Unconstrained Nonlinear Programming

The following conventions apply unless otherwise specified or made clear by the context. A Greek letter represents a scalar. A lower case letter represents a column vector. An upper case letter represents a matrix.

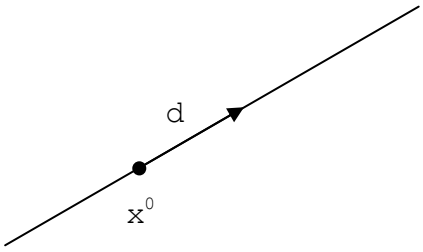
1 Minimizing a multivariate function

Consider a nonlinear function $\pi \in C^2$.

$\pi(x) : R^n \rightarrow R$

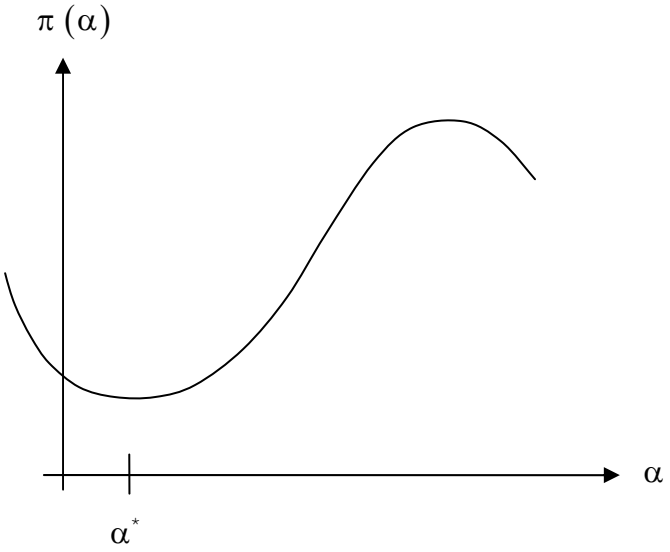
Consider the equation of a line, passing through point x^0 and with direction given by vector d.

$x = x^0 + \alpha d$



Note that for the points belonging to the line, the nonlinear function $\pi(x)$ becomes function of only one variable.

$\pi(x) = \pi(x^0 + \alpha d) = \pi(\alpha)$



In order to find the local minimum points of a nonlinear multivariate function, the general strategy that can be used is: Choose a starting point and move in a given direction such that the function decreases. Find the minimum point in this direction and use it as a new starting point. Continue this way until a local minimum point is reached.

1.1 The gradient vector

The first derivative of $\pi(\alpha)$ can be written as:

$$\pi'(\alpha) = \sum_{i=1}^{n} \frac{\partial \pi}{\partial x_i} \frac{dx_i}{d\alpha} = \sum_{i=1}^{n} \frac{\partial \pi}{\partial x_i} d_i$$

The gradient vector of $\pi(x)$ appears naturally in this equation. The component g_i of the gradient vector of $\pi(x)$ is:

$$g_i = \frac{\partial \pi}{\partial x_i}$$

1.2 The Hessian matrix

The second derivative of $\pi(\alpha)$ can be written as;

$$\pi''(\alpha) = \frac{d\left(\sum_{i=1}^{n} \frac{\partial \pi}{\partial x_i} d_i\right)}{d\alpha} = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 \pi}{\partial x_i \partial x_j} d_i d_j$$

The Hessian matrix of $\pi(x)$ appears naturally in this equation. The component h_{ij} of the Hessian matrix of $\pi(x)$ is:

$$h_{ij} = \frac{\partial^2 \pi}{\partial x_i \partial x_j}$$

1.3 The gradient vector and the Hessian matrix

These two mathematical entities play a key role in the analysis and establishment of numerical methods to find local minimum points of $\pi(x)$. The first and second derivatives of $\pi(\alpha)$ can be written as:

$$\pi'(\alpha) = \sum_{i=1}^{n} g_i d_i = g^T(x)\, d$$

$$\pi''(\alpha) = \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij} d_i d_j = d^T H(x)\, d$$

In structural analysis, the points x are the nodal displacements of the structure, the function $\pi(x)$ is the total potential energy, the vector $g(x)$ is the residue vector and $H(x)$ is the stiffness matrix.

1.4 The Taylor series expansion

The Taylor series expansion of $\pi(\alpha)$ about $\alpha = 0$ can be written as:

$$\pi(\alpha) = \pi(0) + \pi'(0)\,\alpha + \pi''(0)\frac{\alpha^2}{2!} + \cdots$$

However,

$$\pi(0) = \pi\left(x^0\right)$$

$$\pi'(0) = g^T(x^0)\,d$$

$$\pi''(0) = d^T H(x^0)\,d$$

Therefore, the Taylor series expansion of $\pi(\alpha)$ about $\alpha = 0$ can be written as:

$$\pi(x^0 + \alpha d) = \pi(x^0) + \alpha g^T(x^0)\,d + \frac{\alpha^2}{2!} d^T H(x^0)\,d + \cdots$$

1.5 The descent direction

The previous equation can be written in the following form:

$$\pi(x^0 + \alpha d) - \pi(x^0) = \alpha\left[g^T(x^0)\,d + \frac{\alpha}{2!} d^T H(x^0)\,d + \cdots\right]$$

It is sufficient to consider only $\alpha \geq 0$. Note that walking with a negative α in the +d direction is the same as walking with a positive α in the -d direction.

If $g^T(x^0)\,d < 0$ then, for sufficiently small values of $\alpha > 0$ the function π will decrease.

$$g^T(x^0)\,d < 0 \Rightarrow$$
$$\Rightarrow \alpha\left[g^T(x^0)\,d + \frac{\alpha}{2!} d^T H(x^0)\,d + \cdots\right] < 0 \Rightarrow$$
$$\Rightarrow \pi(x^0 + \alpha d) < \pi(x^0)$$

In this situation, the direction given by the vector d is known as descent direction.

1.6 The descent direction and the gradient vector

Let A be any positive definite matrix. A descent direction at a given point x^0 can be defined by the following expression:

$$d = -Ag\left(x^0\right) \Rightarrow$$

$$\Rightarrow g^T\left(x^0\right) d = -g^T\left(x^0\right) Ag\left(x^0\right) < 0$$

It is important to note that the basic numerical methods used to find the local minimum points of a nonlinear function differ mainly by the choice of this matrix.

$A = I \Rightarrow$ Gradient method

$A = \left[H\left(x^0\right) + E\right]^{-1} \Rightarrow$ Newton's method

A diagonal matrix E is added to the Hessian matrix H to ensure that the matrix A is positive definite.

$A \approx H^{-1}\left(x^0\right) \Rightarrow$ Quasi-Newton method

Starting with the unit matrix, a positive definite approximation to the inverse of the Hessian matrix is updated at each iteration. This update is made using only values of the gradient vector.

1.7 Sufficient conditions for a local minimum point

Consider a nonlinear function $\pi \in C^2$, given by:

$$\pi(x) : R^n \rightarrow R$$

If $g(x^*) = 0$ and $H(x^*)$ is positive definite then x^* is strictly a local minimum point.

The Taylor series expansion of $\pi(x)$, on any direction d, about $x = x^*$ can be written as:

$$\pi\left(x^* + \alpha d\right) = \pi\left(x^*\right) + \alpha g^T\left(x^*\right) d + \frac{\alpha^2}{2!} d^T H\left(x^*\right) d + o\left(\alpha^3\right)$$

However,

$$g\left(x^*\right) = 0 \Rightarrow$$

$$\Rightarrow \pi\left(x^* + \alpha d\right) - \pi\left(x^*\right) = \alpha^2\left[\frac{1}{2!} d^T H\left(x^*\right) d + o(\alpha)\right]$$

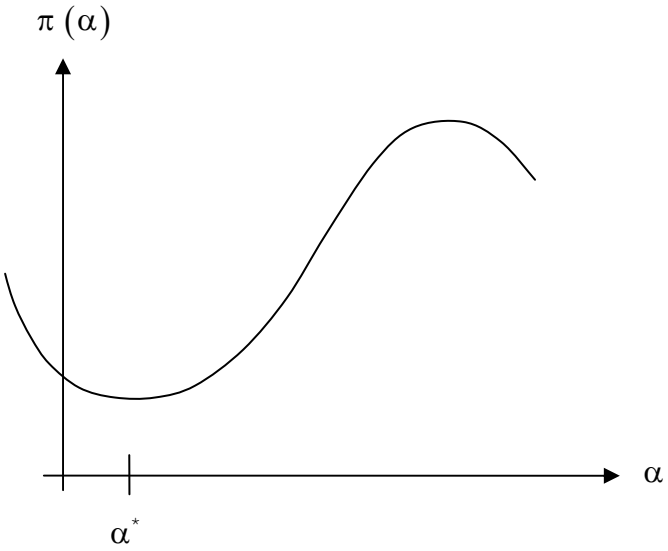
Note that for sufficiently small values of $\alpha > 0$ and considering that $H(x^*)$ is positive definite, the right side of the previous equation will be positive. Therefore,

$$\pi\left(x^* + \alpha d\right) > \pi\left(x^*\right)$$

The point x^* is strictly a local minimum point of $\pi(x)$.

1.8 The line search procedure

The procedure to find a minimum point of $\pi(\alpha)$ for $\alpha \geq 0$ is called line search. This point will be denoted by α^*. Note that a descent direction implies that $\pi'(0) < 0$.



$$\pi\left(\alpha\right) = \pi\left(x^0 + \alpha d\right)$$

$$\pi'\left(\alpha\right) = g^T\left(x^0 + \alpha d\right) d$$

The line search is the essence of basic numerical methods in nonlinear programming. The performance of any method greatly depends on the line search performance. The problem of finding a local minimum point of a multivariate function is reduced to a sequence of problems of finding the minimum point of a univariate function.

The line search procedure can be described in the following way: Find a first interval (α_1, α_2) containing α^*. Calculate α as an approximation for α^* using the limit points of the interval. Redefine the interval that contains α^* with the help of α and the sign of $\pi'(\alpha)$. Continue this way until a given accuracy is reached by satisfying a stopping criterion. Take the last value for α as the value for α^*.

It is important to note that the Quasi-Newton method requires a sufficiently accurate line search. This accuracy is achieved through the following stopping criterion:

$$\left|\pi'(\alpha)\right| \leq -\varepsilon\pi'(0) \mid \varepsilon \in [0, 1)$$

This stopping criterion ensures a positive definite approximation to the inverse of the Hessian matrix, which in turn ensures a descent direction.

It is important to add a counter to limit the number of line search iterations because, depending on the value for ε, the stopping criterion can never be satisfied. Note that the addition of this limit eventually may not lead to a new descent direction. Therefore, $\pi'(0) < 0$ should be checked at the beginning of the line search procedure. If this condition is false then the negative of the gradient vector can be used as a descent direction.

There are several methods that can be used for the line search procedure. However, considering structural analysis, only two methods deserve attention. The line fit method, whose order of convergence is approximately 1.618 and the cubic fit method, whose order of convergence is 2.

1.8.1 Line fit

Find a first interval (α_1, α_2) containing α^*. Fit a line as an approximation for $\pi'(\alpha)$ using α_1, $\pi'(\alpha_1)$ and α_2, $\pi'(\alpha_2)$. Calculate α as the root of the line.

$$\alpha = \frac{\alpha_1\pi'(\alpha_2) - \alpha_2\pi'(\alpha_1)}{\pi'(\alpha_2) - \pi'(\alpha_1)}$$

Redefine the interval that contains α^* with the help of α and the sign of $\pi'(\alpha)$. Continue this way until a given accuracy is reached by satisfying a stopping criterion. Take the last value for α as the value for α^*.

1.8.2 Cubic fit

Find a first interval (α_1, α_2) containing α^*. Fit a cubic function as an approximation for $\pi(\alpha)$ using α_1, $\pi(\alpha_1)$, $\pi'(\alpha_1)$ and α_2, $\pi(\alpha_2)$, $\pi'(\alpha_2)$. Calculate α as the minimum point of the cubic function.

$$u_1 = 3\left[\frac{\pi(\alpha_1) - \pi(\alpha_2)}{\alpha_2 - \alpha_1}\right] + \pi'(\alpha_1) + \pi'(\alpha_2)$$

$$u_2 = \sqrt{u_1^2 - \pi'(\alpha_1)\,\pi'(\alpha_2)}$$

$$\alpha = \alpha_2 - (\alpha_2 - \alpha_1)\left[\frac{\pi'(\alpha_2) + (u_2 - u_1)}{\pi'(\alpha_2) - \pi'(\alpha_1) + 2u_2}\right]$$

Redefine the interval that contains α^* with the help of α and the sign of $\pi'(\alpha)$. Continue this way until a given accuracy is reached by satisfying a stopping criterion. Take the last value for α as the value for α^*.

To avoid severe cancellation, the difference $(u_2 - u_1)$ can be calculated as:

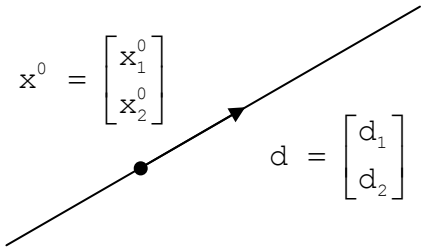
$$\delta = \frac{\pi'(\alpha_1)\,\pi'(\alpha_2)}{u_1^2}$$

$$u_1 > 0 \Rightarrow (u_2 - u_1) = -u_1\,\frac{\delta}{\left(\sqrt{1 - \delta} + 1\right)}$$

$$u_1 < 0 \Rightarrow (u_2 - u_1) = -u_1\left(\sqrt{1 - \delta} + 1\right)$$

2 Appendix

Equation of a line in 2 dimensions



$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1^0 \\ x_2^0 \end{bmatrix} + \alpha \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} x_1^0 + \alpha d_1 \\ x_2^0 + \alpha d_2 \end{bmatrix}$$

Positive definite matrix

A matrix A is positive definite if $\phi(x) = x^T Ax > 0$ for all nonzero vectors x. Note that $\phi(x)$ is a scalar.

3 Examples

3.1 The Rosenbrock function

The Rosenbrock function is used as a test problem for unconstrained nonlinear programming algorithms. The minimum is inside a long, narrow, parabolic shaped flat valley. To find the valley is trivial, however convergence to the minimum is difficult.

Objective function

$$\pi(x_1, x_2) = 100\left(x_1^2 - x_2\right)^2 + (1 - x_1)^2$$

Notice that the function is given by a sum of squares. Therefore there is no possible lower value than zero.

$$x = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow \pi = 0$$

The traditional starting point is (-1,-1).

Gradient of the objective function

$$\frac{\partial \pi}{\partial x_1} = 400\left(x_1^2 - x_2\right)x_1 - 2\left(1 - x_1\right)$$

$$\frac{\partial \pi}{\partial x_2} = -200\left(x_1^2 - x_2\right)$$

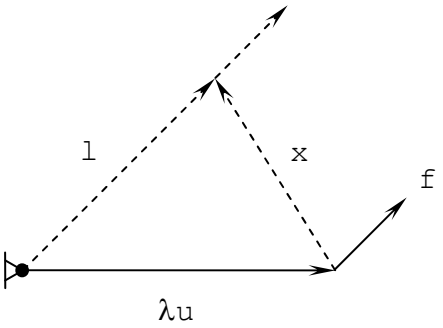
The computer code

```
-- objective
f := 100.0 * (x(1) ** 2 - x(2)) ** 2 + (1.0 - x(1)) ** 2;

-- gradient
g(1) := 400.0 * (x(1) ** 2 - x(2)) * x(1) - 2.0 * (1.0 - x(1));
g(2) := -200.0 * (x(1) ** 2 - x(2));
```

3.2 Truss element 1

The following figure shows a truss element with one node fixed. The modulus of elasticity is given by E and the cross section area of the element is given by A.



The initial position is given by vector λu, where u is a unit vector and λ is the initial length of the element. The final position is given by vector l. The displacement from the initial to the final position is given by vector x. The applied force is given by vector f.

Geometry

For the vectors λu, x and l the following equation can be written:

$\lambda u + x - l = 0 \Rightarrow l = \lambda u + x$

The norm of vector l, which represents the final length of the truss element, can be written as:

$$\|l\| = \left(l^T l\right)^{\frac{1}{2}} = \left[(\lambda u + x)^T (\lambda u + x)\right]^{\frac{1}{2}}$$

$$\|l\| = \left(x^T x + 2\lambda u^T x + \lambda^2\right)^{\frac{1}{2}}$$

Strain

The strain can be written as:

$$\varepsilon = \frac{\|l\|}{\lambda} - 1$$

$$\varepsilon = \frac{\left(x^T x + 2\lambda u^T x + \lambda^2\right)^{\frac{1}{2}}}{\lambda} - 1$$

Severe cancelation

The expression for ε may present severe cancelation. This happens with the subtraction of two approximate numbers. The severe cancelation can be avoided by a mathematical equivalent expression for ε.

$$\varepsilon = \frac{\frac{x^T x + 2\lambda u^T x}{\lambda^2}}{\left(\frac{x^T x + 2\lambda u^T x}{\lambda^2} + 1\right)^{\frac{1}{2}} + 1}$$

Gradient of the strain

$$\nabla\varepsilon = \frac{1}{\lambda}\left(x^T x + 2\lambda u^T x + \lambda^2\right)^{-\frac{1}{2}} (x + \lambda u)$$

Strain energy

$$\phi(x) = \frac{EA\lambda}{2}\varepsilon^2$$

Total Potential energy

Note that The potential energy due to a force is equal to minus the work done by the force.

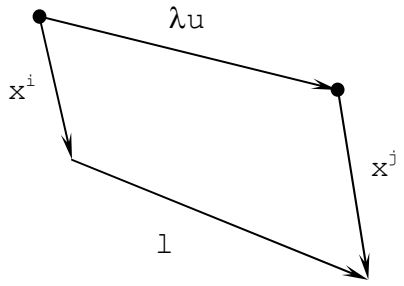
$$\pi(x) = \frac{EA\lambda}{2}\varepsilon^2 - f^T x$$

Gradient of the total potential energy

$$\nabla\pi(x) = EA\lambda\varepsilon\nabla\varepsilon - f$$

3.3 Truss element 2

The following figure shows a truss element connected to nodes i and j. The modulus of elasticity is given by E and the cross section area of the element is given by A. Consider a structure as a a set of connected elements.



The initial position is given by vector λu, where u is a unit vector and λ is the initial length of the element. The final position is given by vector l. The displacements are given by vectors x^i and x^j.

Geometry

For the vectors λu, x^j, l and x^i the following equation can be written:

$$\lambda u + x^j - l - x^i = 0 \Rightarrow l = \lambda u + x^j - x^i$$

$$z = \frac{1}{\lambda}\left(x^j - x^i\right) \Rightarrow l = \lambda(u + z)$$

The norm of vector l, which represents the final length of the truss element, can be written as:

$$\|l\| = \left(l^T l\right)^{\frac{1}{2}}$$

$$l^T l = \lambda^2 \left(1 + 2u^T z + z^T z\right)$$

$$\|l\| = \lambda \left(1 + 2u^T z + z^T z\right)^{\frac{1}{2}}$$

Strain

The strain can be written as:

$$\varepsilon = \frac{\|l\|}{\lambda} - 1$$

$$\varepsilon = \left(1 + 2u^T z + z^T z\right)^{\frac{1}{2}} - 1$$

Severe cancelation

The expression for ε may present severe cancelation. This happens with the subtraction of two approximate numbers. The severe cancelation can be avoided by a mathematical equivalent expression for ε.

$$\varepsilon = \frac{2u^T z + z^T z}{\|u + z\| + 1}$$

Gradient of the strain

$$\begin{bmatrix} \frac{\partial \varepsilon}{\partial x_1^i} \\ \frac{\partial \varepsilon}{\partial x_2^i} \end{bmatrix} = -\frac{1}{\lambda}\left(1 + 2u^T z + z^T z\right)^{-\frac{1}{2}}(u + z)$$

$$\begin{bmatrix} \frac{\partial \varepsilon}{\partial x_1^j} \\ \frac{\partial \varepsilon}{\partial x_2^j} \end{bmatrix} = +\frac{1}{\lambda}\left(1 + 2u^T z + z^T z\right)^{-\frac{1}{2}}(u + z)$$

Strain energy

$$\phi(x) = \frac{EA\lambda}{2}\varepsilon^2$$

Total Potential energy for the structure

Consider that EA is constant for all elements. Consider also that each node of the structure has an applied force in the direction of its displacement. The potential energy due to a force is equal to minus the work done by the force.

$$\pi(x) = \frac{EA}{2}\sum_{\text{elements}} \lambda\varepsilon^2 - f^T x$$

Gradient of the total potential energy for the structure

$$\nabla\pi(x) = EA\sum_{\text{elements}} \lambda\varepsilon \begin{bmatrix} \frac{\partial\varepsilon}{\partial x_1^i} \\ \frac{\partial\varepsilon}{\partial x_2^i} \\ \frac{\partial\varepsilon}{\partial x_1^j} \\ \frac{\partial\varepsilon}{\partial x_2^j} \end{bmatrix} - f$$

The hanging chain problem

The truss element can be used to define the shape of the following structure. Note that the displacements for nodes 4 and 5 are known. Therefore, the derivatives with respect to them are zero.