Modelling Biological Systems using a Parallel Quantized MIMO Channel

Martijn Arts*, Steven Corroy*, Monika Gorin†, Marc Spehr†, Anke Schmeink‡, Rudolf Mathar*
*Institute for Theoretical Information Technology, Email: {arts,corroy,mathar}@ti.rwth-aachen.de
†Department of Chemosensation, Institute for Biology II, Email: {M.Gorin,M.Spehr}@sensorik.rwth-aachen.de
‡UMIC Research Center, Email: schmeink@umic.rwth-aachen.de
RWTH Aachen University, D-52056 Aachen, Germany

***Abstract*—In this work, we investigate a model which is related to the class of stochastic pooling networks (SPN). These networks consist of a parallel structure of noisy and compressive sensors, which observe a common input signal. They have proven to be useful in interdisciplinary research, e.g., in physics and neurobiology. By adding a second source of parallel noise and allowing cross-connections using a channel matrix, we merge these models with the multiple-input multiple-output (MIMO) framework. In contrast to typical wireless communication scenarios, we assume the channel matrix to be changed deliberately in order to study the information processing and interconnection of neurons. We investigate which channel matrix maximizes the mutual information for the MIMO case and a single-input multiple-output (SIMO) special case and present two convex relaxations of the original problems. Based on a modified non-negative matrix factorization (NMF) algorithm, we formulate a heuristic to obtain feasible channel matrices. Finally, we evaluate the performance of the suggested heuristic.**

I. Introduction and Motivation

Biological systems are highly complex and rely on biochemical processes, many of which are not fully understood to date. To analyze such systems it is desirable to model them using mathematical concepts, in order to be capable of predicting the outcome of experiments. One successful example is the use of information theoretical concepts in neurobiology.

In this work, we investigate a model which is related to the class of stochastic pooling networks (SPN) [1]. Such networks exhibit a parallel structure of sensors which observe a common input signal, afflicted by some form of noise. Compression, e.g., by quantization, is present at each sensor and all sensor outputs are fused into a single output by a pooling function. We study a generalization of a model incorporating binary quantization from [2], see Fig. 1. Since the individual nodes behave identically in the absence of noise, their output is a very coarse approximation of the input signal. Interestingly, the maximum of the mutual information between the input and the output of this network occurs for nonzero levels of noise, see [3].

In biological systems, parallel information processing seems to be an inherent principle. Many biological sensor systems bear a strong resemblance to the SPN definition, such as the retina in the human eye, the cochlea in the auditory system or the olfactory system of mammals. The latter example, in particular the accessory olfactory bulb (AOB) of mice, was the main inspiration for our approach.

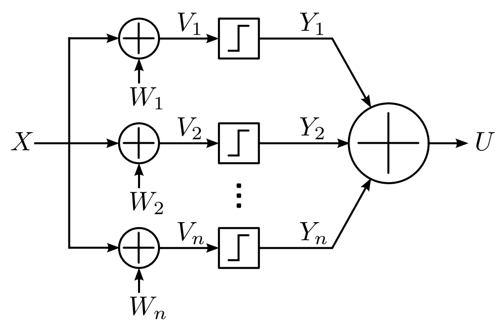

Fig. 1. A model complying with the SPN definition, as studied in [2].

The AOB is the first of only two to three synaptic relay sites in the accessory olfactory pathway — a neuronal information processing stream that extends from sensory neurons to higher order brain cells. Sensory neurons detect environmental pheromone signals, whereas the higher order brain cells determine the endocrine state and behavior of an individual [4]. The rodent AOB is built by a few hundred mitral cells, which receive massive parallel synaptic input from several thousand sensory neurons, see also [5]. Despite the AOB's fundamental biological role, however, the basic rules of information processing in this brain region are still poorly understood.

We will examine two problems in this work, which correspond to two hypotheses about the connectivity within the AOB. The first hypothesis is that different types of sensors are connected to several mitral cells, thus forming a multiple-input multiple-output (MIMO) network. In contrast, the second hypothesis is that sensors detecting a certain class of pheromones are only connected to several mitral cells associated with the same class. This can be seen as single-input multiple-output (SIMO) networks working in parallel. To get quantifiable insights into the information transmission and the connectivity structure, we extend the model in Fig. 1 by a MIMO channel matrix. By assuming the channel matrix, which represents neuronal connections, can be changed, we formulate an optimization problem to find the optimal channel matrix to maximize the mutual information of the non quantized part of the model. Using convex optimization techniques, we relax this problem into a convex problem which can be solved efficiently by numerical methods. Due to a variable change, the optimal channel matrix is not directly obtained by solving the convex problem. Thus, we propose

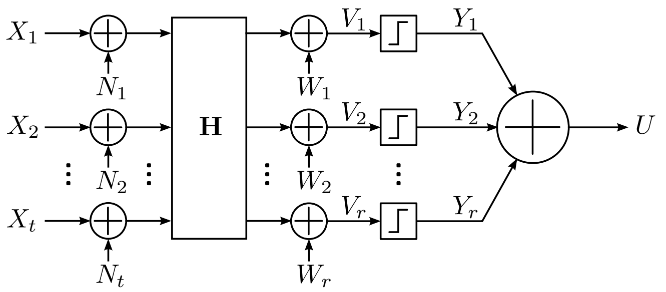

Fig. 2. Our extension of the model depicted in Fig. 1 with MIMO channel matrix $\mathbf{H}$.

a heuristic which approximately recovers the channel matrix. It is based on a modified non-negative matrix factorization (NMF) algorithm and we evaluate the performance of said heuristic. The SIMO special case is treated in a similar fashion.

In Section II we define the system model and necessary notation. Section III covers the optimization problems and their relaxations. The heuristic and the modified non-negative matrix factorization algorithm is introduced in Section IV. The performance of the heuristic is evaluated in Section V. Finally, we conclude the paper in Section VI.

II. System Model and Notation

Extending previous approaches, we propose a generalization of the model shown in Fig. 1 using the multiple-input multiple-output (MIMO) framework, broadening its scope and applicability, see Fig. 2. Instead of a single random variable X, a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_t)^T$ is considered as the models' input. A random vector $\mathbf{N} = (N_1, N_2, \ldots, N_t)^T$ symbolizes additive input noise. A channel matrix $\mathbf{H}$ with real entries of dimension $r \times t$ describes the channel gain of the direct- and cross-links from the transmitter to the receiver stage. After the channel matrix, a second noise vector $\mathbf{W} = (W_1, W_2, \ldots, W_r)^T$ is added. Thus, the input vector of the parallel quantizers $\mathbf{V} = (V_1, V_2, \ldots, V_r)^T$ can be expressed as $\mathbf{V} = \mathbf{H}\,(\mathbf{X} + \mathbf{N}) + \mathbf{W}$.

We consider the input signal and the noise vectors to follow zero mean multivariate normal distributions with full rank covariance matrices $\mathbf{\Sigma_X}$, $\mathbf{\Sigma_N}$ and $\mathbf{\Sigma_W}$ respectively. Furthermore, the input signal vector $\mathbf{X}$ is assumed to be independent of the noise vectors $\mathbf{N}$ and $\mathbf{W}$ as well as the noise vectors $\mathbf{N}$ and $\mathbf{W}$ are assumed to be independent of each other. $\mathbf{\Sigma_N}$ and $\mathbf{\Sigma_W}$ are assumed to be diagonal. The quantizers are binary and their threshold is set to zero.

To establish notations for the following treatment, we define the mutual information between the two random vectors $\mathbf{X}$ and $\mathbf{Y}$ as $\mathrm{I}(\mathbf{X};\mathbf{Y}) = \mathrm{H}(\mathbf{Y}) - \mathrm{H}(\mathbf{Y}|\mathbf{X}) = \mathrm{H}(\mathbf{X}) - \mathrm{H}(\mathbf{X}|\mathbf{Y})$.

Here, $\mathrm{H}(\mathbf{X})$ denotes the entropy of a random vector $\mathbf{X}$ and $\mathrm{H}(\mathbf{Y}|\mathbf{X})$ the conditional entropy of $\mathbf{Y}$ given $\mathbf{X}$ [6].

Note, that we use curly relational operators, e.g., $\mathbf{A} \succeq 0$ to indicate semi-definiteness of a matrix, whereas regular relational operators denote element-wise relations of the matrix entries, i.e., $\mathbf{A} \geq 0$. A single element of the matrix $\mathbf{A}$ will be referenced as a_{ij} and $\mathbf{a_i}$ refers to the i-th column of $\mathbf{A}$.

Analytical formulation of the mutual information between input and output of the system $\mathrm{I}(\mathbf{X}; U)$ is a hard problem for most input and noise distributions (compare results of the original model [3], [2]). Through the use of the information processing inequality [6], we can bound the mutual information between input and output (cf. [7], [3]):

$$\mathrm{I}(X;U) \leq \min\{\mathrm{I}(\mathbf{X};\mathbf{V}),\ \mathrm{I}(\mathbf{V};U)\}, \tag{1}$$

where $\mathrm{I}(\mathbf{V};U) \leq \log(r+1)$. This is the trivial upper bound of the support cardinality of the discrete output U.

We can state the mutual information between $\mathbf{X}$ and $\mathbf{V}$ as:

$$\mathrm{I}(\mathbf{X};\mathbf{V}) = \frac{1}{2}\log\left(\frac{\det(\mathbf{H}(\mathbf{\Sigma_X}+\mathbf{\Sigma_N})\mathbf{H^T}+\mathbf{\Sigma_W})}{\det(\mathbf{H}\mathbf{\Sigma_N}\mathbf{H^T}+\mathbf{\Sigma_W})}\right). \tag{2}$$

Using straightforward matrix identities, Sylvester's determinant theorem and the matrix inversion lemma, we can reformulate (2) as

$$\begin{aligned}\mathrm{I}(\mathbf{X};\mathbf{V}) = \frac{1}{2}\log\det\big(&\mathbf{I} + \mathbf{\Sigma_X}\mathbf{H^T}\mathbf{\Sigma_W^{-1}}\mathbf{H} \\ &- \mathbf{\Sigma_X}\mathbf{H^T}\mathbf{\Sigma_W^{-1}}\mathbf{H}(\mathbf{\Sigma_N^{-1}}+\mathbf{H^T}\mathbf{\Sigma_W^{-1}}\mathbf{H})^{-1}\mathbf{H^T}\mathbf{\Sigma_W^{-1}}\mathbf{H}\big).\end{aligned} \tag{3}$$

III. Optimization Problem Formulation

As opposed to typical communication systems, we will assume in the following that the behavior of the channel, represented by the matrix $\mathbf{H}$, can be deliberately changed. Here, the channel matrix $\mathbf{H}$ describes interconnections between neurons. By evolutionary progress the mammal body has optimized the neural information processing subject to a set of constraints. Hence, also the channel was part of this process.

In this section, we want to find the optimal channel, which maximizes the mutual information $\mathrm{I}(\mathbf{X};\mathbf{V})$. We assume that the channel matrix, i.e., the neuronal connections, are not able to amplify signals and perform only additive processing of the signals. This can be stated in the following optimization problem with solution $p^\star$:

$$\begin{aligned}&\underset{\mathbf{H}}{\text{maximize}} && \mathrm{I}(\mathbf{X};\mathbf{V}) \\ &\text{subject to} && 0 \leq \mathbf{H} \leq 1.\end{aligned} \tag{4}$$

We can make a change of variable in the objective function (3) to obtain an optimization problem which is equivalent to problem (4):

$$\begin{aligned}&\underset{\mathbf{Z}}{\text{maximize}} && \frac{1}{2}\log\det\big(\mathbf{I}+\mathbf{\Sigma_X}\mathbf{Z}-\mathbf{\Sigma_X}\mathbf{Z}(\mathbf{\Sigma_N^{-1}}+\mathbf{Z})^{-1}\mathbf{Z}\big) \\ &\text{subject to} && \mathbf{Z} = \mathbf{H^T}\mathbf{\Sigma_W^{-1}}\mathbf{H},\ \ 0 \leq \mathbf{H} \leq 1.\end{aligned} \tag{5}$$

Problem (5) is not convex. To find an upper bound on the solution efficiently, we will introduce a convex relaxation of this problem in the following. Thereafter, we will examine the single-input multiple-output (SIMO) special case of this problem.

A. Convex Relaxation of the MIMO Problem

Since $\mathbf{\Sigma_W}$ is a diagonal full rank covariance matrix, its inverse $\mathbf{\Sigma_W^{-1}} \succ 0$ and it also holds that $\mathbf{Z} = \mathbf{H^T}\mathbf{\Sigma_W^{-1}}\mathbf{H} \succeq 0$. Therefore we relax problem (5) by replacing the constraint $\mathbf{Z} = \mathbf{H^T}\mathbf{\Sigma_W^{-1}}\mathbf{H}$ by $\mathbf{Z} \succeq 0$. We can relax the element-wise constraints on $\mathbf{H}$ by $0 \leq \mathbf{Z} \leq \mathrm{tr}(\mathbf{\Sigma_W^{-1}})$, where $\mathrm{tr}(\cdot)$ denotes the trace. The minimum element of this set corresponds to $\mathbf{H}=\mathbf{0}$, while the maximum element corresponds to $\mathbf{H}=\mathbf{1}\mathbf{1}^{\mathbf{T}}$. From this relaxation we obtain the following optimization problem:

$$\begin{array}{ll}\underset{\mathbf{Z}}{\text{maximize}} & \frac{1}{2}\log\det\left(\mathbf{I}+\mathbf{\Sigma_X}\mathbf{Z}-\mathbf{\Sigma_X}\mathbf{Z}(\mathbf{\Sigma_N^{-1}}+\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}\right)\\ \text{subject to} & 0\leq\mathbf{Z}\leq\operatorname{tr}(\mathbf{\Sigma_W^{-1}}),\;\; \mathbf{Z}\succeq 0.\end{array} \tag{6}$$

Note that problem (6) is still non-convex since the objective function is not concave. Therefore, we introduce a slack matrix variable $\mathbf{\Gamma}$. By adding the constraint $\mathbf{\Gamma}=\mathbf{Z}(\mathbf{\Sigma_N^{-1}}+\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}$ and replacing the corresponding term in the objective function by $\mathbf{\Gamma}$, we get a problem that is equivalent to problem (6), now maximizing over the variables $\mathbf{H}$ and $\mathbf{\Gamma}$. Relaxing the new constraint to $\mathbf{\Gamma}\succeq\mathbf{Z}(\mathbf{\Sigma_N^{-1}}+\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}$ and using the Schur-complement, we obtain the following problem with solution $q^\star$:

$$\begin{array}{ll}\underset{\mathbf{Z},\mathbf{\Gamma}}{\text{maximize}} & \frac{1}{2}\log\det\left(\mathbf{I}+\mathbf{\Sigma_X}(\mathbf{Z}-\mathbf{\Gamma})\right)\\ \text{subject to} & 0\leq\mathbf{Z}\leq\operatorname{tr}(\mathbf{\Sigma_W^{-1}}),\;\; \mathbf{Z}\succeq 0,\\ & \begin{bmatrix}(\mathbf{\Sigma_N^{-1}}+\mathbf{Z}) & \mathbf{Z}\\ \mathbf{Z} & \mathbf{\Gamma}\end{bmatrix}\succeq 0.\end{array} \tag{7}$$

As the objective function is concave and all constraints are convex, (7) is a convex optimization problem, which can be solved numerically in an efficient manner. The constraint set of problem (7) contains the one of problem (4), thus we have $q^\star \geq p^\star$. We will use the solution $\mathbf{Z}^\star$ of problem (7) in section IV as a basis for a heuristic to find good solutions for problem (5).

B. Convex Relaxation of the SIMO Problem

A notable special case of the MIMO model from Fig. 2 is to restrict the input to a common scalar input X, as depicted in Fig. 3. The random variable X is assumed to follow a zero mean Gaussian distribution with variance σ_X^2 and it is independent of the noise vectors $\mathbf{N}$ and $\mathbf{W}$. In order to gain an optimization problem similar to the MIMO case, we start with problem (6) and substitute $\mathbf{\Sigma_X}$ by $\sigma_X^2\mathbf{1}\mathbf{1}^{\mathbf{T}}$ in the objective function. Here, $\mathbf{1}$ is a column vector with all entries equal to one of appropriate dimension. Application of Sylvester's determinant theorem on the objective function results in a scalar argument of the determinant. The relaxed SIMO problem formulation can thus be stated as:

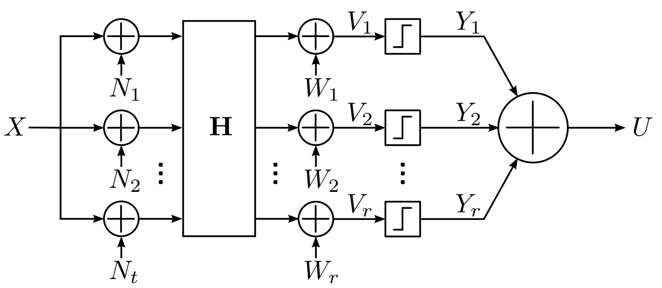

Fig. 3. Single-input multiple-output (SIMO) special case of the original MIMO model from Fig. 2.

$$\begin{array}{ll}\underset{\mathbf{Z}}{\text{maximize}} & \frac{1}{2}\log\left(1+\sigma_X^2(\mathbf{1}^{\mathbf{T}}\mathbf{Z}\mathbf{1}-\mathbf{1}^{\mathbf{T}}\mathbf{Z}(\mathbf{\Sigma_N^{-1}}+\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}\mathbf{1})\right)\\ \text{subject to} & \mathbf{Z}=\mathbf{H}^{\mathbf{T}}\mathbf{\Sigma_W^{-1}}\mathbf{H},\;\; 0\leq\mathbf{H}\leq 1.\end{array} \tag{8}$$

Similarly to the treatment in section III-A, we introduce a slack variable γ. Then we add the constraint $\gamma \geq \mathbf{1}^{\mathbf{T}}\mathbf{Z}(\mathbf{\Sigma_N^{-1}}+\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}\mathbf{1}$ and substitute the corresponding term in the objective function by γ. Finally, application of the Schur-complement results in:

$$\begin{array}{ll}\underset{\mathbf{Z},\gamma}{\text{maximize}} & \frac{1}{2}\log\left(1+\sigma_X^2(\mathbf{1}^{\mathbf{T}}\mathbf{Z}\mathbf{1}-\gamma)\right)\\ \text{subject to} & 0\leq\mathbf{Z}\leq\operatorname{tr}(\mathbf{\Sigma_W^{-1}}),\;\; \mathbf{Z}\succeq 0,\\ & \begin{bmatrix}(\mathbf{\Sigma_N^{-1}}+\mathbf{Z}) & \mathbf{Z}\mathbf{1}\\ \mathbf{1}^{\mathbf{T}}\mathbf{Z} & \gamma\end{bmatrix}\succeq 0.\end{array} \tag{9}$$

Note that γ is scalar and we are maximizing over the variables $\mathbf{Z}$ and γ. Thus, problem (9) is equivalent to problem (8). Furthermore, since problem (9) is convex and contains the constraint set of the original SIMO problem, we can use it to efficiently compute an upper bound on the solution of the original SIMO problem. If each of the noise vectors are i.i.d., so that the covariance matrices can be written as $\mathbf{\Sigma_N}=\sigma_N^2\mathbf{I}$ and $\mathbf{\Sigma_W}=\sigma_W^2\mathbf{I}$, the optimal channel matrix is $\mathbf{H_f^\star}=\mathbf{1}\mathbf{1}^{\mathbf{T}}$. This can be shown by straightforward application of the KKT conditions of problem (9). Noticing that the solution is also feasible for the original problem means that $\mathbf{H_f}$ is also optimal for the original and unrelaxed problem in this case.

IV. Channel matrix heuristic

In the previous section, a change of variable followed by a convex relaxation was used to formulate two convex optimization problems for the MIMO and the SIMO approach. Since we aim at interpreting these models in a biological context, we are not only interested in an upper bound on the solution, but also in the optimal channel matrix itself. However, solving problems (7) or (9) does not provide the optimal channel matrix, but rather a matrix $\mathbf{Z}^\star$ which was substituted as $\mathbf{Z}=\mathbf{H}^{\mathbf{T}}\mathbf{\Sigma_W^{-1}}\mathbf{H}$ before relaxation. We will examine a heuristic in this section, which yields an estimate of the optimal channel matrix.

Given the optimal matrix $\mathbf{Z}^\star$, if a decomposition of the form $\mathbf{Z}^\star=\mathbf{H}^{\star\mathbf{T}}\mathbf{\Sigma_W^{-1}}\mathbf{H}^\star$ exists, where $\mathbf{H}^\star$ satisfies the constraint of the original unrelaxed problem (4), then $\mathbf{H}^\star$ is also optimal for the latter. In case the amount of transmitters and receivers are equal, i.e. $r=t$, one can analytically find a decomposition as $\mathbf{H_d}=\mathbf{\Sigma_W^{\frac{1}{2}}}\mathbf{D}^{\frac{1}{2}}\mathbf{Q}^{\mathbf{T}}$, where $\mathbf{Q}$ and $\mathbf{D}$ form a spectral decomposition $\mathbf{Z}=\mathbf{Q}\mathbf{D}\mathbf{Q}^{\mathbf{T}}$.

However, the matrix $\mathbf{H_d}$ often contains negative entries, which violates the corresponding constraint in problem (4). A second disadvantage is that such a decomposition cannot be explicitly given for the case when $r\neq t$. Thus, we propose a heuristic to find a channel matrix $\mathbf{H_h}$, so that $\mathbf{Z}^\star\approx\mathbf{H_h}\mathbf{\Sigma_W^{-1}}\mathbf{H_h}$, while $\mathbf{H_h}$ satisfies the constraints of the original problem (4).

Decomposing a non-negative matrix approximately into two other non-negative matrices is known as the non-negative matrix factorization (NMF) problem. It was introduced in [8] and became more widely known by the application in a neural network learning context [9]. Given a non-negative $m \times n$ matrix $\mathbf{C}$, the NMF problem can be stated as:

$$\begin{aligned} \underset{\mathbf{A},\mathbf{B}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{C}-\mathbf{A}\mathbf{B}^{\mathbf{T}}\|_F^2 \\ \text{subject to} \quad & \mathbf{A} \geq 0, \;\; \mathbf{B} \geq 0. \end{aligned} \tag{10}$$

Here, the matrix $\mathbf{A}$ is of dimension $m \times o$, $\mathbf{B}$ is of dimension $n \times o$ and $\|\cdot\|_F$ denotes the Frobenius norm. The NMF problem is often associated with a rank reduction, i.e. $o \leq \min(m,n)$. Several different algorithms have been proposed to estimate the optimal solution of the non-convex optimization problem (10), e.g. in [8], [9], [10].

The basic idea behind our heuristic is to find a good approximate decomposition of the optimal matrix $\mathbf{Z}^\star$ of either problem (7) or problem (9). This problem is very similar to the NMF problem explained above and we can state it in the following way. Given a $t \times t$ matrix $\mathbf{Z}$, the diagonal $r \times r$ covariance matrix $\mathbf{\Sigma} = \mathbf{\Sigma_W}$ and the optimization variables $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{t \times r}$, solve:

$$\begin{aligned} \underset{\mathbf{A},\mathbf{B}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{Z}-\mathbf{A}\mathbf{\Sigma}^{-\mathbf{1}}\mathbf{B}^{\mathbf{T}}\|_F^2 + \frac{\alpha}{2}\|\mathbf{A}-\mathbf{B}\|_F^2 \\ \text{subject to} \quad & \mathbf{A} \geq 0, \;\; \mathbf{B} \geq 0. \end{aligned} \tag{11}$$

The second term in the objective function is a penalty with parameter $\alpha > 0$, which is used to get a symmetric decomposition, so that ideally $\mathbf{A} = \mathbf{B}$, see chapter 7 in [10]. To solve this problem we extended the rank-one residue iteration (RRI) algorithm from [10] (Algorithm 13, p. 141) by the diagonal scaling $\mathbf{\Sigma}^{-\mathbf{1}}$. By rewriting the first term of the objective function of problem (11) into $\frac{1}{2}\|\mathbf{Z} - \sum_{i=1}^{r} \sigma_{ii}\mathbf{a_i}\mathbf{b_i^T}\|_F^2$, we see that the matrix product can be interpreted as a sum of rank-one matrices. The main idea of the RRI algorithm is to fix all of those vectors except one and solve the resulting non-negative least square problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{R}-\sigma_{ii}\mathbf{a_i}\mathbf{x}^{\mathbf{T}}\|_F^2 + \frac{\alpha}{2}\|\mathbf{a_i}-\mathbf{x}\|_F^2 \\ \text{subject to} \quad & \mathbf{x} \geq 0, \end{aligned} \tag{12}$$

where $\mathbf{R} = \mathbf{Z} - \sum_{k \neq i} \sigma_{kk}\mathbf{a_k}\mathbf{b_k^T}$ can be interpreted as a residuum matrix. The optimal solution to problem (12) can be analytically found as $\mathbf{x}^\star = \frac{[\sigma_{ii}\mathbf{R}^{\mathbf{T}}\mathbf{a_i}+\alpha\mathbf{a_i}]_+}{\alpha+\sigma_{ii}^2\|\mathbf{a_i}\|_2^2}$, where $[\cdot]_+ = \max(0,\cdot)$. Then, $\mathbf{x}^\star$ is used to update $\mathbf{b_i}$. Likewise, an update rule for $\mathbf{a_i}$ can be created. Iteratively updating $\mathbf{b_i}$ and $\mathbf{a_i}$ to create rank-one approximations of the residuum matrix results in the core part of the RRI algorithm, as can be seen in Algorithm 1 lines 4, 5 and 7. In Line 1 the same random start point is assigned to $\mathbf{A}$ and $\mathbf{B}$ and then both matrices are scaled to get a first crude approximation of $\mathbf{Z}$, cp. [10], p. 85. A balancing step is performed after updating one of the vectors, as suggested in [10], p. 141. Also, to account for update vectors being equal to zero, a suitable procedure to assign a better non-zero update vector is performed in line 9 for a limited number of iterations, see [10], p. 71 - 72. As a stopping condition, it is checked whether $\|[\nabla\mathbf{A}, \nabla\mathbf{B}]\|_F \leq \varepsilon\, G_0$, or if a maximum iteration count is reached. Here, $\nabla\mathbf{A}$ is the difference between the matrices obtained in the current and last iteration and G_0 is the Frobenius norm of the differences from the start point in the first iteration, see [10], p. 62.

Algorithm 1: extended RRI algorithm for problem (11)

Data: $\mathbf{Z}$, $\mathbf{\Sigma}$, α
Result: $\mathbf{A}$, $\mathbf{B}$

1 [$\mathbf{A}$, $\mathbf{B}$] = initialize_startpoint()
2 **repeat**
3   **for** $i \in [1,r]$**:**
4         $\mathbf{R} = \mathbf{Z} - \sum_{k \neq i} \sigma_{kk}\mathbf{a_k}\mathbf{b_k^T}$
5         $\mathbf{b_i} = [\sigma_{ii}\mathbf{R}^{\mathbf{T}}\mathbf{a_i} + \alpha\mathbf{a_i}]_+ \;/\; [\alpha + \sigma_{ii}^2\|\mathbf{a_i}\|_2^2]$
6   [$\mathbf{a_i}, \mathbf{b_i}$] = balancing($\mathbf{a_i}$, $\mathbf{b_i}$)
7         $\mathbf{a_i} = [\sigma_{ii}\mathbf{R}\mathbf{b_i} + \alpha\mathbf{b_i}]_+ \;/\; [\alpha + \sigma_{ii}^2\|\mathbf{b_i}\|_2^2]$
8   [$\mathbf{a_i}, \mathbf{b_i}$] = balancing($\mathbf{a_i}$, $\mathbf{b_i}$)
9   [$\mathbf{a_i}, \mathbf{b_i}$] = fix_zero_vectors($\mathbf{a_i}$, $\mathbf{b_i}$)
10 **until** *stopping condition*

Using the methodology that was introduced in sections III and IV, we can now formulate a heuristic to find feasible matrices $\hat{\mathbf{H}}$ for the original MIMO and SIMO problems. First, we solve problem (7) or (9) to get $\mathbf{Z}^\star$. Then we use Algorithm 1 to get an approximate decomposition. Note that in practice, using a suitable parameter α, the asymmetry between $\mathbf{A}$ and $\mathbf{B}$ is usually very small and can be neglected. Thus, we set the channel matrix to $\hat{\mathbf{H}} = \mathbf{A}^{\mathbf{T}}$. Because of the convex relaxation involving the variable change $\mathbf{Z} = \mathbf{H}^{\mathbf{T}}\mathbf{\Sigma}_{\mathbf{W}}^{-\mathbf{1}}\mathbf{H}$, the upper element-wise constraint $\mathbf{H} \leq 1$ was lost. As a result $\hat{\mathbf{H}}$ typically contains some entries which are bigger than one. Therefore, as a final step of the heuristic, we threshold the matrix entries of $\hat{\mathbf{H}}$ to one, i.e. $\mathbf{H_t} = \min(\hat{\mathbf{H}}, 1)$ to generate a feasible matrix $\mathbf{H_t}$ for the original problem.

V. Heuristic evaluation

In this section, the performance of the heuristic which was introduced above will be evaluated. Lacking other heuristics as a performance comparison and the optimal channel matrix of either the SIMO or the MIMO problem, the performance of the heuristic, which will be labeled as $\mathbf{Z_t}$, will be compared to the solutions of problems (7) or (9), respectively. To obtain numerical solutions of these problems, we used CVX, a software package for solving convex programs [11]. As a second comparison, random channel matrices $\mathbf{H_r}$ with entries following a uniform distribution in the interval $[0,1]$, labeled as $\mathbf{Z_r}$, will be considered. We will additionally compare the heuristic performance for the SIMO special case to the SIMO solution for i.i.d. noise, labeled as $\mathbf{Z_f}$ with $\mathbf{H_f} = \mathbf{1}\mathbf{1}^{\mathbf{T}}$.

For the following simulations, we generated 2000 instances of the problems (7) and (9). The diagonal entries of the covariance matrices $\mathbf{\Sigma_N}$ and $\mathbf{\Sigma_W}$ were independently drawn from a uniform distribution with expectations $E_N = E(\sigma_{Nii})$ and $E_W = E(\sigma_{Wii})$. For the SIMO problem the variance of the input signal was set to one, whereas for the MIMO case the covariance matrices were generated by drawing a matrix $\mathbf{Q}$ from a uniform distribution in the interval $[0,1]$ and setting $\mathbf{\Sigma_X} = \text{const.} \cdot \mathbf{Q^T Q}$. The scaling constant is chosen, so that the expectation of the resulting Beta distribution is $E_X = E(\sigma_{Xii})$.

In Fig. 4 the average of the mutual information $\mathrm{I}(\mathbf{X};\mathbf{V})$ of the 2000 simulated instances is depicted for the MIMO and the SIMO case. As can be seen, the SIMO case is very robust so that the improvement over randomly chosen channel matrices is relatively small, which can be explained by the limited degrees of freedom in the system. Especially in the MIMO case with high noise (lines with squares in Fig. 4), the average deviation to the solution $\mathbf{Z}^\star$ is very small. However, no statement can be made about the deviation that the heuristic achieves when compared to the true optimum, since we only know the optimal solution of the relaxed problems.

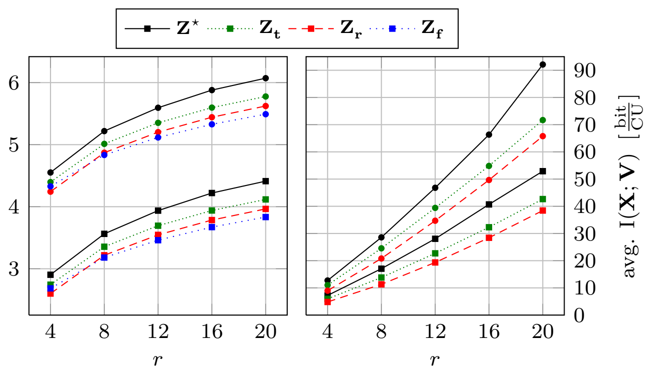

Fig. 4. Average mutual information after 2000 instances, *left:* SIMO model and *right:* MIMO model with $E_X = 1$. For the lines with circles $E_W = E_N = 0.01$, for the lines with squares $E_W = E_N = 0.1$. For all plots, $r = t$. Algorithm 1 was configured with $\varepsilon = 10^{-6}$ with a maximum iteration count of 2500.

In Fig. 5 an approximation of the cumulative distribution function (CDF) of the relative deviation between the mutual information $\mathrm{I}(\mathbf{X};\mathbf{V})|_{\mathbf{Z}=\mathbf{Z_k}}, \mathbf{k} \in \{\mathbf{t},\mathbf{r},\mathbf{f}\}$ and $\mathrm{I}(\mathbf{X};\mathbf{V})|_{\mathbf{Z}=\mathbf{Z}^\star}$ is shown. Clearly, the heuristic generates channel matrices that achieve a higher mutual information with a much higher probability, compared to a randomly chosen matrix. For the MIMO case, about 90% of the generated matrices result in a mutual information value, which deviates less than 22% from the upper bound on the optimal solution, which was obtained by solving problem (7).

VI. Conclusion

In this work, we presented an extension of a well studied SPN model for biological applications, which ties said model to the MIMO framework. To study the interconnections and the information processing of neurons in the olfactory system of mammals using our information theoretic model, we assume the channel matrix $\mathbf{H}$ to be changed deliberately. Using convex optimization techniques we have found convex relaxations of the optimization problems for finding the optimal channel matrix for the MIMO case and the SIMO special case, which can be used to efficiently compute upper bounds on the solution

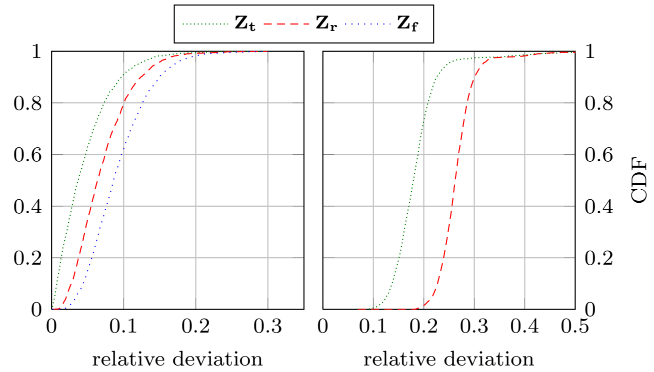

Fig. 5. Approximate cumulative distribution function (CDF) of the relative deviation from $\mathrm{I}(\mathbf{X};\mathbf{V})|_{\mathbf{Z}=\mathbf{Z}^\star}$ *left:* SIMO model with $E_W = E_N = 0.01$, *right:* MIMO model with $E_W = E_N = 0.1$ and $E_X = 0.1$. For both plots, $r = t = 20$. Algorithm 1 was configured with $\varepsilon = 10^{-6}$ with a maximum iteration count of 2500.

of the original problems. Furthermore, we have introduced an extension of a non-negative matrix factorization (NMF) algorithm, which includes a diagonal scaling matrix. Based on the convex relaxation and the NMF algorithm, we formulated a heuristic to generate feasible channel matrices. We found that in about 90% of the cases, the mutual information obtained by the MIMO channel matrix, which was found by the heuristic, deviates less than 22% from the upper bound on the optimal solution obtained by solving problems (7).

Acknowledgment

This work was partly supported by the Deutsche Forschungsgemeinschaft (DFG) SPP 1395: Informations- und Kommunikationstheorie in der Molekularbiologie (InKoMBio) and by the UMIC research cluster of the RWTH Aachen University.

References

[1] M. D. McDonnell, P.-O. Amblard, and N. G. Stocks, “Stochastic pooling networks,” *J Stat Mech-Theory E*, vol. 2009, no. 01, p. P01012, 2009.

[2] N. G. Stocks, “Suprathreshold stochastic resonance in multilevel threshold systems,” *Phys. Rev. Lett.*, vol. 84, pp. 2310–2313, Mar 2000.

[3] M. D. McDonnell, P.-O. Amblard, and N. G. Stocks, “Bio-inspired communication: Performance limits for information transmission and compression in stochastic pooling networks with binary quantizing nodes,” *J Comput Theor Nanos*, vol. 7, no. 5, pp. 876–883, 2010.

[4] R. Tirindelli, M. Dibattista, S. Pifferi, and A. Menini, “From pheromones to behavior,” *Physiol Rev.*, vol. 89, pp. 921–956, 2009.

[5] J. Larriva-Sahd, “The accessory olfactory bulb in the adult rat: a cytological study of its cell types, neuropil, neuronal modules, and interactions with the main olfactory system,” *J Comp Neurol.*, vol. 510, pp. 309–350, 2008.

[6] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing, John Wiley & Sons, 2006.

[7] R. Mathar and A. Schmeink, “A bio-inspired approach to condensing information,” in *IEEE International Symposium on Information Theory (ISIT)*, (Saint-Petersburg), pp. 2524–2528, Aug. 2011.

[8] P. Paatero and U. Tapper, “Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values,” *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.

[9] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, pp. 788–791, Oct. 1999.

[10] N.-D. Ho, *Nonnegative Matrix Factorization Algorithms and Applications*. PhD thesis, Université Catholique de Louvain, École Polytechnique de Louvain, Département d’Ingénierie Mathématique, June 2008.

[11] CVX Research Inc., “CVX: Matlab software for disciplined convex programming, version 1.22.” http://cvxr.com/cvx, Aug. 2012.