Principal-Agent Games  
- Equilibria under Asymmetric Information - 

**Ulrich Horst**\(^1\)

Humboldt-Universität zu Berlin  
Department of Mathematics  
and  
School of Business and Economics

Work in progress - Comments welcome

\(^1\) Based on ongoing work with Francoise Forges and Robert Simon
Outline

• Motivating example

• Repeated games of incomplete information
  • Static games
  • The game (setup, examples, ...)
  • Approachability and individual rationality
  • Joint-plan equilibria

• The mathematical core
  • Contracts and cooperative games
  • The general model
  • Existence of equilibria
    • compact type spaces
    • finite type spaces
  • Example: Risk sharing under asymmetric information
Motivating Example
Motivating Example

Informed player (agent) with random endowment $H$ wants to share his risk exposure with uninformed player (principal).

- Randomness is described by a probability space

$$\hat{\Omega} := K \times \Omega, \quad \hat{\mathcal{F}} := \mathcal{K} \otimes \mathcal{F}, \quad \hat{\mathbb{P}} := p_0 \otimes \mathbb{P}$$

where we assume for the moment that $|K| < \infty$.

- For $X : \hat{\Omega} \to \mathbb{R}$ the agent’s type-dependent utility is

$$U^k(X) = \mathbb{E}_{\delta_k(\cdot) \otimes \mathbb{P}}[u(X)]$$

while the principal’s utility function is

$$\hat{U}(X) = \mathbb{E}_{p_0 \otimes \mathbb{P}}[\hat{u}(X)]$$
Motivating Example

The agent sells contingent claims from a set $\mathcal{X}$ to the principal.

- A contract comprises a stochastic kernel $\mu : K \Rightarrow S$ and a mapping $q : S \rightarrow \mathcal{X}$ where $S$ is a space of signals.
- The principals accepts a claim $q(s) \in \mathcal{X}$ only if

$$\mathbb{E}_{p_s \otimes \mathbb{P}}[\hat{u}(q(s))] \geq 0$$

The agent has a type-dependent reservation utility; $\forall k \in K$:

$$\mathbb{E}_{\delta_k(\cdot) \otimes \mathbb{P}}[u(H + q(s))] \mu(k; ds) \geq \mathbb{E}_{\delta_k(\cdot) \otimes \mathbb{P}}[u(H)]$$

Does a “reasonable”, optimal, ... contract $\left(\mu, q\right)$ exist and what is the general mathematical structure behind this problem?
Part I
Static games
A 2-person (non-cooperative) 1-stage game is a described by:

\[ \Gamma_1 := (I, J, A, B) \]

where \( I \) and \( J \) are the finite action spaces and the matrices \( A_{i,j}, B_{i,j} \) are the players’ payoff functions.
The mixed-extension of the game $\Gamma_1^*$ is defined as

$$\Gamma_1^* := (\Delta(I), \Delta(J), h_1, h_2)$$

where $\Delta(M)$ is the set of probability distributions on the set $M$ and the payoff functions are:

$$h_1(s, t) := \sum_{i \in I, j \in J} s_i \cdot A_{i,j} \cdot t_j = < s, At >$$

$$h_2(s, t) := \sum_{i \in I, j \in J} s_i \cdot B_{i,j} \cdot t_j = < s, Bt > .$$

Notice that payoff functions are bi-linear.
An equilibrium in mixed strategies of the game $\Gamma^*_1$ is any pair of strategies $(s^*, t^*)$ such that

$$< s^*, A t^* > = \max_{s \in \Delta(I)} < s, A t^* >$$

and

$$< s^*, B t^* > = \max_{t \in \Delta(J)} < s^*, B t t >$$
Definition (Zero-Sum Games)

The game $\Gamma_1 := (I, J, A, B)$ is called a zero-sum game if $B = -A$. Its value is given by

$$w := \max_s \min_t < s, At > = \min_t \max_s < s, At >.$$ 

Remark

- The value of a zero-sum game and hence an equilibrium exists.
- The ‘min-max’ approach only works for games with scalar payoffs, not vector payoffs.
Repeated games of incomplete information
Repeated Games

Repeated game of incomplete information on one side ($\Gamma_\infty(p)$):

- there are two players, labelled PI and PII
- there is a finite set $K$ of types; nature chooses $k \in K$ according to a known probability distribution $p \in \Delta(K)$
- PI is informed about $k \in K$, PII is not
- the payoff matrices in state $k \in K$ are $A^k$, respectively $B^k$
- the game with payoff matrices $(A^k, B^k)$ is repeated infinitely often
- after each state of play, both players are informed about each other’s move
- payoffs in the infinitely repeated game are the average payoffs of the stage games.
Theorem (Zero-sum games)

Consider a zero-sum \((B^k = -A^k)\) repeated game of incomplete information and let \(\Gamma_1(p)\) the 1-stage game with payoff matrix

\[
A(p) := \sum_k p^k A^k.
\]

The game \(\Gamma_1(p)\) has a value \(a^*(p)\). The value of the infinitely repeated zero-sum game \(\Gamma_\infty(p)\) is

\[
v_\infty(p) := \text{Cav}(a^*)(p)
\]

where \(\text{Cav}(u)\) is the concave envelope of \(u\), i.e. the smallest concave function dominating \(u\).
Remark

- It is clear, the PI can guarantee himself $Cav(a^*)(p)$ in $\Gamma_\infty(p)$.
- It less obvious that PII can prevent him from getting more. The argument uses Blackwell’s approachability theorem, an extension of v. Neumann’s min-max Theorem to games with vector-valued payoffs.
Approachability
Definition

Let $C$ be a payoff matrix, $\{(i_t, j_t)\}$ be a sequence of moves and put

$$c_n := \frac{1}{n} \sum_{t=1}^{n} C_{i_t,j_t}.$$ 

The sequence of random variables $\{c_t\}$ approaches a set $\mathcal{C}$ if

$$d(c_n, \mathcal{C}) \longrightarrow 0 \quad \text{a.s.}$$

A set of (vector) payoffs $\mathcal{C}$ is approachable for some player if he/she has a strategy such that

$$d(c_n, \mathcal{C}) \longrightarrow 0 \quad \text{a.s.}$$

regardless of his/her opponent’s strategy.
Theorem (Blackwell (1956))

In a repeated game with payoff matrices $A^k$ for Pl, a closed convex set of (vector) payoffs $\mathcal{C}$ is approachable for PII if and only if

$$a^*(q) \leq \max_{c \in \mathcal{C}} (q \cdot c) \quad \text{for all} \quad q \in \Delta(K). \quad (1)$$

Remark

Consider an infinitely repeated game with payoff matrices $A^k$ for Pl and $B^k$ for PII.

- PII can force Pl’s payoff to belong to a set $\mathcal{C}$ if (1) holds.
- PII can guarantee himself $\text{Vex}(b^*(p))$. 
Individual Rationality

A key concept in repeated games is (interim) individual rationality.

- Both players consider “their” respective zero-sums games:

  $$A(p) \text{ and } -B(p)$$

  with respective values $$a^*(p)$$ and $$b^*(p)$$.

- Using the arguments from the zero-sum case one sees that:
  - PII can guarantee herself $$\text{Vex}(b^*(p))$$ in $$\Gamma_\infty(p)$$
  - PII can prevent PI from getting more than $$x^k$$ ($$k \in K$$) in the same game if

    $$x \cdot q \geq a^*(q) \text{ for all } q \in \Delta(K)$$
Definition (Individual rationality)

A payoff vector $y \in R^{|K|}$ is called individually rational for PI if

$$y \cdot q \geq a^*(q) \quad \text{for all} \quad q \in \Delta(K).$$

A payoff $z \in R$ is called individually rational for PII if

$$z \geq \text{Vex}(b^*(p)) \quad \text{for all “possible”} \quad p \in \Delta(K).$$
Joint plans
A joint-plan is a cooperative agreement between the players; the main idea is as follows

• first change the uninformed player’s belief about the distribution of types using signals,
• then play a fixed (joint) action for the rest of the game,
• punish the opponent if he/she does not follow the pre-agreed plan.

More formally, a joint-plan comprises:

• a finite set of signals $S$ and a signaling kernel $\mu : K \Rightarrow S$
• for any posterior probability $p_s$ a joint action $x^s \in \Delta(I \times J)$
• punishment strategies to be implemented when the opponent does not follow $x^s$
A joint-plan describes an equilibrium if there exists $y \in R^{|K|}$ s.t.

- Incentive compatibility (IC) for PI: if $\mu(k; s) > 0$, then
  
  $$y^k = \max_{s'} \sum_{i,j} A^k_{i,j} x^s_{i,j}$$

- Individual rationality (IR) for PI:
  
  $$y \cdot q \geq a^*(q) \quad \forall q \in \Delta(K)$$

- Acceptability (A) for PII: if $\mu(k; s) > 0$ for some $k \in K$, then
  
  $$\sum_{i,j} B_{i,j}(p_s) x^s_{i,j} \geq Vex(b^*)(p_s)$$
Theorem (Sorin ('83); Simon, Spiez & Torunczyk ('95))

A joint plan equilibrium exists

Remark

Sorin (1983) established existence of equilibrium for $|K| = 2$; Simon et al. (1995) extended his result to general finite type spaces using methods and techniques from algebraic topology.
A more abstract point of view

A joint-plan equilibrium can viewed as a special case of a cooperative game where

- one party, the agent, has private information and one, the principal, has not
- the principal suggests a cooperative agreement (contract) to the agent
- as part of the agreement, information may be disclosed
- there are side-constraints (outside options)

One looks for a cooperative agreement in equilibrium. But what is the participation constraint?
Some Comments

In a repeated games framework

- the concept of individual rationality is based on the idea of approachability, a dynamic concept
- and the willingness to cooperate is observable (punishment)
Part II
Preferences and Types

- There is a finite (or compact) set $K$ of agent-types.
- There is a compact set $X_1$ of actions for the agent and a compact set of actions $X_2$ for the principal.
- The preferences are described by concave utility functions $U^k : X_1 \times X_2 \to R$.
- The preferences of the principal are (not yet) important.
Interim Individual Rationality

The agent can accept the principal’s proposal or not.

- Upon acceptance, a signal is sent, an action implemented
- Upon rejection, the following (1-stage) game is played:

\[ U^k(\cdot) \text{ for all } k \in K \text{ and } (X_1, X_2) \]

- A payoff vector \( v = (v^k) \) is interim individually rational if:

\[ \exists x_2 \in X_2 \text{ s.t. } \forall (x_1, k) \in X_1 \times K : U^k(x_1, x_2) \leq v^k \]

In the complete information case, this is just “min max”. 
Lemma (Non-standard (IR) condition)

Assume that the following function is convex for $k = 1, 2$:

$$\phi^k(x_2) = \max_{x_1 \in X_1} U^k(x_1, x_2)$$

and define a mapping $f : \Delta(K) \rightarrow R$ via

$$f(q) := \min_{x_2 \in X_2} \sum_k q^k \phi^k(x_2)$$

Then, a payoff vector $v$ is interim individually rational if and only if

$$q \cdot v \geq f(q) \quad \text{for all} \quad q \in \Delta(K).$$
Example (Cournot game with private information)

Consider a type dependent utility function of the form

\[ U^k(x_1, x_2) = x_1(d - x_1 - x_2) - c^k x_1 \]

A direct calculation shows that that \( \varphi^k : X_2 \rightarrow R \) is convex.

Remark

The individual rationality condition

\[ q \cdot v \geq f(q) \quad \text{for all} \quad q \in \Delta(K) \]

now appears in a static game with concave utility functions. When \( f \) is convex, it reduces to

\[ v^k \geq f(k), \]

that is, to a model with type-dependent outside options.
Equilibria in principal-agent models
A More Abstract Point of View

A joint-plan equilibrium is a special case of the following model:

- there are two players: a principal and an agent
- compact type space $K$; distribution $p_0 \in \Delta(K)$.
- agents have type-dependent concave utility function $U^k(\cdot): X \rightarrow \mathbb{R}$
  on a convex space of joint actions $X$
- a signaling kernel $\mu : K \Rightarrow S$
- a kernel $Q : S \Rightarrow X$ that chooses joint actions

A pair $(\mu, Q)$ will be called a contract.
A More Abstract Point of View: \( (IC) \)

- **Incentive compatibility** for the agent prevails if each type is indifferent between the results of his actions (not signals!):

\[
U^k(\cdot) \equiv u^k \quad (\mu Q)(k; \cdot) \text{-a.s.}
\]

and

\[
U^k(\cdot) \leq u^k \quad (p_0 \mu Q) \text{-a.s.}
\]

- In models of expected utility, this reduces to the condition that types are indifferent between the signals sent.
A More Abstract Point of View: (IR)

Individual rationality in the game is described

• for the principal in terms of a non-empty correspondence

\[ F : \Delta(K) \Rightarrow X \]

• for the agent in terms of a function

\[ f : \Delta(K) \rightarrow \mathbb{R}. \]
Definition (Equilibrium)

A contract \((\mu, Q)\) is an **equilibrium** if the following conditions hold:

- **(A)** for the principal: for almost all \((x, s) \in X \times S\):
  \[
x \in F(p_{x,s})
  \]

- **(IC)** for the agent: for almost all types:
  \[
  U^k(\cdot) = u(k) \ (\mu Q)(k; \cdot) \text{-a.s};
  \]
  \[
  U^k(\cdot) \leq u(k) \ (p_0 \mu Q) \text{-a.s}.
  \]

- **(IR)** for the agent: for \(p_0\)-a.e. \(k \in K\):
  \[
  u \cdot p \geq f(p) \text{ for all } p \in \Delta(K)
  \]
Theorem (Existence of equilibrium - compact type spaces)

Suppose that the following conditions hold:

- \( X, S \) and \( K \) are compact metric spaces; \( K \) is Polish
- The utility function \((k, x) \mapsto U^k(x)\) is continuous
- For all \( k \in K \) the map \( x \mapsto U^k(x) \) is concave
- The correspondence \( F : \Delta(K) \rightrightarrows X \) is u.s.c., non-empty, convex-valued and inversely convex:
  \[ x \in F(p) \text{ for all } p \in \mathcal{P}, \text{ then } x \in F(q) \in \text{Conv}(\mathcal{P}) \]
- The function \( f : \Delta K \to R \) is weakly continuous.

Then an equilibrium exists (there is no reason to assume that \( Q \) is degenerate).
Theorem (Existence of equilibrium - finite type spaces)

Suppose that $K$ is finite and that the following conditions hold:

- $X$ is a compact metric space
- for all $k \in K$ the map $x \mapsto U^k(x)$ is concave
- the correspondence $F : \Delta(K) \rightrightarrows X$ is u.s.c., non-empty, convex-valued.

Then an equilibrium exists with a degenerate signaling kernel

$$Q(s; dx) = \delta_{x(s)}(dx).$$

In particular, no further randomization over actions is needed.
Application: risk sharing under asymmetric information
Back to the Motivating Example

- Chose the action space to be a set of contingent claims:

\[ \mathcal{X} := \left\{ \sum_{i=1}^{m} \lambda_i X^i : \lambda_i \geq 0, \sum_i \lambda_i = 1 \right\} \]

- The correspondence \( F \) is given by

\[ F(p) := \{ X \in \mathcal{X} : \mathbb{E}_{\hat{P} \otimes \mathbb{P}}[\hat{u}(X))] \geq 0 \} \]

- The function \( f \) is given by

\[ f(k) := \mathbb{E}_{\delta_k(\cdot) \otimes \mathbb{P}}[u(H)] \]

Theorem

*If the type space is compact, then an equilibrium exists.*
Preferences of Non-Expected Utility Type

- Same framework as above with $|K| < \infty$, $|\Omega| < \infty$
- The agent’s utility function is

$$U^k(X) = \min_{Q \in \mathcal{Q}} \left\{ \mathbb{E}_{\delta_k(\cdot) \otimes Q}[-X] - \alpha(Q) \right\}$$

- The principal has the utility function:

$$U(X) = \min_{\hat{Q} \in \mathcal{Q}_p} \left\{ \mathbb{E}_{\hat{Q}}[-X] - \hat{\alpha}(\hat{Q}) \right\}.$$  

Furthermore:

$$F(p) = \{X : U(X) \geq 0\}, \quad f(k) = 0.$$  

**Theorem**

*Under mild conditions on the map $p \mapsto \mathcal{Q}_p$, an equilibrium exists.*
(A Conjecture on) Optimality

• Suppose principal can chose $\mu$ and has a preference functional

$$G : \Delta(K) \rightarrow \mathbb{R}$$

such as

$$G(\nu) = \int \{c(y) - \pi\} \nu(d(y, \pi))$$

Then her objective would be

$$\max\{G(p_0\mu Q) : (\mu, Q) \text{ is an equilibrium}\}$$

• For one-dimensional types this is a standard problem addressed by convex analysis methods

**Theorem (Conjectured)**

*Under mild conditions an optimal contract for the principal exists.*
Conclusion

- Proved a general existence of equilibrium result for principal-agent models
- The method allows for multi-dimensional agent-types
- Application to risk-sharing under asymmetric information
- Future research:
  - optimal contract design for multi-dimensional types
  - partial revelation of information and dynamic models
  - (zero-sum) repeated-games with incomplete information with uncountably many types
  - ...


Thanks!