Low-rank forward models: a path to the self-organization of visuo-motor systems.

Ângelo Cardoso[1], Ricardo Ferreira[1], Ricardo Santos[1], Alexandre Bernardino[1]

***Abstract*— Sensorimotor coupling is ubiquitous in living organisms. Sensory and motor systems are utterly useless if left without the presence of the other. One crucial faculty that organisms have developed with tremendous ecological advantages is the ability to discern between the origins of perceptual input as being originated by the environment or the organism itself, provided by resource efficient sensor and motor systems. This ability has been shown to be implemented through a specialized circuit (forward model) receiving a copy of the motor command (corollary discharge). We propose a fast method to derive a resource constrained forward model by framing sensorimotor coupling as a low-rank approximation of an overly detailed forward model. By framing the problem as a factorization approach we can resort to currently available off-the-shelf solvers for matrix factorization. We experimentally show that by solving the problem as a low-rank approximation we obtain more than an order of magnitude speed up relatively to minimizing the objective function with gradient descent methods. The development of resource constrained and ecologically adapted sensorimotor systems is essential for the deployment of low-cost energy efficient autonomous robots for the execution of specific tasks in particular environments.**

I. Introduction

Without perception (the organization, identification and interpretation of sensory information) one is left with little criteria to decide which actions to take, while at the same time there is no purpose in having perception if you cannot act on the world. An ideal rational agent [1] always takes the actions which maximize its performance measure based on its perceptions and built-in knowledge. This definition frames perception as a component used to choose the right action, and not as a goal by itself. Under this light a broad goal is to develop sensorimotor structures which support choosing the right action. Analogously animal brains evolved for sensorimotor control and still retain much of that architecture [2].

Sensorimotor coupling refers to the fact that the sensory and motor systems work in an integrated fashion. In primates, future fields [3] are a good example of this coupling: before a saccade (cortically controlled eye movement) the animal brain already has an expectation of what it should see after the saccade. This is thought to be very important for having stable visual perception despite the eyes performing several saccades per second. Whenever an animal performs an action, the corresponding motor command is not only sent to the muscles to invoke movement but also a copy (corollary discharge) is sent to other brain areas with the goal of generating predictions of future stimuli.

[1] Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon, Portugal – Email: {acardoso,ricardo,rsantos,alex}@isr.ist.utl.pt

In this work we consider an agent that collects visuo-motor experiences in the environment and self-organizes its visual topology (a retina), its motor system (a set of actions) and a predictor system able to forecast the future stimuli resulting from the agent's own motor actions. In previous work [4] we have shown that a special structure in the coupling between the visual, motor and predictive systems enable the development of highly adapted visuo-motor predictive structures, exhibiting simultaneously low computational resources and low prediction error. In that work development was achieved through the minimization, with gradient descent methods, of the prediction error subject to structure constraints.

In this paper we frame the sensorimotor coupling problem as a low-rank approximation of an overly detailed forward model, i.e. a forward model which makes explicit predictions about a high number of sensors and motor actions. The low-rank approximation in the sensors can be seen as finding groups of sensors which are affected by the motor actions in a similar way. While on the motor actions it can be seen as grouping motor actions which affect the sensors in a similar way. By framing the problem as a low-rank approximation, we can resort to efficient methods for matrix factorization available in the literature. By doing so, as we will see in later sections, we also separate the sensorimotor coupling into a set of sub-problems which can be solved in a more efficient manner. We will focus on a particular method to accomplish the low-rank approximation: non negative matrix factorization [5], however others could be used such as non negative sparse coding [6].

II. Related Work

The ability to predict the sensory effects of motor actions is grounded in the corollary discharge (copy of a motor command) which is sent to a brain area for a different reason than producing a movement. The corollary discharge is crucial for the ability to discern the sensory input as being originated by the environment (exafference) or as the result of animal's own movement (reafference) [7]. The ability to discern between these two origins of sensory input requires a forward model [8] to predict the effect a given movement (action) has on its sensory input.

The retinotopic structure of an unknown visual sensor has been reconstructed using an information measure, as well as the optical flow induced by motor actions [9]. A robot with the goal of estimating the distance to objects using motion parallax developed a morphology for the position of movable light sensors which was fit for the task [10].

Guiding the development of a sensorimotor system to maximize the ability of predicting the effect an action has on its sensory input, allows for the emergence of highly regular sensory structures without any prior knowledge. To develop such ability we follow two main principles: the sensory system should capture stimuli which are relevant to motor capabilities and the actions of the motor system should have predictable effects on the sensory system [11].

These principles are related to idea of "morphological computation" in robotics and artificial intelligence, which aims at reducing the computational complexity of a problem by using a specifically designed body to solve it (e.g. [12]).

The non negative matrix factorization (NMF) problem [5] has been proposed to decompose a matrix into the product of two positive matrices and it has found multiple applications in engineering. Multiple efficient methods have been proposed to solve the problem [13].

The human visual system representation of the visual world is progressively differentiated from what is captured through the retina to support complex tasks, e.g. cells which are selective to objects. Also, in machine learning it is known that for recognition tasks there are huge advantages in using specific architectures [14].

III. Finding a low-rank forward model

We consider an agent capable of observing its environment by sensing a light field i which falls on a sensory surface. Similarly this agent is able to interact with its environment by executing a particular primitive action $\mathbf{q}$ on its action space. For implementation purposes, we represent the light field as a vector $\mathbf{i}$ of N_S pixels, and the action space is represented as a vector $\mathbf{q}$ with N_m elements, where a single non-zero entry represents the activated motor primitive. If the m^{th} index of $\mathbf{q}$ is 1, then the m^{th} primitive action is performed (e.g. shift left by a certain amount). Note that no topological assumptions exist on the spatial locations of either the sensors or the motor primitives.

During the learning phase, the agent interacts with the environment by randomly choosing a motor primitive $\mathbf{q}$ while collecting pre-action and post-action sensory stimuli ($\mathbf{i}_0$ and $\mathbf{i}_1$ respectively). A set of ($\mathbf{I_0}^m$, $\mathbf{I_1}^m$,$\mathbf{q}^m$) triplets is collected and the full batch is used as training data. Triplets are obtained from a 2448 by 2448 pixels grayscale image as environment (see Figure III).

To illustrate the overall method and result of each step we we use a retina of 32 by 32 pixels ($N_S = 1024$) and motor commands with 4 degrees of freedom and 5 actions per degree of freedom ($N_M = 625 = 5^4$) being the actions $\{-12,-6,0,6,12\} \times \{-12,-6,0,6,12\} \times \{-90°,-45°,0°,45°,90°\} \times \{0.8,0.9,1,1.1,1.2\}$ (XY translations, rotation and zoom respectively) .

Now considering the Sensory Motor optimization problem [11], which defines an objective function for the problem we address throughout this text,

$$\underset{\mathbf{P}>0,\mathbf{S}>0,\mathbf{M}>0}{\operatorname{argmin}} \sum_{m}^{N_M} \left\| \mathbf{S}^T \left[\mathbf{P}\mathbf{M}^T \mathbf{q}^m \right]_{\square} \mathbf{S}\mathbf{I_0}^m - \mathbf{I_1}^m \right\|^2 \quad (1)$$

Fig. 1. Image used as environment to create the triplets ($\mathbf{I_0}^m$, $\mathbf{I_1}^m$,$\mathbf{q}^m$) – horizontal and vertical axes represent pixels.

Here $\mathbf{S}$ ($N_s \times N_S$) contains the receptive field (N_s elements) representation of the input sensor space (N_S elements), $\mathbf{M}$ ($N_m \times N_M$) the motor field (N_m elements) representation of the motor actions (N_M actions) and $\mathbf{P}$ is a basis for predictors ($N_s \cdot N_s \times N_m$), interpreted as associated with both the motor and receptive fields. $\mathbf{I_0}^m$ ($N_S \times N_a^m$), $\mathbf{I_1}^m$ ($N_S \times N_a^m$) are pre and post sensor input respectively (collected data) for primitive action m over N_a^m samples. The column vector $\mathbf{q}^m$ ($N_M \times 1$) has a single entry different from zero and equal to 1 depending on which of the N_M distinct actions considered in our formulation it represents. The square symbol $\square$ denotes a reshape of the contained vector to a square matrix.

A. Find the high-rank predictors $\mathfrak{P}$

A relaxation of Equation (1) is the simpler problem

$$\underset{\mathfrak{P}^m>0}{\operatorname{argmin}} \sum_{m}^{N_M} \left\| \mathfrak{P}^m \mathbf{I_0}^m - \mathbf{I_1}^m \right\|^2 \quad (2)$$

where a full predictor $\mathfrak{P}^{\mathbf{m}}$ ($N_S \times N_S$) is learned for each kind of primitive action m. Since we consider each action m one at a time, samples $\mathbf{I_0}^m$ and $\mathbf{I_1}^m$ can be discarded after $\mathfrak{P}^{\mathbf{m}}$ is obtained).

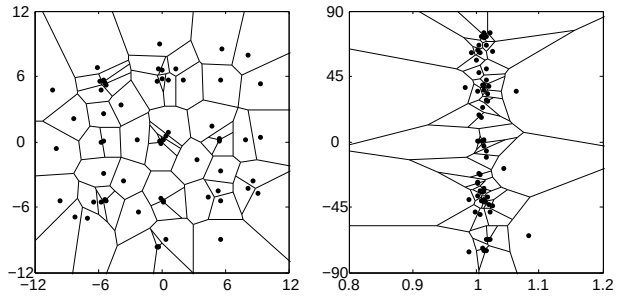
After estimating the full predictors $\mathfrak{P}^{\mathbf{m}}$ a factorization is applied to estimate all the symbols in problem (1) ($\tilde{\mathbf{S}}, \tilde{\mathbf{M}}, \tilde{\mathbf{P}}$) such that for any action m

$$\mathfrak{P}^m \approx \tilde{\mathbf{S}}^T \left[\tilde{\mathbf{P}} \tilde{\mathbf{M}} \mathbf{q}^{\mathbf{m}} \right]_{\square} \tilde{\mathbf{S}} \quad (3)$$

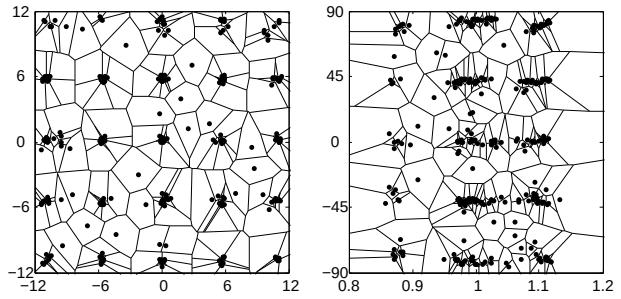
more specifically we will find ($\tilde{\mathbf{S}}, \tilde{\mathbf{M}}, \tilde{\mathbf{P}}$) which are common to all actions such that

$$\underset{\tilde{\mathbf{P}}>0,\tilde{\mathbf{S}}>0,\tilde{\mathbf{M}}>0}{\operatorname{argmin}} \sum_{m}^{N_M} \left\| \tilde{\mathbf{S}}^T \left[\tilde{\mathbf{P}} \tilde{\mathbf{M}} \mathbf{q}^m \right]_{\square} \tilde{\mathbf{S}} - \mathfrak{P}^m \right\|_F^2 \quad (4)$$

After the full set $\{\mathfrak{P}^m\}$ has been sampled, we will factorize this data to recover approximations to $\tilde{\mathbf{S}}$, $\tilde{\mathbf{M}}$ and $\tilde{\mathbf{P}}$ using equation (3) to impose the required structure. We do this one variable at a time. While we do this mostly for simplicity as we do not know how to do it all at once, by solving one variable at a time we also reduce the computational complexity by solving several smaller sub-problems.



(a) XY motor fields, $N_m = 64$ (b) RZ motor fields, $N_m = 64$



(c) XY motor fields, $N_m = 256$ (d) RZ motor fields, $N_m = 256$

Fig. 2. Motor fields ($\tilde{\mathbf{M}}$) found by the proposed method for different N_m values – horizontal and vertical axes represent the actions for each degree freedom (i.e. horizontal and vertical translations for XY; zooms and rotations for RZ respectively) , the center of mass of each motor field is shown in a voronoi partition. Each motor field center of mass is defined by a 4 dimensional point (4 degrees of freedom) which we project for visualization into horizontal and vertical translation degrees of freedom (XY) and rotation and zoom (RZ). Each degree of freedom values are scaled to a common reference to have equal influence in the space partition. We can see that if the number of motor fields available N_m is higher, each one covers a smaller portion of the action space and therefore a larger specialization occurs. We can see for instance that if for $N_m = 64$ there is almost no specialization in different zoom scales while for $N_m = 256$ there are additional motor fields specialized in zoom in and zoom out.

Consider the nonnegative matrix factorization problem of decomposing matrix $\mathbf{V}$ $(n \times d)$ into a product of two positive matrices $\mathbf{W}$ $(n \times k)$ and $\mathbf{H}$ $(k \times d)$ such that $\mathbf{V} \approx \mathbf{WH}$ where both $\mathbf{W}$ and $\mathbf{H}$ are positive. More specifically, the solution of the problem

$$\underset{\mathbf{W}>0,\mathbf{H}>0}{\operatorname{argmin}} \ \|\mathbf{V} - \mathbf{WH}\|_F^2, \tag{5}$$

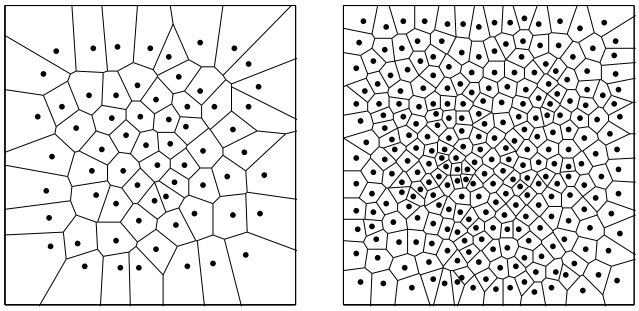
where $\|\|_F^2$ is the Froebenius norm and k is a parameter controlling the dimensionality of matrices $\mathbf{W}$ $(n \times k)$ and $\mathbf{H}$ $(k \times d)$ decomposing $\mathbf{V}$ $(n \times d)$ and is inversely related to the residual error ($\|\mathbf{V} - \mathbf{WH}\|_F^2$). In this paper we use the parameter-free non-negative matrix factorization method from [15].

B. Find the motor fields M

We start by factoring $\tilde{\mathbf{M}}$. We take equation (3) and rewrite it as [1]

$$\text{vec}\,(\mathfrak{P}^m) \approx (\tilde{\mathbf{S}}^T \otimes \tilde{\mathbf{S}}^T)\tilde{\mathbf{P}}\tilde{\mathbf{M}}\mathbf{q}^m \tag{6}$$

[1]Here we use the rule $\text{vec}\,(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}\,(\mathbf{B})$.



(a) receptive fields, $N_s = 64$ (b) receptive fields, $N_s = 256$

Fig. 3. Receptive fields ($\tilde{\mathbf{S}}$) found by the proposed method for different N_s values – horizontal and vertical axes represent pixels, the center of mass of each receptive field is shown in a voronoi partition. We can see that if the number of receptive fields available N_s is higher, each one covers a smaller portion of the action space. The receptive fields near the border are substantially larger for $N_s = 64$ than for $N_s = 256$.

and collect (column-wise) all the samples $\text{vec}\,(\mathfrak{P}^{\mathbf{m}})$ in a single matrix $\mathfrak{P} = \left[\text{vec}\left(\mathfrak{P}^{\mathbf{1}\cdots\mathbf{n_M}}\right)\right]$. This matrix obeys:

$$\mathfrak{P} = \underbrace{(\tilde{\mathbf{S}}^T \otimes \tilde{\mathbf{S}}^T)\tilde{\mathbf{P}}}_{\mathcal{P}}\tilde{\mathbf{M}}[\mathbf{q}^{\mathbf{1}} \dots \mathbf{q}^{N_m}],$$

where $\mathcal{P}$ $(N_s \cdot N_s \times N_m)$ is a set of N_m predictors in the sensory input space. This can be decomposed into a product of two matrices resorting to NMF as defined in equation (5).

$$\underbrace{\mathfrak{P}}_{\mathbf{V}} \approx \underbrace{\mathcal{P}}_{\mathbf{W}}\underbrace{\tilde{\mathbf{M}}[\mathbf{q}^1 ... \mathbf{q}^{N_M}]}_{\mathbf{H}}.$$

We solve this equation using the method from [15] that induces sparsity in $\mathbf{W}$.

We can see in Figure 2 the solutions for $\tilde{\mathbf{M}}$ for different numbers of motor fields N_m. This illustrates how the motor fields spread across each degree of freedom and how some degrees of freedom can be more important for predicting sensory input than others. We can see for this particular data that rotation is much more important than zoom since the motor fields are spread mostly along the rotation axis.

Where the last equation evidences how we factor $\mathfrak{P}$ into $\mathbf{WH}$ by solving the NMF problem defined in (5) with N_m as the parameter controlling the dimensionality. Once factored we recover $\tilde{\mathbf{M}}$ by multiplying $\mathbf{H}$ by $[\mathbf{q}^1 ... \mathbf{q}^{N_m}]^+$ [2]:

$$\tilde{\mathbf{M}} = \mathbf{H}[\mathbf{q}^1 ... \mathbf{q}^m]^+, \tag{7}$$

where $\mathbf{A}^+$ is the pseudo-inverse of $\mathbf{A}$.

C. Find the sensory fields S

Starting from $\mathcal{P}$ obtained in the previous step, which we now factor as

$$\mathcal{P} = (\tilde{\mathbf{S}}^T \otimes \tilde{\mathbf{S}}^T)\tilde{\mathbf{P}},$$

[2]This matrix is known at sampling time and we assume it is full rank due to the nature of the $\mathbf{q}^m$.

we can write for each column i of $\mathcal{P}$ [3] containing a prototype predictor in the input sensor space for a single motor field:

$$\mathcal{P}_i = (\tilde{\mathbf{S}}^T \otimes \tilde{\mathbf{S}}^T)\tilde{\mathbf{P}}_i$$
$$[\mathcal{P}_i]_\square = \tilde{\mathbf{S}}^T[\tilde{\mathbf{P}}_i]_\square\tilde{\mathbf{S}}$$

where $\tilde{\mathbf{S}}$ ($N_s \times N_S$) contains the sensory fields and $\tilde{\mathbf{P}}_i$ represents a prototype predictor in sensor field space for a single motor field. Here $\tilde{\mathbf{S}}^T$ works as the reconstruction operator in an autoencoder framework [16].

We rearrange in matrix form and decompose using NMF as defined in equation (5).

$$\underbrace{\begin{bmatrix}[\mathcal{P}_1]_\square \\ [\mathcal{P}_2]_\square \\ \vdots\end{bmatrix}}_{V} \approx \underbrace{\begin{bmatrix}\tilde{\mathbf{S}^T}[P_1]_\square \\ \tilde{\mathbf{S}^T}[P_2]_\square \\ \vdots\end{bmatrix}}_{W} \underbrace{\tilde{\mathbf{S}}}_{H}$$

this evidences the factorization operation [15] (that induces sparsity in $\mathbf{W}$) where $\mathbf{H}$ yields $\mathbf{S}$.

We can see in Figure 3 the solutions for $\tilde{\mathbf{S}}$ for different numbers of receptive fields N_s. The receptive fields are smaller in the center than in the periphery showing that the center of the image is more relevant to achieve a smaller prediction error for this particular data.

D. Find P

We finally obtain $\tilde{\mathbf{P}}$ by taking $\mathbf{W}$ from the previous step which we rearrange into $\mathbf{W}'$ such that

$$\mathbf{W}' = \left[\tilde{\mathbf{S}}^T[\tilde{\mathbf{P}}_1]_\square\tilde{\mathbf{S}}^T[\tilde{\mathbf{P}}_2]_\square \cdots\right]$$

we then decompose it as $\mathbf{W}' \approx \tilde{\mathbf{S}}^T\mathbf{X}$ by solving a nonnegative least squares problem

$$\underset{\mathbf{X}>0}{\operatorname{argmin}}\|\mathbf{W}' - \tilde{\mathbf{S}}^T\mathbf{X}\|_F^2$$

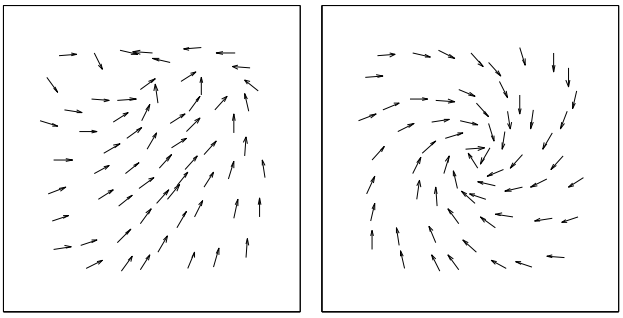
where $\|\|_F^2$ is the Froebenius norm and $\mathbf{X}$ contains

$$\mathbf{X} = \left[[\tilde{\mathbf{P}}_1]_\square^T[\tilde{\mathbf{P}}_2]_\square^T \cdots\right]$$

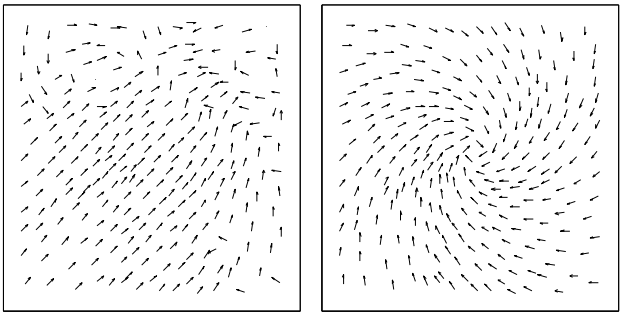
from which, after some index manipulation, we finally obtain

$$\tilde{\mathbf{P}} = \left[\tilde{\mathbf{P}}_1\tilde{\mathbf{P}}_2 \cdots\right].$$

We can see the predictors which more are suited for two particular actions in Figure 4.



(a) motion fields, $N_s = 64, N_m = 64$



(b) motion fields, $N_s = 256, N_m = 256$

Fig. 4. Motion fields ($\tilde{\mathbf{P}}$) found by the proposed method for different N_s and N_m – horizontal and vertical axes represent pixels. Each arrow origin is placed in the center of mass of each receptive field. The arrows are pointing to the receptive field which has a stronger activation produced by the originating receptive field for that prototype predictor ($\tilde{\mathbf{P}}_\mathbf{i}$), the arrow length is constant. On the left figures we show the prototype predictor $\tilde{\mathbf{P}}_\mathbf{i}$ which is more strongly activated (according to $\tilde{\mathbf{M}}$) for the action $(12, 12, 1, 0)$ representing a horizontal and vertical translation of 12 pixels. On the right we plot the same for the action $(0, 0, 1, 90°)$ representing a $90°$ rotation centered in the axis without zoom. We can see that prototype predictors are quite specialized.

IV. Experiments

We will start by comparing the performance of explicitly minimizing the objective function (see Equation (1)) using a gradient method (coordinate descent algorithm) or to use the low-rank approximation of the high-rank forward model.

For each experiment we perform 10 independent runs (i.e. solve the optimization problem). We generate a triplet ($\mathbf{I_0}^m$, $\mathbf{I_1}^m$,$\mathbf{q}^m$) for each of the N_M actions covering 4 degrees of freedom (horizontal and vertical translations, rotation and zoom scale factor) and a set of actions for each degree of freedom. Each triplet contains 100 independent samples of the same action.

The agent is equipped with a square retina of N_S pixels which is used to acquire grayscale images with intensity ranging from 0 to 1. Triplets used ($\mathbf{I_0}^m$, $\mathbf{I_1}^m$,$\mathbf{q}^m$) are obtained from a 2448 by 2448 pixels image as environment (forest image, see Figure III). The triplets simulate an agent placed in a random place in the environment and observing $\mathbf{I_0}^m$, then doing an action $\mathbf{q}^m$ and after doing the action observing $\mathbf{I_1}^m$.

We start by evaluating the quality of the method proposed in Section III to find a low-rank forward model to solve the sensorimotor optimization problem defined in Equation (1). We do so by comparing the objective function value using a coordinate descent gradient algorithm when starting from either a random initialization or from the proposed method. The rationale for the comparison is to evaluate if the approximation is useful in solving the original problem (see Equation (1)) and also if it can be used as a warm-start when the best possible solution is needed. While the initialization using the proposed method (see Section III) already incurred in a computational cost (equivalent to very few gradient iterations) it is absolutely negligible relatively

[3] Here we use the rule vec $(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})$vec $(\mathbf{B})$ in the reverse direction.

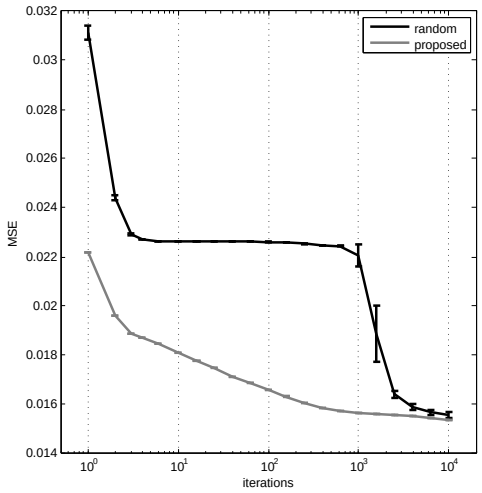


Fig. 5. Average MSE and standard deviation (over 10 runs) per pixel in predicted image after a variable number of gradient iterations starting from a random initialization (random) and using the proposed low-rank approximation as the initialization (proposed). We can see that the proposed method is useful in minimizing the original cost function (Equation 1) since after a few gradient iterations we reach a value which is on the range of the best attainable.

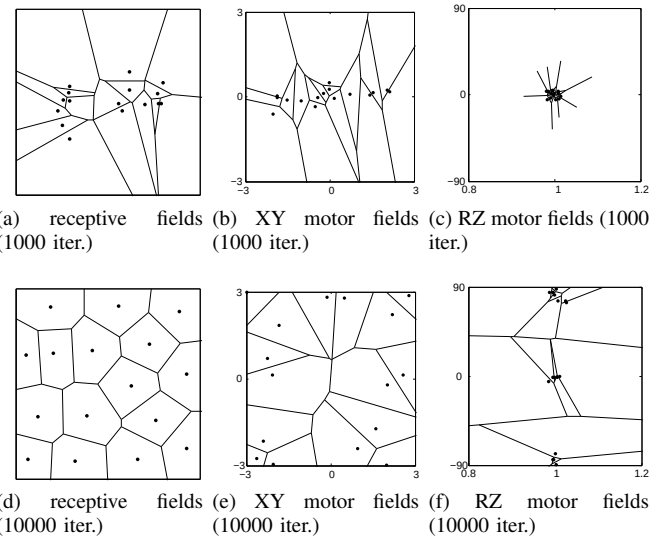
to the number of iterations needed for the gradient descend method to reach a solution.

We use a retina of 9 by 9 pixels ($N_S = 81$) and motor commands with 4 degrees of freedom and 3 actions per degree of freedom ($N_M = 81 = 3^4$) being the actions $\{-3, 0, 3\} \times \{-3, 0, 3\} \times \{-90°, 0°, 90°\} \times \{0.8, 1, 1.2\}$ (XY translations, rotation and zoom respectively) .

We can see in Figure 5 that in terms of objective function value there is huge advantage not only in the first iterations but that this advantage remains highly significant until convergence as shown in Figure in 5 . The speed up provided by the proposed method to reach a near optimal solution comes at the cost of a few iterations and is almost negligible relatively to the total number of iterations that the gradient method requires when starting from a random initialization. Moreover if we look at the overall organization of the intermediate solutions we can see that the difference is even more striking. The intermediate solution for the gradient method (with random initialization) even after 1000 iteration (see Figure 8a and 6a) is still far from the final solution (see Figure 8b and 6e). The solution found by the proposed method (see Figure 9a and 7a) is almost identical to the final solution (see Figure 9b and 7d).

V. Conclusion

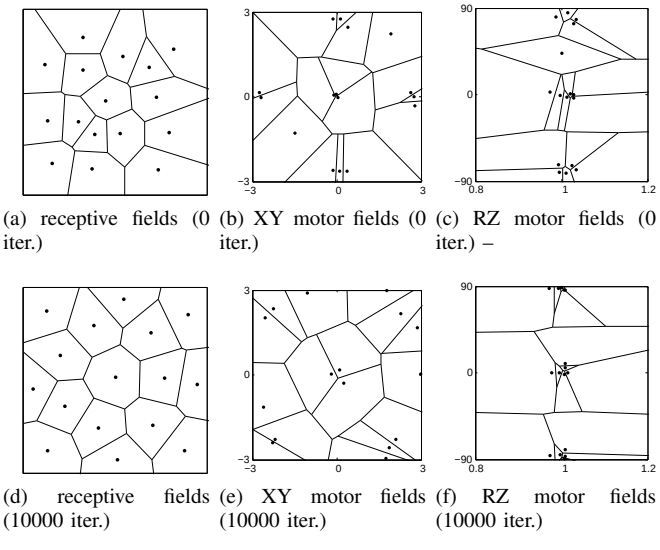
We presented a novel perspective and method for the self-organization of sensorimotor systems and forward models by showing its relation to an extensively studied problem in the literature. We have shown by uncovering the structure of both the sensor and motor inputs that this approach can solve complex problems (i.e. with many sensors and motor actuators). The results show a promising path for the development of more efficient robotic sensorimotor systems which are adapted to specific tasks and particular environments. In the



(a) receptive fields (1000 iter.) (b) XY motor fields (1000 iter.) (c) RZ motor fields (1000 iter.)

(d) receptive fields (10000 iter.) (e) XY motor fields (10000 iter.) (f) RZ motor fields (10000 iter.)

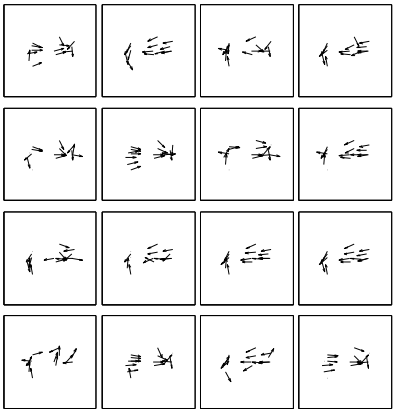
Fig. 6. Intermediate and final receptive and motor fields for the gradient method starting from a random initialization – horizontal and vertical axes represent the actions for each degree freedom (i.e. horizontal and vertical translations for XY; zooms and rotations for RZ respectively). The center of mass of each receptive field is shown in a voronoi partition. The center of mass of each motor field is shown in a voronoi partition. Each center of mass is defined by a 4 dimensional point (4 degrees of freedom) which we group for visualization into horizontal and vertical translation degrees of freedom (XY) and rotation and zoom (RZ). Each degree of freedom values are scaled to a common reference to have equal influence in the space partition. We can see only after 1000 iterations some structure starts to appear while the motor fields regarding the RZ degrees of freedom have still to differentiate.



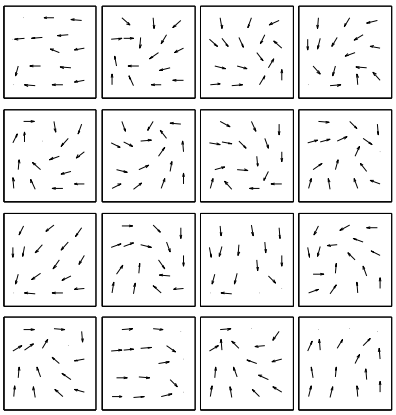
(a) receptive fields (0 iter.) (b) XY motor fields (0 iter.) (c) RZ motor fields (0 iter.) –

(d) receptive fields (10000 iter.) (e) XY motor fields (10000 iter.) (f) RZ motor fields (10000 iter.)

Fig. 7. Intermediate and final receptive and motor fields for the gradient descent using the proposed approximation as initialization – horizontal and vertical axes represent the actions for each degree freedom (i.e. horizontal and vertical translations for XY; zooms and rotations for RZ respectively). The center of mass of each receptive field is shown in a voronoi partition. The center of mass of each motor field is shown in a voronoi partition. Each center of mass is defined by a 4 dimensional point (4 degrees of freedom) which we group for visualization into horizontal and vertical translation degrees of freedom (XY) and rotation and zoom (RZ). Each degree of freedom values are scaled to a common reference to have equal influence in the space partition. We can see that solution found by the proposed method (0 iter.) is already organized and very similar to the final solution after 10000 gradient iterations.

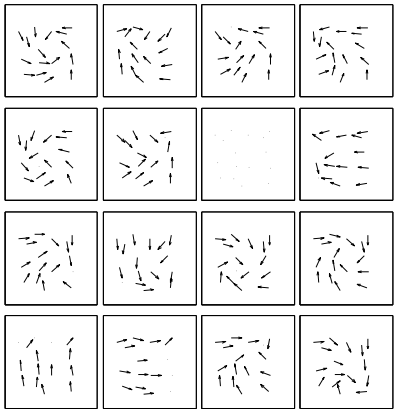


(a) motion fields (1000 iter.)

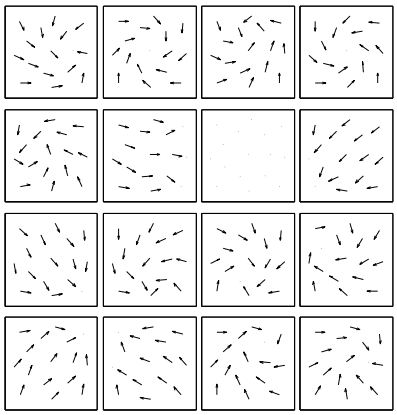


(b) motion fields (10000 iter.)

Fig. 8. Intermediate and final motion fields for the gradient descent starting from a random initialization (horizontal and vertical axes represent pixels).



(a) motion fields (0 iter.)



(b) motion fields (10000 iter.)

Fig. 9. Intermediate and final motion fields for the gradient descent using the proposed method as initialization (horizontal and vertical axes represent pixels).

future we plan to demonstrate the usefulness of the method by applying it in tasks such as the detection of external movement during self-motion in unstructured environments.

ACKNOWLEDGMENTS

This work was supported by the FCT projects BIOMORPH-EXPL/EEI AUT/2175/2013 and Pest-OE/EEI/LA0009/2013 and also by EU Project POETICON++ [FP7-ICT-288382].

REFERENCES

[1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2002.

[2] P. Cisek and J. F. Kalaska, "Neural mechanisms for interacting with a world full of action choices," *Annual review of neuroscience*, vol. 33, pp. 269–298, 2010.

[3] R. H. Wurtz, "Neuronal mechanisms of visual stability," *Vision research*, vol. 48, no. 20, pp. 2070–2089, 2008.

[4] R. Santos, R. Ferreira, A. Cardoso, and A. Bernardino, "Sensorimotor networks vs neural networks for visual stimulus prediction," in *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014 Joint IEEE International Conferences on*, Oct 2014, pp. 287–292.

[5] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.

[6] P. O. Hoyer, "Non-negative sparse coding," in *Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on*. IEEE, 2002, pp. 557–565.

[7] T. B. Crapse and M. A. Sommer, "Corollary discharge across the animal kingdom," *Nature Reviews Neuroscience*, vol. 9, no. 8, pp. 587–600, 2008.

[8] R. C. Miall and D. M. Wolpert, "Forward models for physiological motor control," *Neural networks*, vol. 9, no. 8, pp. 1265 – 1279, 1996.

[9] L. A. Olsson, C. L. Nehaniv, and D. Polani, "From unknown sensors and actuators to actions grounded in sensorimotor perceptions," *Connection Science*, vol. 18, no. 2, pp. 121 – 144, 2006.

[10] L. Lichtensteiger and P. Eggenberger, "Evolving the morphology of a compound eye on a robot," in *Third European Workshop on Advanced Mobile Robots, 1999. (Eurobot '99) 1999* , 1999, pp. 127 – 134.

[11] J. Ruesch, R. Ferreira, and A. Bernardino, "A computational approach on the co-development of artificial visual sensorimotor," *Adaptive Behavior*, vol. 21, no. 6, pp. 452 – 464, 2013.

[12] C. Paul, "Morphological computation: A basis for the analysis of morphology and control requirements," *Robotics and Autonomous Systems*, vol. 54, no. 8, pp. 619 – 630, 2006.

[13] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *The Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[15] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 689–696.

[16] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 8, pp. 1798–1828, 2013.