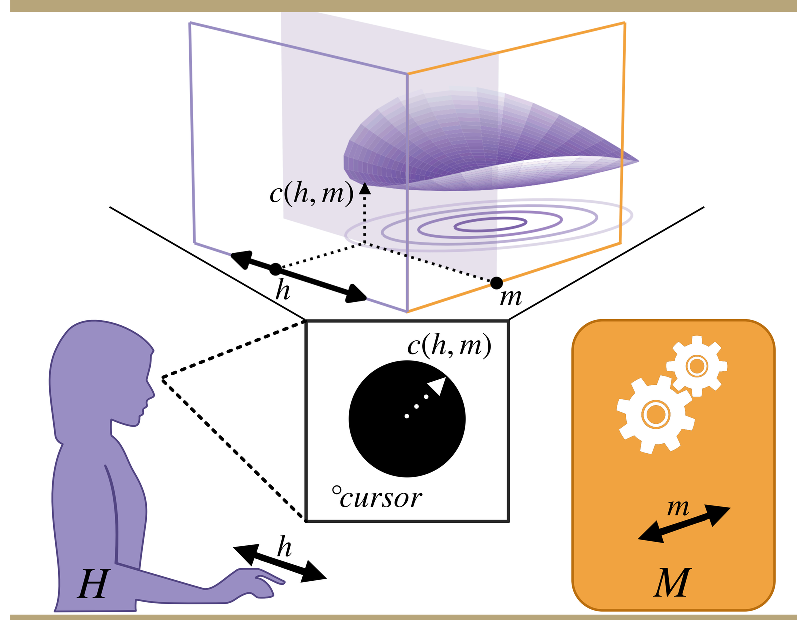
An Algorithm That Attains the Human Optimum in a Repeated Human-Machine Interaction Game

STUDENTS: Jason T. Isa

Introduction



- When humans interact with learning-based control systems, a common goal is to minimize a cost function known only to the human (i.e. exoskeletons and human metabolic cost of transport).
- Conventional approaches impose assumptions or solve an inverse problem to infer human's cost.
- Our algorithm achieves the minimum solely through observations of the human's actions over repeated interactions without any direct knowledge of the cost function.

Machine's Learning Algorithm

Algorithm 1 1×1 Experiments

Require: Initialize $L, \Delta, \alpha, \widehat{h}^*[0], \widehat{m}^*[0]$

for $k = 0, 1, \ldots, K-1$ **do**

  $(h', m') \leftarrow \texttt{trial}\left(L, \widehat{h}^*[k], \widehat{m}^*[k]\right)$

  $(h'', m'') \leftarrow \texttt{trial}\left(L + \Delta, \widehat{h}^*[k], \widehat{m}^*[k]\right)$

  $\widehat{h}^*[k+1] \leftarrow h'$

  $\widehat{m}^*[k+1] \leftarrow \widehat{m}^*[k] + \alpha\,(m'' - \widehat{m}^*[k])$

end

Function $\texttt{trial}\,(\widetilde{L}, \widetilde{h}^*, \widetilde{m}^*)$**:**

  for $t = 0, 1, \ldots, T-1$ **do**

    $h[t] \leftarrow \texttt{get\_manual\_input}\,(t)$

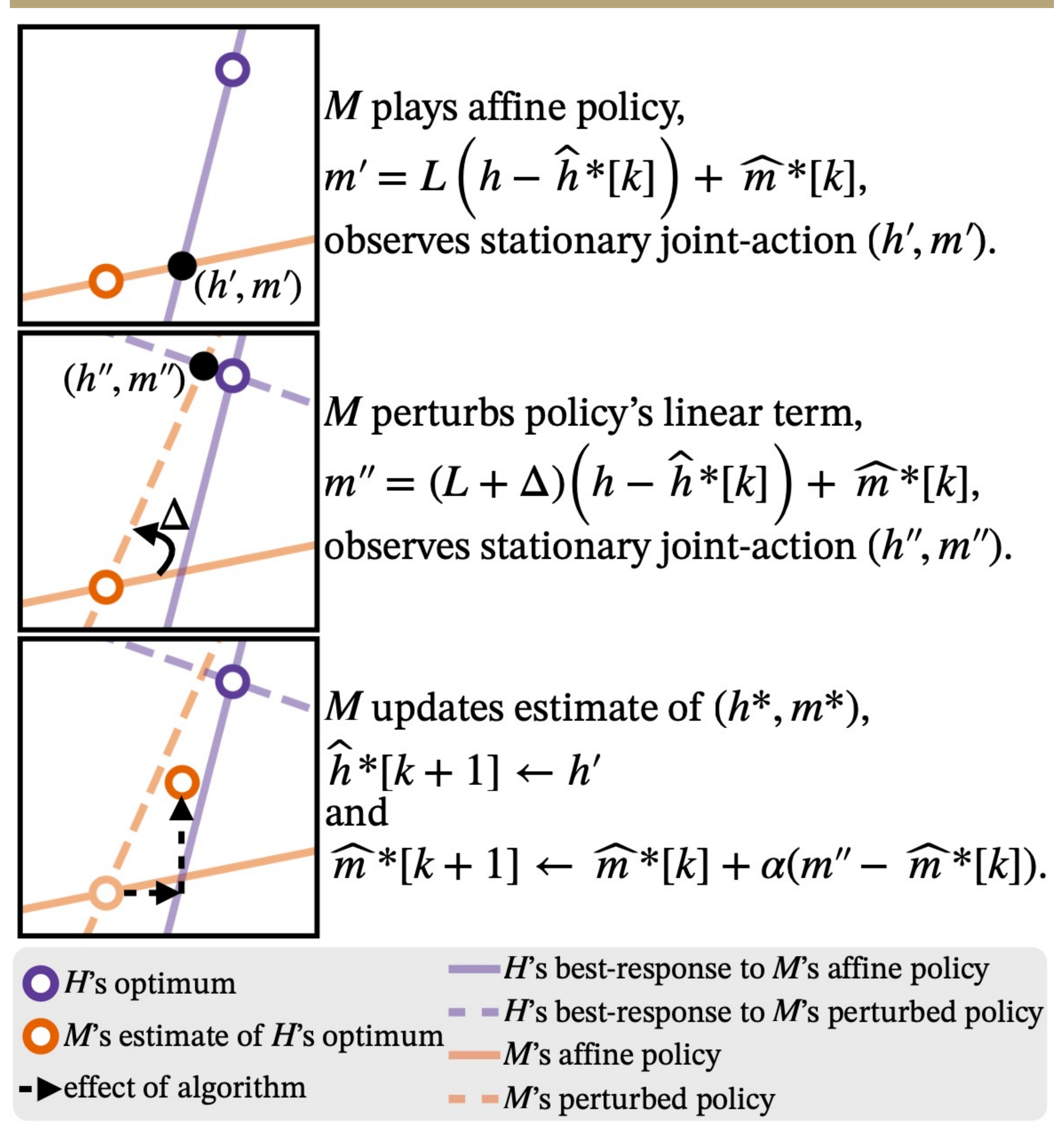
    $m[t] \leftarrow \widetilde{L}\left(h[t] - \widetilde{h}^*\right) + \widetilde{m}^*$

    $\texttt{display\_cost}\,(c\,(h[t], m[t]))$

  end

return mean of last t iterations of h and m

Visualization of Algorithm



Discrete-Time Linear System

We studied quadratic cost, $c(h,m) = \frac{1}{2}h^T h + \frac{1}{2}m^T m$, where (h^*, m^*) denote the minimum of c and the human's best response takes the form,

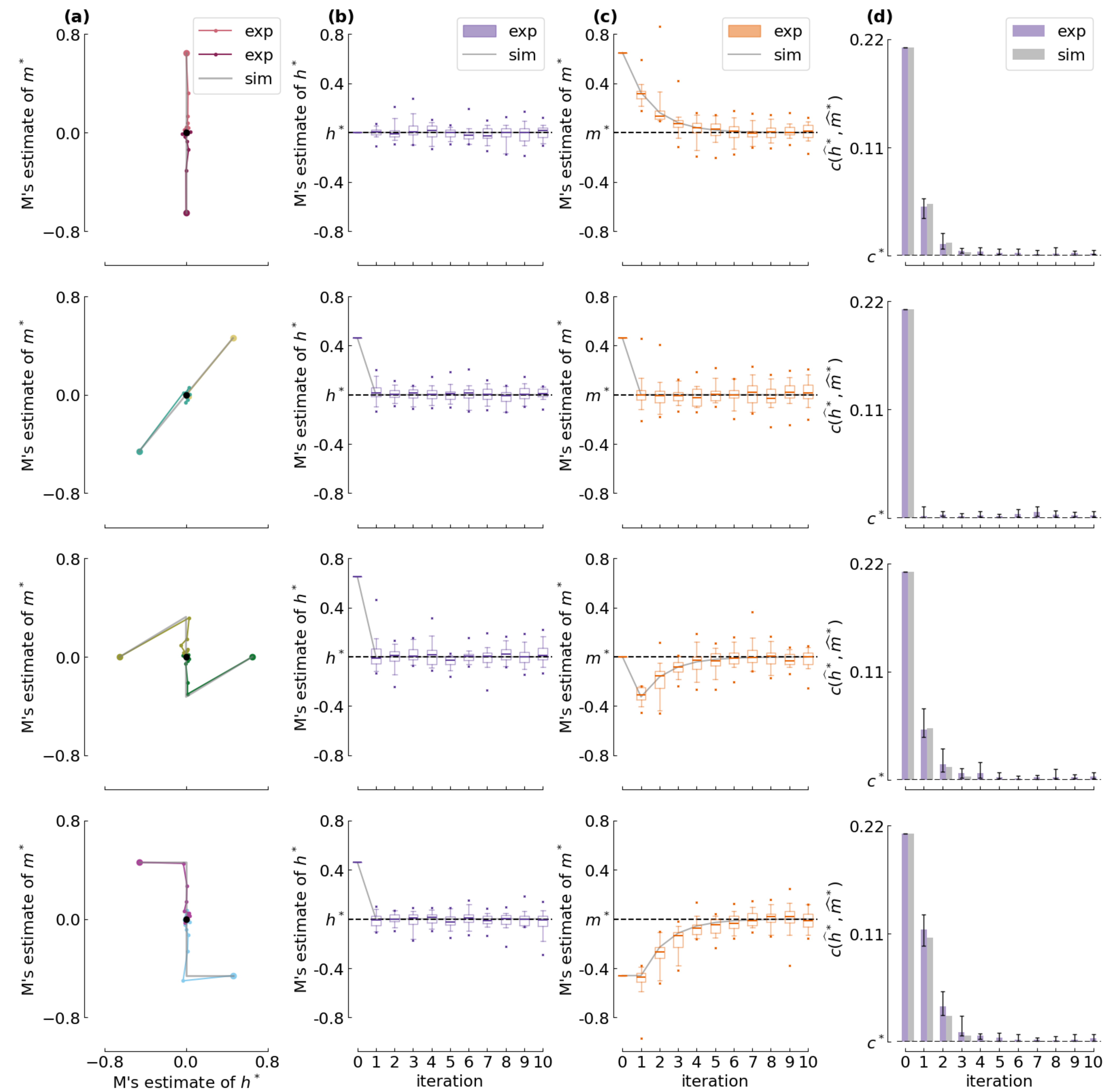
$$\underset{h}{\operatorname{argmin}}\, c\left(h, L(h - \hat{h}^*) + \hat{m}^*\right) = (I + L^T L)^{-1}\left(L^T L \hat{h}^{*T} - L^T \hat{m}^{*T}\right)$$

Combining the human's best response with the formulas from our learning algorithm defines a discrete-time linear system,

$$\begin{bmatrix} \hat{h}^*_+ \\ \hat{m}^*_+ \end{bmatrix} = \begin{bmatrix} (I + L^T L)^{-1} L^T L & -(I + L^T L)^{-1} L^T \\ \alpha L_\Delta \left(I + L_\Delta^T L_\Delta\right)^{-1} L_\Delta^T L_\Delta - \alpha L_\Delta & I - \alpha L_\Delta \left(I + L_\Delta^T L_\Delta\right)^{-1} L_\Delta^T \end{bmatrix} \begin{bmatrix} \hat{h}^* \\ \hat{m}^* \end{bmatrix}$$

where L is the linear term in the machine's affine policy and $L_\Delta = L + \Delta$ is the perturbed linear term.

Empirical vs Simulation Results ($n = 80$, $n = 10$ for each initialization point)



Conclusion

- We present a new learning algorithm that enables a human-machine interaction to converge to the minimum of a share cost function.
- The novelty in our algorithm is that the cost function is only known to the human, leaving the machine to iteratively update its estimate of the minimum solely by observing the human's response to its policies.
- In the case where the cost function is quadratic, our algorithm defines an affine discrete-time linear system, facilitating analysis of convergence.

ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY of WASHINGTON

ADVISERS: Samuel A. Burden, Lillian J. Ratliff

Questions?

email: jisa@uw.edu