Goal Babbling with Direction Sampling for simultaneous exploration and learning of inverse kinematics of a humanoid robot

Rania Rayyes and Jochen Steil

Research Institute for Cognition and Robotics, Bielefeld University,
Universitätsstr, 33615 Bielefeld, Germany
{rrayyes,jsteil}@cor-lab.uni-bielefeld.de

Abstract. Goal Babbling is a recently introduced method for direct learning of the inverse kinematics within few hundred movements even in high-dimensional sensorimotor spaces. This paper investigates if random selection of movement directions in goal space can be used for Goal Babbling without pre-specifying goals, instead, the goals will be generated along the chosen direction. This so-called Direction Sampling was previously developed for a 2D workspace with a simple planar arm model, whereas we scale it to full 3D and a complex 9-DOF humanoid robot (COmpliant huMANoid - COMAN) integrating simplified walking behavior by means of a simulated robot-floating base. The paper evaluates how much of the workspace can be discovered, what the performance of the learned inverse model is, and how the different degrees of freedom can be constrained by changing the exploration noise model. The results show that the combination of Goal Babbling and Direction Sampling works even under these difficult conditions, but has limitations in performance if the workspace is not fully explored.

Keywords: Exploratory learning, Goal Babbling, Humanoid robot

1 INTRODUCTION

With the advent of humanoid and other robots with many degrees of freedom, motion control and in particular movement skill learning has attracted renewed attention recently. Historically, movement skill learning has been a topic in machine learning, robotics and neuroscience since the 90th, where it is widely accepted that human motor control is organized on the basis of forward and inverse models [1]. A number of schemes have been developed for learning of such internal models, among them the seminal work on distal teachers [2] and on feedback error learning [3]. However, these models were applied to simple robots only and assume that first a forward model is learned or is already available which converts actions into predicted outcomes, before learning an inverse model, that converts goals, e.g. positions to reach, into motor commands. These models cannot describe how to learn from scratch, i.e the first phase of motor learning when a good

body coordination is not yet established. Therefore, a number of works have proposed an initial learning phase to obtain a forward model by random exploration of motor commands under the notion of motor babbling [4], [5]. This appears unrealistic, however, for robots with many degrees of freedom. The respective high-dimensional spaces for motor commands cannot be explored randomly or systematically because of a combinatorial explosion. Furthermore, there is an evidence from infant studies that already neonates perform goal directed action from the very beginning of learning [6]. Apparently, they learn how to reach by trying to reach, and they adapt their motion by iterating their tries [7]. These insights motivated researchers to turn to the idea of direct learning of inverse models [5], [7], [8]. Such models directly yield a motor command to achieve a goal and do not depend on a previously learned forward model. But they have to deal with both the problem of redundancy, which is the problem that a redundant robot has many possible ways to achieve a goal and needs to make a selection from these. And they need to assure the scalability in high dimensions. A particularly efficient has been introduced under the notion of Goal Babbling [9]. Goal Babbling follows the approach to explore rather the low-dimensional space of goals, e.g. target positions in space to be achieved for a robot hand. This is in contrast to exploring the much higher dimensional action space of motor commands that motor babbling explores. Goal Babbling systematically generates consistent samples for supervised learning of the inverse model, for which typically a local linear map [7] or a neural network [10] is employed as learner. It has been shown that Goal Babbling scales to high dimensions (up to 50 DoF for a planar arm [7]), it has been applied to learn the body coordination of the humanoid robot ASIMO [9], and its online version [7] has for instance been applied to learn the inverse kinematics of an soft elephant trunk robot [11] in a truly "learning-while-behaving" fashion.

One limitation of Goal Babbling is that the algorithm needs a predefined set of goals to achieve, for instance a grid of positions to reach in the task space. If the workspace is not fully known a priori or unreachable goals are devised, either only parts of the work space are explored or it can be time consuming to ask the robot to achieve unreachable goals. To overcome this drawback, in [12] an extension of Goal Babbling to discover and determine the reachable workspace while learning the inverse model was introduced as "Direction Sampling". The algorithm is based on random selection of movement directions to explore while learning the inverse kinematic mapping along the way. A planar arm was used for evaluation the effectiveness of this direct sampling. In this case, the workspace is 2D and thus very limited, whereas random directions in 2D are easy to follow. The current paper investigates, if direction sampling can be used for a realistic humanoid robot by simulating the robot COMAN (Compliant Humanoid) that can move in space in order to discover its 3D workspace autonomously. This obviously is a harder problem, which is further complicated by the fact that the robot has very different types of movement available. It can 'walk', which we simulate by means of a simple linear x-y translation in space, and reach with its full upper body with nine degrees of freedom.

Algorithm 1 Online Goal Babbling

INPUT: home postures q_{home}, targets X^*, and forward kinematic function FK.

for number of iteration
  for each target x^*
    generate a temporary path
    for each temporary point along the path x_t^*
      estimate joints' value $\hat{q}_t^*$
      add exploratory noise E: $q_t^+ = \hat{q}_t^* + E(x_t^*, t)$
      $x_t^+ = FK(q_t^+)$
    end for
  end for
end for

OUTPUT: $learner \longleftarrow (q_t^+, x_t^+)$

2 The Goal Babbling Algorithm

The algorithm is given in Algo. 1. Goal babbling starts with an initial inverse estimate g, which has parameters θ adaptable by learning, and is initialized in t = 0 such that it always suggests some comfortable home posture: g(x^* ,θ_0) = const = q^{home} . Then, continuous paths of target positions x_t^* are iteratively chosen by interpolating between the K representative points located on the grid of predefined goals. The system then tries to reach for these targets, which roughly corresponds to infants' early goal-directed movement attempts. For that purpose, the current inverse estimate is used to generate a motor command q_t^*.

The command q_t^* is sent to the robot and executed, the outcomes (q_t^+, x_t^+) are observed, and the parameters θ_t of the inverse estimate are updated online before the next example is generated. It is crucial to make the distinction between q_t^* and q_t^+ at this point: the command q_t^* might not be executable, or might not yet be reached at the time of measurement. Hence, only (q_t^+, x_t^+) but not (q_t^*, x_t^*) represents a sample of the ground truth forward function that is useful for learning. The perturbation term $E(x_t^*, t)$ adds exploratory noise in order to discover new positions or more efficient ways to reach for the targets. This allows to unfold the inverse estimate from the home posture and finally find correct solutions for all positions in the volume of targets X^* spanned by the predefined goals [11]. The most efficient movement will be learned by using the weighting scheme, which helps out to solve the redundancy problem.

For learning, a regression mechanism is needed in order to represent and adapt the inverse estimate $g(x^*)$. The goal directed exploration itself does not require particular knowledge about the functioning of this regressor, such that in principal any regression algorithm can be used. For an incremental online learning, a local-linear map has been chosen. The inverse estimate consists of different linear functions $g^k(x)$, which are centered around prototype vectors and active only in its close vicinity which is defined by a radius d. The function $g(x^*)$ is a linear combination of these local linear functions, weighted by a Gaussian responsibility function [7].

2.1 Direction Sampling

Discovering the workspace could be done by using Motor Babbling, i.e. random motor commands are executed, and their outcomes are observed. However, the robot will discover the workspace without learning it. In contrast, the Goal Babbling uses inverse model which suggests a motor command necessary to achieve a desired outcome and learns it. However, a limitation of Goal Babbling is the need to pre-specify the goals. To this aim, targets must be known beforehand or there is a risk to waste time and to distort the learned inverse model by trying to achieve unreachable targets. To tackle this issue, in [12] Direction Sampling was presented, which is an approach to discover the reachable workspace while learning the inverse kinematic mapping during the discovery. It employs Goal Babbling while generating targets in the workspace instead of predefining them. A random direction $\triangle x$ will be chosen, and the targets will be generated along this path as given in (1):

$$x_t^* = x_{t-1}^* + \frac{\varepsilon}{\|\Delta x\|} \cdot \Delta x, \tag{1}$$

where ε is a step-width, t is a time-step, x_t^* is a generated target, and x_{t-1}^* is the previous one. The robot starts exploration from its home position x^{home}, which is corresponding to some initial joints' values q^{home}. It tries to explore along the desired direction until it reaches an unachievable target i.e. the current position deviates from the desired goal by more than 90 degrees, given in (2):

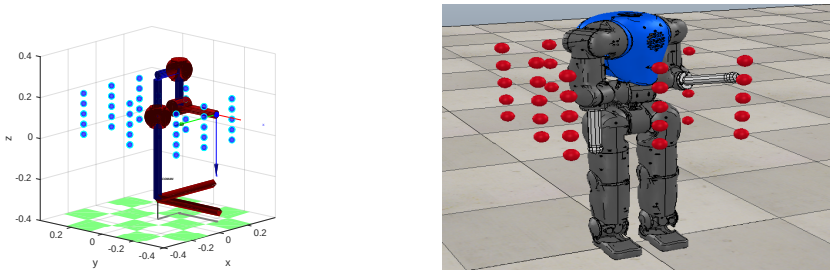
$$(x_t^* - x_{t-1}^*)^T (x_t - x_{t-1}) < 0, \tag{2}$$

where x_t is the current position, and x_{t-1} is the previous observed movement. In this case, a new direction will be chosen and the agent will try to follow it again [12]. Every 100 times the initial position q^{home} is used as a target to avoid drifting. While this mechanism is simple and worked well to explore a 2D workspace, it is not apparent that in full 3D and with a complex robot this mechanism is sufficient to explore a reasonable part of the workspace.

2.2 Noise Scaling

In this section, we introduce a further extension of the Goal Babbling, which is motivated from the idea that not all degrees of freedom should be employed equally much. E.g. walking for a robot can be considered more costly than moving its hand or arm. The previous approach of Goal Babbling already used an efficiency factor to value samples more if they feature more efficient movements. This, however, was purely geometry based, e.g. a shoulder joint needs a smaller deviation to achieve a significant hand movement than an elbow because of the longer lever. But in principle, more factors should be considered such as equilibrium, balance, and motors' synchronization. We therefore try to constrain the learning dynamics to favor solutions that use or avoid certain joints by scaling the exploratory noise for the joints' movement as

$$q_t = g(x_t^*, \theta_t) + E_t(x_t^*) w. \tag{3}$$



(a) (b)

Fig. 1: Compliant humanoid (COMAN) with floating base model in Matlab Robotics toolbox (a) and in VREP (b)

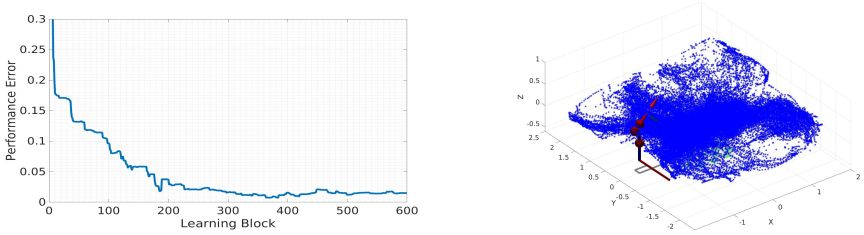
E_t is the exploratory noise weighted by a coefficient vector w. The larger the exploratory noise is in one joint variable i, i.e. the larger the respective w_i, the more likely the learning dynamics will discover a solution for reaching to a point that employs this joint. This implements an implicit, soft constraint. We give highest efficiency for the arm movement, less weight for the torso motion, and the least for the lateral displacement "walking".

3 Setup with the COMAN robot

Unlike standard manipulators, humanoid robots are not physically fixed to a base, there is a so-called floating base. Therefore, the workspace for the humanoid robot is in theory unlimited. However, if we limit the movement to some amount forward and sidewards (in the experiments: $\pm 1.5\ m$), there is a limited reachable workspace around the robot where we can expect interaction of moving, leaning with the upper body and arm motion. We target to discover this reachable workspace with the 3D Direction Sampling approach. Technically, we simulate walking by replacing the actual lower body by two additional degrees of freedom (linear forward, linear sidewards). Therefore, the floating base for the COMAN robot is simplified to move in X-Y plane. The remaining model has 7 DOF: the torso has 3 DOF, the shoulder has 3 DOF, the elbow has 1 DOF. Together with the two virtual DOF for the floating base this is in total a nine dimensional joint space. Note that the types of movement here are very different: linear in the floating base, rotational in the torso and in the arm. The kinematic model has been setup in MATLAB using the Robotic Toolbox [13] and in V-REP for visualization as shown in Fig. 1(a) and Fig. 1(b) respectively.

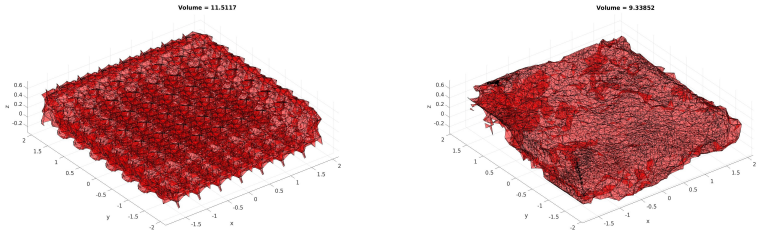
4 Evaluation

In a first step, we verify that Goal Babbling can deal with the complex robot setup and learn to reach 45 targets arranged in a regular 3D grid as illustrated in Fig. 1(a): 15 targets in front of the robot at distance 30 cm, 15 at the coronal plane, and 15 in the back of the robot at distance 30 cm as well. The vertical distance between targets is 5 cm. Fig. 2(a) shows a typical learning curve, the



(a) (b)

Fig. 2: (a) Goal Babbling error in meter, (b) discovered workspace using Direction Sampling



(a) (b)

Fig. 3: Reachable workspace (a) vs Discovered workspace (b)

reaching error drops very fast and already after 200 learning epochs a decent performance on the targets is achieved, i.e. after 800 movements the error drops to 2 mm. The robot leans to use the lateral movement of the floating base to reach to targets behind its body and combines it with the torso and arm movement. Next we turn to Direction Sampling. To obtain a ground truth of the reachable workspace, we use extensive sampling in simulation with a kind of motor babbling to collect 3×10^6 samples. Then the volume of the reachable workspace is estimated using the `alphavol` MATLAB function with radius $R = 0.01$. The estimated volume is 11.5117 m^3 and is illustrated in Fig. 3(a). However, the robot learns nothing about reachable targets in this way. Now, we apply Direction Sampling to explore, discover, and learn the workspace simultaneously. Although the direction sampling is very simple, the robot manages to discover most of the workspace in few thousand steps. Fig. 2(b) illustrates the discovered workspace after 60000 samples. The Direction Sampling algorithm is evaluated after 10^4, 5×10^4, 6×10^4, 10^5, and 10^6 samples. The discovered workspace is again estimated using `alphavol` function. The results are illustrated in Table.1, and the discovered workspace after 10^6 samples is illustrated in Fig. 3(b). As expected, the robot visits an increasing portion of the workspace with more learned samples, and it performs well on the grid targets which were previously used to evaluate the efficiency of standard Goal Babbling, as shown in Table 1.

To gain more insight about the performance relative to the distance from the body, two further target grids for reaching are presented in front of the robot with distance 1 m, and 0.5 m. Then targets are presented in the coronal plane, i.e. some are inside the robot such that it must "walk", i.e. the lateral movement

Table 1: Volume of discovered workspace averaged over 5 runs

Number of Samples	Average Volume Discovered	Percentage Volume Discovered	Average Error for 45 targets
10^4	0.715 ± 0.07	6.211%	0.377 m
5×10^4	2.17 ± 0.2	18.85%	0.0284 m
6×10^4	3.18 ± 0.02	27.62%	0.0484 m
10^5	3.59 ± 0.01	31.816%	0.047 m
10^6	9.338	81.18%	0.036 m
Goal Babbling	-	-	0.02

Table 2: Testing Error Measured for Different No. of Samples.

	Distance				
	Front		On	Behind	
No. of Samples	-1 m	-0.5 m	0 m	0.5 m	1 m
10^4	0.2091 m	0.16 m	0.17 m	0.42 m	0.2517 m
5×10^4	0.2315 m	0.0234 m	0.02 m	0.074 m	0.1256 m
6×10^4	0.14 m	0.127 m	0.03 m	0.158 m	2.37 m
10^6	0.1020 m	0.0123 m	0.0181 m	1.0625 m	7.17 m

Table 3: Discovered workspace after adding noise scaling

Factor of the scaling noise	Percentage Volume of the Discovered Workspace
[1 1 1 1 1 1 1 1 1]	27.62%
[0.15 0.15 0.5 0.5 0.5 1 1 1 1]	12.5%
[0.1 0.1 0.5 0.5 0.5 1 1 1 1]	10.2%
[0.01 0.01 0.5 0.5 0.5 1 1 1 1]	3.3%

in x-y direction. Finally, they are behind the robot at a distance 0.5 m, and 1 m. The performance error is illustrated in Table. 2. Apparently, the targets behind are much more difficult to reach and in the final row, some of the targets were out of the discovered workspace and produced large errors, as the learner extrapolated rather badly because it is a local linear.

The final experiment is on modulating the learning dynamics to use particular joints more or less. The noise is weighted as shown in Table. 3, which scales down exploration with the floating base (i.e. walking) systematically. The discovered workspace after adding the constrains was evaluated after 60000 samples. The robot discovered less workspace, because of the constrains. For example, 0.01 limit the joint movement exploration more than 0.15 illustrated in Table 3.

5 Conclusion

We have shown that Goal Babbling with or without combination with Direction Sampling can be used even in a complex scenario where a 9 DOF humanoid robot discovers its 3D workspace. There were no indications of local minima or

of the algorithm being captured in already explored areas, which is quite remarkable given the complexity of the mapping to be learned. The results also show, however, that a large number of direction changes are needed and the learner naturally performs badly for goals in the undiscovered areas. It is interesting that indirectly, through scaling of the noise, certain degrees of freedom can be preferred. Future work shall improve the direction sampling. A more active choice of directions towards undiscovered areas should yield better performance, however, at the cost of an increased complexity of the algorithm.

ACKNOWLEDGMENT

R. Rayyes received funding from the German Academic Exchange Service (DAAD)- "Research Grants-Doctoral Programme in Germany" scholarship.

References

1. D. Wolpert, R. C. Miall, and M. Kawato, "Internal models in the cerebellum," *Trends Cognit. Sci.*, vol. 2, pp. 338–347, 1998.
2. M. I. Jordan and D. E. Rumelhart, "Forward models: Supervised learning with a distal teacher," *Cognitive Science*, vol. 16, pp. 307–354, 1992.
3. M. Kawato, "Feedback-error-learning neural network for supervised motor learning," *in Advanced Neural Computers. Elsevier*, 1990.
4. Y. Demiris and A. Meltzoff, "The robot in the crib: A developmental analysis of imitation skills in infants and robots," vol. 17, 2008, pp. 43–53.
5. A. Baranes and P. Oudeyer, "Active learning of inverse models with intrinsically motivated goal exploration in robots," *Robot. Auton. Syst.*, vol. 61, no. 1, pp. 49–73, 2013.
6. C. von Hofsten, "An action perspective on motor development," *Trends in CogSci*, vol. 8, p. 266–272, 2004.
7. M. Rolf, J. J. Steil, and M. Gienger, "Online goal babbling for rapid bootstrapping of inverse models in high dimensions," in *IEEE Int. Conf. Development and Learning and on Epigenetic Robotics*, 2011, pp. 1–8.
8. S. V. D'Souza and S. Schaal, "Learning inverse kinematics," *Int. Conf. Intelligent Robots and Systems (IROS)*, vol. 1, pp. 298 – 303, 2001.
9. M. Rolf, J. J. Steil, and M. Gienger, "Goal babbling permits direct learning of inverse kinematics." *IEEE Trans. Autonomous Mental Development*, vol. 2, no. 3, pp. 216–229, 2010.
10. G. bin Huang, Q. yu Zhu, and C. kheong Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, pp. 489–501, 2006.
11. M. Rolf and J. Steil, "Efficient exploratory learning of inverse kinematics on a bionic elephant trunk," in *IEEE Trans. Neural Networks and Learning Systems*, 2014, pp. 1147–1160.
12. M. Rolf, "Goal babbling with unknown ranges: A direction-sampling approach," in *IEEE Int. Conf. on Development and Learning and on Epigenetic Robotics (ICDL)*, 2013, pp. 1–7.
13. P. Corke, "A robotics toolbox for matlab," *IEEE Robotics & Automation Magazine*, vol. 3, no. 1, pp. 24–32, March 1996.