Natural Gradient Shared Control

Yoojin Oh[1], Shao-Wen Wu[1], Marc Toussaint[1,2] and Jim Mainprice[1,3]
firstname.lastname@ipvs.uni-stuttgart.de
[1]Machine Learning and Robotics Lab, University of Stuttgart, Germany
[2]Learning and Intelligent Systems Lab ; TU Berlin ; Berlin, Germany
[3]Max Planck Institute for Intelligent Systems ; IS-MPI ; Tübingen, Germany

***Abstract*— In this paper, we propose a formalism for shared control, which is the problem of defining a policy that blends direct user control and autonomous control. The challenge posed by any shared autonomy system is to maintain user control authority while allowing the robot to support the user. Our proposed solution relies on natural gradients emerging from the divergence constraint between the robot and the shared policy. We approximate the Fisher information by sampling a learned robot policy and computing the local gradient. We use this as a measure to represent how sensitive the autonomous policy changes in the local region and augment the user control when necessary. A user study performed on a manipulation task demonstrates that our approach allows for more efficient task completion while keeping control authority against a number of baseline methods.**

I. INTRODUCTION

Shared control has been studied to exploit the maximum performance of a robot system, by combining human understanding and decision making with robot computation and execution capabilities. A linear blending paradigm introduced in Dragan et. al [1] is still widely applied [2]–[4]. In the approach, the amount of arbitration is dependent on the confidence of user prediction. When the robot predicts the user's intent with high confidence, the user often loses control authority. This has been reported to generate mixed preferences from users, and some users prefered to keep control authority despite longer completion times [5], [6]. When assistance is against the user's intentions, it can aggravate the user's workload [1]; the user "fights" the assistance rather than gain help from it.

We formulate shared control as an optimization problem (see Figure 1). The shared control action is chosen to maximize the user's internal action-value function while constraining it to be close to the autonomous robot policy. We construct the Fisher information matrix F that expresses how sensitive a distribution changes in the local neighborhood of the state. When the autonomous robot policy is represented as a vector field over the state space, the user can maintain more control authority in regions where the field does not diverge. On the contrary, at a state where the robot policy is rapidly changing in the local region (e.g. near an obstacle or near a goal), the inverse Fisher information matrix adjusts the user's actions so that the robot gains more authority. Utilizing the Fisher information matrix we introduce the term "Natural gradient shared control".

We defined a teleoperation task where a user performs pick-and-place tasks with a simulated robot arm and compared the quantitative metrics. We show that our shared control paradigm can assist the user towards accomplishing the goal while enabling more control authority to the user.

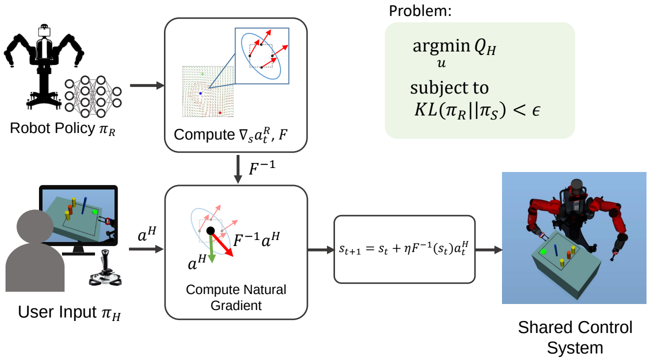

Problem:
$\underset{u}{\text{argmin}}\, Q_H$
subject to
$KL(\pi_R \| \pi_S) < \epsilon$
Robot Policy π_R
Compute $\nabla_s a_t^R, F$
F^{-1}
User Input π_H
a^H
$F^{-1}a^H$
a^H
Compute Natural Gradient
$s_{t+1} = s_t + \eta F^{-1}(s_t) a_t^H$
Shared Control System

Fig. 1: Overview of our method. F is estimated by sampling the robot policy and computing a local gradient. F^{-1} augments the user policy, resulting in a natural gradient step update.

II. METHODS

A. Natural Gradient Shared Control

Let $s \in \mathcal{S}$ be the state of the system, $a^H \in \mathcal{A}^{\mathcal{H}}$ as the user action, $a^R \in \mathcal{A}^{\mathcal{R}}$ be the autonomous robot action, and $u \in \mathcal{U}$ be the shared control action. The human and the robot agent each select actions following their stochastic policies, π_H, and π_R. Our goal is to find a shared control policy π_S that solves the following optimization problem.

$$\begin{aligned} \underset{u_t}{\arg\max} \quad & Q_H(s_t, u_t) \\ \text{subject to} \quad & \mathrm{KL}(\pi_R \| \pi_S) < \epsilon \end{aligned} \tag{1}$$

The shared control policy is chosen to maximize the user's internal action-value function $Q_H(s_t, u_t)$ at each step. Predicting Q_H can be challenging due to interpersonal differences. Instead, we regard the user action as an estimate of $\nabla_s Q_H(s_t, a_t^H)$ at each step.

The constraint on the KL-divergence between the robot policy and the shared policy ensures that the shared policy does not deviate from the autonomous robot policy. The problem can be expressed using a Lagrange Multiplier, assuming a linear approximation of our objective $\nabla_s Q_H(s_t, a_t^H)$ and a quadratic approximation of the KL-divergence constraint. This leads to an update rule which introduces natural gradient adaptation.

$$
\begin{aligned}
s_{t+1} &= s_t + \eta F(s_t)^{-1} \nabla_s Q_H(s_t, a_t^H) && (2)\\
&= s_t + \eta F(s_t)^{-1} a_t^H && (3)\\
&= s_t + \eta u_t && (4)
\end{aligned}
$$

η is the step size and the natural gradient corresponds to the shared control action $u_t \sim \pi_S(\cdot|s_t, a_t^H, a_t^R)$. We utilize the approximation $a_t^H \propto \nabla_s Q_H(s_t, a_t^H)$ in Equation 3. The proportionality constant is absorbed by the step size η.

The Fisher information matrix $F(s_t)$ can be interpreted as the sensitivity of the autonomous robot policy π_R to changes of the parameter. Intuitively, a deterministic robot policy regressed over the whole state space defines a vector field. This vector field integrates information about which optimality and constraint trade-offs are made about the underlying actions. For example when an obstacle is in an environment, it acts as a *source* (positive divergence) in the vector field resulting in a repulsive action. When the policy is goal-directed, the goal acts as a *sink* (negative divergence). The vectors around the goal point inward. F measures how sensitive the field changes and emphasizes or discounts towards certain directions of u_t.

B. Computing the Fisher Information Matrix

We approximate F as the curvature of the robot's action-value function at a given state:

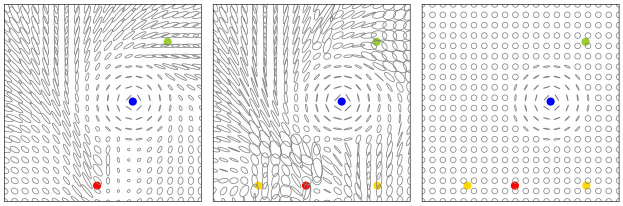
$$
\begin{aligned}
F(s_t) &= \mathop{\mathbb{E}}_{\pi_H} [\nabla_s \log \pi_R(a_t^R|s_t) \nabla_s \log \pi_R(a_t^R|s_t)^T] && (5)\\
&\approx \nabla_s^2 Q_R(s_t, a_t) && (6)\\
&\approx \frac{1}{2}(\nabla_s \tilde{a}_{t,g}^R + \nabla_s \tilde{a}_{t,g}^{R,T}) && (7)
\end{aligned}
$$

π_R is represented as a neural network policy that outputs an optimal velocity towards a known goal given the state of the environment. We use Locally Weighted Regression (LWR) to fit a local model L_g using a set of sampled states and actions inferred using π_R. As we consider action (velocity) as an approximate of the first derivative of the Q-function, we consider the Jacobian of the robot action w.r.t. state $\nabla_s a_t^R$ as the Hessian of the Q-function. $\nabla_s \tilde{a}_t^R$ is the Jacobian computed using the finite difference method with actions $\tilde{a}_{t,g}^R$ from L_g. We decompose the matrix as a sum of symmetric and a skew-symmetric matrix and apply the symmetric matrix.

Figure 2 shows $F(s)^{-1}$ computed over the state space with different assistance modes. The ellipse represents the direction that the user's action is stretched along. When the ellipse is close to a circle the user has more control authority over the system. When the ellipse is narrow, for example near an obstacle, the robot augments the user's action towards one direction.

III. EXPERIMENTS AND RESULTS

A user study with 16 participants (12 male, 4 female) was conducted to assess the efficacy of our method. We defined a teleoperation task where the user controls the robot's gripper using a joystick to grab a cylinder and bring it to a position while avoiding a static obstacle. We hypothesized that our method allows the user to take more control during the task while ensuring safety and efficiency.



(a) Single goal

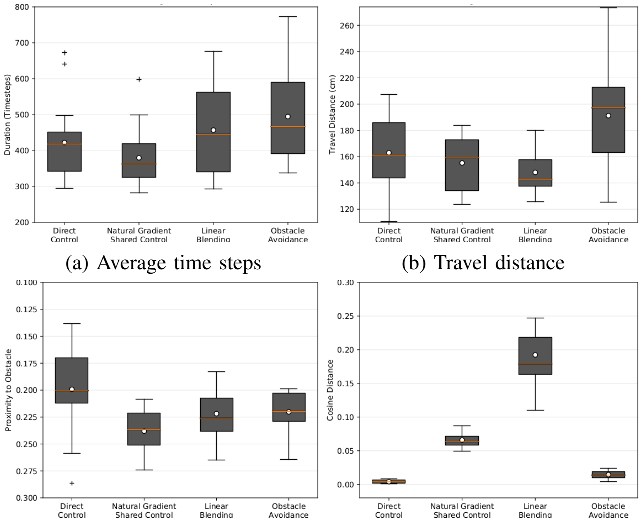
(b) Belief-weighted over goals

(c) Obstacle avoidance

Fig. 2: Ellipse plots from the eigenvalues, eigenvectors of $F(s)^{-1}$ for (a) a single goal, (b) in the presence of multiple goals computed as weighted sum over beliefs, (c) obstacle avoidance

Each user performed three sets of demonstrations, where each set consisted of four different environments repeated over the control methods, a total of 16 episodes. The order of teleoperation methods was random, and the random order was predefined and balanced over the study. The teleoperation methods were as following: Direct Control (**DC**), Natural Gradient Shared Control (**NG**), Linear Blending (**LB**), Obstacle Avoidance (**OB**). For **LB**, we followed the "timid" mode suggested in Dragan et.al. [1]. The method **OA** provides minimal assistance near the obstacle using a signed distance function as shown in Figure 2(c).

Overall, our method showed reliable performance in task execution while still maintaining compliance with user commands. We believe the results are convincing towards positive user experience and we plan to investigate it in our future work.



(a) Average time steps

(b) Travel distance

(c) Proximity to obstacle

(d) Cosine distance

Fig. 3: Boxplots for each control paradigm across all users for (a) time steps, (b) travel distance, (c) proximity to obstacle, (d) cosine distance.

References

[1] A. D. Dragan and S. S. Srinivasa, "A policy-blending formalism for shared control," *The International Journal of Robotics Research*,

vol. 32, no. 7, pp. 790–805, 2013.

[2] A. Goil, M. Derry, and B. D. Argall, "Using machine learning to blend human and robot controls for assisted wheelchair navigation," in *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2013, pp. 1–6.

[3] S. J. Anderson, J. M. Walker, and K. Iagnemma, "Experimental performance analysis of a homotopy-based shared autonomy framework," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 2, pp. 190–199, 2014.

[4] M. Gao, J. Oberländer, *et al.*, "Contextual task-aware shared autonomy for assistive mobile robot teleoperation," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 3311–3318.

[5] D.-J. Kim, R. Hazlett-Knudsen, *et al.*, "How autonomy impacts performance and satisfaction: Results from a study with spinal cord injured subjects using an assistive robot," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 42, no. 1, pp. 2–14, 2011.

[6] S. Javdani, H. Admoni, *et al.*, "Shared autonomy via hindsight optimization for teleoperation and teaming," *The International Journal of Robotics Research*, vol. 37, no. 7, pp. 717–742, 2018.