Department of Industr
and Systems Engineeri
UNIVERSITY OF WISCONSIN–MADISO

Problem Statem

Optimal team formation is crucial in domains like work
preferences for working with others, which influence te
are:

- **Unknown Preferences:** Individual preferences are nc

rial
ing
N

Optimizing

K

ent

places and sports. Each individual has
eam performance. The main challenges

ot directly observable.

Overview: Using
by updating beli
are estimates of
shown in Figure

Team Assignments by P

asper Veje Jakobsen, kjakobsen@wisd.edu

[1]Department of Industrial and Systems Engineering

Learning Preferences: Kalman Filter

g a Kalman filter, the preferences between individuals are dynam
efs based on feedback received after team interactions, using $\hat{\sigma}$
the process noice and feedback noise. An example of the lea
1.

Preferences

u[1]

Simulatio

nically estimated
$_w$ and $\hat{\sigma}_f$ which
rning process is

Setup: The simulation evaluates team a
periods, and $s = 3$ teams. Preference
($\sigma_f = 0.1$). A **Random Assignment** strat
weights β are compared. Results are ave
of performances shown in Figure 3 and

on Setup and Results

ssignment strategies using $n = 10$ individuals, $m = 100$
s evolve with noise ($\sigma_w = 0.1$), and feedback is noisy
egy and **Optimal Assignments** with varying exploration
eraged over 10 simulations, with the mean and variance
Figure 4.

- **Dynamic Evolution:** Preferences evolve continuousl
- **Complexity:** Large action space (all possible team as (preferences and uncertainties).

Objective: Develop a framework to assign individuals to

- **Maximizing** overall **team performance**.
- Balancing **exploration** (learning preferences) and **exp**

Mathematical Not

Individuals: $I = \{1, 2, \ldots, n\}$: Set of n individuals.

Team Assignments: At each period p, individuals are grou

$$\mathcal{T}^p = \{T_1^p, T_2^p, \ldots, T_{s_p}^p\}, \quad T_j^p \cap T_k^p =$$

Preferences: The true (unobserved) preference w_{ij}^p evol

y towards all individuals.

signments) and continuous state space

teams dynamically while:

loitation (using known preferences).

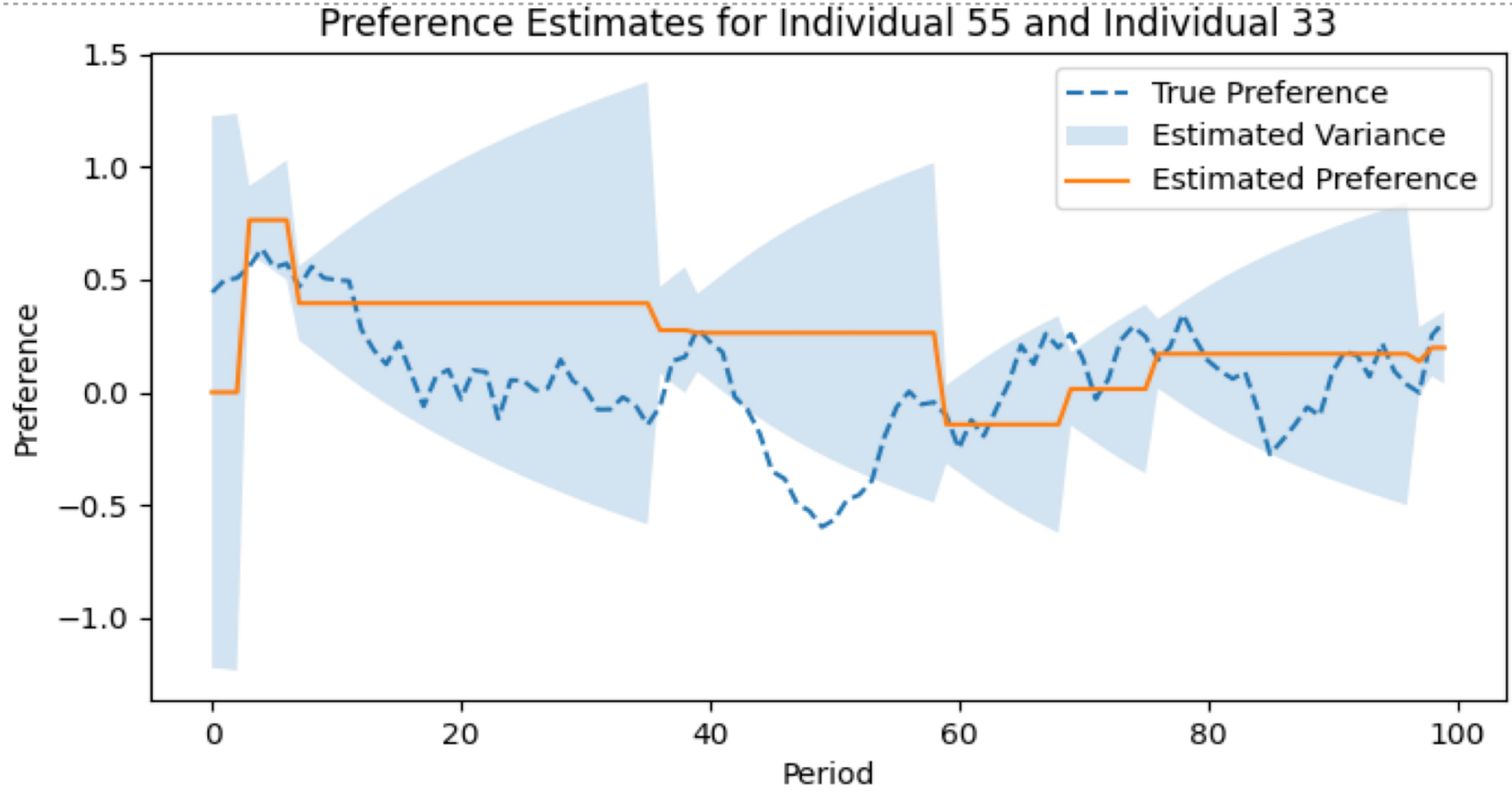
ation

uped into mutually exclusive teams:

$$= \emptyset, \quad \cup_{j=1}^{s_p} T_j^p = I.$$

ves continuously:

Figure 1.

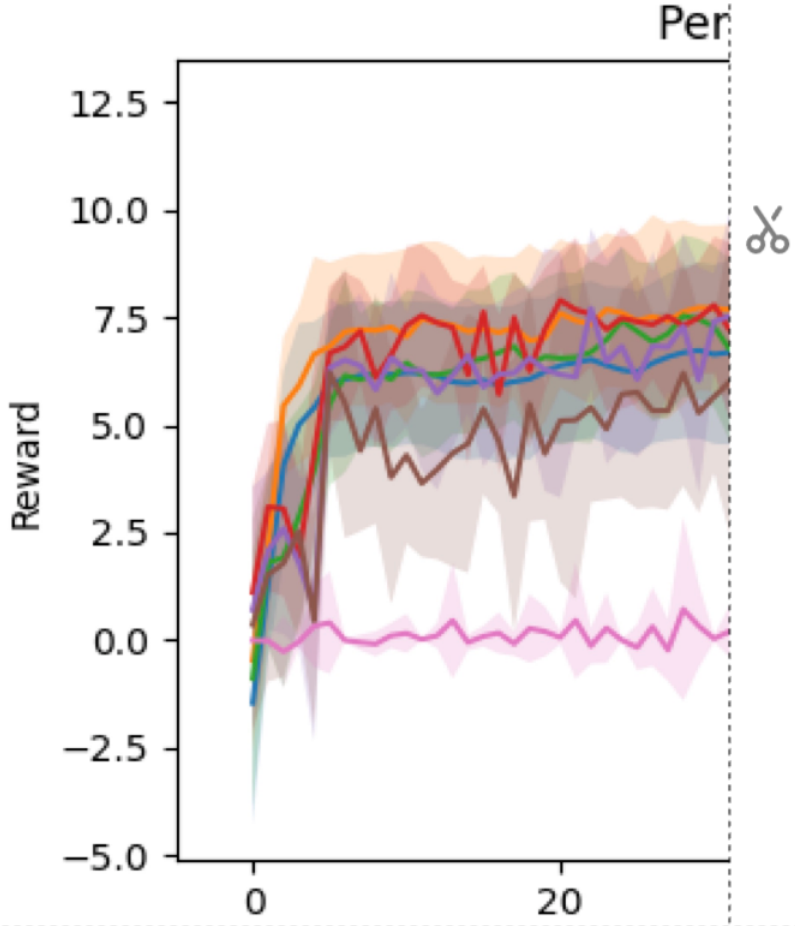


Estimated preferences μ_{ij} compared to true preferences w_{ij}. σ_{ij} reflects unc

Optimal Assignments

of performances shown in Figure 3 and

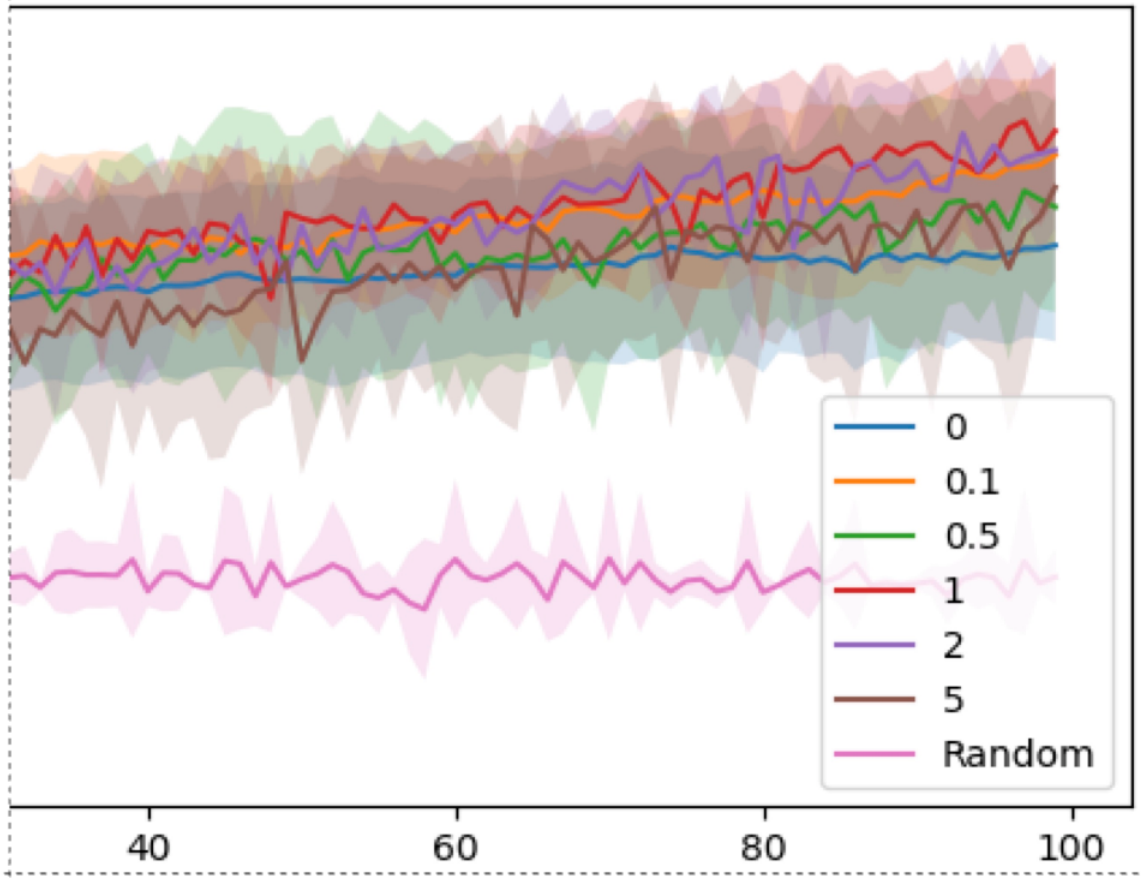
Metrics: The reward measures team pe
assignments. The preference distance e
to the true preferences (w_{ij}) using the L



ertainty.

rformance over periods, reflecting the effectiveness of
valuates the convergence of estimated preferences (μ_{ij})
$_2$-norm, indicating the learning accuracy of the model.

formance of Different Actors



$$w_{ij}^0 = N(0, \sigma_p^2), \quad w_{ij}^{p+1} = w_{ij}^p + \varepsilon_{ij}^p$$

Feedback: Sparse feedback F_{ij}^p is collected for individua

$$F_{ij}^p = w_{ij}^p + \eta_{ij}^p, \quad \eta_{ij}^p \sim$$

DP Formulatio

State Space: The state at each period represents the b probability distribution:

$$\hat{s}_p = \Big\{ P(w_{ij}^p \mid \text{history up to } p$$

This belief can be fully characterized by the current estim ties, where μ_{ij}^p is the estimated preference and σ_{ij}^p is the

$$s_p = \{\mu_{ij}^p, \sigma_{ij}^p \mid i, j \in$$

$$\varepsilon_{ij}^{p} \sim N(0, \sigma_w^2)$$

ls in the same team:

$$N(0, \sigma_f^2)$$

n

---

elief about preferences, expressed as a

$$) \mid i, j \in I \Big\} .$$

ates of preferences and their uncertain-
uncertainty:

$$I\},$$

Objective: Assig
(learning prefere

Upper Confiden
into a score S_{ij}^{p}

Teams are assigr

MILP: Used as a
at each period.

Future Directior
efficiently handl

n individuals to teams to maximize overall performance by balan
nces) and **exploitation** (using known preferences).

ce Bound (UCB) Strategy: Combines mean preference μ_{ij}^p and
using an exploration weight $\beta \geq 0$:

$$S_{ij}^p = \mu_{ij}^p + \beta \cdot \sigma_{ij}^p.$$

ned to maximize the sum of scores:

$$a_p = \arg\max_{a \in \mathcal{A}} \sum_{T \in a} \sum_{i,j \in T, i \neq j} S_{ij}^p.$$

baseline, MILP directly optimizes team assignments by maximizin

ns: Explore scalable methods like multi-agent reinforcement lea
e large action spaces and dynamic team assignment scenarios.

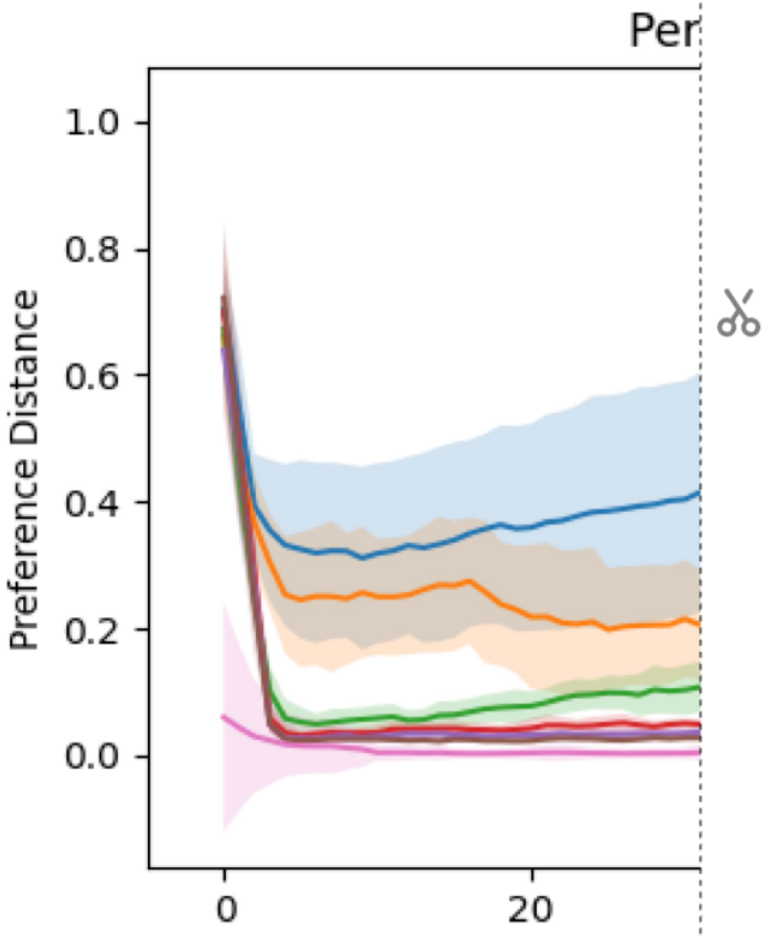
cing **exploration**

uncertainty σ_{ij}^{p}.

g the score (S_{ij}^{p})

rning (MARL) to

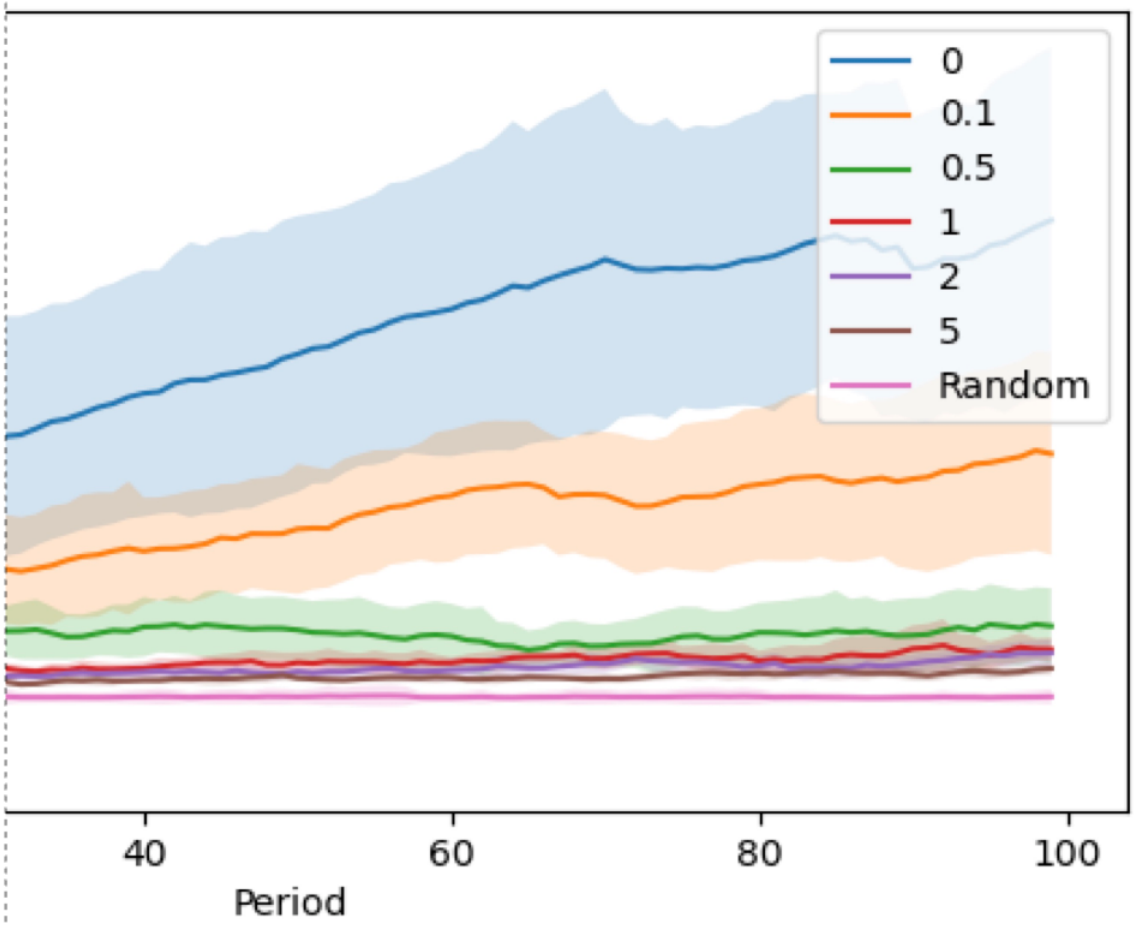
Figure 3. Performance comparison across



Period

strategies. Higher rewards indicate better team assignments.

formance of Different Actors



Action Space: The action $a_p \in \mathcal{A}$ is a valid team assignm

Transition Function: Beliefs about the preferences evolv

$$s_{p+1} = f(s_p, F^p), \quad \text{where } F^p \text{ incorporates feed}$$

Reward Function: The reward reflects team performance

$$r(s_p, a_p) = \sum_{T \in a_p} \left(\sum_{i,j \in T, i=} \right.$$

Objective: Find the policy π^* that maximizes the expecte

$$\pi^* = \arg\max_{\pi} \mathbb{E} \left[\sum_{p=1}^{m} r(s_p \right.$$

ent for the period.

e based on feedback:

dback on evolving preferences.

e, measured by the beliefs:

$$\left(\mu_{ij}^{p} \right)$$

$\neq j$

ed cumulative reward:

$$\left. \right] .$$

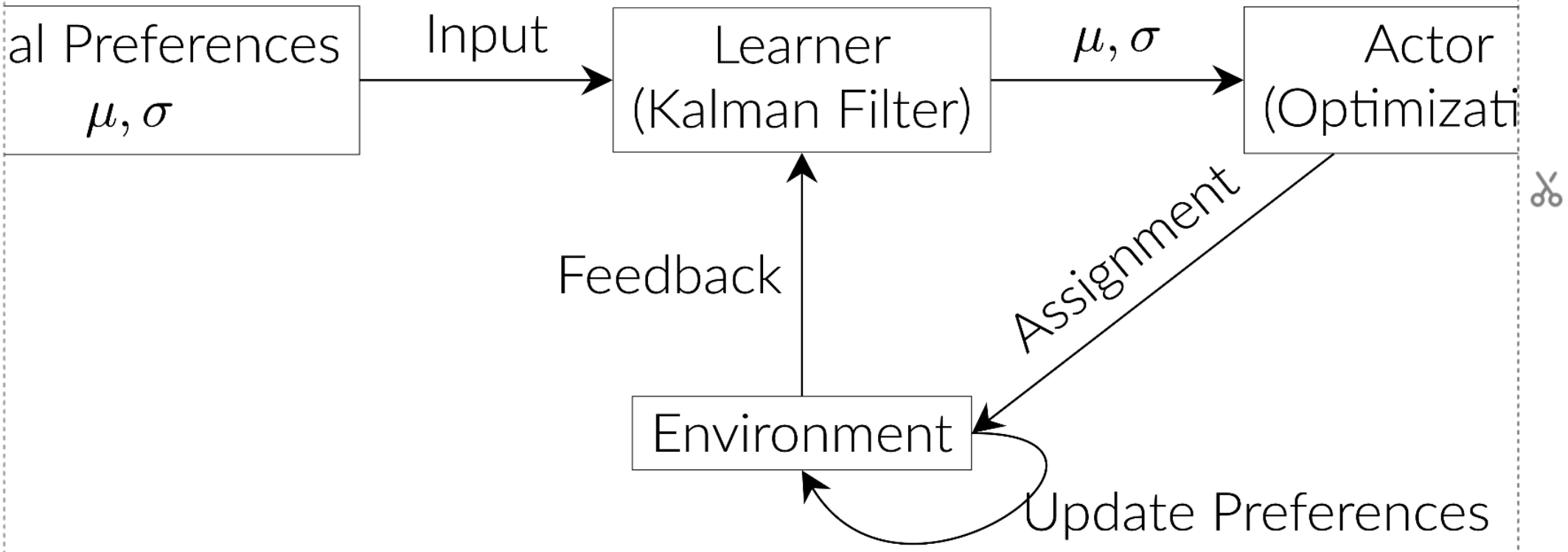
$\flat, a_p)$

The integrated f
signments, and t

Initi

Integrated Framework

ramework, which uses a *learner* for preference estimation, an *ac*
he *environment* for updates and feedback is shown in Figure 2.



tor for team as-

on)

Figure 4. Preference distance between estimate
accuracy.

Discuss

The results demonstrate that incorporati
mean preferences) significantly enhance
The Kalman filter effectively learns indivi
as the number of interactions increases.

The optimization-based approach balan
plexity limits its scalability to larger setti
reinforcement learning or heuristic-base

This framework provides a foundation fo
in corporate, academic, and sports envir

d and true preferences. Lower values show improved learning

sion and Conclusion

ng both **exploration** (via uncertainty) and **exploitation** (via
s team performance compared to random assignments.
idual preferences over time, with convergence improving

ces learning and performance, but computational com-
ngs. Future work will explore **scalable methods** such as
d optimization to handle larger action spaces.

or dynamic team assignment, with potential applications
onments where collaboration dynamics evolve over time.

Department of Industrial and Systems Engineering

Figure 2. Inte

gration of Learner, Actor, and Feedback Loop with Preference Update in the

University of Wisconsin-Madison

Environment.

ISYE 723: Dynamic Programming