A Learning Mechanism for Pareto Optima

Mabel Tidball[1] Alain Jean-Marie[2]

[1]INRA/LAMETA Montpellier
tidball@supagro.inra.fr

[2]INRIA
ajm@sophia.inria.fr

Popeye/GameComp Meeting, Grenoble, 22 may 2008

Outline

Progress

1. The learning model
   - General ideas
   - The specific algorithm
2. Example
3. Extensions

The idea

Game theoretic situation:

- Several players, which can act on a system,
- each seeking to maximize her utility function
- which depends on the actions of all players.

Incomplete information assumptions:

- Players do not know each other's utility functions
- They can observe the past actions of the others
- They have a "limited rationality"

Dynamic assumptions:

- The game is repeated some unspecified number of times
- But players optimize at each time step

Conjectures

Substitute to the unknown motivations of the opponents:

conjectures

Each player assumes that the others will play in function of their own decision: if player i plays e_i, she foresees that player j will play $A_{ij}(e_i)$.

Natural use of the "reaction function" concept.

A sort of "simultaneous Stackelberg" assumption.

Conjectural Optimization

Player i maximizes her profit given her conjecture

$$\max_{e_i} \Pi_i \left(A_{i1}(e_i), A_{i2}(e_i), \ldots, e_i, \ldots, A_{in}(e_i) \right) \ .$$

$\Longrightarrow$ a standard optimization problem

$\Longrightarrow$ produces a *conjectural* reaction functional

$e_i = r_i(A_{i1}(), \ldots, A_{in}())$

Dynamic Conjectures

The game is repeated:

- observations may not match predictions
- $\implies$ need for an adjustment process of conjectures.

Let $A'_{ij}(t)$ be the value player i *thinks she should have used.*
Adjustment process:

$$\begin{aligned} \dot{A}_{ij}(t) &= \mu_i(A'_{ij}(t) - A_{ij}(t)) && \text{continuous time} \\ A_{ij}(t+1) &= (1-\mu_i)A_{ij}(t) + \mu_i A'_{ij}(t) && \text{discrete time} \end{aligned}$$

with μ_i a speed of adjustment.

Consistency

Agents are boundedly rational... but not completely stupid.

- Errors are part of the game: agents can tolerate that their conjectures are not perfectly matched by reality.
- They should be able to detect that their conjectural mechanism is fundamentally wrong.
- But if the conjectures do not move too much or too quickly, they should be conforted in their beliefs.

$\Longrightarrow$ reasonable consistency occurs if the scheme converges.

The learning mechanism

- n players, e_i strategy of i, e profile of strategies,
- e^b a given benchmark strategy,
- Π^i instantaneous payoff of player i.

Player i makes a conjecture about j of the form

$$e_j^? = e_j^b + A_{ij}(e_i - e_i^b), \quad A_{ij} \in \mathbb{R}$$

and solves

$$\max_{e_i} \Pi^i(e_i, (e_j^b + A_{ij}(e_i - e_i^b))_{i \neq j}) \; .$$

Assume there exists a unique solution $e_i = r_i(e^b; A_i)$, $(A_i = (A_{ij})_{i \neq j})$.

Learning model (continued)

i observes that j has played e_j and concludes that her conjecture should have been A'_{ij} /

$$e_j = e_j^b + A'_{ij}\,(e_i - e_i^b), \quad \Longrightarrow \quad A'_{ij} = \frac{e_j - e_j^b}{e_i - e_i^b}$$

Adjustment process of conjectures

$$A_{ij}(t+1) = (1-\mu_i)A_{ij}(t) + \mu_i \frac{e_j(t) - e_j^b}{e_i(t) - e_i^b}$$

with $e_i(t) = r_i(e^b, A_i(t))$.

Fixed points

Consider the slightly more general scheme:

$$A_{ij}(t+1) = (1-\mu_i)A_{ij}(t) + \mu_i \frac{h_j(\bar{x}_j(t), A_j(t))}{h_i(\bar{x}_i(t), A_i(t))} ,$$
$$\bar{x}_i(t+1) = g_i(\bar{x}_i(t), A_i(t)) ,$$

Proposition: convergence

Assume that the evolution converges, that is, when $t \to \infty$:

$$A_{ij}(t) \to A_{ij} \qquad \bar{x}_i(t) \to \bar{x}_i , \quad \text{for all } i \text{ and } j.$$

Assume further that

$$h_i(\bar{x}, A_i) = \beta_i \neq 0 .$$

Then for all i, the vector A_i is proportional to the vector $(\beta_1, \ldots, \beta_N)$.

Properties of fixed points (continued)

Proposition: Pareto optimality

If e is a limit point obtained by the convergence of the adjustment recurrence then e is a candidate Pareto-optimal solution.

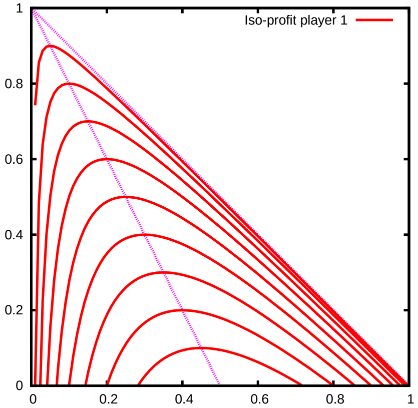
candidate: it verifies necessary optimal conditions.

Proposition

In the case of identical players: $r_i(e^b, r) = r(e^b, r)$, $e_i^b = e^b \; \forall i$; the recurrence converges to 1 for any $0 < \mu < 1$ and any (common) initial condition.

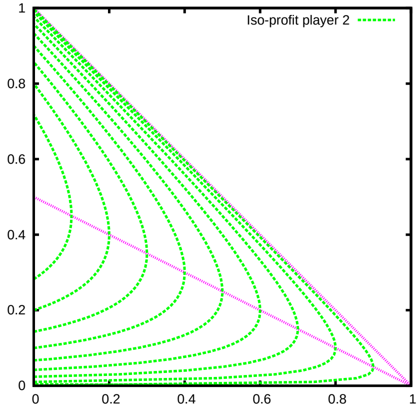
Pareto Optimality

Geometric interpretation of Pareto Optima.



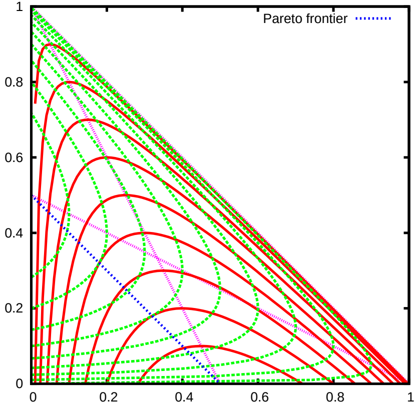
Pareto Optimality

Geometric interpretation of Pareto Optima.



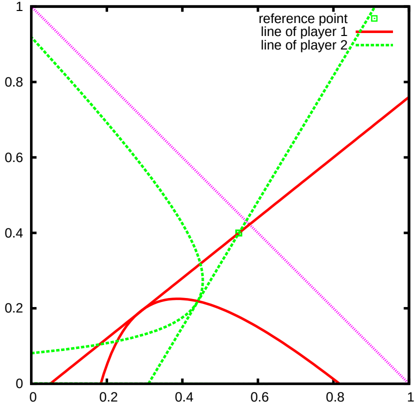
Pareto Optimality

Geometric interpretation of Pareto Optima.



Pareto Optimality (continued)

Geometric interpretation of Conjectural Optimization.



Stability of the process

Are the fixed point stable?

- No hope for global stability: multiplicity of fixed points and complicated attraction basins
- *Local* stability depends on the speed of adjustment.

Illustration for two players: Consider an adjustment scheme of the form:

$$\begin{aligned} \Phi_1(a_1, a_2) &= (1-\rho_1)a_1 + \rho_1 \ f(a_1, a_2) \\ \Phi_2(a_1, a_2) &= (1-\rho_2)a_2 + \rho_2 \ \frac{1}{f(a_1, a_2)} \ . \end{aligned}$$

Local stability for two players (continued)

Define the values:

$$S_1 = -\frac{\partial f}{\partial a_1}(a_1^*, a_2^*) \qquad S_2 = \frac{\partial f}{\partial a_2}\frac{1}{f^2}(a_1^*, a_2^*) \ .$$

where (a_1^*, a_2^*) is some fixed point.

Proposition

The recurrence is locally stable if and only if the following conditions hold:

$$\begin{aligned} \rho_1\rho_2(1+S_1+S_2) &> 0 \\ 4-2(\rho_1(1+S_1)+\rho_2(1+S_2))+\rho_1\rho_2(1+S_1+S_2) &> 0 \\ \rho_1(1+S_1)+\rho_2(1+S_2)-\rho_1\rho_2(1+S_1+S_2) &> 0 \end{aligned}$$

.

Progress

Example: Cournot's oligopoly

Setting of Cournot's oligopoly model:

- firms produce goods and sell them
- price is a given function of total production (inverse demand function)
- utility is net revenue: price $\times$ quantity minus production costs.

Linear inverse demand function and constant marginal costs (linear cost function), symmetric players:

$$\begin{aligned}\Pi^i(e_i, e_{-i}) &= (\alpha - \beta \sum_j e_j)e_i - (be_i + c) \\ &= \beta e_i(\Gamma - \sum_j e_j) - c.\end{aligned}$$

Where $\Gamma = \frac{\alpha - b}{\beta} > 0$.

Standard results

Pareto frontier:

$$\mathcal{P} \;=\; \{(e_1, \ldots, e_n) \mid \sum_j e_j = \frac{\Gamma}{2}\} \;, \quad \text{gain} = \frac{\beta\Gamma}{2}\, e_i$$

Nash equilibrium:

$$e_1 = e_2 = \ldots = e_n \;=\; \frac{\Gamma}{n+1} \;, \quad \text{gain} = \frac{\beta\Gamma^2}{(n+1)^2}$$

Conjectural reaction function:

$$r_i(A_i) \;=\; \frac{\Gamma - Q^b}{2\sum_j A_{ij}} \;+\; \frac{q_i^b}{2} \;.$$

Learning in Cournot's oligopoly

Theorem

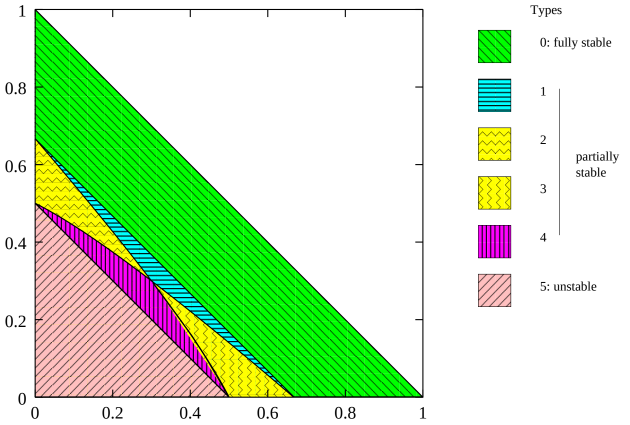
The unique fixed point of the adjustment process are $A_{ij} = e_j^b / e_i^b$ and the corresponding strategies are Pareto optima.

The learning procedure selects among the Pareto outcomes the only one with quantities proportional to that of the reference point: for all i, j,

$$\frac{e_i}{e_i^b} = \frac{e_j}{e_j^b}.$$

Zones of stability

Zones of stability in the Cournot case ($\Gamma = 1$).



Progress

The issue of the reference point

The reference point of the conjecture is considered as exogeneous.

This may make sense if:

- fixed by some authority:
  initial value of some pricing/auction process,
  regulation, from which unruly players try to depart
  ...
- it is the outcome of some internal analysis of the firms: e.g. a Nash equilibrium
- firms have announced objectives some time in advance:
  comparisons are made with respect to this plan

If not, the reference point should move as well.

The Friedman-Mezzetti model

Friedman-Mezzetti (2002) study a discounted repeated game, discrete time, infinite horizon, where agents form fixed conjectures about the others agents but update the reference point.

$$e_j(t+1) = e_j(t) + A_{ij} \times (e_i(t) - e_i(t-1))$$

Optimization:

$$e_i(t) = r_t^i(e(t-1), e_i(t-2)) \ .$$

Optimal policy $\rightarrow r_1^i$ at time $t = 1$, she observes $e(1)$ and applies r_1^i (rolling/receding horizon principle).

$$e_i(t) = r_1^i(e(t-1), e_i(t-2)), \qquad i = 1 \dots n$$

Adapting reference point in our learning model

Assume that the reference point moves in time as in Friedman & Mezzetti:

$$e_j(t+1) = e_j(t) + A_{ij}(e_i(t) - e_i(t-1))$$

For Cournot's duopoly:

$$\Pi^i(e_i, e_{-i}) = \beta e_i(\Gamma - \sum_j e_j) - c.$$

$$e_i = \frac{(1 + A_{ij})\Gamma}{(2 + A_{12})(2 + A_{21}) - 1}$$

Pareto optimality

The fixed point (e_1, e_2) is Pareto iff $A_{12}A_{21} = 1$.

Extension to be analyzed: Adapting conjectures and reference points.

Conclusions

The proposed mechanism

- does not require knowledge on "private" data of opponents (preferences, beliefs...)
- and is apt at selecting one Pareto outcome
- but is not necessarily stable, and may also converge to bad outcomes.

Existing results call for further studies on:

- Ways to enhance stability of the process
- Learning with adaptation of conjectures *and* the reference point.

Bibliography

- C. Figuières, A. Jean-Marie, N. Quérou and M. Tidball, *Theory of Conjectural Variations*, World Scientific Computing, 2004.
- A. Jean-Marie and M. Tidball, "Adapting behaviors in a learning model", *J. Economic Behavior and Organization*, 2006.
- Dixon, H. and Somma, E. (2001), "The Evolution of Consistent Conjectures", Discussion Papers in Economics, N^o 2001/16, University of York, forthcoming in *Journal of Economic Behavior and Organization*.
- Friedman, J.W. and Mezzetti, C. (2002), "Bounded Rationality, Dynamic Oligopoly, and Conjectural Variations", *Journal of Economic Behavior and Organization*, Vol. 49, pp. 287–306.