Robust Mechanism Design and Implementation: A Selective Survey

Dirk Bergemann and Stephen Morris

January 2009
UCI Conference on Adaptive Systems and Mechanism Design

Introduction

- mechanism design and implementation literatures are theoretical successes
- mechanisms seem too complicated to use in practise...
- successful applications of commonly include ad hoc restrictions
- simplicity, non-parametric, detail free, ex post equilibrium...

Weaken Informational Assumptions

- if the optimal solution to the planner's problem is too complicated or sensitive to be used in practice, presumably the original description of the planner's problem was itself flawed
- can improved modelling of the planner's problem endogenously generate the "robust" features of mechanisms that researchers have been tempted to assume?
- weaken informational requirements
- specifically weaken common knowledge assumption in the description of the planner's problem: the "Wilson doctrine"

"Game theory has a great advantage in explicitly analyzing the consequences of trading rules that presumably are really common knowledge; it is deficient to the extent that it assumes other features to be common knowledge, such as one agent's probability assessment about another's preferences or information.

I foresee the progress of game theory as depending on successive reductions in the base of common knowledge required to conduct useful analyses of practical problems. Only by repeated weakening of common knowledge assumptions will the theory approximate reality." Wilson (1987)

Weakening Common Knowledge

- in game theory, Harsanyi (1967/68), Mertens and Zamir (1985) established that relaxing common knowledge assumptions is equivalent to adding types...
- environments with incomplete information can be modeled as a Bayesian game where wlog there is common knowledge among players of (i) each player's type spaces and (ii) each type's beliefs over types of other players
- economic analysis assumes smaller type spaces than universal type space *yet maintains common knowledge of* (i) and (ii)
- are the implicit common knowledge assumptions that come from working with small types spaces problematic? perhaps especially in mechanism design (Neeman (2004))?

Our Agenda (circa 2000)

- introduce rich (higher order belief) types and strategic uncertainty into mechanism design literature
- relax (implicit) common knowledge assumptions by going from "naive" type space to "universal" type space
- find robust mechanism with large type space and obtain comparative statics results across type spaces
- in particular
  1. briefly establish a few easy benchmark "abstract" results
  2. develop a close link between this robust approach and applications
- a decade and seven papers/notes later, we are kind of done with (1)

Seven Papers: A Selective Survey

- since 2000, Stephen Morris and I have written a series of papers on "Robust Mechanism Design":

1. "Robust Mechanism Design", *Econometrica (2005)*
2. "Ex Post Implementation" *Games and Economic Behavior (2008)*
3. "Robust Implementation in Direct Mechanisms" *Review of Economic Studies (2009)*
4. "An Ascending Auction for Interdependent Values" *American Economic Review (2007)*
5. "The Role of the Common Prior Assumption in Robust Implementation" *Journal of European Economic Association (2008)*
6. "Robust Implementation in General Mechanisms" *(2009)*
7. "Robust Virtual Implementation" *Theoretical Economics (2009)*

Payoff Environment

- agent $i \in \mathcal{I} = \{1, 2, ..., I\}$
- i's "payoff type" $\theta_i \in \Theta_i$
- payoff type profile $\theta \in \Theta = \Theta_1 \times \cdots \times \Theta_I$
- social outcome $a \in A$
- utility function $u_i : A \times \Theta \longrightarrow \mathbb{R}$
- social choice function $f : \Theta \longrightarrow A$

Type Spaces

- richer type space T_i than payoff type space Θ_i
- i's *type* is $t_i \in T_i$, t_i includes description of:
- payoff type:

$$\widehat{\theta}_i : T_i \longrightarrow \Theta_i$$

  $\widehat{\theta}_i(t_i)$ is i's payoff type of t_i
- beliefs about types T_{-i} of other players:

$$\widehat{\pi}_i : T_i \longrightarrow \Delta(T_{-i})$$

  $\widehat{\pi}_i(t_i)$ is i's belief type of t_i
- *type space* is a collection $\mathcal{T} = \{T_i, \widehat{\theta}_i, \widehat{\pi}_i\}_{i=1}^{I}$
- type t_i contains information about preferences and information of others agents, i.e. beliefs and higher-order beliefs

Allocating a Single Object

- I agents
- agent i has a payoff type $\theta_i \in \Theta_i = [0, 1]$
- agent i's valuation of the "object" is $v_i(\theta_1, ..., \theta_I)$
- interdependent value model (Maskin (1992), Dasgupta and Maskin (1999))
- all agents have quasi-linear utility
- don't know anything about agent i's beliefs and higher order beliefs about θ_{-i}

- value of i does only depend on payoff type of agent i:

$$v_i(\theta) = \theta_i$$

- "second price sealed bid auction", direct mechanism
- i bids $b_i \in [0,1]$,
- rule of second price auction: highest bid wins, pays second highest bid
- truthful reporting leads to efficient allocation of object: $q^*(\theta)$ (efficient correspondence)

$$q_i^*(\theta) = \begin{cases} \frac{1}{\#\{j:\theta_j \geq \theta_k \text{ for all } k\}}, & \text{if } \theta_i \geq \theta_k \text{ for all } k \\ 0, & \text{if otherwise} \end{cases}$$

- dominant strategy to truthfully report type

- linear example:

$$v_i(\theta) = \theta_i + \gamma \sum_{j \neq i} \theta_j$$

- "generalized VCG mechanism", direct mechanism
- agent bids $b_i \in [0, 1]$, highest bid wins, pays the second highest bid PLUS γ times the bid of others:

$$\max_{j \neq i} \{b_j\} \; + \; \gamma \sum_{j \neq i} b_j$$

- truthful reporting is an ex post equilibrium of the direct mechanism if $\gamma \leq 1$, cf. Maskin (1992)

Detour for Definitions

- "ex post equilibrium": each type of each agent has an incentive to tell truth *if* he expects all other agents to tell the truth
- under private values, ex post equilibrium is equivalent to dominant strategies equilibrium
- if truthtelling is an ex post equilibrium of the direct mechanism for an allocation rule (*including* transfers), then the allocation rule "ex post incentive compatible" [EPIC]

Definition

A scf f is interim incentive compatible on type space $\mathcal{T}$ if

$$\int_{t_{-i}} u_i\left(f\left(t\right),\widehat{\theta}\left(t\right)\right) d\widehat{\pi}_i\left(\left.t_{-i}\right|t_i\right) \geq \int_{t_{-i}} u_i\left(f\left(t_i',t_{-i}\right),\widehat{\theta}\left(t\right)\right) d\widehat{\pi}_i\left(\left.t_{-i}\right|t_i\right)$$

for all i, $t \in T$ and $t_i' \in T_i$.

Definition

A scf f is ex post incentive compatible if, for all i, $\theta \in \Theta$, $\theta_i' \in \Theta_i$:

$$u_i\left(f\left(\theta\right),\theta\right) \geq u_i\left(f\left(\theta_i',\theta_{-i}\right),\theta\right).$$

- Compare: A scf is dominant strategy incentive compatible if for all i and all θ, θ' :

$$u_i\left(f\left(\theta_i,\theta_{-i}'\right),\theta\right) \geq u_i\left(f\left(\theta_i',\theta_{-i}'\right),\theta\right)$$

- when does there exist a mechanism with the property that for any beliefs and higher order beliefs that the agents may have, there exists an equilibrium where an acceptable outcome is chosen?
- in single good example, consider "efficient correspondence" q^* of object and any suitable transfers

Robust Mechanism Design II

- a sufficient condition is that there exists an allocation rule as a function of agents payoff type that is "ex post incentive compatible," i.e., in a payoff type direct mechanism, each agent has an incentive to announce his type truthfully whatever his beliefs about others' payoff types
- the larger the type space, the more incentive constraints there are, the harder it becomes to implement scc
- from smallest type space: "naive type space" to largest type space: "universal type space"

Theorem (2005)

f is ex post incentive compatible if and only if f is interim incentive compatible on every type space $\mathcal{T}$.

- ex post equivalence can be generalized to social choice correspondence with product structure
- ex post equivalence fails to hold for scc in general, e.g. efficient allocation with budget balance
- ex post equilibrium notion incorporates concern for robustness to higher-order beliefs
- in private values case, ex post implementation is equivalent to dominant strategies implementation:
- c.f. Ledyard (1978) and Dasgupta, Hammond and Maskin (1979) in private value environments and dominant strategies

Ex Post Implementation

- when does there exist a mechanism such that, not only is there an ex post equilibrium delivering the right outcome, but *every* ex post equilibrium delivers the right outcome?
- thus there is *full* implementation under the solution concept of ex post equilibrium - and we call this *ex post implementation*
- in addition to ex post incentive compatibility - an ex post monotonicity condition is necessary and almost sufficient
- ex post monotonicity condition neither implies nor is implied by Maskin monotonicity (necessary and almost sufficient for implementation under complete information)
- generalized VCG satisfies ex post monotonicity condition if $I \geq 3$ and $\gamma \neq 0$

- when does there exist a mechanism with the property that for any beliefs and higher order beliefs that the agents may have, *every* interim equilibrium has the property that an acceptable outcome is chosen?
- we call this "robust implementation"
- this is *not* the same as the ex post implementation: to rule out bad equilibria, it was enough to make sure you could not construct a "bad" ex post equilibrium; for robust implementation, we must rule out bad Bayesian, or interim equilibria on all type spaces

Back to the Single Object Example....

- robust implementation fails *even in the private values case*, since truthtelling is only a weak best response and there are many equilibria leading to inefficient outcomes in second price sealed bid auctions.
- robust implementation of the efficient allocation is not possible in the single object example (with private or interdependent values) even if augmented (but well-behaved) mechanisms are allowed.
- but robust implementation is achievable for a nearly efficient allocation under additional restrictions....

...and to Private Values: The Modified Second Price Auction

- with probability

$$1 - \varepsilon$$

  allocate object to highest bidder and pay second highest bid

- for each i, with probability

$$\varepsilon \cdot \frac{b_i}{I}$$

  i gets object and pays $\frac{1}{2} b_i$

- truth-telling is a strictly dominant strategy and ε-efficient allocation is robustly implemented

Interdependent Values: Modified VCG Mechanism

- with probability

$$1 - \varepsilon$$

allocate object to highest bidder i and winner pays

$$\max_{j \neq i} \{b_j\} \; + \; \gamma \sum_{j \neq i} b_j$$

- for each i with probability

$$\varepsilon \cdot \frac{b_i}{I}$$

i gets object and pays

$$\frac{1}{2} b_i + \gamma \sum_{j \neq i} b_j$$

- truth telling is a strict ex post equilibrium

The Modified Generalized VCG Mechanism

- but existence of strict ex post equilibrium does *not* imply robust implementation
- in fact, this mechanism robustly implements the efficient outcome if and only if

$$|\gamma| < \frac{1}{I-1}$$

- and no mechanism robustly implements the efficient outcome if

$$|\gamma| \geq \frac{1}{I-1}$$

- c.f. Chung and Ely (2001)

Rationalizability: A Key Epistemic Result

A message m_i can be sent by an agent with payoff type θ_i in an interim equilibrium on some type space if and only if m_i is "incomplete information rationalizable" in the following sense:

1. First, suppose that every payoff type θ_i could send any message m_i
2. Delete those messages m_i that are not a best response to some conjecture over payoff type - message pairs of the opponents that have not yet been deleted
3. Repeat step 2 until you converge

Illustrate Rationalizability in Generalized VCG mechanism

- direct mechanism: message m_i is report θ_i'
- agent i conjectures that other agents have type θ_{-i} and report θ_{-i}' :

$$\lambda_i \left(\theta_{-i}', \theta_{-i}\right) \in \Delta \left(\Theta_{-i} \times \Theta_{-i}\right)$$

- $\beta_i^k \left(\theta_i\right)$ is set of reports i might send for some conjecture $\lambda_i \left(\theta_{-i}', \theta_{-i}\right)$ over his opponents' types θ_{-i} and reports θ_{-i}', with restriction on conjecture $\lambda_i \left(\theta_{-i}', \theta_{-i}\right)$ that each type θ_j of agent j sends a message in $\beta_i^{k-1} \left(\theta_j\right)$
- initialize at step $k = 0$ by allowing all reports $\beta_i^0 \left(\theta_i\right) = [0, 1]$

- with linear interdependence:

$$v_i(\theta) = \theta_i + \gamma \sum_{j \neq i} \theta_j$$

  ex post incentive compatible transfer $y_i^*(\theta)$ is quadratic in reports θ'
- agent i with type θ_i has linear best response θ_i':

$$\theta_i' = \theta_i + \gamma \sum_{j \neq i} \left(\theta_j - \theta_j'\right)$$

- linear best response allows us characterize set of best response reports by $\beta_i^k(\theta_i)$:

$$\beta_i^k(\theta_i) = \left[\underline{\beta}_i^k(\theta_i), \overline{\beta}_i^k(\theta_i)\right]$$

with linear best response θ_i':

$$\theta_i' = \theta_i + \gamma \sum_{j \neq i} \left(\theta_j - \theta_j'\right)$$

the upper bound $\overline{\beta}^k(\theta_i)$ is determined by:

$$\overline{\beta}^k(\theta_i) = \min\{1, \theta_i + \ \gamma \max_{\left\{\left(\theta'_{-i}, \theta_{-i}\right) : \theta_j' \in \beta_j^{k-1}(\theta_j),\ \forall j \neq i\right\}} \sum_{j \neq i} (\theta_j - \theta_j')\}$$

or through lower bound $\underline{\beta}_j^{k-1}(\theta_j)$:

$$\overline{\beta}^k(\theta_i) = \min\{1, \theta_i + \ \gamma \max_{\theta_{-i}} \sum_{j \neq i} (\theta_j - \underline{\beta}_j^{k-1}(\theta_j))\}$$

rewriting:

$$\overline{\beta}^k(\theta_i) = \min\{1, \theta_i + \ \gamma \max_{\theta_{-i}} \sum_{j \neq i} (\theta_j - \underline{\beta}_j^{k-1}(\theta_j))\}$$

we obtain

$$\overline{\beta}^k(\theta_i) = \min\left\{1, \theta_i + (\gamma(I-1))^k\right\},$$

and likewise

$$\underline{\beta}^k(\theta_i) = \max\left\{0, \theta_i - (\gamma(I-1))^k\right\}.$$

thus

$$\theta_i' \neq \theta_i \Rightarrow \theta_i' \notin \beta^k(\theta_i)$$

for sufficiently large k, provided that $\gamma < \frac{1}{I-1}$

- but now suppose that $\gamma > \frac{1}{I-1}$
- *each* type θ_i convinced (believes with probability one) that others' types are

$$\theta_j = \frac{1}{2} + \frac{1}{\gamma(I-1)}\left(\frac{1}{2} - \theta_i\right)$$

- now the expected value of the object for i is

$$\theta_i + \gamma(I-1)\left[\frac{1}{2} + \frac{1}{\gamma(I-1)}\left(\frac{1}{2} - \theta_i\right)\right] = \frac{1}{2}\left[1 + \gamma(I-1)\right]$$

- types cannot be distinguished (and hence separated) in direct or any other mechanism....

In the single unit auction example:

- robust implementation possible (using the modified generalized VCG mechanism) if $|\gamma| < \frac{1}{I-1}$
- robust implementation impossible (in *any* mechanism) if $\gamma > \frac{1}{I-1}$

In general environment....

- each Θ_i is a compact subset of the real line
- agent i's preferences depend on θ through $h_i : \Theta \rightarrow \mathbb{R}$
- preferences are single crossing in $h_i(\theta)$

Theorem (2009)

1. *Robust implementation is possible in the direct mechanism if strict EPIC and the "contraction property" hold.*
2. *Robust implementation is impossible in any mechanism if either strict EPIC or the "contraction property" fails.*

Contraction Property

- "deception": $\beta = (\beta_1, ..., \beta_I)$; $\beta_i : \Theta_i \to 2^{\Theta_i} / \varnothing$ with $\theta_i \in \beta_i(\theta_i)$
- "truth-telling": $\beta^* = (\beta_1^*, ..., \beta_I^*)$ with $\beta_i^*(\theta_i) = \theta_i$
- the aggregator functions h satisfy the strict contraction property if, $\forall \beta \neq \beta^*$, $\exists i, \theta_i' \in \beta_i(\theta_i)$ with $\theta_i' \neq \theta_i$, such that

$$\operatorname{sign}(\theta_i - \theta_i') = \operatorname{sign}\left(h_i(\theta_i, \theta_{-i}) - h_i(\theta_i', \theta_{-i}')\right),$$

  for all θ_{-i} and $\theta_{-i}' \in \beta_{-i}(\theta_{-i})$
- in the linear model this is equivalent to $|\gamma| \leq \frac{1}{I-1}$

Contraction Property 2

- with linear aggregator for each i:

$$h_i(\theta) = \theta_i + \sum_{j \neq i} \gamma_{ij} \theta_j$$

- the contraction property satisfied if and only if largest eigenvalue of the interaction matrix:

$$\Gamma \triangleq \begin{bmatrix} 0 & |\gamma_{12}| & \cdots & |\gamma_{1I}| \\ |\gamma_{21}| & 0 & & \vdots \\ \vdots & & \ddots & |\gamma_{I-1I}| \\ |\gamma_{I1}| & \cdots & |\gamma_{II-1}| & 0 \end{bmatrix}$$

is less than 1.

The Role of the Common Prior

- what if we know that the common prior assumption holds?
- in the analysis so far, no restrictions were placed on agents' beliefs and higher order beliefs
- consider the role of beliefs and hence intermediate notions of robustness
- remain in the linear valution model with linear best responses
- now not only the size but also the sign of the interdependence, γ, matters

- recall the linear best response in the auction model

$$\theta_i' = \theta_i + \gamma \sum_{j \neq i} (\theta_j - \theta_j')$$

- negative interdependence in agents' types, $\gamma < 0$, gives rise to strategic complementarities in the direct mechanism
- restricting attention to common prior type spaces makes no difference, and the contraction property continues to play the same role as described earlier
- Milgrom and Roberts (1991): with strategic complementarities, there are multiple equilibria if and only if there are multiple rationalizable actions)

- recall the linear best response in the auction model

$$\theta_i' = \theta_i + \gamma \sum_{j \neq i} \left(\theta_j - \theta_j'\right)$$

- positive interdependence in agents' types, $\gamma > 0$, gives rise to strategic substitutability in the direct mechanism, and robust implementation becomes easier
- in particular, it is often possible even if the contraction property failed: if

$$\frac{1}{I-1} < \gamma < 1,$$

  robust implementation is possible if we restrict attention to type spaces satisfying the common prior assumption

Future Questions

- Local, Intermediate Notions of Robustness
- Robust Predictions for Revenue Maximization Problem
- Single Crossing Conditions in Rich Type Spaces
- Beyond Mechanism Design:
  Robust Predictions In Games With Private Information
- If we cannot make unique predictions, can we provide robust bounds on the distribution of outcomes.

add interdependent preferences
add strength of interdependence as an argument
add argument for uniqueness are different
emphasize different values of γ
emphasize different proof techniques
have example of budget balancing as counterexample