ECE-GY 6263 Game Theory	Fall 2019

Lecture 8 — Oct 25, 2019

Prof. Quanyan Zhu — *Scribe: Shutian Liu*

1 Overview

In the last lecture we talked about Folk's theorem, continuous kernel games, and gave an example on economics application.

In this lecture we will continue the part on continuous kernel games, and talk about convergence of Gauss-Seidel algorithms, and then we will introduce robust estimation problems.

2 Continuous Kernel Games

First we restate the payoff function of a continuous kernel game:

$$J^i(u^i, \mu^{-i}) = \frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N} u_j^T R_{jk}^{(i)} \mu_k + \sum_{j=1}^{N} u_j^{(i)T} a_j + c_i, \tag{1}$$

where μ_i is player $j's$ strategy vector, $R_{jk}^{(i)}$ are constant matrices, $r_j^{(i)}$ and c_j are constant matrices. By:

$$\frac{\partial J^i}{\partial u^i} = 0, \tag{2}$$

we get:

$$R_{ii}^i u^i + \sum_{j\neq i} R_{ij}^i u^j + r_i^i = 0, \tag{3}$$

which is equivalent to $Ru = -r$ in the matrix form.
Consider a 2-person game, where matrix $R = \begin{bmatrix} R_{11}^1 & R_{12}^1 \\ R_{21}^2 & R_{22}^2 \end{bmatrix}$, $u = (u_1, u_2)^T$, and $r = (r_1^1, r_2^2)^T$.

Question: Does the solution of the system exist? And if it exists, is it unique?
We construct best-response dynamics to answer the question.

$$u^1(k+1) = C_1 u^{2(k)} + d1, \tag{4}$$

$$u^2(k+1) = C_2 u^{1(k)} + d2, \tag{5}$$

where $C_1 = -(R_{11}^1)^{-1} R_{12}^1$, $d1 = -(R_{11}^1)^{-1} r_1^1$, $C_2 = -(R_{22}^1)^{-1} R_{21}^1$, $d1 = -(R_{22}^1)^{-1} r_2^2$, updates in for pattern of (4)and (5) are called Gauss-Seidel iteration.
By plugging (5) into (4), we can get a linear dynamic system. We combine (4) and (5):

$$u(k+1) = Cu(k) + d, \tag{6}$$

where $u = \begin{bmatrix} u_1(k+1) \\ u_2(k+1) \end{bmatrix}$, $C = \begin{bmatrix} 0 & C_1 \\ C_2 & 0 \end{bmatrix}$, $d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$.
If the algorithm converges, denote the solution by u^*, then we must have:

$$u^* = Cu^* + d, \tag{7}$$

which is:

$$(I - C)u^* = d, \tag{8}$$

which gives the Nash Equilibrium. For when the algorithm will converge, we investigate:

$$u^1(k+1) = C_1 C_2 u^1(k-1) + c_1 d_2 + d_1, \tag{9}$$

and from linear system theory we know that the necessary and sufficient condition for the above iterates to converge is:

$$\rho(C_1 C_2) < 1, \tag{10}$$

where $\rho(A)$ denotes the largest eigenvalue of matrix A.

Example: (Mixed Strategies) Consider the following ZS game on the unit square. $x_1 \in [0,1]$, $x_2 \in [0,1]$, and $P1$ is a maximizer and $P2$ is a minimizer. The utility function is $u(x_1, x_2) = x_1^3 - 3x_1x_2 + x_2^3$. People can verify by taking the second derivatives that the utility funciton is convex in both players' decisions. Since P1 is a maximizer, then $x_1^* = 0$ or $x_1^* = 1$, and we get $u(0, x_2) = x_2^3$, and $u(1, x_2) = 1 - 3x_2 + x_2^3$.
Pick 0: $u(0, x_2) > u(1, x_2)$ we get $x_2 > \frac{1}{3}$, which gives:

$$\begin{aligned} x_1^* &= 1, \text{if } x_2 < \frac{1}{3} \\ x_1^* &= 0, \text{if } x_2 > \frac{1}{3} \\ x_1^* &\in \{0, 1\}, \text{if } x_2 = \frac{1}{3} \end{aligned} \tag{11}$$

P2 is a minimizer, convexity of utility function tells us to use the first-order conditions to find the optimal point. By taking derivative and set to 0, we get $x_2 = \sqrt{x_1}$.
Now if you plot x_1^* and x_2^* on the same figure, you will find that the two curves does not intersect, since one of them is not continuous. This tells us that there is no pure strategy saddle point.
Then, how about mixed strategies?
P1: she can randomizes between $x_1^* = 0$ and $x_2^* = 1$. Then the expected utility:

$$E_\alpha(u) = \alpha x_2^3 + (1 - \alpha)(1 - 3x_2 + x_2^3). \tag{12}$$

What can P2 do?

$$\frac{E_\alpha(u)}{\partial x_2} = 3x_2^2 - 3(1 - \alpha) = 0, \tag{13}$$

which gives us $x_2^* = \sqrt{1 - \alpha}$. P1 best response has to make 0 and 1 indifferent: $u(0, x_2^*) = u(1, x_2^*)$, by some calculation we get $x_2 = 0$ w.p. $\frac{8}{9}$ and $x_2 = 1$ w.p. $\frac{1}{9}$.
The topic of continuous kernel games is discussed in details in Chapter 4 of [2].

3 Robust Estimation

Let's first give an example.

Example: Robust estimation with Gaussians. $y = Ax + w$ is our model, where y is the observation, x is the truth, A is a known matrix, and w is noise. We assume $x \sim N(0, \Sigma_x)$, and $w \sim N(0, \Sigma_w)$.
The estimator $\hat{x} = \mu(y)$ is a function of the observation. And our target is to minimize the expected error of the estimation and the truth, given by the next problem:

$$\min \mathbb{E}(||x - \hat{x}||^2), \tag{14}$$

whose solution is given by the conditional expectation of the truth given the observation under Gaussians, $\hat{x} = \mathbb{E}(x|y) = (\Sigma_w + A^T \Sigma_x A)^{-1} A^T y$. Interested people can check materials on MMSE(minimum min squared error).

Let's come back to the contents of this lecture. Consider the following robust estimation problem:

$$\min_{\mu} [\max_{w,x} \frac{||x - \mu(y)||^2}{||x||^2 + ||w||^2}], \tag{15}$$

where our model is the same as in the example, but we don't have the assumptions of Gaussians anymore, here $w \in (-\infty, +\infty)$. The above problem is a fractional programming problem, for detailed derivation people can read materials on Dinkelback algorithms.
We proceed by forming an associated problem, introducing a parameter γ:

$$\min_{\mu} \max_{x,w} ||x - \mu(y)||^2 - \gamma^2(||x||^2 + ||w||^2). \tag{16}$$

Under the optimal solutions μ^*, x^*, w^*, we must have the objective equals to 0. Use the model $w = y - Ax$, then the above formulation becomes:

$$\min_{\mu} \max_{x,y} ||x - \mu(y)||^2 - \gamma^2(||x||^2 + ||y - Ax||^2), \tag{17}$$

here the maximization over x and y can be done separately, denote:

$$L(\mu(y), y) = \max_{x} ||x - \mu(y)||^2 - \gamma^2(||x||^2 + ||y - Ax||^2), \tag{18}$$

hence our probelm becomes:

$$\min_{\mu} \max_{y} L(\mu(y), y). \tag{19}$$

Now we make an important claim:

$$\min_{\mu} \max_{y} L(\mu(y), y) = \max_{y} \min_{u} L(u, y), \tag{20}$$

note here that $u \in \mathbb{R}^n$, while $\mu(\cdot)$ is a function.

Proof.

$$\max_{y} \min_{u} L(u, y) = \max_{y} L(\mu^*(y), y) \geq \min_{\mu} \max_{y} L(\mu(y), y). \tag{21}$$

At the same time, we also have:

$$\min_{\mu} \max_{y} L(\mu(y), y) \geq \max_{y} \min_{\mu} L(\mu(y), y) = \max_{y} \min_{u} L(u, y). \tag{22}$$

Together proves the claim. □

The power of this claim is obvious, since the difficulty of finding the best functions is much larger than of finding the best vectors. Note that Problem 15 of HW3 tells a similar thing. The proof can be stated in a similar way.
Now with the help of (20), our robust estimation problem becomes:

$$\max_y \min_u \max_x ||x-u||^2 - \gamma^2(||x||^2 + ||y-Ax||^2), \tag{23}$$

denote the objective by $J(u, x)$.
Since we have $\min_u \max_x J(u, x)$, we will want to exchange the order of min and max so that the two max operations can be grouped. To do this, we need $J(u, x)$ to be strictly convex in u and strictly concave in x, which is equivalent to:

$$I - \gamma^2 I - \gamma^2 A^T A < 0, \tag{24}$$

which requires all the eigenvalues of $\gamma^2 A^T A + \gamma^2 I - I > 0$.
We claim that if λ is an eigenvalue of $\gamma^2 A^T A + \gamma^2 I - I$, then $\nu = \frac{\lambda-\gamma^2+1}{\gamma^2}$ is an eigenvalue of $A^T A$.

Proof. If $\lambda x = (\gamma^2 A^T A + \gamma^2 I - I)x$ for some x$\neq 0, thenequivalenty, (\lambda - \gamma^2 + 1)x_{\overline{\gamma^2 = A^T Ax}}$, i.e. $\frac{(\lambda-\gamma^2+1)x}{\gamma^2}$ is an eigenvalue of $A^T A$. □

If $\lambda > 0$, we have equivalently $\gamma^2\nu + \gamma^2 - 1 > 0$, same as $\gamma^2 > \frac{1}{1+\nu}$, which is same as $\gamma^2 > \frac{1}{1+\lambda_{min}(A^T A)}$.
Suppose that $\gamma^2 < \frac{1}{1+\lambda_{min}(A^T A)}$, then $\gamma^2 A^T A + \gamma^2 I - I$ has at least one eigenvalue that is negative. Then the problem $\max_{x_1,x_2} x_1^2 - x_2^2 - 4x_1 x_2$ has no solution.
Suppose the opposite, then $J(u, x)$ is indeed strictly convex in u and strictly concave in x, then $\max_x \min_u J(u, x) = \min_u \max_x J(u, x)$. And we have that $u^* = x$. Then the problem reduces to:

$$\max_x J(x, u = x) = -\gamma^2(||x||^2 + ||y||^2 + x^T A^T Ax - 2y^T Ax). \tag{25}$$

First order condition gives $x = (I + A^T A)^{-1} A^T y$. Then $J(x = x^*, u = x) = -\gamma^2(||x||^2 + ||y - Ax||^2) \leq 0$, so when $y = 0$, we have $J(u, x) = 0$, which is $\max_y \min_u \max_x J = 0$.
Next we give theorems on the existence of solutions of the robust estimation problems, and the proofs can be found in the last chapter of [1].

Theorem 1. *Let* $\gamma^* = \sqrt{\frac{1}{1+\lambda_{min}(A^T A)}}$, *then if* $\gamma > \gamma^*$, $\min_\mu \max_{x,w} ||\mu(y)-x||^2 - \gamma^2(||x||^2 + ||w||^2) = 0$, *and* $\mu^*(y) = (I + A^T A)^{-1} A^T y$; *if* $\gamma < \gamma^2$, *then the problem has no solution.*

Theorem 2. *If* $\gamma > \gamma^*$, *then under* μ^*, $||\mu^*(y) - x||^2 - \gamma^2(||x||^2 + ||w||^2) \leq 0$, *for* $\forall x, w$, *which gives* $\frac{||\mu^*(y)-x||^2}{||x||^2+||w||^2} \leq \gamma^2$, $\forall x, w$.
If $\gamma < \gamma^*$, *it is not possible to derive the results.*

Note, under the case of Gaussians, the procedure introduced above gives exactly the same result as the conditional expectation.
More interesting materials are covered in [1], and topics like robust filtering and Kalman filter are also recommended.

Next lecture: we will talk about auction design and Bayesian games.

References

[1] Başar T, Bernhard P. H-infinity optimal control and related minimax design problems: a dynamic game approach[M]. Springer Science Business Media, 2008.

[2] Basar T, Olsder G J. Dynamic noncooperative game theory[M]. Siam, 1999.