Explorative learning of right inverse functions: theoretical implications of redundancy

Matthias Rolf and Jochen J. Steil

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

Abstract. We investigate the role of redundancy for exploratory learning of inverse functions, where an agent learns to achieve goals by performing actions and observing outcomes. We present an analysis of linear redundancy and investigate goal-directed exploration approaches, which are empirically successful [1], but hardly theorized except negative results for special cases [2], and prove convergence to the optimal solution.

1 Introduction

In many learning scenarios, agents perform actions in some action space, whereas outcomes are measured in a different observation space. We assume that these two spaces are connected by a forward function that turns actions into observations. In order to achieve some desired behavior, an inverse function is needed that returns an appropriate action. A standard example is motor learning, in which action are motor commands like joint angles or forces. A forward function turns the actions into outcomes like effector positions. Learning a corresponding inverse has to rely on exploration schemes that generate examples for supervised learning. One substantial challenge is to deal with the redundancy in such domains: often multiple actions are mapped on the same outcome, such as different joint angles of an arm resulting in the same hand position. In this case learning cannot be phrased as standard regression problem because multiple correct solutions exist. Often, the action space is very high-dimensional, which makes exhaustive exploration unfeasible. Yet, a number of practically efficient schemes have been proposed based on “goal-directed” exploration. This idea has been used for tuning of well initialized inverse functions [3–5]. Goal-directed exploration is particularly beneficial for learning from scratch, because it is applicable in very high-dimensional spaces [1]. Only very few theoretical results are available why and when such schemes can be successful. To the opposite, Sanger [2] proved that certain formulations can fail systematically even in simple domains.

This paper aims to deepen the theoretical understanding of such learning schemes in redundant domains. We first formalize the general problem and discuss its difficulties. Then, we provide a throughout analysis of the linear case with redundancy, which is applied to goal-directed exploration. To our knowledge, we thereby provide the first positive theoretical outcomes on such learning by proving convergence to an optimal solution if exploratory noise is added.
2 Two Spaces and their Gradients

We consider an agent that can execute actions $q$ in the action space $Q \subseteq \mathbb{R}^m$. An action results in an outcome $x \in \mathbf{X} \subseteq \mathbb{R}^n$ in the observation space. Both variables are coupled by the forward function $f(q) = x$. The agent is asked to achieve some “goal” observation $x^* \in \mathbf{X}^* \subseteq \mathbf{X}$. It has to generate an action $\hat{q}$, such that $x = f(\hat{q}) = x^*$. The agent’s selection of an action can be denoted by a function $g(x^*) = \hat{q}$. The learning task is to obtain a function $g$ that can realize all goals:

$$f(g(x^*)) = x^* \forall x^* \in \mathbf{X}^*$$

(1)

Hence, $g$ must be a right inverse function of $f$ on the set of goals $\mathbf{X}^*$. Inverse functions do not always exist, so we need to require that $f$ is surjective with $n \leq m$. For $n < m$ different actions can result in the same outcome, which is referred to as redundancy. An exemplary situation is shown in figure 1a.

2.1 The learning task in the observation space

In the observation space, obtaining a right inverse function can be directly formulated as optimization problem. We parametrize the function $g$ with $W$. For some set of goals $\mathbf{X}^* = \{x^*_0, \ldots, x^*_{K-1}\}$, the performance error $E^X$ naturally measures how much an inverse estimate $g$ deviates from the solution in (1).

$$E^X(W, \mathbf{X}^*) = \frac{1}{2K} \sum_{k=0}^{K-1} ||f(g(x^*_k, W)) - x^*_k||^2$$

(2)
Learning can be formulated as gradient descent on this error [6, 7]. The central difficulty is that computing the performance gradient $\partial E_X / \partial W$ requires analytic knowledge about the forward function: Since $W$ appears inside $f(\cdot)$, differentiating $\partial E_X / \partial W$ requires to know the derivative of $f$. In general, inverse problems do not provide a teacher to indicate such optimal gradient directions.

### 2.2 Explorative learning in the action space

If the performance gradient is not available, a feasible way to probe knowledge is to generate examples $(x, q)$ by exploration [1–5, 8, 9]. The setup starts by performing some action $q_l$, and observing the outcome $x_l = f(q_l)$. The action error on $D = \{(x_l, q_l)\}_l$ measures how well the inverse estimate fits the data:

$$E_Q(W, D) = \frac{1}{2L} \sum_{l=0}^{L-1} ||g(x_l, W) - q_l||^2.$$  (3)

Learning is performed by descending the action gradient $\partial E_Q / \partial W$. Importantly, this scheme is not a data-driven version of minimizing $E_X$. In (3) we can replace $x_l = f(q_l)$ and see that the error evaluates on $g(f(q_l)) - q_l$. Hence, reducing $E_Q$ corresponds to learning a left inverse function $g$, while the learning task is to obtain a right inverse function, which corresponds to minimizing $E_X$ in observation space. Empirical results show that a right inverse function can be learned by minimizing $E_Q$ [1, 3, 8]. Why this is possible is not theoretically understood for the general case. In fact, this kind of learning largely depends on how the data set chosen, whether $f$ is linear or not, and whether the system contains redundancy: For the redundant case, left inverse functions do not exist on general data sets because different $q_l$ can have the same outcome $x_l$. Trying to fit such inconsistent examples results in averaging, leading to invalid results in non-linear domains [6]. Sanger [2] investigated goal-directed exploration in the non-linear case without redundancy and showed that learning is not guaranteed to work.

This paper complements these previous, negative outcomes and investigates redundancy in linear domains. As a first positive result, we show that performance- and action-gradient have a non-negative angle and provide fix-point conditions.

### 3 Gradients in linear domains

In the linear domain, the relation between actions $q \in Q \subseteq \mathbb{R}^m$ and outcomes $x \in X = f(Q) \subseteq \mathbb{R}^n$ is given by the linear forward function:

**Definition 1 (Linear Forward Function).** We define the forward function as $f(q) = M \cdot q$ where $M$ is a matrix $M \in \mathbb{R}^{n \times m}$ with $n \leq m$ and rank$(M) = n$.

Requiring $M$ to have full rank implies solvability of the right inverse problem. Correspondingly, we use a linear inverse estimates, with parameters $W$:

**Definition 2 (Linear Inverse Estimate).** We define the inverse estimate as $g(x^*, W) = W \cdot x^*$ where $W$ is a real-valued parameter matrix $W \in \mathbb{R}^{m \times n}$. 
Using these two definitions, we can re-write the right inverse equation (1) as linear equation. Assuming that the goals $x^*$ span the entire space $\mathbf{X}$ we get:

$$f(g(x^*)) = x^* \quad \forall x^* \iff MW = \mathbf{1}_n.$$  

(4)

Hence, $W$ must be a right inverse matrix of $M$. This equation is exactly solvable in $W$. For $n < m$ it is ill-posed and multiple solutions $W$ exist.

We can now insert these definitions in the error-functionals defined in the last section and compute the gradients. For the performance gradient we get:

$$\frac{\partial E^X(W, \mathbf{X}^*)}{\partial W} = \frac{\partial E^X(W, \mathbf{X}^*)}{\partial W} = M^T(MW - \mathbf{1}_n)\mathbf{X}^*$$

(5)

with $\mathbf{X}^* = \frac{1}{K} \sum_{k=0}^{K-1} x_k^* x_k^{* T} \in \mathbb{R}^{n \times n}$.

Fig. 1b shows the performance gradient in relation to correct right inverse solutions. As an example we have chosen a forward matrix $M = (0.5, 0.5) \in \mathbb{R}^{1 \times 2}$. The figure shows the parameter space of $W \in \mathbb{R}^{2 \times 1}$. Right inverse matrices fulfill $MW = \mathbf{1}_1$ or in scalar notation $MW = 1$. These solutions give $\frac{\partial E^X}{\partial W} = 0$. The performance gradient drives any value of $W$ straight to that solution manifold.

Considering a data set $D = \{(x_i, q_i)\}_l$ with $x_i = Mq_i$, we first define

$$\mathbf{Q} = \sum_{k=0}^{L-1} q_k q_k^T \quad \text{and} \quad \mathbf{X} = \sum_{k=0}^{L-1} x_k x_k^T = MQM^T.$$  

Using this notation the action gradient is:

$$\frac{\partial E^Q(W, D)}{\partial W} = \frac{\partial E^Q(W, \mathbf{Q})}{\partial W} = (WM - \mathbf{1}_m)\mathbf{Q}M^T.$$  

(6)

We can now show a tight relation between minimizing $E^X$ and $E^Q$:

**Theorem 1.** For any data set $D$, the action gradient is related to the performance gradient on the observed $\{x_i\}$ positions by

$$M^T M \frac{\partial E^Q(W, \mathbf{Q})}{\partial W} = \frac{\partial E^X(W, \mathbf{X})}{\partial W}.$$  

(7)

Proof.

$$M^T M \frac{\partial E^Q(W, \mathbf{Q})}{\partial W} \overset{(6)}{=} M^T M(WM - \mathbf{1}_m)\mathbf{Q}M^T = M^T(MWM - M)\mathbf{Q}M^T$$

$$= M^T(MW - \mathbf{1}_n)\mathbf{Q}M^T = M^T(MW - \mathbf{1}_n)\mathbf{X} \overset{(5)}{=} \frac{\partial E^X(W, \mathbf{X})}{\partial W} \square$$

Both gradients have a non-negative angle since $M^T M$ is a positive semi-definite matrix. For $n = m$, $M^T M$ is even positive definite which guarantees a positive angle. Hence, learning a right inverse function is generally possible by minimizing
For $n < m$, $M^TM$ is singular and the action gradient can project in its nullspace. This makes the redundant case mildly more complicated, but not as difficult as the general non-convex case, for which no angle can be guaranteed for arbitrary data sets. However, this theorem does not give a direct relation to the performance on the actual goals $E^X(W,X^*)$. Whether a right inverse can be learned depends on whether the observations $\{x_l\}$ in $D$ span the entire space $X$.

4 Fixpoint analysis

In this section we investigate how the data set $D$, generated by exploration, shapes the learning process. Starting from some initial parameter value $W_0$, the parameters are iteratively updated with the learning equation

$$W_{t+1} = W_t - \eta \frac{\partial E^Q(W,Q)}{\partial W}.$$  \(8\)

Our main concern is whether learning converges to a $W$ that satisfies the right inverse function condition $MW = 1_n$. In order to check for this behavior, we analyze the fixpoints $\partial E^Q/\partial W = 0$ of the learning equation depending on $D$.

The following two theorems give general conditions for which combinations of a parameter value $W$ and data set $D$ the action gradient becomes zero.

**Theorem 2 (Sufficient fixpoint condition).** If $W$ is a partial left inverse of $M$ on the actions $q_l$ (i.e. $WMq_l = q_l \forall q_l \in D$), then $W$ is a fixpoint of eqn. (8).

**Proof.**

$$WMq_l = q_l \forall q_l \in D \Leftrightarrow WMQ = Q \Leftrightarrow (WM - 1_m)Q = 0 \Rightarrow (WM - 1_m)QM^T = 0 = \frac{\partial E^Q}{\partial W} \quad \square$$

Sanger [2] showed for goal-directed exploration that this condition is also sufficient in the non-linear case with $n = m$. In fact, this condition is very general because it indicates that the action error in eqn. (3) is zero. The learner already fits the data. In a linear system with $n = m$, the condition is also necessary because $M$ is square with full rank. Therefore the right-multiplication with $M^T$ in the proof is reversible and we get equivalence between both statements. For redundant systems the condition is not necessary, since a left inverse does not exist on arbitrary data sets. If, for instance, $D$ contains data $q_i \neq q_j$ and $x_i = x_j$, these samples can not be fitted. A more general condition is given by:

**Theorem 3 (Necessary fixpoint condition).** If $W$ is a fixpoint of learning equation (8), then $W$ is a (partial) right inverse of $M$ on the observed positions $x_l$, i.e. $MWx_l = x_l \forall x_l \in D$.

**Proof.**

$$\frac{\partial E^Q(W)}{\partial W} = 0 \Rightarrow M \frac{\partial E^Q(W)}{\partial W} = 0 \Leftrightarrow M(WM - 1_m)Q = 0 \Rightarrow M(WM - 1_m)QMT = (MW - 1_n)Q = 0 \Leftrightarrow MWX = X \Leftrightarrow MWx_l = x_l \forall l \quad \square$$
Like theorem 2 this statement becomes an equivalence for \( n = m \) (here because the left-multiplication with \( M \) is reversible). We can summarize both theorems:

\[
WMq_l = q_l \forall l \Rightarrow \frac{\partial E^Q(W)}{\partial W} = 0 \Rightarrow MWx_l = x_l \forall l
\]

Only for \( n = m \) we get a full equivalence between these conditions. This asymmetry for \( n < m \) is the second result on the impact of redundancy, additionally to the gradients losing their strictly positive relation in theorem 1. According to theorem 3, learning from examples will always result in a right inverse solution on the outcomes \( x_l \) contained in the data set. If the outcomes do not span the entire space \( X^* \), the solution will only be valid in the corresponding subspace.

### 4.1 Goal-directed Exploration

With these fixpoint conditions we can investigate right inverse learning driven by particular exploration processes. Goal-directed exploration has been discussed for the generation of data \( D \) in [3], but using a formulation without exploratory noise that can possibly fail even in non-redundant domains [2].

A data set \( D_t = \{(x_t^k, q_t^k)\}_{k} \) is newly generated for each learning step \( t \). The current inverse estimate \( g(x^*_k, W_t) \) is evaluated on \( X^* = \{x_0^*, ..., x_{K-1}^*\} \) to select actions \( q_t^k \). In order to inject exploratory noise, we follow [1] and add a perturbation function \( E(x_t^*k) \) to the inverse estimate. In the linear case we obtain this by choosing actions with some generating matrix \( W_{\text{gen}} \):

\[
q_t^k = g(x_t^*, W_t) + E(x_t^*k) = W_{\text{gen}}x_t^* \quad \text{with} \quad W_{\text{gen}} \sim W + \varepsilon .
\]

The components of the perturbation \( \varepsilon \in \mathbb{R}^{m \times n} \) are chosen i.i.d. with zero mean and variance \( \sigma^2 \). Examples for multiple perturbations are collected and used for one gradient step according to equation (8). For our analysis we assume that enough data is collected to approximate the learning process by the expectation of this exploration process. First, we derive the expected action matrix:

\[
Q = E[\{(W + \varepsilon)X^*(W + \varepsilon)^T\}] = E [WX^*W^T + WX^*\varepsilon^T + \varepsilon X^*W + \varepsilon X^*\varepsilon^T]_\varepsilon
\]

Here we get \( E[WX^*\varepsilon^T] = E[\varepsilon X^*W] = 0 \) because \( E[\varepsilon] = 0 \). Further we can derive that \( E[\varepsilon X^*\varepsilon^T] = \text{trace}(X^*)\sigma^2 I_m \), which gives the action matrix

\[
Q = WX^*W^T + \text{trace}(X^*)\sigma^2 I_m . \quad (9)
\]

Without noise (\( \sigma = 0 \)) this matrix has a rank of at most \( n \), but can also become zero if \( W \) is zero. This degeneration can cause failures of learning [2]. With noise, it has full rank, which implies that all fixpoints are valid right inverse functions:

**Proposition 1.** For \( \sigma^2 > 0 \): \( \text{rank}(Q) = m , \; \text{rank}(X) = n \)

**Proof.** \( \text{rank}(Q) = m \): The symmetric form \( WX^*W^T \) in eqn. 9 is positive-semidefinite. The second term is positive-definite for \( \sigma^2 > 0 \). The sum of a positive-semidefinite and a positive-definite matrix is also positive-definite, which implies full rank. \( \text{rank}(X) = \text{rank}(M(QM^T)) = n \) then follows from basic linear algebra. \( \square \)
Proposition 2. For $\sigma^2 > 0$, any fixpoint $W$ of the learning equation 8 is a right inverse of $M$.

Proof. We know from theorem 3 that $MWX = X$ for any fixpoint. Since $X$ has full rank we can right-multiply with $X^{-1}$ and get: $MWX = X \Rightarrow MW = 1_n$. □

For a full analysis we insert $Q$ into the gradient equation (6):

$$\frac{\partial \tilde{E}(W)}{\partial W} = (WM-1_m)(WX^*W^T + \text{trace}(X^*)\sigma^2 1_m)M^T.$$  

Using this equation we can show that the exploration results in a unique fixpoint:

Theorem 4. For $\sigma^2 > 0$, the unique fixpoint of the learning equation 8 is the Moore-Penrose pseudoinverse: $W = M^\# = M^T(MM^T)^{-1}$.

Proof. Expanding the gradient first gives for $\alpha = \text{trace}(X^*)\sigma^2 > 0$:

$$0 = \frac{\partial \tilde{E}(W)}{\partial W} = WMWX^*W^TM^T + W\alpha 1_m M^T - W^*W^TM T - \alpha 1_m M^T$$

We can now use the previous result $MW = 1_n$ and substitute $MW$ with $1_n$:

$$WX^* + \alpha WMM^T - W^* - \alpha M^T = \alpha WMM^T - \alpha M^T = 0$$

$$\Leftrightarrow WMM^T = M^T \Leftrightarrow W = M^T(MM^T)^{-1}.$$ □

Fig. 2 illustrates the learning gradient for goal-directed exploration in the example with $M = (0.5, 0.5)$ without (a) and with noise (b). Without noise, the procedure can end up in any of the fixpoints described by theorem 3. It stops in
any correct solution \( MW = 1 \), but also in the entire nullspace of \( M \) \( (MW = 0) \). Gray lines show exemplary trajectories on which \( W_t \) is changed during learning. Without noise, these trajectories are entirely concentric: the exploration never leaves the initial column space and does not allow to orient for new stimuli. With noise, the qualitative behavior is drastically changed (Fig. 2b). Noise removes the erroneous fixpoints on \( MW = 0 \). The gradient is not concentric around \( W = 0 \) anymore and allows \( W \) to change the column space. On the solution manifold \( MW = 1 \) the gradient pulls \( W \) towards the pseudoinverse, which is \( W = M^\# = (\begin{array}{c} 1 \\ 0 \\ 1 \end{array} ) \) in the example.

5 Discussion

We have investigated the explorative learning of right inverse functions, in order to let an agent perform actions that achieve some goal. It turns out that learning from examples corresponds to learning left inverse function. For the redundant linear case we have shown that such learning satisfies a non-negative gradient-relation to the actual right inverse problem. The analysis of goal-directed exploration as previously discussed in [2] shows that learning can lead to the discovery of valid right inverse functions, but may also get stuck in subspaces if the observation matrices lose rank. Redundancy causes additional failure modes in the Nullspace of the forward function. The new results for exploratory noise are particularly encouraging. Obviously, exploratory noise spans the entire observation space which eliminates undesirable fixpoints in subspaces. If applied in a redundant domain, noise even leads to the selection of the least-squares solution among the infinite set of solutions. Here our analysis gives the first affirmative results on goal-directed exploration from a theoretical perspective.

References