Bootstrapping inverse Kinematics with Goal Babbling

Matthias Rolf, Jochen Steil, Michael Gienger

2010

Preprint:

This is an accepted article published in Proceedings of the International Conference on Development and Learning. The final authenticated version is available online at: https://doi.org/[DOI not available]

Bootstrapping inverse Kinematics with Goal Babbling

Matthias Rolf, Jochen J. Steil, Michael Gienger

***Abstract*—We present an approach to learn inverse kinematics of redundant systems without prior- or expert-knowledge. The method allows for an iterative bootstrapping and refinement of the inverse kinematics estimate. We show that the information structure induced by goal-directed exploration enables an efficient resolution of inconsistent samples solely from observable data. The bootstrapped solutions are aligned for a maximum of movement efficiency, i.e. realizing an effector movement with a minimum of joint motion. We derive and illustrate the exploration and learning process with a low-dimensional kinematic example and show that the same procedure scales for high dimensional problems, such as hyperredundant planar arms with up to 50 degrees of freedom.**

***Index Terms*—Motor Exploration, Motor Learning, Inverse Kinematics, Goal Babbling**

I. Introduction

Learning to control the own body is a fundamental problem in human development. Infants need to learn the most basic skills like reaching for an object. The ability to learn control from scratch also allows to master more complex tasks like writing or riding a bicycle [1]. The control of such tasks can be well understood with the notion of internal models [2]. Internal models describe relations between motor commands and their consequences. A forward model predicts the consequence of a motor command, while an inverse model suggests a motor command necessary to achieve a desired outcome.

How can internal models emerge from uncoordinated behavior? Before internal models can be applied for coordinated control, experience must be gained by exploration. The crucial question is how that experience is generated, i.e. how infants explore their body for coordination. Piaget suggested that human (motor-)development progresses in several stages [3]. At first infants react purely reflexive. Meltzoff and Moore [4] suggested the concept of "body babbling" as an initial stage in which experience is gathered. Infants can then use this knowledge to attempt goal-directed action and fine-tune their skills on the fly.

However, evidence over the last decades clearly shows that infants perform goal-directed movements from the very beginning. Von Hofsten [5] has repeatedly highlighted the role of goal-directed action for infant motor development. "Before infants master reaching, they spend hours and hours trying to get the hand to an object in spite of the fact that they will fail, at least to begin with." [5] Statistics revealed that already days after birth, infants attempt goal-directed action by means of arm and finger movements [6], [7]. Early goal-directed actions clearly suggest that "learning by doing" has a central role in infant motor development. Infants learn to reach by trying to reach.

A. The Learning Problem

Before infants, but also robots, can master reaching, inverse models must be learned for their limbs or even the full body. In the present work, we investigate the kinematic control of redundant systems. Formally, we consider the relation between joint angles $q \in \mathbf{Q} \subset \mathbb{R}^m$ and effector poses $x \in \mathbf{X} \subset \mathbb{R}^n$ (e.g. the position of the hand). Thereby m is the number of degrees of freedom (DOF) and n is the dimension of the target variable (e.g. $n = 3$ for the spatial position of a hand). The forward kinematics function $f(q) = x$ uniquely describes the causal relation between both sizes. If the hand needs to be positioned at some desired coordinate x^*, an inverse model $g(x^*) = q$ is needed to find appropriate joint angles q ($f(g(x^*)) = x^*$). Such a function is not uniquely defined if the number of joint angles m exceeds the number of controlled dimensions n.

An example is shown in Fig. 1: a robot arm with two joints ($m = 2$) and a total length of 1m. Since we want to consider a redundant structure, the goal here is to control only the height of the effector ($n = 1$). The redundancy appears in form of manifolds through the joint-space, on which all joint-angles apply the same effector height. An inverse kinematics function in this example must suggest joint angles $q \in \mathbb{R}^2$ for each desired effector height $x^* \in \mathbb{R}^1$. Such an estimate can be visualized by a one dimensional manifold through the joint space. Figure 2 shows an example generated with motor babbling. For several target heights x^*, the joint angle estimates are shown by colored markers on the manifold (the joint angles are furthermore visualized by corresponding postures in the 3D simulation). Small green markers show the examples used for learning. If the estimate is correct, each marker must be positioned on the redundancy manifold that represents the set of joint angles that do indeed realize the desired effector height. An accurate inverse estimate positions all colored markers on the contour with the same color.

Two substantial problems must be solved when an inverse model shall be learned from experience:

1) Inversion of causality. It is difficult to get *at least one* correct solution q for a target x^*. The outcome x for a motor command q can simply be probed by applying q (the cause) and observing x. This probing is not possible for inverse problems.

Matthias Rolf and Jochen J. Steil are with the Research Institute for Cognition and Robotics (CoR-Lab) at Bielefeld University, Germany. Mail: {mrolf, jsteil}@CoR-Lab.Uni-Bielefeld.de

Michael Gienger is with the Honda Research Institute Europe in Offenbach, Germany.

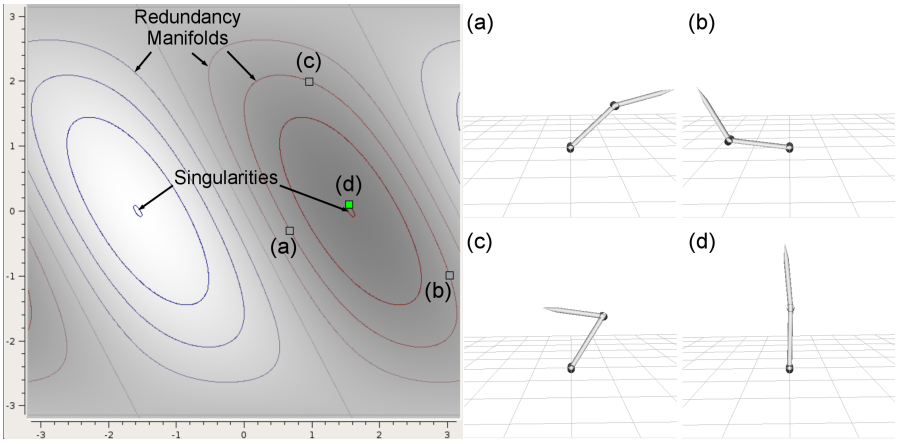


Fig. 1: Robot arm (length $1m$) with two joints. The left side shows the joint space. Multiple configurations (see postures a-c) can be used to apply the same height of the end effector, but can not be averaged without leaving the desired height (see posture d). The sets of joint angles that apply the same height are marked by colored contours in the joint space.

2) Non-convexity. It is also difficult to deal with the possible occurrence of *multiple* solutions. Non-convex solution sets (see Fig. 1) prohibit learning from multiple solutions.

Existing approaches to the exploration and learning of inverse kinematics split into two groups: error-based and example-based methods. *Error-based* methods follow the "learning by doing" approach. An estimate $g(x^*)$ of the inverse kinematics is used for trying to reach for a target position. Using the joint angles $q = g(x^*)$ suggested by the inverse estimate, the resulting position of the effector is evaluated with the forward kinematics function $x = f(q)$. One group of mechanisms is based on the "motor error", which is a correction Δq of the joint angles in order to improve the performance. In *Feedback-error learning* [8], [9] it is simply assumed that a mechanism to compute that motor error is already available. In *Learning with distal teacher* [10], [11] a forward model $\hat{f}(q)$ must be learned in parallel. A motor error can be derived analytically by differentiating the forward model. Both methods can in principle deal with redundant systems. The critical problem is that the motor error is not directly observable, and on its own subject to redundancy. A special case of error-based learning has been developed in [12]. The error in the effector space $x - x^*$ is used directly for learning. The information used in this case is fully observable, but the method has never been shown to work for redundant degrees of freedom ($n < m$) and requires a "good enough" inverse estimate in advance.

Example-based methods use example configurations $(f(q), q)$ for the learning of an inverse estimate $g(x)$. The existing approaches differ in the way how such examples are generated. Motor babbling [13], [14] is a pure random form of exploration. It has been proposed as an implementation of the "body babbling" introduced by Meltzoff and Moore, but was used also before body babbling was introduced [15], [16]. Joint angles are randomly chosen from the set of all possible configurations $q_i \in \mathbf{Q}$. This approach can solve the inversion of causality, if enough examples are generated. However, it is subject to the non-convexity problem. Also

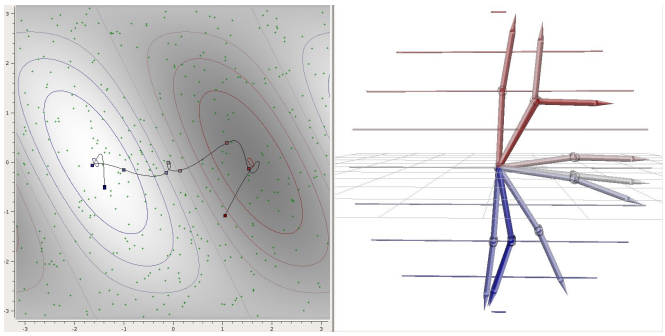


Fig. 2: Direct inverse learning with random motor babbling as exploration process does not yield a correct inverse estimate because of the non-convex redundancy manifolds.

goal-directed exploration approaches have been investigated [17], [18], which we discuss and extend in the next section. The approach – as previously discussed in literature – solves neither the inversion of causality nor the non-convexity problem in a reliable fashion.

Example-based learning of inverse kinematics has only been shown to be successful if training data without inconsistent solutions is already available [19]. Autonomous approaches to learn inverse kinematics based on examples have so far consequently failed on redundant systems.

II. Goal Babbling

With "Goal Babbling" we generally refer to the successful bootstrapping of some motor skill by the (*i*) repeated process of (*ii*) trying to accomplish (*iii*) multiple goals related to that skill. Goal babbling means learning by doing from scratch. We use this terminology in order to highlight the similarities but also differences to previous concepts. The exploration process focuses on the goals of action instead of the means (motor-commands). Intimately related to the original concept of vocal- as well as body-babbling, repetition is everything. A goal

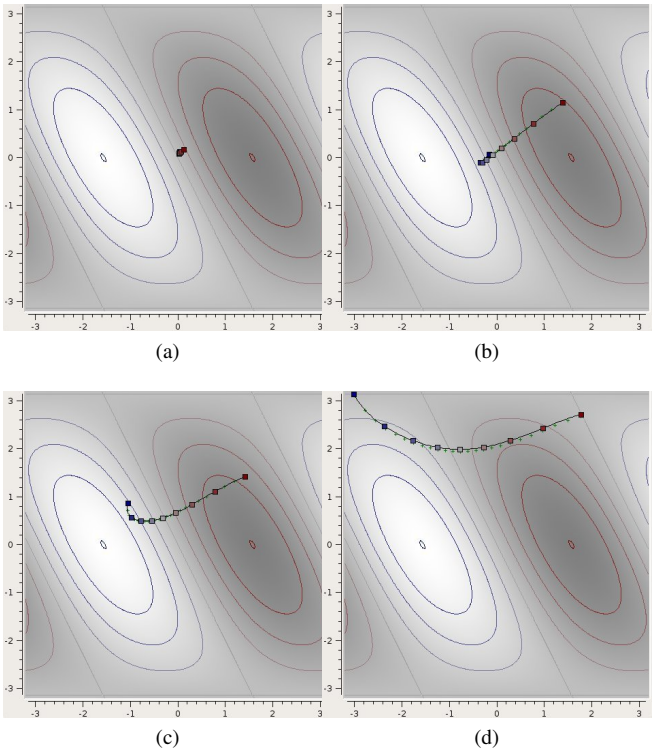


Fig. 3: Learning dynamics with plain goal-directed exploration from (a) to (d). Example data is only generated with the inverse estimate. Learning occurs from inconsistent examples, as the controlled manifold intersects some redundancy manifolds multiple times. The estimate drifts in its orthogonal direction. This method does not find an appropriate inverse estimate.

must be tried to be accomplished again and again in order to succeed.

A. Goal-directed exploration

In a goal-directed exploration [17], [18], examples $(f(q), q)$ are generated with an untrained or inaccurate inverse estimate $g(x, \theta)$, where θ are the parameters adaptable by learning. In principle, any standard machine learning approach can be chosen for $g(x, \theta)$, like neural networks, local learning schemes or polynomial regression. If the parameters are not necessary for the discussion, we will write $g(x)$ for short. Initially, a target motion is chosen and represented as a temporal sequence of target positions: $x_t^* \in \mathbf{X}^* \subseteq \mathbf{X}, t = 1...T$. The inverse estimate is then used for trying to reach for those targets: $q_t = g(x_t^*)$, $x_t = f(q_t)$. After the adaption of the parameters, the process is repeated. We will refer to this method as *plain goal-directed exploration*.

This process does not necessarily discover new target positions [18] and therefore fails to invert causality in a reliable way. Another problem is that the inverse estimate is unstable in the Nullspace of movement, i.e. along its orthogonal direction. Finally, inconsistent examples can also exist under goal-directed exploration. An example of the learning dynamics is shown in Fig. 3.

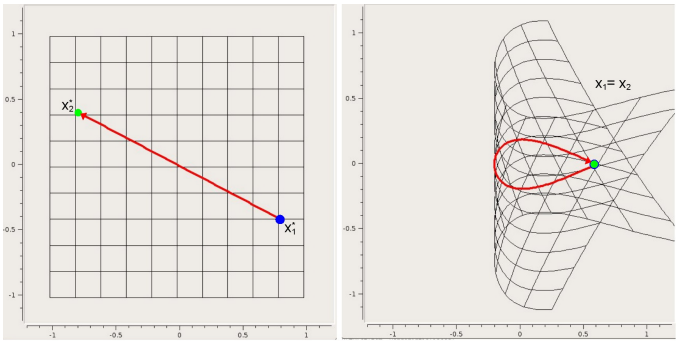


Fig. 4: *Left:* Space of target positions x^*. A linear target motion (bold line) shall be produced between two target points. *Right:* Space of results $x = f(g(x^*))$. An inconsistency occurs e.g. when the grid is folded. The formerly straight line now has a circular shape.

B. Inconsistencies

Two samples (x_1, q_1) and (x_2, q_2) are inconsistent, if they represent the same effector pose $x_1 = x_2$ but different joint angles $q_1 \neq q_2$. Regardless of the kind of exploration that is used to generate samples, two samples with exact same effector pose will rarely be found. A reliable solution must take into account the sample generation method itself.

In goal-directed exploration, examples are generated on a n dimensional manifold inside the m dimensional joint-space. This manifold is defined by the inverse estimate and spanned by the set of target positions: $\mathbf{Q}^{expl} = g(\mathbf{X}^*)$. We assume that two inconsistent samples $q_1 \neq q_2, x_1 = x_2$ may be generated in the exploration ($q_1, q_2 \in \mathbf{Q}^{expl}$). These samples must have been generated for two different target positions $x_1^* \neq x_2^*$. Identical target positions $x_1^* = x_2^*$ would lead to a contradiction, since $g(x_1^*) = g(x_2^*)$ and thus $q_1 = q_2$. We now consider that the inverse estimate is used to attempt a linear target motion between x_1^* and x_2^* (see Fig. 4, left). The system would start from the joint configuration q_1, trying to reach for x_1^*. Then the system would move its joints along some path and end up in joint configuration q_2, trying to reach for x_2^*. At the beginning and end of the movement, the effector has the same pose $x_1 = x_2$. When the effector is observed while trying to follow that straight path, two cases can occur:

1) An effector motion occurs. Since the effector returns to the same position, the observed effector movement must have a closed (e.g. circular) shape (see Fig. 4, right). The goal is to follow a straight line, i.e. keeping the movement direction constant, but the observed movement direction changes.
2) The effector pose remains constant, in spite of the joint movement from q_1 to q_2. This case can occur when the inverse estimate moves exactly along one redundancy manifold. This case is characterized by a minimum of movement efficiency. While the joints are moved, the effect on the effector is zero.

If we generate examples with goal-directed exploration and exclude both unintended changes of movement direction and inefficient movements, the remaining examples must not

contain inconsistencies.

In order to realize this exclusion, we assign weights $w_t \in \mathbb{R}$ for each example (x_t, q_t). Unintended changes of movement direction can be tackled with the following scheme:

$$w_t^{dir} = \frac{1}{2}\left(1 + cos\sphericalangle(x_t^* - x_{t-1}^*,\ x_t - x_{t-1})\right). \quad (1)$$

Thereby $\sphericalangle(x_t^* - x_{t-1}^*, x_t - x_{t-1})$ is the angle between the intended and actual movement direction of the effector. If both are identical the angle is $0.0°$ and the weight becomes $w_t^{dir} = 1.0$. If the observed movement has the exact opposite direction, the angle is $180.0°$ and the weight becomes $w_t^{dir} = 0.0$. If a circular motion occurs for a linear target motion, one half of the motion receives a higher weight than the other one and the inconsistency can be broken.

Inefficient movements can be excluded by weighting with the ratio of effector motion and joint motion, which is 0.0 if the joints move without effector motion:

$$w_t^{eff} = \frac{||x_t - x_{t-1}||}{||q_t - q_{t-1}||} \quad . \quad (2)$$

Since both weights are necessary for inconsistency resolution, they are be combined by multiplication, such that an example is ignored if any of the two criteria assigns a weight zero:

$$w_t = w_t^{dir} \cdot w_t^{eff} \quad . \quad (3)$$

The weighting scheme relies on the temporal order of samples along the trajectory, since the actual and the last sample is taken into account. In particular, it relies on goals: unintended changes of movement direction can only be detected if there is an intended direction.

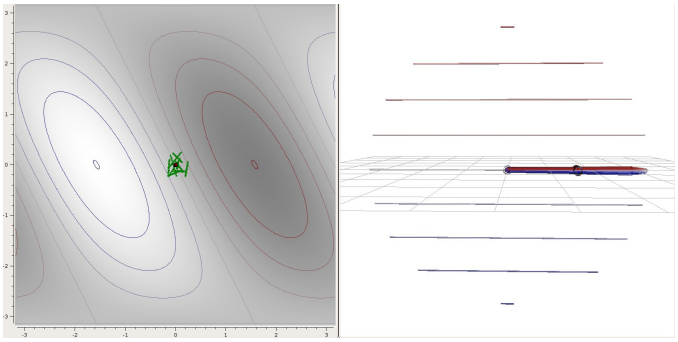
C. More Exploration

In plain goal-directed exploration only those examples are explored that are exactly on the manifold of the inverse estimate. Such behavior is highly unrealistic for human motor development. If a motor command is sent twice, neural and muscular noise as well as external perturbations can cause slightly different outcomes. Such perturbations do not result in noisy and erratic movements in the first place. For instance an imperfect gravity compensation introduces a continuous shift of the observable motion. Noise in the motor system primarily acts upon forces and accelerations and causes smooth deviations when a goal-directed movement is attempted. Human motion can be distorted by neuro-muscular noise and external perturbations, which allows to discover new effector poses by chance. We simulate such perturbations by adding a small disturbance term $E^v(x)$ to the inverse estimate:

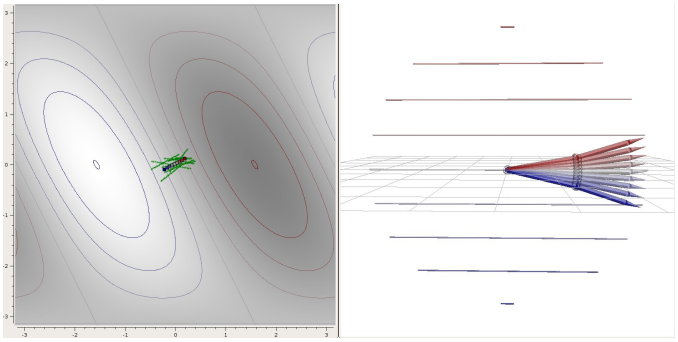
$$g^v(x) = g(x) + E^v(x). \quad (4)$$

Examples are then generated with this variation instead of the actual inverse estimate: $q_t^v = g^v(x_t^*)$, $x_t^v = f(q_t^v)$.

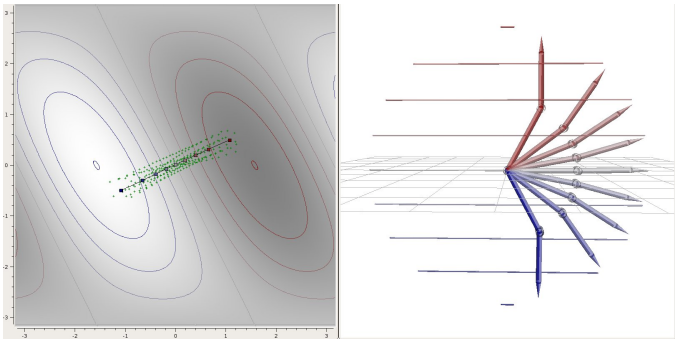
The assumptions and arguments for the inconsistency resolution still hold, since $g^v(x)$ is still a function and spans a n dimensional manifold in the joint-space. For a set of examples, generated with a variation $g^v(x)$, the weighting scheme can be applied as proposed above. The index v is



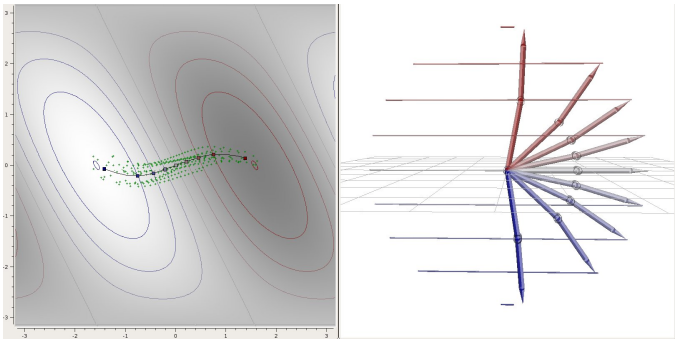
(a) The inverse estimate is initialized around the home posture.



(b) The inverse estimate has aligned with the optimal movement direction and starts to expand.



(c) The performance increases rapidly. Until the ridge of the forward function is hit.



(d) The inverse estimate finds the necessary non-linearities to reach for extreme positions.

Fig. 5: Inverse kinematics learning with Goal Babbling. The images show successive stages of the learning process. The inverse estimate is initialized around a small point in space. It spreads successively and ends up with an accurate solution.

added to identify weights for examples of a specific variation: $w_t^v = w_t^{v^{dir}} \cdot w_t^{v^{eff}}$.

Although exploration is fundamental in infancy, infants do not try to reach for an object forever. At a time, they stop exploration, relax their muscles and rest. Learning is possible from such a "neutral" motor command, since there is still a resulting effector pose. At the level of kinematics, we denote a home posture q^{home} as neutral motor command. The result $f(q^{home})$ can be observed and be used for learning as any other example. We add the example $q_0^v = q^{home}, x_0^v = f(q^{home}) = x^{home}$ to each set generated with goal-directed exploration:

$$D^v \longleftarrow \{(f(q^{home}), q^{home})\} \cup D^v \tag{5}$$

The "home" example receives the full weight $w_0^v = 1.0$.

A home posture is a stable point in exploration, and thus in learning. The inverse estimate will generally tend to reproduce the connection between q^{home} and x^{home} if it is used for learning: $g(x^{home}) \approx q^{home}$. The easiest way to achieve the result of applying the home posture is: applying the home posture. This stable point largely prevents the inverse estimate to drift away. Learning can start around the home posture and proceed to other targets.

D. Learning

Example data (and corresponding weights) from multiple different variations $g^v(x^*), v = 1...V$ is combined for learning, where $V \in \mathbb{N}$ is the number of different variations.

In the learning step, the parameters θ of the inverse estimate $g(x, \theta)$ are updated using the generated examples $(x_t^v, q_t^v), t = 0...T$ (including the home posture) and weights w_t^v in a reward weighted regression manner. Thereby the weighted command error

$$E_w^Q(\theta) = \sum_v \sum_t w_t^v \cdot (g(x_t^v, \theta) - q_t^v)^2 \tag{6}$$

is minimized. Any regression algorithm can be used for this step (e.g. linear regression schemes).

The overall procedure works in epochs. The inverse estimate is initialized with some parameters θ. We generally use a random initialization, but such that the inverse estimate generates joint configurations closely around the home posture for all goal positions. There is no a priori knowledge about the structure of the kinematics. Within one epoch, examples are generated from multiple variations, weights are assigned and the learning is done with the examples. The next epoch repeats the procedure with the updated inverse estimate.

An example of inverse kinematics learning with Goal Babbling on the 2 DOF arm (see Fig. 1) is shown in Fig. 5. The inverse estimate is initialized in a small region around the home posture, which we set to $q^{home} = (0.0, 0.0)$. The next images show the progress of the method after several epochs. The aim is to control the effector's height within the full range from -1.0m to 1.0m. Initially, only heights around $f(q^{home}) = 0$m are reachable. The extrapolation of the inverse estimate than causes a rapid expansion of the inverse estimate in the joint space. Finally, the necessary non-linearities are found to successfully reach for all target positions.

III. Experiments

In this section, we show results of Goal Babbling on a planar arm with varying degrees of freedom and investigate the influence of sensory noise. We use polynomial regression [20] to represent the inverse estimate $g(x^*, \theta)$. The input vector $x \in \mathbb{R}^n$ is expanded by a feature mapping $\Phi^P(x) \in \mathbb{R}^p$ which calculates all polynomial terms of the entries of x. Thereby P is the maximum degree of the polynomial terms and p is the number of polynomial terms that can be calculated from an n dimensional vector. All results are shown for third order polynomials ($P = 3$). A standard linear regression with parameters $\theta = \mathbf{M}$ operates on these features:

$$g(x^*, \mathbf{M}) = \mathbf{M} \cdot \Phi^P(x^*), \quad \mathbf{M} \in \mathbb{R}^{p \times m} \ . \tag{7}$$

The entries of the regression matrix $\mathbf{M}$ are adapted during learning with a gradient descent of the weighted command error as defined in equation 6. We use a learning rate of 0.2. Before exploration and learning proceed, we first set $\mathbf{M}$ to zero and make some random adaptions such that $g(x^*, \mathbf{M})$ produces joint angles close to the home posture.

For the exploration, we use linear disturbance terms:

$$E^v(x) = \mathbf{A} \cdot x + b, \quad \mathbf{A} \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \ . \tag{8}$$

The values of $\mathbf{A}$ and b are chosen randomly, such that the disturbance of any joint-angle never exceeds a range R within the bounded set of target positions $\mathbf{X}^*$:

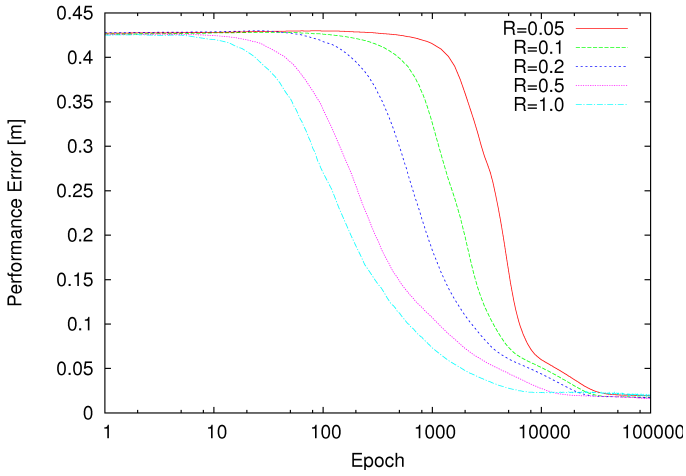
$$E^v(x) = (e_1, ..., e_m)^T, \ |e_i| <= R \ \forall \ i = 1...m, x \in \mathbf{X}^* \ . \tag{9}$$

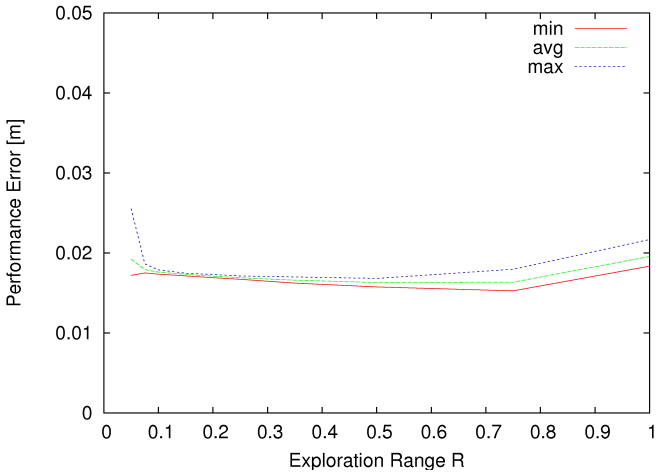
Contrary to the initial example, where only the height of the effector is controlled ($n = 1$), we now consider the 2D position control of the effector ($n = 2$). The aim in this set of experiments is to gain control over a part of the possibly reachable positions as shown in Fig. 7. We evaluate the accuracy of the inverse estimates with *performance error* on those target positions:

$$E^X(\theta) \quad = \quad \sum_t (x_t - x_t^*)^2 \tag{10}$$

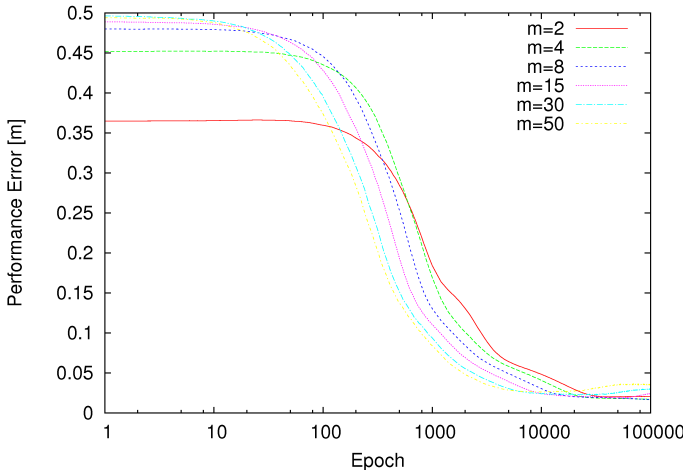
A new sequence of targets x_t^* is generated in each epoch. Ten positions are randomly selected from the target grid shown in Fig. 7. One after the other is connected by a linear target motion with five intermediate target positions.

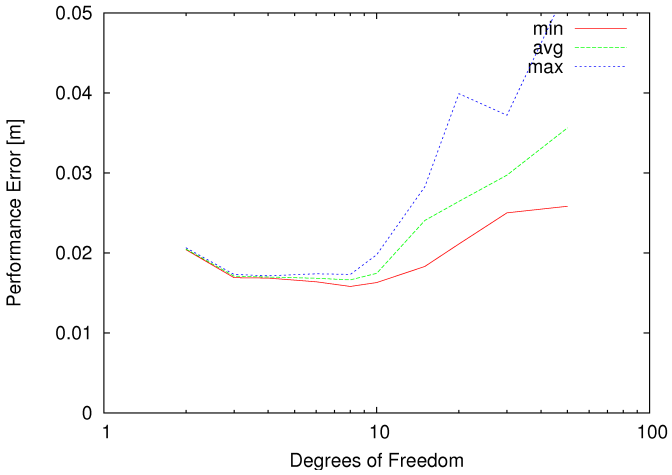
We first investigate the behavior of the exploration range R. Figure 6a shows results for R varying between 0.05 and 1.0 radian over 100000 epochs and for 20 independent trials. The number of variations was set to $V = 20$. The left plot shows the performance error over time for different values of R. The error decreases continuously. High values like $R = 1.0$ display the fastest convergence, but the residual error is slightly increased. Although the speed and the converged error vary, the general success of Goal Babbling is rather insensitive to the concrete exploration range. The performance error is minimized to a small value in all cases. An increase of error is visible for high values of R. Here examples are rather distant and the residual averaging error between the variations has a higher impact compared to small values of R. For $R = 1.0$ the examples are generated in almost the entire joint space.





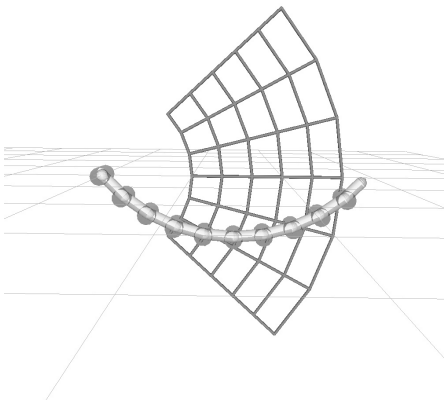
(a) Results for different exploration ranges R. Higher exploration ranges cause a faster convergence, but higher residual error.



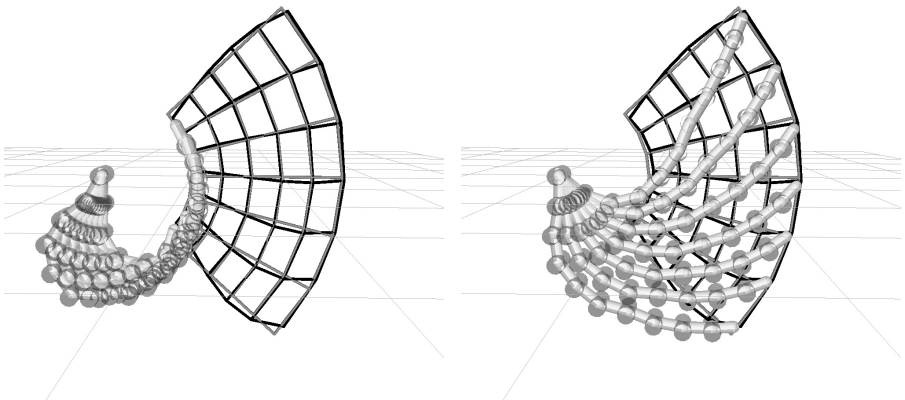


(b) The number of joints m is increased. Successful bootstrapping of inverse kinematics is possible also for 50 DOF.

Fig. 6: Performance of Goal Babbling over 100000 epochs for the planar arm, where the 2D position of the effector is controlled ($n = 2$). The left plots show the performance error over time, averaged over 20 independent trials. The finally reached error is plotted against the varied parameter on the right side. The maximum, average and minimum error of 20 trials are shown.

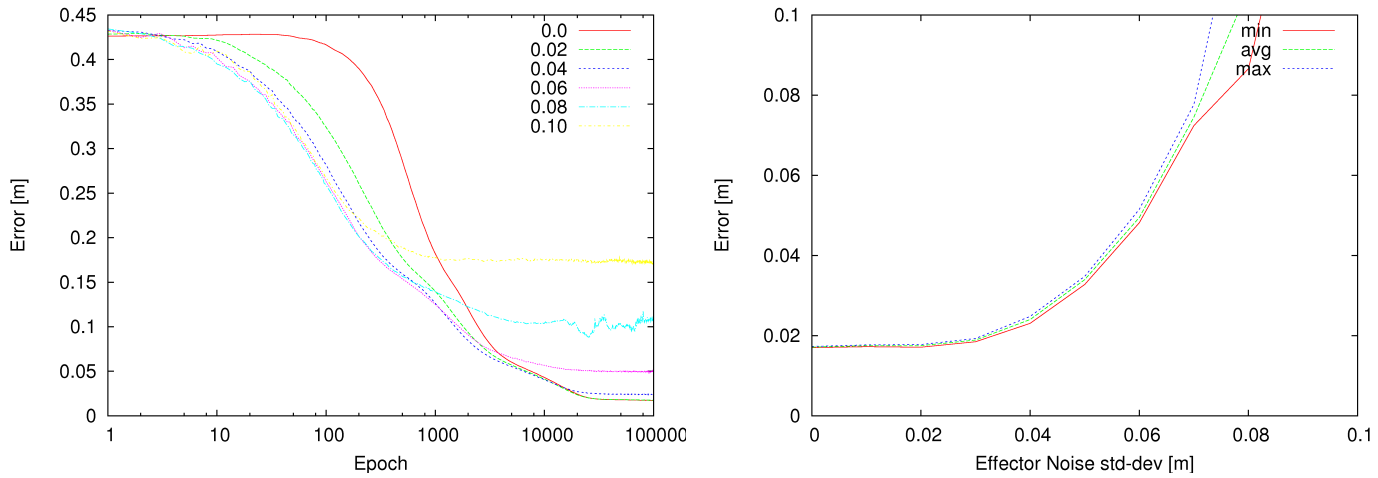


(a) Target positions x^* are shown as gray grid. The arm shows the home posture q^{home}.

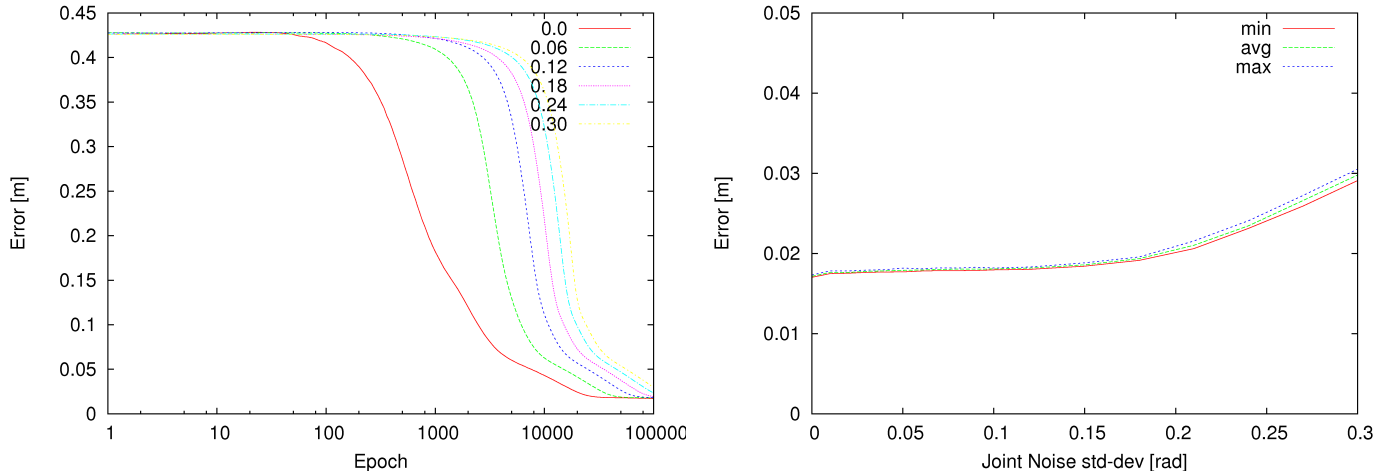


(b) The actually reached positions $f(g(x^*))$ are shown as black grid. Multiple postures $g(x^*)$ are overlaid to show how the redundancy is resolved.

Fig. 7: An inverse estimate for 2D position control of a planar 10 DOF arm generated with Goal Babbling. A third order polynomial was used as approximation model. The inverse estimate is very accurate as the reached positions are close to the target positions. The inverse estimate makes efficient use of all degrees of freedom.



(a) Results for Gaussian white noise with different standard deviations on the effector positions.



(b) Results for Gaussian white noise with different standard deviations on the joint angles.

Fig. 8: Performance of Goal Babbling with sensory noise over 100000 epochs for the planar arm for $n = 2$ and $m = 3$. The left plots show the performance error over time, averaged over 20 independent trials. The finally reached error is plotted against the varied parameter on the right side. The maximum, average and minimum error of 20 trials are shown.

However, the error is – in contrast to motor babbling – still small since the inconsistency resolution filters large portions of the generated examples.

An important question is how Goal Babbling scales with the degrees of freedom m. Results for up to 50 degrees of freedom are shown in Fig. 6b. For each value of m the arm was divided in segments of equal length, whereas we kept the arm length constant at 1m. For instance an arm with $m = 10$ comprises 10 segments with each 10cm length. We used $R = 0.2$ and $V = 20$ for exploration. The results show a reliable decrease of the performance error for all values of m and in all trials. Goal Babbling is systematically successful even for 50 degrees of freedom. An example solution $g(x^*)$ for $m = 10$ is shown in Fig. 7. The target positions are reached accurately.

So far, we evaluated the effector position directly with the analytic forward kinematics function $f(q)$ and assumed that the joint angles q can be applied with perfect accuracy. In contrast to a physical robot system this involves no noise. On a robot, the effector position might as well be measured with a stereo vision system. Thereby the analytic forward kinematic function would be fully replaced. In order to assess the influence of sensory noise, which is unavoidable in such systems, we added Gaussian white noise with different standard deviations to the effector positions x_t^v. This noise acts on the learner (Eqn. 6), but also affects the weight computation (Eqn. 1 and 2). Fig. 8a shows results for standard deviations ranging from 0cm up to 10cm (0.1m). The noise speeds up the initial bootstrapping significantly. In the first epochs, the effect of sensor noise on the effector positions is similar to a higher exploration range R: effector positions are observed, that are more distant to the home position $f(q^{home})$, causing a steeper learning gradient and accelerating the learning. Since such noisy examples do not reflect the true relation $f(q)$, very high amplitudes of noise cause a degeneration of the learning. An increase of the performance error is visible for standard deviations higher than 4cm. However, this amplitude is substantially higher than typical noise in a stereo vision system [21].

Sensory noise on the joint angles has a different effect. We applied Gaussian white noise, again with different standard deviations, to the joint angles q_t^v that are used for the weight computation and the learning. Fig. 8b shows results for standard deviations ranging from 0 radian up to 0.3 radian per joint. Joint noise slows down the initial bootstrapping. The final performance is very stable and the performance error increases only very slowly with increasing joint noise. We can conclude that Goal Babbling works reliably also with sensory noise.

IV. Discussion

We have presented an approach to bootstrap inverse kinematics for redundant systems without prior- or expert knowledge. We have shown theoretical insights about the structure of inconsistencies in goal-directed exploration [17], [18]. Based on that insights we have proposed a weighting scheme that resolves inconsistent solutions which occur in redundant systems with non-convex solution sets. To our knowledge this is the first successful approach of direct (example-based) learning that can solve the non-convexity problem. Moreover, it is the only successful approach to learn inverse kinematics exclusively from observable information. Methods based on the motor-error can in principle deal with redundant degrees of freedom, but the motor-error is not observable. Feedback-error learning [8], [9] assumes the prior knowledge about motor-errors. Learning with distal teacher [10], [11] relies on a complex mathematical derivation of the motor-error, which is neurally implausible [12]. In contrast, the information needed for Goal Babbling is fully observable – actually reached positions as well as movement directions and velocities.

Goal-directedness is essential for the success of autonomous motor learning. The comparison of intended movement directions with actually observed movement directions allows to detect and resolve one type on inconsistencies. Striving for optimal movement efficiency allows to resolve the other type of inconsistencies that can occur in goal-directed exploration. Optimality is necessary to learn correctly. The introduction of a "structured" noise in the simulation allows to find previously unreachable positions and better solutions in terms of efficiency, while maintaining the information structure that is necessary to resolve inconsistencies.

Goal Babbling is sufficient as exploration strategy to learn inverse kinematics. Forms of unstructured, not goal-directed motor-exploration (like motor babbling) are not only insufficient for redundant systems, they are even unnecessary. Admittedly, target effector positions are rather low-level goals. The important aspect, however, is the change of perspective: the exploration does not focus on the means of action (e.g. joint-angles), but on the action itself. Contrary to suggestions of distinct exploration mechanisms in infant motor development, exploration and control may be based on one mechanism.

What do infants "babble" in body babbling? Possibly goals instead of motor commands. Goal-directed action may not be the only form of exploration in infants. However, "learning by doing", or Goal Babbling can be successful in learning control from the very beginning.

Acknowledgements

Matthias Rolf gratefully acknowledges the financial support from Honda Research Institute Europe for the Project "Neural Learning of Flexible Full Body Motion".

References

[1] D. M. Wolpert, Z. Ghahramani, and J. R. Flanagan, "Perspectives and problems in motor learning," *Trends in Cog. Sci.*, vol. 5, no. 11, 2001.

[2] D. Wolpert, R. C. Miall, and M. Kawato, "Internal models in the cerebellum," *Trends in Cognitive Sciences*, vol. 2, no. 9, 1998.

[3] J. Piaget, *The Origin of Intelligence in the Child*, 1953.

[4] A. Meltzoff and M. Moore, "Explaining facial imitation: A theoretical model," *Early Development and Parenting*, vol. 6, pp. 179–192, 1997.

[5] C. von Hofsten, "An action perspective on motor development," *Trends in Cognitive Sciences*, vol. 8, no. 6, pp. 266–272, 2004.

[6] ——, "Eye-hand coordination in the newborn," *Developmental Psychology*, vol. 18, no. 3, pp. 450–461, 1982.

[7] L. Ronnquist and C. von Hofsten, "Neonatal finger and arm movements as determined by a social and an object context," *Early Development and Parenting*, vol. 3, no. 2, pp. 81–94, 1994.

[8] M. Kawato, "Feedback-error-learning neural network for supervised motor learning," in *Advanced Neural Computers*, R. Eckmiller, Ed. Elsevier, 1990.

[9] D. Wolpert and M. Kawato, "Multiple paired forward and inverse models for motor control," *Neural Networks*, pp. 1317–1329, 1998.

[10] M. Jordan and D. Rumelhart, "Forward models: supervised learning with distal teacher," *Cognitive Science*, vol. 16, pp. 307–354, 1992.

[11] M. I. Jordan, "Computational aspects of motor control and motor learning," in *Handbook of Perception and Action: Motor Skills*. Academic Press, 1996.

[12] J. Porrill, P. Dean, and J. V. Stone, "Recurrent cerebellar architecture solves the motor-error problem," *Proc Biol Sci*, vol. 271, no. 1541, pp. 789–796, 2004.

[13] Y. Demiris and A. Dearden, "From motor babbling to hierarchical learning by imitation: A robot developmental pathway," in *Proceedings of the fifth international workshop on epigenetic robots (EPIROB): modeling cognitive development in robotic systems*, 2005, pp. 31–37.

[14] Y. Demiris and A. Meltzoff, "The robot in the crib: A developmental analysis of imitation skills in infants and robots," *Infant Child Development*, vol. 17, no. 1, pp. 43–53, 2008.

[15] P. Gaudiano and D. Bullock, "Vector associative maps unsupervised real-time error-based learning and control of movement trajectoires," *Neural networks*, vol. 4, no. 2, pp. 147–183, 1991.

[16] D. Bullock, S. Grossberg, and F. H. Guenther, "A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm," *Journal of Cognitive Neuroscience*, vol. 5, no. 4, pp. 408–435, 1993.

[17] E. Oyama and T. M. S. Tachi, "Goal-directed property of on-line direct inverse modeling," in *IEEE Int. Jnt. Conf. on Neural Networks*, 2000.

[18] T. D. Sanger, "Failure of motor learning for large initial errors," *Neural Computation*, vol. 16, no. 9, pp. 1873–1886, 2004.

[19] M. Rolf, J. J. Steil, and M. Gienger, "Efficient exploration and learning of full body kinematics," in *ICDL*, 2009.

[20] T. Poggio and F. Girosi, "Networks for approximation and learning," in *IEEE*, vol. 78, no. 9, 1990, pp. 1481–1497.

[21] G. D. Hager, W.-C. Chang, and A. S. Morse, "Robot hand-eye coordination based on stereo vision," *IEEE Control Systems Magazine*, vol. 15, no. 1, 1995.