Dimensionality Reduction and Prioritized Exploration for Policy Search

Only update and explore effective policy parameters!

Marius Memmel[1], Puze Liu[1], Davide Tateo[1] & Jan Peters[1]
[1]Intelligent Autonomous Systems, TU Darmstadt, Germany

Problem formulation

Black-box Optimization Objective (with KL constraint):

$$\max_{p_k} \mathcal{J}(p_k), \quad \text{s.t.} \;\; \mathrm{KL}\left(p_k \| p_{k-1}\right) \leq \epsilon$$

$$\mathcal{J}(p) = \mathop{\mathbb{E}}_{\boldsymbol{\theta}\sim p}\left[J(\pi_{\boldsymbol{\theta}})\right], \quad p_k(\boldsymbol{\theta}) = \mathcal{N}\left(\cdot|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$$

$$J(\pi_\theta) = \mathop{\mathbb{E}}_{(\boldsymbol{x}_t,\boldsymbol{u}_t)\sim\pi_\theta,\mathcal{P},\iota}\left[\sum_{t=0}^{T}\gamma^t r(\boldsymbol{x}_t, \boldsymbol{u}_t)\right]$$

The sample complexity of this approach scales poorly with the dimensionality of $\boldsymbol{\theta}$, particularly when using a full covariance matrix $\boldsymbol{\Sigma}_k$. Furthermore, samples are extremely costly in a Robot Learning context.

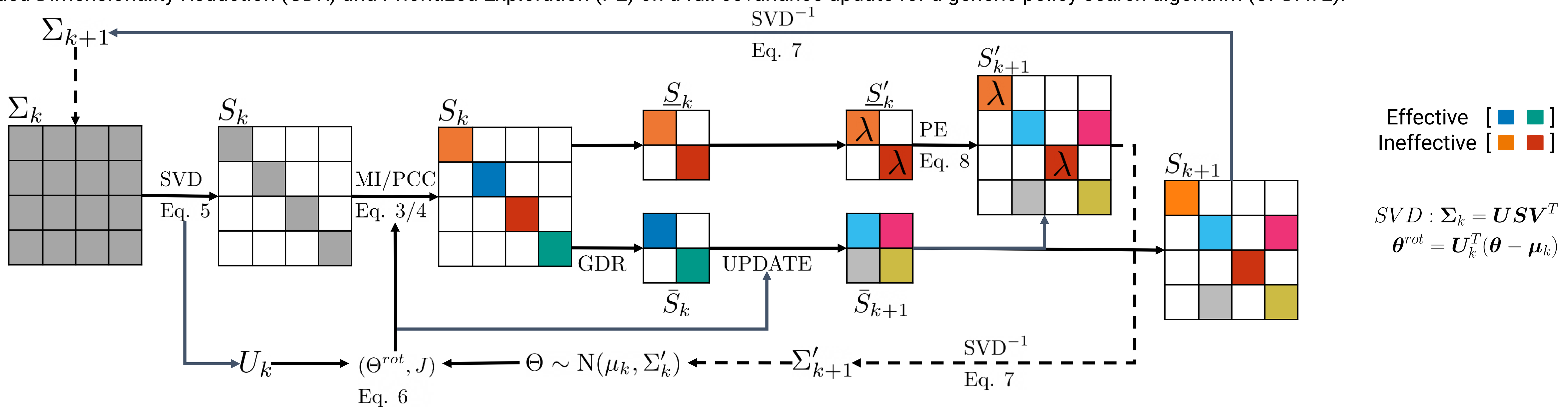
Approach

Assumption: *Not all policy parameters contribute equally to the return!*

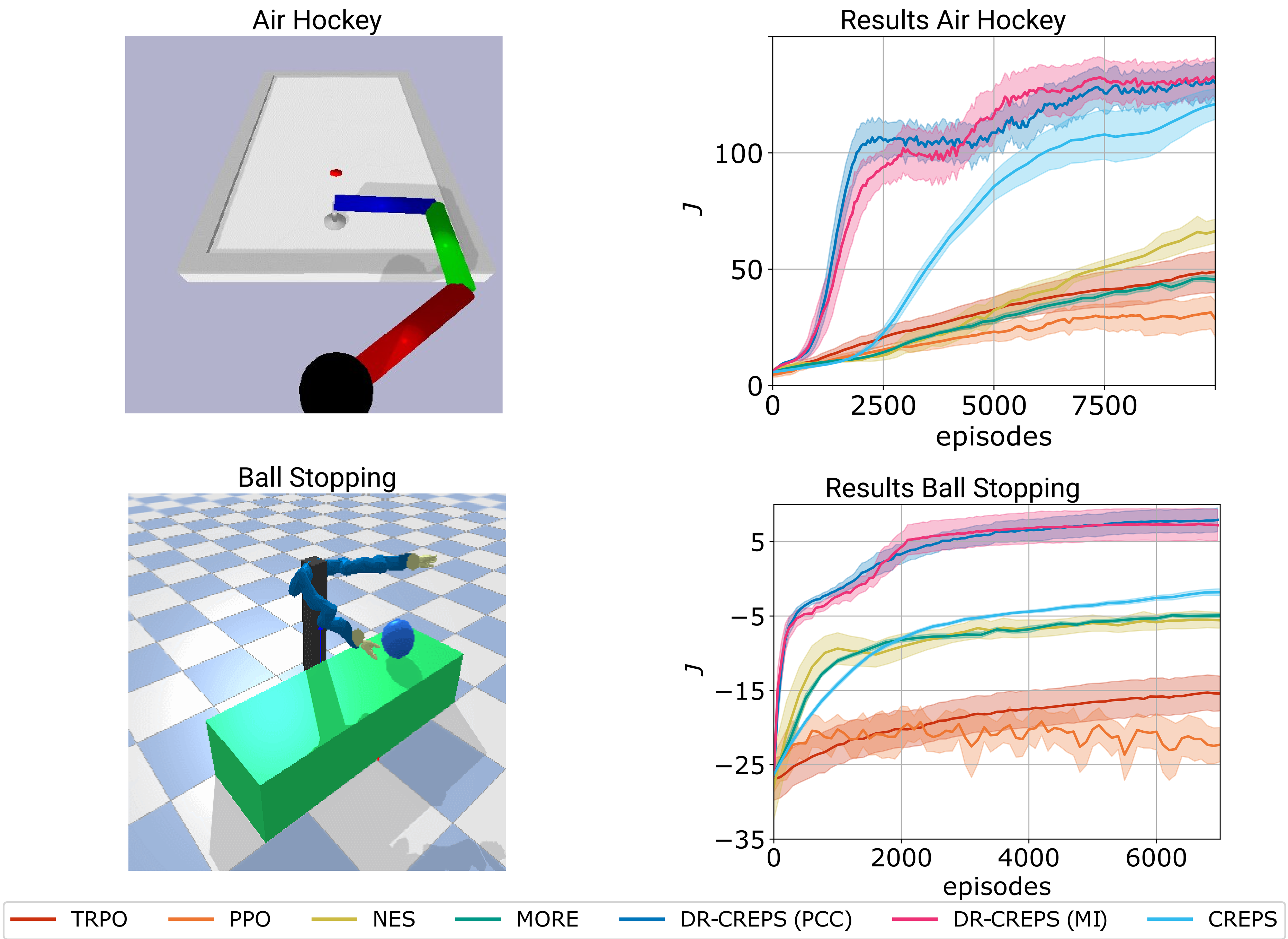
1. Diagonalize covariance matrix $\boldsymbol{\Sigma}_k$ via *SVD* and continue work in rotated space
2. Identify important parameters with Parameter Effectiveness Metric $C[\boldsymbol{\Theta}^i; \boldsymbol{J}]$
   linear correlation: *Pearson Correlation Coefficient* (PCC)
   non-linear correlation: *Mutual Information* (MI)
3. **Prioritized Exploration**: decrease variance of ineffective parameters by scaling factor $0 \leq \lambda \leq 1$ which increases exploration of effective parameters
4. **Guided Dimensionality Reduction**: update partial covariance matrix of effective parameters $\bar{S}_k$ with *CREPS* and substitute result back into original covariance matrix to obtain S_{k+1}

Implementation Details

Guided Dimensionality Reduction (GDR) and Prioritized Exploration (PE) on a full covariance update for a generic policy search algorithm (UPDATE).



Experimental Evaluation



Conclusions

- Exploiting parameter effectiveness leads to faster learning
- More updates with fewer samples are beneficial
- Structured policies can reduce task complexity in Robot Learning

We acknowledge also the support provided by China Scholarship Council (No. 201908080039).