Non-linear optimization
Exercises

Giovanni Righini

1 Mono-dimensional unconstrained optimization

Approximate the local minimum of $f(x) = 4x^3 + x^2 - 10x + 1$ which falls in the range [0..2]. A precision $\Delta = 0.1$ is required.
To facilitate the computations, assume the resolution is $\epsilon = 0$ until the very last iteration.

Fibonacci numbers method. We look for a Fibonacci number at least equal to $\frac{I_0}{\Delta}$, that is such that $F(n+2) \geq 20$: $F(8) = 21$ satisfy this condition. Therefore we perform $n = 6$ iterations to obtain a final range of width $I_6 = I_0/F(8) = \frac{2}{21}$.

Then we start with

$$I_0 = F(8)I_6 = 21\frac{2}{21} = 2$$

$$I_1 = F(7)I_6 = 13\frac{2}{21} = \frac{26}{21}$$

and then

$$I_2 = F(6)I_6 = 8\frac{2}{21} = \frac{16}{21}$$

$$I_3 = F(5)I_6 = 5\frac{2}{21} = \frac{10}{21}$$

$$I_4 = F(4)I_6 = 3\frac{2}{21} = \frac{6}{21}$$

$$I_5 = F(3)I_6 = 2\frac{2}{21} = \frac{4}{21}$$

and finally

$$I_6 = \frac{2}{21}$$

At the extreme points of the initial range [0..2] we have $f(0) = 1$.
$f(2) = 17$.

The first two internal evaluation points provide the values
$f(\frac{26}{21}) = -\frac{20899}{9261} \approx -2.25667$.
$f(\frac{16}{21}) = -\frac{39539}{9261} \approx -4.26941$.
Therefore we discard the sub-range $[\frac{26}{21}..2]$.
We are left with the range $[0..\frac{26}{21}]$ with width $I_1 = \frac{26}{21}$, including an evaluated point at $x = \frac{16}{21}$.
The point symmetric of $x = \frac{16}{21}$ in the range $[0..\frac{26}{21}]$ is the point $x = \frac{10}{21}$.

The evaluation of $f(x)$ in $x = \frac{10}{21}$ provides the value
$f(\frac{10}{21}) = -\frac{28739}{9261} \approx -3.10323$.
Therefore we discard the sub-range $[0..\frac{10}{21}]$.
We are left with the range $[\frac{10}{21}..\frac{26}{21}]$ with width $I_2 = \frac{16}{21}$, including an evaluated point at $x = \frac{16}{21}$.

The point symmetric of $x = \frac{16}{21}$ in the range $[\frac{10}{21}..\frac{26}{21}]$ is the point $x = \frac{20}{21}$.

The evaluation of $f(x)$ in $x = \frac{20}{21}$ provides the value
$f(\frac{20}{21}) = -\frac{38539}{9261} \approx -4.16143$.
Therefore we discard the sub-range $[\frac{20}{21}..\frac{26}{21}]$.
We are left with the range $[\frac{10}{21}..\frac{20}{21}]$ with width $I_3 = \frac{10}{21}$, including an evaluated point at $x = \frac{16}{21}$.
The point symmetric of $x = \frac{16}{21}$ in the range $[\frac{10}{21}..\frac{20}{21}]$ is the point $x = \frac{14}{21}$.

The evaluation of $f(x)$ in $x = \frac{14}{21}$ provides the value
$f(\frac{14}{21}) = -\frac{109}{27} \approx -4.03704$.
Therefore we discard the sub-range $[\frac{10}{21}..\frac{14}{21}]$.
We are left with the range $[\frac{14}{21}..\frac{20}{21}]$ with width $I_4 = \frac{6}{21}$, including an evaluated point at $x = \frac{16}{21}$.
The point symmetric of $x = \frac{16}{21}$ in the range $[\frac{14}{21}..\frac{20}{21}]$ is the point $x = \frac{18}{21}$.

The evaluation of $f(x)$ in $x = \frac{18}{21}$ provides the value
$f(\frac{18}{21}) = -\frac{1481}{343} \approx -4.31778$.
Therefore we discard the sub-range $[\frac{14}{21}..\frac{16}{21}]$.
We are left with the range $[\frac{16}{21}..\frac{20}{21}]$ with width $I_5 = \frac{4}{21}$, including an evaluated point at $x = \frac{18}{21}$.
The point symmetric of $x = \frac{18}{21}$ in the range $[\frac{16}{21}..\frac{20}{21}]$ is the point itself. Under the assumption $\epsilon = 0$, we cannot discard an interval at the last iteration. Hence let assume that $\epsilon > 0$.

The evaluation of $f(x)$ in $x = \frac{18}{21} + \epsilon$ provides the value
$f(\frac{18}{21} + \epsilon) = -\frac{1491}{343} + \frac{26\epsilon+79\epsilon^2+4\epsilon^3}{343} > -\frac{1491}{343}$.
Therefore we discard the sub-range $[\frac{18}{21}..\frac{20}{21}]$.
We are left with the range $[\frac{16}{21}..\frac{18}{21}]$ with width $I_6 = \frac{2}{21}$. This final range does not include any already evaluated internal point and its width meets the requirement (it is not larger than Δ).

The final uncertainty range is $[\frac{16}{21}..\frac{18}{21}] \approx [0.76190..0.85714]$.
The minimum value found for $f(x)$ is $f(\frac{18}{21}) \approx -4.31778$.

Bisection method. Assume we know the first derivative $f'(x) = 12x^2 + 2x - 10$. We verify that $f'(0) = -10$ is negative and $f'(2) = 42$ is positive. So, there must be a minimum between these two extreme points.

The initial uncertainty is 2. To reduce it to no more than $\Delta = 0.1$ we need to halve it at least $n = \lceil \log_2 \frac{2}{0.1} \rceil = 5$. Hence we need 5 iterations.
We evaluate $f'(x)$ in the midpoint of the current range $[0..2]$: $f'(1) = 4$ is positive. Therefore we discard the range $[1..2]$ and we keep the range $[0..1]$.

We evaluate $f'(x)$ in the midpoint of the current range $[0..1]$: $f'(\frac{1}{2}) = -6$ is negative. Therefore we discard the range $[0..\frac{1}{2}]$ and we keep the range $[\frac{1}{2}..1]$.

We evaluate $f'(x)$ in the midpoint of the current range $[\frac{1}{2}..1]$: $f'(\frac{3}{4}) = -\frac{7}{4}$ is negative. Therefore we discard the range $[\frac{1}{2}..\frac{3}{4}]$ and we keep the range $[\frac{3}{4}..1]$.

We evaluate $f'(x)$ in the midpoint of the current range $[\frac{3}{4}..1]$: $f'(\frac{7}{8}) = \frac{15}{16}$ is positive. Therefore we discard the range $[\frac{7}{8}..1]$ and we keep the range $[\frac{3}{4}..\frac{7}{8}]$.

We evaluate $f'(x)$ in the midpoint of the current range $[\frac{3}{4}..\frac{7}{8}]$: $f'(\frac{13}{16}) = \frac{547}{64}$ is positive. Therefore we discard the range $[\frac{13}{16}..\frac{7}{8}]$ and we keep the range $[\frac{3}{4}..\frac{13}{16}]$.

The final uncertainty range is $[\frac{3}{4}..\frac{13}{16}]$, i.e. $[0.7500..0.8125]$.
The evaluated point with the smallest value of the first derivative is $x = \frac{13}{16}$, where the value of $f(x)$ is $f(\frac{13}{16}) = -\frac{17692}{4096} \approx -4.31934$.

Newton method. To apply Newton method we must know the first and the second derivative of $f(x)$:
$f'(x) = 12x^2 + 2x - 10$.
$f''(x) = 24x + 2$.
We must also verify that $f''(x)$ is always positive in the range of interest, that is $[0..2]$. The condition holds.

Let start from the midpoint $x^{(0)} = 1$.

We evaluate the derivatives in $x^{(0)}$: $f'(1) = 4$; $f''(1) = 26$.
The next point is $x^{(1)} = x^{(0)} - \frac{f'(x^{(0)})}{f''(x^{(0)})} = 1 - \frac{4}{26} = \frac{11}{13} \approx 0.84615$.

We evaluate the derivatives in $x^{(1)}$: $f'(\frac{11}{13}) = \frac{48}{169}$; $f''(\frac{11}{13}) = \frac{290}{13}$.
The next point is $x^{(2)} = x^{(1)} - \frac{f'(x^{(1)})}{f''(x^{(1)})} = \frac{11}{13} - \frac{48\ 13}{169\ 290} = \frac{1571}{1885} \approx 0.83342$.

We evaluate the derivatives in $x^{(2)}$: $f'(\frac{1571}{1885}) = \frac{6912}{3553225}$; $f''(\frac{1571}{1885}) = \frac{41474}{1885}$.
The next point is $x^{(3)} = x^{(2)} - \frac{f'(x^{(2)})}{f''(x^{(2)})} = \frac{1571}{1885} - \frac{6912\ 1885}{3553225\ 41474} \approx 0.83351$.

The step is very small (much smaller than Δ); so we can stop. The final approximated minimum is $x^{(3)} \approx 0.83351$ and the corresponding value of the function is $f(x^{(3)}) \approx -4.32407$.

2 Multi-dimensional unconstrained optimization

Find the minimum of the function $f(x) = 3x_1^2 + 2x_1x_2 + 3x_2^2 - 14x_1 - 10x_2$ with the gradient method, starting from $x^{(0)} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$.

The gradient vector has the following general expression:

$$\nabla f(x) = \begin{bmatrix} 6x_1 + 2x_2 - 14 \\ 2x_1 + 6x_2 - 10 \end{bmatrix}.$$

Iteration 1. The gradient evaluated in $x^{(0)} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$ is

$$\nabla f(x^{(0)}) = \begin{bmatrix} 0 \\ -16 \end{bmatrix}.$$

Choosing the direction opposite to it and normalizing the components to have unit norm, we obtain

$$d^{(1)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Depending on the step α_1, the next point is

$$x^{(1)} = x^{(0)} + \alpha_1 d^{(1)} = \begin{bmatrix} 3 \\ -2 \end{bmatrix} + \alpha_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 + \alpha_1 \end{bmatrix}.$$

To compute the optimal step, we must find the minimum of the function

$$f(\alpha_1) = 27 + 6(-2 + \alpha_1) + 3(-2 + \alpha_1)^2 - 42 - 10(-2 + \alpha_1) = 3\alpha_1^2 - 16\alpha_1 + 5.$$

To find the minimum, we compute its first derivative with respect to α_1:

$$\frac{\partial f}{\partial \alpha_1} = 6\alpha_1 - 16$$

which is null for $\alpha_1 = \frac{8}{3}$. This is a minimum because the second derivative is equal to 6, which is positive. Therefore the optimal step is

$$\alpha_1 = \frac{8}{3}.$$

Hence we have

$$x^{(1)} = \begin{bmatrix} 3 \\ -2 + \frac{8}{3} \end{bmatrix} = \begin{bmatrix} 3 \\ \frac{2}{3} \end{bmatrix}.$$

Iteration 2. The gradient evaluated in $x^{(1)} = \begin{bmatrix} 3 \\ \frac{2}{3} \end{bmatrix}$ is

$$\nabla f(x^{(1)}) = \begin{bmatrix} \frac{16}{3} \\ 0 \end{bmatrix}.$$

Choosing the direction opposite to it and normalizing the components to have unit norm, we obtain

$$d^{(2)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

Depending on the step α_2, the next point is

$$x^{(2)} = x^{(1)} + \alpha_2 d^{(2)} = \begin{bmatrix} 3 \\ \frac{2}{3} \end{bmatrix} + \alpha_2 \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3-\alpha_2 \\ \frac{2}{3} \end{bmatrix}.$$

To compute the optimal step, we must find the minimum of the function

$$f(\alpha_2) = 3(3-\alpha_2)^2 + \frac{4}{3}(3-\alpha_2) + \frac{4}{3} - 14(3-\alpha_2) - \frac{20}{3} = 3\alpha_2^2 - \frac{16}{3}\alpha_2 - \frac{49}{3}.$$

To find the minimum, we compute its first derivative with respect to α_2:

$$\frac{\partial f}{\partial \alpha_2} = 6\alpha_2 - \frac{16}{3}$$

which is null for $\alpha_2 = \frac{8}{9}$. This is a minimum because the second derivative is equal to 6, which is positive. Therefore the optimal step is

$$\alpha_2 = \frac{8}{9}.$$

Hence we have

$$x^{(2)} = \begin{bmatrix} 3-\frac{8}{9} \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{19}{9} \\ \frac{2}{3} \end{bmatrix}.$$

Iteration 3. The gradient evaluated in $x^{(2)} = \begin{bmatrix} \frac{19}{9} \\ \frac{2}{3} \end{bmatrix}$ is

$$\nabla f(x^{(2)}) = \begin{bmatrix} 0 \\ -\frac{16}{9} \end{bmatrix}.$$

Choosing the direction opposite to it and normalizing the components to have unit norm, we obtain

$$d^{(3)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Depending on the step α_3, the next point is

$$x^{(3)} = x^{(2)} + \alpha_3 d^{(3)} = \begin{bmatrix} \frac{19}{9} \\ \frac{2}{3} \end{bmatrix} + \alpha_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{19}{9} \\ \frac{2}{3}+\alpha_3 \end{bmatrix}.$$

To compute the optimal step, we must find the minimum of the function

$$f(\alpha_3) = 3(\frac{19}{9})^2 + \frac{38}{9}(\frac{2}{3}+\alpha_3) + 3(\frac{2}{3}+\alpha_3)^2 - 14\frac{19}{9} - 10(\frac{2}{3}+\alpha_3) = 3\alpha_3^2 - \frac{16}{9}\alpha_3 + \ldots.$$

To find the minimum, we compute its first derivative with respect to α_3:

$$\frac{\partial f}{\partial \alpha_3} = 6\alpha_3 - \frac{16}{9}$$

which is null for $\alpha_3 = \frac{8}{27}$. This is a minimum because the second derivative is equal to 6, which is positive. Therefore the optimal step is

$$\alpha_3 = \frac{8}{27}.$$

Hence we have

$$x^{(3)} = \begin{bmatrix} \frac{19}{9} \\ \frac{2}{3} + \frac{8}{27} \end{bmatrix} = \begin{bmatrix} \frac{19}{9} \\ \frac{26}{27} \end{bmatrix}.$$

The algorithm asymptotically converges towards the minimum at $x^* = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

3 Multi-dimensional constrained optimization

Find the minimum of the function $f(x) = -(x_1 - 4)^2 - (x_2 - 3)^2$ subject to the constraints

$$\begin{cases} x_1 + x_2 - 6 \leq 0 \\ -x_1 \leq 0 \\ -x_2 \leq 0 \end{cases}$$

with the feasible directions method, starting from $x^{(0)} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$.

The gradient of $f(x)$ has the following form:

$$\nabla f(x) = \begin{bmatrix} -2x_1 + 8 \\ -2x_2 + 6 \end{bmatrix}.$$

Iteration 1. In $x^{(0)} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ we have

$$\nabla f(x^{(0)}) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

and an empty set of active constraints

$$\Omega(x^{(0)}) = \emptyset.$$

Since we want to minimize $f(x)$, the best direction is the one that minimizes the scalar product with the gradient vector $\nabla f(x)$ at the current point. All directions are feasible because no constraints are active. So, we solve the sub-problem

$$\begin{aligned} \text{minimize } \nabla f(x^{(0)}) d^{(1)} =& 2d_1^{(1)} + 2d_2^{(1)} \\ \text{s.t. } & (d_1^{(1)})^2 + (d_2^{(1)})^2 = 1 \end{aligned}$$

The optimal solution is obviously

$$d^{(1)} = \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}.$$

The next point $x^{(1)}$ is a function of the step α_1 to be taken along direction $d^{(1)}$:

$$x^{(1)} = x^{(0)} + \alpha_1 d^{(1)} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} + \alpha_1 \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 3 - \frac{\sqrt{2}}{2}\alpha_1 \\ 2 - \frac{\sqrt{2}}{2}\alpha_1 \end{bmatrix}.$$

The value of $f(x)$ in the new point $x^{(1)}$ is a function of α_1:

$$f(x^{(1)}) = -(-1 - \frac{\sqrt{2}}{2}\alpha_1)^2 - (-1 - \frac{\sqrt{2}}{2}\alpha_1)^2 = -\alpha_1^2 - 2\sqrt{2}\alpha_1 - 2$$

and the following constraints must be satisfied

$$\begin{cases} (3 - \frac{\sqrt{2}}{2}\alpha_1) + (2 - \frac{\sqrt{2}}{2}\alpha_1) - 6 \leq 0 \\ -3 + \frac{\sqrt{2}}{2}\alpha_1 \leq 0 \\ -2 + \frac{\sqrt{2}}{2}\alpha_1 \leq 0 \end{cases}$$

To find the optimal step $\alpha_1 \geq 0$ we have to solve the sub-problem:

$$\begin{aligned} \text{minimize } f(\alpha_1) = & -\alpha_1^2 - 2\sqrt{2}\alpha_1 - 2 \\ \text{s.t. } & \alpha_1 \geq -\frac{\sqrt{2}}{2} \quad \text{(redundant)} \\ & \alpha_1 \leq 3\sqrt{2} \\ & \alpha_1 \leq 2\sqrt{2} \\ & \alpha_1 \geq 0 \end{aligned}$$

The first derivative of $f(\alpha_1)$ is

$$\frac{\partial f}{\partial \alpha_1} = -2\alpha_1 - 2\sqrt{2}$$

which is never null for $\alpha_1 \geq 0$. Therefore no minimum is reached by moving along $d^{(1)}$; the step is only limited by the constraints. The first constraint that becomes active is the third one: $\alpha_1 \leq 2\sqrt{2}$. This is the binding constraint. Therefore we have

$$\alpha_1 = 2\sqrt{2}.$$

The new point is

$$x^{(1)} = x^{(0)} + \alpha_1 d^{(1)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and $\Omega(x^{(1)}) = \{3\}$.

Iteration 2. In $x^{(1)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ we have

$$\nabla f(x^{(1)}) = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

and a set of active constraints

$$\Omega(x^{(1)}) = \{3\}.$$

Since we want to minimize $f(x)$, the best direction is the one that minimizes the scalar product with the gradient vector $\nabla f(x)$ at the current point. However the direction must be feasible with respect to the active constraint, i.e. its scalar product with the gradient of the active constraint must be non-negative. The gradient of constraint $x_2 \geq 0$ is

$$\nabla g_3(x) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

everywhere. So, we solve the sub-problem

$$\begin{aligned} \text{minimize } & \nabla f(x^{(1)}) d^{(2)} = 6d_1^{(2)} + 6d_2^{(2)} \\ \text{s.t. } & d_2^{(2)} \geq 0 \\ & (d_1^{(2)})^2 + (d_2^{(2)})^2 = 1 \end{aligned}$$

The optimal solution is

$$d^{(2)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

The next point $x^{(2)}$ is a function of the step α_2 to be taken along direction $d^{(2)}$:

$$x^{(2)} = x^{(1)} + \alpha_2 d^{(2)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 - \alpha_2 \\ 0 \end{bmatrix}.$$

The value of $f(x)$ in the new point $x^{(2)}$ is a function of α_2:

$$f(x^{(2)}) = -(-3 - \alpha_2)^2 - 9$$

and the following constraints must be satisfied

$$\begin{cases} (1 - \alpha_2) + 0 - 6 \leq 0 \\ -1 + \alpha_2 \leq 0 \end{cases}$$

To find the optimal step $\alpha_2 \geq 0$ we have to solve the sub-problem:

$$\begin{aligned} \text{minimize } & f(\alpha_2) = -\,(-3 - \alpha_2)^2 - 9 \\ \text{s.t. } & \alpha_2 \geq -5 \ \ \text{(redundant)} \\ & \alpha_2 \leq 1 \\ & \alpha_2 \geq 0 \end{aligned}$$

The first derivative of $f(\alpha_2)$ is

$$\frac{\partial f}{\partial \alpha_2} = -2\alpha_2 - 6$$

which is never null for $\alpha_2 \geq 0$. Therefore no minimum is reached by moving along $d^{(2)}$; the step is only limited by the constraints. The first constraint that becomes active is the second one: $\alpha_2 \leq 1$. This is the binding constraint. Therefore we have

$$\alpha_2 = 1.$$

The new point is

$$x^{(2)} = x^{(1)} + \alpha_2 d^{(2)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

and $\Omega(x^{(2)}) = \{2, 3\}$.

Iteration 3. In $x^{(2)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ we have

$$\nabla f(x^{(2)}) = \begin{bmatrix} 8 \\ 6 \end{bmatrix}$$

and a set of active constraints

$$\Omega(x^{(1)}) = \{2, 3\}.$$

Since we want to minimize $f(x)$, the best direction is the one that minimizes the scalar product with the gradient vector $\nabla f(x)$ at the current point. However the direction must be feasible with respect to the active constraints, i.e. its scalar product with the gradient of the active constraints must be non-negative. So, we solve the sub-problem

$$\begin{aligned} \text{minimize } \nabla f(x^{(2)})d^{(3)} &= 8d_1^{(3)} + 6d_2^{(3)} \\ \text{s.t. } & d_1^{(3)} \geq 0 \\ & d_2^{(3)} \geq 0 \\ & (d_1^{(3)})^2 + (d_2^{(3)})^2 = 1 \end{aligned}$$

Disregarding the normalization constraint, the optimal solution is

$$d^{(3)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which means that no improving direction exists at the current point. Therefore the algorithm stops at the local minimum $x^{(2)}$.