VARIABLE METRIC MINIMIZATION

Sorin R. Straja, Ph.D., FRM
Montgomery Investment Technology, Inc.
200 Federal Street
Camden, NJ 08103
Phone: (610) 688-8111
sorin.straja@fintools.com
www.fintools.com

Suppose that the function **f(x)**, **f: $\mathbf{R}^n$ → R**, can be locally approximated by the following quadratic form:

$$f(x) = c - b \cdot x + \frac{1}{2} x \cdot A \cdot x$$

The basic idea of the variable metric algorithm is to build up, iteratively, a good approximation of the inverse Hessian matrix $\mathbf{A}^{-1}$, i.e. to build up a sequence of matrices $\mathbf{H_i}$ with the following property:

$$\lim H_i = A^{-1} \quad for \quad i \rightarrow \propto$$

Consider finding the minimum by using Newton's method. Near the current point $\mathbf{x_i}$ we have the following second order approximation:

$$f(x) = f(x_i) + (x - x_i) \cdot \nabla f(x_i) + \frac{1}{2} (x - x_i) \cdot A \cdot (x - x_i)$$

Therefore:

$$\nabla f(x) = \nabla f(x_i) + A \cdot (x - x_i)$$

From the minimum condition (i.e., the gradient is zero) we get the next iteration point:

$$x - x_i = - A^{-1} \cdot \nabla f(x_i)$$

or:

$$x_{i+1} - x_i = H_{i+1} \cdot (\nabla f(x_{i+1}) - \nabla f(x_i))$$

Using the above mentioned approximation for the inverse Hessian matrix it results:

$$x_{i+1} - x_i = A^{-1} \cdot (\nabla f(x_{i+1}) - \nabla f(x_i))$$

The left-hand side of the above equation is the finite step to take in order to reach the exact minimum. The right-hand side is defined once we have an accurate **H**. The variable metric method is labeled also as a quasi-Newton method because it uses an approximation instead of the actual Hessian matrix. The descent direction requires **A** to be a positive definite matrix. Far from the minimum, this condition can be violated. The quasi-Newton method may be better than the original Newton method because it approximates the inverse Hessian matrix using positive definite matrices.

The Davidon-Fletcher-Powell algorithm updates the **H** matrix as follows:

$$H_{i+1} = H_i + \frac{(x_{i+1} - x_i) \otimes (x_{i+1} - x_i)}{(x_{i+1} - x_i) \otimes (\nabla f(x_{i+1}) - \nabla f(x_i))} -$$

$$- \frac{[H_i \cdot (\nabla f(x_{i+1}) - \nabla f(x_i))] \otimes [H_i \cdot (\nabla f(x_{i+1}) - \nabla f(x_i))]}{(\nabla f(x_{i+1}) - \nabla f(x_i)) \cdot H_i \cdot (\nabla f(x_{i+1}) - \nabla f(x_i))}$$

Where $\otimes$ denote the "outer" (or "direct") product of two vectors. The Broyden-Fletcher-Goldfarb-Shanno algorithm updates the **H** matrix as follows:

$$H_{i+1} = H_i + \frac{(x_{i+1} - x_i) \otimes (x_{i+1} - x_i)}{(x_{i+1} - x_i) \otimes (\nabla f(x_{i+1}) - \nabla f(x_i))} -$$

$$- \frac{[H_i \cdot (\nabla f(x_{i+1}) - \nabla f(x_i))] \otimes [H_i \cdot (\nabla f(x_{i+1}) - \nabla f(x_i))]}{(\nabla f(x_{i+1}) - \nabla f(x_i)) \cdot H_i \cdot (\nabla f(x_{i+1}) - \nabla f(x_i))}$$

$$+ [(\nabla f(x_{i+1}) - \nabla f(x_i)) \cdot H_i \cdot (\nabla f(x_{i+1}) - \nabla f(x_i))]\, u \otimes u$$

Where:

$$u = \frac{(x_{i+1} - x_i)}{(x_{i+1} - x_i) \cdot (\nabla f(x_{i+1}) - \nabla f(x_i))} -$$

$$- \frac{H_i \cdot (\nabla f(x_{i+1}) - \nabla f(x_i))}{(\nabla f(x_{i+1}) - \nabla f(x_i)) \cdot H_i \cdot (\nabla f(x_{i+1}) - \nabla f(x_i))}$$

REFERENCES

Dennis, J.E., and Schnabel, R.B. “Numerical Methods for Unconstrained Optimization and Nonlinear Equations”, Prentice-Hall, Englewood Cliffs, NJ, 1983

Brodli, K.W. “The State of the Art in Numerical Analysis”, Chapter III, (Jacobs D.A.H. editor), Academic Press, London, 1977

Polak, E. “Computational Methods in Optimization”, Academic Press, New York, NY, 1971