Value-Based Deep RL Scales Predictably

Oleh Rybkin, Michal Nauman, Preston Fu, Charlie Snell,
Pieter Abbeel, Sergey Levine, Aviral Kumar

Question

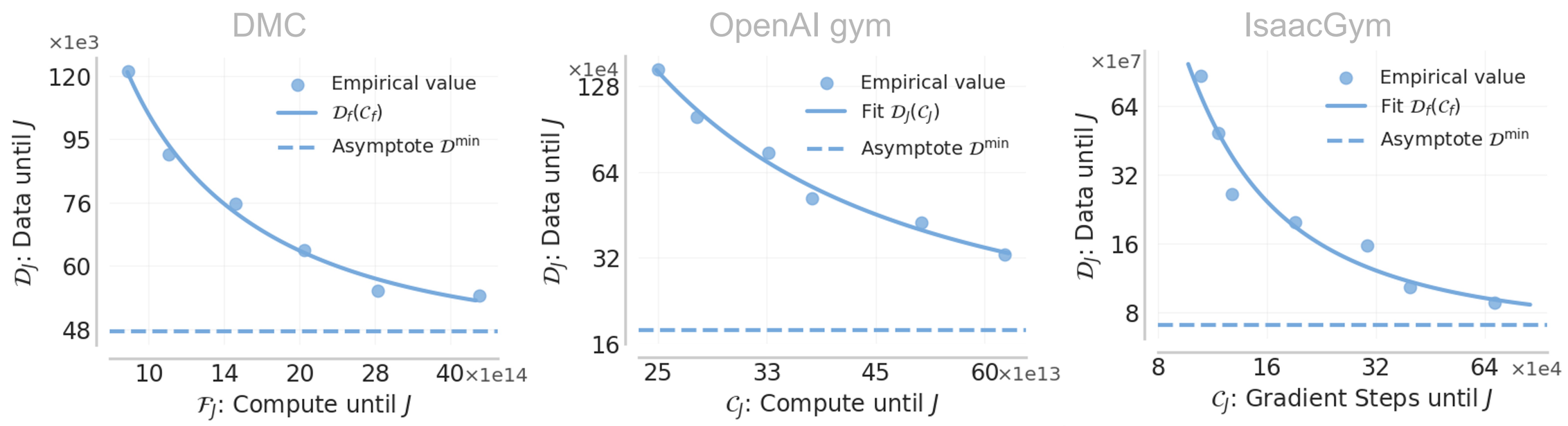
Can we predict the optimal settings for a large-scale run from small-scale runs?

- In RL, need to account both for online *data* ($\mathcal{D}$) and *compute* ($\mathcal{C}$) requirements
- *Budget* requirement reflects the total cost $\mathcal{F} = \mathcal{C} + \delta \cdot \mathcal{D}$
- Which hyperparameters to use for the large-scale run?

Notation

- $\mathcal{C}$ - FLOPs
- $\mathcal{D}$ - data points
- σ - UTD (gradient steps per data point)
- B - batch size
- η - learning rate

Data-Compute Pareto Frontier

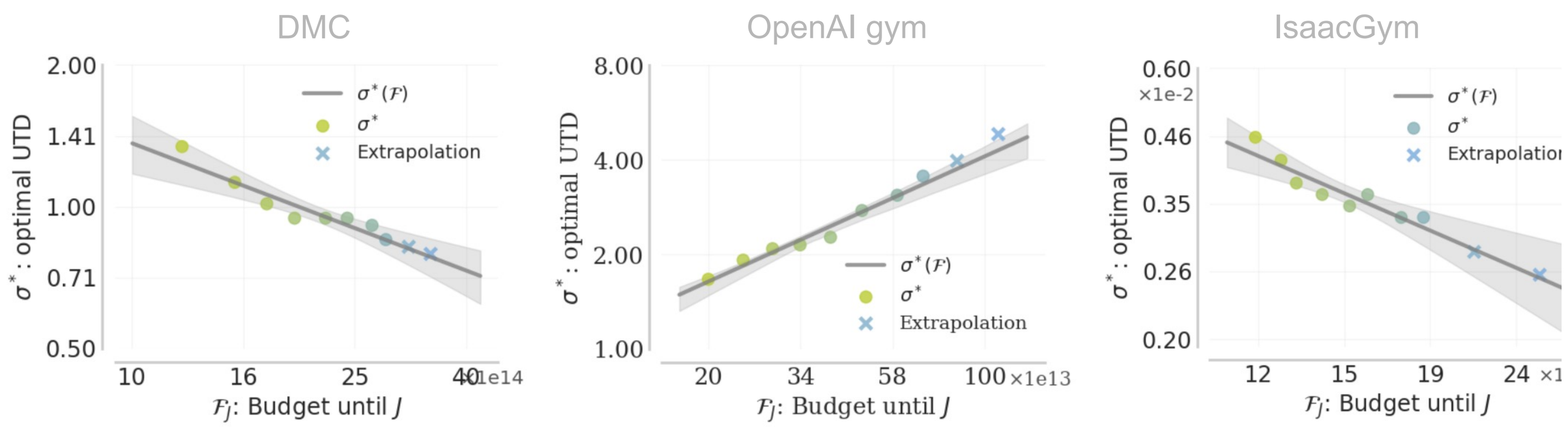


- Compute and data are predictable functions of UTD
- The tradeoff between the two is the Pareto Frontier

$$\mathcal{D}_J(\sigma) \approx \mathcal{D}_J^{\min} + \left(\frac{\beta_J}{\sigma}\right)^{\alpha_J}$$

$$\mathcal{C}_J(\sigma) \approx 10 \cdot N \cdot B(\sigma) \cdot \sigma \cdot \mathcal{D}_J(\sigma)$$

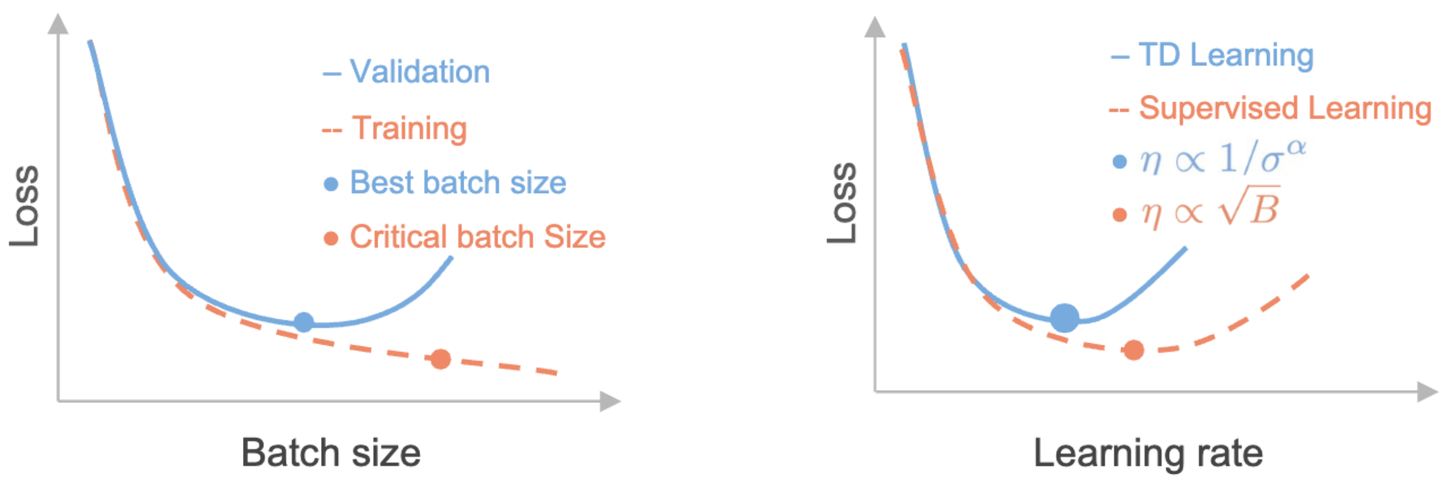
Budget Extrapolation



- We can extrapolate optimal tradeoff to higher budgets and performance levels

$$\sigma^*(\mathcal{F}_0) \approx \left(\frac{\beta_\sigma}{\mathcal{F}_0}\right)^{\alpha_\sigma}$$

Hyperparameter Dependencies



$$B^*(\sigma) \approx \left(\frac{\beta_B}{\sigma}\right)^{\alpha_B}$$

$$\eta^*(\sigma) \approx \left(\frac{\beta_\eta}{\sigma}\right)^{\alpha_\eta}$$

- Batch size is controlled by overfitting
- Learning rate needs to decrease with UTD