Sparse Latent Space Policy Search

Kevin Sebastian Luck, Joni Pajarinen, Erik Berger, Ville Kyrki, Heni Ben Amor
Arizona State University, Aalto University, Technical University Freiberg

Motivation

Most robotic systems have a high number of degree-of-freedom, while most tasks in robotics are intrinsically low dimensional, for example,

- grasping
- walking
- human arm movements

Hence, we want to

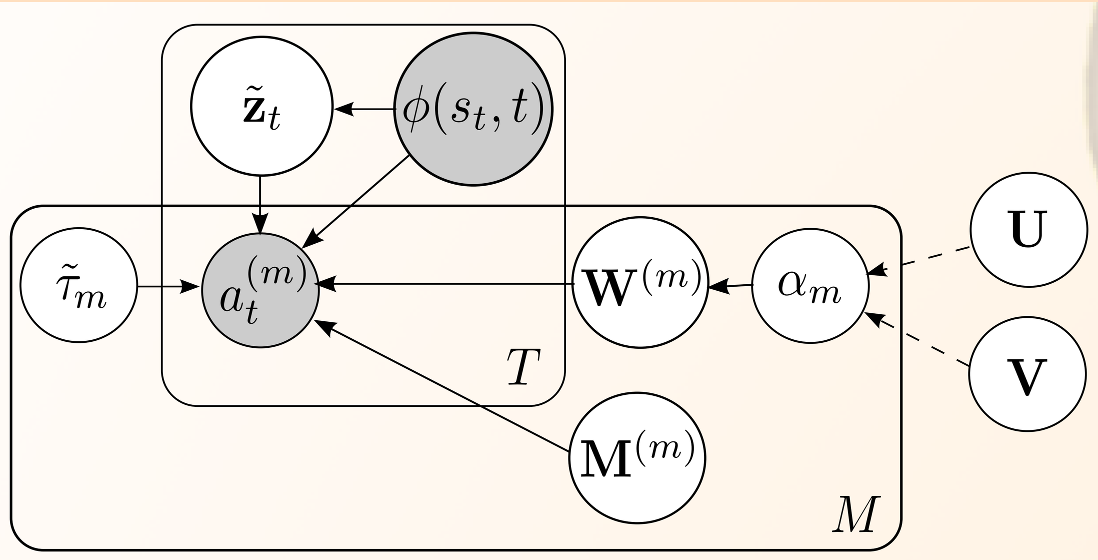
exploit the low dimensional nature of the tasks

and, furthermore,

use prior structural knowledge

to learn movements in an efficient and meaningful way.

Graphical Model



$\mathbf{a}_t$	Action	$\mathbf{M}$	Mean matrix
$\tilde{\tau}$	(Isotropic) precision	α	Sparse structure prior
$\tilde{\mathbf{z}}$	Latent variable	T	Time
$\boldsymbol{\phi}$	Basic functions	$\mathbf{s}, t$	State and time
$\mathbf{W}$	Transformation matrix	M	Number of groups

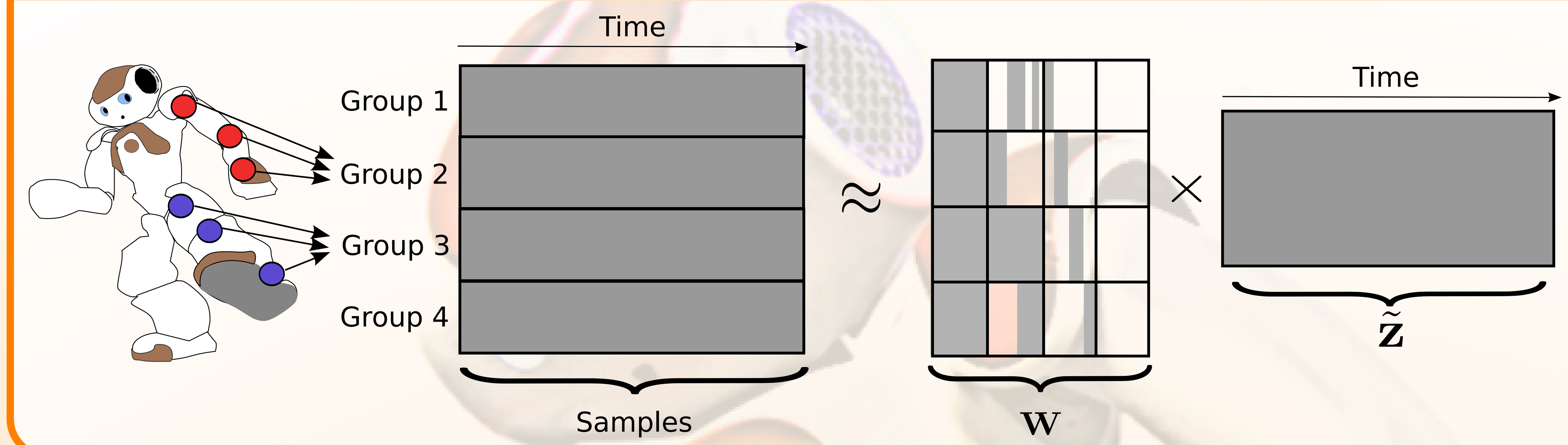
Contact Information

Web www.kevin-luck.net
Email ksluck@asu.edu

ASU Interactive Robotics Lab
Aalto Intelligent Robotics Group

Data lab.engineering.asu.edu/interactive-robotics

Main Idea



Variational Inference

By following a Variational Bayes approach on a lower bound we can derive the following formula for the estimation of the approximated q-distributions.

$$\log q_j(\theta_j) = const + \int_{\theta_{-j}} \prod_{i \neq j} q_i(\theta_i) \int p_{\text{old}}(\boldsymbol{\tau}) \log \prod_{t=1}^{T} \pi(\mathbf{a}_t, \theta | \mathbf{s}_t) \frac{p(r = 1|\boldsymbol{\tau})}{\hat{R}} \mathrm{d}\boldsymbol{\tau}\mathrm{d}\theta_{-j}. \quad (1)$$

Group Factor Policy Search

The **model equation** for the actions (of a robot) is

$$\mathbf{a}_t^{(m)} = \left(\mathbf{W}^{(m)}\mathbf{Z}_t + \mathbf{M}^{(m)} + \mathbf{E}_t^{(m)}\right) \boldsymbol{\phi}(\mathbf{s}_t, t) . \quad (2)$$

The **probabilistic policy** is given by

$$\pi(\mathbf{a}_t|\theta, \mathbf{s}_t) = \prod_{m=1}^{M} \mathcal{N}\left(\mathbf{a}_t^{(m)} \middle| \mathbf{W}^{(m)}\tilde{\mathbf{z}}_t + \mathbf{M}^{(m)}\boldsymbol{\phi}, \frac{\operatorname{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)}{\tilde{\tau}_m}\mathbf{I}\right) . \quad (3)$$

The prior of the transformation matrix incorporating structural information and sparsity is

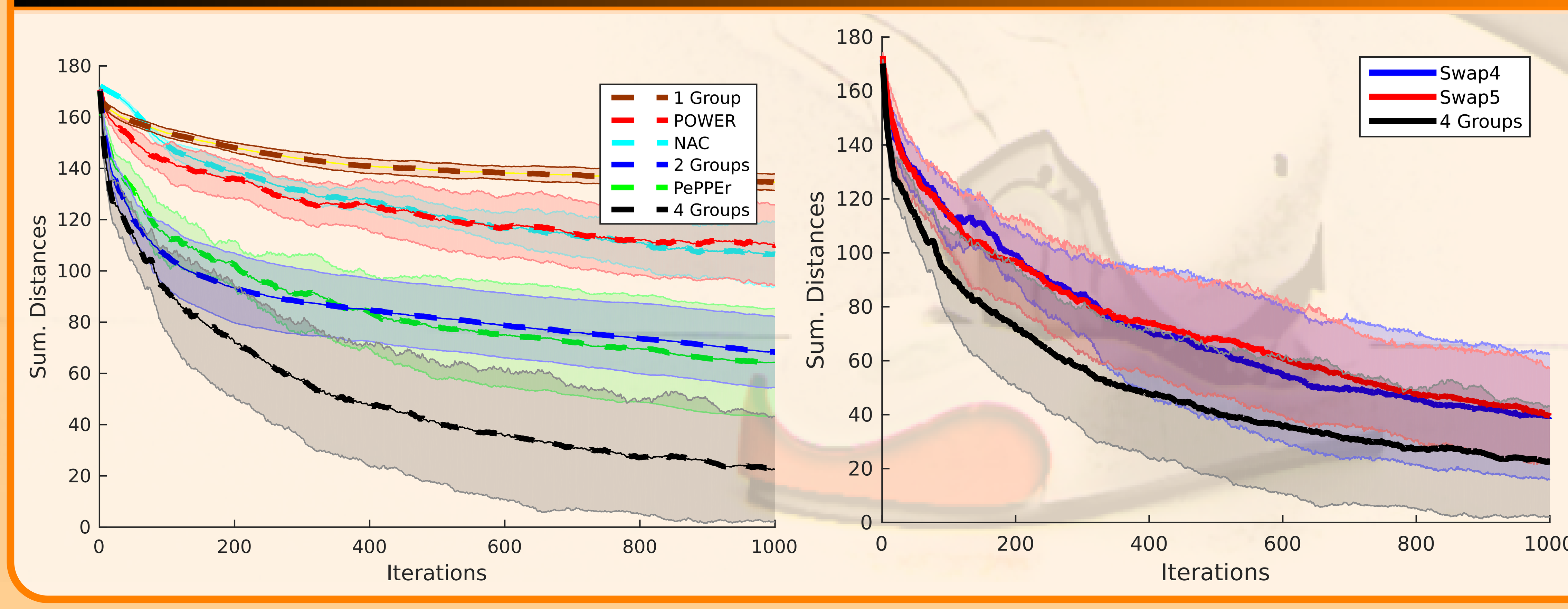
$$p(\mathbf{W}|\boldsymbol{\alpha}) = \prod_{m=1}^{M}\prod_{k=1}^{K}\prod_{d=1}^{D_m} \mathcal{N}\left(\mathbf{w}_{d,k}^{(m)} \middle| \mathbf{0}, \alpha_{m,k}^{-1}\right) . \quad (4)$$

The distributions of other parameters are either normal or gamma distributions with

$$\mathbf{M} \sim \mathcal{N}\left(\mathbf{M}_{\text{old}}, \sigma^2\mathbf{I}\right), \qquad \tilde{\mathbf{z}} \sim \mathcal{N}\left(\mathbf{0}, \operatorname{Tr}\left(\boldsymbol{\phi}\boldsymbol{\phi}^{\mathrm{T}}\right)\mathbf{I}\right),$$

$$\alpha_{m,k} \sim \mathcal{G}\left(a^{\alpha}, b^{\alpha}\right), \qquad \tilde{\tau}_m \sim \mathcal{G}\left(a^{\tilde{\tau}}, b^{\tilde{\tau}}\right).$$

Comparison - Movement of two sim. robot arms



Conclusion

We derived a novel reinforcement learning algorithm that integrates dimensionality reduction and policy search by using sparse structural prior distributions.
Resulting factors of the latent space model specific behaviour of (joint) groups in robot arms.

Algorithm

Input: Reward function $R(\cdot)$ and choose number of latent dimension n. Set fixed hyper-parameters $a^{\tilde{\tau}}, b^{\tilde{\tau}}, a^{\alpha}, b^{\alpha}$, σ^2 and define groupings.

---

while *reward not converged* **do**
  for *h=1:H* **do** # Sample H rollouts
    for *t=1:T* **do**
      $\mathbf{a}_t = \mathbf{W}_i\mathbf{Z}\boldsymbol{\phi} + \mathbf{M}\boldsymbol{\phi} + \mathbf{E}\boldsymbol{\phi}$ with $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\mathbf{E} \sim \mathcal{N}(\mathbf{0}, \tilde{\boldsymbol{\tau}})$, where $\tilde{\boldsymbol{\tau}}^{(m)} = \tilde{\tau}_m\mathbf{I}$
      Execute action $\mathbf{a}_t$
    Observe and store reward $R(\boldsymbol{\tau})$
  Initialization of q-distribution
  while *not converged* **do**
    Update $q(\mathbf{M})$
    Update $q(\mathbf{W})$
    Update $q\left(\tilde{\mathbf{Z}}\right)$
    Update $q(\boldsymbol{\alpha})$
    Update $q(\tilde{\boldsymbol{\tau}})$
  $\mathbf{M} = \mathbb{E}_{q(\mathbf{M})}[\mathbf{M}]$
  $\mathbf{W} = \mathbb{E}_{q(\mathbf{W})}[\mathbf{W}]$
  $\boldsymbol{\alpha} = \mathbb{E}_{q(\boldsymbol{\alpha})}[\boldsymbol{\alpha}]$
  $\tilde{\boldsymbol{\tau}} = \mathbb{E}_{q(\tilde{\boldsymbol{\tau}})}[\tilde{\boldsymbol{\tau}}]$

Result: Linear weights $\mathbf{M}$ for the feature vector $\boldsymbol{\phi}$, representing the final policy. The columns of $\mathbf{W}$ represents the factors of the latent space.

References

[1] Arto Klami, Seppo Virtanen, Eemeli Leppaaho, and Samuel Kaski. Group factor analysis. 2014.

[2] Kevin Sebastian Luck, Gerhard Neumann, Erik Berger, Jochen Peters, and Heni Ben Amor. Latent space policy search for robotics. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 1434–1440. IEEE, 2014.