PRELIMINARY!
NOT FOR CITATION
WITHOUT PERMISSION.

Approximation Methods and Projection Methods in Economic Analysis

KENNETH L. JUDD*

January 10, 1993

ABSTRACT. This article examines local and global approximation methods which have been used and have potential future value in economic and econometric analysis. We first review the foundations of regular and singular perturbation analysis, and asymptotic evaluation of integrals. We then discuss their applications to dynamic economic models and finite-sample econometrics. We also compare the perturbation methods used in public finance with linear-quadratic models and the ad hoc procedures used by macroeconomists. We next discuss global approximation methods, including orthogonal polynomials, interpolation theory, shape-preserving splines, and neural networks, and the related projection method for solving operator equations. We illustrate their application to dynamic economic analysis and equilibrium with asymmetric information. Finally, we discuss how the hybrid perturbation-Galerkin method combines the complementary strengths of local approximation procedures and the projection method.

CONTENTS

1. INTRODUCTION

In many economic analyses, the key technical problem is the determination of some unknown function, such as policy functions in dynamic models, equilibrium strategies in games, and inference rules in asymmetric information problems. The usual approach is to make functional form assumptions on the structural elements which

*Hoover Institution and National Bureau of Economic Research. Supported by NSF Grant SES-??.

lead to closed-form solutions. This, unfortunately, restricts the analysis to a few special cases. While these special cases may suffice for some purposes, they are generally inadequate for a robust analysis of most problems.

The alternative is to use approximation ideas to compute functions which are "close" to the true solution. In this paper, we will review two basic approaches to the approximation of functions and the approximate solution of operator equations, representing two different kinds of data and objectives. *Local approximations* take as data the value of f and its derivatives at a point x_0 and constructs a function which matches those properties at x_0. These constructions rely on various versions of Taylor's theorem and the implicit function theorem, and lead to the construction of Taylor or Padé series. These methods are called *perturbation*, or *asymptotic*, methods. The basic idea of asymptotic methods is to formulate a general problem, find a particular case which has a known solution, and then use that particular case and its solution as a starting point for computing approximate solutions to "nearby" problems. These methods are widely used in mathematical physics, particularly in quantum mechanics and general relativity theory, with much success. While economists have often used special versions of perturbation techniques, such as linearizing around a steady state and computing asymptotic distributions, they often proceed in an ad hoc and potentially invalid fashion, and have generally not exploited the full range and power of asymptotic techniques. We will discuss the differences between the mathematical literature and the economics literature, and indicate directions where perturbation analysis can advance economic analysis.

L^p Approximation takes a given function f and finds a "nice" function g which is "close to" f in the sense of some L^p norm. To compute an L^p approximation of f, one ideally needs the entire function, whereas we generally have information about f at only a finite number of values. *Interpolation* is any procedure which finds a "nice" function which exactly fits a finite set of prescribed conditions. *Regression* lies between L^p approximation and interpolation in that it uses a finite collection of data, but produces an approximation which "nearly" satisfies the data. These approximation methods form the basis for *projection* methods for solving functional equations. Projection methods have been used to solve various economic problems, ranging from dynamic growth models, dynamic games, and asset market equilibria with incomplete information.

The perturbation and projection methods of solution differ substantially in their focus and procedures. However, their strengths and weaknesses are complementary, and a combined method, called the *hybrid perturbation-Galerkin* procedure is one which holds out much promise for solving complex economic models.

These methods are important because of the increasing importance of computation in economic analysis. Many economists eschew sophisticated approximation techniques, believing that a supercomputer will solve any problem they might have.

This is not the attitude taken in other computationally intensive fields; in fact, advances in numerical software have made as much a contribution to algorithm speed as hardware advances. One reason for this improvement has been the application of basic approximation ideas, many of which we present below. It is clear from examination of the mathematical and economic literature that even a modest application of modern approximation techniques can substantially improve the efficiency of most computational methods in economics.

In this paper we shall outline asymptotic approximation techniques, L^p approximation procedures, the projection method and the hybrid perturbation-Galerkin approach, and their application to some simple growth models. While the focus on dynamic models parallels the usual applications in economics of these methods, the reader should keep in mind that these methods are potentially useful for any problem which reduces to a functional equation. The objective is to discuss basic approximation methods in an integrated fashion and give examples of their use. We will discuss both examples in the existing literature of the use of approximation methods, and also methods and applications which are potentially useful[1].

2. The Mathematical Foundations of Local Approximation Methods

Local approximation methods are based on a few basic theorems. They include the well-known Taylor's theorem and the implicit function theorem for R^n as well as the Bifurcation theorem and extensions to operators on infinite-dimensional spaces. Closely related is the notion of an asymptotic series, particularly the asymptotic evaluation of integrals. We will first state the basic theorems in this section, and give examples of their use in the next section.

2.1. Taylor Series Approximation.

The most basic local approximation is described by Taylor's Theorem:

Theorem 1. *(Taylor's Theorem) If* $f \in C^{n+1}[a, b]$ *and* $x, x_0 \in [a, b]$, *then*

$$\begin{aligned} f(x) &= f(x_0) + (x - x_0) f'(x_0) + \tfrac{(x-x_0)^2}{2} f''(x_0) \\ &\quad + \cdots + \tfrac{(x-x_0)^n}{n!} f^{(n)}(x_0) + R_{n+1}(x) \end{aligned} \tag{1}$$

where

$$\begin{aligned} R_{n+1}(x) &= \tfrac{1}{n!} \int_{x_0}^{x} (x - t)^n f^{(n+1)}(t)\, dt \\ &= \tfrac{(x-x_0)^{(n+1)}}{(n+1)!} f^{(n+1)}(\xi) , \end{aligned}$$

[1] **A note to the reviewer:** This draft focusses on the substantive issues which I want to cover and illustrative examples. It is slim on citations to economic applications of these techniques. Future versions will be more complete in terms of these citations.

for some ξ between x and x_0.

The Taylor series approximation of $f(x)$ based at x^0, (1), uses derivative information at x^0 to construct a polynomial approximation. If f is analytic on $[a, b]$ then this approximation converges to f on $[a, b]$ as n increases. Generally, this approximation is good only near x^0 and decays rapidly away from x^0.

2.2. Rational Approximation. *Padé approximation* uses the same derivative information as does a Taylor series approximation, but instead constructs a rational function to approximate f. The (m, n) Padé approximant of f at x_0 is a rational function

$$r(x) = \frac{p(x)}{q(x)} \tag{2}$$

where $p(x)$ and $q(x)$ are polynomials of degree m and n, and

$$\frac{d^k}{dx^k}(p - f\ q)(x_0), \quad k = 0, \cdots, m+n \tag{3}$$

The $m + n + 1$ derivative conditions in (3) suffice since $q(x^0)$ can be normalized to be 1. The problem of computing the coefficients of p and q is a (generally nonsingular) linear problem.

The experience is that Padé approximants are better global approximants than Taylor series approximations, that is, the error grows less rapidly as we move away from x_0. There are strong theorems confirming this for analytic functions; see Petrushev and Popov.

2.3. Implicit Function Theorem. The next important tool is the Implicit Function Theorem.

Theorem 2. *(Implicit Function Theorem) If $H(x, y) : R^n \times R^M \to R^m$ is C^1 and $H_y(x_0, y_0)$ is not singular, then there is a unique function C^0 function $h : R^n \to R^m$ such that for (x, y) near (x_0, y_0)*

$$H(x, h(x)) = 0.$$

Furthermore, if H is C^n then h is C^{n-1} and its derivatives can be computed by implicit differentiation of the identity $H(x, h(x)) = 0$.

The Implicit Function Theorem states that h can be uniquely defined for x near zero by a relation of the form $H(x, h(x)) = 0$, whenever $H_y(0, h(0))$ is not singular. This allows us to implicitly compute the derivatives of h with respect to x as a functions of x. When we combine Taylor's theorem and the Implicit Function theorem, we have a way to compute a locally valid polynomial approximation of the implicit function $h(x)$.

2.4. Bifurcation Methods. Sometimes we will want to compute an approximation to an implicitly defined function at a point where the conditions of the Implicit Function theorem do not hold, in particular when $H_y(x_0, y_0)$ is singular. In some cases, there is additional structure which can be exploited by *bifurcation* methods, to which we now turn.

Suppose that $H(x, \epsilon)$ is C^2. One way to view the equation $H(x, \epsilon) = 0$ is that for each ϵ it defines a collection of x which solves the equation. We say that ϵ_0 is a *bifurcation point* if the number of solutions to $H(x, \epsilon) = 0$ changes as ϵ passes through ϵ_0. Two situations are summarized in the following theorem.

Theorem 3. *(Bifurcation Theorem) Suppose* $H(x, 0) = 0$ *for all* x. *Furthermore, suppose that*

$$H_x(x_0, 0) = 0 = H_\epsilon(x_0, 0), \; H_{x\epsilon}(x_0, 0) \neq 0$$

for some $(x_0, 0)$. *Then, if* $H_{\epsilon\epsilon}(x_0, 0) \neq 0$, *there is an open neighborhood* $\mathcal{N}$ *of* $(x_0, 0)$ *and a function* $h(\epsilon), h(\epsilon) \neq 0$ *for* $\epsilon \neq 0$, *such that*

$$H(h(\epsilon), \epsilon) = 0 \;\; on \;\; \mathcal{N}$$

and locally $H(x, \epsilon)$ *is diffeomorphic to* $\epsilon(\epsilon - x)$ *or* $\epsilon(\epsilon + x)$. *Otherwise, if* $H_{\epsilon\epsilon}(x_0, 0) = 0 \neq H_{\epsilon\epsilon\epsilon}(x_0, 0)$, *then there is an open neighborhood* $\mathcal{N}$ *of* $(x_0, 0)$ *and a function* $h(\epsilon), h(\epsilon) \neq 0$ *for* $\epsilon \neq 0$, *such that*

$$H(h(\epsilon), \epsilon) = 0 \;\; on \;\; \mathcal{N}$$

and $H(x, \epsilon)$ *is locally diffeomorphic to* $\epsilon^3 - x\epsilon$ *or* $\epsilon^3 + x\epsilon$. *In both cases,* $(x_0, 0)$ *is a bifurcation point.*

Figure 1 shows that $\epsilon^2 - x\epsilon$ and $\epsilon^3 - x\epsilon$ have bifurcations at $(x, \epsilon) = (0, 0)$. That is, as ϵ passes through 0, the number of x which satisfy $\epsilon^2 - x\epsilon$ ($\epsilon^3 - x\epsilon$) changes. In the cases specified in the Bifurcation Theorem, we can still examine the nondegenerate branch and, as long as there are enough derivatives, use implicit differentiation to compute a polynomial approximation for the implicit function.

2.5. Generalizations to Function Spaces. To solve dynamic economic problems, we need generalizations of these theorems to functional spaces. It is necessary, therefore, to first introduce some terminology from nonlinear functional analysis. We will state a generalization which is quite similar to the implicit function for R^n, and implies a straightforward computational implementation. Suppose that X and Y are Banach spaces, i.e., normed complete vector spaces. A map $M : X^k \to Y$ is *k-linear* if it is linear in each of its k arguments. It is a *power* map if it is symmetric and

k-linear, in which case it is denoted by $Mx^k \equiv M(x, x, \ldots, x)$. The norm of M is constructed from the norms on X and Y, and is defined by

$$||M|| = \sup_{||x_i||=1,\ i=1,2,\ldots,k} ||M(x_1, x_2, \ldots, x_k)||$$

For any fixed x_0 in X, consider the infinite sum in Y:

$$Tx = \sum_{k=1}^{\infty} M_k(x - x_0)^k$$

where each of the M_k is a k-linear power map from X to Y. When the infinite series converges, T is a map from X to Y. It will be convenient to associate a real valued series, called its *majorant series*, with T

$$\sum_{k=0}^{\infty} ||M_k||\, ||x - x_0||^k$$

The important connection between the power series for T and its majorant series is that T will converge whenever its majorant series does.

Definition 4. *T is analytic at x_0 if and only if it is defined for some neighborhood of x_0 and its majorant series converges for some neighborhood of x_0.*

With these definitions, we can now state the analytic operator version of the Implicit Function Theorem.

Theorem 5. *(Implicit Function Theorem for Analytic Operators) Suppose that*

$$F(\epsilon, x) = \sum_{n,k=0}^{\infty} \epsilon^n M_{nk} x^k \tag{4}$$

defines an analytic operator, $F : U \subset X \to Y$, where U is a neighborhood of $(0, 0)$ in $R \times X$. Furthermore, assume that $F(0, 0) = 0$ and that the operator $M_{01} : X \to Y$, representing the Frechet cross-partial derivative at $(0, 0)$, is invertible. Consider the equation

$$F(\epsilon, x(\epsilon)) = 0 \tag{5}$$

implicitly defining a function $x(\epsilon) : R \to X$. The following are true:

1. *There is a neighborhood of $0 \in R$, V, and a number, $r > 0$, such that (5) has a unique solution $x(\epsilon)$ with $||x(\epsilon)|| < r$ for each $\epsilon \in V$.*

2. *The solution, $x(\epsilon)$, of (5) is analytic at $\epsilon = 0$, and, for some sequence of x_n in X, can be expressed as*

$$x(\epsilon) = \sum_{n=1}^{\infty} x_n \epsilon^n \tag{6}$$

where the coefficients x_n can be determined by substituting (6) into (5) and equating coefficients of like powers of ϵ.

3. *The radius of convergence of the power series representation in (6) is no less than that of the analytic map, $z(\epsilon) : R \to R$, defined implicitly for some neighborhood of 0 by*

$$0 = \sum_{n,k=0}^{\infty} \epsilon^n \, ||M_{nk}|| \, z(\epsilon)^k \tag{7}$$

Furthermore, for some sequence z_n of real numbers,

$$z(\epsilon) = \sum_{n=0}^{\infty} \epsilon^n \, z_n$$

represents the solution to (7) and $|z_n| > ||x_n||$.

Proof: See Zeidler (1986).

The mathematics of applying this method turns out to be elementary since the task is reduced to recursive computation of x_n terms, in term-by-term approach described above. The only requirement is to set up the problem so that it is expressed as an analytic operator with a nondegenerate radius of convergence.

There are weaker infinite-dimensional generalizations of the implicit function theorem; see Zeidler. The basic idea continues to hold: find a point where (5) holds, and implicitly differentiate the implicit function at that point to compute a local approximation. The Bunch Theorem (see Zeidler) generalizes the Bifurcation Theorem to Banach spaces. Space limitations prevent our discussing it here, but below we will see that economic applications are obvious.

2.6. Gauge Functions. The methods described above, commonly referred to as regular perturbations, compute expansions of the form $\sum_{i=1}^{\infty} a_i \epsilon^i$. There are many cases where we will want to compute different expansions. In general, a system of *gauge functions* is a sequence of functions, $\{\delta_n(\epsilon)\}_{n=1}^{\infty}$, such that

$$\lim_{\epsilon \to 0} \frac{\delta_{n+1}(\epsilon)}{\delta_n(\epsilon)} = 0$$

An *asymptotic expansion* of $f(x)$ near $x = 0$ is denoted

$$f(x) \sim f(0) + \sum_{i=1}^{\infty} a_i \, \delta_i(x)$$

where, for each n,

$$\lim_{x \to 0} \frac{f(x) - (f(0) + \sum_{i=1}^{n} a_i \, \delta_i(x))}{\delta_n(x)} = 0 \, .$$

In regular perturbations, the sequence of gauge functions is $\delta_k(\epsilon) = \epsilon^k$. Another example of a gauge system is $\delta_k(\epsilon) = \epsilon^{k/2}$. In many problems, part of the problem is in determining the correct gauge system. The next sections present examples of this more general problem.

2.7. Asymptotic Expansions of Integrals. In economic and econometric problems, integrals frequently take the form

$$\int_D e^{-\lambda g(x)} \, f(x) \, dx \tag{8}$$

where λ is a large parameter. Simply differentiating (8) with respect to λ will not work here. Laplace's method provides a useful way to approximation (8). The basic idea is that the major contribution of the integrand is at the minimum of $g(x)$. Suppose $g(x)$ is minimized at $x = a$. For large λ, if $x \neq a$ then $e^{-\lambda g(x)} \ll e^{-\lambda g(a)}$. As long as $f(x)$ does not offset this for $|x - a| \gg 0$, the integral (8) is determined largely by the behavior of the integrand, $e^{-\lambda g(x)} f(x)$, for x near a.

The one-dimensional case is easy to state. Assume that g and f satisfy the asymptotic series

$$g(x) \sim g(a) + \sum_{i=0}^{\infty} a_i (x - a)^{i+\mu}, \quad f(x) \sim \sum_{i=0}^{\infty} b_i (x - a)^{i+\alpha-1}$$

Under modest assumptions (see Wong) if the integral $I(\lambda) = \int_a^b f(x) e^{-\lambda g(x)} \, dx$ converges absolutely for sufficiently large λ, and if g is minimized on $[a, b]$ at a, then

$$I(\lambda) \sim e^{-\lambda g(a)} \sum_{i=0}^{\infty} \Gamma\left(\frac{i+\alpha}{\mu}\right) \frac{c_i}{\lambda^{(i+\alpha)/\mu}} \tag{9}$$

where $\Gamma(\lambda) \equiv \int_0^\infty e^{-x} x^{\lambda-1} \, dx$, and the c_i depend on the a_i and b_i. In particular

$$c_0 = \frac{b_0}{\mu a_0^{\alpha/\mu}}, \quad c_1 = \left(\frac{b_1}{\mu} - \frac{(\alpha+1) a_1 \, b_0}{\mu^2 a_0}\right) a_0^{-(\alpha+1)/\mu}$$

To compute these coefficients and others one essentially expands the integrand in terms of λ and matches like powers. Note that the gauge functions of λ in (9) depend on the asymptotic expansions of f and g.

One important application of Laplace's method is Stirling's formula for $n!$. Recall that $n! = \Gamma(n+1)$. We would like to approximate $\Gamma(n)$ for large n. To use Laplace's method, let $x = y\lambda$; then

$$\Gamma(\lambda) = \lambda^{\lambda} \int_0^{\infty} e^{-\lambda(y-\ln y)} y^{-1}\, dy$$

The minimum of $y - \ln y$ is at $y = 1$. Break the integral into two integrals over $[0,1]$ and $[1,\infty)$, and add the two one-sided approximations to get

$$\Gamma(\lambda) \sim \sqrt{2\pi}\ \lambda^{\lambda-\frac{1}{2}} e^{-\lambda} \left(1 + \frac{1}{12\lambda}\right)$$

Stirling's formula is just the one-term expansion $n! \simeq \sqrt{2\pi}(n+1)^{n+\frac{1}{2}} e^{-(n+1)}$. While the operating assumption is that λ is a large parameter, these expansions do very well even when λ is not small; Stirling's approximation for 1! is .9595 and the two–term expansion yields .9995.

There is a multidimensional extension of Laplace's method. Suppose $D \subset R^n$, $f, g \in C^2[D]$. Suppose the minimum of $g(x)$ for $x \in D$ is achieved at x_0 which is in the interior of D. Then the leading term of the expansion is

$$I(\lambda) = \frac{e^{-\lambda g(x_0)}}{|H|^{1/2}} \left(\frac{2\pi}{\lambda}\right)^{n/2} f(x_0)$$

where $H \equiv (g_{x_i x_j})$ is the Hessian of g. Higher order terms can be computed; while such an expansion would be very tedious to construct, symbolic languages such as Mathematica, Maple, and Macsyma, are ideally suited to do this.

2.8. Singular Perturbations. Regular perturbations and the related bifurcation procedures are relatively simple procedures, relying on standard implicit function theorems. More difficult are *singular perturbations.* Singular perturbations are problems where the fundamental nature of the problem changes as ϵ moves away from zero.

A simple example of this is the second-order differential equation

$$\epsilon\, \ddot{x} + \dot{x} + x = 0 \tag{10}$$

with boundary conditions

$$x(0) = 0, \quad x(1) = 1$$

Note that if $\epsilon = 0$ then (10) reduces to a first-order differential equation with the solution Ce^{-t}, but there is no constant C which will satisfy both boundary conditions. Therefore, there is no $\epsilon = 0$ case around which we can apply the regular perturbation method.

There are special methods for solving such problems. They revolve around using specially devised gauge families. In the case of this problem, one postulates an approximation of the form

$$\sum_{i=0}^{n} a_i(t)\epsilon^i + \sum_{i=1}^{n} b_i(\frac{t}{\epsilon})\epsilon^i + \sum_{i=1}^{n} c_i(\frac{1-t}{\epsilon})\epsilon^i \tag{11}$$

and computes the coefficient functions a_i, b_i, and c_i essentially by substituting (11) into (10), collecting terms of like powers, and solving the resulting linear ordinary differential equations. The result is a solution which approximates the true solution to $o(\epsilon^n)$.

There is no general theory for singular perturbation methods. Their application tends to be much more problem specific than regular perturbations. Fortunately, some of the problems which have been studied include optimal control problems. Bensoussan is an excellent source for this material.

2.9. The Meaning of "Approximation". We often use the phrase "$f(x)$ approximates $g(x)$ for x near x_0", but the meaning of this phrase is seldom made clear. One trivial sense of the term is that $f(x_0) = g(x_0)$. While this is certainly a necessary condition, it is generally too weak to be a useful concept. Approximation usually means at least that $f'(x_0) = g'(x_0)$ as well. In this case, we say that "f is a first-order (or linear) approximation to g at $x = x_0$". In general, "f is an n'th order approximation of g at $x = x_0$" if and only if

$$\lim_{x \to x_0} \frac{\| f(x) - g(x) \|}{\| x - x_0 \|^n} = 0$$

While this seems rather obvious, these definitions are not always used in the economics literature, as we will see below.

3. Applications of Local Methods to Economics

There have been many uses of local approximations in economics, implicit and explicit. We will review some basic applications which have appeared and give examples of some possible future uses.

Since it will be frequently used below, we will now describe a simple continuous-time[2] model of economic growth. Let k be the capital stock, c the rate of consumption, and $f(k)$ the rate of output. Assume that the intertemporal utility function of the

[2]The examples will focus on continuous-time models. Because it is obvious that all of these methods can be applied in the same way to discrete-time models, there is no substantive distinction between the discrete-time and continuous-time literatures, and I will discuss continuous-time and discrete-time papers together. The continuous-time choice is made here since the notation is cleaner.

representative agent is $\int_0^\infty e^{-\rho t}u(c(t))dt$, and that the capital stock evolves according to $\dot{k} = f(k) - c$. The optimal growth problem is

$$\begin{aligned} V(k) &\equiv \max_{c(t)} \int_0^\infty e^{-\rho t}\, u(c)\, dt \\ \dot{k} &= f(k) - c \end{aligned} \tag{12}$$

where $V(k)$ is the value function. Our examples will study the solution to the optimal growth problem as well as equilibrium growth under taxation in this model.

3.1. Perturbing Dynamic Growth Equilibria. We can apply this procedure to study the effects of policy changes in a dynamic model of equilibrium. Turnovsky and Brock (1980) shows that equilibrium with taxation in the simple growth model described above solves the system

$$\begin{aligned} \dot{c} &= \gamma(c)\, c\, (\rho - f'(k)(1-\tau)) \\ \dot{k} &= f(k) - c - g \end{aligned} \tag{13}$$

where $\gamma(c)$ is the rate of intertemporal substitution in consumption, $\tau(t)$ is the tax on capital income at time t, and $g(t)$ is government expenditure (on goods which do not affect utility) at t. The tax rates are exogenous, and c and k are the unknowns to be determined. The boundary conditions are an initial condition on the capital stock

$$k(0) = k_0 \tag{14}$$

and a stability condition on consumption

$$0 < \mid \lim_{t\to\infty} c(t) \mid < \infty \tag{15}$$

Judd (1985) showed how to analyze various policy "shocks" in this model. The conceptual experiment is as follows. We assume that the "old" tax policy was constant, $\tau(t) = \bar{\tau}$, and that it has been in place so long that, at $t = 0$, the economy is at the steady state corresponding to $\bar{\tau}$. Note that this also assumes that for $t < 0$, agents assumed that $\tau(t) = \bar{\tau}$ for all t, even $t > 0$. Hence, at $t = 0$, $k(0) = k^{ss}$. Suppose, however, that at $t = 0$, agents are told that future tax policy will be different. Say that they find out that the new tax rates are $\bar{\tau} + \tau(t)$, $t \geq 0$, that is $\tau(t)$ will be the change in the tax rate at time t. Similarly, they are told that the new expenditure policy is $\overline{g} + g(t)$. The new system is

$$\begin{aligned} \dot{c} &= \gamma(c)\, c\, (\rho - f'(k)\, (\bar{\tau} + \tau(t))) \\ \dot{k} &= f(k) - c - (\overline{g} + g(t)) \end{aligned}$$

together with $k(0) = k_0$ and (15). We will use perturbation methods to approximate the effects of the new policies τ and g.

We need to parameterize the new policy so that it fits the perturbation approach. We do this by defining

$$\tau(t,\epsilon) = \bar{\tau} + \epsilon\tau(t), \quad g(t,\epsilon) = \bar{g} + \epsilon g(t)$$

and the corresponding continuum of BVP's

$$\begin{aligned} c_t(t,\epsilon) &= \gamma(c(t,\epsilon))\, c(t,\epsilon)\, (\rho - f'(k(t,\epsilon))(1-\tau(t,\epsilon))) \\ k_t(t,\epsilon) &= f(k(t,\epsilon)) - c(t,\epsilon) \end{aligned} \tag{16}$$

where we explicitly indicate the dependence of c and k on both ϵ and t.

We differentiate (16) with respect to ϵ, evaluate the resulting differential equation at $\epsilon = 0$, and arrive at a linear differential equation. We then solve for $c_\epsilon(t,0)$ and $k_\epsilon(t,0)$, the linear approximation of the response to c and k to the shock, after imposing the boundary conditions on the initial capital stock and the asymptotic consumption level. The result will give a linear approximation for $c(t,1)$ and $k(t,1)$, the consumption and capital paths under the tax and spending changes.

The resulting solutions can be very informative. For example, in this model (see Judd (1985)) the initial shock to net investment (denoted by the derivative of $I \equiv f(k) - c$ with respect to ϵ at $t = 0$) is

$$I_\epsilon(0) = -\frac{\gamma\, c\, \rho}{1-\bar{\tau}}\, T(\mu) + \mu G(\mu) - g(0) \tag{17}$$

where

$$\mu = \frac{\rho}{2(1-\bar{\tau})}\left(1 + \sqrt{1 + \frac{4\gamma(1-\bar{\tau})\theta_L\theta_c}{\sigma\theta_K}}\right)$$

is the positive eigenvalue of the linearized system, θ_K is capital's share of income, θ_L is labor's share, and θ_c is the steady state share of output which goes to consumption. $G(s)$ and $T(s)$ are the Laplace transforms[3] of $g(t)$ and $\tau(t)$.

Perturbation methods yield algebraic formulas for quantities of interest. For example, the formula (17) tells us many things. First, future tax increases reduce investment. Second, government spending has an ambiguous impact on investment – current government spending depresses investment and future spending increases investment. Third, since investment and output are related, we also know the initial impact of this policy shock on output; that is, if current investment falls, output in

[3]If $f(t) : R^1 \to R^n$, then the Laplace transform of $f(t)$ is $L\{f\} : R^1 \to R^n$, where $L\{f\}(s) \equiv \int_0^\infty e^{-st} f(t)dt$..

the future will also fall. Note that the shock could be nonconstant, allowing us to consider partially anticipated shocks. These simple calculations address basic issues in macroeconomics; in particular, they contrast with Barro's???

While this example is quite simple, the robustness is obvious. One can add labor supply, and other tax instruments. In tax theory, Judd (1985, 1987) calculated the marginal efficiency cost of various tax innovations and related impulse responses for important macroeconomic variables. In monetary models, Wilson and Obstfeld have also applied these methods in monetary models. Laitner also applied this to tax problems.

We can also use this method to approximate solutions to the optimal growth model. If taxes and government spending is zero, then the problem reduces to the social planner's problem. We can use the same approach to compute the effect of changes in the initial capital stock on consumption; more precisely, we make k_0 in (14) a parameter and differentiate (13) with respect to k_0 near a steady state. This procedure yields the local linear approximation to the optimal policy function for consumption. Similarly, such a perturbation can also yield a local linear approximation to the equilibrium policy function with taxation. It is clear that other distortions, such as externalities and imperfect competition can also be modeled in this fashion. This method can be extended to several state variables, as Laitner has done. This simple example is just the simplest application of this method.

3.2. Perturbing Dynamic Functional Equations. A large variety of economic problems can be reduced to various kinds of functional equations, some more complex than the simple differential equations we saw above. In this section we shall take a more functional approach to a simple growth model to illustrate the general applicability of perturbation methods to those functional problems arising from dynamic programming and recursive equilibrium.

Stationary, Deterministic Growth. We will first look at a single-sector, single good, continuous-time optimal growth problem, (12). The Bellman equation defining $V(k)$ is

$$\rho V(k) = \max_c \; u(c) + V'(k)(f(k) - c). \tag{18}$$

By the concavity of u and f, at each k there is a unique optimal choice of c, which satisfies the first order condition $u'(c) = V'(k)$. We will let $C(k)$, the policy function, denote that choice. (18) implies a differential equation for $C(k)$:

$$u''(C(k))\,C'(k)(f - C(k)) + u'(C(k))(f'(k) - \rho) = 0 \tag{19}$$

At the steady state, k^{ss}, $f(k^{ss}) = C(k^{ss})$, which, when substituted into (19) implies the condition $\rho = f'(k^{ss})$ which determines k^{ss}.

Our goal is to compute the Taylor series expansion of the policy function around the steady state. Specifically, we want to compute the coefficients of

$$C(k) \doteq C(k^{ss}) + C'(k^{ss})(k - k^{ss}) + C''(k^{ss})(k - k^{ss})^2/2 + ... \tag{20}$$

We have so far computed k^{ss}, $C(k^{ss})$, and $f'(k^{ss})$. We next move to $C'(k^{ss})$. At this point we must assume that $C(k)$ is C^∞. This assumption is clearly excessive, but not unrealistic. if we also assume that $u(c)$ and $f(k)$ are also C^∞. In fact, Santos and Vila [56] shows that if u and f are C^k then the policy function is C^{k-2} near any stable steady state.

Differentiating (19) with respect to k yields[4]

$$0 = u'''C'C'(f - C) + u''C''(f - C) + u''C'(f' - C') + u''(f' - \rho) + u'f'' \tag{21}$$

which holds at each k and at the steady state, k^{ss}, reduces to

$$0 = -u''(C')^2 + u''C'f' + u'f'' \tag{22}$$

Hence $C'(k^{ss})$ must solve (22), implying

$$C' = \frac{u''f' \pm \sqrt{(u''f')^2 + 4u''u'f''}}{2u''} \tag{23}$$

where all derivatives are evaluated at the steady-state levels for the capital stock and consumption. Since u and f are increasing and concave, (23) has two real solutions of opposite signs. Since $C' > 0$ is known, we choose the positive root.

To demonstrate the ease with which higher-order terms can be calculated, we next $C''(k^{ss})$. Differentiating (21) with respect to k and imposing the steady state conditions yields an equation linear in the unknown $C''(k^{ss})$. Therefore, solving for $C''(k^{ss})$ is easier than solving for $C'(k^{ss})$. In fact, the solution for $C''(k^{ss})$ is

$$C''(k^{ss}) = \frac{2(\rho - C')u'''C'C' + 3u''C'f'' + u'f'''}{u''(3C' - 2\rho)}$$

where all functions are evaluated at k^{ss}. Note that the solution for $C''(k^{ss})$ involves $C'(k^{ss})$. The critical simplifying feature is that once we have solved the quadratic equation for $C'(k^{ss})$, we have a linear equation for $C''(k^{ss})$. Similarly, continued differentiation of (19) shows that every other derivative of C can be defined linearly in terms of the steady–state derivatives of u, f, and lower order derivatives as long as there are sufficient derivatives of u and f.

[4]We drop arguments when they can be understood from context.

Judd [38] shows that the 100 degree polynomial approximation to C is easily computed via a recursive formula, and shows that it is quite accurate even for capital stocks far away from the steady state. Judd and Guu [34] present Mathematica programs which compute arbitrary order Taylor and Padé expansions based on the derivatives of C at the steady state. The Padé expansions were found to do better than the Taylor expansions at capital stocks far away from the steady state, consistent with the standard experience for Padé approximations. They also demonstrate that high-order Taylor and Padé expansions do a very good job in approximating $C(k)$, and that the additional terms sufficiently increase the range over which the approximations are good.

Single-Sector, Stochastic Growth. We next take the deterministic model above, add uncertainty, and show how to use the approximation to the deterministic policy function around k^{ss} in the deterministic case to compute an approximate policy function in the model with a small amount of uncertainty. While the assumption of small shocks may seem limiting, it is sensible in many applications, such as macroeconomic and related financial analysis.

The stochastic problem is

$$\begin{aligned} V(k) &= \sup\ E\{\textstyle\int_0^\infty e^{-\rho t}\, u(c)\, dt\} \\ & \\ & dk = (f(k) - c) + \tfrac{1}{2}\sqrt{\sigma}\ k\, dz \end{aligned} \tag{24}$$

The Bellman equation becomes

$$0 = \max_c [-\rho V(k) + u(c) + V_k(k)\ (f(k) - c) + \frac{1}{2}\sigma\, k^2\, V_{kk}(k)]$$

It is straightforward to show that $C(k)$ solves

$$0 = \alpha(k)u'''(C(k)) + \phi(k)\ u''(C(k)) + \gamma(k)\ u'(C(k)) \tag{25}$$

where

$$\begin{aligned} \alpha(k) &= \sigma k^2\ [C'(k)]^2/2 \\ \phi(k) &= [f(k) + \sigma\ k - h(k)]\ C'(k) + \sigma k^2\ C''(k)/2 \\ \gamma(k) &= f'(k) - \rho \end{aligned}$$

Formally, we are again looking for the terms of the Taylor expansions of C,

$$\begin{aligned} C(k,\sigma) \doteq\ & C(k^{ss},0) + C_k(k^{ss},0)(k - k^{ss}) + C_\sigma(k^{ss},0)\sigma \\ & + C_{kk}(k^{ss},0)(k - k^{ss})^2/2 + C_{\sigma k}(k^{ss},0)\sigma(k - k^{ss}) \\ & + C_{\sigma\sigma}(k^{ss},0)\sigma^2/2 + \ldots \end{aligned} \tag{26}$$

Before proceeding as before, we should note that the validity of these simple methods in this case is surprising. Note that (25) is a second order differential equation when $\sigma \neq 0$, but that it degenerates to a first-order differential equation when $\sigma = 0$. Changing σ from zero to a nonzero value is said to induce a *singular perturbation* in the problem because of this change of order. Normally much more subtle and sophisticated techniques must be used to use the $\sigma = 0$ case as a basis of approximation for nonzero σ. The remarkable feature of stochastic control problems, proved by Fleming (1971), is that this is not the case, that perturbations of σ, the instantaneous variance can be analyzed as a regular perturbation in σ.

With Fleming's analysis in hand, we will now proceed. We assume that we know all the k derivatives of C at $k = k^{ss}$ and $\sigma = 0$. This is what the previous section on deterministic problems produced. We now move to computing C_σ by differentiating (25) with respect to σ. When we impose the deterministic steady state conditions $f(k^{ss}) = C(k^{ss})$, $f'(k^{ss}) = \rho$, and $\sigma = 0$, we arrive at a linear equation which implies that

$$C_\sigma = \frac{u''' C_k^2 + C_{kk}}{u'' C_k} \tag{27}$$

where all the derivatives of C are evaluated at $k = k^{ss}$ and $\sigma = 0$. Note that the solution for C_σ is a function not only of the deterministic steady state value of u, u', and u'', it also depends on u''', and C_{kk}, which in turn depends on f'''. If u were quadratic and f linear, then (27) shows that $C_\sigma = 0$, as we expect from the certainty equivalence results for linear-quadratic control. Again, continued differentiation of (25) with respect to σ and k leads to solutions for $C_{\sigma\sigma}$, $C_{\sigma k}$, $C_{k\sigma\sigma}$, etc. Judd and Guu (1992) present Mathematica programs for computing these coefficients. They also show that the approximations are valid over a substantial range of values for σ and k.

Term Structure of Interest Rates. We next examine a less standard application of the ideas of regular perturbation, the term structure of interest rates. This will be an example of perturbing a partial differential equation. It will also serve as an example of where third-order terms in u affect the first-order analysis of an quantity of direct economic importance, the interest rate, and be valuable in our discussion below of alternative ad hoc procedures.

Define $R(k, s, \sigma)$ to be the value today of a contract which promises the delivery of one dollar s periods in the future in the competitive (hence, Pareto optimal) equilibrium of the model (24); we include the parameter σ in the arguments of R to make explicit the dependence on that parameter. Application of Ito's Lemma and the martingale properties of asset prices implies that $R(k, s, \sigma)$ satisfies the linear

parabolic partial differential equation (see Judd [38] for the details)

$$R_s = R_k(f - C) - R(f' + A(C)C_k\sigma k + A(C)\frac{1}{2}\sigma k^2 C_{kk}) \tag{28}$$

where $A(c) \equiv u''(c)/u'(c)$ is the coefficient of absolute risk aversion. The boundary condition is implied by the fact that a dollar delivered today is worth a dollar at any capital stock and variance.

$$R(k, 0, \sigma) = 1$$

At the deterministic steady state, $R(k^{ss}, s, 0) = e^{-\rho s}$, the familiar equation for the price of a forward dollar when the interest rate is constant. Differentiating (28) with respect to k and σ will lead to linear differential equations for $\frac{dR}{dk}(k^{ss}, s, 0)$, $\frac{dR}{d\sigma}(k^{ss}, s, 0)$, and higher derivatives. The solutions are

$$R_k(k^*, s, 0) = \frac{f''(k^*)}{C_k(k^*, 0) - \rho}\left(e^{-C_k(k^*,0)s} - e^{-\rho s}\right)$$

Since $C_k(k^{ss}, 0) > \rho$, $R_k(k^*, s, 0) > 0$, reflecting the intuition that an increase in the capital stock reduces the marginal product of capital and, in a deterministic model, the interest rate, which then increases the present value of a forward dollar.

The dependence of R on σ is a much more complex quantity, and is locally given by

$$\begin{aligned} R_\sigma(k^{ss}, s, 0) &= -C_\sigma \int_0^s e^{\rho(\tau - s)} R_k(k^{ss}, \tau, 0)\, d\tau \\ &\quad -(AC_k\sigma k + AC_{kk}\sigma) \int_0^s e^{\rho(\tau - s)} R(k^{ss}, \tau, 0)\, d\tau \end{aligned} \tag{29}$$

The first term represents the effect of the shift in $C(k, \sigma)$ due to the uncertainty. If $C_\sigma \neq 0$, the capital stock no longer tends to drift towards k^{ss}, but some other level; hence, the interest rate is altered to reflect the changed marginal product, R_k. The second term represents a pure risk effect since it remains if $C_\sigma = 0$. Since $\sigma > 0 > A$, if C_{kk} is not too negative, this term increases R, implying a higher bond price and lower interest rate. When viewed as a function of s, $R + \sigma R_\sigma$ is the first-order approximation of the term structure of bond prices, showing that (29) allows us to isolate the importance of uncertainty and various parameters for bond prices. Similarly, one can compute first-order approximations for other asset prices, risk premia, their cyclical behavior, and any other feature of asset prices in this model.

Adjustment Cost Models. The problems above were based essentially on first-order conditions. We can apply perturbation methods to other problems which are not as simple. Dixit ?? studied the dynamics of models where a controller incurs linear adjustment costs whenever he adjusts the state. This leads to (S, s) rules. Dixit used perturbation methods to study the dependence of S and s on structural parameters when the linear cost is small.

Multidimensional Dynamics. The methods used above can be extended to the case of several state variables by applying basic linear algebra and differential equation theory. The basic idea just follows from the stable manifold theorem. Suppose we have a dynamic system

$$\dot{z} = g(z)$$

where $z = (x, y)$, x a vector of predetermined variables with initial conditions $x(0) = x_0$, and y a vector of variables whose initial values are chosen to keep z bounded. This is a common situation in dynamic growth models, with and without distortions. The predetermined variables are the state variables, such the distribution of the capital stock across sectors, or the distribution of wealth. The free variables are the decision variables and prices which are endogenous at each moment. Suppose that there is a stationary point at x^{ss}. Then the local behavior of the system is linearly approximated by the linear system

$$\dot{z} = A\, z$$

where $A = g_z(x^{ss})$. The solution is $z(t) = e^{At}\ (x_0, y_0)$, where y_0 is chosen to keep z bounded. This is done by computing the eigenvalues and eigenvectors of A and choosing y_0 so as to kill any unstable eigenvalues of A. In economic models, the dependence of y_0 on x_0 generates much valuable information, such as the dependence of prices and decisions variables on the state variables. As in the one-dimensional case, these methods can be used to compute equilibrium policy functions.

Bensoussan [4] presents the mathematical foundations for these methods in the finite-horizon case. This procedure for computing a linear approximation is presented in detail in Chapter 6 of Stokey and Lucas[59]. Judd [38] extends the procedure to higher-order terms, extending the Einstein notation to make the techniques more tractable. These linear approximations have been used in the public finance literature, such as Laitner's examination of taxation in OLG models. Judd [38], following Fleming (1971), further extends the multidimensional case to include uncertainty. These procedures have not been exploited much, but can be obviously applied to problems in the real business cycle, finance, public finance, and dynamic general equilibrium literatures.

The Macroeconomic "Linear–Quadratic Approximation". The perturbation methods described have been used to approximate a wide variety of optimal control and economic equilibrium problems, and can be used much more extensively. While macroeconomists have also studied stochastic growth models, they have eschewed the procedures above and instead either assume linear–quadratic models, as advocated in Hansen and Sargent [25], or use ad hoc procedures which replace nonlinear growth models with linear-quadratic models. Since the latter strategy bears some similarity to perturbation methods and often uses the same terminology – "linear

approximation" – we will next describe it and discuss the many differences between it and perturbation methods.

As discussed in Kydland and Prescott (1982), the basic idea to replace a stochastic nonlinear control problem with a "similar" linear-quadratic control problem which "approximates" the nonlinear model, and then apply linear-quadratic methods to solve the model. While there are some minor differences among different writers, they generally follow the procedure described precisely in McGrattan (1992). She takes the nonlinear stochastic optimal control problem

$$\begin{aligned} V(x_0) &\equiv \max_{u_t} E\left\{\sum_{t=0}^{\infty} \beta^t \pi(u,x)\right\} \\ x_{t+1} &= g(x_t, u_t, \epsilon_t) \end{aligned} \tag{30}$$

where x is a vector of state variables, u is a vector of controls, and π is concave. She solves for the steady state of the deterministic version of (30), and replaces (30) with the linear regulator problem

$$\begin{aligned} V(x_0) &\equiv \max_{u_t} E\{\sum_{t=0}^{\infty} \beta^t (x_t' Q x_t + u_t' R u_t + 2 x_t' W u_t)\} \\ x_{t+1} &= A x_t + B u_t + C \epsilon_t \end{aligned} \tag{31}$$

where both $x'Qx + u'Ru + 2x'Wu$ is the second-order Taylor expansion of π, and $Ax + Bu$ is the first-order Taylor expansion of g, both taken at the deterministic steady state.

The linear-quadratic procedure outlined in McGrattan differs from the perturbation method in its approach, objective, and results. Despite using the term "linear approximation," the objective is not to compute a locally valid Taylor series for the equilibrium rules. In fact, this procedure may produce an "approximation" which differs substantially from the Taylor series produced by perturbation methods. This is immediately seen by applying it to (24): $f''(k^{ss})$ appears in the solution to $C'(k^{ss})$ in (23), but appears nowhere in (31) after we apply McGrattan's procedure to u and f in (24); therefore, the linear decision rule computed by McGrattan would not be the linear approximation, (20), of the true decision rule even in the deterministic model[5]. In fact, those who use this procedure generally make no claim that they are computing the the linear approximation of the true decision rule.

[5]If one were to use investment instead of consumption as the decision variable in (24) then the result from McGratton's procedure and the perturbation method is the same in some cases. This does not say that McGratton's procedure is saved. Instead it points out an undesirable sensitivity to inessential details. In fact, the perturbation method is not sensitive to such changes. Many who apply the linear-quadratic procedure (as in McGratton and Christiano) do choose the variables so that g is linear in u; however, that may not always be easy to do.

Even if the McGrattan procedure were to compute the correct linear approximation for consumption near the steady state for the deterministic model, it does not always provide a correct linear approximation for the stochastic model, which is the focus of the Real Business Cycle literature. The key fact is that the linear regulator problem is certainty equivalent, that is, the variances of disturbances do not affect decision rules. Another way to interpret the McGratton procedure is to compute the decision rules in a tractable deterministic problem and use them to study a nonlinear stochastic problem. In contrast, the consumption rule in (24) generically depends on the variance, implying that $C_\sigma \neq 0$ the first-order portion of (26), implying that any linear-quadratic procedure is first-order valid only if the model is actually a linear-quadratic model to a high order. This is important for many questions; for example, we saw above that bond prices depend locally on third-order properties of utility which are absent in any linear-quadratic models. Therefore, when we replace the nonlinear control models with "similar" linear-quadratic models, some aspects of the nonlinear model are not going to be approximated reliably.

This intuitive way of approaching the problem has lead to some conceptual problems in thinking about approximations. For example, it suggests that the way to do a third-order approximation would be to take a third-order polynomial approximation around the deterministic steady state and solving exactly that control problem. Of course, there is no exact solution for third-order problems, making it appear difficult to compute a third-order approximation.

In contrast, the perturbation methods described above show that the higher-order terms are in fact easy to compute. This is a point of some controversy in the literature. Marcet [45], in his discussion of higher-order expansions, states that "perturbation methods of order higher than one are considerably more complicated than the traditional linear-quadratic case; the reason is that in linear-quadratic case we can apply certainty equivalence, and the problem is nearly a deterministic one in terms of computational costs, while the same is not true with higher order Taylor approximations". Marcet [45] gives no reason why he believes that the linear problems which produce the higher-order terms of a Taylor series approximation are "considerably more complicated" than the matrix Ricatti equations which must be solved in the linear-quadratic approximations; most numerical analysts consider linear equations simpler to solve than Ricatti equations.

Dotsey et al.[15], Christiano[10], and McGrattan[46] have documented the quality of some implementations of the macroeconomic linear-quadratic approach. The results follow what one would expect from the perturbation analysis. This method does fairly well when it comes to modeling movements of quantities, but not as well with asset prices. The approximation also breaks down as the variance of the productivity shocks increases. This is not surprising since the linear-quadratic approach ignores the effects of the variance on the decision rules. Furthermore, as we saw above, the

movements of quantities only involve first and second derivatives of the critical functions whereas first-order accurate approximations of asset prices involve higher order derivatives which are fixed at zero in linear-quadratic models.

Perturbation methods have also been overlooked when it comes to the analysis of linear-quadratic models. Many equilibrium problems do not reduce to optimal control problems, such as dynamic equilibria with taxation or money. Macroeconomists have devised complex iterative schemes for solving such problems even after imposing a linear-quadratic structure (see, for example, Cooley and Hansen[12]. These procedures are offered without any rigorous justification, and offer no reason why they are used instead of standard linearization methods. As pointed out above, the standard perturbation methods used in Laitner (1984, 1990), Judd (1982, 1985, 1987), and described in Lucas and Stokey (1989) will compute first-order valid linear approximations in nonlinear equilibrium models, and do so in a fast, nonrecursive fashion. As a special case, they will also yield first-order linear approximations in linear models where, as long as the equilibrium is globally linear, the approximations are also globally valid. McGrattan (1993) outlines a nonrecursive procedure for linear-quadratic equilibrium models with linear equilibrium rules, but gives no discussion as why it is a different or superior nonrecursive procedure than the standard perturbation approach.

Dynamic Games. Perturbation techniques can also be used to analyze dynamic games. Because of the notational burden of a formal treatment, I will here just give the basic idea behind the perturbation approach. Suffice it to say here that we are discussing dynamic game equilibrium concepts which can be written as solutions to ordinary or partial differential equations, or some similar system of functional equations.

As with any perturbation method, we begin with a "point" (possibly in a function space) where we know the solution. In game theory, such cases do arise. For example, suppose that we have two players who each influence the state variable, but that the payoff functions and the laws of motion are such that neither player is affected by the actions of the other. This would be the case of two differentiated duopolists where the cross-elasticity of demand is zero, and the state variable of the game is the vector of the firm's capital stocks. Then the equilibrium of such a "game" is trivial, reducing to an optimal control problem for each player. Those problems will possibly have stable steady states. Using the techniques above, we can compute local approximations for each player's strategy around such steady states in the degenerate game.

Now suppose that the payoffs and/or laws of motion are slightly perturbed so that each player now cares about the other's actions. By differentiating the functional equations which define equilibrium with respect to the perturbation parameter and imposing the implicit function theorem and Taylor's theorem, we will be able to compute how equilibrium is affected by the alteration.

There have been few applications of perturbation methods to game analyses thus far, but they do indicate the potential of the method. Srikant and Basar(1992) develops regular perturbation methods to various dynamic games. Their methods presume existence of equilibrium. Judd (1985) applied the Analytic Implicit Functions theorem to patent races by beginning with a patent race with a zero prize; he proves local existence as well as constructs local approximations. Budd et al. (1992) used singular perturbation methods to study a dynamic stochastic duopoly game, which reduced to a system similar to (10) above. Given the general applicability of these methods and the difficulties of game theory computation, one suspects that these procedures will become increasingly popular.

3.3. Portfolio Choices for Small Risks: A Bifurcation Application. A simple example of a bifurcation is the basic portfolio problem. Suppose that an investor has W in wealth to invest in two assets. The safe asset yields R per dollar invested and the risky asset yields $\tilde{Z}$ per dollar invested. If a proportion ω of his wealth is invested in the risky asset, final wealth is $\tilde{Y} = W((1-\omega)R + \omega\tilde{Z})$. We assume that he chooses ω to maximize $E\{u(\tilde{Y})\}$ for some concave utility function $u(\cdot)$.

One way in which economists have gained insight into this problem is to approximate u with a quadratic function and solve the resulting quadratic optimization problem. It is argued that this is valid for small risks. The bifurcation approach allows us to examine this rigorously. We first create a continuum of portfolio problems by assuming

$$\tilde{Z} = R + \epsilon\tilde{z} + \epsilon^2\pi$$

At $\epsilon = 0$, $\tilde{Z}$ is degenerate and equal to R. If $\pi > 0$, we model the intuitive case of risky assets paying a premium. Note that we multiply $\tilde{z}$ by ϵ and π by ϵ^2. Since the variance of ϵz is $\epsilon^2\sigma_z^2$, this models the observation that risk premia are the same order as the variance.

The first-order condition for ω is

$$0 = E\{u'(R + \omega(\epsilon z + \epsilon^2\pi))(z + \epsilon\pi)\} \equiv G(\omega, \epsilon) \tag{32}$$

We want to analyze this problem for small ϵ. We cannot apply the implicit function theorem since $0 = G(\omega, 0)$ for all ω implying that ω is indeterminate at $\epsilon = 0$. Since we want to solve for ω as a function of ϵ near 0, we first need to compute which of these ω values is the "correct" solution to the $\epsilon = 0$ case; specifically, we want to compute

$$\omega_0 \equiv \lim_{\epsilon \to 0} \omega(\epsilon)$$

Implicit differentiation of (32) implies

$$0 = G_\omega\,\omega' + G_\epsilon \tag{33}$$

Differentiating G we find

$$G_\epsilon = E\{u''(\tilde{Y})\,(\omega z + 2\omega\epsilon\pi)\,(z + \epsilon\pi) + u'(\tilde{Y})\pi\}$$

$$G_\omega = E\{u''(\tilde{Y})\,(z + \epsilon\pi)^2\epsilon\}$$

At $\epsilon = 0, G_\omega = 0$. $\omega'(0)$ can be well-defined in (33) only if $G_\epsilon(\omega, 0) = 0$ also. Therefore, we look for a bifurcation point, ω_0, defined by $0 = G_\epsilon(\omega_0, 0)$. At $\epsilon = 0$, this reduces to $0 = u''(R)\,\omega_0\sigma_z^2 + u'(R)\pi$, which implies

$$\omega_0 = -\frac{u'(R)}{u''(R)}\,\frac{\pi}{\sigma_z^2}$$

This is the simple portfolio rule indicating that ω is the product of risk tolerance and the risk premium per unit variance. If ω_0 is well-defined, then this must be its value. Since the conditions of the Bifurcation Theorem are satisfied at $(\omega_0, 0)$, there is a function $\omega(\epsilon)$ which goes satisfies (32) and goes through $(\omega_0, 0)$.

This is not an approximation to the portfolio choice at any particular variance. Instead, ω_0 is the limiting portfolio share as the variance vanishes. If we want the linear approximation (recall our discussion of what "approximation" means) of $\omega(\epsilon)$ at $(\omega_0, 0)$, we must go one more step since the linear approximation is

$$\omega(\epsilon) \doteq \omega(0) + \epsilon\,\omega'(0).$$

To calculate $\omega'(0)$ we need to do one more round of implicit differentiation. Differentiating (33) with respect to ϵ yields

$$0 = G_{\omega\omega}\,\omega'\omega' + 2G_{\omega\epsilon}\,\omega' + G_\omega\omega'' + G_{\epsilon\epsilon}$$

At $(\omega_0,0)$,

$$G_{\epsilon\epsilon} = u'''(R)\omega_0^2\,E\{z^3\}, \quad G_{\omega\omega} = 0, \quad G_{\omega\epsilon} = u''(R)\,E\{z^2\}$$

Therefore,

$$\omega'(0) = -\frac{1}{2}\,\frac{u'''(R)}{u''(R)}\,\frac{E\{z^3\}}{E\{z^2\}}\,\omega_0^2\,.$$

This formula tells us how the share of wealth invested in the risky asset changes as the riskiness increases, highlighting the importance of the third and second derivatives of utility and the ratio of skewness to variance. While this is a simple application of the ideas of bifurcation approximations, it is clear that it can be quite useful. For example, one could assume two types of investors and compute the equilibrium risk premium and asset allocation as a function of ϵ and initial endowment of the risky

asset. The result will show how asset prices depend in tastes, returns, and wealth distribution.

The possible uses of bifurcation methods are numerous. Another application demonstrated in Judd[38] examined the adverse selection model of Rothschild-Stiglitz-Wilson. Nor is this approach limited to relatively simple, static problems. The Bunch Theorem (see Zeidler) which generalizes the Bifurcation Theorem could presumably be used to analyze similar problems in dynamic contexts.

3.4. Econometric Applications of Local Approximations. Laplace's method focuses on computing integrals of the form (8) where the parameter λ is taken to be large. One area where this method is used is the study of finite-sample properties of estimators. In this case, the integral is the moment generating function and the parameter λ is the sample size. Phillips [51] used Laplace's method to approximate small sample marginal densities of instrumental variables estimators. Ghysels and Lieberman [20] use Laplace's method to compute small sample biases which arise from using filtered data in dynamic regressions. Laplace's method has been more popular among statisticians; see the citations in [20]. Padé approximation methods have also been used to develop finite-sample theory. Phillips [50] describes various extensions of the simple Padé approximation scheme described above and their applications to econometrics.

4. The Mathematics of L^p Approximations

We will often want to approximate functions over a broad range of values with relatively uniform accuracy. In this case, we turn to L^p approximations. *L^p approximations* finds a "nice" function g which is "close to" a given function f in the sense of a L^p norm. To compute an L^p approximation of f, one ideally needs the entire function, an informational requirement which is generally infeasible. *Interpolation* is any procedure which finds a "nice" function which goes through a collection of prescribed points. *Regression* lies between L^p approximation and interpolation in that it uses a finite collection of data, but produces an approximation which only goes near the prescribed points and has many degrees of freedom. In all cases, we need to formalize the notions of "nice" and "close to." We now examine all three basic approaches.

4.1. Orthogonal Polynomials. We will next use basic vector space ideas to construct representations of functions which will lead to good approximations. Since the space of continuous functions is spanned by the polynomials, x^n, it is natural to think of the ordinary polynomials as a basis for the space of continuous functions. However, recall that good bases for vector spaces possess useful orthogonality properties. We will develop those orthogonality ideas to construct *orthogonal polynomials*.

Definition 6. *A weighting function, $w(x)$, on $[a, b]$ is any function which is positive and has a finite integral on $[a, b]$. Given a weighting function $w(x)$, we define an inner product on integrable functions over $[a, b]$:*

$$< f, g >= \int_a^b f(x)\, g(x)\, w(x)\, dx$$

The family of polynomials $\{\varphi_n(x)\}$ are mutually orthogonal with respect to the weighting function $w(x)$ if and only if

$$< \varphi_n, \varphi_m >= 0, \quad n \neq m$$

There are several examples of orthogonal families of polynomials, each defined by a different weighting function and interval. Some common ones used in economics are Legendre, Chebyshev, Laguerre, and Hermite polynomials. Legendre Polynomials assume $w(x) = 1$ on the interval $[-1, 1]$; the n'th Legendre polynomial is

$$P_n(x) \equiv \frac{(-1)^n}{2^n n!} \cdot \frac{d^n}{dx^n} \left[(1 - x^2)^n\right]$$

The Chebyshev polynomials arise from $w(x) = (1-x^2)^{-\frac{1}{2}}$ on $[-1, 1]$; the n'th Chebyshev polynomial is

$$T_n(x) \equiv \cos(n \cos^{-1} x)$$

The Chebyshev and Legendre polynomials are useful in solving problems which live on compact sets. The Laguerre polynomials correspond to $w(x) = e^{-x}$ on $[0, \infty)$; the n'th member is

$$L_n(x) \equiv \frac{e^x}{n!} \frac{d^n}{dx^n} (x^n e^{-x})$$

Laguerre polynomials are used when one needs to approximate time paths of variables in a deterministic analysis. Hermite polynomials arise from $w(x) = e^{-x^2}$ on $(-\infty, \infty)$; the n'th member is

$$H_n(x) \equiv (-1)^n e^{x^2} \frac{d^n}{dx^n} (e^{-x^2})$$

Hermite polynomials are used to approximate functions of normal random variables.

4.2. Least-Squares Orthogonal Polynomial Approximation. Given $f(x)$ defined on $[a, b]$, one approximation concept is least-squares with respect to the weighting function $w(x)$. That is, given $f(x)$, the least-squares polynomial approximation of f with respect to weighting function $w(x)$ is the degree n polynomial which solves

$$\min_{\deg\,(p) \leq n} \int_a^b (f(x) - p(x))^2\, w(x)\, dx.$$

In this problem, the weighting function $w(x)$ indicates how we care about approximation errors as a function of x. For example, if one has no preference over where the approximation is good (in a squared-error sense) then we take $w(x) = 1$. If one cared more about the error around $x = 0$ we should choose a $w(x)$ which is larger near 0.

The connections between orthogonal polynomials and least-squares approximation are immediately apparent in solving for the coefficients of $p(x)$ in the least-squares approximation problem. If $\{\varphi_n\}_{n=1}^{\infty}$ is an orthogonal sequence with respect to $w(x)$, and we define $< f, g > \equiv \int_a^b f(x)\, g(x)\, w(x)\, dx$ the induced metric is $\| f \| \equiv < f, f >$, the least-squares solution minimizes $\| f - p \|$, and can be expressed

$$p(x) = \sum_{k=0}^{n} \frac{< f, \varphi_k >}{< \varphi_k, \varphi_k >} \varphi_k(x).$$

Note the similarity between least-squares approximation and regression. The formula for p is essentially the same as regressing the function f on $n + 1$ orthogonal regressors φ; the coefficient of the k'th regressor equals the covariance between f and the k'th regressor divided by the variance of the k'th regressor. This is no accident since regression is a least-squares approximation. The difference in practice is that regression is limited to those regressors which real data gives the analyst, whereas in approximation the analyst can choose the basis functions upon which the approximation is built.

Fourier Approximation. A particularly important form of least-squares approximation is Fourier approximation. By Fourier Approximation theorems, if f is continuous on $[0, 2\pi]$ and $f(0) = f(2\pi)$, then

$$f(\theta) = \sum_{n=0}^{\infty} A_n \cos(n\theta) + \sum_{n=0}^{\infty} B_n \sin(n\theta) \tag{34}$$

where

$$A_n = \frac{1}{\pi} \int_0^{2\pi} \cos(n\theta)\, f(\theta)\, d\theta$$

$$B_n = \frac{1}{\pi} \int_0^{2\pi} \sin(n\theta)\, f(\theta)\, d\theta$$

and convergence in (34) is uniform.

Chebyshev Approximation. We will next describe some of the features of Chebyshev approximation since they play an important role in many applications.

Theorem 7. *(Chebyshev Approximation Theorem) Assume $f \in C^r[-1, 1]$. Let*

$$C_n(x) \equiv \frac{1}{2} c_0 + \sum_{j=1}^{n} c_j T_j(x)$$

where

$$c_j \equiv \frac{2}{\pi}\int_{-1}^{1} \frac{f(x)\,T_j(x)\,dx}{\sqrt{1-x^2}}.$$

Then there is a b such that, for all $n \geq 2$

$$\| f - C_n \|_\infty \leq \frac{b \ln n}{n^r}$$

Hence $C_n \to f$ uniformly as $n \to \infty$. Furthermore, there is a constant c such that

$$|c_j| \leq c/j^r, \quad j \geq 1$$

This theorem will help us assess the quality of an approximation since both the error and the coefficients eventually drop off rapidly for smooth functions. Note that the convergence is uniform, a stronger form of convergence that the L^2 sense which initially motivated the Chebyshev approximation. Therefore, Chebyshev approximation also works for us when we want to approximate f uniformly well with polynomials.

4.3. Interpolation. Interpolation is any method which takes a finite set of pointwise restrictions and finds a function $f : R^n \to R^m$ satisfying those restrictions.

Lagrange Interpolation. Lagrange interpolation takes a collection of n points in R^2, (x_i, y_i), $i = 1, \cdots, n$, where the x_i are distinct, and finds a degree $n-1$ polynomial, $p(x)$, such that $y_i = p(x_i)$, $i = 1, \cdots, n$. The Lagrange formula demonstrates that there is such interpolating polynomial. Define

$$\ell_i(x) = \prod_{j \neq i} \frac{x - x_j}{x_i - x_j}$$

Note that $\ell_i(x)$ is unity at $x = x_i$ and zero at $x = x_j$ for $i \neq j$. This property implies that the polynomial

$$p(x) = \sum_{i=1}^{n} y_i \ell_i(x)$$

interpolates the data, that is, $y_i = p(x_i)$, $i = 1, \cdots, n$. Furthermore, this is the unique solution.

Hermite Interpolation. We may want to find a polynomial p which fits slope as well as level requirements. Suppose we have data

$$p(x_i) = y_i, \; p'(x_i) = y_i' \; , \; i = 1, \cdots, n$$

where the x_i are distinct. Since we have $2n$ conditions, we are looking for at least a degree $2n-1$ polynomial which satisfies the conditions above.

We will construct the unique solution, $p(x)$. First define the functions

$$\tilde{h}_i(x) = (x - x_i)\,\ell_i(x)^2$$

$$h_i(x) = (1 - 2\ell_i'(x)\,(x - x_i))\,\ell_i(x)^2$$

The critical facts are that h_i is a function which is zero at all x_j nodes except at x_i, where it is unity, and its derivative is zero at all x_j, and the reverse is true for $\tilde{h}_i(x)$. The unique solution to the Hermite interpolation problem is

$$p(x) = \sum_{i=1}^{n} y_i\,h_i(x) + \sum_{i=1}^{n} y_i'\,\tilde{h}_i(x)$$

4.4. Approximation Through Interpolation. Interpolation is extremely powerful since it uses a minimal amount of information to construct an approximation. However, we want the approximation to be valid generally, not just at the interpolation nodes. Consider the function $f(x) = \frac{1}{1+x^2}$ over the interval $[-5, 5]$. Let $p_n(x)$ be the n'th degree polynomial which agrees with f at the $n+1$ uniformly spaced (including the endpoints) nodes. Not only does p_n not converge to f, but for $|x| > 3.64$, $\limsup_{n\to\infty} |f(x) - p_n(x)| = \infty$. Therefore, for a seemingly well-behaved C^∞ function, interpolation at the uniformly spaced nodes does not improve as we use more points.

Interpolation Error. The last example may discourage one from approximating a function through interpolation. While the example does indicate that caution is necessary, there are some facts which reduce the likelihood of perverse behavior by interpolants. Recall that we defined $\ell_i(x) = \prod_{j\neq i} \frac{x-x_j}{x_i-x_j}$, and that the Lagrange polynomial interpolating f at points x_i is $p_n(x) = \sum_{i=1}^{n} f(x_i)\ell_i(x)$. Define

$$\Psi(x; x_1, \cdots, x_n) = \prod_{k=1}^{n} (x - x_k).$$

The following theorem uses smoothness conditions to compute a bound on the interpolation error of the Lagrange interpolation polynomial.

Theorem 8. *Assume $a = x_0 < x_1 < \cdots < x_n = b$. Then*

$$\sup_{x\in[a,b]} |\, f(x) - p_n(x)\,| \leq \|\, f^{(n+1)}\,\|_\infty\,(n!)^{-1} \sup_{x\in[a,b]} \Psi(x; x_1, \cdots, x_n) \qquad (35)$$

Chebyshev Interpolation. We will next determine a good collection of interpolation nodes. Note that our choice of $\{x_i\}_{i=1}^n$ affects only the maximum value of $\Psi(x)$, which in turn does not depend on f. So if we want to choose interpolation points so as to minimize their contribution (35), the problem is

$$\min_{x_1,\cdots,x_n} \max_x \Pi_{k=1}^n (x - x_k)$$

The solution to this problem on $[-1, 1]$ is

$$x_k = \cos\left(\frac{2k-1}{2n}\pi\right), \ k = 1, \cdots, n$$

which are the zeros of T_n. Therefore, the interpolation nodes which minimize the error bound computed above are the zeros of a Chebyshev polynomial adapted to the interval.

This shows that the Chebyshev interpolant is the best in terms of minimizing the worst-case error. Furthermore, it also keeps the maximum error, $\| f - p_n \|_\infty$, small.

Theorem 9. *(Chebyshev Interpolation Theorem) Suppose $f \in C^k[a, b]$. If I_n^f is the degree n Chebyshev interpolant, then there is some d_k such that for all n*

$$\| f - I_n^f \|_\infty \leq \left(\frac{2}{\pi} \log(n+1) + 2\right) \frac{d_k}{n^k} \| f^{(k)} \|_\infty .$$

This theorem says that the Chebyshev interpolant converges to f rapidly as we use more Chebyshev zeros. Convergence may seem to be an unremarkable property, but interpolation at uniformly spaced points does not necessarily converge as we use more points. Given these properties, Chebyshev methods are valuable whenever the approximated function is smooth.

4.5. Piecewise Polynomial Interpolation. Lagrange interpolation computes a C^∞ function to interpolate the given data. An alternative is to construct a function which is only piecewise smooth. Two common schemes are Hermite polynomials and splines.

Step Function Approximation. One common approximation strategy in economics is to use step functions. Step function approximations on $[a, b]$ are generated by a basis of step functions, $\{\varphi_i : i = 1, \cdots, n\}$ where $h = \frac{a-b}{n}$ and

$$\varphi_i(x) = \begin{cases} 0, & a \leq x \leq a + (i-1)h \\ 1, & a + (i-1)h \leq x \leq a + ih \\ 0, & a + ih \leq x < b \end{cases}$$

If the interpolation data are (x_i, y_i) and $\varphi_i(x_i) = 1$, then the step function $\sum_i y_i \varphi_i(x)$ interpolates the data. To get better approximations, one increases n.

Piecewise Linear Approximation. Piecewise linear approximations take a sequence of data, (x_i, y_i), and creates a piecewise linear function which interpolates the data. If the x_i are uniformly distributed, then they are generated by a basis of *tent functions*, that is, functions of the form, for $i = 0, \cdots, n$,

$$\varphi_i(x) = \begin{cases} 0, & a \le x \le a + (i-1)h \\ (x - (a + (i-1)h))/h, & a + (i-1)h \le x \le a + ih \\ 1 - (x - (a + (i-1)h))/h, & a + ih \le x \le a + (i+1)h \\ 0, & a + (i+1)h \le x \le b \end{cases}$$

These are called tent functions since $\varphi_i(x)$ is zero to the right of $a + (i-1)h$, rises linearly to a peak at $a + ih$, and then falls back to zero at $a + (i+1)h$, and remains zero. While both step function and piecewise linear approximations fit into our general linear approach, they differ in that the basis elements are zero over most of the domain, and at each point in the domain most basis functions are zero. This is the defining feature of *finite element* approaches to approximation. While the resulting bases are not strictly orthogonal, they are close to being so since the inner product between most basis elements is zero.

Hermite Interpolation Polynomials. Next, suppose that we have this level and slope information at $x_1, \cdots, x_n$. Within each $[x_i, x_{i+1}]$ interval, we construct the Hermite interpolation polynomial given the level and slope information at x_i and x_{i+1}. The collection of interval-specific interpolations constitute our piecewise polynomial approximation. The resulting function is a cubic polynomial almost everywhere. However, at the interpolation nodes, it is only C^1. This lack of smoothness is often undesirable and is addressed by splines.

Splines. Another piecewise smooth scheme is to construct a *spline*. A spline is any smooth function which is piecewise polynomial but also very smooth where the polynomial pieces connect. Formally, a function $s(x)$ on $[a, b]$ is a spline of order n if s is C^{n-2} on $[a, b]$, and there is a grid of points (called nodes) $a = x_0 < x_1 < \cdots < x_k = b$ such that s is a polynomial of degree at most $n - 1$ on each subinterval $[x_i, x_{i+1}]$, $i = 0, \ldots, n-1$. Note that order 2 splines are just the common piecewise linear functions.

The cubic spline (that is, of order 4) is popular. Suppose that we have Lagrange interpolation data $\{(x_i, y_i) \mid i = 0, \cdots, n\}$. The x_i will be the nodes of the spline, and we want to construct a spline, $s(x)$, such that $s(x_i) = y_i, i = 0, \cdots, n$. On each interval $[x_i, x_{i+1}]$, $s(x)$ will be a cubic $a_i + b_i x + c_i x^2 + d_i x^3$. We, therefore, have n intervals and $4n$ unknown coefficients. The interpolating conditions plus continuity at nodes implies $2n$ conditions on the coefficients:

$$y_i = a_i + b_i x_i + c_i x_i^2 + d_i x_i^3 \ , \ i = 0, \cdots, n$$

$$y_i = a_{i-1} + b_{i-1}\, x_i + c_{i-1}\, x_i^2 + d_{i-1}\, x_i^3\ ,\ i = 1, \cdots, n-1.$$

We next force the function to be C^2 at the interior nodes of the interpolation, implying $2n-2$ more conditions:

$$b_i + 2c_i\, x_i + 3d_i\, x_i^2 = b_{i-1} + 2c_{i-1}\, x_i + 3d_{i-1}\, x_i^2,\ i = 1, 2, \cdots, n-1.$$

This exhausts the interpolation and smoothness conditions, leaving 2 degrees of freedom. Various splines are differentiated by the two additional conditions imposed. One way to fix the spline is to pin down $s'(x_0)$ and $s'(x_n)$. For example, the *natural spline* imposes $s'(x_0) = 0 = s'(x_n)$. Other splines make other choices, such as setting $s'(x_0)$ and $s'(x_n)$ equal to the slope of the secant lines over $[x_0, x_1]$ and $[x_{n-1}, x_n]$. One's choice depends on the properties desired for the spline.

Splines are excellent for approximations for two general reasons. First, evaluation is cheap since splines are locally cubic. To evaluate a spline at x you must first find which interval $[x_i,\ x_{i+1}]$ contains x, then find the coefficients for the particular cubic polynomial used over $[x_i,\ x_{i+1}]$, and evaluate that cubic at x. The second reason for using splines is that good fits are possible even for functions which are not highly differentiable, or are C^∞ but have regions where the curvature is high.

Which Method to Use?. While all of these interpolation schemes will asymptotically approximate any C^2 function, each has its advantages. Chebyshev interpolation is asymptotically superior to piecewise linear and piecewise constant approximations for $f \in C^3[a, b]$. While these asymptotic results are strong, real-life computing also needs to watch the proportionality constants. Geometrically, if the curvature of f changes rapidly and one can use only a few basis elements, a piecewise linear or more general spline approximation will do better than Chebyshev interpolation.

These considerations motivate the following rule: if high-order derivatives are not large and f is smooth, use Chebyshev polynomial basis; otherwise, use a piecewise polynomial approach.

4.6. Shape-Preserving Interpolation. Above we have focussed on the pointwise convergence properties of various approximation schemes. Sometimes we will want to both interpolate data and preserve some shape in the data. For example, if the interpolation data indicates an increasing function, we may want to compute an approximation which is increasing everywhere, not only node-to-node but also between the interpolation nodes. We will discuss one simple example of shape-preserving interpolation.

Schumaker [1983] presents a particularly simple way to construct shape-preserving quadratic splines. Suppose that we have $z_1 < z_2, s_1 > s_2$ and want to find a piecewise quadratic function $s \in C^1[t_1, t_2]$ such that

$$s(t_1) = z_i,\ s'(t_i) = s_i,\ i = 1, 2$$

and, furthermore, we want s to be concave, as the data indicates is possible. Schumaker shows how to do this by adding one interpolation node in $[t_i, t_{i+1}]$. By piecing together these functions, he shows how to preserve shape globally. The shape properties which can be preserved are nonnegativity, monotonicity, and curvature. If one does not have slope information, one need only to choose the slope parameters so as to be consistent with the shape of the data.

4.7. Multidimensional Approximation. Most economic problems involve several dimensions – physical and human capital, capital stocks of competitors, wealth distribution, etc. When we attempt to approximate functions of several variables, many difficulties present themselves. We will discuss multidimensional interpolation and approximation methods, first by generalizing the one-dimensional methods via product formulations, and then by constructing inherently multidimensional schemes.

Tensor Product Bases. *Tensor product* methods build multidimensional basis functions up from simple one-dimensional basis functions. If $\{\varphi_i(x)\}_{i=1}^{\infty}$ is a basis for functions of one real variable, then the set of pairwise products, $\{\varphi_i(x)\varphi_j(y)\}_{i,j=1}^{\infty}$ is the tensor product basis for functions of two variables. To handle n dimensional problems in general, one can take all the n-wise products, and create the *n-fold tensor product* of a one-dimensional basis. The tensor approach can extend orthogonal polynomials and spline approximation methods to several dimensions. One advantage of the tensor product approach is that if the one-dimensional basis is orthogonal in a norm, the tensor product is orthogonal in the product norm. The disadvantage is that the number of elements increases exponentially in the dimension.

Complete Polynomials. There are many ways to form multidimensional bases and avoid the "curse of dimensionality." One way is to use *complete polynomial* bases, which grow only polynomially as the dimension increases. To motivate the complete polynomials, recall Taylor's theorem for R^n:

Theorem 10. *(Taylor's Theorem): Suppose $f : R^n \to R^1$, and is C^{k+1}. Then for $x^0 \in R^n$*

$$\begin{aligned} f(x) \;=\; & f(x^0) + \textstyle\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^0)\,(x_i - x_i^0) \\ & + \tfrac{1}{2} \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x^0)\,(x_i - x_i^0)\,(x_j - x_j^0) \\ & \vdots \\ & + \tfrac{1}{k!} \textstyle\sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(x^0)\,(x_{i_1} - x_{i_1}^0) \cdots (x_{i_k} - x_{i_k}^0) \\ & + \mathcal{O}\left(\| x - x^0 \|^{k+1}\right) \end{aligned}$$

Notice the terms used in the k'th degree Taylor series expansion. For $k = 1$, Taylor's theorem uses the linear functions $\mathcal{P}_1 \equiv \{1, x_1, x_2, \cdots, x_n\}$. For $k = 2$, Taylor's theorem uses

$$\mathcal{P}_2 \equiv \mathcal{P}_1 \cup \{x_1^2, \cdots, x_n^2, x_1x_2, x_1x_3, \cdots, x_{n-1}x_n\}.$$

$\mathcal{P}_2$ contains some cross-product terms, but not all; for example, $x_1x_2x_3$ is not in $\mathcal{P}_2$. In general, the k'th degree expansion uses functions in

$$\mathcal{P}_k \equiv \{x_1^{i_1} \cdots x_n^{i_n} \mid \sum_{\ell=1}^{n} i_\ell \leq k,\ 0 \leq i_1, \cdots, i_n\}.$$

The set $\mathcal{P}_k$ is the *complete set of polynomials of total degree k.*

Complete sets of polynomials are often superior to tensor products for multivariate approximation. The n-fold tensor product of $\{1, x, \cdots, x^k\}$ contains $(k+1)^n$ elements, far more than $\mathcal{P}_k$. For example, $\mathcal{P}_2$ contains $1 + n + n(n+1)/2$ elements compared to 3^n for the tensor product. Taylor's Theorem tells us that many of the tensor product elements add little to the approximation, saying that the elements of $\mathcal{P}_k$ will yield a k'th order approximation near x^0, and but that the n-fold tensor product of $\{1, x, \cdots, x^k\}$ can do no better than k'th order convergence since it does not contain all degree $k+1$ terms. This suggests that the a complete family of polynomials will give us nearly as good an approximation as the tensor product of the same order, but with far fewer elements.

Finite Element Approaches. Finite element methods use bases whose elements have small support. One simple example is bilinear interpolation. Suppose we have the values of $f(x, y)$ at $(x, y) = (\pm 1, \pm 1)$. Then, the following 4 functions form a cardinal interpolation basis:

$$\begin{array}{ll} \varphi_1(x,y) = \frac{1}{4}(1-x)(1-y), & \varphi_2(x,y) = \frac{1}{4}(1+x)(1-y) \\ \varphi_3(x,y) = \frac{1}{4}(1+x)(1+y), & \varphi_4(x,y) = \frac{1}{4}(1-x)(1+y) \end{array} \tag{36}$$

and the bilinear approximation to f on the square $[-1,1]\times[-1,1]$, which is an example of an *element*, is

$$f(-1,-1)\phi_1(x,y) + f(1,-1)\phi_2(x,y) + f(1,1)\phi_3(x,y) + f(-1,1)\phi_4(x,y).$$

The approximation is linear at each edge, but generally has a saddlepoint curvature on the interior. To interpolate data on a two-dimensional lattice, we create the bilinear approximation on each square.

Finite element methods consist of partitioning a domain into several elements, and patching together the local approximations on the elements, but this is not easy. Since we generally want the result to be a continuous function, care must be taken that resulting approximation is continuous across element boundaries. With bilinear

interpolation, this will hold since any two approximations overlap only at the edges of rectangles, and on those edges the approximation is the linear interpolant between the common vertices.

An alternative approach is to use triangular elements. Let $P_1P_2P_3$ be the canonical linear triangular element, $P_1 = (0,0)$, $P_2 = (0,1)$, and $P_3 = (1,0)$. We are looking for linear functions, $\varphi_i(x,y)$, such that $\varphi_i(P_i) = 1$ and $\varphi_i(P_j) = 0$ for $i \neq j$, $i,j = 1,2,3$. The functions

$$\varphi_1(x,y) = 1 - x - y, \quad \varphi_2(x,y) = y, \quad \varphi_3(x,y) = x$$

satisfy the cardinal interpolation conditions. The resulting function is linear in (x,y). The global approximation is obtained by triangulating the space and constructing this approximation on each triangle. One advantage of this piecewise planar approximation approach is that the result does preserve shape, both monotonicity and concavity, whereas bilinear interpolation will not preserve concavity.

If we know that we are approximating a smooth function, then the kinks at the edges of the elements may make both bilinear and piecewise planar approximation unappealing. Assuring smoothness at element boundaries is an increasingly difficult problem as we increase the desired degree of differentiability and the dimension. There is a large literature on finite element approximations of multidimensional functions (see Burnett[8]). Because the finite element method is largely aimed at engineers, it is most highly developed for two- and three-dimensional problems. For higher dimensions, economists will have to adapt existing methods, but the basic ideas will still hold: divide the space into elements, and construct low-order approximations within elements which satisfy the desired smoothness conditions across the boundaries of elements.

Neural Networks. The previous approximation procedures are based on linear combinations of polynomial and trigonometric functions. Neural networks provide us with an alternative and inherently nonlinear functional form for approximation. A *single-layer* neural network is a function of the form

$$F(x;\beta) \equiv h(\sum_{i=1}^{n} \beta_i g(x_i)) \tag{37}$$

where $x \in R^n$ is the vector of inputs and h and g are scalar functions. A common form chooses $g(x) = x$, reducing (37) to the form $h(\beta^T x)$. A *single hidden-layer feedforward* network is the form

$$F(x;\beta,\gamma) \equiv f(\sum_{j=1}^{m} \gamma_j h(\sum_{i=1}^{n} \beta_i^j g(x_i))). \tag{38}$$

The data for a neural network consists of (y, x) pairs such that y is supposed to be the output of a neural network if x is the input. This requirement imposes conditions on the parameters β in (37) and β and γ in (38). One fits single-layer neural networks by finding β to solve

$$\min_{\beta} \sum_{j} (y_j - F(x^j; \beta))^2$$

and the objective of a single hidden-layer feedforward network is to solve

$$\min_{\beta,\gamma} \sum_{j} (y_i - F(x^j; \beta, \gamma))^2.$$

which are just instances of nonlinear least squares fitting.

The approximating power of neural network approximation is indicated by theorems of Horni, Stinchcombe and White (see White [62]). Let G be a continuous function, $G : R \to R$, such that $\int_{-\infty}^{\infty} G(x)dx$ is finite and nonzero and G is L^p for $1 \leq p < \infty$. Let

$$\Sigma^n(G) = \{g : R^n \to R \mid \ g(x) = \textstyle\sum_{j=1}^{m} \beta_j G(w^j \cdot x + b_j),\ b_j, \beta_j \in R,$$

$$w^j \in R^n, w^j \neq 0,\ m = 1, 2, \cdots\}$$

be the set of all possible single-hidden layer feedforward neural networks using G as the hidden layer activation function.

Theorem 11. *Let $f : R^n \to R$ be continuous. Then for all $\epsilon > 0$ probability measure μ, and compact sets $K \subset R^n$, there is a $g \in \Sigma^n(G)$ such that*

$$\sup_{x \in K} |f(x) - g(x)| \ \leq \epsilon$$

and

$$\int_R |f(x) - g(x)|\, d\mu \quad \leq \epsilon$$

This also holds when G is a squashing functions, i.e., $G : R \to [0, 1]$, G is nondecreasing, $\lim_{x \to \infty} G(x) = 1$, and $\lim_{x \to -\infty} G(x) = 0$

Note that any squashing function is a cumulative distribution function and vice-versa. A common choice for G is the sigmoid function

$$G(x) = \frac{1}{1 + e^{-x}}$$

These are universal approximation results which justifies the use of neural network approximation, and helps explain its success. Note the simplicity of the functional forms; this simplicity makes neural network approximations easy to evaluate. The theoretical development of neural networks is proceeding, but is inherently difficult because of the nonlinearity of this approach.

5. Applications of Approximation to Dynamic Programming

Appriximation methods are a key part of most numerical procedures. They are particularly important in discrete-time dynamic programming problems. These problems are among the most useful and basic of dynamic economic problems, with well-understood theoretical properties (see Bertsekas (1976), and Bertsekas and Shreve (1978)). We will briefly discuss them and the approximation aspects of numerical dynamic programming.

Let $\pi(u, x)$ be profit flow if the state is x and the control is u. Suppose the law of motion is

$$x_{t+1} = g(x_t, u_t)$$

Then the value function, $V(x)$, solves

$$V(x) = \max_u \ \pi(u, x) + \beta V(g(x, u)) \equiv (TV)(x) \tag{39}$$

The standard theoretical procedure is to iterate on the basic functional equation. If we could handle arbitrary functions, we would start with a guess, V_0, and then compute the sequence $\{V_n\}$ generated by

$$V_n = TV_{n-1} \tag{40}$$

This procedure converges when viewed as a mapping in the space of value functions. On the computer, however, one cannot store arbitrary functions. There are several details which need to be decided to compute approximations to (40).

Since we cannot deal directly with the space of continuous functions, we focus on a finite-dimensional subspace. We will approximate $V(x)$ as a finite linear sum of basis functions.

$$V(x) \doteq \sum_{i=1}^{N} a_i \varphi_i(x) \equiv \hat{V}(x\,,\ \vec{a}) \tag{41}$$

Numerical procedures construct a $\hat{V}(x)$ which approximately satisfies the Bellman equation, (39). More specifically, the objective is to find a vector, $\vec{a} \in R^N$, such that $\hat{V}$ solves (39) as closely as possible.

The basic task is to replace T, an operator mapping continuous functions to continuous functions, with a finite-dimensional approximation, $\hat{T}$, which, since we are focusing on approximations to V of the form (41), maps R^N, the space of coefficients, into itself. We construct that map in two steps. First, we choose a finite number of points $x \in X$, and evaluate $(T\hat{V})(x)$ at $x \in X$. We will refer to this as the maximization step. The resulting values are points on the function $T\hat{V}$. Since we are approximating a continuous value function, we use that information to choose a value function of form (41) which "best" summarizes the information generated concerning $T\hat{V}$. This is the critical approximation step, and we denote the result $\hat{T}\hat{V}$. In essence,

$\hat{T}$ takes a function of form (41), $\hat{V}$, and maps it to another function of the same form, and is therefore a mapping in the space of coefficients, and the objective is to find a fixed point for $\hat{T}$ in the space of coefficients.

The details of the approximation aspects of this procedure – choosing a basis for the expression of V, choosing points X to evaluate $T\hat{V}$, and fitting the data – are important. We next discuss some basic approaches.

5.1. Discretization Methods. The simplest approximation procedure is to discretize the state space, that is, they replace the problem on a continuous state space with one with a finite number of points. This has the advantage of reducing the problem to one of finite matrices. The other advantage is that the resulting analysis does exactly solve for the solution to some similar economic problem.

While the discretization method does not obviously fit the description above, it is generally equivalent to approximating the value function with a step function[6]. However, step functions are highly inefficient ways to approximate a smooth value function. Because of this, the discretized state space method is unlikely to be of much value in economic analysis outside of some simple cases, such as those discussed in Imrohoroglu[27], Christiano[10], and Rust[55]. The impracticality of discretization is indicated by the fact that supercomputers are often used. Multidimensional problems are practically impossible, even for supercomputers, since the "curse of dimensionality" is particularly vexing for this method; if N points are used for a one-dimensional problem, then N^d points will be used for a d-dimensional problem.

5.2. Multilinear Approximation. While the discretization approach has been popular in macroeconomics, many OR researchers and economists have moved instead to continuous approximations of the value function. The simplest example of this is the DYGAM package discussed in Dantzig, et al. which used multilinear interpolation on hypercube elements when computing $\hat{V}_{n+1}$ from the information generated by $T\hat{V}_n$. In economics, Zeldes has used piecewise linear approximations.

This procedure has several advantages. Far fewer nodes are needed compared to a discretization method since the continuity of V is being exploited. There are some difficulties. First, the kinks make the optimization step more difficult, and are unrepresentative of V if V is C^2. Second, multilinear approximation generates curvature properties which may cause multiple local optima in the optimization step.

5.3. Polynomial Approximations. If a little continuity is good, then more should be better if V is sufficiently continuous. In this spirit, Bellman and Drey-

[6]After computing the solution to (40), many users then use linear interpolation to estimate the value function at points not part of the discretized state space. Since this linear interpolation is done only after the value iteration is completed, it does not affect these comments and it's contribution to improving the algorithm's accuracy is limited.

fus (1962) and Bellman et al. (1963) proposed the use of polynomials, and Daniel (1976) discussed the use of splines. Kotlikoff used polynomials in his study of saving and bequests. Judd presents and example of using a tensor-product basis of Chebyshev polynomials to solve a three-dimensional optimal growth problem. The advantages of polynomial approximations are that fewer points are evaluated and increased smoothness makes the optimization step more rapid.

There are, however, some problems which may arise with polynomial approximation which don't arise with discretization or multilinear approximation. The difficulty is that many interpolation schemes do not preserve shape. Even if we use the best possible interpolation scheme, the resulting approximation may not be good in between the nodes in X, and can lead to instabilities in the value iteration algorithm. To deal with this, Judd (1994) proposes the use of shape-preserving polynomials to construct value function approximations, and computed upper bounds on the error which are superior to those from the discretization approach. In fact, for problems with a concave (or convex) value function, polynomial approximations are clearly superior to discretization methods and multilinear approximations.

6. Projection Methods

The approximation ideas discussed naturally leads to algorithms for solving the kind of operator equations which arise in economics. They are called *projection methods.* We will describe the general projection approach for solving problems, and show that most of the techniques currently used by economists are also projection methods when viewed from the general perspective. With this common framework, we can discuss and compare many numerical methods, In particular, numerical analytic ideas will show why some methods outperform others and how to devise efficient algorithms.

The basic idea of projection techniques is to first express equilibrium as a zero of an operator, $\mathcal{N} : B_1 \to B_2$, where B_1 and B_2 are function spaces. That operator can be an ordinary differential equation, as in optimal control problems, a partial differential equation, as in continuous-time dynamic programming, or a more general functional equation, as in Euler equations expressing necessary conditions for recursive equilibria (as formulated in Prescott and Mehra[26]). Of course, space and time limitations make it impossible for computers to store and evaluate all possible elements of B_1. To make the problem tractable, projection methods focus on a finite-dimensional subspace of candidates in B_1 which can be easily represented on a computer and is likely to contain elements "close" to the true solution. The selection of this finite-dimensional space naturally exploits approximation methods. It may be difficult for the computer to compute $\mathcal{N}$, in which case we find a computable operator, $\hat{\mathcal{N}}$, which is "similar" to $\mathcal{N}$. Within the space of candidates solutions, we then find an element which is "almost" a zero of $\hat{\mathcal{N}}$.

While the basic idea is natural, there are many details. The key details are speci-

fying the subspace within which we look for an approximate solution and its computer representation, defining what "close" and "almost" mean, and finding the approximate solution. By studying these details, we will see how to implement these ideas efficiently to solve numerically interesting dynamic nonlinear economic problems.

6.1. General Projection Algorithm. We next describe the projection method in a general context. One begins with an operator equation representation of the problem, that is, one reduces the economic problem to finding an operator $\mathcal{N}$ and a function f such that equilibrium is represented by the solution to

$$\mathcal{N}(f) = 0$$

where $f : D \subset R^N \to R^M$, $\mathcal{N} : B_1 \to B_2$, and the B_i are function spaces. Typically $\mathcal{N}$ is a composition of algebraic operations, differential and integral operators, and functional compositions, and is frequently nonlinear.

We shall show how to implement the canonical projection technique in a step-by-step fashion. We first give an overview of the approach, then highlight the critical issues for each step, and discuss how the steps interact.

The first step is to decide how to represent approximate solutions. One general way is to assume that our approximation, $\hat{f}$, is built up as a linear combination of simple functions. We will also need a concept of when two functions are close. Therefore, the first step is to choose a basis and an appropriate concept of distance:

Step 1) Choose bases, $\Phi_j = \{\varphi_i\}_{i=1}^{\infty}$, and inner products, $< \cdot, \cdot >_j$, over B_j, $j = 1, 2$.

The basis over B_1 should be flexible, capable of yielding a good approximation for the solution, and the inner products should induce useful norms on the spaces.

Next, we decide how many basis elements to use and how to implement $\mathcal{N}$:

Step 2) Choose a degree of approximation n for f, a computable approximation $\hat{\mathcal{N}}$ of $\mathcal{N}$, and a collection of n functions from B_2, $p_i : D \to R^M$, $i = 1, \cdots, n$.

The approximate solution will be $\hat{f} \equiv \Sigma_{i=1}^{n} a_i \varphi_i(x)$. The convention is that the ϕ_i increase in "complexity" and "nonlinearity" as i increases, and that the first n elements are used. The best choice of n cannot be determined *a priori*. Generally, the only "correct" choice is $n = \infty$. Larger n should yield better approximations, but one is most interested in the smallest n which yields an acceptable approximation. One initially begins with small n and increases n until some diagnostic indicates little is gained by continuing. Similar issues arise in choosing $\hat{\mathcal{N}}$. Sometimes, as in section 2, we can take $\hat{\mathcal{N}} = \mathcal{N}$. The p_i will be the projection directions used to determine $\vec{a}$.

Step 1 lays down the topological structure of our approximation and Step 2 fixes the degrees of freedom of the approximation. Once we have made these basic decisions, we begin our search for an approximate solution to the problem. Since the true solution f satisfies $\mathcal{N}(f) = 0$, we will choose as our approximation some $\hat{f}$ which

makes $\hat{\mathcal{N}}(\hat{f})$ "nearly" equal to the zero function. Since $\hat{f}$ is parameterized by $\vec{a}$, the problem reduces to finding a $\vec{a}$ which makes $\hat{\mathcal{N}}(\hat{f})$ nearly zero. This search for $\vec{a}$ is the focus of Steps 3–5.

Step 3) For a guess $\vec{a}$, compute the approximation, $\hat{f} \equiv \Sigma_{i=1}^{n} a_i \varphi_i(x)$, and the residual function,

$$R(x\,;\,\vec{a}) \equiv (\hat{\mathcal{N}}(\hat{f}))\,(x).$$

The first guess of $\vec{a}$ should reflect some initial knowledge about the solution. After the initial guess, further guesses are generated in Steps 4 and 5, where we see how we use the inner product, $< \cdot, \cdot >_2$, to define what "near" means.

Step 4) For each guess of $\vec{a}$, compute the n projections,

$$P_i(\cdot) \equiv < R(\cdot\,;\,\vec{a})\,, p_i(\cdot) >_2, i = 1, \cdots, n.$$

Step 5) By iterating over steps 3 and 4, find $\vec{a}$ which sets the n projections equal to zero.

This general algorithm breaks the numerical problem into several distinct steps. It points out the many distinct techniques of numerical analysis which are important. First, in Steps 1 and 2 we choose the finite-dimensional space wherein we look for approximate solutions, hoping that within this set there is something "close" to the real solution. These steps require us to think seriously about approximation theory methods. Second, Step 4 will involve numerical integration if we cannot explicitly compute the integrals which define the projections. Third, Step 5 is a distinct numerical problem, involving the solution of a nonlinear set of simultaneous equations or the solution of a minimization problem. We shall now consider each of these numerical problems in isolation.

Choice of Basis and Inner Product. There are many criteria which the basis and inner product should satisfy. The full basis Φ_1 for the space of candidate solutions should be "rich"; in particular, it should be complete in B_1. We will generally use inner products of the form

$$< f(x)\,,\, g(x) > \equiv \int_D f(x) g(x) w(x)\, dx$$

for some weighting function $w(x) \geq 0$.

Computational considerations also play a role in choosing a basis. The φ_i should be simple to compute. They should be similar in size to avoid scaling problems. While asymptotic results such as the Chebyshev Interpolation Theorem may lull one into accepting polynomial approximations, practical success requires a basis where only a few elements will do the job. This requires that the basis elements should "look something like" the solution. In particular, our discussion of approximation

methods above shows that we should use smooth functions to approximate smooth functions. We will also see that the use of orthogonal bases will enhance efficiency and accuracy. Because of its special properties, a generally useful choice is the Chebyshev polynomial family. If, on the other hand, one has a basis which is known to efficiently approximate the solution, one should use that instead or combine it with the Chebyshev polynomials. A good choice of basis can substantially improve algorithmic performance.

Choice and Evaluation of Projection Conditions. Projection techniques include a variety of special methods. Generally we use $< \cdot, \cdot >$ to measure the "size" of the residual function, R. The general strategy is to find an $\vec{a}$ which makes R small. There are several ways to proceed.

First, we have the *least-squares* approach which chooses $\vec{a}$ so as to minimize the "weighted sum of squared residuals":

$$\min_{\vec{a}} \quad < R(x; \vec{a}),\ R(x; \vec{a}) > .$$

This replaces an infinite-dimensional operator equation with a nonlinear minimization problem in R^n. The standard difficulties may arise; for example, there may be local minima which are not global minima. However, there is no reason for these problems to arise more often here than in any other context, such as maximum likelihood estimation, where extremal problems are solved numerically.

While the least-squares method is a direct approach to making the error of the approximation small, most projection techniques find approximations by fixing n projections and making the projection of the residual function in each of those n directions zero. Formally, these methods find $\vec{a}$ such that $< R, p_i >_2 = 0$ for some specified collection of functions, p_i. Different choices of the p_i defines different implementations of the projection method.

One such technique is the *Galerkin* method. In the Galerkin method we use the first n elements of the basis for the projection directions. Therefore, $\vec{a}$ is chosen to solve the equations:

$$P_i(\vec{a}) \equiv < R(x; \vec{a}),\ \varphi_i(x) >= 0, \qquad i = 1, \cdots, n$$

Notice that here we have reduced the problem of solving a functional equation to solving a set of nonlinear equations. In some cases the Galerkin projection equations are the first-order conditions to some minimization problem, in which case the Galerkin method is also called the *Rayleigh-Ritz* method. This is not as likely to happen in economics problems because of their inherent nonlinearities.

There are obviously many ways to implement the projection idea. A collocation method takes n points from the domain D, $\{x_i\}_{i=1}^n$, and chooses $\vec{a}$ to solve

$$R(x_i; \vec{a}) = 0, \qquad i = 1, \cdots, n$$

This is a special case of the projection approach since $R(x_i\,;\vec{a})$ equals the projection of $R(x;\vec{a})$ against the Dirac delta function at x_i, $< R(x;\,\vec{a}),\ \delta(x-x_i) >_2$. *Orthogonal collocation* chooses the collocation points in a special way. The chosen x_i are the zeros of the n'th basis element, where the basis elements are orthogonal with respect to the inner product. The Chebyshev Interpolation Theorem suggests its power. Suppose we have found an $\vec{a}$ such that $R(z_i^n\,;\,\vec{a}) = 0$, $i = 1, \cdots, n$, where the z_i^n are the n zeros of T_n. As long as $R(x;\vec{a})$ is smooth in x, the Chebyshev Interpolation Theorem says that these zero conditions force $R(x;\vec{a})$ to be close to zero for all x, and that these are the best possible points to use if we are to force $R(x;\vec{a})$ to be close to zero. Even after absorbing these considerations, it is not certain that even orthogonal collocation is a reliable method; fortunately, its performance turns out to be surprisingly good.

Choosing the projection conditions is a critical decision since the major computational task is the computation of those projections. The collocation method is fastest in this regard since it only uses the value of R at n points. More generally, the projections will involve integration. In some cases one may be able to explicitly perform the integration. This is generally possible for linear problems, and possible for special nonlinear problems. However, our experience is that this will generally be impossible for nonlinear economic problems. We instead need to use numerical quadrature techniques to compute the integrals associated with evaluating $< \cdot, \cdot >$. A typical quadrature formula approximates $\int_a^b f(x)\,g(x)\,dx$ with a finite sum $\sum_{i=1}^n \omega_i f(x_i)$ where the x_i are the quadrature nodes and the ω_i are the weights. Since these formulas also evaluate R at just a finite number of points, quadrature-based projection techniques are essentially weighted collocation methods. The advantage of quadrature formulas over collocation is that information at more points is used to compute the approximation, hopefully yielding a more accurate approximation of the projections.

Finding the Solution. Step 5, which determines $\vec{a}$ by solving the projection conditions computed in Step 4, uses either a minimization algorithm (in the least-squares approach) or a nonlinear equation solver to solve the system $P(\vec{a}) = 0$. Many alternatives exist, including successive approximation, Newton's method, and homotopy methods, all of which have been used in the economics applications of the projection method. We won't discuss these details here since much of that material is covered in Chapter [the CGE chapter].

Coordination among Steps 1–5: The Importance of Conditioning. We now see what is needed for efficiency. We need basis functions which are easy to evaluate since they will be frequently evaluated. The integration in Step 4 must be accurate but fast. This can be helped by using quadrature formulas which work well with the basis. Finally, the nonlinear equation solver in Step 5 needs to be efficient and should be able to use all the information arising from Step 4 calculations. Step 5 will typically use gradient information about the integrals of Step 4. It is therefore

important to do those gradient calculations quickly, doing them analytically when possible.

A particularly important interaction is that between the formulation of $\mathcal{N}$, the choice of a basis and inner product, and the technique for solving the projection conditions. Newton-style methods for solving the system $P(\vec{a}) = 0$ will invert its Jacobian, $P_{\vec{a}}(\vec{a})$. This inversion makes the method sensitive to conditioning problems. The *spectral condition number*, defined to be the ratio of the largest and smallest (in magnitude) eigenvalues of a matrix, is a commonly used index of being nearly singular and indicates how sensitive matrix inversion is to error. If a Jacobian is nearly singular near the solution, the accuracy of the inversion will be poor due to round-off error and convergence will be slow. In particular, a condition number of 10^k tells you that an error of ϵ in specifying an inversion can yield an error of up to $10^k\epsilon$ in the computed inverse; in particular, you can lose up to k significant digits when you solve for $\vec{a}$ in a Newton step.

We now see why an orthogonal basis is going to be important. If a basis is nearly collinear, then the rows of $P_{\vec{a}}(\vec{a})$ will likely be nearly collinear, $P_{\vec{a}}(\vec{a})$ will likely have a large condition number, and large errors will likely arise in computing its inverse. Bases with just the first six ordinary polynomials can easily generate Jacobians with condition numbers on the order of 10^{10}, in which case one has possibly lost almost all significant digits on, say, a 13-digit machine, that is a machine where the machine round-off error ϵ is 10^{-13}. By choosing a basis which is orthogonal with respect to the inner product used in defining $P(\vec{a})$, one reduces the chances of poor conditioning in the Jacobian of P.

The form chosen for $\mathcal{N}$ will also have a dramatic influence on conditioning, accuracy and speed. If $\mathcal{N}$ is linear then $P(\vec{a}) = 0$ is a linear equation in $\vec{a}$, and Newton's method converges in just one step. In our economic problems, there are typically several economically equivalent operators which represent equilibrium, typically differing by nonlinear transformations. The more linear we can make $\mathcal{N}$, the better Newton's method will perform. This "linearization" idea helps us find a good form for our problems.

7. Applications of Projection Methods to Rational Expectations Models

Most methods used in numerical analysis of economic models fall within the general description above. We will see this below when we compare how various methods attack growth problems. The key fact is that the methods differ in their choices of basis, fitting criterion, and quadrature techniques. With the general method laid out, we will now report on a particularly important application to show its usefulness.

7.1. Discrete-Time Deterministic Optimal Growth. We examine optimal growth problems in discrete time and show how projection techniques can be adapted to calculate solutions. The stochastic case is one which has been studied by many others with various numerical techniques. In fact, one point we make below is that most of these procedures are really projection methods. By recognizing the common projection approach underlying these procedures, we can better understand their differences, particularly in accuracy and speed. We conjecture that the comparative performances of these various implementations of projection ideas in the discrete-time stochastic optimal growth problem is indicative of their relative value in other future problems.

We first examine a deterministic growth problem. We want to choose consumption, c_t, to maximize

$$\sum_{t=0}^{\infty} \beta^t u(c_t)$$

and where capital obeys the law of motion

$$k_{t+1} = f(k_t) - c_t$$

To calculate the optimal consumption policy, $h(k)$, it is enough to focus on the Euler equation,

$$0 = u'(h(k)) - \beta u'(h(f(k) - h(k)))\, f'(f(k) - h(k)) \equiv (\mathcal{N}(h))(k) \qquad (42)$$

We shall now describe the details of a projection approach to the problem. The domain D of our approximation will be $[k_m, k_M]$. k_m and k_M are chosen so that the solution will have k confined to $[k_m, k_M]$. In particular, $[k_m, k_M]$ must contain the steady state, a point which we can determine before calculations begin. Our approximation to h is parametrically given by

$$\hat{h}(k\,;\,\vec{a}) = \sum_{i=1}^{n} a_i \psi_i(k)$$

where n is the number of terms used. We could choose the Chebyshev polynomials $\psi_i(k) \equiv T_{i-1}(2\frac{k-k_m}{k_M-k_m} - 1)$, or the tent functions.

In this problem, $\mathcal{N}$ is a simple operator using only arithmetic operations and composition. Therefore, we can take $\hat{\mathcal{N}} = \mathcal{N}$. Since h is continuous, we define $\mathcal{N}$ to have domain and range in $C^0[k_m, k_M]$. Hence, $B_1 = B_2 = C^0[k_m, k_M]$, the continuity of $\mathcal{N}$ in the L^∞ norm following from the u, f, and $\hat{h}$ being C_1 in all their arguments. Given the Euler equation, (42), the residual function becomes

$$R(k\,;\vec{a}) = u'(h(k\,;\vec{a})) - \beta u'(h(f(k) - h(k\,;\vec{a});\vec{a}))f'(f(k) - h(k\,;\vec{a})) = \mathcal{N}(\hat{h})$$

To compute $\vec{a}$, we can do one of several things. First, we consider orthogonal collocation. We choose n values of k, denoted by k_i, $i = 1, ..., n$. We then choose $\vec{a}$ so that $R(k_i\,;\vec{a}) = 0$ for each i. Orthogonal collocation chooses the k_i to be the n zeros of ψ. The Chebyshev Interpolation Theorem strongly argues for using Chebyshev polynomials in this case. If $R(k_i\,;\vec{a}) = 0$ for each k_i, then we would like to conclude that $R(k_i\,;\vec{a})$ is the zero function on the domain D. The Chebyshev Interpolation Theorem says that this is most justified if the k_i were the Chebyshev zeros, and that if we use Chebyshev zeros, we are very likely to $R(k_i\,;\vec{a})$ to be nearly zero.

The tent function approach was used in Bizer and Judd. There the interpolation nodes were chosen to be uniformly distributed in D. The advantage of this approach is that the resulting interpolation is shape-preserving. This may be useful since we also know that h is monotone increasing. The policy functions computed in Judd(1992) using Chebyshev polynomials were increasing, and using tent functions reduced the algorithm's efficiency.

We could also implement the Galerkin method. If we use Chebyshev polynomials as a basis, then we use projections with the inner product

$$< f(k),\, g(k) > \equiv \int_{k_m}^{k_M} f(k)g(k)w(k)dk$$

where

$$w(k) \equiv (1 - (2\frac{k - k_m}{k_M - k_m} - 1)^2)^{-\frac{1}{2}}.$$

With this choice of inner product, the basis is orthogonal. The Galerkin method computes the n projections

$$P_i(\vec{a}) \equiv \int_{k_m}^{k_M} R(k;\vec{a})\,\psi_i(k)\,w(k)\,dk, \qquad i = 1, \ldots, n$$

and chooses $\vec{a}$ so that $P(\vec{a}) = 0$. Here the difficulty is that each $P_i(\vec{a})$ is an integral which needs to be computed numerically. The form of $w(k)$ implies the use of Gauss-Chebyshev quadrature. That is, we approximate $P_i(\vec{a}) = 0$ conditions with

$$\sum_{j=1}^{m} R(k_j;\vec{a})\,\psi_i(k_j)$$

for some $m > n$, with the k_j being the m zeros of ψ_{m+1}.

When we have calculated our estimate of $\vec{a}$, we would like to check if this procedure yields reliable approximations. Several diagnostics can be used to see if the proposed solution is acceptable. First, the a_k coefficients decline rapidly in k, as predicted by the Chebyshev approximation theorem. Second, the low-order coefficients should be insensitive to the choice of n. While these facts do not prove that the approximation

is good, we would be uncomfortable if the high-order coefficients were not small, or if the coefficient estimates were not stable as we increase n. We also want to examine test cases to see if the results from the projection method agree with the answer from another method known to be accurate. Judd (1992) performs these tests on a variety of empirically interesting cases, finding that the projection method applied to this model is very accurate and very fast.

7.2. Stochastic Optimal Growth. We next turn to a stochastic optimal growth model. This example will show us how to handle multidimensional problems and the conditional expectations which arise in stochastic dynamic problems. We will also be able to describe the parameterized expectations method of solving rational expectations models.

More specifically, we examine the problem

$$\max \; E\left\{\Sigma_{t=0}^{\infty} \beta^t \, u(\tilde{c}_t)\right\}$$

$$k_{t+1} = \theta_t f(k_t) - c_t$$

where θ_t is a stationary AR(1) multiplicative productivity parameter. We will assume that the productivity shock obeys $\ln \theta_{t+1} = \rho \ln \theta_t + \epsilon_{t+1}$ and that the $\epsilon_t \sim N(0, \sigma^2)$ are independent. The multiplicative specification of the shock is clearly not essential.

In this problem, both the beginning-of-period capital stock and the current value of θ are needed for a sufficient description of the state. Hence, the Euler equation is

$$u'\left(h(k,\theta)\right) = \beta \, E\left\{u'(h(\theta f(k) - h(k,\theta), \tilde{\theta}))\, \tilde{\theta} f'(\theta f(k) - h(k,\theta)) \mid \theta\right\} \qquad (43)$$

At this point, we will rewrite the Euler equation to make it more linear. We know that projection algorithms work well for linear problems. Perhaps our algorithm will do better if we make it more like a linear problem. To that end, rewrite (43) as

$$0 = h(k,\theta) - (u')^{-1}(\beta \, E\left\{u'(h(\theta f(k) - h(k,\theta), \tilde{\theta}))\, \tilde{\theta} f'(\theta f(k) - h(k,\theta)) \mid \theta\right\}) \qquad (44)$$

Note that (44) has two terms, one linear in $h(k,\theta)$, and the other is similar to a CRTS function of next period's potential consumption values.

The procedure for finding $\hat{h}$ is similar to the deterministic case. First of all, we need to approximate the policy function. Judd(1992) and Coleman(1991) use approximations of the form

$$\hat{h}(k,\theta;\mathbf{a}) = \sum_{i=1}^{n_k} \sum_{j=1}^{n_\theta} a_{ij} \psi_{ij}(k,\theta)$$

where the ψ_{ij} functions are Chebyshev functions of k and θ in Judd(1992), and tent functions of $\ln k$ and $\log \theta$ in Coleman(1991). Judd(1992) also considered complete

polynomials. Comparisons followed the considerations outlined above. Since the policy function is smooth, the smooth approximation procedures did better with the complete polynomial approach doing best, that is, the greatest accuracy per unit of computer time. Coleman's choice of a finite element approach used far more basis elements and cannot switch to a complete polynomial approach.

In their approach to the stochastic growth model, den Haan and Marcet ([24]) parameterized the policy function to be

$$h(k,\theta) = (k^{\delta_2}\theta^{\delta_3}e^{\delta_1})^{1/\gamma} = (\exp\{\delta_1 + \delta_2 \ln k + \delta_3 \ln \theta\})^{\frac{1}{\gamma}}$$

that is, they assume that log consumption is a linear function of $\ln k$ and $\log\theta$. However, this basis is not orthogonal. When they tried to improve the approximation to a quadratic form in $\ln k$ and $\log\theta$, the lack of orthogonality lead to difficulties which prevented them from improving on the linear approximation. They even argue that the collinearity of their basis elements is "a fortunate situation." and justifies their focus on the linear case. In contrast, the use of orthogonal bases in Judd (1992) and the use of a finite element approach in Coleman(1991) leads to no difficulties in finding substantially better approximations beyond the linear case. Therefore, the lack of orthogonality is just an example of a *bad* basis, not as den Haan-Marcet argues, a justification for looking at a smaller basis.

7.3. Applications to Distorted Economies. The Euler Equation approach is useful also because it can handle equilibrium problems which are *not* solutions to dynamic optimization problems. There are many such examples. In the rest of this chapter we shall explore several of them. First, there are taxes. Bizer-Judd shows that if a tax of τ is imposed on net output, $f'(k) - 1$, and the revenues are lump-sum rebated to agents, then equilibrium is the solution to

$$u'(h(k)) = \beta u'(h(f(k) - h(u))) \cdot (f'(f(k) - h(k))(1-\tau) - \tau)$$

Note that this is mathematically of the same form as the Euler equation for the optimal growth policy. Actually we can get much more complicated by having investment tax credits, random taxes, and noneconomic depreciation.

Another example is that of externalities. Suppose

$$y = f(k)\, g(K)$$

where k is the capital-labor ratio at a firm, y is output per unit labor, and K is the social average capital-labor ratio. Then optimality and the condition $k = K$ would imply that the socially optimal plan would satisfy the condition

$$u'(h(k)) = \beta u'(h(k^+))\,(f'(k^+)g(k^+) + f(k^+)g'(k^+))$$

$$k^+ = f(k)g(k) - h(k)$$

On the other hand, competitive equilibrium wherein individuals would ignore the effect their private capital has on others' productivity, would imply the condition

$$u'(h(k)) = \beta u'(h(k^+))\, f'(k^+)g(k^+)$$

the difference arising from the fact that individual agents would not take into account the marginal externality of investment, $f(k^+)g'(k^+)$. Pete Klenow ([39]) used projection methods to analyze this problem.

7.4. Equilibrium Growth in Continuous Time. Projection methods have been used to solve Euler equations for dynamic models. This includes both optimal growth and equilibrium growth with distortions. One simple example is the canonical continuous-time optimal growth problem described above in (19), which reduced to solving the differential equation:

$$0 = C'(k)\,(f(k) - C(k)) - \frac{u'(C(k))}{u''(C(k))}(\rho - f'(k)) \equiv \mathcal{E}\,(k;\, C)$$

Judd (1990) used a basis of Chebyshev polynomials to approximate $C(k)$, $\hat{C}(k,a) \equiv \sum_{i=0}^{n-1} a_i T_i(k)$, on an interval of capital stocks bounded below by capital stock equal to a quarter of the steady state and above by capital stock equal to twice the steady state. He then used a collocation method to solve for $\vec{a}$ in the system of equations $\mathcal{E}(k_i\ ;\ \hat{C}(k_i, a)) = 0$ where the k_i are the n zeros of the n'th degree Chebyshev polynomial adapted to the interval of capital stocks examined. Again, the performance of the algorithm was very good, independent of the details of the implementation. In fact, it easily outperformed the more commonly used shooting approach to the problem. Judd also extended this model to allow for taxation and uncertainty in continuous time. In all cases, accurate results were obtained quickly.

7.5. Information and Asset Markets. All of the examples discussed above reduced to applying the projection method to standard mathematical problems – ordinary and partial differential equations and integral equations. To demonstrate the flexibility of the projection method, we next examine a very different kind of problem– asset market equilibrium with imperfect information. In fact, this problem does not reduce to any of the standard operator problems discussed in applied mathematical literature.

Asset market equilibrium with imperfect information have been rigorously studied in recent years. Grossman (1976) and Grossman and Stiglitz (1980) began a long literature on the partial equilibrium analysis of security markets with asymmetric

information. However, much of this literature makes very special and simple assumptions about the distribution of returns, the information asymmetries, investor tastes, and asset structure. Furthermore, most analyses add a group of traders, called "noise" or "liquidity" traders whose demand is insensitive to prices and their information content. Together, these restrictions substantially limit the generality of the results and the range of questions which can be addressed. Recently, projection methods have been used to analyze these models without special functional form assumptions.

Information and Asset Demand. A simple one-period investment problem illustrates the method. Each investor may invest in two assets. The safe asset pays out R dollars per dollar invested, and the basic risky asset (we will call it stock) pays out $\tilde{Z}$ dollars per share. We also assume that each investor of type has an initial endowment of cash and shares, and that there are only two periods: a trading period followed by consumption. Therefore, if an investor begins the first period with W dollars in cash and ω_0 shares of stock, and ends the first period with ω shares of stock which trade at a price of p dollars per share, his second period consumption will be

$$\tilde{c} = (W - (\omega - \omega_0)p)\,R + \omega\tilde{Z}.$$

The first-order condition for the choice of ω will be

$$0 = E\{u'(\tilde{c})\,(\tilde{Z} - pR) \mid I\} \tag{45}$$

where I is the investor's information set. This says that the excess return should be uncorrelated with the marginal utility of consumption when conditioned on an investor's information set.

While this structure is rather simple, it is arbitrary in the number of investors, the distribution of $\tilde{Z}$, and the information allocation of investors.

Computing Conditional Expectations. The conditional expectation in (45) implies that our equilibrium concept involves a conditional expectation. Numerical implementation of the conditional expectation conditions is the most challenging aspect of this problem. We use Gaussian quadrature methods combined with basic projection ideas to implicitly compute conditional expectations.

To solve this problem, we use the following definition of conditional expectation:

$$Z(X) = E\{Y \mid X\}$$

if and only if

$$E\{(Z(X) - Y)f(X)\} = 0$$

for all bounded measurable functions, $f(X)$, of X. Intuitively, this says that the prediction error of the conditional expectation, $E\{Y \mid X\}$, is uncorrelated with any

measurable function of the conditioning information, X. This definition replaces the conditional expectation with an infinite number of unconditional expectation conditions.

Computing an Asymmetric Information Rational Expectations Equilibrium. We will assume three types of investors, all with different information and possibly different tastes. Type i investors observe y_i. The uncertain second-period consumption of a type i investor is given by:

$$\tilde{c}_i = \theta_i \tilde{Z} + (W_i + p(\bar{\theta}_i - \theta_i))R$$

where p is the share price of stock, θ_i is the number of shares held by type i agents after trading, and $\bar{\theta}_i$ is the type i endowment of stock. A type i investor's problem is to solve

$$\max_{\theta_i} E\{U(\tilde{c}_i) \mid y_i, p\} \tag{46}$$

where a type i investor's conditioning information at period 1 includes his private signal, y_i, and the price, p.

The state of the market includes all private signals, $y = (y_1, y_2, y_3)$, but each investor sees only the market-clearing price and his own information. Therefore, a rational expectations equilibrium includes a price function $p(y)$ and type-specific demand policy functions, $\theta_i(p, y_i)$ for $i = 1, 2, 3$, such that given $p(y)$, θ_i solves (46) for $i = 1, 2, 3$, and $\sum_{i=1}^{3} \theta_i(y_i, p(y)) = 1$ for all states y.

In their solution to one specification of this model[7], Judd and Bernardo (1993) approximate the price law with the polynomial

$$p(y_1, y_2, y_3) = \sum_{0 \le j+k+l \le N_p} a_{jkl} H_j(y_1) H_k(y_2) H_l(y_3)$$

where $H_i(\cdot)$ denotes the degree i Hermite polynomial and N_p represents the total degree of the polynomial approximation. Similarly, they represented the stock demand for a type i investor by

$$\theta_i(p(y), y_i) = \sum_{0 \le j+k \le N_\theta} b^i_{jk} H_j(p(y)) H_k(y_i), \quad i = 1, 2, 3 \tag{47}$$

To determine the unknown a_{jkl} and b^i_{jk} coefficients, they impose projection conditions on the investors' first-order conditions. The first-order-condition for a type i investor is given by

$$E_{y,Z}\{U'(\tilde{c}_i)(\tilde{Z} - pR) \mid y_i, p\} = 0, \quad i = 1, 2, 3. \tag{48}$$

[7]The approach in Judd and Bernardo is similar to the theoretical analysis of Anderson and Sonnenschein [1].

Using the definition of conditional expectation given above they impose projection conditions of the form

$$E_{y,Z}\{U'(\tilde{c}_i)(\tilde{Z} - p(y)R)H_j(p(y))H_k(y_i)\} = 0, \quad . \tag{49}$$

for various choices of $j, k \geq 0$. The condition in (49) states that the product of the excess return and the marginal utility of consumption for a type i agent is uncorrelated with $H_j(p(y))H_k(y_i)$. To impose market clearing, they imposed $\theta_3(p(y), y_3) = 1 - \theta_1(p(y), y_1) - \theta_2(p(y), y_2)$.

After imposing a sufficient number of such conditions, the result is a system of projection conditions constituting a finite nonlinear system of algebraic equations, thereby reducing an infinite dimensional functional problem to a finite-dimensional algebraic problem. The projection conditions given in equations ((49) are only part of the conditional expectation condition given in equation (48). The hope is that a small number of projections can yield a useful approximation. Judd and Bernardo (1993) document the accuracy for this approximation method for a variety of distributions. Overall, their experience is that this method is fast and reliable.

At this point, we should also note the applicability of Laplace's method to this problem. The projections (49) are often of moderate dimension, making conventional numerical integration rather costly. Since high accuracy is needed to compute the unknown coefficients, sampling methods of integration will not suffice. One way to economize on the numerical integration is to use Laplace's method, which is natural here since the projections are all expectations and the variance is often small. One can use as a control variate the approximation for the integrand which Laplace's method implicitly constructs, and then use conventional quadrature procedures to estimate the residual.

7.6. Convergence Properties and Accuracy of Projection Methods. When using numerical procedures, it is desirable to know something concerning its errors. An important focus of theoretical numerical analysis is the derivation of bounds on errors. Two kinds of error results are desirable. First, it is desirable to derive an upper bound on the error for a given level of approximation. Second, if such upper bounds are not possible, it may still be valuable to know that the error goes to zero asymptotically, that is, as one lets the degree of approximation become arbitrarily large. The first kind of error information is rarely available. More typical in numerical algorithms for differential equations are asymptotic results. There has been little work on proving that the algorithms used by economists are asymptotically valid.

When using numerical procedures, it is desirable to know something concerning the error of the solution. An important focus of theoretical numerical analysis is the derivation of bounds on errors. Two kinds of error results are desirable. First, it is desirable to derive an upper bound on the error for a given level of approximation.

Second, if such upper bounds are not possible, it may still be valuable to know that the error goes to zero asymptotically, that is, as one lets the degree of approximation become arbitrarily large. For most methods, the first kind of error information is rarely available. More typical in numerical algorithms for differential equations are asymptotic results. There has been little work on proving that the algorithms used by economists are asymptotically valid. Miranda (1977) has made progress in this direction with finite-element methods in some models.

Fortunately, there are general theorems concerning the consistency of the Galerkin method. Recall that the Galerkin method takes projections of the residual function against the basis elements, and the integrals are theoretically exact. Zeidler [36, 37] proves consistency for the Galerkin method assuming that the nonlinear operator $\mathcal{N}$ is monotone, coercive, and satisfies a growth condition. Galerkin methods are quite natural for computational purposes since a common theoretical way to prove the existence of a solution to an operator is to prove the existence of a solution to an infinite collection of projection conditions. In fact, Zeidler shows that if these conditions are satisfied one simultaneously proves the existence of a (weak) solution and the consistency of the Galerkin method. Similarly, using degree theory, Krasnosel'skii and Zabreiko [20] demonstrate consistency for a more general set of projection methods (possibly including Galerkin methods which use numerical quadrature).

It is unknown whether the operators used above satisfy the sufficient conditions discussed in Zeidler, Krasnosel'skii and Zabreiko, and elsewhere. Even though it remains to be seen whether these theorems do cover our problems, they do indicate that projection methods are potentially valid for our economic problems. They also point us in potential fruitful directions for proving both existence results for the underlying operators, and consistency results for alternative solution methods.

8. Hybrid Perturbation-Galerkin Method

We have discussed both perturbation and projection methods for solving economic models. While they are different approaches to approximation problems, we will next describe a method, the *hybrid perturbation-Galerkin* method which synergestically exploits their differences and similarities.

Suppose that there are a continuum of problems indexed by a parameter ϵ. Suppose that the continuum of problems to be solved has the form

$$\mathcal{N}(f(x,\epsilon);\epsilon) = 0$$

and that we can solve the $\epsilon = 0$ instance. The result of applying perturbation methods is the calculation of a series of the form

$$f(x,\epsilon) \sim \sum_{i=0}^{n} \delta_i(\epsilon)\,\varphi_i(x) \tag{50}$$

where the $\varphi_i(x)$ functions are computed by the perturbation calculations and the $\delta_i(\epsilon)$ are the (often prespecified) gauge functions. Similarly, the result of a projection approach is an approximation of the form

$$f(x,\epsilon) \simeq \sum_{i=0}^{n} a_i(\epsilon)\varphi_i(x) \tag{51}$$

where the $\varphi_i(x)$ functions are the (prespecified) basis elements of the approximation system and the $a_i(\epsilon)$ coefficients computed by the projection method. The strength of perturbation methods is that the approximations are quite good for small ϵ, but the weakness is that the quality may not hold up globally. The projection approach tries to be good for any ϵ, but the difficulty is finding good bases which will allow the series in (51) to be short. Therefore, the strengths and weaknesses of these methods are complementary.

This observation turns out to be substantive. The idea of the hybrid perturbation-Galerkin method is to use the $\varphi_i(x)$ functions from perturbation calculations as the basis functions to be used in a projection method. We know that these functions constitute an optimal basis for small ϵ, and that the optimal weight on $\varphi_i(x)$ on is $\delta_i(\epsilon)$ for small ϵ. The conjecture is that the φ_i functions still form a good basis for approximating $f(x,\epsilon)$ but that the weight on φ_i should not be the prespecified $\delta_i(\epsilon)$ but rather should be computed by (51).

Our continuous-time growth model gives a simple example of this approach. Consider the continuum of problems

$$0 = C'(k,\epsilon)\left(f(k,\epsilon) - C(k,\epsilon)\right) + \gamma C(k,\epsilon)\left(\rho - f'(k,\epsilon)\right)$$

where γ is the constant relative risk aversion parameter, and

$$f(k,\epsilon) = (1-\epsilon)\rho k + \epsilon k^{\alpha}\rho/\alpha$$

At $\epsilon = 0$, we have a linear production function with a marginal product of capital equal to ρ, the pure rate of time preference; in this degenerate case, the solution is $C(k,\epsilon) = \rho k$, that is, consumption equals output. At all other values for ϵ, the production function is concave and the unique steady state is $k = 1$. Suppose that the case we are really interested in is the $\epsilon = 1$ case, where f is the standard Cobb-Douglas production function. The hybrid perturbation-Galerkin approach is to use the perturbation results around $\epsilon = 0$ to generate basis elements which can be used in a projection method to solve the Cobb-Douglas case.

The first perturbation implies that

$$0 = C'_{\epsilon}(f(k,\epsilon) - C) + C'(f_{\epsilon} - C_{\epsilon}) + \gamma C_{\epsilon}(k,\epsilon)\left(\rho - f'(k,\epsilon)\right) + \gamma C(k,\epsilon)\left(-f'_{\epsilon}(k,\epsilon)\right)$$

which at $\epsilon = 0$ reduces to imply

$$C_\epsilon = k^\alpha \rho(\alpha^{-1} - \gamma) + (\gamma - \rho)k$$

Note that this function has a singularity at $k = 0$, a feature which is probably also true of the solution. This feature is absent in the orthogonal bases we discussed above. We see here already that this procedure has produced a basis element which has some advantages. Judd (1994) discusses further the usefulness of this approach to producing bases. While this method may not be useful in simple one-dimensional problems, it has substantial potential in multidimensional problems where economizing on the basis size is important.

Continuing the perturbation approach will generate a series of functions which can be used as a basis for a projection approach. These basis elements are possibly going to be collinear. However, for any specified inner product, we can use a standard Gram–Schmidt procedure to construct a basis which spans the same space and is orthogonal. In this way, we can combine the conditioning advantages of orthogonal bases with the desirable shape properties of the perturbation functions.

The hybrid perturbation-Galerkin method also points out the value of combining methods. Since economics problems do not fit into standard mathematical classifications, it is likely that skillful combinations of various techniques will prove to be a powerful technique.

9. Conclusions

We have shown that a general class of techniques from the numerical partial differential equations literature can be usefully applied and adapted to solve nonlinear economic growth problems. Despite the specificity of the applications discussed here, the general description makes clear the general usefulness of projection methods for economics. The application of perturbation and projection methods and the underlying approximation ideas have already substantially improved the efficiency of economic computations. Further exploitation of these ideas will surely lead to further progress.

References

[1] Anderson, R. M. and H. Sonnenschein (1982): "On the Existence of Rational Expectations Equilibrium," *Journal of Economic Theory* 26, 261–278.

[2] M. Baxter, M. J. Crucini, and K. G. Rouwenhorst, Solving the stochastic growth models by a discrete-state-space, euler-equation approach, *Journal of Business and Economic Statistics*, 8 (1990), 19–21.

[3] Budd, Christopher, Christopher Harris, and John Vickers. "A Model of the Evolution of Duopoly: Does the Asymmetry between Firms Tend to Increase or Decrease?" *Review of Economic Studies* 60 (No. 3, July 1993): 543–573.

[4] A. Bensoussan, "Perturbation Methods in Optimal Control," John Wiley and Sons, 1988.

[5] D. S. Bizer, and K. L. Judd, Uncertainty and taxation, *American Economic Review*, **79** (1989), 331-336.

[6] W. A. Brock, and L. J. Mirman, Optimal economic growth and uncertainty: the discounted case, *Journal of Economic Theory* **4** (1972), 479–513.

[7] de Boor, C. *A Practical Guide to Splines*, New York, NY: Springer, 1978.

[8] D. S. Burnett, "Finite Element Analysis," Addison-Wesley, Reading, Massachusetts, 1987.

[9] Cheney, E.W. *Introduction to Approximation Theory*, New York, NY: McGraw-Hill, 1966.

[10] L. J. Christiano, Solving the stochastic growth model by linear-quadratic approximation and by value-function iteration, *Journal of Business and Economic Statistics*, **8** (1990), 23–26.

[11] W. J. Coleman, II, Solving the stochastic growth model by policy function iteration, *Journal of Business and Economic Statistics*, **8** (1990), 27–29.

[12] "The Inflation Tax in a Real Business Cycle Model," *American Economic Review* **79** (1989): 733–748.

[13] Cuyt, A. and Wuytack, L. *Nonlinear Numerical Methods: Theory and Practice*, Amsterdam: North-Holland, 1986.

[14] Diaz-Gimenez, Javier, and Prescott, Edward C. "Asset Returns in Computable General Equilibrium: Heterogeneous Agent Economies." Manuscript. (March, 1990).

[15] Cotsey, Michael, adn Ching Sheng Mao. "How Well Do Linear Approximation Methods Work?" *Journal of Monetary Economics* 29 (1992): 25–58.

[16] C. A. J. Fletcher, "Computational Galerkin Techniques" Springer-Verlag, New York, 1984.

[17] Fleming, Wendell. "Stochastic Control for Small Noise Intensities," *SIAM Journal of Control* 9 (No. 3, August, 1971).

[18] Friedman, Avner. "Stochastic differential games." *Journal of Differential Equations* 11 (1972): 79–108.

[19] Gallant, A.R. (1987): "Identification and Consistency in Seminonparametric Regression," in T. Bewlely ed., *Advances in Econometrics Fifth World Congress.* New York: Cambridge University Press, pp. 145–170.

[20] Ghysels, Eric, and Offer Lieberman. "Dynamic Regression and Filtered Data Series: A Laplace Approximation to the Effects of Filtering in Small Samples," University of Montreal, 1993, mimeo.

[21] Grossman, Sanford. "On the Efficiency of Competitive Stock Markets Where Agents Have Diverse Information," *Journal of Finance* 18 (1976): 81–101.

[22] Grossman, S. J., and J. E. Stiglitz (1980): "On the Impossibility of Informationally Efficient markets," *American Economic Review*, 70: 393–408.

[23] R. L. Gustafson, Carryover levels for grains: a method for determining amounts that are optimal under specified conditions, USDA Technical Bulletin 1178, 1958.

[24] W. J. den Haan, and A. Marcet, Solving the stochastic growth model by parameterizing expedtations, *Journal of Business and Economic Statistics*,**8** (1990), 31–34.

[25] Hansen, Lars P., and Thomas J. Sargent. *Recursive Linear Models of Dynamic Economies*, manuscript, Hoover Institution, 1990.

[26] Huffman, Gregory W. "A Dynamic Equilibrium Model of Asset Prices and Transaction Volume," *Journal of Political Economy* 95 (No. 1, February 1987): 138–159.

[27] Imrohoroglu, Ayse. "The Welfare Cost of Inflation"

[28] Judd, Kenneth L. "An Alternative to Steady-State Comparisons in Perfect Foresight Models," *Economics Letters* 10 (1982): 55–59.

[29] Judd, Kenneth L. "Short-run Analysis of Fiscal Policy in a Simple Perfect Foresight Model", Journal of Political Economy 93 (April, 1985), 298-319.

[30] Judd, Kenneth L. "Debt and Distortionary Taxation in a Simple Perfect Foresight Model", *Journal of Monetary Economics* 20 (1987): 51–72

[31] Judd, Kenneth L. "Welfare Cost of Factor Taxation in a Perfect Foresight Model", *Journal of Political Economy* 95 (August, 1987).

[32] K. L. Judd, Asymptotic methods in dynamic economic models, mimeo, Hoover Institution, October, 1990.

[33] K. L. Judd, Minimum weighted residual methods for solving dynamic economic models, Hoover Institution, 1990.

[34] Judd, Kenneth L., and Sy-Ming Guu. "Perturbation Solution Methods for Economic Growth Models," in *Economic and Financial Modeling with Mathematica*, Hal Varian, ed., Springer-Verlag: New York, 1993.

[35] Judd, Kenneth L. "Projection Methods for Solving Aggregate Growth Models", *Journal of Economic Theory* 58 (No.2, December, 1992): 410–452.

[36] Judd, Kenneth L., and Antonio Bernardo. "Projection Methods for Computing Conditional Expectations," (1993) mimeo.

[37] Judd, Kenneth L. "Numerical Dynamic Programming with Shape-Preserving Approximation Methods," (1994), mimeo.

[38] Judd, Kenneth L. *Numerical Methods in Economics*, MIT Press (forthcoming).

[39] P. J. Klenow, "Externalities and Business Cycles," Ph.D. thesis, Department of Economics, Stanford University, June 1991.

[40] M. A. Krasnosel'skii, and P. Zabreiko, "Geometrical Methods of Nonlinear Analysis," Springer-Verlag, 1984.

[41] Kydland, Finn E., and Edward C. Prescott. "Time to Build and Aggregate Fluctuations," *Econometrica* 50 (1982), 1345-1370.

[42] Kyle, A. S. (1985): "Continuous Auctions and Insider Trading." *Econometrica*, Vol. 53, 1315–1335.

[43] Laitner, John. "Tax Changes and Phase Diagrams for an Overlapping Generatios Model," *Journal of Political Economy* **96** (No. 1, February 1990):193–220.

[44] Lucas, Robert E. "Asset Prices in an Exchange Economy." *Econometrica* 46 (1978): 1429–1445.

[45] Marcet, Albert. "Simulation Analysis of Dynamic Stochastic Models: Applications to Theory and Estimation," mimeo, 1993.

[46] McGrattan, Ellen R. "Solving The Stochastic Growth Model by Linear-Quadratic Approximation." *Journal of Business and Economic Statistics* 8 (No. 1, January 1990): 41–43.

[47] McGrattan, Ellen R. A Note on Computing Competitive Equilibria in Linear Models," *Journal of Economic Dynamics and Control* 18 (1994): 149–160.

[48] J. P. M. Magill, "A Local Analysis of N-Sector Capital Accumulation under Uncertainty," *Journal of Economic Theory*, **15** No. 1, June (1977), 211–219.

[49] Petrushev and Popov. "*Nonlinear Approximation*".

[50] Phillips, Peter C.B. "Best Uniform and Modified Padé Approximants to Probability Densities in Econometrics" in *Advances in Econometrics*, Werner Hildebrand, ed., Cambridge University Press: Cambridge, 1982.

[51] Phillips, Peter C. B. "Marginal Densities of Instrument Variable Estimators in the General Single Equation Case", *Advances in Econometrics* **2**: 1-24.

[52] E. C. Prescott, and R. Mehra, Recursive competitive equilibrium: the case of homogeneous agents, *Econometrica* **48** (1980), 1365–1379.

[53] Rivlin, T.J. *An Introduction to the Approximation of Functions*, Waltham, Massachusetts: Blaisdell Publishing, 1969.

[54] T. J. Rivlin, "Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory," Wiley-Interscience, New York, 1990.

[55] J. Rust, Maximum likelihood estimation of discrete control processes, *SIAM Journal on Control and Optimization* **26** (1988), 1006-1023.

[56] Santos, Manual S., and Jean-Luc Vila, "Smoothness of the Policy Function in Continuous Time Economic Models: The One-Dimensional Case," mimeo, 1988.

[57] C. A. Sims, Solving the stochastic growth model by backsolving with a particular nonlinear form for the decision rule, *Journal of Business and Economic Statistics* **8** (1990), 45–47.

[58] Schumaker, Larry L. "On Shape-Preserving Quadratic Spline Interpolation," *SIAM Journal of Numerical Analysis* 20 (No. 4, August 1983): 854–864.

[59] Stokey, Nancy L., and Robert E. Lucas, Jr. *Recursive Methods in Economic Dynamics*, Harvard University Press: Cambridge. 1989.

[60] G. Tauchen, Solving the stochastic growth model by using quadrature methods and value-function iterations, *Journal of Business and Economic Statistics* **8** (1990), 49–51.

[61] J. B. Taylor, and H. Uhlig, Solving nonlinear stochastic growth models: a comparison of alternative solution methods, *Journal of Business and Economic Statistics*, **8** (1990), 1–18.

[62] White, Halbert. *Artificial Neural Networks,* Blackwell Publishers: Cambridge. 1992.

[63] J. Williams, and B. Wright, "Storage and Commodity Markets," Cambridge University Press, January, 1991.

[64] Wong, R. *Asymptotic Approximation of Integrals*, Academic Press: San Diego. 1989.

[65] Srikant, R., and T. Basar. "Iterative Computatoin of Noncooperative Equilibria in Nonzero-Sum Differential Games with Weakly Coupled Players," *Journal of Optimization Theory and Applications*, 71(No. 1, October 1991): 137–168.

[66] E. Zeidler, "Nonlinear Functional Analysis and Its Applications: Volume I," Springer-Verlag, New York, 1986.

[67] E. Zeidler, "Nonlinear Functional Analysis: Volume II," Springer-Verlag, New York, 1989.