Adaptive Incentive Design with Regret Minimization

24th European Control Conference, Reykjavík, Iceland

Georgios Vasileiou, Lantian Zhang, Prof. Silun Zhang

Motivation: Strategic Participation in the Control Loop

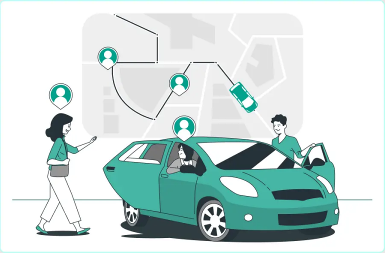
Cyber-physical systems interact with **strategic participants** $\rightarrow$ humans, firms, algorithms.

Infrastructure systems:

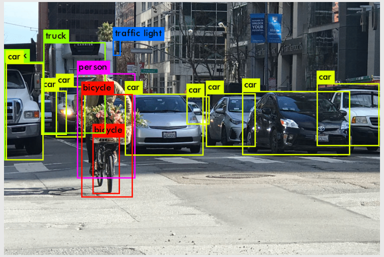
Congestion control

Power-grid operation

Economic markets:



Ride-sharing markets



Data crowdsourcing

images: Getty Images, Unsplash, Wikimedia Commons

Motivation: Strategic Participation in the Control Loop

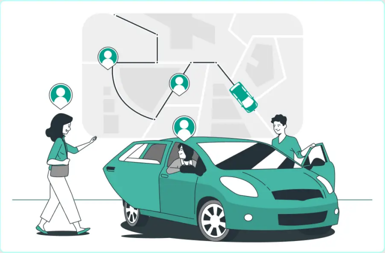
Cyber-physical systems interact with **strategic participants** $\rightarrow$ humans, firms, algorithms.

Infrastructure systems:

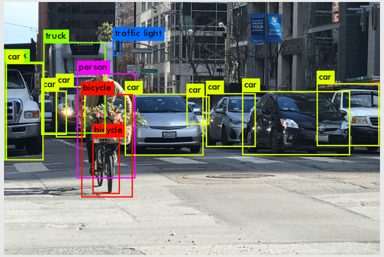
Congestion control

Power-grid operation

Economic markets:



Ride-sharing markets



Data crowdsourcing

Characteristics:

- Self-interested agents
- **Indirectly influenced** $\rightarrow$ incentives
- Heterogeneous, private preferences
- Interactions repeated over time

Game theory

Adaptive Control

Adaptive Incentive Design:
How do we choose incentives to steer agents to desirable outcomes?

Contents

Incentive Design Problems

System Planner
$\min_{x \in \mathcal{X}, u \in \mathbb{R}^n} \Psi(x, u)$

1 Announce
Contract γ

Agents
$\min_{x_i \in \mathcal{X}_i} \ (\ell_i(x) + \gamma_i(x))$

2 Competition
$x^* = \mathrm{NE}\ (\gamma)$
$u^* = \gamma(x^*)$

Reverse Stackelberg Game:
Choose $\gamma^* \in \mathcal{G}$: $(x^*, u^*) \in \arg\min_{(x,u)} \Psi(x, u)$

If ℓ_i are known,
Reverse Stackelberg game is one-shot solvable
(though optimizing over $\mathcal{G}$ is hard.)

Zheng, Basar, *Existence and Derivation of Optimal Affine Incentive Schemes for Stackelberg Games with Partial Information*, Int J Control, 1982

Adaptive Incentive Design: Closing the loop

$\hat{\Theta}_t$

System Planner

$\min_{x \in \mathcal{X}, u \in \mathbb{R}^n} \Psi(x, u)$

3 **Estimator**

$\hat{\Theta}_t = F_t(x_t^*, \gamma_t, \hat{\Theta}_{t-1})$

1 Announce

Contract γ_t

Agents

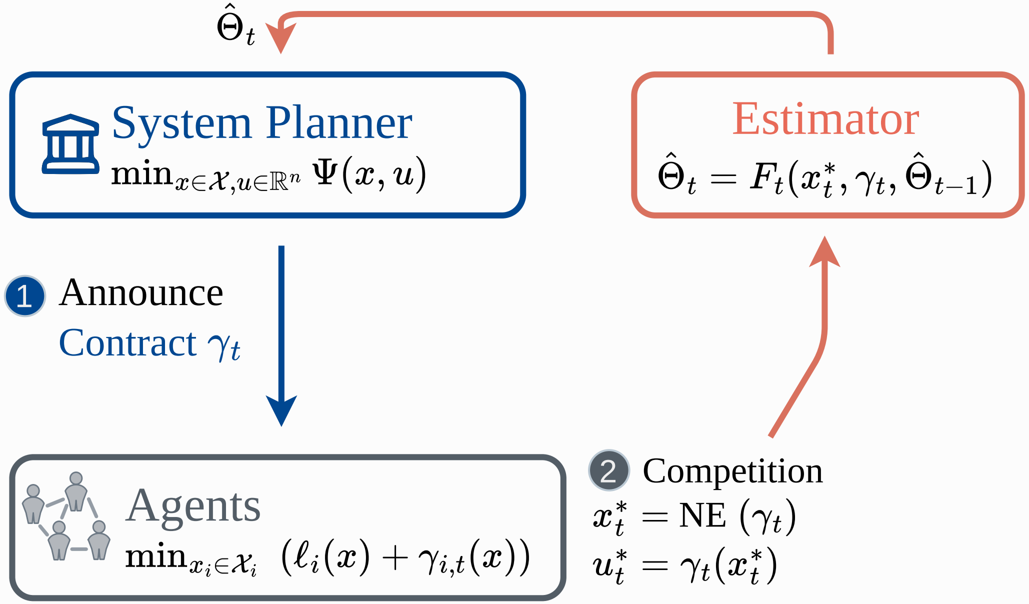
$\min_{x_i \in \mathcal{X}_i} \; (\ell_i(x) + \gamma_{i,t}(x))$

2 Competition

$x_t^* = \text{NE}\,(\gamma_t)$

$u_t^* = \gamma_t(x_t^*)$

Adaptive Incentive Design: Closing the loop



3

Design parameters:

1) **Estimator** $(\gamma_t, x_t^*) \mapsto \hat{\Theta}_t$
2) **Policy** $\hat{\Theta}_t \mapsto \gamma_{t+1}$

Adaptive Incentive Design:
Perform estimation + mechanism design over an unknown game, in feedback so that

$$\gamma_t \xrightarrow{t} \gamma^*$$

Connection to Literature

Relation to various perspectives

- **Stackelberg / Reverse Stackelberg Games.**
  - (Fudenberg et al., MIT Press, 1991) | Hierarchical leader-follower structure
  - (Ho et al., Automatica, 1982), (Basar et al., IEEE TAC, 1979) | Feedback, incentive controllabiltiy, credibility
  - (Jing et al., Springer, 1988), (Ho et al., IEEE TAC, 1981) | Information structure, multi-follower systems, and others...
- **Inverse Game Theory/ Inverse Variational Inequalities / Utility Learning.**
  - (Kuleshov et al., IEEE Conf WINE, 2015) | Consistent equilibria in succinct games
  - (Bertsimas et al., Mathematical Programming Ser. A, 2015) | Data-driven estimation of VI map
  - (Varian, Economic Journal, 2006) | Rationalizable behavior, axioms of revealed preference
- **Equilibrium Steering and Stackelberg Learning** (recent, relevant)
  - (Ratliff et al., IEEE TAC, 2021) | *Adaptive incentive design*: Parametric cost identification as least-squares
  - (Balcan et al., ACM Conf Econ Comp, 2015), (Lauffer et al., IEEE TAC, 2024), | No-regret policies on finite preference or action spaces.
  - (Yorulmaz et al., CDC, 2025) | Equilibrium steering in Bayesian normal-form games
  - (Zhang et al., ACM Conf Econ Comp, 2024), (Zhang et al., ICLR, 2026) | Identification and steering in extensive form games with no-regret learners

Contents

Agents' Game

Agents minimize $c_i^{\gamma}(x) = \ell_i(x_i, x_{-i}) + \gamma_i(x_i, p_i)$ subj. to $x_i \in \mathcal{X}_i$

Nominal costs

- **Convex game:** ℓ_i is C^2, convex and
- (Assumption) **Strongly Monotone.**
- Parameterization of agent costs

$$\frac{\partial \ell_i(x)}{\partial x_i} = \theta_i^{*\top} \Phi(x), \ \Phi(\cdot) \text{ monomial basis.}$$

Affine incentives

$$\gamma_i(x_i, p_i) = p_i(x_i - x_i^{\dagger}) + \text{constant},$$

$x_i^{\dagger}$: planner's target.

Lemma (GV, LZ, SZ):
Under strong monotonicity, affine incentives are **point-wise optimal** in

$$\mathcal{G} = \{\gamma : C^2, \text{ separable, convex in } x_i\}$$

G. Vasileiou, L. Zhang, S. Zhang, *A No-Regret Framework for Adaptive Incentive Design*, Preprint arXiv:2606.02529, 2026.

Planner's Performance

Compared to an oracle

- Objective $\min \Psi(x, u)$ s.t. $x \in \mathcal{X}$, $u \in \mathbb{R}^n$ $\quad \rightarrow \quad (x^\dagger, u^\dagger) \in \operatorname{int} \mathcal{X} \times \mathbb{R}^n$
- In T rounds of interaction, the **regret of policy** $\{p(t)\}_{t=1}^T$ is

$$\mathcal{R}_T := \sum_{t=1}^{T} \big(\Psi(x(t), u(t)) - \Psi(x^\dagger, u^\dagger) \big)^2.$$

RAID Problem: Design

$$\hat{\Theta}(t) = F_t\big(p(t), x(t), \hat{\Theta}(t-1)\big), \quad \text{(estimator)}$$

$$p(t+1) = G_t\big(\hat{\Theta}(t), p(t), x^\dagger, u^\dagger\big), \quad \text{(incentive policy)}$$

such that, almost surely,

1. $\lim_{t\to\infty} \hat{\Theta}(t) = \Theta^*$ **(strong consistency)** , and 2. $\mathcal{R}_t = o(t)$ **(sublinear regret)**.

Intermission: How do we learn from NE?

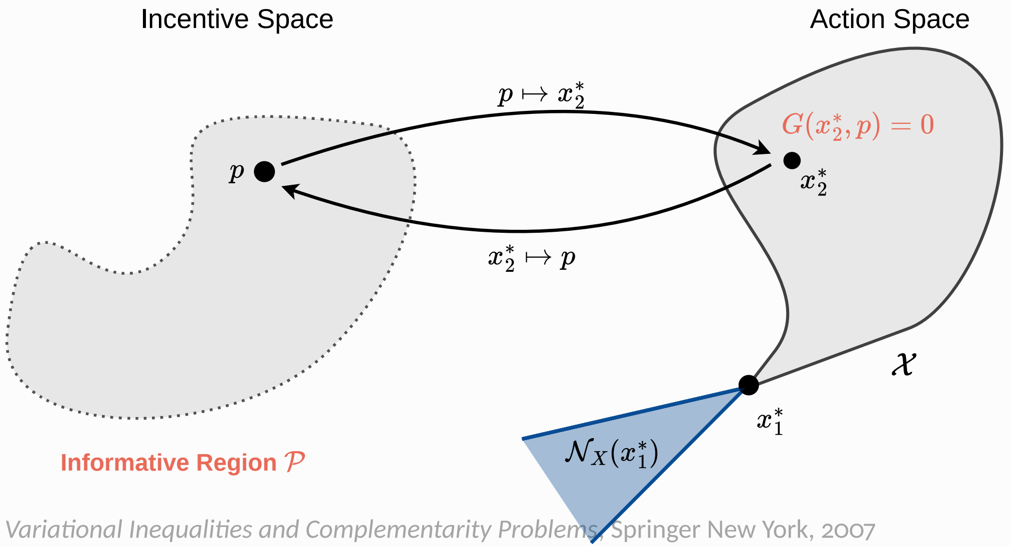
- **Definition (Incentivized Nash Equilibrium)**

Given incentive γ, strategy x^* is an INE if

$$c_i^\gamma(x_i^*, x_{-i}^*) \leq c_i^\gamma(x_i, x_{-i}^*),\ \forall x_i \in \mathcal{X}_i, \forall i \in [n]. \quad \Longleftrightarrow$$

- Solves the **variational inequality** (VI)

$$\langle G(x^*, p), x - x^* \rangle \geq 0, \forall x \in \mathcal{X},$$
$$G(x, p) := \Theta^* \Phi(x) + p.$$



Facchinei et al., *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer New York, 2007

Observation Model

$p(t)$: incentive issued
$\hat{p}(t)$: incentive observed

- **Unknowns:** Informative region $\mathcal{P}$, preferences Θ^*.
- **Available:** Action Space $\mathcal{X}$.
- **Noisy observations** $\{(\hat{p}(t), \hat{x}(t))\}_{t=1}^{T}$

$$\text{whenever } \hat{x}(t) \in \operatorname{int} \mathcal{X} : \quad \begin{cases} p(t) + v(t) & = -\Theta^* \Phi(\hat{x}(t)), \\ p(t) + e(t) & = \hat{p}(t), \end{cases}$$

where

$e(t)$ is a **measurement noise**

$v(t)$ is an **endogenous noise** $\rightarrow$ uncertainty on *best-response model*

Observation Model

$p(t)$: incentive issued
$\hat{p}(t)$: incentive observed

- **Unknowns:** Informative region $\mathcal{P}$, preferences Θ^*.
- **Available:** Action Space $\mathcal{X}$.
- **Noisy observations** $\{(\hat{p}(t), \hat{x}(t))\}_{t=1}^{T}$

$$\text{whenever } \hat{x}(t) \in \operatorname{int} \mathcal{X} : \quad \begin{cases} p(t) + v(t) & = -\Theta^* \Phi(\hat{x}(t)), \\ p(t) + e(t) & = \hat{p}(t), \end{cases}$$

where

$e(t)$ is a **measurement noise**

$v(t)$ is an **endogenous noise** $\rightarrow$ uncertainty on *best-response model*

Remark: $\{v(t)\}_t$ creates *correlation between the noise and the NE* $\rightarrow$ error-in-variables (EIV) regression problem. (Under review, 2026)

G. Vasileiou, L. Zhang, S. Zhang, *A No-Regret Framework for Adaptive Incentive Design*, Preprint arXiv:2606.02529, 2026

Contents

- Recursive least-squares (RLS), conditioned on *informative observations*

$$\hat{\Theta}(t) := \arg\min_{\Theta} \sum_{\tau \leq t} \delta(\tau) \left\| \hat{p}(\tau) + \Theta \Phi(\hat{x}(\tau)) \right\|^2, \text{ where } \delta(\tau) = \mathbf{1}\{\hat{x}(\tau) \in \operatorname{int} \mathcal{X}\}.$$

- Use *diminishing-excitation* sufficient condition

Lemma: Suppose strong monotonicity and $\{e(t)\}_t$ structure. Let $\lambda_{\min}(t) := \lambda_{\min}(\sum_\tau \delta(\tau) \Phi(\hat{x}(\tau)) \Phi(\hat{x}(\tau))^\top)$.
If $\lambda_{\min}(t) = \omega(\log t)$, then

$$\lim_{t \to \infty} \hat{\Theta}(t) = \Theta^* \text{ a.s.}$$

Lai et al., The Annals of Statistics, 1982 | L. Zhang et al., *Stochastic Adaptive Control for Systems with Nonlinear Parameterization*, Preprint arXiv:2604.06980, 2026

RLS Estimator

For the measurement-noise model

Lemma: Suppose strong monotonicity and $\{e(t)\}_t$ structure. Let $\lambda_{\min}(t) := \lambda_{\min}(\sum_\tau \delta(\tau)\Phi(\hat{x}(\tau))\Phi(\hat{x}(\tau))^\top)$. If $\lambda_{\min}(t)) = \omega(\log t)$, then

$$\lim_{t\to\infty} \hat{\Theta}(t) = \Theta^* \text{ a.s.}$$

Theorem: Under previous assumptions, if $\{p(t)\}_t$ is i.i.d. Gaussian, then $\lambda_{\min}(t) = \Omega(t)$ almost surely.

Faster than sufficient condition!

Key Technical Challenge $\to$ **Filtered regressors** $\mathbf{1}\{\hat{x}(t) \in \mathcal{X}\}\Phi(\hat{x}(t))$

- $\delta(t)$ filtering: Conditional expectation over *unknown* $\mathcal{P}$.
- Excitation must be preserved through the nonlinearity $p \mapsto \Phi \circ x^*(p)$.

The Incentive Policy

Use probing incentives whenever information is insufficient

Σ_t: RLS covariance matrix
$\lambda_{\min}(t)$: spectrum information matrix

- Switching policy:

$$p(t) = \begin{cases} \sim \mathcal{N}(0, \sigma_p^2 \mathbb{I}_n), & \text{if } \operatorname{tr}(\Sigma_t) > A(t)^{-1}, \\ -\hat{\Theta}(t-1)\Phi(x^\dagger), & \text{otherwise} \end{cases}$$

Exploration

Exploitation (Centrainty-equivalence design)

where

$$\Sigma_t := \left(\sum_{\tau \le t} \delta(\tau) \Phi(\hat{x}(\tau)) \Phi(\hat{x}(\tau))^\top \right)^{-1},$$

and $A(t) := t^\gamma \log(t+1)$, $\gamma \in [1/2, 1)$.

The Incentive Policy

Use probing incentives whenever information is insufficient

Σ_t: RLS covariance matrix
$\lambda_{\min}(t)$: spectrum information matrix

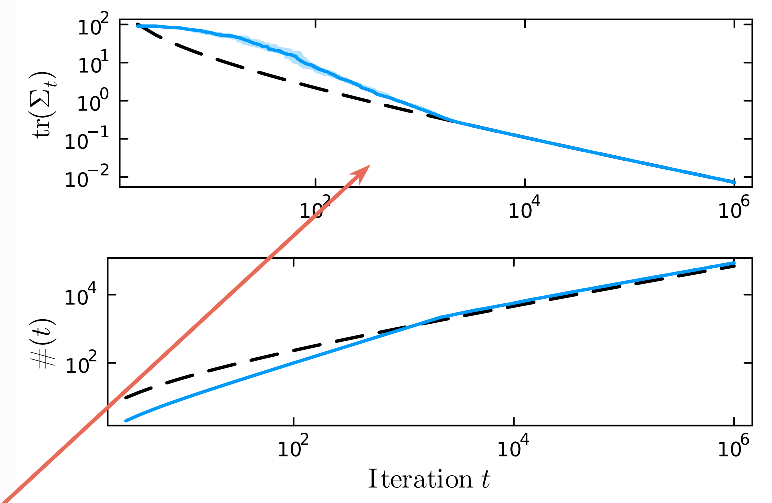
- Switching policy:

$$p(t) = \begin{cases} \sim \mathcal{N}(0, \sigma_p^2 \mathbb{I}_n), & \text{if } \operatorname{tr}(\Sigma_t) > A(t)^{-1}, \\ -\hat{\Theta}(t-1)\Phi(x^\dagger), & \text{otherwise} \end{cases}$$

where

$$\Sigma_t := \left(\sum_{\tau \le t} \delta(\tau)\Phi(\hat{x}(\tau))\Phi(\hat{x}(\tau))^\top \right)^{-1},$$

and $A(t) := t^\gamma \log(t+1)$, $\gamma \in [1/2, 1)$.



Switching threshold enables sufficient exploration
$\lambda_{\min}(t) = \Omega(t^\gamma \log t)$.

Convergence Results

The RAID Algorithm Converges with given rates

Theorem: Suppose strong monotonicity and $\{e(t)\}_t$ structure. For all parameters $\gamma \in [1/2, 1)$, the estimates $\{\hat{\Theta}(t)\}_t$ generated satisfy

$$\|\hat{\Theta}(t) - \Theta^*\|_F^2 = O\left(t^{-\gamma}\right) \text{ a.s.}$$

Moreover, the planner's regret over the generated incentive sequence $\{p(t)\}_t$ satisfies

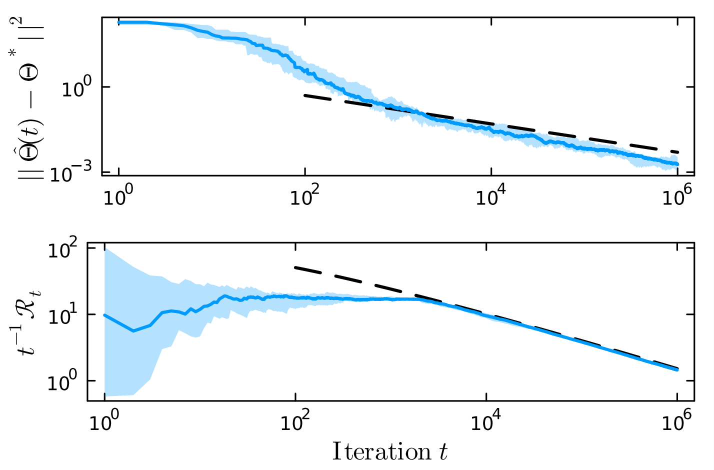
$$\mathcal{R}_t = O\left(t^{\gamma} \log t\right) \text{ a.s.}$$

G. Vasileiou, L. Zhang, S. Zhang, *A No-Regret Framework for Adaptive Incentive Design*, Preprint arXiv:2606.02529, 2026

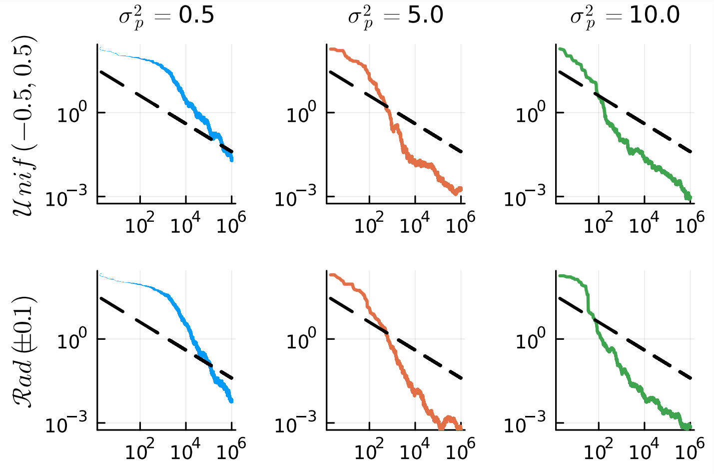
Numerical Verification

- Identification on 3-player game with 3rd-order polynomial costs

$$\ell_1(x) = 3(\tfrac{1}{6}x_1^3 + x_1^2 + x_1x_2 - x_1x_3 + x_1), \quad \ell_2(x) = 3(\tfrac{1}{6}x_2^3 + x_2^2 - x_1x_2 + x_2x_3 - x_2), \quad \ell_3(x) = 3(\tfrac{1}{6}x_3^3 + x_3^2 + x_1x_3 - x_2x_3 + x_3)$$



Convergence rates of the estimator and regret (γ = 1/2, black line indicates predicted rate).



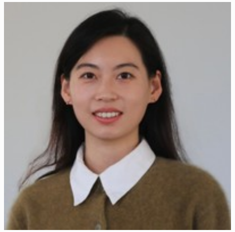
Performance under different measurement noise and probing variance.

Contents

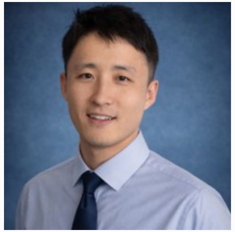
Takeaways

1. *Adaptive Incentive Design* is a closed-loop regulator for strategic agents under information asymmetry: **regulate Nash equilibrium behavior** + **generate data for learning**.
2. This work proposes the **RAID Algorithm**:
    1. Achieves $O(t^{-\gamma})$ parameter estimation error and $O(t^{\gamma} \log t)$ regret.
    2. Active probing guarantees excitation.
    3. Sparsity of probing incentives produces strong consistency and sublinear regret.

Acknowledgements



Lantian Zhang
KTH Royal Institute of Technology
`lantian@kth.se`



Silun Zhang
Assistant Professor
KTH Royal Institute of Technology
`silunz@kth.se`

Thank you to my collaborators, and the KTH Optimization and Systems Theory group for many useful discussions.

Further details

Contact me

WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

KTH
VETENSKAP
OCH KONST

References I

[1] Y.-P. Zheng and T. Basar, "Existence and Derivation of Optimal Affine Incentive Schemes for Stackelberg Games with Partial Information: A Geometric Approach," *International Journal of Control*, vol. 35, no. 6, pp. 997–1011, 1982.

[2] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge Massachusetts: MIT press, 1991, ISBN: 978-0-262-06141-4.

[3] Y.-C. Ho, P. B. Luh, and G. J. Olsder, "A Control-Theoretic View on Incentives," *Automatica*, vol. 18, no. 2, pp. 167–179, 1982.

[4] T. Basar and H. Selbuz, "Closed-loop Stackelberg strategies with applications in the optimal control of multilevel systems," *IEEE TAC*, vol. 24, no. 2, pp. 166–179, Apr. 1979.

[5] Y.-W. Jing and S.-Y. Zhang, "The solution to a kind of stackelberg game systems with multi-follower: Coordinative and incentive," in *Analysis and Optimization of Systems*. Berlin, Heidelberg: Springer-Verlag, 1988, pp. 593–602, ISBN: 3540192379.

[6] Yu-Chi Ho, P. Luh, and R. Muralidharan, "Information structure, Stackelberg games, and incentive controllability," *IEEE TAC*, vol. 26, no. 2, pp. 454–460, Apr. 1981.

[7] V. Kuleshov and O. Schrijvers, "Inverse game theory: Learning utilities in succinct games," in *International Conference on Web and Internet Economics*, Springer, 2015, pp. 413–427.

[8] D. Bertsimas, V. Gupta, and I. C. Paschalidis, "Data-driven estimation in equilibrium using inverse optimization," *Mathematical Programming*, vol. 153, no. 2, pp. 595–633, Nov. 2015, ISSN: 0025-5610, 1436-4646.

[9] H. R. Varian, "Revealed Preference," in *Samuelsonian Economics and the Twenty-First Century*, Oxford University Press, Aug. 2006, pp. 99–115.

[10] L. J. Ratliff and T. Fiez, "Adaptive Incentive Design," *IEEE TAC*, vol. 66, no. 8, pp. 3871–3878, 2021.

References II

[11] M.-F. Balcan, A. Blum, N. Haghtalab, and A. D. Procaccia, "Commitment Without Regrets: Online Learning in Stackelberg Security Games," in *Proc. of the 16th ACM Conf Econ Comp*, Portland Oregon USA, Jun. 2015, pp. 61–78, ISBN: 978-1-4503-3410-5.

[12] N. Lauffer, M. Ghasemi, A. Hashemi, Y. Savas, and U. Topcu, "No-Regret Learning in Dynamic Stackelberg Games," *IEEE Transactions on Automatic Control*, vol. 69, no. 3, pp. 1418–1431, Mar. 2024, ISSN: 0018-9286, 1558-2523, 2334-3303.

[13] A. E. Yorulmaz, R. K. Velicheti, M. Bastopcu, and T. Başar, "A soft inducement framework for incentive-aided steering of no-regret players," in *Proc. of the 64th Conference on Decision and Control (CDC)*, Rio de Janeiro, Brazil, 2025, pp. 4396–4401.

[14] B. H. Zhang, G. Farina, I. Anagnostides, *et al.*, "Steering no-regret learners to a desired equilibrium," in *Proc. 25th ACM Conf Econ Comp*, ser. EC '24, New Haven, CT, USA: Association for Computing Machinery, 2024, pp. 73–74, ISBN: 9798400707049.

[15] B. H. Zhang, T. Lin, Y. Chen, and T. Sandholm, "Learning a game by paying the agents," arXiv:2503.01976, 2026.

[16] F. Facchinei and J.-S. Pang, Eds., *Finite-Dimensional Variational Inequalities and Complementarity Problems* (Springer Series in Operations Research and Financial Engineering). Springer New York, 2004.

[17] T. L. Lai and C. Z. Wei, "Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Dynamic Systems," *The Annals of Statistics*, vol. 10, no. 1, pp. 154–166, 1982.

[18] L. Zhang, B. Wahlberg, and S. Zhang, "Stochastic adaptive control for systems with nonlinear parameterization: Almost sure stability and tracking," arXiv:2604.06980, 2026.