Procedia Computer Science
Volume 88, 2016, Pages 259–271

7th Annual International Conference on Biologically Inspired
Cognitive Architectures, BICA 2016

Cartesian Abstraction Can Yield ‘Cognitive Maps’

András Lőrincz

Eötvös Loránd University, Budapest, Hungary

Abstract

It has been long debated how the so called cognitive map, the set of place cells, develops in rat hippocampus. The function of this organ is of high relevance, since the hippocampus is the key component of the medial temporal lobe memory system, responsible for forming episodic memory, declarative memory, the memory for facts and rules that serve cognition in humans. Here, a general mechanism is put forth: We introduce the novel concept of Cartesian factors. We show a non-linear projection of observations to a discretized representation of a Cartesian factor in the presence of a representation of a complementing one. The computational model is demonstrated for place cells that we produce from the egocentric observations and the head direction signals. Head direction signals make the observed factor and sparse allothetic signals make the complementing Cartesian one. We present numerical results, connect the model to the neural substrate, and elaborate on the differences between this model and other ones, including Slow Feature Analysis [17].

Keywords: cognitive map, Cartesian factorization, autoencoder, comparator hypothesis

1 Introduction

What is mammalian intelligence about? What are the key components necessary for the scientific and technological progress of mankind in the last 20,000 years or so? Are those very special to the human race, or for primates, or for mammals? Recent review suggests that the basic mechanisms, or algorithms are similar in rats and humans [8]. In their paper, Buzsáki and Moser propose that planning has evolved from navigation in the physical world *and* that navigation in real and mental space are fundamentally the same. They also underline the hypothesis that the entorhinal cortex and the hippocampus, the hippocampal entorhinal complex (EHC) support navigation and memory formation.

The importance of this complex was discovered many years ago [44]. Now, it is widely accepted that the EHC is responsible for episodic memory, see, e.g., [46] and [35] for an earlier review and for a recent one, respectively. The intriguing puzzle is that people are able to describe autobiographic events, can discover rules, in spite of the many dimensional inputs, such as the retina (millions of photoreceptors), the ear (cca. 15,000 inner plus outer hair cells), the large number of chemoreceptors as well as proprioceptive, mechanoreceptive, thermoceptive and nociceptive sensory receptors. This looks like an impossible mission, since the number of the sensors influence the size of the state space observed in the exponent and make it enormous. This number is gigantic even if the basis of exponent is only two. How is it possible to remember for anything in such a huge space?

Selection and peer-review under responsibility of the Scientific Programme Committee of BICA 2016

doi:10.1016/j.procs.2016.07.433

An illuminating observation to us is the fact that the brain develops low dimensional topographic maps manifested by retinotopy in the visual system, tonotopy in the auditory system, somatotopy in the somatosensory system, among many others. Such maps are related to some metric of the sensed space, being visual, auditory, or body related. The dimensionality of the maps are low unlike the number of the sensors that give rise to the map. One may say that (i) the representation, i.e., the topographic map, discretizes a low dimensional variable, (ii) both the space and the actual state are described by such variables, and (iii) the variables are like Cartesian coordinates at different cognitive levels, examples including the 'where' and 'what' system, or, the form and the color of an object, or, the face and facial expressions, or, the position and the direction of an animal in the world among others. In turn, we distinguish two factor types

Type 1 factors: These factors make no (or minor) assumptions on each other. Non-negative matrix factorization (NMF), for example, originates from chemistry: it is used in mass spectrometry and radiology among other fields, where absorbing or radiating components sum up. In a given environment and for a given detector system, the presence of different Radon isotopes depends on the environment and the detector, but they do not influence each others spectrum except that to a good approximation they sum up.

Type 2 factors: These factors assume each other and they are supposed to characterize objects and episodes. For example, texture, shape, weight, material components are all relevant when considering the value of a sword like a Damascus Khanjar. Nonetheless, the set of such factors is insufficient for providing a full description of the state of a sword; the state of the atoms or molecules. On the other hand, these Cartesian factors can give a fairly good and highly compressed description, suitable for communicating the usage and the value of the sword.

Sensory information brings about two interdependent task types: Information fusion and the related pattern completion make the first type. For example, when searching for food, the animal may use either visual information, or smell, or both. Thus, information fusion is about the Cartesian product of sensory modalities and pattern completion occurs in this product space. This task can be accomplished efficiently with sparse compressed sensing methods [13] that may have neuronal implications [42]. The other task is the formation of new, low dimensional maps, at least in a discretized form that are *not* having dedicated sensors. This is a kind of abstraction, when irrelevant details, judged from the point of view of the tasks, are cleaned off.

The concept of numbers is such an abstraction; it eliminates material properties. Two plus two equals four, no matter if we are concerned with apples or peaches. Such abstractions enable concise formulations for many similar tasks. The world of numbers is one dimensional and material properties are orthogonal to it, so it is like a coordinate system: quantity is one coordinate and the material substance is another. We are concerned with such generalized Cartesian factors.

Geometry uses concepts like points and straight lines, disregards physical interactions, which can be described by a different set of Cartesian factors. Geometry can serve many tasks including the computation of homing distance [9] and the prediction of the paths of planets. Cartesian coordinates or concepts are typically low dimensional. In the brain, the computations of Cartesian factors, like the estimation of homing distance, should be based on high dimensional sensors, including visual and vestibular information and the representation of novel Cartesian factors that are derived from those.

Laboratory coordinate system also called allothetic representation is the landmark of geometry representation. Such representation appears already in rats and supports path planning and navigation, possibly because it detaches the egocentric direction from the other parameters of the environment. The allothetic description is robust against certain changes in the environment, e.g., (i) the abstraction can be used both in light and in dark and (ii) it can be used efficiently in obstacle avoidance, since it is not subject to occlusions as opposed to visual observations. In addition, such abstract concepts offer highly compressed Cartesian description of facts and episodes.

Here, we put forth a novel method for the development of a Cartesian factor. We demonstrate the method in the derivation of the allothetic representation of the world; we develop a set of direction independent *place cells* [38] from sensory information. The set of place cells is called '*the cognitive map*' [39], because they can be used in path planning. We will exploit direction sensors called head direction cells, an allothetic signal (see, e.g.,[53] and the references therein) together with idiothetic visual information. We will apply deep learning methods in an autoencoder to derive direction insensitive place cells. Due to the autoencoder part of our model, we can connect our method to the comparator hypothesis of the hippocampus [58], where place cells are abundant.

Both the mathematical background and the demonstrative examples have relevance in goal oriented framework and, in part, they have been submitted elsewhere [31]. Here, we review them and embed the results into the neurobiological context. In turn, our contribution is a meta-level functional model of the entorhinal-hippocampal neuronal circuits that may shed light to the algorithmic principles behind the development of discretized low dimensional representations that can support cognition.

The paper is organized as follows. In the next section (Sect. 2), we review related works, including algorithmic models that can produce place cells and distinct findings in neuroscience that indicate the required components for developing such cells. Section 3) deals with our computational architecture capable of developing cognitive maps, while satisfying relevant constraints of neuroscience. Section 4 lists our results followed by the discussion in Sect. 5. Conclusions are drawn in Sect. 6. The Appendix contains the mathematical details.

2 Related Works

The number of place cell models is considerable. Neural representation of trajectories travelled and the connectivity structure developed during such paths have been suggested as the method for place cell formation [41]. Sensory information includes external cues and internally generated signals. They are fused to develop place cells in [2]. Place cells can be derived [45] by linear combinations of entorhinal grid cells [18] and vice versa, neuronal level model can derive grid cell firing from place cell activities [7]. Time plays the key role in the slow feature analysis model of place cells [17, 43]. Time does not play a role in the independent component analysis based place cell model, except in the novelty detection step of the autoencoder of the model [30, 32].

We believe that all of these models, i.e., navigation based models, models based on interaction between representations, models that search for components that change slowly in time, and models that consider novelty detection, have their merits: the development of low-dimensional representation of Cartesian factors is hard and all possible clues should be exploited for developing them. Beyond learning, the different mechanisms can be used during task execution: navigation in partially observed environments, like the Morris maze or when in dark, can be supported by temporal integration, or novelty detection may support the separation of a rotating platform from remote, non-rotating cues [48, 49].

Another point of reference comes from neuroscience indicating that both the entorhinal grid representation and the place cell representation of the hippocampus depend strongly on the vestibular information. It has been a long standing question what comes first, namely if the idiothetic head direction representation, the representation of place cells, or the representation of grid cells is the prerequisite of the others. For example, grid cells require hippocampal input [4]. There are indications [60] that head direction cells may be critical for grid cell formation, but it may not be so for place place cells since those can be controlled by environmental cues, like visual landmarks. However, very recently, it turned out that the disruption of head direction cells can impair grid cell signals and are thus crucial for the formation of the allothetic representation including both place cells and grid cells [59].

It is clear that information both from the environment and self-motion should be combined for an efficient and precise neural representation of space [16] and that different sensory signals can serve the purpose. Object recognition is similar in many respects;, it can exploit different mechanisms such as stereoscopic information, structure from motion, shape from shading, gradient of the optica texture, and occlusion contours, for example, all being important for disambiguation of the visual information under different conditions [55].

Due to the critical nature of the vestibular input, our goal is to derive place cells under the assumption that only this component of the Cartesian representation, namely the egocentric direction relative to an allothetic coordinate system is available and we ask if the allothetic representation of space can be derived by using only (i) the egocentric direction and (ii) the egocentric visual information.

3 Methods

3.1 Autoencoder

An autoencoder is the unsupervised version of the Multilayer Perceptron (MLP) and may have *deep* versions [22, 57]. For the sake of general formulation, the deep version is described below although deep studies are limited to a single case.

Consider a series of non-linear mappings (layers) of the form:

$$\boldsymbol{H} = f_N\Big(\cdots f_2\big(f_1(\boldsymbol{X}\boldsymbol{W}_1)\boldsymbol{W}_2\big)\cdots \boldsymbol{W}_N\Big), \tag{1}$$

where $\boldsymbol{X} \in \mathbb{R}^{I\times J}$ is the matrix of I inputs of size J, $\boldsymbol{W}_n \in \mathbb{R}^{Q_{n-1},Q_n}$ are parameters with $Q_0 = J$, and f_n are non-linear almost everywhere differentiable element-wise functions ($n = 1,\dots,N; N \in \mathbb{N}$). Then $\boldsymbol{H} \in \mathbb{R}^{I\times Q}$ is called the feature map ($Q_N = Q$). Typically, one takes two such mappings with reversed sizes — an encoder and a decoder — and composes them. Thereupon one can define a so-called reconstruction error between the encoder input $\boldsymbol{X}$ and the decoder output $\widehat{\boldsymbol{X}} \in \mathbb{R}^{I\times J}$ (normally the ℓ_2 or Frobenius norm of the difference, i.e., $\frac{1}{2}\|\boldsymbol{X}-\widehat{\boldsymbol{X}}\|_F^2 = \frac{1}{2}\sum_{i=1,\dots,I}\sum_{j=1,\dots,J}(\boldsymbol{X}_{i,j}-\widehat{\boldsymbol{X}}_{i,j})^2$) and try to find a local minimum of it in terms of parameters $\boldsymbol{W}_n$ after random initialization, by taking advantage of a step-size adaptive mini-batch subgradient descent method [14, 61, 28]. The non-linearity can be chosen to be the rectified linear function $f_n(x) = x\cdot\mathbb{I}(x>0)$ for $x \in \mathbb{R}$ [36, 11] to avoid the vanishing gradient problem [24, 25], where $\mathbb{I}$ designates the indicator function.

3.2 Spatial and Lifetime Sparsity

Deep Autoencoders are often used as a pretraining scheme [15] or as a part of supervised algorithms [40], but they lack the ability to learn a meaningful or simple data representation without prior knowledge [50]. To obtain such a description, one might add regularizers or constraints to the objective function [20, 3], or employ a greedy procedure [56, 12]. It is well known that minimizing the sum of ℓ_2 norms of parameters $\boldsymbol{W}_n$ can reduce model complexity by yielding a dense feature map, and similarly, the ℓ_1 variant may result in a sparse version [54, 37]. An alternative possibility is to introduce constraints in the non-linear function f_n. For example, one may utilize a k-sparse representation by keeping solely the top k activation values in feature map $\boldsymbol{H}$, and letting the rest of the components zero [33]. This case, when features, i.e., the components of the representation, compete with each other is referred to as *spatial sparsity*. Input indices of the representation may also go up against each other and this case is called *lifetime sparsity* [34].

3.3 Problem Formulation

We assume that a latent random variable Z (e.g., the discretized allothetic representation of the state, that is, the place cells) and an observed random variable Y (e.g., the head direction, that is, a compass) are continuous and together they fully explain away – by means of saved memories – another observed binary random variable X (e.g., the egocentric view with pixel values either one or zero taken in the direction of the head). The ranges of Z and Y are supposed to be grid discretized finite r- and one-dimensional intervals, respectively. For more details, see Fig. 1 and the Appendix.

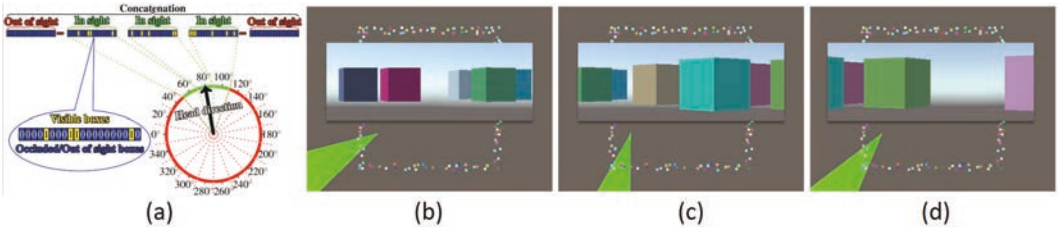

Figure 1: Numerical experiment. (a): input is concatenated from sub-vectors, which belong to different allothetic directions. A given index corresponds to the same box, the ‘remote visible cue’, in all sub-vectors. The value of the a component of a sub-vector is 1 (0) if the box is visible (non-visible) in the corresponding direction. Three directions are visible (green). Some boxes may be present in neighboring sub-vectors, since they are large. (b)-(d): the ‘arena’ from above with the different boxes around it. Shaded green areas in (b), (c), and (d), show the visible portions within the field of view at a given position with a given head direction. Insets show the visual information for each portion to be transformed to 1s and 0s in the respective components of the sub-vectors. Components of out-of-view sub-vectors are set to zero.

3.4 Demonstration

For our study, we generated a squared ‘arena’ surrounded by $d = 150$ boxes (Fig. 1). The ‘arena’ had no obstacles. Boxes were placed pseudo-randomly: they did not overlap. The ‘arena’ was discretized by an $M \times M = 36 \times 36$ grid. From each grid point and for every 20°, a 28° field of view was created (i.e., $L = \frac{360^{\circ}}{20^{\circ}} = 18$, overlap: 4° between regions), and the visibility — a binary value (0 for occlusion or out of the angle of view) — for each box was recorded, according to Eq. (2); we constructed a total of $I = 37 \cdot 37 \cdot 18 = 24{,}642$ binary ($\boldsymbol{x}^{(m,l)}$) vectors.

These vectors were processed further. Beyond the actual direction of the center of the viewing angle, we introduced some degree of closeness about the input regarding the direction, but not the position: we varied the viewing angle between 28° and 360°. Formally, for various experiments, we defined masks $\boldsymbol{V}_{i,\cdot}$ summing to $v = 1, 3, \ldots, 17, 18$, for which we carried out the concatenation method (see, Fig. 2 below and Eq. (3) in the Appendix): for each visible 20° sector, while for all non-visible sectors, an all-zero vector were appended. $\mathbb{R}^r$. The procedure is summarized in Fig. 2.

In some experiments we normalized the inputs to unit ℓ_2 norm for each $d = 150$ dimensional components, provided that at least one of the components differed from zero. This is called normalized experiment. We used spatial sparsification with $k = 1$ and lifetime sparsification with $p = 3.33\%$ and $p = 6.66\%$. In the error of the autoencoder we considered two options: (a) error of the full output and (b) error only on the visible components that belonged to the viewing angle as in Eq. (4). This latter is called masked experiment. We used them in combination. We also tried 3 and 5 layer autoencoders,

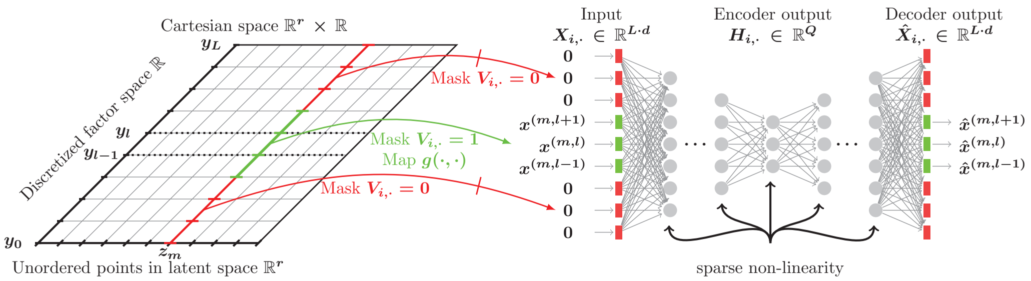

Figure 2: General architecture. In the numerical experiments the notations correspond to the following quantities: Z latent positions, Y discretized 'compass' values. Input to the network: red: not visible, green: visible, i.e., the viewing angle range of the viewing direction and some neighboring viewing angle ranges. Each viewing angle range provides inputs about all boxes visible within that range. No topography is assumed within the ranges; box indices are ones if they are visible and zero otherwise. The full input equals to the 'No. of boxes × No. of viewing angle ranges'.

with the middle layer representing the latent variables.

The size of the middle layer was always $Q = 30$. This means that spatial (i.e., latent component-wise) sparsity gave rise to 3.33% lifetime sparsity. On the other hand, $p = 3.33\%$ lifetime sparsity was effectively larger than 3.33% since it was possible that none of the latent unit was selected for a given input (and thus all of them were set to zero), when backpropagation became ineffective. The same holds for $p = 6.66\%$ lifetime sparsity, which, on the average, would give rise to 1, 2, 3, or more non-zero latent units with average above 2. The sizes of the hidden layers were spaced linearly between 2,700 and 30 for the 5 layer autoencoder $(2700,\ 1335,\ 30,\ 1335,\ 2700)$.

4 Results

The dependencies of the responses in the hidden representation vs. space and direction are shown in Fig. 3 and Fig. 4. Linear responses of randomly selected latent units for different algorithms are depicted in Fig. 3, illustrating the extent that the responses were localized. Figure 4 shows the direction (in)dependence of the responses. For each input, we chose the highest activity latent component and in each position we computed the number of directions (out of the 18 possible) that a neuron was the winner in the dataset. We computed these numbers for all units, selected the largest values at each position and combined them in a single figure that we color coded (Fig. 4): Black color at one position means that unit won in all angles at that position. The lighter the colors, the smaller the number of winning directions is.

One should ask (i) if the linear responses are local and activities far from the position of the peak activity are close to zero; (ii) if the number of dead latent units is small, (iii) if responses are direction independent, that is, if we could derive the discretization of space in allothetic coordinates, the complementary component of the egocentric direction. We found that spatial sparsity with the 3 layer network and the 5 layer network with dense 2^{nd} and 4^{th} layers rendered the output of some or sometimes all hidden units to zero (Table 1). On the othe hand, lifetime sparsity with the 5 layer network produced excellent results. Lifetime sparsity $p = 6.66\%$ can still produce place fields. Note that local responses appear without the mask, but only for very large viewing angles. For the sake of completeness, we also provide the ICA responses in Fig. 3. We discuss the relevance, the limitations and the promises below.

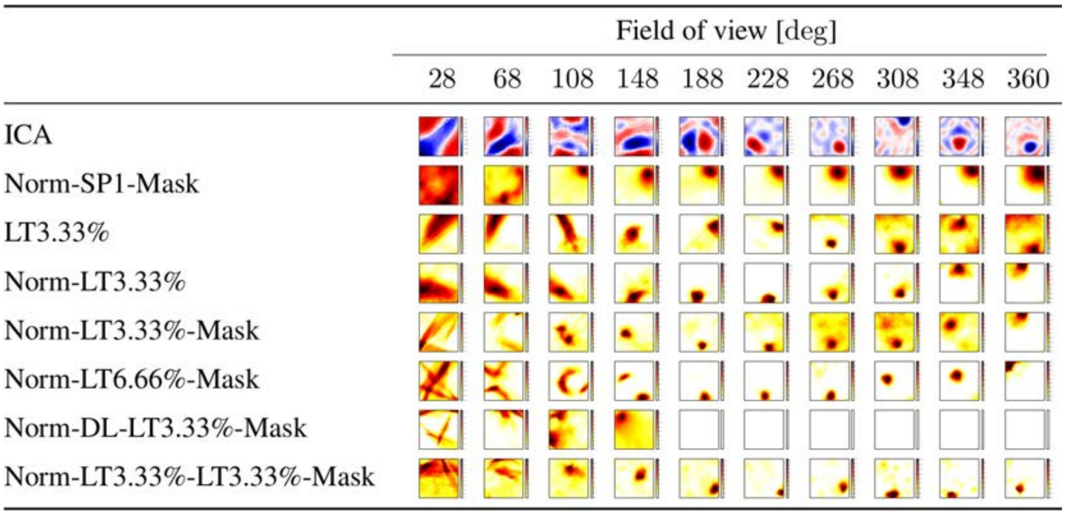

Figure 3: Linear responses of individual latent units selected randomly: we chose neuron with index 2 from the latent layer. ICA: values may take positive and negative values. Other experiments: all units are ReLUs, except the output, which is linear. Color coding represents the sum of responses for all directions at a given point. SP1: spatial sparsity with $k = 1$, LT3.3%: lifetime sparsity = 3.3%, Norm: for each 150 components, the ℓ_2 norm of input is 1 if any of the components is non-zero, Mask: autoencoding error concerns only the visible part of the scene, DL: dense layer. 'Norm-LT3.3%-LT3.3%-Mask' means normed input, masked error, 5 layers: input layer, 3 layers with LT sparsity equals 3.3%, output layer.

Table 1: Dead neuron count: number of non-responsive computational units.

	Field of view [deg]									
	28	68	108	148	188	228	268	308	348	360
Norm-SP1-Mask	2	0	5	5	10	12	16	18	15	18
LT3.33%	0	0	0	0	0	2	2	6	8	9
Norm-LT3.33%	0	0	0	1	1	3	2	4	9	11
Norm-LT3.33%-Mask	0	0	0	0	0	0	1	2	7	11
Norm-LT6.66%-Mask	0	0	0	0	0	0	1	4	13	13
Norm-DL-LT3.33%-Mask	0	3	1	29	30	30	30	30	30	30
Norm-LT3.33%-LT3.33%-Mask	0	0	0	0	0	0	0	0	0	0

5 Discussion

We elaborate on two aspects below. First, we consider temporal dynamics, integrated neural implementations, and control. Then, we relate our algorithm to the neural substrate.

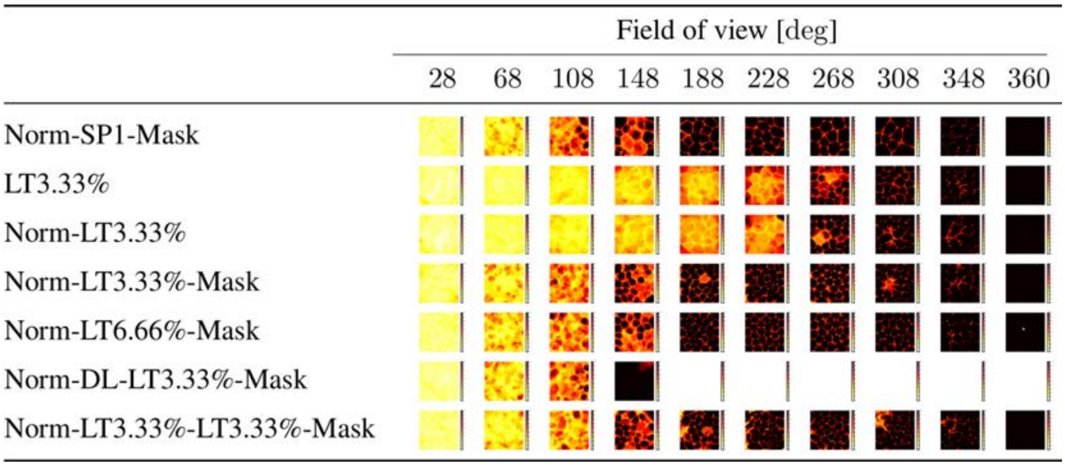

Figure 4: Angle independence. Notations are the same as in Fig. 3. The highest activity (winning) unit was selected for each input at each position in each direction. We counted the number of wins at each position for each unit and selected the largest number. Results are color coded. Black (18): there is a single winner for all angles at that position. White (0): no response at that point from any neuron in any direction. Values between 1 and 17: the darker the color the larger the direction independence for the best winner at that position.

5.1 General consideration

Our goal was to find a discretization of hidden Cartesian coordinates, an abstraction, provided that we already have the complementer one. Due to the nature of such coordinates, the one that we have must have metrics. The abstraction that we are after is similar to geometrical abstractions or algebraic abstractions: they cannot be sensed directly, so they are latent and they are Cartesian, i.e., they are like coordinates in an abstract space. In turn, they promote highly compressed description, since in problem solving the may eliminate irrelevant variables, an example being path planning that can be accomplished in an allothetic abstraction. This can be of high importance for goal-oriented activities, since reinforcement learning scales with the number of variables in the exponent [27, 5, 52] and the abstraction decreases that exponent.

Any metrics may show up in the temporal domain or, alternatively, it may manifest itself implicitly, via discretized spatial information (such as neighboring viewing ranges). Temporal information has been exploited by the SFA procedure [17]. It was found that in realistic conditions that include large viewing angles, direction independent place fields can be formed [17, 43] by means of temporal information.

Here, we wanted to neglect temporal information and, instead, to include some metrical one. In our demonstration, we included the Cartesian (metrical) factor; the head direction. We could go down to viewing angle of about 100°, or so. Further improvements can be expected for deeper networks. We found in our simulations that for deeper networks sparsity should be kept at least for some of the layers.

It is worth noting that we are not opposing the exploitation of temporal information, but it may not be available for every type of data. Temporal information, if available, should improve the capabilities. In this respect, our aim was to develop a general cognitive mechanism for developing Cartesian factors and the demonstration was inspired by the neural representation found in the entorhinal-hippocampal complex of rodents, where the temporal information might lower the minimal viewing angle needed for the formation of place cells further. Our model predicts that rats can develop place cells with viewing angles much lower than 300^0, which they have. Such large viewing angles may be necessary for observations, but they are not needed for the development of the cognitive map.

We highlight that high quality reconstruction is corrupted without the mask, i.e. if the autoencoder has to match the zero inputs, too. Large viewing angles (cca. 300^o) were needed for developing direction independent representation in our numerical experiments for this case (see rows in Fig. 4 not having 'Mask' marks).

Consider, path planning. Place fields uncover neighboring relations in linear mode by computing the correlations between the activities. The neighboring graph can be used for navigation in real and mental space, i.e., for control and for path planning. The neural architecture that can accomplish integrated control and path planning and suits place fields formulation has been described some years ago, see, e.g., [51] and the references therein.

Lifetime sparsity seems important in the development of place fields, whereas it can not be applied in real time. Real time operation requires spatial sparsity, or possibly some thresholding, or even the linear mode, since responses are fairly local for the linear mode for lifetime sparsity $p = 6.66\%$ and for the 5 layer network. Thresholded and non-thresholded modes can both uncover the neighbor relations supporting path planning. In turn, two types of operations may be favorable, one for learning off-line that enables lifetime sparsity, and one for developing the neighboring relations that can work real-time.

Before turning to the features that can be related to the neural substrate, we mention that a transformation of the result of a path planning procedure accomplished in an abstract allothetic coordinate system to idiothetic control *is needed* when the path is being traversed. This transformation is similar to the symbol grounding problem (for a review, see [21]) at each instant during execution with the specific property that the representation abstracted can have a metric. This specific feature, however, is not exploited in our algorithm, neither through time, nor through neighboring relations.

5.2 Relating structure and function: some features of the neural substrate

Our architecture requires cells that represent the direction of the head. Such head direction (HD) cells have been found in the brain. For a recent review, see [23] and the cited references. Our candidate is the so called 'head direction pathway' [1] that passes through the medial entorhinal cortex before reaching the hippocampus.

We were able to develop direction independent locally responding cells. Cells with similar properties are called place cells and they can be found in the CA1 and CA3 regions of the hippocampus, in the dentate gyrus, the parasubiculum the entorhinal and postrhinal cortices (see, [6] and the references therein).

The model of this paper needs neurons that can multiply and can produce conjunctive representations, e.g., between the visual cues and the head direction cells. There are at least three candidates for such computations

1. Deep networks should be able to implement the multiplicative function
2. Logical operations, such as the AND operation as well as others are made possible by synchronized synaptic inputs, called coincidence detection (for a recent review, see [47])
3. The interplay between distal and proximal dendritic regions when the proximal input enhances the propagation of the distal dendritic spikes can also support a multiplicative function [29, 26].

The EHC has sophisticated interconnections between distant and proximal regions [19] and this propagation enhancing mechanism listed above as the last option is our proposed candidate mechanism for place cell formation: We exploited this multiplicative feature in our representation by using the product space and using masking.

It has been suggested that the autoencoder mechanism of Vinogradova [58] should be modified: it is the full EHC that works as an autoencoder [30, 10]. We note that novelty detection can be used for the development of direction insensitive grids [32], a subject that we have not explored in the present studies. Temporal information with slow feature analysis have been used for developing place cells, head-direction cells, and spatial-view cells [17, 43]. These mechanisms may promote the learning of Cartesian factors.

6 Conclusions

One of the mysteries of intelligence is the development of novel concepts that are not directly sensed. For example, the concept of an infinitesimal point, a straight line, or the concept of numbers are such creatures. Such features that we call Cartesian factors originate from generalizations and enable one to solve complex problems in lower dimensional spaces. An example is the allothetic representation. It is of high relevance since path planning is simplified in such representation. It is known that some neurons in the brains of rats form allothetic codes; they are the place cells of the hippocampus.

We put forth a novel method for the development Cartesian factors and demonstrated it via the forming of place cells. We exploited the complementary information, the head direction cells. Our proposed cognitive mechanism does not work in the absence of such information. We note that upon destroying the vestibular system, which is critical for having head direction cells, no place cell is formed [53, 60].

Acknowledgments

This research was supported by the EIT Digital grant (Grant No. 16257) on *CPS for Smart Factories*. Computer runs of the demonstration are gratefully acknowledged to Zoltán Á. Milacski and András Sárkány.

Appendix: Problem Formulation

Assume that a latent random variable Z and an observed random variable Y are continuous and together they fully explain away another observed binary random variable X. The ranges of Z and Y are supposed to be grid discretized finite r- and one-dimensional intervals, respectively. We denote the resulting grid points by $(z^{(m)}, y^{(l)}) \in \mathbb{R}^r \times \mathbb{R}$; $l = 0, \ldots, L$, $m = 1, \ldots, (M+1)^r$, $L, M, r \in \mathbb{N}$. The indices $m = 1, \ldots, (M+1)^r$ are supposed to be scrambled throughout training (i.e., we assume no topology between $z^{(m)}$). Then observation $\boldsymbol{x}^{(m,l)} \in \{0,1\}^d$ is generated by a highly non-linear function $g\colon \mathbb{R}^r \times \{1, \ldots, L\} \to \{0,1\}^d$ from grid point $z^{(m)}$ and grid interval $[y^{(l-1)}, y^{(l)})$ as

$$\boldsymbol{x}^{(m,l)} = g(\boldsymbol{z}^{(m)}, l) \tag{2}$$

for $m = 1, \ldots, (M+1)^r$; $l = 1, \ldots, L$. For each fixed m, one is given masks $\boldsymbol{V}_{i,\cdot} \in \{0,1\}^L$; $\sum_{l=1}^{L} \boldsymbol{V}_{i,l} = v \in \mathbb{N}$ indexing pairs of the form $(l, \boldsymbol{x}^{(m,l)})$, where $i = 1, \ldots, I$ is a global index. Provided such a sample from Y and X, we aim to approximate the discretized version of Z.

We formulated the above problem as a multilayer feedforward *lifetime sparse autoencoding* [34] procedure with input matrix $\boldsymbol{X} \in \{0,1\}^{I \times J}$ utilizing two novelties: concatenated input vectors and a masked loss function are motivated by the input structure. In order to construct the inputs $\boldsymbol{X}_{i,\cdot}$; $i = 1, \ldots, I$ of size $J = L \cdot d$, we coupled each v-tuple of $\boldsymbol{x}^{(m,l)}$ vectors for fixed m into a single block-vector using the $\boldsymbol{V}_{i,\cdot}$ values as follows:

$$\boldsymbol{X}_{i,\cdot} = \left[\boldsymbol{V}_{i,1} \cdot \boldsymbol{x}^{(m,1)}, \quad \ldots, \boldsymbol{V}_{i,l} \cdot \boldsymbol{x}^{(m,l)}, \quad \ldots, \boldsymbol{V}_{i,L} \cdot \boldsymbol{x}^{(m,L)}\right]. \tag{3}$$

Then, we used the ℓ_2 reconstruction error as the loss, but on a restricted set of elements, namely, on the v non-zero blocks for each input:

$$l(\boldsymbol{X}, \widehat{\boldsymbol{X}}, \boldsymbol{V}) \colon = \frac{1}{I} \sum_{\substack{i=1,\ldots,I \\ j=1,\ldots,J}} \boldsymbol{V}_{i,\lfloor \frac{j-1}{d} + 1 \rfloor} \cdot (\boldsymbol{X}_{i,j} - \widehat{\boldsymbol{X}}_{i,j})^2 \tag{4}$$

where $\widehat{\boldsymbol{X}}$ denotes the output of the decoder network. Finally, a sparse non-linearity was imposed on top of each encoder layer, which selected the k percent topmost activations across one component. We applied both lifetime [34] and spatial sparsification [33]. Multilayer autoencoders with rectified linear units, $k = 1$ spatial sparsity, $p\%$-sparse lifetime sparsity, and linear decoder output layer make the non-linear units of the network.

References

[1] John Patrick Aggleton, Seralynne Denise Vann, Catherine JP Oswald, and M Good. Identifying cortical inputs to the rat hippocampus that subserve allocentric spatial processes: a simple problem with a complex answer. *Hippocampus*, 10(4):466–474, 2000.

[2] Angelo Arleo and Wulfram Gerstner. Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity. *Biological Cybernetics*, 83(3):287–299, 2000.

[3] Stephen R Becker, Emmanuel J Candès, and Michael C Grant. Templates for convex cone problems with applications to sparse signal recovery. *Math. Prog. Comp.*, 3(3):165–218, 2011.

[4] Tora Bonnevie, Benjamin Dunn, Marianne Fyhn, Torkel Hafting, Dori Derdikman, John L. Kubie, Yasser Roudi, Edvard I. Moser, and May-Britt Moser. Grid cells require excitatory drive from the hippocampus. *Nature Neuroscience*, 16(3):309–317, 2013.

[5] Craig Boutilier, Richard Dearden, and Moisés Goldszmidt. Stochastic dynamic programming with factored representations. *Artif. Intel.*, 121(1):49–107, 2000.

[6] Joel E Brown and Jeffrey S Taube. Neural representations supporting spatial navigation and memory. In *Representation and Brain*, pages 219–248. Springer, 2007.

[7] Neil Burgess and John OKeefe. Models of place and grid cell firing and theta rhythmicity. *Current Opinion in Neurobiology*, 21(5):734–744, 2011.

[8] György Buzsáki and Edvard I Moser. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, 16(2):130–138, 2013.

[9] Elizabeth R Chrastil, Katherine R Sherrill, Michael E Hasselmo, and Chantal E Stern. There and back again: hippocampus and retrosplenial cortex track homing distance during human path integration. *The Journal of Neuroscience*, 35(46):15442–15452, 2015.

[10] James J Chrobak, András LH øríncz, and György Buzsáki. Physiological patterns in the hippocampo-entorhinal cortex system. *Hippocampus*, 10(4):457–465, 2000.

[11] George E Dahl, Tara N Sainath, and Geoffrey E Hinton. Improving deep neural networks for LVCSR using rectified linear units and dropout. In *Acoust., Speech Sign. Proc. (ICASSP), 2013*, pages 8609–8613. IEEE, 2013.

[12] Wei Dai and Olgica Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *Info. Theo.*, 55(5):2230–2249, 2009.

[13] Marco F Duarte and Yonina C Eldar. Structured compressed sensing: From theory to applications. *Signal Processing, IEEE Transactions on*, 59(9):4053–4085, 2011.

[14] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12:2121–2159, 2011.

[15] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio. Why does unsupervised pre-training help deep learning? *J. Mach. Learn. Res.*, 11:625–660, 2010.

[16] Talfan Evans, Andrej Bicanski, Daniel Bush, and Neil Burgess. How environment and self-motion combine in neural representations of space. *Journal of Physiology*, 2016.

[17] Mathias Franzius, Henning Sprekeler, and Laurenz Wiskott. Slowness and sparseness lead to place, head-direction, and spatial-view cells. *PLoS Comput Biol*, 3(8):e166, 2007.

[18] Marianne Fyhn, Sturla Molden, Menno P. Witter, Edvard I. Moser, and May-Britt Moser. Spatial representation in the entorhinal cortex. *Science*, 305(5688):1258–1264, 2004.

[19] John Gigg. Constraints on hippocampal processing imposed by the connectivity between ca1, subiculum and subicular targets. *Behavioural Brain Research*, 174(2):265–271, 2006.

[20] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, March 2014.

[21] Stevan Harnad. Symbol-grounding problem. *Encyclopedia of Cognitive Science*, 2003.

[22] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[23] Martin Hitier, Stephane Besnard, and Paul F Smith. Vestibular pathways involved in cognition. *Front. Integr. Neurosci.*, 8(59.10):3389, 2014.

[24] Sepp Hochreiter. Untersuchungen zu dynamischen neuronalen netzen. *Master's thesis, Institut für Informatik, Technische Universität, München*, 1991.

[25] Sepp Hochreiter, Yoshua Bengio, and Paolo Frasconi. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. In J. Kolen and S. Kremer, editors, *Field Guide to Dynamical Recurrent Networks*. IEEE Press, 2001.

[26] Tim Jarsky, Alex Roxin, William L Kath, and Nelson Spruston. Conditional dendritic spike propagation following distal synaptic activation of hippocampal ca1 pyramidal neurons. *Nature Neuroscience*, 8(12):1667–1676, 2005.

[27] Michael Kearns and Daphne Koller. Efficient reinforcement learning in factored MDPs. In *IJCAI*, volume 16, pages 740–747, 1999.

[28] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv:1412.6980*, 2014.

[29] Matthew E Larkum, J Julius Zhu, and Bert Sakmann. Dendritic mechanisms underlying the coupling of the dendritic with the axonal action potential initiation zone of adult rat layer 5 pyramidal neurons. *Journal of Physiology*, 533(2):447–466, 2001.

[30] András Lőrincz and György Buzsáki. Two-phase computational model training long-term memories in the entorhinal-hippocampal region. *Annals of the New York Academy of Sciences*, 911(1):83–111, 2000.

[31] András Lőrincz, András Sárkány, Zoltán Á. Milacski, and Zoltán Tősér. Estimating Cartesian compression via deep learning. submitted, 2016.

[32] András Lőrincz and Gábor Szirtes. Here and now: How time segments may become events in the hippocampus. *Neural Networks*, 22(5):738–747, 2009.

[33] Alireza Makhzani and Brendan Frey. k-sparse autoencoders. *arXiv:1312.5663*, 2013.

[34] Alireza Makhzani and Brendan J Frey. Winner-take-all autoencoders. In *Adv. Neural Info. Proc. Sys.*, pages 2773–2781, 2015.

[35] Morris Moscovitch, Roberto Cabeza, Gordon Winocur, and Lynn Nadel. Episodic memory and beyond: The hippocampus and neocortex in transformation. *Annual review of psychology*, 67:105–134, 2016.

[36] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted Boltzmann machines. In *Proc. 27th Int. Conf. Mach. Learn.*, pages 807–814, 2010.

[37] Andrew Y Ng. Feature selection, l1 vs. l2 regularization, and rotational invariance. In *Proc. 21st Int. Conf. Mach. Learn.*, page 78. ACM, 2004.

[38] John O'Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1):171–175, 1971.

[39] John O'Keefe and Lynn Nadel. *The Hippocampus as a Cognitive Map*, volume 3. Clarendon Press Oxford, 1978.

[40] Antti Rasmus, Mathias Berglund, Mikko Honkala, Harri Valpola, and Tapani Raiko. Semi-supervised learning with ladder networks. In *Adv. Neural Info. Proc. Sys.*, pages 3532–3540, 2015.

[41] A. David Redish and David S. Touretzky. The role of the hippocampus in solving the morris water maze. *Neural Computation*, 10(1):73–111, 1998.

[42] Christopher J Rozell, Don H Johnson, Richard G Baraniuk, and Bruno A Olshausen. Sparse coding via thresholding and local competition in neural circuits. *Neural Computation*, 20(10):2526–2563, 2008.

[43] Fabian Schönfeld and Laurenz Wiskott. Modeling place field activity with hierarchical slow feature analysis. *Frontiers in Computational Neuroscience*, 9, 2015.

[44] William Beecher Scoville and Brenda Milner. Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20(1):11, 1957.

[45] Trygve Solstad, Edvard I. Moser, and Gaute T. Einevoll. From grid cells to place cells: a mathematical model. *Hippocampus*, 16(12):1026–1031, 2006.

[46] Larry R Squire and Stuart M Zola. Episodic memory, semantic memory, and amnesia. *Hippocampus*, 8(3):205–211, 1998.

[47] Greg J Stuart and Nelson Spruston. Dendritic integration: 60 years of progress. *Nature Neuroscience*, 18(12):1713–1721, 2015.

[48] A. Stuchlik and J. Bures. Relative contribution of allothetic and idiothetic navigation to place avoidance on stable and rotating arenas in darkness. *Behavioural Brain Research*, 128(2):179–188, 2002.

[49] A. Stuchlik, T. Petrasek, I. Prokopova, K. Holubova, H. Hatalova, K. Vales, S. Kubik, C. Dockery, and M. Wesierska. Place avoidance tasks as tools in the behavioral neuroscience of learning and memory. *Physiological Research*, 62:S1, 2013.

[50] Yanan Sun, Hua Mao, Yongsheng Sang, and Zhang Yi. Explicit guiding auto-encoders for learning meaningful representation. *Neural Comp. Appl.*, pages 1–8, 2015.

[51] Csaba Szepesvári and Andras Lőrincz. An integrated architecture for motion-control and path-planning. *J. Robot. Syst.*, 15(1):1–15, 1998.

[52] István Szita and András Lőrincz. Optimistic initialization and greediness lead to polynomial time learning in factored MDPs. In *Proc. 26th Int. Conf. Mach. Learn.*, pages 1001–1008. ACM, 2009.

[53] Jeffrey S. Taube. The head direction signal: origins and sensory-motor integration. *Annual Rev.. Neuroscience*, 30:181–207, 2007.

[54] Robert Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc. Ser. B (Meth.)*, pages 267–288, 1996.

[55] James T. Todd. The visual perception of 3d shape. *Trends in Cognitive Sciences*, 8(3):115–121, 2004.

[56] Joel A Tropp and Anna C Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *Info. Theo.*, 53(12):4655–4666, 2007.

[57] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders. *J. Mach. Learn. Res.*, 11:3371–3408, 2010.

[58] Olga S. Vinogradova. Hippocampus as comparator: role of the two input and two output systems of the hippocampus in selection and registration of information. *Hippocampus*, 11(5):578–598, 2001.

[59] Shawn S. Winter, Benjamin J. Clark, and Jeffrey S. Taube. Disruption of the head direction cell network impairs the parahippocampal grid cell signal. *Science*, 347(6224):870–874, 2015.

[60] Shawn S. Winter and Jeffrey S. Taube. Head direction cells: from generation to integration. In *Space, Time and Memory in the Hippocampal Formation*, pages 83–106. Springer, 2014.

[61] Matthew D Zeiler. Adadelta: an adaptive learning rate method. *arXiv:1212.5701*, 2012.