Meta-Cognition. An Inverse-Inverse Reinforcement Learning Approach for Cognitive Radars

Kunal Pattanayak (Cornell University),
Vikram Krishnamurthy (Cornell University),
Christopher M. Berry (Lockheed Martin).

(Research funded by **Lockheed Martin** and the **Army Research Office**)

Main Idea. Detecting utility maximization $\equiv$ Checking linear feasibility

How to make checking linear feasibility difficult?

Cognitive radar $\rightarrow$ Choose optimal waveform for target tracking

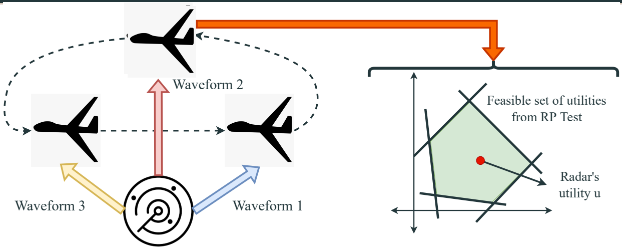
Adversarial Target $\rightarrow$ Malicious maneuvers to 'estimate' radar's utility

How to spoof adversarial attacks on radar's utility function?

Ans. **Cognition Masking**

Intelligently perturbed radar actions successfully hide radar's utility

Background. Cognitive Radar and Revealed Preference



Cognitive Radar (Utility u)

Adversarial Target

Cognitive Radar [1–3]: Optimal waveform adaptation.
For target maneuvers (**probe**) $\{\alpha_k\}_{k=1}^{K}$, radar chooses waveforms (**response**) $\{\beta_k\}_{k=1}^{K}$ that maximize utility u:

$$\beta_k = \operatorname{argmax}_{\beta \in \mathbb{R}_+^m} u(\beta), \ \ \alpha_k'\beta \leq 1 \tag{1}$$

Radar Bayesian tracker: Linear Gaussian dynamics
(i) α_k: state noise covariance
(ii) β_k: observation noise covariance
(iii) $\alpha_k'\beta_k \leq 1$ (1): Bound on radar SNR $\equiv$ Bound on radar's asymptotic predicted Kalman **precision** [3]
'Choose best waveform subject to resource constraints'

Utility Estimation via Revealed Preference (RP):
RP Test [4, 5] : For dataset $\mathbb{D} = \{\alpha_k, \beta_k\}_{k=1}^{K}$, linear feasibility test is **equivalent** to checking for utility maximization (1):

$$\mathrm{RP}(u, \mathbb{D}) \leq 0, \ \ u = \{u_k, \lambda_k\} \in \mathbb{R}_+^{2m}, \tag{2}$$

$$u_{\text{est}}(\beta) = \min_k \{u_k + \lambda_k \alpha_k'(\beta - \beta_k)\} \tag{3}$$

What if $\mathbb{D}$ is noisy?
RP Test (2) generalizes to statistical hypothesis test to detect feasibility [6] (discussed in slide 4).

Cognition Masking

How to mitigate adversarial RP test and ensure poor reconstruction of radar's utility function

Result 1. Deterministic Inverse RP for Masking Cognition

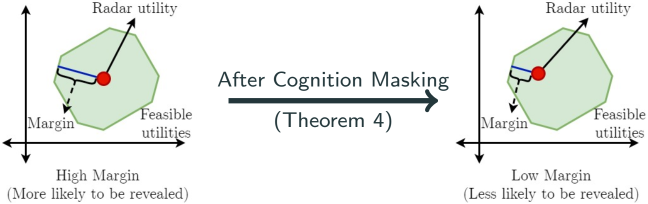
Assumption: "Radar and adversary have accurate probe-response measurements."

Adversarial target $\overset{\text{IRL}}{\rightarrow}$ RP Feasibility test (2) (Set-valued estimate of radar's utility)

How to rank utility functions in the feasible set?

Rank via **Margin** of RP test - **max. perturbation to fail RP test** (based on [7])

$$\boxed{\text{Margin}_{\mathbb{D}}(u) = \max_{\epsilon \geq 0} \epsilon, \ \ \text{RP}(u, \mathbb{D}) + \epsilon \geq 0, \ \ u \in \text{Feasible set}}$$



- Margin: Closeness to edge of feasible set (infeasibility of RP test)
- Center of feasible set: **max. margin**, edge of feasible set: **zero margin**
- $\uparrow$ Margin $\iff$ $\uparrow$ Goodness-of-fit to RP test
- **Deterministic** Cognition masking: Deliberately perturb radar's response to push radar's utility **towards** edge of feasible set from RP test

Deterministic Inverse IRL for Masking Cognition (Theorem 4)

Suppose radar faces adversarial constraints $\{\alpha_k'\beta \leq 1\}_{k=1}^{K}$. The radar's *deterministic* I-IRL algorithm to hide its utility u is:

Step 1. Choose margin $\epsilon_{\text{thresh}} \in \mathbb{R}_+$
Step 2. Compute naive response β_k^* (1)
Step 3. Compute optimal perturbation $\{\delta_k^*\}$ for I-IRL:

$$\{\delta_k^*\} = \underset{\{\delta_k\}\in\mathbb{R}^m}{\text{argmin}} \sum_{k=1}^{K} \|\delta_k\|_2^2, \quad \underbrace{\text{Margin}_{\{\alpha_k,\beta_k^*+\delta_k\}}(u) \leq \epsilon_{\text{thresh}}}_{\text{(Mitigating adversarial RP Test)}} \tag{4}$$

Step 4. Transmit engineered sub-optimal responses $\{\beta_k^* + \delta_k^*\}$.

Summary

Deterministic I-IRL: Small margin ϵ_{thresh}

$\iff$ Closer to failing RP test (2)

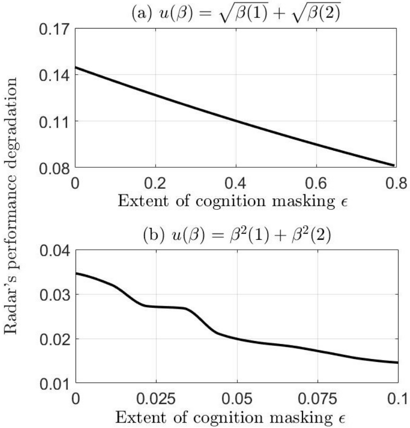
$\iff$ Larger deviation from radar's optimal strategy

- Margin Constraint in (4) is non-convex (bilinear).

Current research: *Formulate convex relaxations of bilinear constraints in* (4).

Numerical Results: Deterministic Inverse IRL

- Simulation-based datasets to illustrate I-IRL for 2 utility functions
- Parameters: Time horizon $K = 50$, Probe/Response dimension $m = 2$



Key Insights:

- **Small deviation** from *optimal strategy* masks utility by a large extent.
- Radar's performance degradation $\uparrow$ with ϵ.

Result 2. Stochastic Inverse RP for Masking Cognition

Assumption: "Adversary has *noisy* measurements of the radar's response."

$$\text{(Adversary side)}: \hat{\beta}_k = \beta_k + w_k, \;\; w_k \sim f_w \;\; (f_w \text{ known to radar}) \tag{5}$$

Adversarial target $\overset{\text{IRL}}{\rightarrow}$ Feasibility *Detector* (see also [3] for details)

H_0 : RP Test (2) has a feasible solution for $\{\alpha_k, \beta_k\}$

H_1 : RP Test (2) has NO feasible solution for $\{\alpha_k, \beta_k\}$

IRL Feasibility Detector : $\boxed{\phi^*(\widehat{\mathbb{D}}) \lessgtr_{H_0}^{H_1} F_L^{-1}(1-\eta)}$ $(\widehat{\mathbb{D}} = \{\alpha_k, \hat{\beta}_k\})$, (6)

$\phi^*(\widehat{\mathbb{D}})$: $\max_{\{\bar{u}>0, \bar{u}(\beta_1)\}} \text{Margin}_{\bar{u}}(\widehat{\mathbb{D}})$, r.v. $L := \max_{j,k} \alpha_j'(w_j - w_k)$,

η : Adversary chosen bound for $\mathbb{P}(H_1|H_0)$

"*Radar is non-cognitive if margin is under a threshold*"

- Radar **can no more** manipulate margin of RP test.
- Can *at best* manipulate $\mathbb{P}(H_1|\{\alpha_k, \beta_k\}, u)$ (Cond. Type-I error prob.)
- **Stochastic** Cognition masking: Deliberately perturb radar's response to mitigate IRL detector (**increase** conditional Type-I error probability).

Stochastic Inverse IRL for Masking Cognition (Theorem 5)

Adversary's sensor is noisy; everything else the same as deterministic case. Radar's *stochastic* I-IRL algorithm is:

Step 1. Choose sensitivity parameter $\lambda > 0$

Step 2. Compute naive response β_k^* (1)

Step 3. Compute optimal perturbation $\{\delta_k^*\}$ for I-IRL:

$$\{\delta_k^*\} = \underset{\{\delta_k\}\in\mathbb{R}^m}{\operatorname{argmin}} \sum_{k=1}^{K} (u(\beta_k^*) - u(\beta_k^* + \delta_k)) - \lambda \underbrace{\mathbb{P}(H_1 | \{\alpha_k, \beta_k^* + \delta_k\}, u)}_{\text{(Mitigating adversarial IRL detector)}} \quad (7)$$

Step 4. Transmit engineered sub-optimal responses $\{\beta_k^* + \delta_k^*\}$

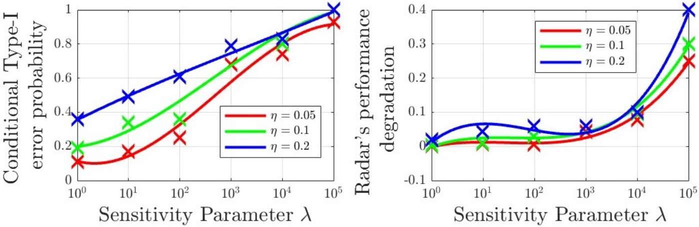
(7): Ensuring low margin of RP Test with high probability

Summary

- **Stochastic I-IRL**: Trade-off between $\uparrow$ *QoS* and $\uparrow$ *adversarial obfuscation*.
- Radar can only <u>estimate</u> $\mathbb{P}(H_1|H_0, u)$ (7) via Monte-Carlo methods.
- Stochastic approximation based algorithms like **SPSA [8]** can be used for implementing optimization problem (7).
- SPSA $\rightarrow$ Fewer (only 2) computations/update wrt finite diff. methods.

Numerical Results: Stochastic Inverse IRL

- Simulations for a single utility function $u(\beta) = \sqrt{\beta_1} + \sqrt{\beta_2}$
- Parameters: Time horizon $K = 50$, Probe/Response dimension $m = 2$



Key Insights:

- Small *performance loss* sufficiently confuses IRL detector (large cond. Type-I error).
- **Both** adversarial confusion and radar's performance degradation $\uparrow$ with λ.
- Interestingly, performance degradation $\downarrow$ with η (error bound).

Conclusion and Extensions

Summary:

- Radar **counter**-countermeasure to mitigate an adversarial countermeasure
- Cognition Masking: *Deliberately perturb optimal radar waveforms to sufficiently reduce margin of RP test and 'hide' radar's utility.*
- Sub-optimality in response trades-off between Privacy and Performance
- Methodology inspired from adversarial obfuscation [9] in deep learning and differential privacy [10]

Applications of Inverse IRL:

Online Ad Design. Deliberately tweak meta-data to incentivize user clicks

Survey Design. Deliberate abnormality in questions to incentivize truthfulness

Extensions (Current research):

1. Finite sample results for spoofing the adversary's IRL detector
2. Convex relaxations of the I-IRL objective function
3. **Counter**-(counter-)^{n}measure: What if adversary knows radar's spoofing strategy? *Game theoretic approach?*

Thank You!

Miscellaneous

FAQs

- **How justified is the constrained utility maximization abstraction for radar operation?**

Quite prevalent in literature:

(i) Multi-UAV network [11]: Utility $\rightarrow$ Fairness and downlink data rate, Constraint $\rightarrow$ Transmission power, Cramer-Rao bound on localization accuracy
(ii) Q-RAM (Resource Allocation) [12]: Utility $\rightarrow$ QoS for tracking and search, Constraint $\rightarrow$ Bandwidth, Short-term and Long-term constraints
(iii) Radar Tracking with ECM [13]: Utility $\rightarrow$ Neg. of weighted mean of radar energy and dwell time, Constraint $\rightarrow$ 4% Cap on lost tracks due to ECM

FAQs

- **Is conditional Type-I probability the only I-IRL metric for adversarial obfuscation in stochastic I-IRL?**

No fixed formula, does need more work. Some intuitive alternatives: (a) Use deterministic I-IRL as is. Formulate concentration inequalities for margin of the noisy dataset.

(b) Manipulate the average margin instead of margin. BUT, might be underplaying robustness of IRL detector.

(c) [**Speculative**] Use a neural network to learn IRL method on the fly and disrupt.

Remark: I-IRL hinges delicately on IRL methodology.

Other heuristic ideas to hide utility?

FAQs

- **What's next after IRL, and inverse IRL? I2-IRL?**

Game-theoretic formulation.

Key challenge: Formulate a utility function in terms of both adversary probes and radar response.

Anticipated outcome: Inverse game theory - Detecting play from the Nash equilibrium of a game between adversary and radar.

References

[1] Simon Haykin et al. "Cognitive tracking radar". In: *2010 IEEE Radar Conference*. IEEE. 2010, pp. 1467–1470.

[2] Kristine L Bell et al. "Cognitive radar framework for target detection and tracking". In: *IEEE Journal of Selected Topics in Signal Processing* 9.8 (2015), pp. 1427–1439.

[3] Vikram Krishnamurthy et al. "Identifying cognitive radars-inverse reinforcement learning using revealed preferences". In: *IEEE Transactions on Signal Processing* 68 (2020), pp. 4529–4542.

[4] S. N. Afriat. "The construction of utility functions from expenditure data". In: *International economic review* 8.1 (1967), pp. 67–77.

[5] H. R. Varian. "Revealed preference and its applications". In: *The Economic Journal* 122.560 (2012), pp. 332–338.

[6] Vikram Krishnamurthy and William Hoiles. "Afriat's test for detecting malicious agents". In: *IEEE Signal Processing Letters* 19.12 (2012), pp. 801–804.

[7] Hal R Varian et al. *Goodness-of-fit for revealed preference tests*. Citeseer, 1991.

[8] James C Spall. "An overview of the simultaneous perturbation method for efficient optimization". In: *Johns Hopkins apl technical digest* 19.4 (1998), pp. 482–492.

[9] Varun Chandrasekaran et al. "Face-off: Adversarial face obfuscation". In: *arXiv preprint arXiv:2003.08861* (2020).

[10] Rathindra Sarathy and Krishnamurty Muralidhar. "Evaluating Laplace noise addition to satisfy differential privacy for numeric data.". In: *Trans. Data Priv.* 4.1 (2011), pp. 1–17.

[11] Xinyi Wang et al. "Constrained utility maximization in dual-functional radar-communication multi-UAV networks". In: *IEEE Transactions on Communications* 69.4 (2020), pp. 2660–2672.

[12] Jeffery Hansen et al. "Resource management for radar tracking". In: *2006 IEEE Conference on Radar*. IEEE. 2006, 8–pp.

[13] WD Blair et al. "Benchmark for radar allocation and tracking in ECM". In: *IEEE Transactions on Aerospace and Electronic Systems* 34.4 (1998), pp. 1097–1114.