LaGrange Multiplier Method for MDPs

Reference: G.C. Chow (1997)

Dynamic Economics: Optimization by the Lagrange Method

0. Infinite Horizon MDP problem

$$V(s) \equiv \max_{\alpha} E_{\alpha} \left\{ \sum_{t=0}^{\infty} \beta^t u(s_t, a_t) | s_0 = s \right\}, \tag{1}$$

1. "State Space Representation" of Law of Motion:

$$s_{t+1} = m(s_t, a_t) + \epsilon_t. \tag{2}$$

where $\{\epsilon_t\}$ is an *IID* "forcing process" with density $f(\epsilon)$.

2. Relationship to Transition Probability Formulation

$$p(s'|s, a) = f\left(s' - m(s, a)\right). \tag{3}$$

3. "Lagrangian" for MDP problem *Heuristic:* Introduce sequence of Lagrange Mutlipliers $\{\lambda_t\}$ which λ_{t+1} is the LM associated with "constraint" in equation (2). Lagrangian is:

$$\mathcal{L} = \max_{\{a_t\},\{s_t\}} E\left\{ \sum_{t=0}^{\infty} \beta^t u(s_t, a_t) + \beta^{t+1} \lambda_{t+1} \left[s_{t+1} - m(s_t, a_t) - \epsilon_t \right] \right\}. \tag{4}$$

- First Order Conditions are:

$$\frac{\partial \mathcal{L}}{\partial a_t} = 0 = \frac{\partial u}{\partial a_t}(s_t, a_t) - \beta \frac{\partial m}{\partial a_t}(s_t, a_t) E_t \left\{ \lambda_{t+1} \right\} \tag{5}$$

$$\frac{\partial \mathcal{L}}{\partial s_t} = 0 \Longrightarrow \lambda_t = \frac{\partial u}{\partial s_t}(s_t, a_t) - \beta \frac{\partial m}{\partial s_t}(s_t, a_t) E_t \left\{ \lambda_{t+1} \right\} \tag{6}$$

- Note $\{\lambda_t\}$ is a sequence of it functions. Conjecture that $\lambda_t = \lambda(s)$.

a. In this case (5) and (6) become a pair of *functional equations* determining the unknown functions $\{\lambda, \alpha\}$.

b. Given α equation (6) is equivalent to a *Fredholm Integral Equation of the Second Kind:*

$$\lambda(s) = \frac{\partial u}{\partial s}(s, \alpha(s)) - \beta \frac{\partial m}{\partial s}(s, \alpha(s)) \int \lambda(s') f(s' - m(s, \alpha(s))) ds' \tag{7}$$

Use infinite-dimensional analog of *policy iteration* to solve for $\{\lambda, \alpha\}$.

1. **Policy evaluation step:** Given trial policy α_k solve (7) for λ_{α_k}.

2. **Policy improvement step:** Given λ_{α_k} compute updated policy α_{k+1} which solves:

$$\begin{aligned} 0 = & \frac{\partial u}{\partial a}(s, \alpha_{k+1}(s)) - \\ & \beta \frac{\partial m}{\partial a}(s, \alpha_{k+1}(s)) \int \lambda_{\alpha_k}(s') f(s' - m(s, \alpha_{k+1}(s))) ds' \end{aligned} \tag{8}$$

Repeat steps (1) and (2) until given convergence tolerance for $\|\alpha_k - \alpha_{k+1}\|$ is attained.

Proposition:

$$\lambda(s) = \frac{\partial V(s)}{\partial s}. \tag{9}$$

Proof: Write Bellman equation for MDP problem:

$$V(s) = \max_{a \in A(s)} \left[u(s, a) + \beta \int V(s' - m(s, a)) f(s') ds' \right] \tag{9}$$

Assuming V is continuously differentiable, first order conditions for optimal $a = \alpha(s)$ are:

$$0 = \frac{\partial u(s, a)}{\partial a} + \beta \frac{\partial m(s, a)}{\partial a} \int \frac{\partial V(s' - m(s, a))}{\partial s} f(s') ds' \tag{10}$$

The *Envelope Theorem* implies:

$$\frac{\partial V(s)}{\partial s} = \frac{\partial u(s,\alpha(s))}{\partial a} - \beta\frac{\partial m(s,\alpha(s))}{\partial s}\int\frac{\partial V(s' - m(s,\alpha(s)))}{\partial s}f(s')ds' \tag{11}$$

It is easy to see that the pair of functional equations (10) and (11) determining $\{\partial V/\partial s, \alpha\}$ are identical to the functional equations (7) and (8) determining $\{\lambda, \alpha\}$. By uniqueness of solution to the Fredholm Integral Equation, we must have $\partial V/\partial s = \lambda$.

Is Lagrangian Approach "Easier" than Value Function Approach?

- **In special cases yes:** Lagrangian or "Euler equation" can allow us to get analytical solutions for α in problems where we can't get analytical solution for V from Bellman's equation.

- **In general case no:** since λ is the derivative of V, it is generally less smooth than V. Worst case computational complexity of solving Fredholm Integral Equation is *higher* than complexity of Bellman equation.

- In particular, *randomization can't break the curse of dimensionality for solving the FII but randomization can break the curse of dimensionality of solving the Bellman equation* (a result critically dependent on the contraction property of the Bellman equation).

- Nemirovsky and Yudin's work on complexity of optimization implies that Lagrangian approach can't help to find ϵ-approximation to global optimum in smooth, non-concave optimization problems such as

  $$\max_x u(x) \quad \text{s.t.} \quad p'x \leq y \tag{12}$$

  Consider Lagrangian saddle-point solution:

  $$\mathcal{L}(x,\lambda) = u(x) + \lambda(p'x - y) \tag{13}$$

Lagrangian is unconstrained but also has an extra variable λ. Consider Kuhn-Tucker conditions corresponding to (13):

$$\frac{\partial u(x)}{\partial x} - \lambda p \leq 0 \tag{14}$$

$$\lambda(p'x - y) = 0 \tag{15}$$

Question: Is it easier to solve (14), (15) for (x^*, λ^*) as a system of nonlinear equations or to solve (12) by hill-climbing methods?

Answer: Complexity of constrained optimization in d-dimensions is:

$$comp(\epsilon, d) = \Theta\left(\frac{1}{\epsilon^d}\right) \tag{16}$$

so *no method* (not even randomization) cannot break the worst case complexity of the general optimization problem.

Remark: Complexity of zero finding in d dimensions is at least as difficult as optimization in d-dimensions, i.e. both have complexity functions of form (16): see Sikorski 1984.

Exception to the rule: *Concave optimization.* Complexity function for concave optimization problems is given by:

$$comp(\epsilon, d) = \Theta\left(d\, log(1/\epsilon)\right) \tag{16}$$

So concave optimization is not subject to a curse of dimensionality.