Universitetet i Oslo / Økonomisk institutt / NCF; this version October 28, 2011

A brief note on contraction mappings

The purpose of this topic is to give some principles for computer implementation. We are only covering basic approaches, and no code nor pseudocode will be given.

Update October 28, 2011: More precise language, and some typos fixed. As this note was used in ECON5300: students at that course need not take notice of any changes.

Learning outcomes: You should be able to understand the fixed point property, including *functions as fixed points of operators*, and describe how to iterate a procedure to find an unknown function with a fixed point property, assuming convergence. You should know only basic issues concerning convergence (i.e. the role of discounting in Blackwell's condition, and some conditions which ensure that the convergence is monotoneous); nothing concerning the efficiency in terms of convergence speed is covered in this note.

The idea of fixed point iteration – the one dimensional case

First, a definition: a *fixed point* $\tilde{x}$ for a function f is one for which $f(\tilde{x}) = \tilde{x}$.
Fixed point iteration is connected to difference equations in that the basic idea is to start at some (suitable) point x_0, and then define

$$x_{n+1} = f(x_n).$$

If this iterative procedure converges to some limit L – i.e. $\lim_n x_n = L$ – then also of course $\lim_n x_{n+1} = L$. Therefore $L = f(L)$, so the limit is a fixed point. More:[1] `http://en.wikipedia.org/w/index.php?title=Fixed_point_iteration&oldid=340133049`. So if we wish to compute a certain $\tilde{x}$ within a certain error margin, then we could try the following approach:

- Find a function f which has $\tilde{x}$ as a fixed point
- Iterate, i.e. apply f in succession

Of course, this «if» catch has to be sorted out. That could be done ad hoc – for example, run the algorithm until either the accuracy is as desired, or until a certain test indicates it is time to give up; or there are sometimes theorems granting that it will indeed converge (and to the right value). See below.

A basic example: Newton's method for finding zeroes

See `http://en.wikipedia.org/w/index.php?title=Newton's_method&oldid=344625996`. that the function we use to iterate, is not f, but the function g given in the above link, the «Applications» section.

[1] you all know that Wikipedia articles are published and edited by anyone, without any scientific quality assurance – but I am happy to use articles where I have checked the accuracy of the content. Note the perma-link, ensuring that the article is precisely the one I have checked.)

Convergence, and to the right value?

It suffices either that $|f'(\tilde{x})| < 1$ and that we start sufficiently close – or, assuming a fixed point exists, that $|f'| < 1$ everywhere. See FMEA (the Maths 3 textbook), last subsection of chapter 11, and more practical conditions are given in the next section. For an illustration see `http://en.wikipedia.org/w/index.php?title=Logistic_map&oldid=342666210`, the animated cobweb diagram; observe how the fixed point loses stability when r exceeds 3, where the derivative becomes < -1.

Beyond one dimension

If e.g. x is an n-dimensional vector, then f must also take n-dimensional values; f could for example be multiplication by some $n \times n$ matrix. The conditions on $|f'|$ to ensure convergence could be amended accordingly, see the contraction mappings subsection below.

However, if we are to solve e.g. the Bellman equation, we are not looking for some n-dimensional $\tilde{x}$ – we are looking for a function v, preferably the value function. Then the function to iterate must be some $f(v)$, but (usually) not a pointwisely defined $f(v(x))$. We want to iterate an *operator*, which we shall denote F to distinguish from a function f. An operator[2] is a «function of functions», a black box which takes a function as an input and returns out something, in this case another function. An example is the differentiation operator $F = \frac{\mathrm{d}}{\mathrm{d}x}$; it takes a function as input and returns the derivative (which is also a function; furthermore, the derivative at a point x_0 depends on more than the value at x_0).

Iteration can be done in the same way: We start with an initial function v_0 and define recursively

$$v_{n+1} = F(v_n).$$

A fixed point (yes, one still uses the term «point» although somewhat confusing) is similarly defined as a function L for which $L = F(L)$. For example, ce^x is a fixed point for the differentiation operator $\mathrm{d}/\mathrm{d}x$, for any constant c.
So now we are looking for some F to use for iteration. The Bellman equation is of the form

$$v = F(v)$$

which gives a natural candidate.

Convergence and contraction mappings

As hinted on, we need an infinite-dimensional analogue to the $|f'| < 1$ type criterion. The following is a slightly stronger one, analogous to $|f'| < \beta < 1$ (that is: «uniformly less than 1»). The essential part is not really the derivative, but the property that $|f(x) - f(y)| < \beta|x - y|$. Here, the absolute value is the natural «distance measure» on the real line.

[2] the previous version was not very clear about the distinction between «operator» and «functional». An operator returns an element of a vector space – in this case a function – while a functional returns a number. For example, the «black box» which takes a function and returns its derivative, is an operator, while the black box which returns its slope at a given number, is a functional. The below procedure returns a given number for each x, and running x through the grid, we get a (discretely defined) function.

The more general «distance measure» concept is the one of a *metric*. A metric is a function $d(x, y)$ with the properties that $d > 0$ except $d(x, x) = 0$, that $d(x, y) = d(y, x)$, and finally the triangle inequality: $d(x, y) + d(y, z) \geq d(x, z)$ (i.e., going by way of a third point is never a short cut). There is some inconsistency in the literature whether on whether a metric can attain the value $+\infty$; for our purposes, we shall assume it is finite.

Now consider a more general space $\mathcal{S}$, which in our case will be a set of functions. An $F : \mathcal{S} \mapsto \mathcal{S}$ then takes as input a member of $\mathcal{S}$ and returns a member of $\mathcal{S}$. If S has a metric d, it may be applied both to two members w, v in $\mathcal{S}$, and to $F(w),\ F(v)$. A *contraction mapping*[3] is one F for which there is a $\beta < 1$ such that the inequality

$$d(F(v), F(w)) < \beta d(v, w)$$

holds for all w, v in $\mathcal{S}$. That is, F contracts the distance down by a factor of at most $\beta < 1$. Now assume we start at v_0 and define $v_{n+1} = F(v_n)$. We claim that v_n converges to a fixed point v for F provided that (I) F is a contraction mapping, and (II) the space $\mathcal{S}$ is «without holes» – i.e. like the reals, in contrast to the rationals, where it is easy do «diverge by tending to a hole in the space»[4]:

- Assume $h > 0$ and take n as given. Then successive applications of the triangle inequality yields

  $$d(v_{n+h}, v_n) \leq d(v_{n+1}, v_n) + d(v_{n+2}, v_{n+1}) + \cdots + d(v_{n+h}, v_{n+h-1}) + d(v_{n+h-1}, v_{n+h-2})$$

  ($h - 1$ terms). Now each pair (v_{m+1}, v_m) equals $(F(v_m), F(v_{m-1}))$. Successive replacements all the way down to v_n yields

  $$\begin{aligned} &= d(v_{n+1}, v_n) + d(F(v_{n+1}), (v_n)) + \cdots \\ &\quad + d(F^{h-2}(v_{n+1}), F^{h-2}(v_n)) + d(F^{h-1}(v_{n+1}), F^{h-1}(v_n)) \end{aligned}$$

  where F^m just denotes F applied m times. Using the contraction property m times, we then get

  $$\begin{aligned} &\leq d(v_{n+1}, v_n) + \beta d(v_{n+1}, v_n) + \cdots + d(v_{n+1}, v_n) + \beta^{h-1} d(v_{n+1}, v_n) \\ &= d(v_{n+1}, v_n) \cdot (1 + \beta + \beta^2 + \cdots + \beta^{h-1}) \\ &\leq d(v_{n+1}, v_n) \cdot \frac{1}{1 - \beta}. \end{aligned}$$

  Apply now the contraction property n times, to get

  $$\leq d(v_1, v_0) \cdot \frac{\beta^n}{1 - \beta}.$$

[3] a more precise name would be «β-contraction» to indicate that it does not only contract, but with a factor uniformly less than 1. But then mathematicians usually use the term «k-contraction» (k being in some sense a reserved word here), and I understand that your curriculum uses that letter for capital stock.

[4] More precisely, if you are curious:
`http://en.wikipedia.org/w/index.php?title=Complete_metric_space&oldid=349745167`

- The $d(v_1, v_0)$ is just a number, call it Q. So once at n, the entire rest of the sequence is within distance $\beta^n Q/(1-\beta)$ (and in particular, once we have made one step, we know that the sequence is bounded). So by moving one step forward, we are bounding the sequence by intersections of successively shrinking balls («ball» = a set of points within given distance), shrinking to radius 0. There will be one and only one limit point, namely the intersection of all these balls.

Notice that the contraction property need not hold for any metric – it suffices to find *one* metric such that the property holds. A particularly interesting case of metrics are those defined in terms of a vector norm $||\cdot||$, so that $d(w,v) = ||w-v||$. Vectors may here be e.g. functions[5]; examples of possible norms on functions are $\max_x |v(x)|$ (or, better, «sup»[6]), or $\int_{-\infty}^{\infty} |v(x)| \mathrm{d}x$ if v takes one-dimensional values. Indeed, existence and uniqueness results for differential equations are proved this way, by finding a suitable norm. One is in some sense *constructing* a solution.

Now, to the Bellman equation. This is a lucky case:

- *If payoff is bounded and there is discounting, then this procedure works, with β being the discount factor, when starting with a bounded function (e.g., $v_0 = 0$)* (The «Blackwell condition».)
- Should boundedness fail, then practitioneers need not worry too much: *If boundedness fails, then there is a wide range of problems for which the procedure converges monotonically to the value function if the value is finite, and diverges monotonically to infinity when the value is infinite.*

To sketch a proof of the contraction property, consider the iteration

$$v_{n+1} = \sup_{k\in K} \mathsf{E}[u(k,x) + \beta v_n(g(x,k))]$$

where x is the state, k is the control taking values in K, and $g(x,k)$ is the (stochastic!) state next period.

- The right-hand side is an F applied to v_n, returning another function.

- Suppose that there exists an optimal control k for the RHS maximization. This will depend on both v_n and state x. Call it $\kappa_{v_n}(x)$ where the subscript denotes the dependence of the function.

- Now to establish a contraction mapping property, consider two bounded functions w and v. Each of these will be used as input on the right-hand side, and the corresponding control will be κ_w resp. κ_v. Consider the difference

$$F(w) - F(v) = \mathsf{E}[u(\kappa_w, x) + \beta w(g(x,\kappa_w))] - \mathsf{E}[u(\kappa_v, x) + \beta v(g(x,\kappa_v))].$$

[5] Functions are in some sense no more than infinite-dimensional vectors; an n-vector $(a_1,\dots,a_n)$ returns a value for each $i = 1,\dots,n$, and a function $v(x)$ returns a value for each x in its domain. If the domain is $\{1,\dots,n\}$ they are precisely the same, except the notation with parentheses instead of subscripts.

[6] sup – short for «supremum» – means «the least upper-bound of». Example: the Gaussian cdf Φ ranges $(0,1)$ but does not attain 1. It therefore does not have a max. But the sup is 1. Similarly, the min does not exist, but the corresponding greatest upper-bound inf – «infimum» – is 0.

Since κ_v maximizes the last expectation, it becomes smaller if replaced by κ_w. With the minus sign in front, we obtain

$$\begin{aligned} F(w) - F(v) &\leq \mathsf{E}[u(\kappa_w, x) - u(\kappa_w, x) + \beta w(g(x, \kappa_w) - \beta v(g(x, \kappa_w))] \\ &= \beta \mathsf{E}[w(g(x, \kappa_w) - v(g(x, \kappa_w))]. \end{aligned}$$

Notice that in the last expression, the input argument of w and v are the same. Therefore, the expectation is $\leq \beta \sup_x |w - v|$ (the sup norm!).

- We therefore have $F(w) - F(v) \leq \beta \sup_x |w - v|$, all x. But if we just reverse rôles of w and v, we obtain as well that $F(v) - F(w) \leq \beta \sup_x |w - v|$, all x. We therefore have

$$\begin{aligned} |F(w) - F(v)| &\leq \beta \sup_x |w - v|, \quad \text{all } x \\ \text{so that} \quad \sup_x |F(w) - F(v)| &\leq \beta \sup_x |w - v| \\ \text{i.e.} \quad ||F(w) - F(v)|| &\leq \beta ||w - v|| \end{aligned}$$

  which is precisely the contraction property.

- This works if the iterates are all bounded functions (otherwise, $|v_{n+1} - v_n|$ might be unbounded, the sup norm will yield infinity and all that we know is that we are «within $+\infty \cdot \beta^n/(1 - \beta)$ of the limit»). Now suppose you start with a bounded function. If payoff u is bounded, then the next iterate will also be bounded, and the procedure works.

It should be noted that there are very simple problems where the above fails to apply. Consider for example the problem of saving vs. consumption, where the utility u from consumption is not bounded from above. Start with a constant v_0 (say 0); then in the first iteration, the control will be to consume everything. The next iterate $v_1(x)$ will therefore equal $u(x)$, and if this is unbounded, we are already outside the result. That is not to say that the iteration will not produce the solution, just that the proof does not work, but there are certainly cases where the procedure cannot converge to a finite value function, because the value function need not be finite (consider the case where the return on savings exceeds the discount rate). But for a wide range of problems encountered in practice, the procedure will diverge to infinity precisely when it should.