EC316a: Advanced Scientific Computation, Fall 2003

Notes Section 1

Background In this module, we will study a variety of computational techniques used to solve dynamic models in economics and finance. These models are dynamic because they explicitly consider the passage of time—as opposed to static models, which are perhaps the most common objects to anyone who has taken an economics principles course. Structures such as the simple supply and demand model of microeconomics, or the "Keynesian cross" model of elementary macroeconomics, are static representations of a market or an economy, and the analysis of how they react when a particular factor is altered—e.g., shifting one of the curves—is known as "comparative statics." Comparative static analysis is concerned with the analysis of two equilibria, and does not consider the path taken between those equilibrium points (or even whether such a path exists). When economists construct dynamic models, they are concerned with this transition path, and the trajectory that the system will take when going from one equilibrium to another.

We will use the Miranda and Fackler text *Applied Computational Economics and Finance* and the MATLAB toolbox those authors have provided, *CompEcon*, in our study of dynamic economic models. As we have a very limited time for this module, we will not explicitly cover the background material in the first six chapters of the text in any detail. I will provide a quick summary of that material, since some of its elements are referenced in the later chapters.

First, we must consider the need for computation in the context of economic modeling. Many models in economics and finance possess closed–form, or analytical solutions. If an analytical solution exists, why should we need to turn to computational techniques, beyond those related to symbolic algebra and calculus? Why is it necessary to become familiar with techniques of numerical analysis, root–finding, numeric integration and differentiation and interpolation?

Many interesting economic models cannot be analytically solved. These include applied economics models that attempt to capture real–world complexities in economic behavior. For instance, the simple Marshallian market model of supply and demand, customarily depicted as a supply curve and demand curve intersecting in quantity, price space, is too simplified to reflect many aspects of producers' and consumers' behavior which play an important role in the workings of markets. Such factors as multiple goods (or goods differentiated by bundled features), interregional trade and intertemporal storage may be relevant for many markets, as may the interventions of governments in the form of taxes, subsidies, tariffs and trade quotas. A realistic model of a particular market may have to take some of these factors into consideration—but in doing so may be rendered analytically intractable.

A second class of models that are not amenable to analytical solutions includes some of those we will consider here: stochastic, dynamic models of rational, forward–looking economic behavior. These models are dynamic because they explicitly include the timing of decisions; they are stochastic because they incorporate the effects of uncertainty. What is "rational, forward–looking behavior"? Merely the notion that economic agents will do the best they can with the available information to make an optimal decision, and that the decision may be based on their evaluation of future events. For instance, a young worker's decision to contribute toward her pension, the choice of how much of her salary should be directed toward retirement savings and the mix of assets should be purchased as investments in the plan clearly reflect forward–looking behavior, expressing a concern over the worker's well–being in distant years, even though increased contributions to the plan reduce her current income and welfare. This type of model typically gives rise to functional equations where the unknown is not simply a vector (e.g., the amount to contribute to the pension plan each year of working life) but an entire function. Except in very special cases, the functional equation lacks a known closed–form solution, even though the solution can be shown theoretically to exist and be unique.

Models lacking analytical solutions are hardly special to economics nor finance; many models requiring numerical solutions appear in the biological and physical sciences and engineering. But in those fields the need for explicit study of computational modeling and analysis has been more quickly accepted as a key component of the disciplines. Economists are beginning to realize the importance of these techniques; a society to which I belong, the Society for Computational Economics, has just held its 10th annual conference, and the professional journal *Computational Economics* is in its 11th year of publication.

Although an analytical solution is surely to be preferred on the grounds of elegance and generality, it would be a mistake to shun development of a more sophisticated model, incorporating key elements of economic behavior, merely because it required the modeler to eschew a closed–form solution and turn to numerical techniques for its solution. Models, although abstractions from reality by definition, are most useful when they capture the essential features of the relationship being modeled. In a behavioral science such as economics, those features often include the notion that economic agents' reactions to economic events are not perfectly predictable, and indeed that their actions may include "herd instincts" that lead to seemingly irrational outcomes such as "bubbles" in asset prices.

Let us consider a very simple example of analytical intractability in the context of a market model. In a free market, the equilibrium price and quantity is obtained by the juxtaposition of supply and demand functions. The supply function reflects the producer's costs of offering the good to the market. Consumers of the good are characterized by a demand function, which expresses how many units of the good they would be willing to buy at various prices. Let us consider a so–called constant elasticity demand function

$$q = p^{-0.2} \tag{1}$$

which is a simple function linking the quantity demanded (q) to the price of the good (p). It is said to be a constant–elasticity function (as opposed to a linear demand function) since a one percent increase in the price will yield a change in the quantity demanded of -0.2 per cent (in economic terms, the good is said to be inelastic, since the price elasticity of demand is greater than -1.0).

If we know how many units of the good are available at a point in time, we can invert the demand function to determine the market–clearing price:

$$p = q^{-5} \tag{2}$$

So that in this case the so–called inverse demand function (IDF) exists, and we could replace -0.2 with $-\eta$ and represent the IDF analytically. What if we wanted to consider a more general demand function, in which there are two terms expressing the demand for the good from domestic consumers and from foreign consumers (i.e., export demand)? Such a demand function might be

$$q = 0.5p^{-0.2} + 0.5p^{-0.5} \tag{3}$$

That is, foreign consumers are more responsive to price changes, presumably reflecting their option to purchase locally–produced substitutes. What can we say about this demand function? It is continuous, differentiable, and strictly decreasing in p. We can readily calculate the quantity demanded q for any level of price, and graph the resulting functional relationship. But what if we wanted to compute its IDF—to consider, for instance, what price would clear the market for a quantity of two units? On a theoretical level, the Intermediate Value and Implicit Function theorems guarantee existence of the IDF, and indicate that it is continuous and strictly decreasing, and capable of generating a unique price that would clear the market for a specific quantity. But can algebra and calculus answer this question, and indicate what that price will be? No. A numerical solution would be required, and could be obtained by a root–finding algorithm such as Newton's method (which utilizes a first–order Taylor expansion of the function) as 0.154 equalling the market–clearing price.

This is merely one example of a very simple economic relationship in which the introduction of a realistic feature, the notion that a good may be purchased by more than one group of consumers, who may exhibit different reactions to changes in the good's price, has turned the problem into one requiring numerical analysis to achieve a solution.

Let us now consider some of the building blocks used in computational economics and finance. The most elementary problem that arises in computational economic analysis is the system of linear equations, in which an n-vector x that satisfies $Ax = b$ is sought. Even when complex nonlinear problems arise, their solution often includes a system of linear equations, since many solution techniques rely on linearization or log-linearization around a trial solution. Some of the computational methods associated with systems of linear equations are LU-factorization, Gaussian elimination, and a special form of the latter—Cholesky factorization, which may be employed when the matrix A is symmetric and positive definite (such as a matrix of second partials under appropriate conditions). These methods are collectively known as exact methods, and generally dominate the direct computation of the matrix inverse (or that construct favored by economists, the application of "Cramer's rule") in terms of computational burden and accuracy, especially in the case of ill–conditiioned matrices (i.e., those whose numerical determinants are close to machine zero). There are also iterative methods for solution of a system of linear equations such as the Gauss–Jacobi and Gauss–Seidel methods. The latter method is often employed to solve a system of nonlinear equations.

Many interesting problems in economics and finance cannot be expressed as a system of linear equations, but rather contain one or more nonlinear aspects. Sets of nonlinear equations give rise to a root–finding problem (the solution x to the functional relationship $f(x) = 0$) or a fixed–point problem (the vector x satisfying the relationship $g(x) = x$). The two forms are equivalent, in that a root–finding problem may be rewritten as a fixed–point problem by rewriting $g(x) = x - f(x)$, while the converse may be achieved by expressing $f(x) = g(x) - x$. A related problem is the complementarity problem, where we have two n-vectors a and b, $a < b$, a function f, and a solution x that satisfies

$$x_i > a_i \Rightarrow f_i(x) \geq 0 \quad \forall i = 1, \ldots, n \tag{4}$$

$$x_i < b_i \Rightarrow f_i(x) \leq 0 \quad \forall i = 1, \ldots, n \tag{5}$$

where f_i is the partial derivative of f with respect to the i^{th} element of x. The complementarity problem imposes constraints on the optimization problem, so that a root of f (or a fixed point of x) may imply that one of the bounds a or b are breached. In this case, the unconstrained optimum will not equal the constrained optimum. This sort of problem will arise, for instance, in the case of a market model where there are price supports, quantity quotas, price caps, nonnegativity conditions, or limited production capacity. For instance, in a standard market model, excess demand $E(p)$ will be zero at the equilibrium price. What if government imposes a price ceiling $\bar{p}$, and that price is below the free–market equilibrium? Then we may solve the complementarity problem $CP(E, 0, \bar{p})$, which will imply that excess demand could be positive if the price ceiling is binding.

Standard root–finding methods may be considered as a special case of the complementarity problem with $a_i = -\infty$ and $b_i = \infty$ for all i. Conversely, the complementarity problem may be recast as a "minmax" root–finding problem: x solves the $CP(f, a, b)$ iff it solves the root–finding problem

$$\hat{f}(x) = \min(\max(f(x), a - x), b - x) = 0 \quad (6)$$

where min and max are applied row–wise. Thus, the CP may be solved with standard root–finding algorithms by noting that the Jacobian of the function $\widehat{f}$ may be written as

$$\hat{J}_i(x) = \begin{array}{ll} J_i(x), \text{ for} & a_i - x_i < f_i(x) < b_i - x_i \\ I_i, & \text{otherwise} \end{array} \tag{7}$$

where I_i is the i^{th} row of the identity matrix.

Methods designed for the solution of root–finding, fixed–point and complementarity problems may be derivative–free, requiring only specification of the functions involved, or may require that derivatives of the functions be either analytically specified or be computable. The derivative–free methods include the bisection method (where a root is located by successive bisection of an interval in which a root exists) and function iteration (in which a function is successively iterated to locate a fixed point).

Most nonlinear problems are solved with Newton's method or one of its variants, which are based on the principle of successive linearization. In these methods, a difficult nonlinear problem is recast as a sequence of simpler linear problems which lead to the solution. Newton's method is usually considered as a root–finding method, but as described above can be applied to fixed–point problems with a change of functions. It relies on a first–order Taylor approximation to the function, so that the Jacobian of the function must be computable. Let us consider a simple application of Newton's method. Let the inverse demand function for a good, which is produced by two competing firms, equal

$$P(q) = q^{-1/\eta} \tag{8}$$

and let the firms face cost functions

$$C_i(q_i) = \frac{1}{2} c_i q_i^2, i = 1, 2 \tag{9}$$

The profit for firm i is defined as

$$\pi_i(q_1, q_2) = P(q_1 + q_2)q_i - C_i(q_i) \tag{10}$$

In a so–called Cournot duopoly model, each firm makes its decision taking the other firm's choice of output as given. In this setup, firm i will choose its output level so as to solve the first–order conditions for a maximum of the profit function:

$$\partial\pi_i/\partial q_i = P(q_1+q_2)q_i + P'(q_1+q_2)q_i - C_i'(q_i) = 0 \tag{11}$$

The market equilibrium levels of output, q_1 and q_2, are the roots of the two nonlinear equations

$$f_i(q) = (q_1+q_2)^{-1/\eta} - (1/\eta)(q_1+q_2)^{-1/\eta-1}q_i - c_iq_i = 0, \tag{12}$$

for $i = 1, 2$. These equations are nonlinear in q_1, q_2 space, and their intersection denotes the optimal level of output for each Cournot duopolist. Although these equations do not possess an analytical solution, the solution may readily be found by iterative methods, employing Newton's method since the first derivatives of the function are readily calculated. For instance, if we assume that $\eta = 1.6$, $c_1 = 0.6$ and $c_2 = 0.8$, and provide initial guesses of $0.2, 0.2$ for the solution values, Newton's method will yield the equilibrium quantities $q_1 = 0.8396$ and $q_2 = 0.6888$ (the intuitive result that the firm with higher marginal costs will produce at a lower level).

While Newton's method requires analytical first derivatives, a variety of methods based on Newton's method have been developed as derivative–free alternatives; the so-called quasi–Newton (QN) methods. The most widely used univariate QN method is the secant method, which replaces the analytical derivative with an approximation constructed from two successive function evaluations. Its multivariate analogue is Broyden's method, which approximates the root of f and the Jacobian f' at the root, and will generally converge if the functions are reasonably well behaved.

Now let us turn to finite–dimensional optimization methods. We seek

$$\max_{x \in X} f(x) \tag{13}$$

where f is a real-valued objective function, X the feasible set, and a solution $x^* \in X$ (if it exists) satisfies $f(x^*) \geq f(x)$ for all $x \in X$. We can always recast minimization problems as $\min -f(x)$, so that we need only consider maximization. An optimization problem involves first–order conditions, which may be viewed as a root–finding problem, as discussed earlier; the first–order necessary conditions (known as the Kuhn–Tucker conditions) for a constrained optimization problem are a complementarity problem. Problems of this nature, both unconstrained and constrained, are ubiquitous in economics and finance: e.g., consumers maximizing their utility, firms maximizing profit, financial analysts optimizing a portfolio of assets, or governments choosing policies designed to maximize societal welfare. Derivative–free methods include the univariate golden search method (thusly named because it chooses interior evaluation points by utilizing the golden ratio), or the multivariate Nelder-Mead algorithm which operates on the n–dimensional simplex. The equivalent of Newton's method for root–finding is the Newton–Raphson method, which uses successive quadratic approximations to the objective to locate the optimum. It is intimately related to Newton's method, as it amounts to computing the root of the gradient of the objective function, working with a second–order Taylor approximation to the objective function (requiring that the objective function is twice continuously differentiable). The burden of setting up the Newton–Raphson method, and coding both first and second derivatives, is onerous. For this reason, quasi–Newton optiimization methods are often employed. The method of steepest ascent is perhaps the simplest, but does not take into account any information on the curvature of the objective function. Two methods that do the latter are the Davidson–Fletcher–Powell (DFP) and Broydon–Fletcher–Goldfarb–Shanno (BFGS) methods, which employ different updating formulas for the numerical Hessian, but are otherwise identical. These two "workhorse" methods are found in almost all econometric software containing maximum likelihood estimation routines. For more difficult optimization problems, line search methods are often employed to control the length of the Newton step.

We now consider some of the methods used for numerical integration (or quadrature) and numerical differentiation. Where analytic derivatives are unavailable or their construction is burdensome, numerical methods are often used. Univariate Newton–Cotes quadrature methods such as the trapezoid rule and Simpson's rule are used to approximate the definite integral of a real–valued continuous function. They may be generalized to higher–dimensional integration problems. Gaussian quadrature is perhaps most commonly employed; this method employs the notion that any real–valued function may be approximated to arbitrary precision by fitting a sufficiently high–order polynomial (equivalent to the other Gaussian technique, ordinary least squares (OLS) regression, where n data points can be perfectly fit by a polynomial of order $n-1$). Gaussian quadrature employs a weight function; if that is chosen as the identity function, we have Gauss–Legendre quadrature, which is appropriate for computing the area under a curve. Monte Carlo and quasi–Monte Carlo integration routines are also widely used.

Numerical differentiation involves computing finite differences of the function at adjacent points; higher–order numerical Taylor expansions may be used to increase accuracy. Higher–order derivatives are also computed numerically: for instance, many optimization methods require a numerical Hessian (matrix of second partials). Differential equations, used widely in expressing many economic and financial problems, also require numerical derivatives, which are often calculated via Runge–Kutta methods.

As a last topic in our survey of computational methods, consider issues of function approximation. These include interpolation methods, where a computationally intractible function f

is approximated by a computationally tractable function $\hat{f}$. In the functional equation problem, we wish to find a function f which satisfies $Tf = 0$ where T is an operator that maps a vector space of functions onto itself. Closely related is the functional fixed–point problem, with $f = Tf$. These equations are common in dynamic economic models; for instance, the well–known Bellman equation characterizes the solution of an infinite–horizon dynamic optimization problem. These equations are difficult to solve because the unknown is not simply a vector in $\mathbf{R}^n$, but rather an entire function f (in economic problems, often known as the policy function) whose domain contains an infinite number of points, with an infinite number of conditions imposed on the solution f. These equations almost always lack explicit closed–form solutions, and thus cannot be solved exactly; one must seek a numerical solution that satisfies the functional equation to a close approximation. Two generally practical techniques are Chebychev polynomial approximation and polynomial spline interpolation. These methods may be generalized to a multivariate setting in which a number of choice variables' trajectories must be defined by a solution. A generalization of these methods, presented by the authors in section 6.8, is the collocation method, which effectively replaces the difficult infinite–dimensional functional equation problem with a simpler finite–dimensional root–finding problem that can be solved by standard techniques, such as Newton's method or Broyden's method. The finite dimension of the resulting problem refers to the number of collocation nodes chosen; the collocation method requires that the approximating function satisfies the functional equation, not at all possible points of the domain, but at the n prescribed points (the collocation nodes). As an example, again drawing from Cournot oligopoly: the firm equates marginal revenue to marginal cost to maximize profit. Realizing that its actions affect price, the firm takes marginal revenue to be $p+q\frac{dp}{dq}$, where p is price, q is quantity produced, and the derivative is the marginal impact of quantity on market price (which for an infinitesimal perfect competitor would be zero). The Cournot assumption implies that the firm acts as if any change in its output will be unmatched by competitors, so that $\frac{dp}{dq} = \frac{1}{D'(p)}$ where $D(p)$ is the market demand curve. What is the effective supply function for the firm, which specifies what it will supply at any market price?

$$p + \frac{S(p)}{D'(p)} - MC(S(p)) = 0 \qquad (14)$$

for all positive prices p, and MC is the marginal cost function. In simple cases, this function has an explicit solution (e.g. if $MC(q) = c$ and $D(p) = p^{-\eta}$). But for that demand function and a more realistic marginal cost function, $MC(q) = \alpha\sqrt{q} + q^2$, the functional equation does not have a closed–form solution. Collocation methods could be used to solve the equation over a grid of points on the q axis for given values of α and η, and determine how sensitive the resulting approximate solution might be to, for example, the number of firms competing in the industry. Collocation methods also may be used to solve boundary value problems, which often arise in economics and finance as deterministic optimal control problems, such as arise in modeling the market for a periodically produced storable commodity.