Learning in Proportional Allocation Auctions

Younes Ben Mazziane (University of Avignon)



AVIGNON UNIVERSITÉ

Outline

1. Proportional allocation (Kelly) auctions
2. Kelly game
3. Research Gap and Contributions
4. Best-response dynamics
5. No-regret learning
6. Numerical results

Proportional allocation (Kelly) auctions

1. Users submit bids $z_i \geq 0$
2. Resource owner allocates $x_i(\mathbf{z}) = \dfrac{w_i(z_i)}{\sum_j w_j(z_j)}$

Revenue: $\sum_i z_i$

Agent Gains: $V_i(x_i(\mathbf{z}))$

Social Welfare (SW): $\sum_i V_i(x_i(\mathbf{z}))$

Kelly Game

Players set: $\mathcal{I}$

Actions set: $\prod_{i \in \mathcal{I}} [\epsilon, c_i]$

Utilities $\phi_i(z_i, z_{-i}) = V_i(x_i(z_i, z_{-i})) - z_i$

Kelly Game

Players set: $\mathcal{I}$

Actions set: $\prod_{i \in \mathcal{I}} [\epsilon, c_i]$

Utilities $\phi_i(z_i, z_{-i}) = V_i(x_i(z_i, z_{-i})) - z_i$

Definition (Nash equilibrium): $\mathbf{z}^*$ is NE iff $\forall \mathbf{z}, \forall i,\ \phi_i(z_i^*, z_{-i}^*) \geq \phi_i(z_i, z_{-i}^*)$

Kelly Game

Players set: $\mathcal{I}$

Actions set: $\prod_{i \in \mathcal{I}} [\epsilon, c_i]$

Utilities $\phi_i(z_i, z_{-i}) = V_i(x_i(z_i, z_{-i})) - z_i$

Definition (Nash equilibrium): $\mathbf{z}^*$ is NE iff $\forall \mathbf{z}, \forall i,\ \phi_i(z_i^*, z_{-i}^*) \geq \phi_i(z_i, z_{-i}^*)$

Nice properties of Kelly:

→ Good social welfare at NE: 3/4 (no budgets) 1/2 (with budgets)

Research Gap and Contributions

Can we reach NE and good social welfare in dynamic settings?

Repeated Kelly game (Synchronous):

$$z_i^{\mathcal{A}}(t) = f(\mathbf{z}(1), \ldots, \mathbf{z}(t-1))$$

$$max \sum_t \phi_i(z_i^{\mathcal{A}}(t), z_{\_i}(t))$$

Research Gap and Contributions

Can we reach NE and good social welfare in dynamic settings?

Repeated Kelly game (Synchronous):

$$z_i^{\mathcal{A}}(t) = f(\mathbf{z}(1), \ldots, \mathbf{z}(t-1))$$

$$max \sum_t \phi_i(z_i^{\mathcal{A}}(t), z_{-i}(t))$$

Previous results: For linear V_i →

1. Convergence of Fictitious play to NE
2. Convergence, in utility, to NE under no-regret

Research Gap and Contributions

Can we reach NE and good social welfare in dynamic settings?

Repeated Kelly game (Synchronous):

$$z_i^{\mathcal{A}}(t) = f(\mathbf{z}(1), \ldots, \mathbf{z}(t-1))$$

$$max \sum_t \phi_i(z_i^{\mathcal{A}}(t), z_{-i}(t))$$

Previous results: For linear V_i

1. Convergence of Fictitious play to NE
2. Convergence, in utility, to NE under no-regret

Our results (informal): For a family of $V_i's$ including $\ln$

1. Motivation for $V_i \propto \ln$
2. Uniqueness NE under general conditions
3. Converge to NE under OGD, DA, and BR

Best-response dynamics

Best response operator $\mathrm{BR}_i(s) = \arg \max_{z_i \in [\epsilon, c_i]} \varphi_i(z_i, s)$

BR updates: $z_i^{\mathrm{BR}}(t) = \mathrm{BR}(z\_{i}(t-1))$

Best-response dynamics

Best response operator $\mathrm{BR}_i(s) = \arg \max_{z_i \in [\epsilon, c_i]} \varphi_i(z_i, s)$

BR updates: $z_i^{\mathrm{BR}}(t) = \mathrm{BR}(z_{-i}(t-1))$

Asynchronous BR updates
in potential games with finite action sets

Convergence to NE

Best-response dynamics

Best response operator $\mathrm{BR}_i(s) = \arg \max_{z_i \in [\epsilon, c_i]} \varphi_i(z_i, s)$

BR updates: $z_i^{\mathrm{BR}}(t) = \mathrm{BR}(z_{-i}(t-1))$

Asynchronous BR updates
in potential games with finite action sets

→ Convergence to NE

Synchronous BR are fixed point iterations

→ Contraction via Jacobian matrix: $\sup_{\mathbf{z}} ||J_{BR}(\mathbf{z})|| < 1$

No-regret learning

At every iteration $t \in \{1,2,\ldots,T\}$

1. The agent selects a decision $z(t) \in \mathcal{X} \subset \mathbb{R}^d$
2. Nature selects a reward function $u_t : \mathcal{X} \mapsto \mathbb{R}$, $u_t \in \mathcal{F}$
3. Agent observes u_t and gains $u_t(x_t)$

No-regret learning

At every iteration $t \in \{1,2,\ldots,T\}$

1. The agent selects a decision $z(t) \in \mathcal{X} \subset \mathbb{R}^d$
2. Nature selects a reward function $u_t : \mathcal{X} \mapsto \mathbb{R}$, $u_t \in \mathcal{F}$
3. Agent observes u_t and gains $u_t(x_t)$

Objective: $\mathrm{Reg}_T(\mathscr{A}_i) = o(T)$

$$\mathrm{Reg}_T(\mathscr{A}) \triangleq \max_{z \in \mathcal{X}} \sum_{t=1}^{T} u_t(z) - \sum_{t=1}^{T} u_t\left(z^{\mathscr{A}}(t)\right).$$

No-regret learning

At every iteration $t \in \{1,2,\ldots,T\}$

1. The agent selects a decision $z(t) \in \mathcal{X} \subset \mathbb{R}^d$
2. Nature selects a reward function $u_t : \mathcal{X} \mapsto \mathbb{R}$, $u_t \in \mathcal{F}$
3. Agent observes u_t and gains $u_t(x_t)$

Objective: $\mathrm{Reg}_T(\mathcal{A}_i) = o(T)$

$$\mathrm{Reg}_T(\mathcal{A}) \triangleq \max_{z \in \mathcal{X}} \sum_{t=1}^{T} u_t(z) - \sum_{t=1}^{T} u_t\left(z^{\mathcal{A}}(t)\right).$$

In a game : $u_t(z) = \phi_i(z, z_{-i}(t))$

No-regret learning

Define gradient at each step: $g_t^{(i),\mathscr{A}_i} \triangleq \partial_i u_i^t(z_i^{\mathscr{A}_i}(t)),$

Online Gradient Descent (OGD):

$$z_i^{OGD}(t+1) = \Pi_{[\epsilon, c_i]}\left(z_i^{OGD}(t) + \eta_{t+1}^{(i)} g_t^{(i),OGD}\right),$$

No-regret learning

Define gradient at each step: $g_t^{(i),\mathcal{A}_i} \triangleq \partial_i u_i^t(z_i^{\mathcal{A}_i}(t)),$

Online Gradient Descent (OGD):

$$z_i^{OGD}(t+1) = \Pi_{[\epsilon,c_i]}\left(z_i^{OGD}(t) + \eta_{t+1}^{(i)} g_t^{(i),OGD}\right),$$

Dual Averaging with quadratic regularizer (DAQ):

$$z_i^{DAQ}(t+1) = \Pi_{[\epsilon,c_i]}\left(\eta_{t+1}^{(i)} g_{1:t}^{(i),DAQ}\right)$$

If $V_i = a_i \ln$ → $\eta_t^{(i)} = \mathcal{O}\left(\frac{c_i \epsilon}{a_i \sqrt{T}}\right)$

No-regret learning

- Relevant results in the literature

$$\forall x,\ H(x) + H^T(x) \prec 0$$

$$H(\mathbf{x}) \triangleq \left(\frac{\partial^2 U_i(\mathbf{x})}{\partial x_i \partial x_j} \right)_{i,j}$$

Convergence of OGD and DA to NE of the game with utilities U_i

No-regret learning

- Relevant results in the literature

$$\forall x,\ H(x) + H^T(x) \prec 0$$

$$H(\mathbf{x}) \triangleq \left(\frac{\partial^2 U_i(\mathbf{x})}{\partial x_i \partial x_j}\right)_{i,j}$$

→ Convergence of OGD and DA to NE of the game with utilities U_i

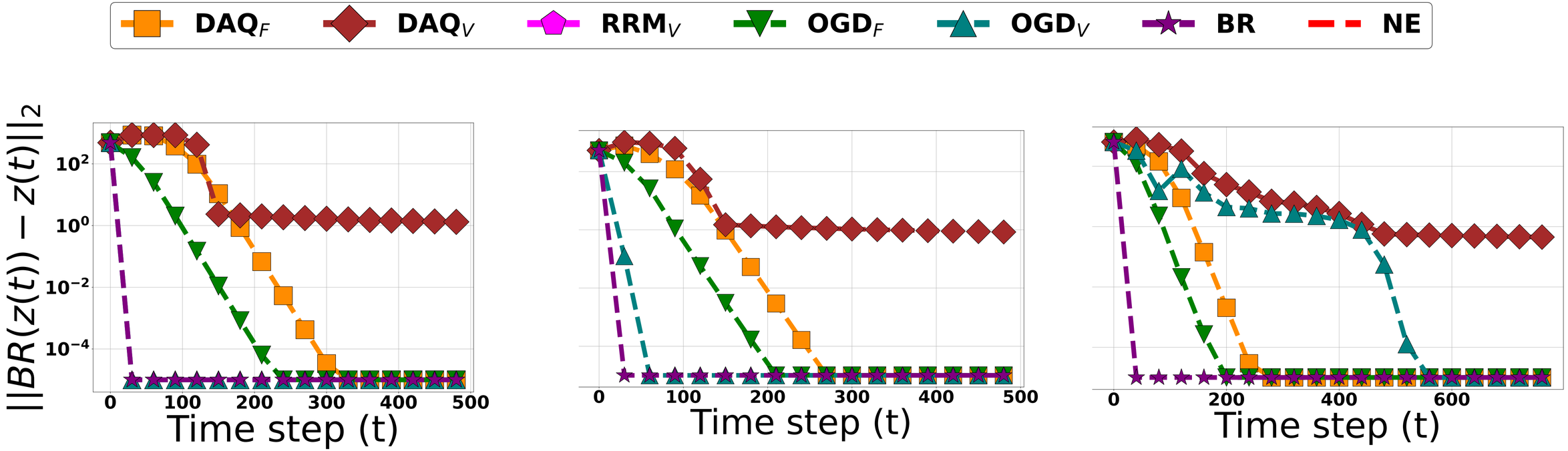
---

- $U_i(x_i, \cdot\,)$ is convex
- $\exists \boldsymbol{\lambda} > 0 : \sum_i \lambda_i U_i(\mathbf{x})$ is concave

→ Convergence of the average action to NE

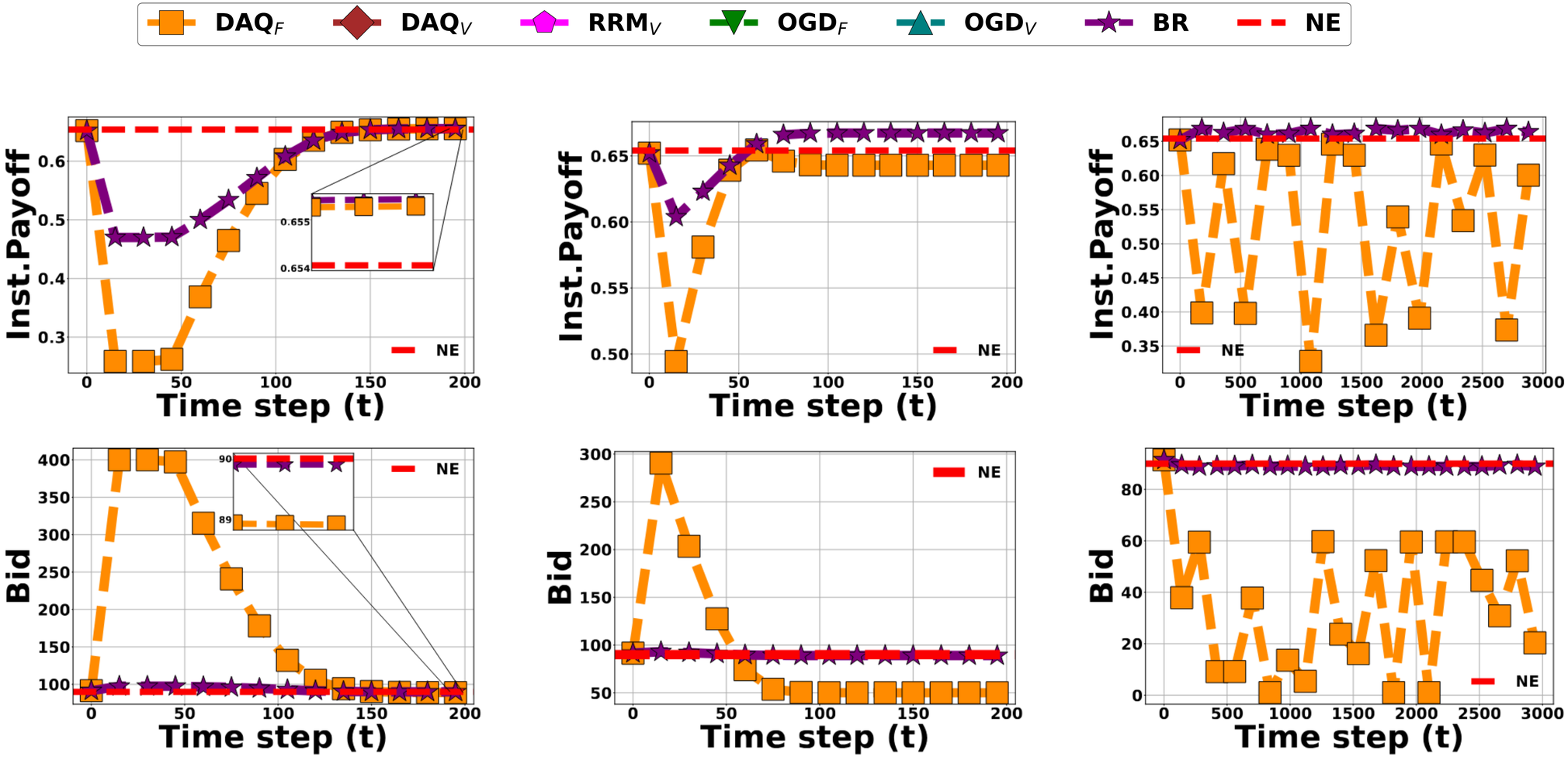
Numerical Results (Homogeneous updates)

Experimental setup: $V_i(x) = (100 - \gamma i)\ln(x), \forall i \in [10], \epsilon = 1, c_i = 400$



$\gamma = 0$ $\gamma = 5$ $\gamma = 10$

Numerical Results (Heterogeneous updates)

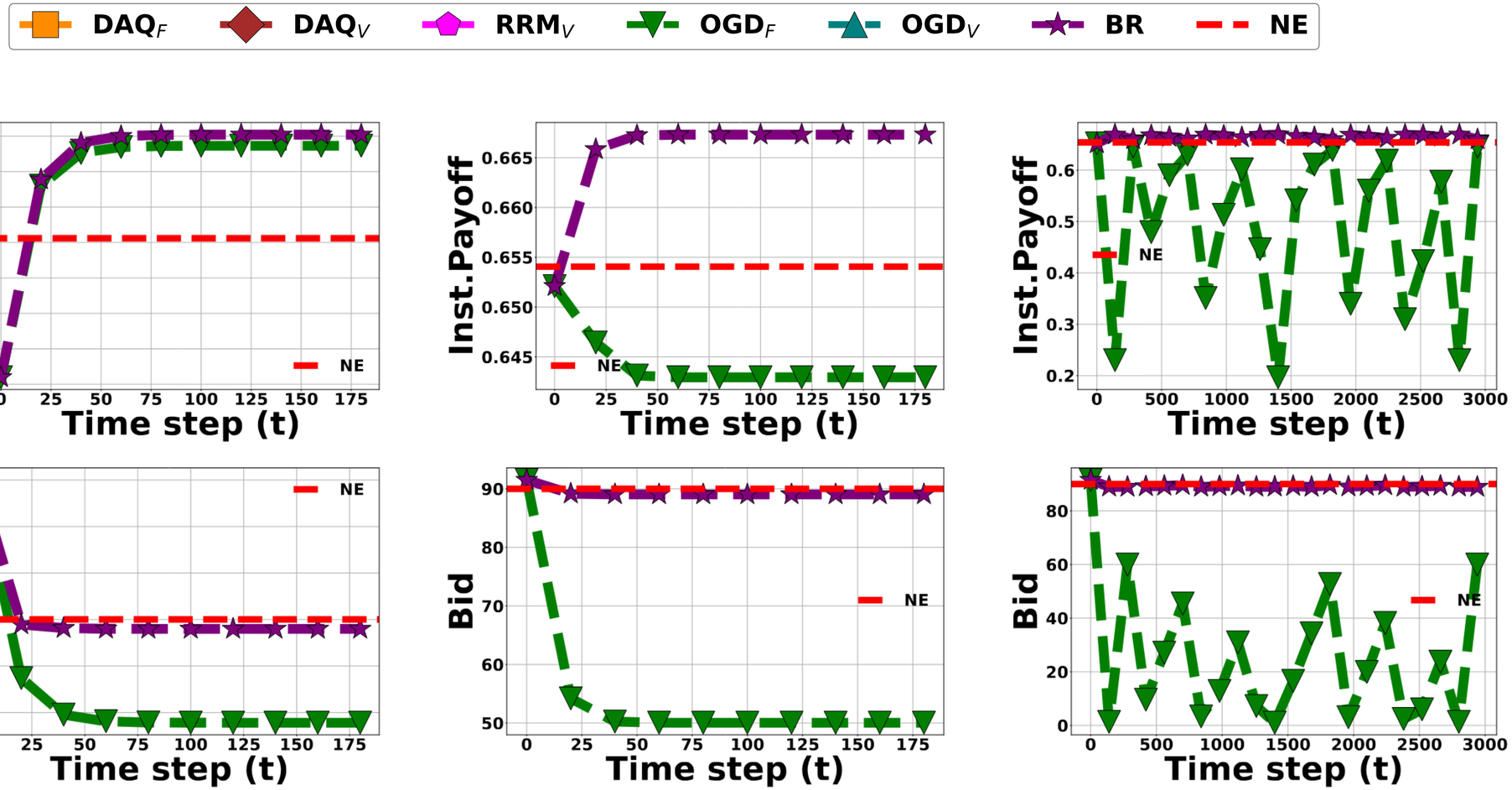


(a) BR vs. $\mathrm{DAQ_F}$ ($\alpha_{\mathrm{BR}} = 10\%$)

(b) BR vs. $\mathrm{DAQ_F}$ ($\alpha_{\mathrm{BR}} = 80\%$)

(c) BR vs. $\mathrm{DAQ_F}$ ($\alpha_{\mathrm{BR}} = 90\%$)

Numerical Results (Heterogeneous updates)



(d) BR vs. OGD$_F$ ($\alpha_{BR} = 10\%$)

(e) BR vs. OGD$_F$ ($\alpha_{BR} = 80\%$)

(f) BR vs. OGD$_F$ ($\alpha_{BR} = 90\%$)

Future work

- Extension to multiple resources
- Fixed budget over all rounds: Online Knapsack problem?
- Heterogeneous updates, e.g., Bimatrix games

Thank you.