# COMPUTATIONAL FINANCE

## Chapter 1: The Solution of Nonlinear Equations

An essential component in the quantitative modeling of financial processes is the solution of equations and inequalities. If the model leads to a system of linear equations of the form

$$Ax = b,$$

where $A$ is an $m \times n$ matrix, then the tools of linear algebra can be brought to bear to analyze this system, and the highly developed software of numerical linear algebra can be employed to find the actual solution or an approximation to it.

The situation is usually considerably more complicated when the model leads to $m$ nonlinear equations in $n$ unknowns, written conveniently as

$$F(x) = 0$$

where $x = (x_1, \ldots, x_n)$ and $F = (f_1, \ldots, f_m)$. There is in general no guarantee that such a system will have a solution, and if there are solutions, that any of them can be found numerically. If the system can in fact be solved then the usual methods are iterative and require repeatedly the solution of linear systems.

Many of the analytical and numerical problems disappear when we have the special case of one equation in one unknown, written as

$$f(x) = 0.$$

We shall begin our discussion by considering this case.

### 1. The numerical solution of $f(x) = 0$.

In order to have a concrete case in mind let us consider the following example from the world of equity options:

A "European call" is an option which gives its holder (owner) the right to buy an asset for a specified amount $\$K$ at a specified time $T$. $K$ is known as the strike price and $T$ is the time of maturity of the option. If at time $T$ the asset can be bought at a cheaper price than \$$K$ then the option will not be exercised. But if the asset trades for more than \$$K$ then the holder will exercise the option. The option confers a right but not an obligation and hence has value. The writer (seller) of the option, on the other hand, is obligated to sell the asset on demand at time $T$ for \$$K$ regardless of its value at that time. This creates risk for which the writer must be compensated by charging for the option.

Pricing options is one of the central topics of computational finance and will dominate this course. For the above European call there is the famous Black-Scholes formula which explicitly gives its price

$$C = SN(d_1) - Ke^{-rT}N(d_2)$$

where $S$ is the price of the asset at the time $t = 0$ when the option is sold. Here

$$d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln(S/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-s^2/2} ds.$$

The parameter $r$ is the riskless interest rate at which money can be invested today with payout at time $T$ and is generally assumed known. The second parameter $\sigma$ is the so-called volatility of the asset and is a measure of the day to day changes in the value $S$ of the asset. There is a great deal of uncertainty of how to choose $\sigma$ because it is not an observable quantity. What is observable is the market price $C$ of the option. So one possible approach to determining $\sigma$ is to find that value of $\sigma$ which for given $K$, $r$, $T$ and $S$ produces a value $C$ which is identical to the quoted market price. This value of $\sigma$ is called the "implied volatility." Mathematically, we need to solve the problem

$$f(\sigma) \equiv C - SN(d_1(\sigma)) + Ke^{-rT}N(d_2(\sigma)) = 0$$

where $K$, $r$, $T$, $S$ and $C$ are given. It is clear that this is a highly nonlinear equation in the unknown $\sigma$ which can only be solved numerically (or approximately). Implied volatility calculations are said to take place around the clock in the financial derivative industry.

With this example in the background let us now consider feasible numerical methods for solving the general problem

$$f(x) = 0.$$

We shall denote a root of this equation by $x^*$.

If this equation has to be solved only a few times then efficiency in its solution is of secondary concern. One may as well plot $f$ versus $x$, read off the screen an approximate value for $x^*$ and then refine this value by replotting $f$ over a small interval centered at this approximation. It is clear that such an interactive approach is expensive in terms of man-hours and function evaluations.

This intuitive approach is made formal by the method of bisection which under mild conditions on the function $f$ is guaranteed to give a numerical solution arbitrarily close to the analytic solution $x^*$.

**Bisection algorithm:** Let $f$ be continuous on the interval $[a, b]$. Suppose that $f(a)f(b) \leq 0$. Then

i) $f(x) = 0$ has at least one solution $x^* \in [a, b]$.

ii) let $x_L = a$, $x_R = b$; for $n = 1, 2, \ldots$ set

$$z_n = \frac{x_L + x_R}{2}.$$

Then $|z_n - x^*| < \dfrac{x_R - x_L}{2^n}$

If $f(z_n) \neq 0$ and $z_n$ is not acceptably close to $x^*$ then set

$x_L = z_n$ if $f(z_n)f(x_R) \leq 0$

$x_R = z_n$ if $f(z_n)f(x_R) > 0$

and go to $n + 1$.

The continuity of $f$ and a sign change over the interval $[a, b]$ guarantee that $f$ has at least one root in the interval. In every step of the algorithm we insure that $f$ changes sign over $[x_L, x_R]$. Since in every step of the algorithm we halve the interval we see that at step $n$

$$x_R - x_L = \frac{b - a}{2^n}$$

so that

$$|z_n - x^*| < \frac{b - a}{2^n}.$$

Suppose we wish to insure that $|z - x^*| < 10^{-6}$. This is guaranteed if

$$\frac{b^n - a}{2^n} < 10^{-6}$$

or

$$n > \frac{\ln(b-a) + 6\ln 10}{\ln 2}.$$

For example, if $b - a = 1$ then

$$n > 19.93$$

so that as many as 20 function evaluations may be required to achieve the stated accuracy.

Suppose we wish to apply the bisection method to the implied volatility calculation. By inspection $d_1$ and $d_2$ are continuous functions of $\sigma$ and since $N$ is a continuous function of $x$ it follows that $f$ is a continuous function of $\sigma$. Moreover,

$$\lim_{\sigma \to \infty} N(d_1(\sigma)) = \lim_{x \to \infty} N(x) = 1$$

$$\lim_{\sigma \to \infty} N(d_2(\sigma)) = \lim_{x \to -\infty} N(x) = 0$$

so that

$$\lim_{\sigma \to \infty} f(\sigma) = C - S.$$

This is a negative quantity since if the call were more expensive than the asset one may as well buy the asset itself.

The condition as $\sigma \to 0$ is a little trickier. We note that

$$\lim_{\sigma \to 0} d_1(\sigma) = \lim_{\sigma \to 0} d_2(\sigma) = \lim_{\sigma \to 0} \frac{\ln S/K + rT}{\sigma\sqrt{T}}.$$

If $\ln S/K + rt < 0$ then $d_1$ and $d_2 \to -\infty$ as $\sigma \to 0$ and

$$\lim_{\sigma \to 0} f(\sigma) = C > 0.$$

If $\ln S/K + rt > 0$ then $d_1$ and $d_2 \to \infty$ as $\sigma \to 0$ and

$$\lim_{\sigma \to 0} f(\sigma) = C - (S - Ke^{-rT}).$$

For a correctly priced call this quantity is also positive, for if

$$\ln S/K - rT > 0, \quad \text{i.e., } S > Ke^{-rT}$$

and

$$C < (S - Ke^{-rT})$$

then an investor can sell short the asset for \$$S$ and buy the call for $C$. The value of the contract at time $T$ is then

$$(S - C)e^{rT}$$

which would exceed the strike price $K$ required to repurchase the asset sold short before. Hence

$$\lim_{\sigma \to 0} f(\sigma) > 0 \quad \text{and} \quad \lim_{\sigma \to \infty} f(\sigma) < 0$$

so that the bisection method will succeed.

A search method like the method of bisection is not useful if efficiency of the computation is of concern. Then more sophisticated methods must be employed which, unfortunately, tend to be more delicate and require deeper mathematical insight to insure that they work and work well.

Let us again consider the problem

$$f(x) = 0.$$

We shall rewrite the equation generically in the form of a so-called fixed point equation

$$x = g(x)$$

where $g$ is chosen such that any $x^*$ which satisfies

$$x^* = g(x^*)$$

is also a solution of $f(x) = 0$. The solution $x^*$ is known as a fixed point of $g$.

There are infinitely many fixed point equations associated with a given $f(x) = 0$. For example, one may write trivially

$$x = g(x) \equiv x - \alpha f(x)$$

for any non-zero scalar $\alpha$. More commonly, a $g$ is obtained by solving $f$ for $x$ in terms of a function of $x$. For example,

$$f(x) \equiv x - x^2 = 0$$

can rewritten as the fixed point equations

$$x = x^2$$

or

$$x = \sqrt{x}.$$

The numerical solution of

$$x = g(x)$$

is obtained by simple substitution or, what is the same, a fixed point iteration. We assume that we have an initial guess $x^0$ and compute iteratively the sequence $\{x^n\}$ from

$$x^{n+1} = g(x^n), \qquad n = 0, 1, 2, \ldots$$

If the sequence converges to some $x^*$ then we have found a root of $f$.

Convergence of the iteration requires a particular structure of $g$. If $g$ does not have the right properties then the iteration will not converge regardless of how close $x^0$ is to a root of $f$. For example,

$$f(x) \equiv x - x^2 = 0$$

has a root at $x^* = 1$ and at $x^* = 0$. But if we try to solve

$$x = x^2$$

with an initial guess of

$$x^0 = 1 + \epsilon, \qquad \epsilon > 0$$

then $x^n \to \infty$ as $n \to \infty$ no matter how small $\epsilon$ is chosen. The desirable property for a fixed point $x^*$ is that it be a point of attraction which is defined as follows.

**Definition:** Let $x^*$ be a fixed point of $g$. $x^*$ is a *point of attraction* of the fixed point iteration if there is a neighborhood $N(x^*, \delta)$ of $x^*$ (i.e., an interval of radius $\delta$ around $x^*$) such that for any $x^0 \in N(x^*, \delta)$ the fixed point iteration converges to $x^*$.

The discussion above shows that $x^* = 1$ is not a point of attraction of

$$g(x) = x^2.$$

A moment's reflection will show that the fixed point $x^* = 0$ is a point of attraction. Similarly, $x^* = 1$ is a point of attraction of the alternate fixed point equation

$$x = g(x) = \sqrt{x}$$

but $x^* = 0$ is not. In many applications one has a reasonable idea of $x^*$ and it is important to have a fixed point formulation for $f(x) = 0$ for which $x^*$ is a point of attraction so that a good initial guess will lead to convergence. This raises the question of what property of $g$ makes a fixed point a point of attraction. We have the following theoretical criterion.

**Theorem:** *Let $x^*$ be a fixed point of the function $g$. Suppose that $g$ is continuously differentiable in a neighborhood of $x^*$ and that*

$$|g'(x^*)| < 1.$$

*Then $x^*$ is a point of attraction.*

**Proof:** If $g'$ is continuous and $|g'(x^*)| < 1$ then $|g'(x)| \leq 1 - \alpha$ for some $\alpha > 0$ and all $x$ in some interval $N(x^*, \delta)$ around $x^*$ of radius $\delta$. (We don't know $x^*$ or the length $\delta$, but we know from analysis that such a $\delta > 0$ exists.) By the mean value theorem for differentiable functions we also know that

$$g(x) - g(x^*) = g'(\xi)(x - x^*)$$

for some $\xi$ between $x$ and $x^*$. Hence if $x \in N(x^*, \delta)$ then

$$|g(x) - g(x^*)| \leq (1 - \alpha)|x - x^*|.$$

Now suppose that $x^0$ is an arbitrary point in $N(x^*, \delta)$. Then

$$|x^1 - x^*| = |g(x^0) - g(x^*)| < (1 - \alpha)|x^0 - x^*|$$

so that $x^1 \in N(x^*, \delta)$; similarly $|x^{n+1} - x^*| \leq (1 - \alpha)|x^n - x^*|$ and hence

$$|x^n - x^*| \leq (1 - \alpha)^n |x^0 - x^*|$$

which guarantees that $x^n \to x^*$ as $n \to \infty$.

If we examine the two fixed point equations

$$x = g_1(x) = x^2$$

and

$$x = g_2(x) = \sqrt{x}$$

associated with

$$f(x) = x - x^2 = 0$$

then we see that

$$g_1'(1) = 2$$

$$g_2'(1) = 1/2$$

so $x^* = 1$ is not a point of attraction for $g_1$ but is a point of attraction for $g_2$.

---

**Newton's method**

---

For the method of bisection we only required a continuous function $f$ and an interval over which $f$ changes sign. The algorithm itself asks for the evaluation of $f$ at given points and does not demand that $f$ be given analytically. For example, the evaluation of $f$ may involve a table-look up or $f$ may require the solution of a differential equation whose solution depends on the independent variable $x$. Hence $f$ may be very general and complicated.

If $f$ is given analytically and $f'$ is continuous then we can solve

$$f(x) = 0$$

by a fixed point method called Newton's method (sometimes Newton Raphson method) which is in general much more efficient than bisection. The idea is as follows: Given an initial guess $x^0$ then for $n = 0, 1, 2, \ldots$ we linearize $f$ around $x^n$ and find $x^{n+1}$ as the solution of the linear problem. The linearization is obtained from the first two terms of the Taylor expansion of $f$, i.e. the linearization of $f$ around $x^n$ is

$$L_n x = f(x^n) + f'(x^n)(x - x^n)$$

so that $x^{n+1}$ is the solution of $L_n x = 0$, or

$$x^{n+1} = x^n - \frac{f(x^n)}{f'(x^n)}\,, \qquad n = 0, 1, 2.$$

There is, of course, a natural geometric interpretation of this method. The equation

$$y = L_n x$$

is the tangent to $f$ at $x^n$ and $x^{n+1}$ is the point where the tangent crosses the $x$-axis. The hope is that $x^{n+1}$ is a better approximation to $x^*$ than the preceding iterate $x^n$.

Newton's method has two exceedingly important properties. Under mild conditions we are guaranteed convergence from a good initial $x^0$ and once $x^n$ is sufficiently close to $x^*$ the convergence is very rapid. These properties are easy to establish in view of our discussion of fixed point iterations.

We observe that Newton's method is a fixed point iteration for

$$x = g(x)$$

where

$$g(x) = x - \frac{f(x)}{f'(x)}\,.$$

Let $x^*$ be a root of $f$ and suppose, as is usually the case, that $f''$ exists and that $f'(x^*) \neq 0$. Then it follows that

$$g'(x) = 1 - \frac{f'(x^*)^2 - f(x^*)f''(x^*)}{f'(x^*)^2} = 0.$$

Hence $x^*$ is a point of attraction and Newton's method will converge from a good initial guess. Moreover, it follows from the identity

$$g(y) = g(x) + g'(x)(y - x) + \frac{1}{2}\,g''(\xi)(y - x)^2$$

for some $\xi$ between $x$ and $y$ that

$$|x^{n+1} - x^*| = |g(x^n) - g(x^*)| = |g(x^*) + g'(x^*)(x^n - x^*) + \frac{1}{2}\,g''(\xi)(x^n - x^*)^2 - g(x^*)|$$

so that in view of $g'(x^*) = 0$ we obtain

$$|x^{n+1} - x^*| \leq K|x^n - x^*|^2$$

for some constant $K$ related to $g''(x)$. Thus, if in the iteration the error $|x^n - x^*|$ is $10^{-2}$ then the next iterate gives an error of order $10^{-4}$. This convergence is called quadratic and usually insures that only two or three iterations are needed to have an acceptable approximation to $x^*$.

The dominant draw-back to using Newton's method is that a good initial guess is required since convergence to a point of attraction is a local property. If one were to provide, say, a Newton's method based program for solving the implied volatility problem one would have to guard against a bad choice of the initial guess $\sigma^0$, unless, of course, one can establish that Newton's method will automatically converge. This will require additional conditions on $f$.

**Theorem:** *Let $f$ be defined on the interval $[a, b]$. Suppose that*

$$\begin{aligned} f(x^*) &= 0 \qquad \text{for some } x^* \in [a, b] \\ f'(x) &> 0 \\ f''(x) &\geq 0 \end{aligned}$$

*Then Newton's method will converge from $x^0 = b$.*

The proof of this result can be given rigorously, but a geometric argument will make clear what happens. First of all, the root $x^*$ is unique in $[a, b]$ because of $f' > 0$. Then we observe that the tangent to $f$ at $x^0 = b$ will lie below the graph of $f$ because $f'' > 0$ implies that $f$ is concave downward. Hence the tangent crosses the $x$-axis at a point $x$ which satisfies

$$x^* \leq x^1 < x^0$$

because the tangent has positive slope. The same observation applies to all subsequent tangents. As a consequence we generate a decreasing sequence of numbers $\{x^n\}$ which is bounded below by $x^*$. Hence the sequence must converge and from

$$x^{n+1} = x^n - \frac{f(x^n)}{f'(x^n)}$$

follows that $x^n$ must converge to a root of $f$, hence to $x^*$. Similar arguments are used if $f$ is decreasing or convex downward. The geometry tells us whether monotone convergence can be guaranteed.

Let us look at an application of this result to the implied volatility calculation.

$$f(\sigma) \equiv C - SN(d_1(\sigma)) + Ke^{-rT}N(d_2(\sigma)) = 0$$

where

$$d_1(\sigma) = \frac{A}{\sigma} + b\sigma, \qquad d_2(\sigma) = d_1(\sigma) - 2b\sigma$$

with

$$A = \frac{\ln(S/K) + rT}{\sqrt{T}} \quad \text{and} \quad b = \sqrt{T}/2.$$

Our discussion of the problem in connection with the bisection method already established that under reasonable assumption we may assume that

$$f(\sigma) = 0$$

has a solution. We compute:

$$f'(\sigma) = -SN'(d_1)d_1' + Ke^{-rT}N'(d_2)d_2'.$$

From the definition of $N(d_2)$ we find

$$N'(d_2) = \frac{1}{\sqrt{2\pi}}\, e^{-d_2^2/2}$$

and if we substitute $d_2 = d_1 - 2b$ then simple algebra leads to

$$N'(d_2) = N'(d_1)\frac{S}{K}\, e^{rT}$$

so that

$$f'(\sigma) = -2bSN'(d_1) < 0.$$

Furthermore,

$$f''(\sigma) = -\frac{2bS}{\sqrt{2\pi}}\,(-d_1 d_1')e^{-d_1^2/2}.$$

But

$$d_1' = -\frac{1}{\sigma}\, d_2$$

so that

$$f''(\sigma) = -\frac{2bS}{\sigma\sqrt{2\pi}}\,(d_1 d_2)e^{-d_1^2/2}.$$

From the definition of $d_1$ and $d_2$ we see that

$$d_1 d_2 = \left(\frac{A}{\sigma} + b\sigma\right)\left(\frac{A}{\sigma} - b\sigma\right).$$

If we set

$$\sigma^0 = \sqrt{|A/b|}$$

then $f''$ is negative for $\sigma < \sigma^0$ and positive for all $\sigma > \sigma^0$. A look at successive tangents to the graph of $f$ shows that starting from $\sigma^0$ Newton's method will converge. If $f(\sigma^0) > 0$ we obtain a monotone increasing sequence, if $f(\sigma^0) < 0$ we generate a monotone decreasing sequence. Hence this choice of initial value is sufficient to guarantee monotone convergence as long as the call is correctly priced so that $f(\sigma) = 0$ has a solution. While in general quadratic convergence only sets in close to the correct volatility in practice this approach is quite efficient compared to the bisection method.

**A few final comments:** The strengths and weaknesses of Newton's method carry over to the solution of the system

$$F(x) = 0.$$

Given $x^n$ the system is again linearized around $x^n$

$$L_n x = F(x^n) + F'(x^n)(x - x^n)$$

where $F'$ is the $n \times n$ matrix

$$F'(x) = (\delta f_i / \delta x_j)$$

$x^{n+1}$ is the solution of

$$L_n x = 0,$$

i.e, formally

$$x^{n+1} = x^n - F'(x^n)^{-1} F(x^n)$$

but computed from

$$F'(x^n)\delta = -F(x^n)$$

$$x^{n+1} = x^n + \delta$$

in order to avoid the inverse of $F'(x)$. The multi-dimensional Newton method will again converge for a good initial guess as long as $F'(x^*)$ is non-singular, and convergence close to the solution remains quadratic. In general, the choice of the initial condition is more critical than in the scalar case and has lead to some sophisticated methods for choosing $x^0$.

Lack of convergence of Newton's method for a good initial guess is usually due to the incorrect calculation of $F'(x)$. In addition, $F$ may not be explicitly given but only in terms of an input-output algorithm (given $x$ one can calculate $F(x)$) so that $F'$ is not calculable. One can avoid $F'(x)$ altogether if its entries are replaced by difference quotients. For example, the $j$th column of $F'(x)$ can be approximated by

$$\frac{F(x + h\hat{e}_j) - F(x)}{h}$$

for small $h$ where $\hat{e}_j$ is the $j$th unit vector. By definition the limit of this difference quotient as $h \to 0$ is the $j$th column of $F'(x)$. This discrete approximation to Newton's method and many variations thereof are closely related to secant methods and interpolation methods which are discussed in texts on the numerical solution of nonlinear systems.