2 + 8 \end{pmatrix}.$$

k	x^k		$\|\nabla f(x^k)\|$
0	10.000000	5.000000	8189.6317378
1	6.655450	3.298838	2429.6437291
2	4.421132	2.149158	721.6330686
3	2.925965	1.361690	214.6381594
4	1.923841	0.811659	63.7752575
5	1.255001	0.428109	18.6170045
6	0.823359	0.209601	5.0058040
7	0.580141	0.171251	1.0538969
8	0.492175	0.179815	0.1022945
9	0.481639	0.180914	0.0013018
10	0.481502	0.180928	0.0000002

Newton method – basic version

Drawbacks of Newton method:

- at each iteration we need to compute both the gradient $\nabla f(x^k)$ and the hessian matrix $\nabla^2 f(x^k)$
- local convergence: if x^0 is too far from the optimum x^*, then the generated sequence can be not convergent to x^*

Example. Let $f(x) = -\frac{1}{16}x^4 + \frac{5}{8}x^2$.
Then $f'(x) = -\frac{1}{4}x^3 + \frac{5}{4}x$ and $f''(x) = -\frac{3}{4}x^2 + \frac{5}{4}$.
$x^* = 0$ is a local minimum of f with $f''(x^*) = 5/4 > 0$.
The sequence does not converge to x^* if it starts from $x^0 = 1$:
$x^1 = -1$, $x^2 = 1$, $x^3 = -1$, ...

Newton method with line search

If f is strongly convex, we get global convergence because d^k is a descent direction: $\nabla f(x^k)^\mathsf{T} d^k = -\nabla f(x^k)^\mathsf{T} [\nabla^2 f(x^k)]^{-1} \nabla f(x^k) < 0$.

Newton method with line search

Set $\alpha, \gamma \in (0,1)$, $\bar{t} > 0$. Choose $x^0 \in \mathbb{R}^n$, set $k = 0$
while $\nabla f(x^k) \neq 0$ **do**
    [search direction] Solve the linear system $\nabla^2 f(x^k) d^k = -\nabla f(x^k)$
    $t_k = \bar{t}$
    while $f(x^k + t_k\, d^k) > f(x^k) + \alpha\, t_k\, (d^k)^\mathsf{T} \nabla f(x^k)$ **do**
        $t_k = \gamma t_k$
    end
    $x^{k+1} = x^k + t_k\, d^k$, $k = k + 1$
end

Theorem (Convergence)

If f is strongly convex, then for any starting point $x^0 \in \mathbb{R}^n$ the sequence $\{x^k\}$ converges to the global minimum of f. Moreover, if $\alpha \in (0, 1/2)$ and $\bar{t} = 1$ then the convergence is quadratic.

Newton method with line search

Exercise 7.4. Solve the problem

$$\begin{cases} \min \ 2\,x_1^4 + 3\,x_2^4 + 2\,x_1^2 + 4\,x_2^2 + x_1\,x_2 - 3\,x_1 - 2\,x_2 \\ x \in \mathbb{R}^2 \end{cases}$$

by means of the Newton method with line search setting $\alpha = 0.1$, $\gamma = 0.9$, $\bar{t} = 1$ and starting from the point $(0, 0)$. [Use $\|\nabla f(x)\| < 10^{-3}$ as stopping criterion.]

Quasi-Newton methods

At each iteration $[\nabla^2 f(x^k)]^{-1}$ is approximated by a positive definite matrix H_k

Quasi-Newton method

Choose $x^0 \in \mathbb{R}^n$, a positive definite matrix H_0, $k = 0$

while $\nabla f(x^k) \neq 0$ **do**

$d^k = -H_k \nabla f(x^k)$

Compute step size t_k

$x^{k+1} = x^k + t_k d^k$

update H_{k+1}

$k = k + 1$

end

How to update matrix H_k?

Quasi-Newton methods

$$\nabla f(x^k) \simeq \nabla f(x^{k+1}) + \nabla^2 f(x^{k+1})\,(x^k - x^{k+1}).$$

Set $p^k = x^{k+1} - x^k$ and $g^k = \nabla f(x^{k+1}) - \nabla f(x^k)$, then

$$\nabla^2 f(x^{k+1})\,p^k \simeq g^k, \text{ i.e. } [\nabla^2 f(x^{k+1})]^{-1}\,g^k \simeq p^k.$$

We choose H_{k+1} such that

$$H_{k+1}\,g^k = p^k.$$

Davidon-Fletcher-Powell (DFP) method:

$$H_{k+1} = H_k + \frac{p^k\,(p^k)^{\mathsf T}}{(p^k)^{\mathsf T} g^k} - \frac{H_k\,g^k (g^k)^{\mathsf T}\,H_k}{(g^k)^{\mathsf T}\,H_k\,g^k},$$

Quasi-Newton methods

Another approach: find a matrix $B_k = (H_k)^{-1}$ approximating $\nabla^2 f(x^k)$.

Since $\nabla^2 f(x^{k+1})\, p^k \simeq g^k$, we impose that $B_{k+1}\, p^k = g^k$

Update B_k as

$$B_{k+1} = B_k + \frac{g^k (g^k)^\mathsf{T}}{(p^k)^\mathsf{T} g^k} - \frac{B_k\, p^k (p^k)^\mathsf{T} B_k}{(p^k)^\mathsf{T} B_k\, p^k},$$

hence

$$H_{k+1} = H_k + \left(1 + \frac{(g^k)^\mathsf{T} H_k\, g^k}{(p^k)^\mathsf{T} g^k}\right) \frac{p^k (p^k)^\mathsf{T}}{(p^k)^\mathsf{T} g^k} - \frac{p^k (g^k)^\mathsf{T} H_k + H_k\, g^k (p^k)^\mathsf{T}}{(p^k)^\mathsf{T} g^k}.$$

(Broyden–Fletcher–Goldfarb–Shanno (BFGS) method).

Derivative-free methods

There are situations where derivatives of the objective function do not exist or are computationally expensive.

Derivative-free methods sample the objective function at a finite number of points at each iteration, without any explicit or implicit derivative approximation.

Definition

A positive basis is a set of vectors $\{v^1, \dots, v^p\} \subset \mathbb{R}^n$ such that:

- any $x \in \mathbb{R}^n$ is a conic combination of $v^1, \dots, v^p$, i.e., there exist $\alpha_1, \dots, \alpha_p \geq 0$ such that $x = \sum_{i=1}^{p} \alpha_i v^i$
- for any $i = 1, \dots, p$, v^i is not a conic combination of others $v^1, \dots, v^p$.

Examples: $\{e_1, \dots, e_n, -e_1, \dots, -e_n\}$ is a positive basis of $\mathbb{R}^n$; $\{(1,0), (0,1), (-1,-1)\}$ is a positive basis of $\mathbb{R}^2$.

Proposition. If $\{v^1, \dots, v^p\}$ is a positive basis, then for any $w \in \mathbb{R}^n \setminus \{0\}$ there is $i \in \{1, \dots, p\}$ such that $w^T v^i < 0$.

Directional direct-search method

Directional direct-search method

Choose starting point $x^0 \in \mathbb{R}^n$, step size $t_0 > 0$, $\beta \in (0,1)$, tolerance $\varepsilon > 0$ and a positive basis D. Set $k = 0$.

while $t_k > \varepsilon$ **do**
  Order the poll set $\{x^k + t_k d, \quad d \in D\}$
  Evaluate f at the poll points following the chosen order
  If there is a poll point s.t. $f(x^k + t_k d) < f(x^k)$
    then $x^{k+1} = x^k + t_k d$, $t_{k+1} = t_k$ (successful iteration)
  else $x^{k+1} = x^k$, $t_{k+1} = \beta t_k$ (step size reduction)
  end
  $k = k + 1$
end

The method is called coordinate-search method if $D = \{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}$.

Directional direct-search method

Theorem

Assume that all the vectors of the positive basis D are in $\mathbb{Z}^n$. If f is coercive and continuously differentiable, then the generated sequence $\{x^k\}$ has a cluster point x^* such that $\nabla f(x^*) = 0$.

Remark 1. The assumption that vectors of D are in $\mathbb{Z}^n$ can be deleted if we accept new iterates which satisfy a "sufficient" decrease condition:

$$f(x^{k+1}) \leq f(x^k) - t_k^2.$$

Remark 2. If a complete poll step is performed, i.e.,

$$f(x^{k+1}) \leq f(x^k + t_k d) \qquad \forall\ d \in D,$$

then *any* cluster point of $\{x^k\}$ is a stationary point of f and $\lim_{k\to\infty} \|\nabla f(x^k)\| = 0$.

Variants of the Directional direct-search method

Directional direct-search method

Choose starting point $x^0 \in \mathbb{R}^n$, step size $t_0 > 0$, $\beta \in (0,1)$, tolerance $\varepsilon > 0$, $\gamma \geq 1$ and a set of positive bases $\mathcal{D}$. Set $k = 0$.

while $t_k > \varepsilon$ **do**

    Choose a positive basis $D \in \mathcal{D}$
    Order the poll set $\{x^k + t_k d, \quad d \in D\}$
    Evaluate f at the poll points following the chosen order
    If there is a poll point s.t. $f(x^k + t_k d) < f(x^k)$
        then $x^{k+1} = x^k + t_k d$, $t_{k+1} = \gamma t_k$ (successful iteration)
    else $x^{k+1} = x^k$, $t_{k+1} = \beta t_k$ (step size reduction)
    end
    $k = k + 1$

end

Directional direct-search method - Exercise

Exercise 7.5.

a) Solve the problem

$$\begin{cases} \min \ 2\,x_1^4 + 3\,x_2^4 + 2\,x_1^2 + 4\,x_2^2 + x_1\,x_2 - 3\,x_1 - 2\,x_2 \\ x \in \mathbb{R}^2 \end{cases}$$

by means of the directional direct-search method setting $x^0 = (0,0)$, $t_0 = 5$, $\beta = 0.5$, $\varepsilon = 10^{-5}$ and the positive basis $D = \{(1,0),(0,1),(-1,0),(0,-1)\}$.

b) Solve the previous problem by means of the directional direct-search method setting $x^0 = (0,0)$, $t_0 = 5$, $\beta = 0.5$, $\varepsilon = 10^{-5}$ and the positive basis $D = \{(1,0),(0,1),(-1,-1)\}$.