

Numerical Optimization

Zi Wang
HKBU

Spring 2025

Example: Generalized Method of Moments (GMM)

- Sample size n . Moment conditions

$$\hat{m}(\theta) \equiv \frac{1}{n} \sum_{i=1}^n m_i(\theta) = 0 \quad (1)$$

- GMM estimator:

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \hat{m}(\theta)' \hat{W} \hat{m}(\theta) \quad (2)$$

- Example:

- Linear IV: $m_i(\theta) = z_i (y_i - x_i' \theta)$
- Logistic IV: $m_i(\theta) = z_i \left(y_i - \frac{1}{1 + \exp[-x_i' \theta]} \right)$

Example: Generalized Method of Moments (GMM)

- Asymptotic:

$$\sqrt{n} (\hat{\theta} - \theta_0) \rightarrow_d N(0, V) \quad (3)$$

where $V = (G'WG)^{-1} G'WV_mWG (G'WG)^{-1}$, $G = G(\theta_0) = \nabla_{\theta'} m(\theta_0)$ and $V_m = \text{Var}[m(\theta_0)]$

- Estimate V :

- $\hat{G} = \frac{1}{n} \sum_{i=1}^n \nabla_{\theta'} m_i(\hat{\theta})$

- $\hat{V}_m = \frac{1}{n} \sum_{i=1}^n m_i(\hat{\theta}) m_i(\hat{\theta})' - \left[\frac{1}{n} \sum_{i=1}^n m_i(\hat{\theta}) \right] \left[\frac{1}{n} \sum_{i=1}^n m_i(\hat{\theta}) \right]'$

- Optimal W :

- If $W = V_m^{-1}$, then $V = [G'V_m^{-1}G]^{-1}$

Newton-Raphson Method

- One-dimension unconstrained optimization: $\min_x f(x)$
 - The second-order Taylor expansion around initial guess x_0 :
 $f(x_0 + t) = f(x_0) + f'(x_0)t + \frac{1}{2}f''(x_0)t^2$
 - FOC: $f'(x_0) + f''(x_0)t = 0 \Rightarrow t = -\frac{f'(x_0)}{f''(x_0)}$
 - Iteration: $x_k = x_{k-1} - \frac{f'(x_{k-1})}{f''(x_{k-1})}$ for $k = 1, 2, \dots$
- Higher dimensions:

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} - \underbrace{\left[\nabla_{\mathbf{x}^2}^2 f(\mathbf{x}^{(k-1)}) \right]^{-1}}_{\text{Hessian}} \underbrace{\nabla_{\mathbf{x}} f(\mathbf{x}^{(k-1)})}_{\text{Jacobian}} \quad (4)$$

Broyden-Fletcher-Goldfarb-Shano (BFGS) Method

- Approximate the inverse of the Hessian matrix: initial guess $B_0 = \mathbf{I}$
- Let $\mathbf{h} = \mathbf{x}^{(1)} - \mathbf{x}^{(0)}$ and $\mathbf{y} = \nabla_{\mathbf{x}} f(\mathbf{x}^{(1)}) - \nabla_{\mathbf{x}} f(\mathbf{x}^{(0)})$. By the definition of the Hessian matrix, we want to obtain B_1 such that $B_1 \mathbf{y} = \mathbf{h}$
- Update $B_1 = B_0 - \frac{B_0 \mathbf{y} \mathbf{y}' B_0}{\mathbf{y}' B_0 \mathbf{y}} + \frac{\mathbf{h} \mathbf{h}'}{\mathbf{y}' \mathbf{h}}$. It is straightforward to verify that $B_1 \mathbf{y} = \mathbf{h}$
- Update $\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} - B_{k-1} \nabla_{\mathbf{x}} f(\mathbf{x}^{(k-1)})$ for $k = 1, 2, \dots$

Newton and Quasi-Newton Methods: Example

- Solve $\min_{x_1, x_2} f(x_1, x_2) = -\log(1 - 2x_1 - x_2) - \log x_1 - \log x_2$
- Jacobian matrix:

$$J(x) = \begin{bmatrix} \frac{2}{1-2x_1-x_2} - \frac{1}{x_1} \\ \frac{1}{1-2x_1-x_2} - \frac{1}{x_2} \end{bmatrix} \quad (5)$$

- Hessian matrix:

$$H(x) = \begin{bmatrix} \frac{4}{(1-2x_1-x_2)^2} + \frac{1}{x_1^2} & \frac{2}{(1-2x_1-x_2)^2} \\ \frac{1}{(1-2x_1-x_2)^2} & \frac{1}{(1-2x_1-x_2)^2} + \frac{1}{x_2^2} \end{bmatrix} \quad (6)$$

Example: Optimal Tariffs in the Armington Model

- Mathematical programming with equilibrium constraints (MPEC): for some country n^*

$$\max_{(\widehat{1+t_{in}^*}; \hat{w}_i, \hat{X}_i, \hat{P}_i)_{i=1}^N} \hat{U}_{n^*} \equiv \frac{\hat{X}_{n^*}}{\hat{P}_{n^*}}$$

s.t.

$$\hat{w}_i w_i L_i = \sum_{n=1}^N \frac{1}{1+t'_{in}} \hat{\lambda}_{in} \hat{X}_n \lambda_{in} X_n, \quad \hat{\lambda}_{in} = \frac{(\hat{w}_i \hat{\kappa}_{in})^{1-\sigma}}{\sum_{k=1}^N \lambda_{kn} (\hat{w}_k \hat{\kappa}_{kn})^{1-\sigma}}, \quad \hat{\kappa}_{in} = \widehat{1+t_{in}} \quad (7)$$

$$\hat{X}_n X_n = \hat{w}_n w_n L_n + \sum_{i=1}^N \frac{t'_{in}}{1+t'_{in}} \hat{\lambda}_{in} \hat{X}_n \lambda_{in} X_n$$

$$\hat{P}_i = \left[\sum_{k=1}^N \lambda_{ki} (\hat{w}_k \hat{\kappa}_{ki})^{1-\sigma} \right]^{\frac{1}{1-\sigma}}$$

Constrained Optimization

- General form of MPEC: policies \mathbf{s} ; equilibrium outcomes \mathbf{x}

$$\begin{aligned} \max_{\mathbf{s}, \mathbf{x}} \quad & O(\mathbf{x}, \mathbf{s}) \\ \text{s.t.} \quad & F(\mathbf{x}, \mathbf{s}) = 0 \end{aligned} \tag{8}$$

- General form of constrained optimization:

$$\begin{aligned} \min_{\mathbf{x}} \quad & O(\mathbf{x}) \\ \text{s.t.} \quad & \\ & F_i(\mathbf{x}) = 0, \quad i = 1, \dots, I \\ & G_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, J \end{aligned} \tag{9}$$

Constrained Optimization

- Lagrange function:

$$\mathcal{L}(\mathbf{x}, \mu, \lambda) \equiv O(\mathbf{x}) + \mu' F(\mathbf{x}) + \lambda' G(\mathbf{x}) \quad (10)$$

- Karush–Kuhn–Tucker (KKT) conditions:

- FOC: $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \mu^*, \lambda^*) \equiv \nabla_{\mathbf{x}} O(\mathbf{x}^*) + \sum_{i=1}^I \mu_i^* \nabla_{\mathbf{x}} F_i(\mathbf{x}^*) + \sum_{j=1}^J \lambda_j^* \nabla_{\mathbf{x}} G_j(\mathbf{x}^*) = 0$

- $F_i(\mathbf{x}^*) = 0, \quad i = 1, \dots, I$

- $\lambda_j^* \geq 0$ and $\lambda_j^* G_j(\mathbf{x}^*) = 0$

- In practice: Constrained \Rightarrow Unconstrained

- Interior point (Barrier) methods: Barrier function

- Active set methods: KKT

Constrained Optimization: Example

- Problem:

$$\begin{aligned} & \min_{x_1 \geq 0, x_2 \geq 0} 2x_1^2 - 12x_1 + 3x_2^2 - 18x_2 + 45 \\ & \text{s.t.} \\ & 3x_1 + x_2 \leq 12 \\ & x_1 + x_2 \leq 6 \end{aligned} \tag{11}$$

Mutual Optimization: Nash Equilibria

1. Initial guess $(t'_{in})^{(0)}$ for all i, n
2. For each n , solve for the welfare-maximizing (t'_{in}) , given $(t'_{i,-n})^{(0)}: (t'_{in})^{(1)}$
3. Iterate until $(t'_{in})^{(k)} = (t'_{in})^{(k-1)}$

Example: Nash Tariffs between the U.S. and China

For some country k :

$$\begin{aligned}
 \max_{(t'_{ik})_{i \neq k}; (\hat{w}_i, \hat{X}_i, \hat{P}_i)_{i=1}^N} \hat{U}_k &\equiv \frac{\hat{X}_k}{\hat{P}_k} \\
 \text{s.t.} \\
 \hat{P}_n &= \left[\sum_{i=1}^N \lambda_{in} (\hat{w}_i \hat{\kappa}_{in})^{1-\sigma} \right]^{\frac{1}{1-\sigma}} \\
 \hat{w}_i w_i L_i &= \sum_{n=1}^N \frac{1}{1+t'_{in}} \hat{\lambda}_{in} \hat{X}_n \lambda_{in} X_n, \quad \hat{\lambda}_{in} = \frac{(\hat{w}_i \hat{\kappa}_{in})^{1-\sigma}}{\sum_{i'=1}^N \lambda_{i'n} (\hat{w}_{i'} \hat{\kappa}_{i'n})^{1-\sigma}}, \quad \hat{\kappa}_{in} = \widehat{1+t_{in}} \\
 \hat{X}_n X_n &= \hat{w}_n w_n L_n + \sum_{i=1}^N \frac{t'_{in}}{1+t'_{in}} \hat{\lambda}_{in} \hat{X}_n \lambda_{in} X_n
 \end{aligned} \tag{12}$$

Nested Fixed Point (NFXP)

- MPEC:

$$\begin{aligned} & \max_{\mathbf{s}, \mathbf{x}} O(\mathbf{x}, \mathbf{s}) \\ & \text{s.t. } F(\mathbf{x}, \mathbf{s}) = 0 \end{aligned} \tag{13}$$

- NFXP

- $F(\mathbf{x}, \mathbf{s}) = 0 \Rightarrow \mathbf{x} = H(\mathbf{s})$ (inner loop)

- $\max_{\mathbf{s}} O(H(\mathbf{s}), \mathbf{s})$ (outer loop)

- FOC: for $\mathbf{x}^* = H(\mathbf{s}^*)$

- $\nabla_{\mathbf{x}} O(\mathbf{x}^*, \mathbf{s}^*) \nabla_{\mathbf{s}} H(\mathbf{s}^*) + \nabla_{\mathbf{s}} O(\mathbf{x}^*, \mathbf{s}^*) = 0$

- $\nabla_{\mathbf{s}} H(\mathbf{s}^*)$: Implicit Differentiation

$$\nabla_{\mathbf{s}} H(\mathbf{s}^*) = - [\nabla_{\mathbf{x}} F(\mathbf{x}^*, \mathbf{s}^*)]^{-1} \nabla_{\mathbf{s}} F(\mathbf{x}^*, \mathbf{s}^*) \tag{14}$$

Example: Optimal Spatial Transfers

- N regions. Total labor \bar{L} , freely mobile
- Each region produces a distinctive variety. Representative consumer in region n has a CES utility:

$$U_n = B_n \left[\sum_{i=1}^N c_{in}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \quad \sigma > 1 \quad (15)$$

where amenity $B_n \equiv \bar{B}_n L_n^{-\beta}$ and L_n is the labor in region n

- Each variety is produced using labor under perfect competition. Unit cost:

$$c_i = \frac{w_i}{A_i}, \quad A_i = \bar{A}_i L_i^{\alpha} \quad (16)$$

- Iceberg trade cost $\tau_{in} \geq 1$. Price index in region n :

$$P_n^{1-\sigma} = \sum_{i=1}^N \left(\frac{w_i \tau_{in}}{A_i} \right)^{1-\sigma} \quad (17)$$

Example: Optimal Spatial Transfers

- Wage tax/subsidy+Welfare equalization:

$$U = U_n = B_n \frac{(1 + s_n) w_n}{P_n} \quad (18)$$

- Labor allocation and welfare:

$$\frac{L_n}{\bar{L}} = \frac{\left[\bar{B}_n \frac{(1+s_n)w_n}{P_n} \right]^{\frac{1}{\beta}}}{\sum_{k=1}^N \left[\bar{B}_k \frac{(1+s_k)w_k}{P_k} \right]^{\frac{1}{\beta}}} \quad U = \left[\sum_{k=1}^N \left[\bar{B}_k \frac{(1+s_k)w_k}{P_k} \right]^{\frac{1}{\beta}} \right]^{\beta} \quad (19)$$

- Governmental budget balance:

$$\sum_{n=1}^N s_n w_n L_n = 0 \quad (20)$$

Example: Optimal Spatial Transfers

- Optimal spatial transfers: (s_n^*)

$$\begin{aligned}
 \max_{(s_n; w_n, P_n, L_n)_{n=1}^N} U &= \left[\sum_{k=1}^N \left[\bar{B}_k \frac{(1+s_k) w_k}{P_k} \right]^{\frac{1}{\beta}} \right]^{\beta} \\
 \text{s.t.} \quad & \\
 w_i L_i &= \sum_{n=1}^N \lambda_{in} (1+s_n) w_n L_n, \quad \lambda_{in} \equiv \frac{X_{in}}{X_n} = \frac{\left(\frac{w_i \tau_{in}}{\bar{A}_i L_i^{\alpha}} \right)^{1-\sigma}}{\sum_{k=1}^N \left(\frac{w_k \tau_{kn}}{\bar{A}_k L_k^{\alpha}} \right)^{1-\sigma}} \\
 \frac{L_n}{\bar{L}} &= \frac{\left[\bar{B}_n \frac{(1+s_n) w_n}{P_n} \right]^{\frac{1}{\beta}}}{\sum_{k=1}^N \left[\bar{B}_k \frac{(1+s_k) w_k}{P_k} \right]^{\frac{1}{\beta}}}, \quad P_n^{1-\sigma} = \sum_{i=1}^N \left(\frac{w_i \tau_{in}}{\bar{A}_i L_i^{\alpha}} \right)^{1-\sigma} \\
 \sum_{n=1}^N s_n w_n L_n &= 0
 \end{aligned} \tag{21}$$

Multiple Equilibria and Global Optimization

- When $\alpha > \beta$, there could be multiple equilibria
- Local optima under multiple equilibria: how to find the globally optimal spatial transfers?
 - In general, a very difficult problem
 - Simulated annealing: a probabilistic technique for approximating the global optimum of a given function

Simulated Annealing

- Annealing:
 - Heating a metal to a high temperature, allowing its atoms to move freely within the structure
 - As the metal is slowly cooled, the atoms gradually settle into a low-energy crystalline configuration
- Optimization: $\min_{\mathbf{x}} f(\mathbf{x})$
 1. Initial guess \mathbf{x}_0
 2. Given a constant $c_0 > 0$
 - 2.1 Generate \mathbf{x}_1 from the neighborhood of \mathbf{x}_0 (perturbation techniques)
 - 2.2 If $f(\mathbf{x}_1) < f(\mathbf{x}_0)$, \mathbf{x}_1 becomes the current solution
 - 2.3 If $f(\mathbf{x}_1) \geq f(\mathbf{x}_0)$, \mathbf{x}_1 becomes the current solution with probability $\exp\left[\frac{f(\mathbf{x}_0) - f(\mathbf{x}_1)}{c_0}\right]$
 3. Decrease c_0 based on a pre-specified rule: e.g. $c_1 = 0.85 * c_0$
 4. Iterate until $c_k \simeq 0$

Simulated Annealing

- Example: $f(x) = \sin(x) + \sin\left(\frac{10}{3}x\right)$ for $x \in [-\pi, \pi]$
 - The global minima is $x \simeq -2.30$ with $f(x) \simeq -1.73$
 - Starting from $x = 0$, search the global minima using simulated annealing

Summary

- Unconstrained optimization
- Constrained optimization: MPEC
- Globally optimal: simulated annealing