Goal Babbling Permits Direct Learning of Inverse Kinematics

Matthias Rolf, Jochen J. Steil, and 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. The essential novelty lies in a path-based sampling approach: we generate training data along paths, which result from execution of the currently learned estimate along a desired path towards a goal. The information structure thereby induced enables an efficient detection and resolution of inconsistent samples solely from directly observable data. We derive and illustrate the exploration and learning process with a low-dimensional kinematic example that provides direct insight into the bootstrapping process. We further show that the method scales for high dimensional problems, such as the Honda humanoid robot or hyperredundant planar arms with up to 50 degrees of freedom.**

***Index Terms*—Goal babbling, inverse kinematics, motor exploration, motor learning.**

I. Introduction

LEARNING to control our own body is a fundamental problem in human development. In early childhood, infants need to learn the most basic skills like reaching for an object. The ability to learn control from scratch also allows us to master the change induced by body growth and to learn 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. Once internal models are established for a certain task, 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 initially uncoordinated behavior? Before internal models can be applied for coordinated control, experience must be gained by exploration. The crucial question is how to acquire that experience, i.e., how infants explore their bodies for coordination. Piaget suggested that human– (motor) development progresses in several stages [3]. At first, infants react purely reflexive. From an age of six weeks to approximately four months, the development is characterized by primary circular reactions. Infants try to reproduce observations that initially only occur by chance. Infants repeat those actions over and over again. At the age of eight months, infants then intentionally reach for objects. This finding inspired Meltzoff and Moore [4] to derive the concept of "body babbling," related to the vocal babbling [5] of young infants. They describe body babbling as an initial stage in which experience is gathered. Infants then use this experience to attempt goal-directed action and fine-tune their skills on the fly. Similar to Piaget's work, a conceptual difference is introduced between exploration (gathering experience) and control (application of experience).

Contrary to Piaget's suggestions, evidence over the last decades clearly shows that infants perform goal-directed movements from the very beginning. For instance, von Hofsten 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," [6]. Statistics revealed that already days after birth, infants attempt goal-directed action by means of arm and finger movements [7], [8]. Even behaviors that were previously regarded as reflexes have been rediscovered as goal-directed actions [9], [10]. These findings of early goal-directed actions clearly suggest that "learning by doing" plays a central role in infant motor development. Infants learn to reach by trying to reach. Whether learning by doing—or rather exploration by trying to do—is a sufficient exploration strategy or other strategies are needed as well is, however, not clear.

A. The Learning Problem

Before infants, but also robots, can master deliberate reaching, inverse models must be learned for their limbs or even the full body. All joints must be coordinated in order to move the hand. 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 $f(q) = x$ describes the causal and uniquely determined relation between both sizes. It cannot be used directly for coordination and control, because to position the hand at some desired target x^*, an inversion mechanism is needed to find appropriate joint angles q that apply the desired position $(f(q) = x^*)$.

An inverse function, however, is not uniquely defined if the number of joint angles m exceeds the number of controlled dimensions n. Even for $n = m$, there are typically multiple so-

Manuscript received February 19, 2010; revised April 27, 2010 and July 07, 2010; accepted July 17, 2010. Date of publication August 03, 2010; date of current version September 10, 2010. This work was supported in part by the Honda Research Institute Europe Project "Neural Learning of Flexible Full Body Motion."

M. Rolf and J. J. Steil are with the Research Institute for Cognition and Robotics (CoR-Lab), Bielefeld University, Bielefeld 33611, Germany (e-mail: mrolf@CoR-Lab.Uni-Bielefeld.de; jsteil@CoR-Lab.Uni-Bielefeld.de).

M. Gienger is with the Honda Research Institute Europe, Offenbach 63073, Germany (e-mail: michael.gienger@honda-ri.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAMD.2010.2062511

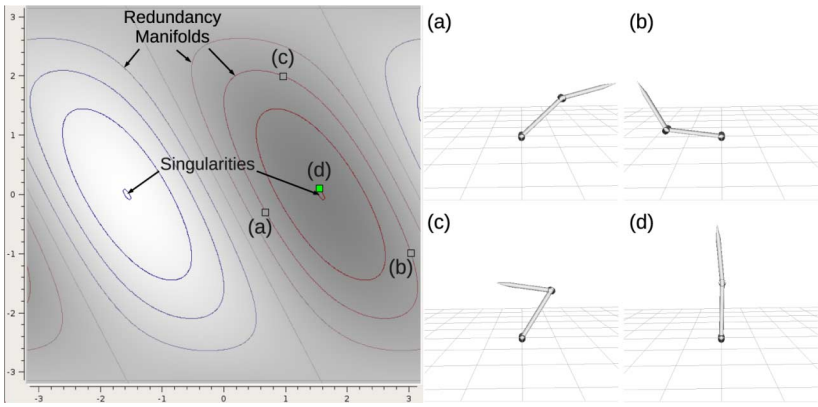


Fig. 1. Robot arm (length 1 m) with two joints. The left display shows the joint space. The bottom axis encodes the angle of the first joint between $-\pi$ and $+\pi$ radiant and the left-hand axis encodes the angle of the second joint with the same range. Nonconvex sets of configurations [see, e.g., postures (a)–(c)] can be used to reach the same height of the end effector and are marked by colored contours in the joint space. Multiple configurations that apply the same height [e.g., (a)–(c)] must not be averaged, because the average may result in a different height [see posture (d), the average of (a)–(c)].

lutions q for a target x^*. If there are more degrees of freedom $(n < m)$, an infinite number of joint angles exists for the same target. Several learning schemes to find appropriate joint angles have been proposed including feedback-based learning schemes (e.g., [11]) which resemble Jacobian-based controllers [12], associative procedures (e.g., [13]) as well as feed-forward-based schemes (e.g., [14]). We focus on the fundamental task of learning a single inverse function/model $g(x^*) = q$ that returns joint angles for a given target such that $f(g(x^*)) = x^*$. Evidence for the relevance of this task comes, for instance, from prism-glass experiments on human learning of new sensorimotor maps [15]. The direct inverse function $g(x^*)$ here, selects exactly one of these joint angles, which describes a developmentally plausible path by first learning one valid solution before trying to remember all solutions. Despite being seemingly simple in terms of control, the earlier proposed methods for the learning of a direct inverse function $g(x^*)$ are either unplausible from a developmental point of view or fail in the case of redundancy. It is the aim of this paper to provide a method that both can deal with redundancy and is developmentally plausible.

A minimal example of redundant control is shown in Fig. 1: a robot arm with two joints $(q = (q_{(1)}, q_{(2)}), m = 2)$ and a total length of 1 m is controlled to achieve a certain height of the effector $(n = 1)$. Left/right movements of the effector are ignored in this example. The redundancy appears in form of manifolds through the 2-DOF joint-space, on which all joint angles apply the same effector height. Some of these manifold are visualized by colored contours (see Fig. 1).

The geometry of the arm defines the forward kinematics function $f(q)$ as

$$f(q) = 0.5 \cdot \sin(q_{(1)}) + 0.5 \cdot \sin(q_{(1)} + q_{(2)}). \qquad (1)$$

An inverse kinematics function in this example must return 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. Figs. 2(a) and 2(b) show two examples. For several target heights x^*, the joint angle estimates are shown by colored markers on the manifold and visualized by corresponding postures in the 3-D simulation. Small green markers show the examples used for learning. An accurate inverse estimate positions all markers on the contour with the same color.

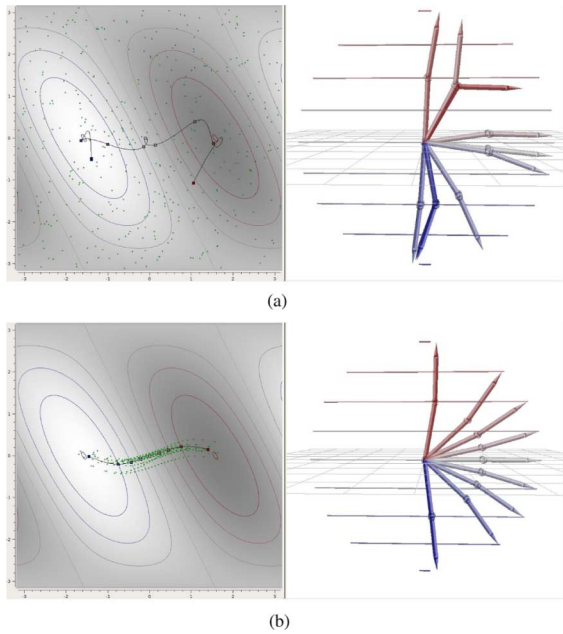


Fig. 2. Direct inverse learning with random motor babbling as exploration process does not yield a correct inverse estimate because of the nonconvex redundancy manifolds. Goal babbling finds an accurate solution. The small green crosses show the position of the example configurations used for learning. (a) Motor Babbling and (b) Goal Babbling.

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^*. In the case of forward kinematics, or generally forward modeling, the correct outcome x for a model input q can simply be probed by applying q (the cause) and observing x. This probing is not possible for inverse problems, where a cause is searched for a desired outcome.
2) Nonconvexity. It is also difficult to deal with the presence of *multiple* solutions. The sets of solutions in redundant systems typically form nonconvex sets (see Fig. 1).

Learning algorithms that average between multiple correct solutions consequently fail [14].

Existing approaches to the exploration and learning of inverse kinematics split into two groups: error-based and example-based methods.

Error-Based Learning: 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 $x^* \in \mathbf{X}^* \subseteq \mathbf{X}$. Using the joint angles $q = g(x^*)$ returned by the inverse estimate, the resulting position x of the effector is evaluated with the forward kinematics function $x = f(q)$, and will generally differ from the target position x^*. Error-based approaches then aim at improving the inverse estimate at the target position x^*. One group of mechanisms is based on the "motor error." The motor error is a correction Δq of the joint angles that is added to the estimated joint angles in order to improve the performance. If such a value is available, it can directly be used to improve the inverse estimate. In *feedback-error learning* [16], [17], it is simply assumed that a mechanism to compute that motor error is already available. In *learning with distal teacher* [14], [18], an estimated forward model $\hat{f}(q)$ is used for learning. A motor error can be derived analytically by differentiating the forward model. The forward model must be pretrained with an exhaustive, nongoal-directed exploration of the joint space [14], which is very inefficient for many degrees of freedom. Both methods can, in principle, deal with redundant systems. The critical problem is that the motor error is not directly observable. The prior existence of a module for computing the motor error is not plausible for problems that exceed the control of a single muscle. Neither is the analytic differentiation of a forward-model.

A special case of error-based learning has been developed in [19] and [20]. The error in the effector space $x - x^*$ is used directly for learning. The idea is to correct mistakes *a priori* by shifting the target positions. The information used in this case is fully observable, but the method has not been shown to work for redundant degrees of freedom $(n < m)$, and requires a rather accurate inverse estimate in advance.

Example-Based Learning: Example-based methods use example configurations $(x, q) = (f(q), q)$ for the learning of an inverse estimate $g(x)$. This kind of learning has also been named *direct learning* of inverse kinematics. The existing approaches differ in the way how such examples are generated. Motor babbling [21], [22] 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 [23], [24]. Joint angles are randomly chosen from the set of all possible configurations $q_k \in \mathbf{Q}$, and the outcome $x_k \in \mathbf{X}$ is observed. This approach can solve the inversion of causality, if enough examples are generated such that q will come close to any desired (and possible) effector pose x^*. However, it is subject to the nonconvexity problem and the curse of dimensionality. An outcome of motor-babbling for inverse kinematics learning is shown in Fig. 2(a).

A few goal-directed exploration approaches have also been investigated. Experience is generated with an initially chosen inverse estimate $g(x^*)$ that is used for trying to reach for target positions x^*. Using the joint angles $q = g(x^*)$ computed by the inverse estimate, the resulting effector pose $x = f(q)$ is evaluated. Samples (x_k, q_k) are generated for several target positions x_k^* and the inverse estimate is iteratively updated with those samples. It was shown that such exploration and learning processes can be successful for discrete redundancies $(n = m)$ if a "good enough" initial estimate is available [25]. In general, no success can be guaranteed even if the system is not redundant [26]. As the entire exploration depends on the inverse estimate, the success depends on the extrapolation of the used function approximation method. If an outcome x is observed, the model is not necessarily improved at the target position and the same mistake might be repeated. The approach—as previously discussed in literature—is unable to invert causality in a reliable fashion. Moreover, also the goal-directed approach is subject to the nonconvexity problem.

Example-based learning of inverse kinematics has only been shown to be successful if training data (x_k, q_k) without inconsistent solutions is already available [27]. This requires an expert to generate such training data—either another controller that moves the robot or, e.g., a human caregiver by kinesthetic teaching. 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: 1) repeated process of; 2) trying to accomplish; and 3) 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 the differences to previous concepts. The exploration process focuses on the goals of action instead of the means. The emphasis in this approach is on "trying to accomplish," which means to generate paths towards the given goal with the currently learned system and to evaluate samples along this path. It turns out that this path-based approach remedies the major flaw all previously proposed methods share: they all consider samples in isolation. We show that we can exploit additional information provided by the fact that the executed motion, and therefore, the samples are continuous along the paths generated by "trying to reach" for goals. In the present work, we use a random selection of target positions, but which may ultimately be selected by a higher cognitive mechanism. Intimately related to the original concept of vocal- as well as body-babbling, repetition is important. A goal must be tried to be accomplished again and again in order to succeed.

In the remainder of this paper, we introduce and evaluate a computational model for goal babbling, inspired by the findings of goal-directed action in infants. We show that the developmentally plausible method is a successful bootstrapping strategy for the inverse kinematics of redundant systems.

A. Goal-Directed Exploration

As starting point, we introduce goal-directed exploration as in [25] and [26]. 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 $g(x, \theta)$, e.g., 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 set of target positions is chosen: $x_k^* \in \mathbf{X}^* \subseteq \mathbf{X}, k = 1 \dots K$. The inverse estimate is then used to acquire respective training data: $q_k = g(x_k^*), x_k = f(q_k)$. The set of K generated examples is denoted as

$$D = \{(f(g(x_k^*, \theta)), g(x_k^*))\}_k. \tag{2}$$

Note that this set has no particular predictable structure in the joint space, because x_k^* were chosen arbitrarily. The parameters θ of the inverse estimate are then updated to minimize the *command error* [26]

$$E^Q(\theta) = \sum_k (g(x_k, \theta) - q_k)^2.$$

After the adaption of the parameters, the process is repeated. We will refer to this method as *plain goal-directed exploration*. The overall goal of learning inverse kinematics is to minimize the *performance error*

$$E^X(\theta) = \sum_k (f(g(x_k^*, \theta)) - x_k^*)^2. \tag{3}$$

Plain goal-directed exploration does not necessarily reduce the performance error. It fails to invert causality in a reliable way. For instance, if the output of the inverse estimate is constant $g(x^*) = c$, only one effector pose $f(c)$ will be observed. The command error $E^Q(\theta)$ is zero in this case, since $g(f(c)) = c$ is already achieved. The performance error is not zero if other goal positions than $f(c)$ exist. A further problem is that the inverse estimate is unstable in the nullspace of movement, i.e., along its orthogonal direction. Any drift of the inverse estimate in that direction may self-reinforce in the next step and cause the inverse estimate to drift away. An example of the learning dynamics is shown in Fig. 3. The inverse estimate drifts away into the upper regions of the joint-space visualization. Finally, inconsistent examples can also exist under goal-directed exploration.

B. Path-Based Inconsistency Detection and Resolution

Two samples (x_A, q_A) and (x_B, q_B) are inconsistent, if they represent the same effector pose $x_A = x_B$, but different joint angles $q_A \neq q_B$. Regardless of the kind of exploration that is used to generate samples, two samples with exact same effector pose will rarely be found. Resolving inconsistencies solely based on the samples is therefore hardly possible. We argue that inconsistency resolution becomes feasible if we take into account the sample generation method itself, which is not the case in plain goal-directed exploration.

Structure of Inconsistencies: For redundant systems, inconsistent configurations $q_A, q_B \in \mathbf{Q}^{\text{expl}}$ generally exist, where $\mathbf{Q}^{\text{expl}}$ is the set of possibly generated joint configurations. To gain further insight, we assume that two inconsistent samples $q_A \neq q_B, x_A = x_B$ are generated in the goal-directed exploration ($q_A, q_B \in \mathbf{Q}^{\text{expl}}$). These samples must originate from two different target positions $x_A^* \neq x_B^*$ (for identical target positions $x_A^* = x_B^*$, it follows that $q_A = q_B$ because $g(x_A^*) = g(x_B^*)$)).

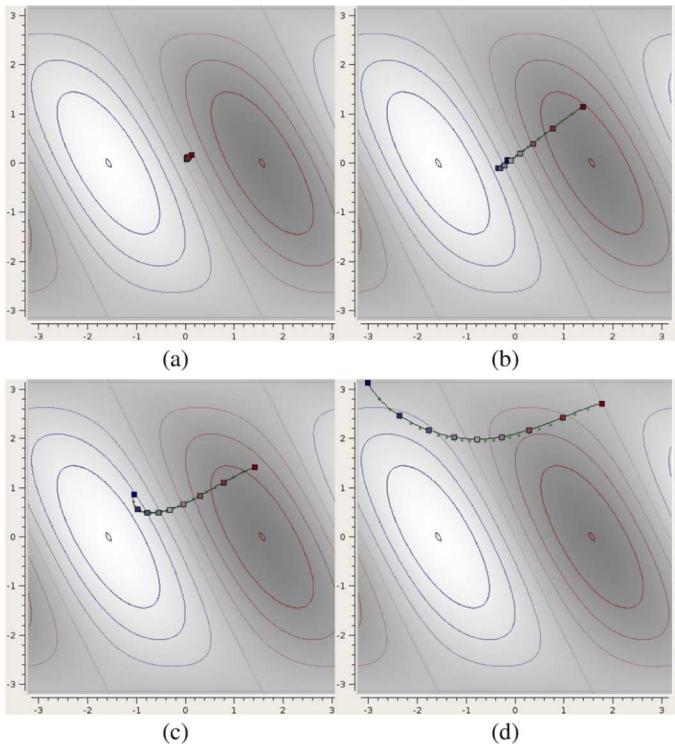


Fig. 3. Learning dynamics with plain goal-directed exploration from (a) to (d). Learning occurs from inconsistencies, as the controlled manifold intersects some redundancy manifolds multiple times. The estimate drifts in its orthogonal direction, where no training data is available.

We now perform the crucial step in our analysis, which motivates to evaluate samples along paths. We use the inverse estimate to attempt a linear target motion between x_A^* and x_B^* [see Fig. 4, (left)], i.e., perform "trying to reach" for x_B^*. The system starts from the joint configuration q_A, corresponding to x_A^*, moves its joints along some path and ends up in joint configuration q_B. At the beginning and end of the movement, the effector has the same pose $x_A = x_B$. When the effector is observed while trying to follow that straight path, two cases can occur.

1) An effector motion occurs while using the inverse estimate $g(x^*)$ to follow a linear target motion between x_A^* and x_B^*. Since the effector returns to the same position, the observed effector movement must have a closed shape [see Fig. 4, (right)]. The goal is to follow a straight line, i.e., to keep the movement direction constant, but the observed movement direction changes.
2) The effector pose remains constant, in spite of the joint movement from q_A to q_B. This case can occur when the inverse estimate moves exactly along one redundancy manifold. This case is characterized by a minimum of movement efficiency as defined in (5). While the joints are moved, the effect on the effector is zero.

We conclude that along a path between two target points, inconsistencies occur only if either: 1) unintended changes of movement direction; or 2) movements with zero efficiency are present, which both can be detected from observation of the movement. For goal babbling, we consequently propose to sample data along paths between the target positions $x_k^\star$,

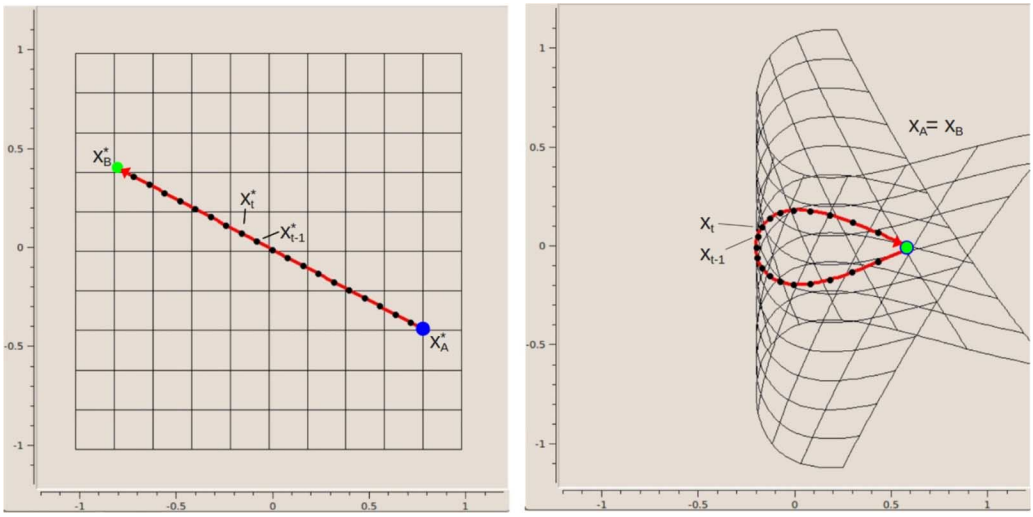


Fig. 4. (Left) Space of target positions x^*. A linear target motion shall be produced between two targets. (Right) Space of results $x = f(g(x^*))$. An inconsistency occurs, e.g., when the grid is folded. The formerly straight line now has a circular shape.

e.g., using L intermediate samples between $x_k^\star$ and $x_{k+1}^\star$. Data are therefore connected on a n-dimensional manifold inside the m dimensional joint-space that is defined by the inverse estimate. Along this manifold, we obtain $K \cdot L$ temporally ordered targets $x_t^\star$ and data points $x_t, t = 1 \dots K \cdot L$, where $x_t, t = kL, k = 1 \dots K$ are the samples corresponding to the original targets $x_k^\star$.

Inconsistency Resolution: We now provide a mechanism to exclude inconsistent samples along the generated paths using the insights from the last section. We assign weights $w_t \in \mathbb{R}$ for each example (x_t, q_t), and take into account the temporal order of the examples. Unintended changes of movement direction can be tackled with a scheme that bases upon a special case: If the observed movement direction never deviates by 90° or more from the intended movement direction, circular shapes as shown in Fig. 4 can not occur. We utilize this fact in the following weighting scheme:

$$w_t^{\text{dir}} = \frac{1}{2}(1 + \cos \sphericalangle(x_t^* - x_{t-1}^*,\ x_t - x_{t-1})). \tag{4}$$

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 $w_t^{\text{dir}} = 1.0$. If the observed movement has the exact opposite direction, the angle is 180.0° and the weight becomes $w_t^{\text{dir}} = 0.0$. 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. If the estimate $g(x^*)$ is rather accurate and the intended movement direction can always be realized, all samples receive full weight 1.0.

The second case of an inconsistency (low movement efficiency) can be resolved by weighting with the ratio of effector motion and joint motion, which becomes 0.0 if the joints move without effector motion

$$w_t^{\text{eff}} = ||x_t - x_{t-1}|| \cdot ||q_t - q_{t-1}||^{-1}. \tag{5}$$

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

$$w_t = w_t^{\text{dir}} \cdot w_t^{\text{eff}}. \tag{6}$$

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. The path-based exploration generates an n-dimensional manifold within the joint space, where the information about continuity along this manifold allows for evaluation of the movement directions. It is this very information structure that allows for a resolution of inconsistencies and distinguishes our scheme from all previous ones. The rules are local in space and time, since only the immediate temporal and spatial context is considered. Therefore, both rules are imperfect, since only one movement direction can be observed at a time. However, we show experimentally that the rules are sufficient to resolve inconsistencies.

C. Structured Variation for Efficient Exploration

The proposed resolution of inconsistencies is not yet sufficient to find an accurate inverse estimate, since it does not solve the inversion of causality. Again we use a developmentally plausible principle: 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 erratic movements in the first place and rather cause smooth deviations when a goal-directed movement is attempted. We simulate this at the kinematics level by adding a small disturbance term $E^v(x)$ to the inverse estimate

$$g^v(x) = g(x) + E^v(x). \tag{7}$$

Examples are then generated with this variation instead of the actual inverse estimate$q_t^v = g^v(x_t^*)$,, $x_t^v = f(q_t^v)$. We denote the set of examples generated for a variation v as

$$D^v = \{(f(g^v(x_t^*)),\ g^v(x_t^*))\}_t. \tag{8}$$

The assumptions and arguments for the inconsistency resolution still hold, since $g^v(x)$ is again a function and spans a n dimensional manifold in the joint-space along the respective path. For a given set of examples D^v, the weighting scheme can be applied as proposed above. The index v is added to identify weights for examples of a specific variation

$$w_t^{v^{\text{dir}}} = \frac{1}{2}\left(1 + \cos \sphericalangle \left(x_t^* - x_{t-1}^*,\ x_t^v - x_{t-1}^v\right)\right) \tag{9}$$

$$w_t^{v^{\text{eff}}} = \frac{\left\|x_t^v - x_{t-1}^v\right\|}{\left\|q_t^v - q_{t-1}^v\right\|} \tag{10}$$

$$w_t^v = w_t^{v^{\text{dir}}} \cdot w_t^{v^{\text{eff}}}. \tag{11}$$

The Home Posture: 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 $f(q^{\text{home}})$ can be observed and used for learning as any other example. We add the example $q_0^v = q^{\text{home}}, x_0^v = f(q^{\text{home}}) = x^{\text{home}}$ to each set generated with goal-directed exploration

$$D^v \longleftarrow \{(f(q^{\text{home}}), q^{\text{home}})\} \cup D^v. \tag{12}$$

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: $g(x^{\text{home}}) \approx q^{\text{home}}$. 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.

Algorithm 1: Goal Babbling Pseudocode

Require: Forward kinematics: $f(q)$

Require: Set of target positions: $\mathbf{X}^*$

Initialize learner: $\theta \leftarrow \theta_0, g(x^*) \leftarrow g(x^*, \theta)$

for Number of epochs **do**

Select target sequence from $\mathbf{X}^*$: $x_t^*,\ t = 1 \ldots T$

$D \leftarrow \varnothing$

for $v = 1 \ldots V$ **do**

Select disturbance term: $E^v(x^*)$

Get variation: $g^v(x^*) = g(x^*) + E^v(x^*)$

Generate examples: $D^v \leftarrow \{(f(g^v(x_t^*)),\ g^v(x_t^*))\}_t$

Compute weights w_t^v ((9), (10) and (11))

Add home posture: $D^v \leftarrow D^v \cup (f(q^{\text{home}}), q^{\text{home}})$

$D \leftarrow D \bigcup D^v$

end for

Reduce error $E_w^Q(\theta)$ on D using gradient descent

end for

D. Learning

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

$$D = \bigcup_v D^v = \bigcup_v \{(\ f(g^v(x_t^*)),\ g^v(x_t^*))\}_{t=0 \ldots T}. \tag{13}$$

The multiple variations allow to discover new poses by chance which solves the problem of plain goal-directed exploration to reliably invert causality. Furthermore, it ultimately solves the instability problem of goal-directed exploration if the number of variation V exceeds the joint dimension m, since all directions in the joint-space are locally covered.

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

$$E_w^Q(\theta) = \sum_v \sum_t w_t^v \cdot \left(g\left(x_t^v, \theta\right) - q_t^v\right)^2. \tag{14}$$

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 use a random initialization 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 or the parameters of the kinematic function. Within one epoch, examples are generated from multiple variations, weights are assigned, and the learning is performed with the examples. The next epoch repeats the procedure with the updated inverse estimate. The entire procedure is also detailed in Algorithm 1.

The introduction of multiple variations in the exploration locally adds multiple solutions. However, if the disturbance terms $E^v(x)$ have numerically small values, these solutions are located in a small region in the joint space. Therefore, the error induced by the nonconvexity problem is generally very small and can safely be neglected. The weighting based on intended movement directions prevents learning from significantly inconsistent examples. The efficiency weighting allows to "select" examples generated by different variations. Good solutions in terms of the movement efficiency criterion [see (5)] will dominate the learning and cause the inverse estimate to be aligned along such optimal joint configurations. The averaging is therefore constructive (compared to the destructive averaging

in motor babbling) which is only possible due to the combination of variation *and* weighting. Striving for such optimal movement efficiency is not a luxury. It is necessary to resolve inconsistent solutions and guide the exploration systematically towards new targets.

E. Example

An example of inverse kinematics learning with goal babbling for the minimal 2-DOF problem [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 $q^{\text{home}} = (0.0, 0.0)$. The next images show the progress of the method after several epochs. Each image shows the current inverse estimate together with the currently generated example data. The aim is to control the effector's height within the full range from -1.0 m to 1.0 m. Initially, only heights around $f(q^{\text{home}}) = 0$ m are reachable. Target positions between the extremes -1.0 m and 1.0 m are tried to reach from the very beginning, although these attempts are not successful at first.

Three qualitative stages can be observed in the progress of bootstrapping the inverse kinematics. These stages are not preprogrammed, but they arise naturally from the learning dynamics. In the *first stage* (orientation), the manifold spanned by the inverse estimate remains close to the home posture. Only a small set of effector poses x can be observed, such that the command error E^Q is rather small (similar to the case of a constant function $g(x^*)$). Triggered by the exploration of variations, the inverse estimate starts to align with the correct movement directions and for optimal movement efficiency. Thereby the weights of the examples slowly increase.

Once the inverse estimate is aligned with optimal directions, the *second stage* (expansion) can be observed. The extrapolation of the inverse estimate causes a rapid expansion of the inverse estimate in the joint space. This stage is characterized by a rapid decrease of the performance error. Due to the efficiency weighting, the expansion directly follows the steepest, most efficient direction. The inverse estimate is aligned nearly orthogonal to the redundancy manifolds.

The expansion saturates when the ridge of the forward kinematics is hit. More expansion would not discover new effector positions, but only introduce more inconsistencies since the same redundancy manifolds would be crossed again. Samples generated beyond the ridge are, however, filtered out by the weighting of correct movement directions (9). Then the *third stage* (tuning) can be observed. The inverse estimate finds the nonlinearities that are necessary to reach for the extreme positions and to further optimize the movement efficiency. The performance error decreases slowly until it converges.

F. Influence of the Home Posture

The home posture is an open parameter of the exploration procedure, which can be used to shape the inverse estimate. Goal babbling works robustly for a wide range of home postures. An example of a different home posture is shown in Fig. 6(a). The inverse estimate aligns with the optimal efficient movement direction with respect to the home posture, which acts as origin. Goal babbling can even be successful, if the home posture is placed in a singularity, as shown in Fig. 6(b). In that case, multiple ways exist to leave the singularity with optimal movement efficiency (two in the example). This symmetry is broken by the randomized exploration of variations. The learning can get stuck if the inverse estimate hits the joint limits, such that a further local improvement of the inverse estimate is not possible, see Fig. 6(b). Goal babbling shares this problem with feedback-error learning and learning with distal teacher. All three approaches operate iteratively, based on local improvements. Although motor-error-based methods do not make explicit use of a home posture, they require an initial placement of the inverse estimate and cannot proceed if local improvements are not possible. Home postures that cause such ill-posedness are, however, biologically not plausible for systems that need to bootstrap their motor repertoire. Also, they are easy to avoid in en-

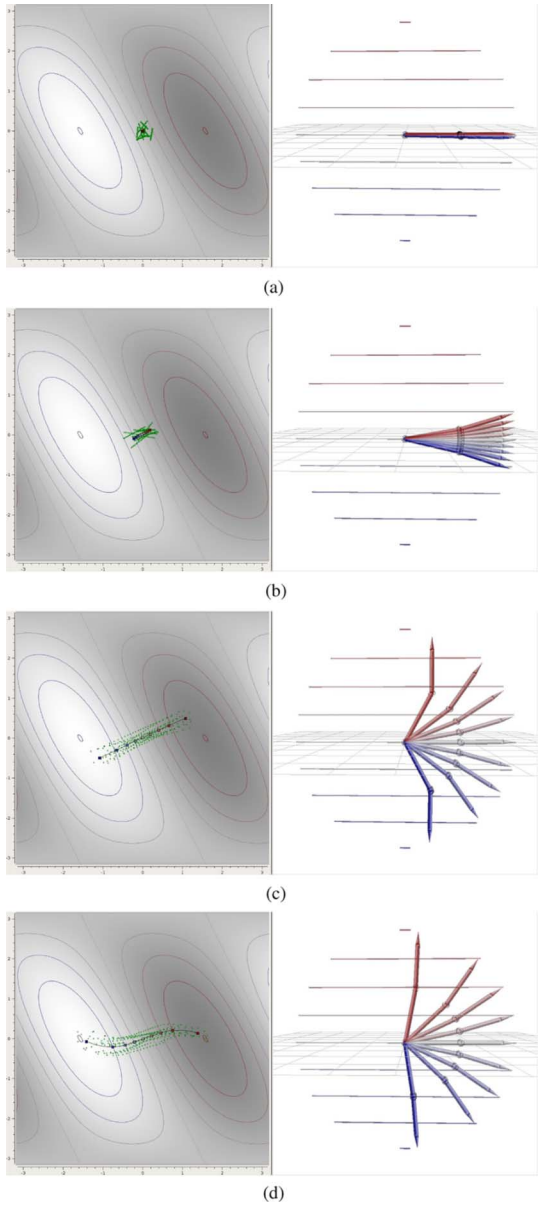
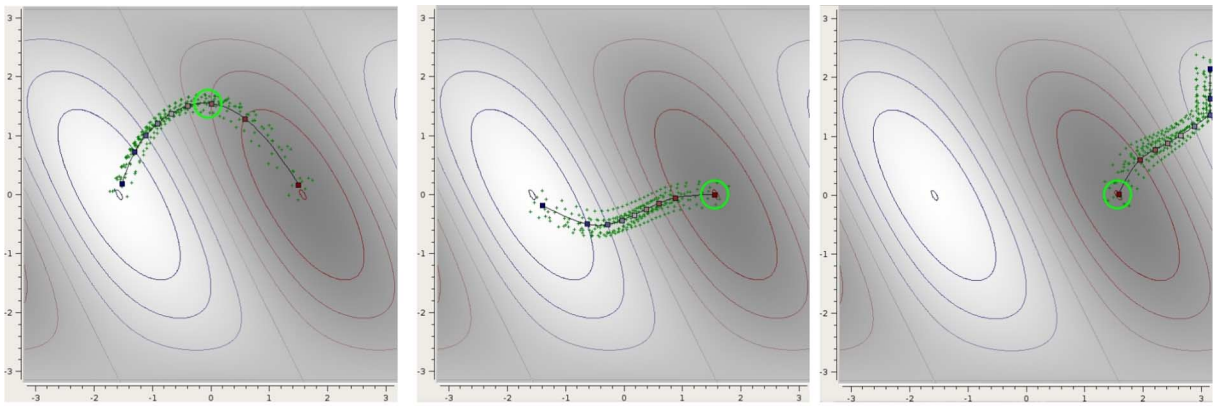


Fig. 5. Inverse kinematics learning with goal babbling. The images show successive stages of the learning process. The inverse estimate is initialized around zero in joint space. It unfolds successively and finally reaches an accurate solution. (a) The inverse estimate is initialized around the home posture. (b) Orientation: the inverse estimate has aligned with the steepest direction. (c) Expansion: the performance error decreases rapidly. (d) Tuning: the inverse estimate finds the necessary nonlinearities to reach for extreme positions.



(a) (b)

Fig. 6. The inverse estimate can be shaped by the choice of the home posture, which is shown as green circle. Learning is still possible from a singularity as start point. However, learning can no longer proceed if the joint limits are hit. (a) Outcome for $q^{\text{home}} = (0.0, (\pi/2))$. (b) Two possible outcomes for $q^{\text{home}} = ((\pi/2), 0.0)$, which is a singularity of the forward kinematics.

gineering by choosing a position in the center of the joint space and nearby the target positions that are tried to reach.

III. Experiments

In this section, we show results of goal babbling for different robot morphologies. We start by extending the simple 2-DOF arm (see Fig. 1) by more degrees of freedom and finish with goal babbling on the Honda humanoid research robot.

A. General Setup

In all experiments, we use polynomial regression [28] 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. For a two dimensional input vector, $x = (x_{(1)}, x_{(2)})$ and a polynomial degree $P = 2, \Phi^P(x)$ calculates the terms $(1.0, x_{(1)}, x_{(2)}, x_{(1)}^2, x_{(1)} \cdot x_{(2)}, x_{(2)}^2)^T \in \mathbb{R}^6$. A linear regression with parameters $\theta = \mathbf{M}$ operates on these features

$$g(x^*, \mathbf{M}) = \mathbf{M} \cdot \Phi^P(x^*), \quad \mathbf{M} \in \mathbb{R}^{p \times m}. \tag{15}$$

The entries of the regression matrix $\mathbf{M}$ are adapted by gradient descent in the weighted command error as defined in (14). We use a learning rate of 0.2. Before exploration and learning proceed, we initialize $\mathbf{M}$ to zero and perform random adaptions such that $g(x^*, \mathbf{M})$ produces joint angles in a range of 0.1 radian around the home posture.

For the exploration we use linear disturbance terms

$$E^v(x) = \mathbf{A} \cdot x + b, \quad \mathbf{A} \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m. \tag{16}$$

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 $\mathbf{X}^*$

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

B. Planar Arm: 1-D Control

We start our experiments with the simulated robot arm in Fig. 1. The arm with initially two degrees of freedom ($m = 2$) is used to control only the height of the end effector ($n = 1$). If only one dimension is controlled, one linear target motion is enough to cover the whole space of targets. The target motion x_t^* is a linear sequence from $x_1^* = -1.0$ to $x_T^* = 1.0$ with $T = 25$ steps in all epochs. The home posture is $q^{\text{home}} = \vec{0}$ such that the arm is stretched and at height 0.0.

We first investigate the behavior of the exploration range R. Fig. 7(a) shows results for R varying between 0.05 and 1.0 radian over 10 000 epochs and for 20 independent trials. The number of variations was set to $V = 20$, and we used third order polynomials ($P = 3$). The left plot shows the performance error [see (3)] over time for different values of R. The error decreases continuously. The qualitative stages orientation, expansion, and tuning can be identified in all curves, where the expansion shows a rapid decrease of the performance error. High values like $R = 1.0$ display the fastest convergence, but remain at a higher absolute error. The right plot shows the final error reached after 10 000 epochs. For $R = 0.05$, not all inverse estimates are converged after that time, depending on the initialization. For $R = 0.15$ or higher, all trials have converged and show a very low performance error from –1 to 2 cm with an arm length of 1 m. 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. However, the error is—in contrast to motor babbling [see Fig. 2(a)]—still small since the inconsistency resolution filters large portions of the generated examples. Although the speed varies, the general success of goal babbling is rather insensitive to the concrete exploration range.

Fig. 7(b) shows the same setup, but the exploration range is fixed to $R = 0.2$, and the polynomial degree P is varied. The temporal characteristics of the performance error do not differ significantly for different polynomial degrees. Higher polynomial degrees allow a more accurate approximation of the examples. While first and second order polynomials do not yield a

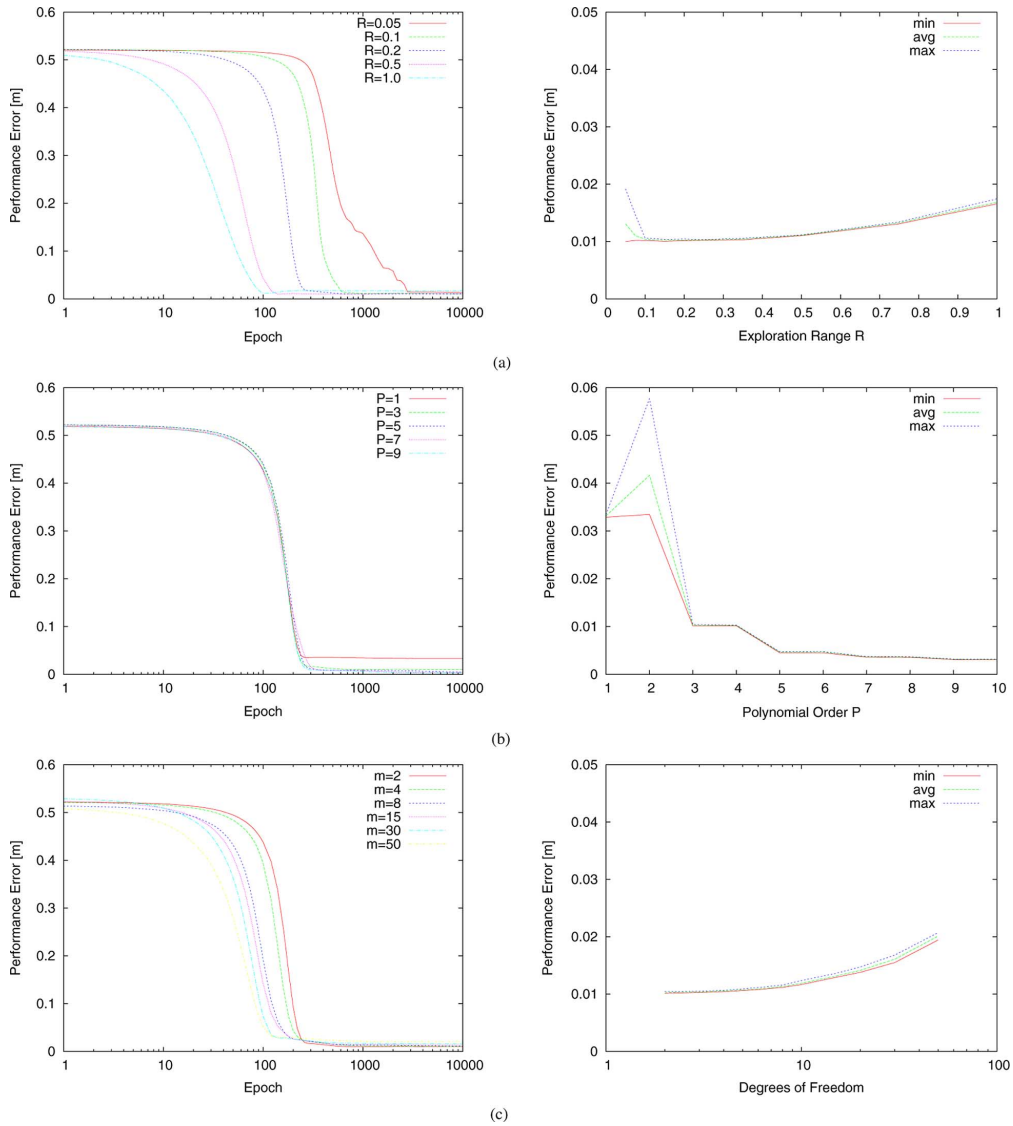


Fig. 7. Performance of goal babbling over 10 000 epochs for the planar arm, whereas only the height is controlled ($n = 1$). The left plots show the performance error over time, averaged over 20 independent trials. The finally reached error is plotted against the varied parameter on the right side. The maximum, average, and minimum error of 20 trials are shown. (a) Results for different exploration ranges R. Higher exploration ranges cause a faster convergence, but higher residual error. (b) Different polynomial degrees P are used for regression. More expressive regression models (higher order polynomials) reach more accuracy, while the speed of convergence does not significantly differ. (c) The number of joints m is increased. Successful bootstrapping of inverse kinematics is possible also for 50 DOF

very accurate inverse estimate, the error reaches few millimeters for higher polynomial degrees (ca. 3 mm error for $P = 10$). The averaging error between variations must therefore be smaller than 3 mm. The error has converged in all cases and shows—depending on the expressiveness of the polynomials—a good performance. Goal babbling was successful for all values of P and in all independent trials.

The next question is how goal babbling scales with the degrees of freedom m. Results for up to 50 degrees of freedom are shown in Fig. 7(c). For each value of m the arm was divided in segments of equal length, whereas we kept the arm length constant at 1 m. For instance, an arm with $m = 10$ comprises 10 segments with each 10 cm length. We used $R = 0.2$ and $V = 20$ in order to compare the results to the previous experiments. The results show a rapid and reliable decrease of the performance error for all values of m and in all trials. The simulated arm with 50 degrees of freedom can be controlled with an accuracy of 2 cm after 10 000 epochs. Goal babbling is systematically successful for such hyperredundant setups.

C. Planar Arm: 2-D Control

We continue the experiments with the simulated planar arm, but increase the dimension of the control task. Instead of controlling only the height ($n = 1$), we now consider 2-D position control of the effector ($n = 2$). The position is encoded in cartesian coordinates with origin in the base of the arm. The step from $n = 1$ to $n = 2$ is essential to show the validity of the movement direction weighting for redundancy resolution (9). In 1-D, the angle between intended and actual movement direction can only be 0.0° or 180.0°. In $n = 2$, arbitrary angles can occur. Since the weighting scheme (10) only uses the immediate temporal and spatial context, each goal position must be passed from different directions for a correct resolution of inconsistencies.

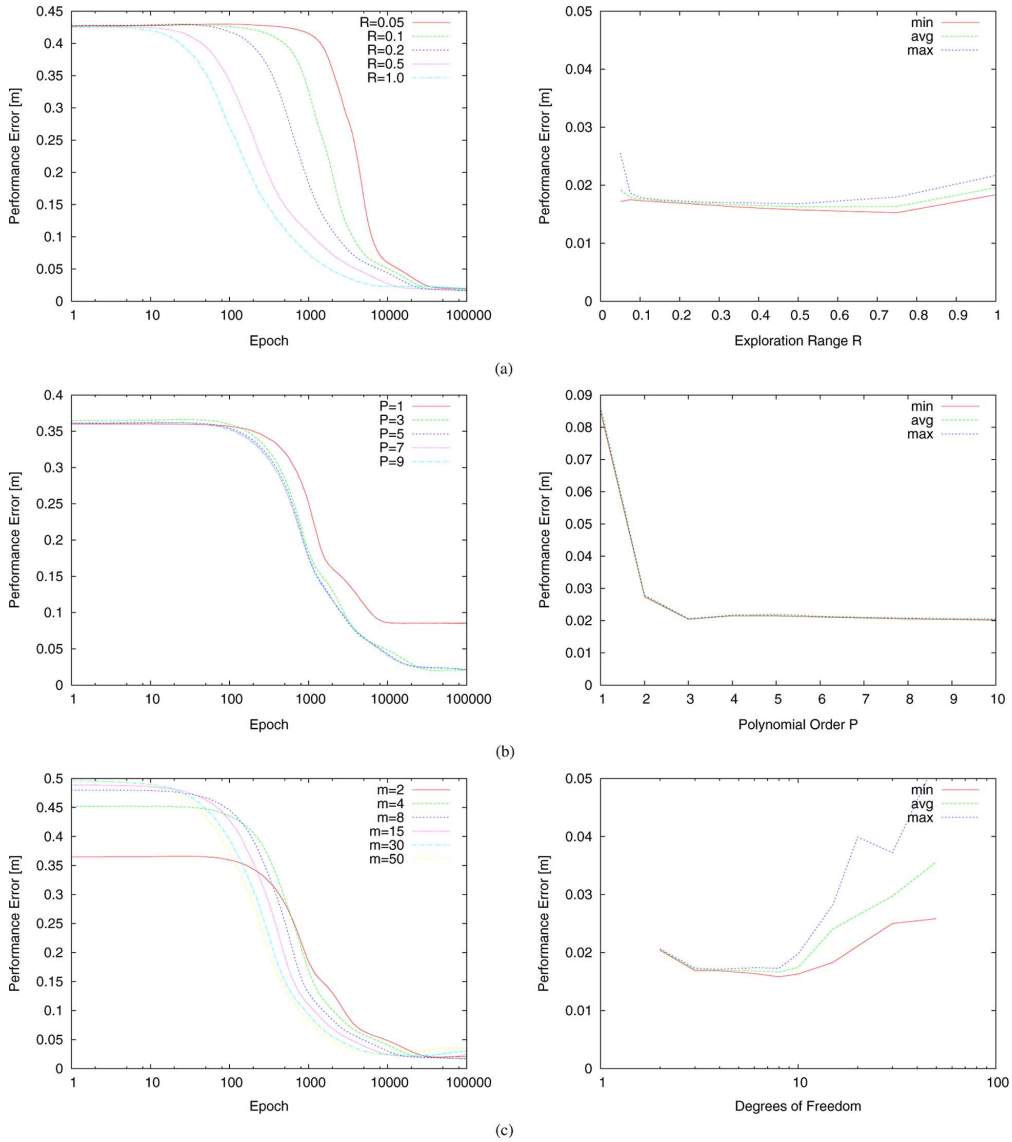


Fig. 8. Performance of goal babbling over 100 000 epochs for the planar arm, where the 2-D position of the effector is controlled ($n = 2$). The left plots show the performance error over time, averaged over 20 independent trials. The finally reached error is plotted against the varied parameter on the right side. The maximum, average, and minimum error of 20 trials are shown. (a) Results for different exploration ranges R. Higher exploration ranges cause a faster convergence, but higher residual error. (b) Different polynomial degrees P are used for regression. More expressive regression models (higher order polynomials) reach more accuracy, while the speed of convergence does not significantly differ. (c) The number of joints m is increased. Successful bootstrapping of inverse kinematics is possible also for 50 DOF.

The aim in this set of experiments is to gain control over a part of the possibly reachable positions as shown in Fig. 9. The set of goal positions is shown as a grid. The home posture is set to a slightly curved shape, since a stretched position corresponds to a singularity in the 2-D task. Learning would still be possible from that position, as either an "elbow-up" or "elbow-down" configuration could be chosen. However, it takes more time for the exploration to leave the singularity. A new sequence of targets x_t^* is generated in each epoch. $K = 15$ positions $x_{k \cdot L}^* \in \mathbf{X}^*, k = 0 \ldots K-1$ are randomly selected from the target grid shown in Fig. 9. One after the other is connected by a linear target motion with $L = 7$ intermediate target positions $(l = 0 \ldots L-1)$

$$x_{k \cdot L + l}^* = \frac{L-l}{L} \cdot x_{k \cdot L}^* + \frac{l}{L} \cdot x_{(k+1) \cdot L}^* \cdot \qquad (17)$$

As in the $n = 1$ experiment, the target selection does not depend on learning progress and $\mathbf{X}^*$ does not change over time. However, in $n = 1$ one linear motion covers the entire target space. In $n = 2$, multiple linear series are required.

With this setup, we repeated all experiments of the $n = 1$ case. The results are summarized in Fig. 8. Goal babbling requires more epochs for convergence than in the $n = 1$ setup. Except for the speed, all results can be reproduced for $n = 2$. Again we used $P = 3, R = 0.2$ and $V = 20$ as default values for the parameters and one redundant degree of freedom $(m = 3)$. Higher exploration ranges [see Fig. 8(a)] result in a faster convergence. The converged performance error only shows marginal differences across values of R. In all cases, the error converges below 2 cm. Only for very small exploration ranges the error has not yet converged after 100 000 epochs. The variation of polynomial degrees P [see Fig. 8(b)] shows a good and re-

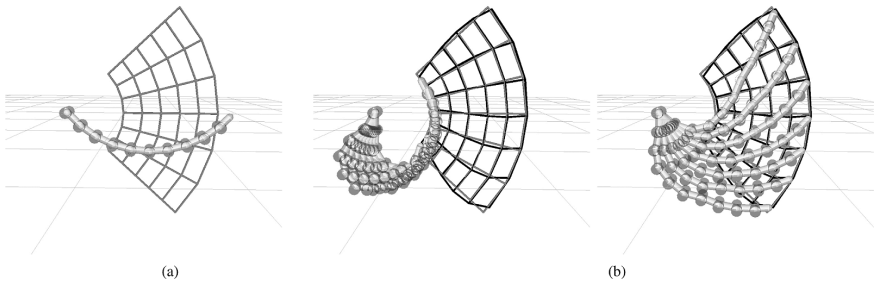


Fig. 9. An inverse estimate for 2-D position control of a planar 10 DOF arm generated with goal babbling. A third order polynomial was used as approximation model. The inverse estimate is very accurate as the reached positions are close to the target positions. The inverse estimate makes efficient use of all degrees of freedom. (a) 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.

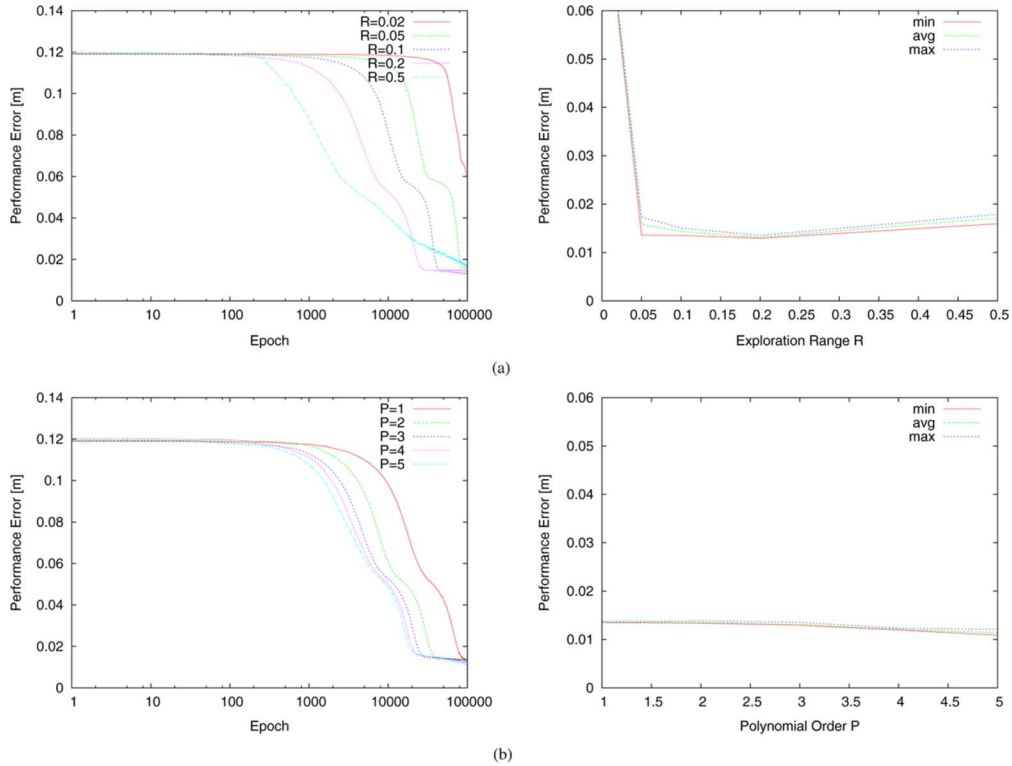


Fig. 10. Performance of goal babbling over 100 000 epochs on the humanoid robot, where the 3-D position of the right hand is controlled ($n = 3$). The left plots show the performance error over time, averaged over five independent trials. The finally reached error is plotted against the varied parameter on the right side. The maximum, average, and minimum error of five trials are shown. (a) Results for different exploration ranges R. Higher exploration ranges cause a faster convergence, but higher residual error. (b) Different polynomial degrees P are used for regression. More expressive regression models (higher order polynomials) reach more accuracy, while the speed of convergence does not significantly differ.

liable performance for all $P \geq 2$. In the case of 2-D position control, linear models $(P = 1)$ are not expressive enough to represent an accurate inverse solution.

Also in the 2-D control case, goal babbling is successful for hyperredundant setups. Fig. 8(c) shows the results for up to 50 degrees of freedom. An example solution $g(x^*)$ for $m = 10$ is shown in Fig. 9. The target positions are reached accurately.

Goal babbling reliably yields accurate inverse estimates for $n = 2$. The results confirm that the resolution of inconsistencies as proposed in the weighting scheme [see (9) and (10)] is valid, although it only uses local information.

D. Honda's Humanoid Research Robot: 3-D Control

We further increase complexity with goal babbling on a kinematic simulation of the Honda humanoid robot [29], where we use $m = 15$ degrees of freedom. Five joint angles are controlled in each arm: three rotational joints in the shoulder, one in the elbow, and the rotation of the hand around the forearm axes. Four virtual joints are controlled in the hip: its orientation around all three spatial axes and the height over ground. The hip degrees of freedom are implemented by means of leg motion, whereas the leg joints are automatically adjusted to realize the desired hip pose [30]. As additional degree of freedom, the head-pan direction is controlled. This joint is, like the joints in the left arm, irrelevant for the task. The kinematic structure is rather complex compared to the planar arm, as the joints have offsets and rotate the hands around different axis. Since the ranges of the possible angles differ significantly between different joints, we normalize the range to $q_i \in [-1.0;\ 1.0]\ \forall i = 1 \ldots 15$.

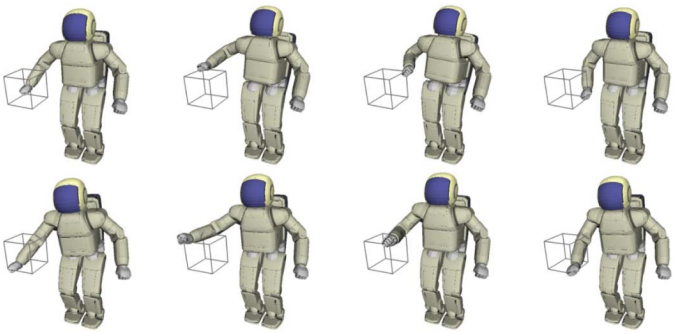


Fig. 11. An inverse estimate for the Honda humanoid robot generated with goal babbling. Target positions are located in a cube with a 20-cm-edge length in front of the body. Several postures show how the inverse estimate reaches for the corners. All relevant degrees of freedom are effectively used. Irrelevant degrees of freedom (e.g., in the left arm) stay approximately fixed.

In this experiment, we aim at a control of the 3-D spatial position of the right hand $(n = 3)$. Nine degrees of freedom are relevant for this task (five in the arm and four in the hip). The set of target positions is defined in cube with 20 cm edge length in front of the upper body (see Fig. 11). A sequence of targets x_t^* is generated newly in each epoch and in the same manner as in the $n = 2$ experiment [see (17)]. $K = 50$ positions are randomly selected from the target set and connected by $L = 10$ intermediate positions. We used $P = 3, R = 0.2$, and $V = 25$ as default parameters.

The results are shown in Fig. 10. The performance error is shown over time for different exploration ranges R and polynomial degrees P. For $n = 3$, it takes more time for the inverse estimate to orient with the correct movement direction. The performance error decreases slowly, but continuously. The temporal curves, but also the converged errors have the same characteristics as in the planar arm experiments. Higher exploration ranges cause faster convergence but higher residual errors. The performance error benefits from higher polynomial degrees, indicating that the full expressiveness of the model can be used. Already linear models $(P = 1)$ yield accurate inverse estimates with performance errors around 1.5 cm inside the cube of targets. An example solution with a third-order polynomial is shown in Fig. 11. The task-relevant degrees of freedom in the hip and in the right arm are used effectively. The task-irrelevant joints are stabilized in an approximately fixed position, which is the most efficient way to deal with irrelevant joints. Goal babbling shows a reliable performance also on humanoid robots with complex kinematic structure in three dimensions.

E. Sensory Noise

So far, we evaluated the effector position 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 (14), but also affects the weight computation [(4) and (5)]. We evaluated the influence of sensory noise exemplary for a planar arm with three joints $(n = 2, m = 3)$. Fig. 12(a) shows results for standard deviations from 0 cm up to 10 cm. 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 $f(q^{\text{home}})$, which accelerates 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 4 cm. However, this amplitude is substantially higher than typical noise in a stereo vision system [31].

Sensory noise on the joint angles has a different effect. We applied Gaussian white noise to the joint angles q_t^v that are used for the weight computation and the learning. Fig. 12(b) shows results for standard deviations 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 slowly with increasing 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 [25], [26]. An efficient detection and resolution of inconsistencies is possible, considering paths on the low-dimensional manifold spanned by the inverse estimate. To our knowledge, this is the first successful approach of direct (example-based) learning that can solve the nonconvexity problem. Moreover, it is the only successful approach to learn a direct inverse kinematics mapping 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 [16], [17] assumes the prior knowledge about motor-errors. Learning with distal teacher [14], [18] relies on a complex mathematical derivation of the motor-error, which is neurally implausible [19]. In contrast, the information needed for goal babbling is fully observable from movement paths—actually reached positions, as well as movement directions and velocities. The method is therefore more plausible as model for human motor development.

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. The introduction of a "structured" noise in the exploration allows to find previously unreachable positions and better solutions in terms of movement efficiency, while maintaining the information structure that is necessary to resolve inconsistencies. Goal babbling is therefore also successful in inverting causality.

Goal babbling is sufficient as exploration strategy to learn inverse kinematics. Unstructured motor-exploration (like motor

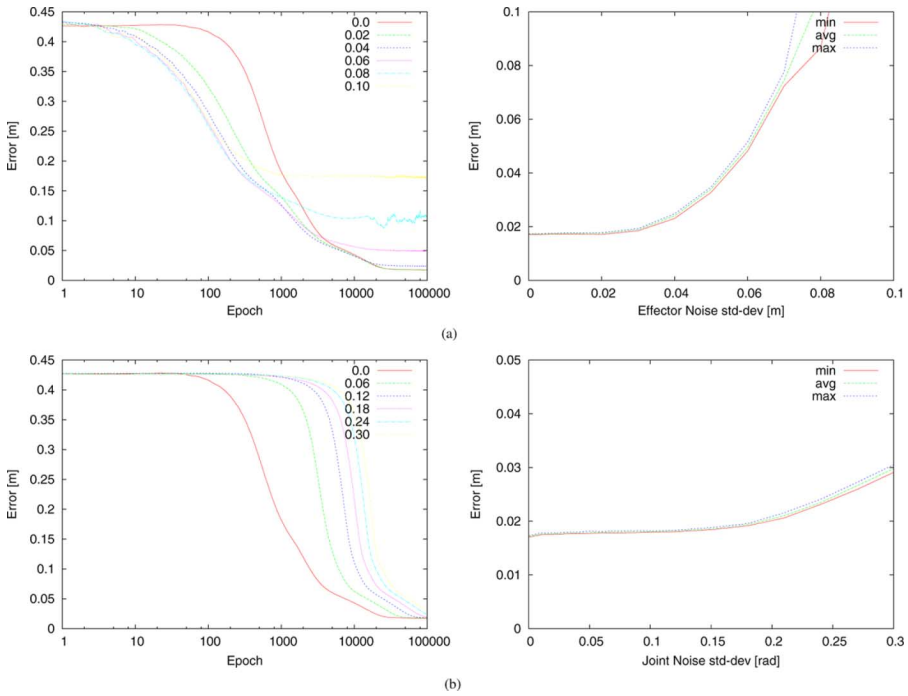


Fig. 12. Performance of goal babbling with sensory noise over 100 000 epochs for the planar arm for $n = 2$ and $m = 3$. The left plots show the performance error over time, averaged over 20 independent trials. The finally reached error is plotted against the varied parameter on the right side. The maximum, average, and minimum error of 20 trials are shown. (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.

babbling) is not only insufficient for redundant systems, it is 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 [4], [32], exploration and control may be based on one mechanism. Our computational model can explain the transition from uncoordinated to coordinated behavior with a single mechanism. This goal-directedness allows to focus exploration on the surrounding of a low dimensional subspace. In contrast, motor babbling aims at an exploration of the entire joint space, which is impractical up to impossible for many degrees of freedom. The focus on behaviorally relevant data is the distinct difference between both exploration methods and constitutes the excellent scalability of goal babbling in high dimensions. In fact, learning only one solution for each target is highly beneficial for the exploration. Not trying to learn all possible solutions also means, that not all solutions must be known and that large portions of the joint space can be ignored. It describes an efficient, developmentally plausible pathway on which, at first, one solution is obtained for a given problem, before other solutions might discovered on demand.

Our simulation experiments reveal a reproducible stage-like behavior. This may provide a possible explanation for infant learning trajectories with rapid transitions between observable stages. The same kind of trajectories can also be found in adult motor learning. When adults have to learn entirely new visuomotor skills, the same qualitative behavior can be found. Sailer and Flanagan [33] found stage like characteristics of the learning progress: an initial plateau, a rapid transition to successful control and fine tuning.

What do infants "babble" in body babbling? Possibly goals instead of motor commands. Our results clearly support the function and relevance of early goal-directed action in infants investigated by von Hofsten and others [6]–[10]. At the same time the approach is fully compatible with Piaget's concept of circular reactions [3]: New experiences can occur by chance, are reproduced, and built into the repertoire of skills. Thereby, the motor commands are not just repeated, but the result becomes a goal which is tried to achieve. The exploration is therefore shaped by prior experience and learning. This view of an incremental, ongoing process goes in line with the dynamic systems view on infant development [34], [35]. Goal-directed action may not be the only form of exploration in infants. Reflexes do certainly play an important role and other forms of exploration may exist. However, "learning by doing," or goal babbling can be successful in learning control from the very beginning. Infants learn to reach by trying to reach. Robots can do so, too.

REFERENCES

[1] D. Wolpert, Z. Ghahramani, and J. R. Flanagan, "Perspectives and problems in motor learning," *Trends Cogn. Sci.*, vol. 5, no. 11, pp. 487–494, 2001.
[2] D. Wolpert, R. C. Miall, and M. Kawato, "Internal models in the cerebellum," *Trends Cogn. Sci.*, vol. 2, no. 9, pp. 338–347, 1998.
[3] J. Piaget, *The Origin of Intelligence in the Child*. London, U.K.: Routledge, 1953.
[4] A. Meltzoff and M. Moore, "Explaining facial imitation: A theoretical model," *Early Develop. Parent.*, vol. 6, pp. 179–192, 1997.

[5] M. M. Vihman, M. A. Macken, R. Miller, H. Simmons, and J. Miller, "From babbling to speech: A re-assessment of the continuity issue," *Language*, vol. 61, no. 2, pp. 397–445, 1985.
[6] C. von Hofsten, "An action perspective on motor development," *Trends Cogn. Sci.*, vol. 8, no. 6, pp. 266–272, 2004.
[7] C. von Hofsten, "Eye-hand coordination in the newborn," *Develop. Psychol.*, vol. 18, no. 3, pp. 450–461, 1982.
[8] L. Ronnquist and C. von Hofsten, "Neonatal finger and arm movements as determined by a social and an object context," *Early Develop. Parent.*, vol. 3, no. 2, pp. 81–94, 1994.
[9] A. van der Meer, F. van der Weel, and D. Lee, "The functional significance of arm movements in neonates," *Science*, vol. 267, no. 5198, pp. 693–695, 1995.
[10] A. van der Meer, "Keeping the arm in the limelight: Advanced visual control of arm movements in neonates," *Eur. J. Paediatric Neurol.*, vol. 1, no. 4, pp. 103–108, 1997.
[11] S. V. Aaron D'Souza and S. Schaal, "Learning inverse kinematics," in *Proc. Int. Conf. Intell. Robot. Syst. (IROS)*, Wailea, Hawaii, 2001, pp. 298–303.
[12] M. Gienger, H. Janssen, and C. Goerick, "Task-oriented whole body motion for humanoid robots," in *Proc. Int. Conf. Humanoid Robot.*, Tsukuba, Japan, 2005, pp. 238–244.
[13] M. Lopes and B. Damas, "A learning framework for generic sensory-motor maps," in *Proc. Int. Conf. Intell. Robot. Syst. (IROS)*, San Diego, CA, 2007, pp. 1533–1538.
[14] M. Jordan and D. Rumelhart, "Forward models: Supervised learning with distal teacher," *Cogn. Sci.*, vol. 16, pp. 307–354, 1992.
[15] J. Baily, "Adaptation to prisms: Do proprioceptive changes mediate adapted behaviour with ballistic arm movements?," *Quart. J. Exp. Psychol.*, pp. 8–20, 1972.
[16] M. Kawato, "Feedback-error-learning neural network for supervised motor learning," in *Advanced Neural Computers*. Amsterdam, The Netherlands: Elsevier, 1990, pp. 365–372.
[17] D. Wolpert and M. Kawato, "Multiple paired forward and inverse models for motor control," *Neural Netw.*, pp. 1317–1329, 1998.
[18] M. I. Jordan, "Computational aspects of motor control and motor learning," in *Handbook of Perception and Action: Motor Skills*. New York: Academic, 1996, pp. 71–120.
[19] 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.
[20] J. Porrill and P. Dean, "Recurrent cerebellar loops simplify adaptive control of redundant and nonlinear motor systems," *Neural Comput.*, vol. 19, no. 1, pp. 170–193, 2007.
[21] Y. Demiris and A. Dearden, "From motor babbling to hierarchical learning by imitation: A robot developmental pathway," in *Proc. Int. Conf. Epig. Robot. (EpiRob)*, Nara, Japan, 2005, pp. 31–37.
[22] Y. Demiris and A. Meltzoff, "The robot in the crib: A developmental analysis of imitation skills in infants and robots," *Inf. Child Develop.*, vol. 17, no. 1, pp. 43–53, 2008.
[23] P. Gaudiano and D. Bullock, "Vector associative maps unsupervised realtime error-based learning and control of movement trajectoires," *Neural Netw.*, vol. 4, no. 2, pp. 147–183, 1991.
[24] D. Bullock, S. Grossberg, and F. H. Guenther, "A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm," *J. Cogn. Neurosci.*, vol. 5, no. 4, pp. 408–435, 1993.
[25] E. Oyama and T. M. S. Tachi, "Goal-directed property of on-line direct inverse modeling," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Como, Italy, 2000, pp. 383–388.
[26] T. D. Sanger, "Failure of motor learning for large initial errors," *Neural Comput.*, vol. 16, no. 9, pp. 1873–1886, 2004.
[27] M. Rolf, J. J. Steil, and M. Gienger, "Efficient exploration and learning of whole body kinematics," in *Proc. Int. Conf. Develop. Learn.*, Shanghai, China, 2009, pp. 1–7.
[28] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, no. 9, pp. 1481–1497, Sep. 1990.
[29] Y. Sakagami, R. Watanabe, and C. Aoyama, "The intelligent asimo: System overview and integration," in *Proc. Int. Conf. Intell. Robot. Syst. (IROS)*, Lausanne, Switzerland, 2002, pp. 2478–2483.
[30] T. Takenaka, "The control system for the honda humanoid robot," in *Age and Ageing*. London, U.K.: Oxford Univ. Press, 2006, pp. 24–26.
[31] G. D. Hager, W.-C. Chang, and A. S. Morse, "Robot hand-eye coordination based on stereo vision," *IEEE Contr. Syst. Mag.*, pp. 30–39, 1995.
[32] L. B. Smith and M. Gasser, "The development of embodied cognition: Six lessons from babies," *Artif. Life*, vol. 11, no. 1–2, pp. 13–30, 2005.
[33] U. Sailer, J. R. Flanagan, and R. S. Johansson, "Eye-hand coordination during learning of a novel visuomotor task," *J. Neurosci.*, vol. 25, no. 39, pp. 8833–8842, 2005.
[34] E. Thelen, "Motor development: A new synthesis," *Amer. Psychol.*, vol. 50, no. 2, pp. 79–95, 1995.
[35] L. B. Smith and E. Thelen, "Development as a dynamic system," *Trends Cogn. Sci.*, vol. 7, no. 8, pp. 343–348, 2003.

Matthias Rolf received the Diploma degree (with distinction) in computer science from the Bielefeld University, Bielefeld, Germany, in 2008. The topic of his diploma thesis was the guidance of visual attention with audiovisual synchrony. Since 2008, he has been working towards the Ph.D. degree at the Research Institute for Cognition and Robotics, Bielefeld University, where his research topic is motor learning and control with neural networks.

His research interests include developmental robotics, machine-learning applications in robotics and vision, multimodal perception, and computational neuroscience.

Jochen J. Steil received the Diploma degree in mathematics from Bielefeld University, Bielefeld, Germany, in 1993 and spent one year as the Chair for Automatic Control of the St. Petersburg Electrotechnical University, St. Petersburg, Russia, supported by a German Academic Exchange (DAAD) grant in 1995/1996. He received the Ph.D. degree with a dissertation on "Input-Output Stability of Recurrent Neural Networks" from Bielefeld University in 1999.

He has been the Managing Director of the Research Institute for Cognition and Robotics (CoR-Lab) since its inauguration in 2007, and adjunct professor for Neuroinformatics at the Faculty of Technology at Bielefeld University since 2008. He is coordinator of the FP7-IP project "Adaptive Modular Architectures for Rich Motor Skills." From 2002–2007, he worked as senior researcher in the Bielefeld Neuroinformatics Group in projects on robot learning architectures, stability and learning of recurrent neural networks, and visual online learning. His research interest include motion learning, learning architectures for complex cognitive robots, visual learning, and behavior organization of humanoid robots. As responsible investigator of the Center of Excellence in Cognitive Interaction Technology, he also pursues projects in computational modeling of visual attention and motor learning.

Michael Gienger received the Diplom-Ingenieur degree from the Technical University of Munich (TUM), Munich, Germany, in 1998. He received the Ph.D. degree also from TUM with a dissertation on "Design and Realization of a Biped Walking Robot" in 2003.

From 1998 to 2003, he was a Research Assistant at the Institute of Applied Mechanics of the TUM, addressing issues in design and realization of biped robots. He is currently a Principal Scientist at the Honda Research Institute Europe, Offenbach, Germany. He also serves as a Scientific Coordinator for the Research Institute for Cognition and Robotics (CoR-Lab) of the Bielefeld University, Bielefeld, Germany. His research interests include mechatronics, robotics, control systems, and cognitive systems, with a particular affection for humanoid robots.