Lecture 2: Non-Linear Equations

Yuya Takahashi

April 4, 2019

1 Introduction

- Although I showed several economic models that lead to a linear equation system, most of models generate a system of non-linear equations.

- Today we introduce several basic methods that are used in application.

- Basic idea: Generate a sequence of guesses that (hopefully) converges to the solution.

- Different methods construct a sequence in different ways, using different information.

2 Univariate Nonlinear Equation

- A nonlinear equation is of the form

$$f(x) = 0$$

- Sometimes in the form of fixed point:

$$g(x) = x$$

- These are equivalent as we can just redefine the function:

$$f(x) = g(x) - x$$

2.1 Bisection

- Suppose $f(x)$ is continuous and there are a and b such that

$$f(a) < 0 < f(b)$$

  for $a < b$. The Intermediate Value Theorem says there is at least one root for the problem $f(x) = 0$.

- Take $c = \frac{1}{2}(a+b)$. If $f(c) = 0$, you are done.

- If $f(c) < 0$, then we know there is a root in (c, b). Then, take $d = \frac{1}{2}(c+b)$.

- If $f(c) > 0$, then we take $d = \frac{1}{2}(a+c)$.

- Repeat the same "test" by calculating the sign of $f(d)$.

- Algorithm

  1. Bracket a zero: find x_1 and x_2 such that $f(x_1)f(x_2) < 0$. Choose some stopping criteria $\varepsilon, \delta > 0$.

  2. Compute the midpoint $x_m = \frac{1}{2}(x_1 + x_2)$.

  3. Shrink the bounds: If $f(x_1)f(x_m) < 0$, then set $x_2 = x_m$. Otherwise $x_1 = x_m$.

  4. If $x_2 - x_1 \leq \varepsilon(1 + |x_1| + |x_2|)$ or $|f(x_m)| \leq \delta$ stop. Otherwise, go back to step 1.

- The bisection will find a root (if it exists) for sure.

- But its convergence is slow because it does not use almost any information of f (such as slope).

- Linearly convergent iteration.

2.2 Newton's Method

- Suppose $f(x)$ is differentiable.

- Construct a linear approximation to $f(x)$ around a initial guess of x^k and call it $g(x)$

$$g(x) \equiv f'(x_k)(x - x_k) + f(x_k).$$

- Our new guess x_{k+1} is chosen such that $g(x_{k+1}) = 0$: implying

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

- Algorithm

  1. Choose some stopping criteria $\varepsilon, \delta > 0$ and a starting point x_0.

  2. Compute

  $$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

  3. If $|x_k - x_{k+1}| \le \varepsilon (1 + |x_{k+1}|)$ go to the next stop. Otherwise, go back to step 2.

  4. If $|f(x_{k+1})| \le \delta$, stop successfully. Otherwise, stop and report "failure".

- It will converge if the function is well-behaved and the initial guess is not "too bad".

- Newton's method may fail (diverging sequence, cycling, etc).

- Quadratically convergent sequence

2.3 Quasi-Newton Method

- To use the Newton's method, we need to calculate $f'(x)$, which can be very costly.

- Instead of using the exact value of $f'(x)$, we can approximate it.

- **Secant** method approximates it using the slope of the secant of f between x_k and x_{k-1}:

$$x_{k+1} = x_k - f(x_k)\left(\frac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})}\right).$$

- Unlike the Newton method, the secant method requires two starting values.

- Convergence is slower than the Newton's method.

3 Multivariate Nonlinear Equations

- In this case, $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ and we want to solve $f(x) = 0$. (Note that both sides are of n dimensions)

- A system of n equations with n unknowns:

$$\begin{aligned} f^1(x_1, x_2, ..., x_n) &= 0 \\ f^2(x_1, x_2, ..., x_n) &= 0 \\ &\vdots \\ f^n(x_1, x_2, ..., x_n) &= 0 \end{aligned}$$

- From here on, the subscript for each variable denotes the location of the variable in a vector. Superscripts for vectors denote the number of iterations.

3.1 Gauss-Jacobi

- Given the k-th iteration x^k, we update x^{k+1} element-by-element using

$$
\begin{aligned}
f^1\left(x_1^{k+1}, x_2^k, ..., x_n^k\right) &= 0 \\
f^2\left(x_1^k, x_2^{k+1}, ..., x_n^k\right) &= 0 \\
&\vdots \\
f^n\left(x_1^k, x_2^k, ..., x_n^{k+1}\right) &= 0.
\end{aligned}
$$

- We have decomposed the entire system into a series of n single nonlinear equations. For each equation, we can apply methods discussed above.

- Since we are going to solve each equation many times, there is no point in solving them very accurately.

- We could approximately solve each equation: **linear Gauss-Jacobi** takes one single Newton step:

$$x_i^{k+1} = x_i^k - \frac{f^i\left(x^k\right)}{f^i_{x_i}\left(x^k\right)}, \qquad i = 1, ..., n$$

where $f^i_{x_i}$ is the partial derivative of f^i with respect to the i-th argument.

3.2 Gauss-Seidel

- Same as before (case of linear equations), use the new guess for x_i as soon as it becomes available.

- For a given k-th iteration x^k, update x^{k+1} sequentially using

$$\begin{aligned} f^1\left(x_1^{k+1}, x_2^k, x_3^k ..., x_n^k\right) &= 0 \\ f^2\left(x_1^{k+1}, x_2^{k+1}, x_3^k ..., x_n^k\right) &= 0 \\ f^3\left(x_1^{k+1}, x_2^{k+1}, x_3^{k+1} ..., x_n^k\right) &= 0 \\ &\vdots \\ f^n\left(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, ..., x_n^{k+1}\right) &= 0. \end{aligned}$$

- Notice that we solve $f^1, f^2, ..., f^n$ in sequence, but we use each component (solution) in the subsequent steps.

- The same logic of solving only approximately to construct a **linear Gauss-Seidel** applies:

$$x_i^{k+1} = x_i^k - \left(\frac{f^i}{f^i_{x_i}}\right)\left(x_1^{k+1}, ..., x_{i-1}^{k+1}, x_i^k, ..., x_n^k\right), \quad i = 1, ..., n.$$

3.3 Fixed-Point Iteration

- Simplest procedure to solve $x = f(x)$ is

$$x^{k+1} = f(x^k).$$

- When will it converge?

Def: A differentiable contraction mapping on D is any C^1 $f : D \longrightarrow \mathbb{R}^n$ defined on a closed, bounded, convex set $D \subset \mathbb{R}^n$ such that

1. $f(D) \subset D$, and
2. $\max_{x \in D} ||J(x)||_\infty < 1$ where $J(x)$ is the Jacobian of f.

- We can apply the contraction mapping theorem: If f is a differentiable contraction map on D, then

1. The fixed point problem $x = f(x)$ has a unique solution $x^* \in D$

2. The sequence defined by $x^{k+1} = f(x^k)$ converges to x^*

3. There exists a sequence $\varepsilon_k \rightarrow 0$ such that

$$||x^* - x^{k+1}||_\infty \leq (||J\left(x^*\right)||_\infty + \varepsilon_k)\,||x^* - x^k||_\infty$$

3.4 Newton's Method

- Apply Taylor's theorem to the first order around some initial guess x^k and call the series $g(x)$.

$$g(x) \equiv f\left(x^k\right) + J\left(x^k\right)\left(x - x^k\right),$$

where J is a Jacobian.

- As before, our new guess x^{k+1} is chosen such that $g(x^{k+1}) = 0$: implying

$$x^{k+1} = x^k - J\left(x^k\right)^{-1} f\left(x^k\right).$$

- Algorithm

1. Choose some stopping criteria $\varepsilon, \delta > 0$ and a starting point x^0.
   (a) Compute Jacobian $A_k = J(x^k)$.
   (b) Solve the matrix equation $A_k s^k = -f(x^k)$ for s^k.
   (c) Set $x^{k+1} = x^k + s^k$.
2. If $||x^k - x^{k+1}|| \leq \varepsilon\left(1 + ||x^{k+1}||\right)$, go to the next step. Otherwise, go back to step 2.
3. If $||f\left(x^{k+1}\right)|| \leq \delta$, stop and declare "success". Otherwise, stop and report "failure".

3.5 Quasi Newton (Broyden)

- Again, instead of calculating the Jacobian, we approximate it.

- This is the n-dimensional analogue of the secant method.

- Idea: let's solve the secant equation iteratively. For some guess of the Jacobian A_k, solve for the Newton step s^k by $A_k s^k = -f(x^k)$ and set $x^{k+1} = x^k + s^k$.

- Then, instead of directly calculating A_{k+1} we approximate it. How do we do that?

- In the one dimension case we approximate the derivative near two points y, z by the solution m to the secant equation

$$f(y) - f(z) = m \times (y - z).$$

- A naive analogue is to find a Jacobian J that satisfies

$$f(y) - f(z) = J(y - z). \tag{1}$$

- But this imposes n restrictions, while the Jacobian has n^2 elements.

- Broyden's method: the "best possible" choice for A_{k+1} is a minimal modification of A_k, which is given by $A_{k+1}q = A_k q$ whenever $< s^k, q >= 0$. This way, A_{k+1} is "closest" to A_k in a certain sense.

- This uniquely pins down A_{k+1} :

$$A_{k+1} = A_k + \frac{\left(y_k - A_k s^k\right)\left(s^k\right)^\top}{\left(s^k\right)^\top s^k}$$

where $y_k \equiv f(x^{k+1}) - f(x^k)$.

- "Intuition"

  - As A_k converges, (1) is satisfied
  - Choose q such that $< s^k, q >= 0$. If you post-multiply such q on both sides, the second term of the RHS vanishes, and the "minimal requirement" is satisfied.
  - $\left(y_k - A_k s^k\right)\left(s^k\right)^\top$ is of rank one, the change is small

3.6 Global Convergence

- Any nonlinear equation problem can be converted to an optimization problem.

- Any solution to the system $f(x) = 0$ is also a global solution to

$$\min_x SSR(x)$$

where

$$SSR(x) = \sum_{i=1}^{n} f^i(x)^2$$

and any global minimum of $\sum_{i=1}^{n} f^i(x)^2$ is a solution to $f(x) = 0$.

- This method will converge to something. But the process may be very slow, and converge to a point other than a solution to the nonlinear system.

- Powell's hybrid method: Accept the Newton step only if it reduces SSR; otherwise choose a direction equal to a combination of the Newton step and the gradient of $-SSR$.

- This will converge to a solution of $f(x) = 0$ or stop at a local minimum of SSR.

3.7 Preliminary Transformation

- Transformations are useful. The idea is to transform the problem to a simpler one that has the same solution as the one we are trying to solve

- For example, in Newton's method, the key assumption is that the system can be well-approximated by a linear function.

- Simple dumb but powerful example. Solve the highly non-linear equation

$$x^{13} - 1 = 0.$$

- If we use Newton's method, how many iterations do we need to get convergence?

- Simple transformation: solve

$$x - 1 = 0.$$

- Another example:

$$x^{0.2} + y^{0.2} - 2 = 0$$
$$x^{0.1} + y^{0.4} - 2 = 0$$

with solution $x = 1, y = 1$. If we start at $x = 2, y = 2$, we converge. If we start at $x = 3, y = 3$, we do not.

- What if we try to make it look "more linear"

$$\left(x^{0.2} + y^{0.2}\right)^5 - 32 = 0$$
$$\left(x^{0.1} + y^{0.4}\right)^4 - 16 = 0$$

4 Games of Incomplete Information

- Consider a simple static entry model where 2 players (i and j) simultaneously choose between entering or not.

- Entry of firm j affects (arguably reduces) firm i's profit.

- Without loss of generality, we normalize the profit of not entering to zero for both firms.

- Profits are given by

$$\Pi_i = \begin{cases} h_i(X_i) + \alpha_i y_j - \varepsilon_i & \text{if } y_i = 1 \\ 0 & \text{if } y_i = 0 \end{cases},$$

where $y_j = 1$ if firm j enters the market and $y_j = 0$ otherwise.

- Let Ω_i denote firm i's information set (i.e. state variables when making a decision) and let $\pi_j \equiv E\left(y_j = 1|\Omega_i\right)$

- The optimal choices are given by

$$y_i = \mathbf{1}\left\{h_i\left(X_i\right) + \alpha_i \pi_j - \varepsilon_i \geq 0\right\},$$

where $\mathbf{1}\left\{a\right\}$ is an indicator function

4.1 Independent Private Shocks (IPS)

- Assume $\Omega_i = \left(X_i, X_j, \varepsilon_i\right)$.

- A Bayesian-Nash equilibrium of this game is given by

$$\begin{aligned} y_1 &= \mathbf{1}\left\{h_1\left(X_1\right) + \alpha_1 \pi_2^* - \varepsilon_1 \geq 0\right\} \\ y_2 &= \mathbf{1}\left\{h_2\left(X_2\right) + \alpha_2 \pi_1^* - \varepsilon_2 \geq 0\right\}, \end{aligned}$$

where (π_1^*, π_2^*) is a fixed point of $\varphi = (\varphi_1, \varphi_2) = \mathbf{0}$ with

$$\begin{aligned}\varphi_1(\pi_1, \pi_2) &= \pi_1 - G_{\varepsilon_1}(h_1(X_1) + \alpha_1 \pi_2) \\ \varphi_2(\pi_1, \pi_2) &= \pi_2 - G_{\varepsilon_2}(h_2(X_2) + \alpha_2 \pi_1).\end{aligned}$$

- This implies that both π_1^* and π_2^* are functions of only $\mathbf{X} = (X_1, X_2)$.

- Work with an example

- $\varepsilon_1, \varepsilon_2$ follow iid standard normal

- Use

$$y_1 = \mathbf{1}\{1.2 - 0.5\pi_2^* - \varepsilon_1 \geq 0\} \tag{2}$$
$$y_2 = \mathbf{1}\{1.2 - 0.5\pi_1^* - \varepsilon_2 \geq 0\}, \tag{3}$$

where π_1^* and π_2^* are the fixed point of

$$\pi_1 - \Phi\left(1.2 - 0.5\pi_2\right) = 0 \tag{4}$$
$$\pi_2 - \Phi\left(1.2 - 0.5\pi_1\right) = 0. \tag{5}$$

- How to compute an equilibrium?

1. Draw ε_1 and ε_2.
2. Find the equilibrium probabilities by finding the fixed point to (4) and (5). Use a variant of fixed-point algorithms (Gauss-Jacobi?).
   (a) Start the fixed point search at $\pi_2 = 1$. Let π_1^1 be the solution to (4). Using π_1^1, let π_2^1 be the solution to (5).
   (b) Iterate until we get $|\pi_1^k - \pi_1^{k+1}| < \epsilon$ and $|\pi_2^k - \pi_2^{k+1}| < \epsilon$ for sufficiently small ϵ. Call the fixed point π_1^* and π_2^*.
3. Using these values, determine (y_1, y_2) from the threshold crossing model given by (2) and (3).

- What if we have multiple roots? (for example, instead use $\alpha_1 = \alpha_2 = -3$)

- If we focus on a symmetric equilibrium, let $\pi_1 = \pi_2$. Then, the fixed-point problem is reduced to a univariate case:

$$\pi = \Phi\left(1.2 - 0.5\pi\right).$$

- Since Φ is strictly increasing, the RHS is strictly decreasing in π. Unique solution.

- To use bisection, how to bracket zero?

4.2 Correlated Private Shocks (CPS)

- Let $G_{\varepsilon_1,\varepsilon_2}(\cdot,\cdot)$ be the joint distribution of $(\varepsilon_1,\varepsilon_2)$ and let $g_{\varepsilon_1|\varepsilon_2}(\varepsilon_1|\varepsilon_2)$ denote the density of ε_1 conditional on ε_2.

- Note that $\Omega_i = \left(X_i, X_j, \varepsilon_i\right)$.

- Since now the realization of the privately observed shock ε_1 contains information about the realized ε_2, the equilibrium beliefs will be functions of shock realizations.

- A Bayesian-Nash equilibrium of this game is given by

$$y_1 = \mathbf{1}\left\{h_1(X_1) + \alpha_1\pi_2^* - \varepsilon_1 \geq 0\right\} \quad (6)$$

$$y_2 = \mathbf{1}\left\{h_2(X_2) + \alpha_2\pi_1^* - \varepsilon_2 \geq 0\right\}, \quad (7)$$

where $\left(\pi_1^*, \pi_2^*\right)$ is a solution to the following system of functional equations:

$$\pi_1^*\left(\mathbf{X}, \varepsilon_2\right) = \int \mathbf{1}\left\{h_1\left(X_1\right) + \alpha_1 \pi_2^*\left(\mathbf{X}, \varepsilon_1\right) - \varepsilon_1 \geq 0\right\} g_{\varepsilon_1|\varepsilon_2}\left(\varepsilon_1|\varepsilon_2\right) d\varepsilon_1 \tag{8}$$

$$\pi_2^*\left(\mathbf{X}, \varepsilon_1\right) = \int \mathbf{1}\left\{h_2\left(X_2\right) + \alpha_2 \pi_1^*\left(\mathbf{X}, \varepsilon_2\right) - \varepsilon_2 \geq 0\right\} g_{\varepsilon_2|\varepsilon_1}\left(\varepsilon_2|\varepsilon_1\right) d\varepsilon_2. \tag{9}$$

Work with an example: $h_1\left(X_1\right) = h_2\left(X_2\right) = 1.2$ and $\alpha_1 = \alpha_2 = -0.5$.

- Assume

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ & 1 \end{pmatrix}\right).$$

- Calculating the fixed point for (8) and (9) is computationally demanding.

- Remember that there may be multiple equilibria.

- Algorithm:

  1. Approximate (8) and (9) in a certain way (topic of numerical integration). For now, let's just assume ε can take one of N points in $\{z_1, z_2, ..., z_N\}$ with weights $\{w_1, w_2, ..., w_N\}$.

  2. Solve the fixed-point problem. Use a variant of fixed-point algorithm (follow Aradillas-Lopez (2010)). Set $\pi_1^0(\cdot) = 1$ and $\pi_2^0(\cdot) = 0$. For all $\varepsilon_2 \in \{z_1, z_2, ..., z_N\}$, update $\pi_1^1(\varepsilon_2)$ using

$$\pi_1^{k+1}(\varepsilon_2) \approx \sum_{s=1}^{N_s} \mathbf{1}\left\{1.2 - 0.5\pi_2^k(z_s) - z_s \geq 0\right\} \phi\left(z_s; \rho\varepsilon_2, 1-\rho^2\right) w_s,$$

     where $\phi(\cdot; a, b)$ is the PDF of a normal distribution with mean a and variance b. Likewise, for all $\varepsilon_1 \in \{z_1, z_2, ..., z_N\}$, update $\pi_2^1(\varepsilon_1)$ using

$$\pi_2^{k+1}(\varepsilon_1) \approx \sum_{s=1}^{N_s} \mathbf{1}\left\{1.2 - 0.5\pi_1^k(z_s) - z_s \geq 0\right\} \phi\left(z_s, \rho\varepsilon_1, 1-\rho^2\right) w_s.$$

  3. Iterate the procedure until convergence.

4. Determine (y_1, y_2) using (6) and (7).

4.3 Application: Seim (2006)

- Consider several estimation procedures using Seim (2006)'s framework

- In the static game literature, we typically collect outcomes ($\#$ of entrants and other observable variables) from geographically separated markets

- Assume these are outcomes of independent repetitions of the same game

4.3.1 Model

- Consider a large number of potential entrants: N_{max}

- Parameterization:

$$\Pi(N, S) = S \cdot (\alpha - \gamma(N - 1)) - F$$

- (α, γ, F) are parameters and (N, S) are observed in each market

- ϵ_i is private information: Observed by firm i but not other firms or econometrician. Assume iid across players.

- Hence, firm i must form expectation about other firms entering

- Firm i enters if

$$E(\Pi(N,S)|a_i = enter) + \epsilon_i \geq 0$$

or (linearity allows to move through expectations operator)

$$S \cdot (\alpha - \gamma E(N-1|a_i = enter)) - F + \epsilon_i \geq 0$$

- Assume symmetry

- Let

$$\Pr(Entry) = p = G_\epsilon(E[\Pi(N,S)|a_i = enter])$$

- Hence, firm i enters if

$$S \cdot (\alpha - \gamma[p \cdot (N_{max} - 1)]) - F + \epsilon_i \geq 0$$

- A Bayesian-Nash Equilibrium is given by a fixed point:

$$p = G_{\epsilon}(S \cdot (\alpha - \gamma[p \cdot (N_{max} - 1)]) - F)$$

- If ϵ_{it} iid extreme value:

$$p = \frac{\exp(S \cdot (\alpha - \gamma[p \cdot (N_{max} - 1)]) - F)}{1 + \exp(S \cdot (\alpha - \gamma[p \cdot (N_{max} - 1)]) - F)}$$

- Likelihood contribution of market with N active firms and market demand S:

$$\ell(N, S, p) = \frac{N_{max}!}{(N_{max} - N)!N!} p^{N} (1 - p)^{(N_{max} - N)}$$

4.3.2 Estimation

- "Nested Fixed Point Algorithm":

  1. Fix parameter vector $\theta = (\alpha, \gamma, F)$

  2. Find fixed point in

  $$p_m = G_\epsilon(S_m \cdot (\alpha - \gamma[p_m \cdot (N_{max} - 1)]) - F)$$

     for every market $m = 1...., M$, yielding $p_m(\theta)$

  3. Evaluate log-likelihood

  $$L_M(\theta) = \frac{1}{M} \sum_{m=1}^{M} \log(\ell(N_m, S_m, p_m(\theta)))$$

  4. Use numerical algorithm to find θ that maximizes $L_M(\theta)$

- "Nested" because the inner loop solves for equilibrium p_m and the outer loop maximizes likelihood

- Alternatively, we could form an moment estimator:

    1. Estimate probability of entry in each market

    $$\hat{p}_m = \frac{N_m}{N_{\mathsf{max}}}$$

    2. Moment condition:

    $$\frac{1}{M}\sum_{m=1}^{M} Z_m \otimes [\hat{p}_m - G_\epsilon(S_m \cdot (\alpha - \delta[\hat{p}_m \cdot (N_{max} - 1)]) - F)]$$

    where Z_m is a vector of exogenous variables (e.g., S_m). The sample analogues is

    $$E[\hat{p}_m - G_\epsilon\left(S_m \cdot (\alpha - \gamma[\hat{p}_m \cdot (N_{\mathsf{max}} - 1)]) - F\right)|Z_m] = 0$$

- Advantage of two-step estimators:

    - No need to solve for equilibrium for each trial of parameter vector

- Avoid problem of multiple equilibria

  - Even with multiplicity, we can calculate the equilibrium probability in each market when N_{max} is large ("picked by the data") as in step 1 above (see social interaction literature, e.g., Brock and Durlauf, 2001)

  - When N_{max} is small, we need to pool different markets

  - Either focusing on unique outcomes or specifying an equilibrium selection rule is needed

- Two step idea will be what we use for estimating dynamic models without solving for optimal policy/equilibrium