MASARYK UNIVERSITY
Faculty of Science
Department of Mathematics and Statistics

DISSERTATION

Ondřej POKORA   **Brno 2009**

MASARYK UNIVERSITY
Faculty of Science
Department of Mathematics and Statistics

Ondřej POKORA

MATHEMATICAL MODELS
IN NEUROPHYSIOLOGY
MODELS OF OLFACTORY NEURONS

Dissertation

Bibliographic entry

Acknowledgements

Abstract

In the first part, some basic properties of Fisher information about a scalar parameter are recalled and the relation to the variance of estimator of the parameter is explained. Some approximations of Fisher information are introduced and their properties are derived. Further, several theoretical models for coding of odor intensity in olfactory sensory neurons are investigated. Behavior of the models is described by stochastic processes of binding the odorant molecules to the receptors and activation of receptor-odorant complexes. Characteristics how well the odorant concentration can be estimated from the knowledge of response, the concentration of activated neurons, are studied. Two different approaches in search for optimal odorant concentration are used, the first one uses Fisher information and its approximations, the other one suggests to use some information-theory measures. These measures of optimality are computed and applied to locate the odorant concentration which is most suitable for coding. The results are compared with the classical deterministic approach which judges the odorant concentration optimality via steepness of the input-output (transfer) function. The conclusion is, that the most suitable signal, from the point of view of its identification, is not unique and the same holds for most informative signal. The methods are illustrated on realistic models of olfactory sensory neurons, too.

Abstrakt

V první části této práce jsou zmíněny některé základní vlastnosti Fisherovy míry informace skalárního parametru a je vysvětlena spojitost s rozptylem odhadu tohoto parametru. Jsou zadefinovány některé aproximace Fisherovy míry informace a odvozeny jejich vlastnosti. Dále jsou zkoumány některé teoretické modely pro kódování intenzity odorantu v olfaktorních sensorických neuronech. Chování těchto modelů je popsáno pomocí náhodných procesů vázání molekul odorantu na receptory a aktivace vzniklých komplexů. Jsou studovány charakteristiky, popisující jak přesně lze koncentraci odorantu určit ze znalosti odpovědi, koncentrace aktivovaných receptorů. K tomu jsou užity dva odlišné přístupy: první z nich používá Fisherovu míru informace a její aproximace, druhý navrhuje použití některých veličin z teorie informace. Tato kritéria optimality jsou spočítána a aplikována k určení koncentrace odorantu, která je nejvhodnější pro kódování. Výsledky jsou porovnány s klasickým deterministickým přístupem, který posuzuje optimalitu koncentrace odorantu podle strmosti přenosové funkce. Závěrem je, že nejvhodnější signál není určen jednoznačně, a to jak z hlediska jeho identifikace, tak z hlediska přenosu maxima informace. Metody jsou také demonstrovány na realistických modelech olfaktorních neuronů.

Table of contents

List of symbols

$\mathsf{E}(X)$	expected value, mean of random variable X
$\mathsf{E}_\theta(X)$	expected value of random variable X by given parameter θ
$\mathsf{Var}(X)$	variance, dispersion of random variable X
$\mathsf{SD}(X)$	standard deviation of X, $\mathsf{SD}(X) = \sqrt{\mathsf{Var}(X)}$
$\mathsf{N}(\mu, \sigma^2)$	normal (Gaussian) distribution with mean μ and variance σ^2
$\mathsf{U}(a, b)$	continuous uniform distribution on interval (a, b), $a < b$
$\mathsf{B}(a, b)$	Beta distribution with parameters $a > 0$ and $b > 0$
$\mathsf{Bi}(n, \theta)$	binomial distribution with parameters $n \in \mathbb{N}$, $\theta \in [0, 1]$
$\mathsf{P}\{E\}$	probability of event E
$\{X_t; t \geq 0\}$	continuous-time random process
$X\|y; \ X\|Y{=}y$	random variable X conditioned by event $Y{=}y$
$g(x; \theta)$	probability density function of random variable X with distribution dependent on scalar parameter θ
$g(x\|y)$	probability density function of $X\|Y{=}y$
$J(\theta); \ J^X(\theta)$	Fisher information about parameter θ (in random variable X)
$J_k(\theta); \ J_k^X(\theta)$	approximation of $J^X(\theta)$ specified by index k
$\ln x$	natural logarithm (of base e) of x
$\mathrm{e}^x; \ \exp\{x\}$	natural exponential function of x
${}_p\mathrm{F}_q(\boldsymbol{a}, \boldsymbol{b}; x)$	generalized hypergeometric function at x
$\Gamma(x)$	Gamma function at x
$\mathrm{o}(g(t))$	"little-o", represents function $f(t)$ such that $\lim\limits_{t \to \infty} f(t)/g(t) = 0$
$(\Omega, \mathcal{A}, \mathsf{P})$	probability space
$L^2(\Omega, \mathcal{A}, \mathsf{P})$	set of real random variables defined upon $(\Omega, \mathcal{A}, \mathsf{P})$ with finite second moments
$\langle X, Y \rangle$	inner product of random variables X, Y in $L^2(\Omega, \mathcal{A}, \mathsf{P})$ space
$\int_M \mathrm{d}\mu$	Lebesgue integral with respect to measure μ
A	odorant
$\bar{\mathrm{A}}$	degraded odorant
R	free (unbound) receptor
C	activated signaling receptor-odorant complex
C*	inactive receptor-odorant complex

Introduction

Characterization of the input-output properties of sensory neurons and their models is commonly done by using the so called input-output *response functions*, $R(s)$, in which the output is plotted against the strength of the *signal*, s. The output is usually the spiking frequency, or rate of firing, but it can be also the activity or level of any intermediate variable in the transduction cascade, e.g. concentration of activated receptors, effector enzyme concentration, ionic channel activity or receptor potential. The response curves are usually monotonously increasing functions (most often of sigmoid shape) assigning a unique response to an input signal (see Figure 1.1 for illustration). In these curves, there are two special points – the threshold below which the neuron does not respond or only a spontaneous activity, $r_{\min}$, is detected and the first level of the signal at which the response, $r_{\max}$, is saturated. The range of signals between threshold and saturation is called *dynamic (coding) range D*. For formal treatment it is convenient to scale the range of responses, $[r_{\min}, r_{\max}]$, into interval $[0, 1]$ or $[0, r_{\max}]$.

The intuitive concept of *just noticeable difference* has been deeply studied in psychophysics. This concept is also implicitly involved in understanding of signal optimality in neurons. Having the transfer function $R(s)$ and minimum detectable increment ε of the response, we can calculate Δ_s which is the just noticeable difference in the signal. Following from the model given by the response curve, Δ_s depends on the slope of the response curve. In the case of an abstract linear response curve, $R(s) = s/D$, we have $\Delta_s = D\varepsilon$. If the response curve is nonlinear (for example sigmoidal as in Figure 1.1) we can see that Δ_s varies along D and the smallest values of the just noticeable difference in the signal are achieved where the response curve is steepest. Therefore the stimulus intensity for which the signal is optimal, that is the best detectable, is where the slope of the transfer function is highest. This measure of signal optimality is based on deterministic transformation of the signal into a response. In other words it applies to situations in which to each signal s corresponds a deterministically unique response $R(s)$.

However, in practice, an identical signal does not always yield the same response. The variability can be intrinsic (on the way from signal to response) or extrinsic (in the stimulus or when reading the response). These two kinds of
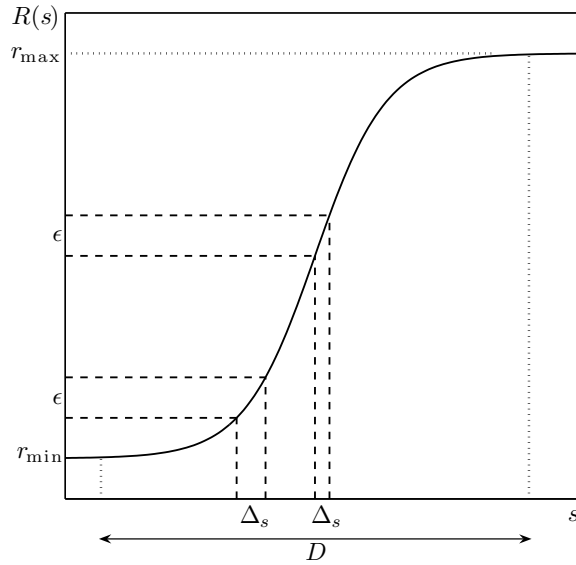
Figure 1.1: A schematic example of transfer function. The dynamic range D, threshold response $r_{\min}$ and maximal discharge $r_{\max}$ are given. The size of the just noticeable difference Δ_s in the signal corresponding to the just noticeable difference ε in the response depends on the slope of the transfer function and is smallest where the slope is highest.

variability are not distinguished here and are collectively considered as noise acting upon the signal-to-response transfer. The presence of noise complicates the concept of signal optimality based on the just noticeable difference. Not only a fixed response (e.g., firing rate) is assigned to every level of the stimulus (as in the classical frequency coding schema), but also a probability distribution of the responses. The noise causes two situations which should be avoided: difference in the responses is due to the noise and not to the signal-false alarm, the signals are different but the difference is not noticed in the response-missed signal. To quantify these effects a new measures for signal optimality is required.

The main aim of this thesis is to propose and investigate alternative measures of signal optimality, that can be applied to noisy transfer in olfactory sensory neurons. It means, the measures have to take into account both factors – the slope of the transfer function and the amplitude of the noise. As a general measure of signal optimality in this case, *Fisher information*, which has become a common tool in computational neuroscience, is proposed. This measure is extended by some *approximations of Fisher information*, which are easier for computation. Other group of applied criteria consists of different *measures of transferred information*, defined in theory of information.

In Chapter 2, some essential theoretical properties of Fisher information are recalled. The inverse of Fisher information $J(\theta)$ about parameter θ gives lower bound for variance of any regular estimator $\hat{\theta}$ of the parameter θ of distribution of

the response. This fact is used as a recipe how to find the signal estimable with the highest precision. In general, the computation of Fisher information is difficult. Some computationally simpler approximations $J_k(\theta)$ of Fisher information are suggested and their relations with Fisher information are derived. Some of the formal mathematical tools reappear in Chapter 3 and Chapter 4 again. This is due to the fact that these chapters are based on already published articles and their structure was designed as individual presentation.

Chapter 3 consists of contents of articles [1, 2], extended by computation of a new criterion. Three theoretical models of concentration detector and a model of flux detector are described. Behavior of the systems is described by stochastic processes of binding the odorant molecules to the receptors and their activation. Characteristics how well the odorant concentration can be estimated from the knowledge of response, the number of activated neurons, are studied. For each model, the steady-state distribution of the number of activated receptors is derived and the criteria of the signal optimality are computed and compared with deterministic result. The contribution of the author is derivation of steady-state distribution of the number of activated receptors in dependency on the odorant intensity and computation of criteria of the signal optimality, introduced in Chapter 2.

Chapter 4 contains mainly extended text of article [4]. Some earlier ideas can be found in article [3], too. The aim of this chapter is to determine the stimulus intensities which can be considered as the most important from two different points of view: transferring as much information as possible and coding the intensity as precisely as possible. These two problems are very different because, for example, an informative signal may be difficult to identify. It is shown, that the role of noise is crucial in both problems. To obtain the range of stimuli which are the best identified, measures based on Fisher information are proposed, as known from the theory of statistical inference. To classify the most important stimuli from the point of view of information transfer, methods based on information theory are suggested. It is presented, that both the most identifiable signal and the most informative signal are not unique. To study this, a generic model of input-output transfer function is analyzed under the influence of several different types of noise. Finally, the methods are illustrated on a biophysical model and empirical data pertaining to olfactory sensory neurons. The contribution of the author is the analysis of quality of the signal with respect to statistical inference based on Fisher information and its approximation described in Chapter 2, the comparison of these criteria with the measures based on information-theory and verification of the methodology on realistic models.

Fisher information and its approximation

2

In this chapter, theoretical statements about Fisher information are given and their connections with variance of a parameter estimator are explained. Further, some new approximations of Fisher information are introduced and their relation with Fisher information is derived.

2.1 Fisher information and its properties

In this section, only basic definitions and theorems from theory of estimation are recalled; see [7, 49] for details. Let us denote $L^2(\Omega, \mathcal{A}, \mathsf{P})$ the set of real random variables defined upon the same probability space $(\Omega, \mathcal{A}, \mathsf{P})$, which have finite second moments, i.e.

$$|\mathsf{E}(X\,Y)| < \infty \text{ for } X, Y \in L^2(\Omega, \mathcal{A}, \mathsf{P}) \ . \tag{2.1}$$

Two random variables $X, Y \in L^2(\Omega, \mathcal{A}, \mathsf{P})$ are equivalent, if they are equal almost everywhere with respect to measure μ, i.e.

$$\mu\left(\{\omega \in \Omega; X(\omega) \neq Y(\omega)\}\right) = 0 \ . \tag{2.2}$$

In this space, for $X, Y \in L^2(\Omega, \mathcal{A}, \mathsf{P})$ let us define inner product

$$\langle X, Y \rangle = \mathsf{E}(X\,Y) \ . \tag{2.3}$$

In this dissertation it is dealt with such random variables $X \in L^2(\Omega, \mathcal{A}, \mathsf{P})$ which have probability density function $f(x; \theta)$ with respect to some countably additive measure μ. The probability density $f(x; \theta)$ is assumed to be dependent on a scalar parameter $\theta \in \Theta$. Thus, the inner product defined by (2.3) can be rewritten explicitly as

$$\langle X, Y \rangle = \mathsf{E}_\theta(X\,Y) \ . \tag{2.4}$$

Definition 1. Class of probability density functions $\{f(x; \theta); \theta \in \Theta\}$ is called *regular* if following conditions hold:

(i) parametric space Θ is open set and $\Theta \neq \emptyset$,

(ii) support $M = \{x \in (-\infty, \infty); f(x; \theta) > 0\}$ does not depend on parameter θ,

(iii) for almost all $x \in M$ with respect to measure $\mu(x)$, finite derivative $\dfrac{\partial f(x; \theta)}{\partial \theta}$ exists, $\left| \dfrac{\partial f(x; \theta)}{\partial \theta} \right| < \infty$,

(iv) for all $\theta \in \Theta$: $\displaystyle\int_M \frac{\partial f(x; \theta)}{\partial \theta}\, \mathrm{d}\mu(x) = 0$,

(v) value $J^X(\theta) = \mathsf{E}_\theta \left(\left(\dfrac{\partial \ln f(X; \theta)}{\partial \theta} \right)^2 \right)$ holds $0 < J^X(\theta) < \infty$.

An estimator $\hat\theta$ of parameter θ in is some function $\hat\theta = H(X)$ of random variable X. It is a random variable again and it depends on actual value of parameter θ.

Definition 2. Estimator $\hat\theta = H(X)$ of parameter θ in random variable X with probability density function $f(x; \theta)$ is called *regular* if following conditions hold:

(vi) the class of probability density functions $\{f(x; \theta); \theta \in \Theta\}$ is regular, i.e. it satisfies conditions (i)–(v) on page 13,

(vii) $\hat\theta$ is unbiased, $\mathsf{E}_\theta \left(\hat\theta \right) = \theta$,

(viii) for all $\theta \in \Theta$: $\displaystyle\int_M H(x) \frac{\partial f(x; \theta)}{\partial \theta}\, \mathrm{d}\mu(x) = \frac{\partial}{\partial \theta} \int_M H(x) f(x; \theta) \mathrm{d}\mu(x)$.

Condition (iv) of Definition 1 says that the exchange of order of derivative and integral in relation

$$0 = \frac{\partial}{\partial \theta} \int_M f(x; \theta) \mathrm{d}\mu(x) = \frac{\partial}{\partial \theta} \left\langle f(X; \theta), \frac{1}{f(X; \theta)} \right\rangle = \left\langle \frac{\partial f(X; \theta)}{\partial \theta}, \frac{1}{f(X; \theta)} \right\rangle \quad (2.5)$$

is justified. Similarly, (viii) of Definition 2 can be rewritten in form

$$\frac{\partial}{\partial \theta} \langle H(X), 1 \rangle = \left\langle H(X) \cdot \frac{\partial f(X; \theta)}{\partial \theta}, \frac{1}{f(X; \theta)} \right\rangle \quad . \quad (2.6)$$

Definition 3. The value

$$J^X(\theta) = \mathsf{E}_\theta \left(\left(\frac{\partial \ln f(X; \theta)}{\partial \theta} \right)^2 \right) = \left\langle \frac{\partial \ln f(X; \theta)}{\partial \theta}, \frac{\partial \ln f(X; \theta)}{\partial \theta} \right\rangle \quad (2.7)$$

is called *Fisher information* about parameter θ in random variable X.

Fisher information is not measure of information in the sense of the theory of information (e.g. like entropy). However, it gives how much "information" is transferred into the distribution of X when the parameter θ changes. In other words, it indicates how precisely the change in parameter can be identified (estimated) from the knowledge of the changed distribution. This point of view is induced by following theorem, which was published in [18].

Theorem 4 (Cramér-Rao). *Let $\hat{\theta} = H(X)$ be regular estimator of parameter θ with finite second moment. Then, for all $\theta \in \Theta$ following statement is fulfilled,*

$$\frac{1}{J^X(\theta)} \le \mathsf{Var}\left(\hat{\theta}\right) \ . \tag{2.8}$$

Inequality (2.8) is called *Cramér-Rao inequality* and gives the lower bound for variance of any regular estimator of the parameter.

Proof. Let us start with Cauchy-Schwarz inequality

$$\left\langle \hat{\theta} - \mathsf{E}_\theta\left(\hat{\theta}\right), \frac{\partial \ln f(X;\theta)}{\partial \theta} \right\rangle^2 \le$$

$$\le \left\langle \hat{\theta} - \mathsf{E}_\theta\left(\hat{\theta}\right), \hat{\theta} - \mathsf{E}_\theta\left(\hat{\theta}\right) \right\rangle \cdot \left\langle \frac{\partial \ln f(X;\theta)}{\partial \theta}, \frac{\partial \ln f(X;\theta)}{\partial \theta} \right\rangle \ , \tag{2.9}$$

where the right hand side is equal to product $\mathsf{Var}\left(\hat{\theta}\right) \cdot J^X(\theta)$. Using relations (2.5) and (2.6), computation of the inner product on the left hand side of (2.9) gives

$$\left\langle \hat{\theta} - \mathsf{E}_\theta\left(\hat{\theta}\right), \frac{\partial \ln f(X;\theta)}{\partial \theta} \right\rangle = \left\langle H(X), \frac{\partial \ln f(X;\theta)}{\partial \theta} \right\rangle - \theta \left\langle 1, \frac{\partial \ln f(X;\theta)}{\partial \theta} \right\rangle =$$

$$= \left\langle H(X) \cdot \frac{\partial f(X;\theta)}{\partial \theta}, \frac{1}{f(X;\theta)} \right\rangle - \theta \underbrace{\left\langle \frac{\partial f(X;\theta)}{\partial \theta}, \frac{1}{f(X;\theta)} \right\rangle}_{0} =$$

$$= \frac{\partial}{\partial \theta}\langle H(X), 1 \rangle = \frac{\partial}{\partial \theta}\mathsf{E}_\theta(H(X)) = \frac{\partial}{\partial \theta}\theta = 1 \ . \tag{2.10}$$

After assignment into (2.9) and rearrangement of the terms, inequality (2.8) is obtained. □

Some authors give more general statement for $\hat{\theta} = H(X)$ which does not need to be unbiased estimator of θ. The proof of this statement is straightforward analogy of the proof of the previous Theorem.

Theorem 5 (Cramér-Rao)**.** *Let $\hat{\theta} = H(X)$ be an estimator of parametric function of θ and let it have finite second moment. If conditions* (vi) *and* (viii) *on page* 13 *are fulfilled and derivative $\frac{\partial \mathsf{E}_\theta(\hat{\theta})}{\partial \theta}$ exists for all $\theta \in \Theta$, then, for all $\theta \in \Theta$ following statement is fulfilled,*

$$\frac{1}{J^X(\theta)} \left(\frac{\partial \mathsf{E}_\theta\left(\hat{\theta}\right)}{\partial \theta} \right)^2 \leq \mathsf{Var}\left(\hat{\theta}\right) \ . \tag{2.11}$$

Assuming we know the best estimator $\hat{\theta} = H(X)$ of θ in the sense of minimal variance (even if there is no general method how to find this estimator), Cramér--Rao Theorem 4 can be seen as relation which gives the quality of estimator $\hat{\theta}$ as a function of the true value of parameter θ. The idea of analysing Fisher information $J^X(\theta)$ as a function of θ to find the *optimal* value of θ, i.e. the value for which the best estimator $\hat{\theta}$ has the lowest variance, is applied and further discussed in Chapters 3 and 4.

Necessary condition on the distribution of X dependent on a parameter θ for existence of an estimator of θ, which holds equality in Cramér-Rao inequality, gives next theorem. Its proof can be found e.g. in [7].

Theorem 6. *If regular estimator $\hat{\theta} = H(X)$ of parameter θ satisfies equality*

$$\mathsf{Var}\left(\hat{\theta}\right) = \left(J^X(\theta)\right)^{-1} \quad \text{for all} \quad \theta \in \Theta \ , \tag{2.12}$$

then the probability density function has form

$$f(x;\theta) = \exp\left\{ c(\theta)H(x) - b(\theta) + a(x) \right\} \ , \tag{2.13}$$

where $a(x), b(\theta), c(\theta)$ are some functions, fulfilling relation

$$\frac{\mathrm{d}c(\theta)}{\mathrm{d}\theta} \cdot \frac{\mathrm{d}b(\theta)}{\mathrm{d}\theta} = \frac{1}{\theta} \ . \tag{2.14}$$

Having the form (2.13), probability density function $f(x;\theta)$ belongs to *exponential class* with respect to parameter θ, for details see Definition 12 and Remark 13. Let us only briefly recall the sufficient condition for achieving equality (2.12). Theorems of *Lehmann-Scheffé* and *Rao-Blackwell* give a recipe where to search for such an estimator: this regular estimator $\hat{\theta}$ should be a *complete sufficient statistics* for θ in X.

2.2 Approximation of Fisher information

In general, it is difficult task to compute the Fisher information analytically. Usually the integral has to be computed numerically. Moreover, having only measured data without the knowledge of their distribution (which is a typical situation), it is impossible to compute the Fisher information without estimation of

the probability density function. These reasons lead to search for some approximation of the Fisher information. Following definition introduce a sequence of such approximations. It is an extension of definition of approximation $J_2^X(\theta)$, which was already used by several authors. For references, see Chapters 3 and 4.

Definition 7. Let X be random variable with probability density function $f(x; \theta)$ dependent on a scalar parameter $\theta \in \Theta$. For $k = 2, 3, \ldots$ the sequence of approximations is defined,

$$J_k^X(\theta) = \frac{1}{\mathsf{Var}(X^{k-1})} \left(\frac{\partial \mathsf{E}(X^{k-1})}{\partial \theta} \right)^2 . \tag{2.15}$$

For $k = 2, 3, \ldots$, let us introduce a sequence of conditions similar to relation (2.6),

$$\frac{\partial}{\partial \theta} \left\langle X^{k-1}, 1 \right\rangle = \frac{\partial}{\partial \theta} \left\langle X^{k-1} \cdot f(x; \theta), \frac{1}{f(x; \theta)} \right\rangle = \left\langle \frac{\partial}{\partial \theta} \left(X^{k-1} \cdot f(x; \theta) \right), \frac{1}{f(x; \theta)} \right\rangle . \tag{2.16}$$

Following theorem says that approximations $J_k^X(\theta)$ given by Definition 7 are lower bounds for Fisher information $J^X(\theta)$.

Theorem 8. *Let $X \in L^2(\Omega, \mathcal{A}, \mathsf{P})$ be random variable with probability density function $f(x; \theta)$ dependent on a scalar parameter $\theta \in \Theta$. If the class $\{f(x; \theta); \theta \in \Theta\}$ satisfies regularity conditions* (i)–(v) *of Definition 1, then, for those $k = 2, 3, \ldots$ for which relation (2.16) is satisfied for all $\theta \in \Theta$, there is inequality*

$$J_k^X(\theta) \leq J^X(\theta) \qquad \text{for all } \theta \in \Theta . \tag{2.17}$$

Proof. The principal idea of proof is similar to that one of proof of Cramér-Rao Theorem 4 and it uses Cauchy-Schwarz inequality

$$\left\langle X^{k-1} - \mathsf{E}_\theta(X^{k-1}), \frac{\partial \ln f(X; \theta)}{\partial \theta} \right\rangle^2 \leq$$
$$\leq \left\langle X^{k-1} - \mathsf{E}_\theta(X^{k-1}), X^{k-1} - \mathsf{E}_\theta(X^{k-1}) \right\rangle \cdot \left\langle \frac{\partial \ln f(X; \theta)}{\partial \theta}, \frac{\partial \ln f(X; \theta)}{\partial \theta} \right\rangle \tag{2.18}$$

for $k = 1, 2, \ldots$, where $f(X; \theta)$ denotes the probability density function of X with respect to the parameter θ. It is easy to verify that the right hand side of (2.18) is equal to the product $\mathsf{Var}(X^{k-1}) J^X(\theta)$ of variance of X^{k-1} and Fisher information

about θ in X. Further, let us compute the inner product on the left hand side of (2.18) using relations (2.5) and (2.16),

$$
\left\langle X^{k-1} - \mathsf{E}_\theta\big(X^{k-1}\big), \frac{\partial \ln f(X;\theta)}{\partial \theta} \right\rangle =
$$

$$
= \left\langle X^{k-1}, \frac{\partial \ln f(X;\theta)}{\partial \theta} \right\rangle - \left\langle \mathsf{E}_\theta\big(X^{k-1}\big), \frac{\partial \ln f(X;\theta)}{\partial \theta} \right\rangle =
$$

$$
= \left\langle X^{k-1} \cdot \frac{\partial f(X;\theta)}{\partial \theta}, \frac{1}{f(X;\theta)} \right\rangle - \mathsf{E}_\theta\big(X^{k-1}\big) \cdot \underbrace{\left\langle \frac{\partial f(X;\theta)}{\partial \theta}, \frac{1}{f(X;\theta)} \right\rangle}_{0} =
$$

$$
= \left\langle X^{k-1} \cdot \frac{\partial f(X;\theta)}{\partial \theta}, \frac{1}{f(X;\theta)} \right\rangle = \left\langle \frac{\partial}{\partial \theta}\big(X^{k-1} \cdot f(X;\theta)\big), \frac{1}{f(X;\theta)} \right\rangle =
$$

$$
= \frac{\partial}{\partial \theta} \left\langle X^{k-1} \cdot f(X;\theta), \frac{1}{f(X;\theta)} \right\rangle = \frac{\partial}{\partial \theta} \left\langle X^{k-1}, 1 \right\rangle = \frac{\partial \mathsf{E}_\theta\big(X^{k-1}\big)}{\partial \theta} \quad . \tag{2.19}
$$

Finally, substituting (2.19) into left hand side of (2.18) and dividing by $\mathsf{Var}\big(X^{k-1}\big)$ the following inequality is obtained,

$$
\frac{1}{\mathsf{Var}(X^{k-1})} \left(\frac{\partial \mathsf{E}_\theta\big(X^{k-1}\big)}{\partial \theta} \right)^2 \le \left\langle \frac{\partial \ln f(X;\theta)}{\partial \theta}, \frac{\partial \ln f(X;\theta)}{\partial \theta} \right\rangle \quad , \tag{2.20}
$$

which gives the statement of the theorem. $\qquad\square$

Lemma 9. *Let $X \in L^2(\Omega, \mathcal{A}, \mathsf{P})$ be random variable with probability density function $f(x;\theta)$ dependent on a scalar parameter $\theta \in \Theta$. If the class $\{f(x;\theta); \theta \in \Theta\}$ satisfies regularity conditions* (i)–(v) *of Definition 1 and if for all $\theta \in \Theta$ relation (2.16) is fulfilled for some $k \in \{2, 3, \dots\}$, then, there is equality*

$$
J_k^X(\theta) = J^X(\theta) \qquad \text{for all } \theta \in \Theta \tag{2.21}
$$

if and only if

$$
\frac{\partial \ln f(x;\theta)}{\partial \theta} = r(\theta) \cdot \big(x^{k-1} - \mathsf{E}_\theta\big(X^{k-1}\big)\big) \tag{2.22}
$$

for some function $r(\theta)$.

Proof. Equality in Cauchy-Schwarz inequality (2.18) is achieved if and only if functions $X^k - \mathsf{E}_\theta\big(X^k\big)$ and $\dfrac{\partial \ln f}{\partial \theta}$ are linearly dependent. Omitting the irrelevant case of null function (which means that $f(x;\theta) = f(x)$ does not depend on parameter θ), it means that

$$
\frac{\partial \ln f(x;\theta)}{\partial \theta} = r(\theta) \cdot \big(x^{k-1} - \mathsf{E}_\theta\big(X^{k-1}\big)\big) \quad , \tag{2.23}
$$

where $r(\theta)$ is a constant or a function dependent on θ only. $\qquad\square$

Lemma 10. *Under the same conditions as in Lemma 9, the following conditions on the probability density function are equivalent:*

(i) *for some function $r(\theta)$,*

$$\frac{\partial \ln f(x;\theta)}{\partial \theta} = r(\theta)\cdot\big(x^{k-1} - \mathsf{E}_\theta\big(X^{k-1}\big)\big) \ ; \tag{2.24}$$

(ii) *for some functions $a(x)$, $b(\theta)$ and $c(\theta)$, $f(x;\theta)$ has form*

$$f(x;\theta) = \exp\big\{x^{k-1}c(\theta) - b(\theta) + a(x)\big\} \ . \tag{2.25}$$

Proof. (i) $\Rightarrow$ (ii): Let us assume that the derivative of $\ln f(x;\theta)$ with respect to θ can be rewritten in form (2.24). Integrating (2.24) with respect to θ (over the support M of x) and using notation

$$c(\theta) = \int_M r(\theta)\mathrm{d}\theta \ , \qquad b(\theta) = \int_M r(\theta)\mathsf{E}_\theta\big(X^{k-1}\big)\,\mathrm{d}\theta \tag{2.26}$$

it gives

$$\ln f(x;\theta) = x^{k-1}c(\theta) - b(\theta) + a(x) \ , \tag{2.27}$$

where $a(x)$ is the integral constant (can be a function of x). Taking the exponential of (2.27), the expression (2.25) is obtained.

(ii) $\Rightarrow$ (i): Assuming the form (2.25) of $f(x;\theta)$, derivative of its logarithm is directly computed,

$$\frac{\partial \ln f(x;\theta)}{\partial \theta} = x^{k-1}\frac{\mathrm{d}c(\theta)}{\mathrm{d}\theta} - \frac{\mathrm{d}b(\theta)}{\mathrm{d}\theta} \ . \tag{2.28}$$

Let us recall the relation (2.5), which implies that

$$\left\langle \frac{\partial \ln f(x;\theta)}{\partial \theta}, 1 \right\rangle = \frac{\partial}{\partial \theta} \langle 1, 1 \rangle = 0 \ . \tag{2.29}$$

This property applied on (2.28) gives relation

$$\frac{\mathrm{d}b(\theta)}{\mathrm{d}\theta} = \mathsf{E}_\theta\big(X^{k-1}\big) \frac{\mathrm{d}c(\theta)}{\mathrm{d}\theta} \ . \tag{2.30}$$

Finally, substituting (2.30) into (2.28) and denoting $r(\theta) = \dfrac{\mathrm{d}c(\theta)}{\mathrm{d}\theta}$, the expression (2.24) for $\dfrac{\partial \ln f(x;\theta)}{\partial \theta}$ is obtained. $\qquad\square$

Theorem 11. *Let $X \in L^2(\Omega, \mathcal{A}, \mathsf{P})$ be random variable with probability density function $f(x;\theta)$ dependent on a scalar parameter $\theta \in \Theta$. If the class $\{f(x;\theta); \theta \in \Theta\}$ satisfies regularity conditions* (i)–(v) *of Definition* 1, *then, for those $k = 2, 3, \ldots$ for which relation* (2.16) *is satisfied for all $\theta \in \Theta$, equality*

$$J_k^X(\theta) = J^X(\theta) \qquad \text{for all } \theta \in \Theta \tag{2.31}$$

is fulfilled if and only if the probability density function $f(x;\theta)$ of random variable X has the form

$$f(x;\theta) = \exp\left\{ x^{k-1} c(\theta) - b(\theta) + a(x) \right\} \tag{2.32}$$

for some functions $a(x), b(\theta), c(\theta)$.

Proof. This theorem is direct consequence of joint statements of Lemma 9 and Lemma 10. The way of proof follows the idea of equivalent conditions for holding the equality $J^X(\theta) = J_2^X(\theta)$ published in [44]. $\qquad\qquad\square$

Definition 12. Random variable X has a distribution belonging to *exponential class* with respect to parameter θ and with power k, $k = 1, 2, \ldots$, if probability density function of X takes the form

$$f(x;\theta) = \exp\left\{ x^k c(\theta) - b(\theta) + a(x) \right\} \tag{2.33}$$

for some functions $a(x), b(\theta), c(\theta)$.

Let us note, that form (2.33) is equal to (2.13) with estimator $H(x) = x^k$. Definition 12 allows us to formulate the statement of Theorem 11 by equivalence: $J_k^X(\theta) = J^X(\theta)$ for all $\theta \in \Theta$ if and only if $f(x;\theta)$ belongs to the *exponential class* with respect to parameter θ and with power $k - 1$.

Remark 13. In [53], the exponential class is defined as probability density functions of the form

$$f(x;\theta,\phi) = \exp\left\{ \frac{x\theta - b(\theta)}{a(\phi)} + c(x,\phi) \right\}, \tag{2.34}$$

where ϕ is nuisance parameter. Knowing ϕ, it is called exponential class with canonical parameter θ. After necessary change of notation, it leads to exponential class with power 1 by Definition 12 with $k = 1$.

Example 14. Let us suppose that random variable X has Gaussian distribution $X \sim \mathsf{N}(\mu, \sigma^2)$ with known variance σ^2. Fisher information about the unknown mean value μ,
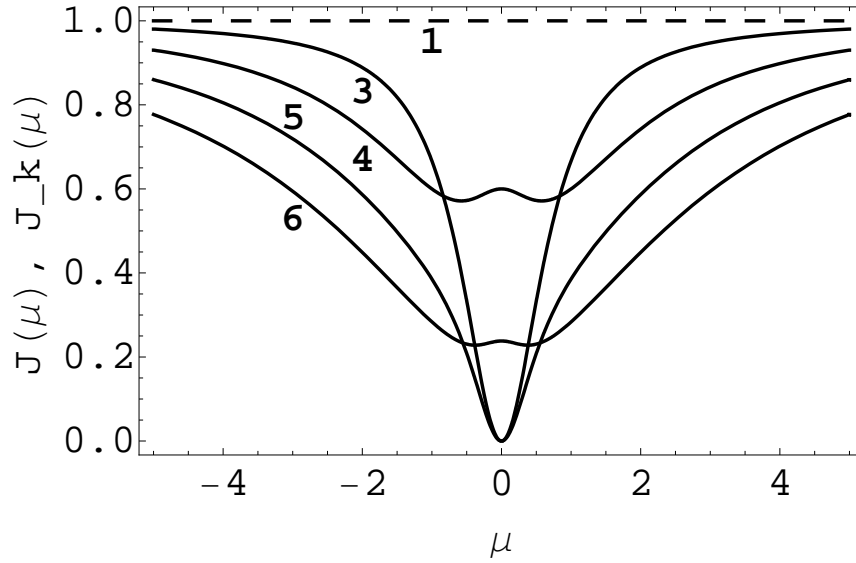
$$J^X(\mu) = \frac{1}{\sigma^2}, \tag{2.35}$$

Figure 2.1: Fisher information $J^X(\mu)$ (dashed curve 1) and its approximations $J_k^X(\mu)$ for $k = 3, 4, 5, 6$ (curves 3–6) computed from random variable $X \sim \mathsf{N}(\mu, \sigma^2 = 1)$.

does not depend on the true value μ (it means, all values of mean are estimable with equal accuracy, which only depends on the variance) and its approximations for $k = 2, 3, 4$ are equal to

$$J_2^X(\mu) = \frac{1}{\sigma^2} \quad, \tag{2.36}$$

$$J_3^X(\mu) = \frac{2\mu^2}{2\mu^2\sigma^2 + \sigma^4} \quad, \tag{2.37}$$

$$J_4^X(\mu) = \frac{3\left(\mu^2 + \sigma^2\right)^2}{3\mu^4\sigma^2 + 12\mu^2\sigma^4 + 5\sigma^6} \quad. \tag{2.38}$$

Fisher information $J^X(\mu)$ and approximations $J_k^X(\mu)$ for $k = 3, 4, 5, 6$ are depicted in Figure 2.1. The approximation $J_2^X(\mu)$ is accurate, $J_2^X(\mu) = J^X(\mu) = \sigma^{-2}$. This result corresponds with the statement of Theorem 11, because Gaussian distribution belongs to the exponential class with respect to parameter μ with power $k = 1$, e.g. for $\sigma^2 = 1$, probability density function is

$$f(x; \mu) = \exp\left\{x^1\mu - \frac{\mu^2}{2} - \frac{x^2}{2} - \frac{\ln 2\pi}{2}\right\} \quad. \tag{2.39}$$

Example 15. Let us suppose that random variable X has Gaussian distribution $X \sim \mathsf{N}(\mu, \sigma^2)$ with known mean μ. Fisher information about the unknown standard deviation σ,

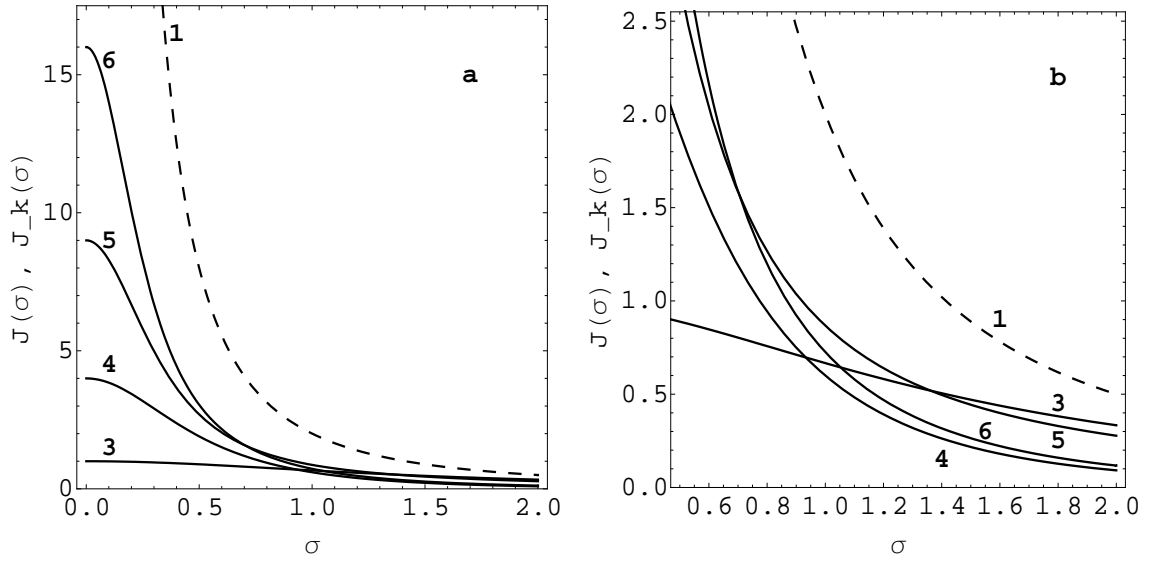$$J^X(\sigma) = \frac{2}{\sigma^2} \quad, \tag{2.40}$$

Figure 2.2: (a) Fisher information $J^X(\sigma)$ (dashed curve 1) and its approximations $J_k^X(\mu)$ for $k = 3, 4, 5, 6$ (curves 3–6) computed from random variable $X \sim \mathsf{N}(\mu = 1, \sigma^2)$. (b) Detail of (a).

depends on the true value σ (the larger the true variance is the worse estimate of the standard deviation can be achieved) and its approximations for $k = 2, 3, 4$ are equal to

$$J_2^X(\sigma) = 0 \ , \tag{2.41}$$

$$J_3^X(\sigma) = \frac{2}{2\mu^2 + \sigma^2} \ , \tag{2.42}$$

$$J_4^X(\sigma) = \frac{12\mu^2}{3\mu^4 + 12\mu^2\sigma^2 + 5\sigma^4} \ . \tag{2.43}$$

Fisher information $J^X(\mu)$ and approximations $J_k^X(\mu)$ for $k = 3, 4, 5, 6$ are depicted in Figure 2.2. The approximation $J_2^X(\mu)$ gives no information about the quality of estimate of σ, $J_2^X(\mu)$ is accurate for $\mu = 0$.

Example 16. As the last example, let us assume random variable X with Gaussian distribution $X \sim \mathsf{N}(\mu(\theta), \sigma^2(\theta))$ with both the mean and the variance dependent on a parameter $\theta > 0$,

$$\mu(\theta) = \frac{\theta}{\theta + 1} \ , \qquad \sigma^2(\theta) = \frac{\theta}{(\theta + 1)^2} \ . \tag{2.44}$$

Fisher information about the parameter θ and its approximations for $k = 2, 3$ are
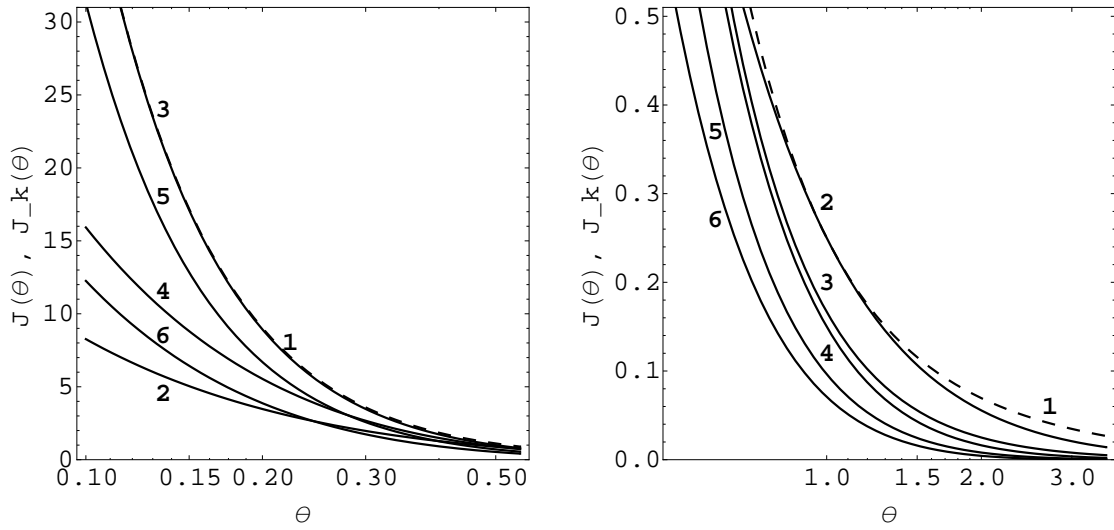
Figure 2.3: Both figures depict Fisher information $J^X(\theta)$ (dashed curve 1) and its approximations $J_k^X(\theta)$ for $k = 2, 3, 4, 5, 6$ (curves 2–6) computed from random variable $X \sim \mathsf{N}\left(\frac{\theta}{\theta+1}, \frac{\theta}{(\theta+1)^2}\right)$ with coefficient of variation $1/\sqrt{\theta}$ dependent on parameter $\theta > 0$. The horizontal axes have logarithmic scales.

equal to

$$J^X(\theta) = \frac{\theta^2 + 1}{2\theta^2(\theta+1)^2} \quad , \tag{2.45}$$

$$J_2^X(\theta) = \frac{1}{\theta(\theta+1)^2} \quad , \tag{2.46}$$

$$J_3^X(\theta) = \frac{1}{2\theta^2 + 4\theta^3} \quad . \tag{2.47}$$

Fisher information $J^X(\mu)$ and approximations $J_k^X(\mu)$ for $k = 2, 3, 4, 5, 6$ are depicted in Figure 2.3.

Considering the Definition 7 and Theorem 11 one could expect that the approximations $J_k^X(\theta)$ would be ordered (and would converge to the Fisher information). But, this is not true in general, as can be seen on results of preceding examples. However, some kinds of asymptotical ordering of $J^X(\theta)$ and convergence to $J^X(\theta)$ can be seen in Figure 2.1, Figure 2.2 and Figure 2.3. These ordering and convergence differ case to case.

Models of olfactory neurons **3**

This chapter consists of contents of articles [1, 2], extended by computation of a new criterion. Some theoretical models of olfactory neurons are described using stochastic processes, the steady-state distribution of number of activated receptors and criteria of the signal optimality are derived and compared with known results.

3.1 Introduction

Signal processing in olfactory systems is initialized by binding of odorant molecules to receptor molecules embedded in the membranes of sensory neurons. Binding of odorants and receptor activation trigger a sequence of biochemical events that result in the opening of ionic channels, the generation of receptor potential which triggers a train of action potentials. Our models of the binding and activation of receptor sites are based on models proposed by Beidler [9], Brown and Rothery [13], Cleland and Linster [17], Kaissling [30, 29], Lánský et al. [42, 41], Lauffenburger and Linderman [46], Mankin and Mayer [50], Maurin [51], Rospars et al. [62, 60]. The models can be, in general, classified into two categories, concentration detectors and flux detectors. In the concentration detector models it is assumed that the neuronal membrane is directly exposed to the odorant molecules present in the external space. In other words, it is assumed that the transfer of odorant molecules between the external space and the perireceptor space is very fast and reversible. Thus the odorant concentrations in both compartments are the same. We investigate three types of the concentration detectors which differ in the binding and activation mechanisms. In the flux detector model it is assumed that the transfer of odorants from the external to the perireceptor space is relatively slow and irreversible. Here, the degradation must be included to compensate for the fact that no possible outflow of the odorant occurs (see [30] for details).

All presented models aim to study theoretically how the concentration of an odorant influences the steady-state number of activated receptor proteins and how precisely the number of activated receptors can code the intensity of the
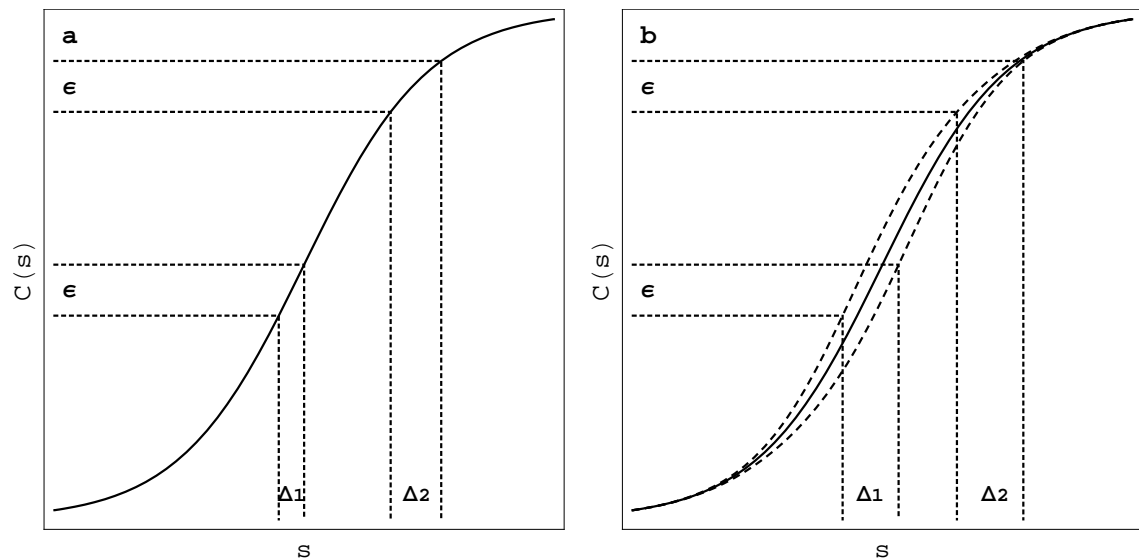
Figure 3.1:   Two concepts in search for the optimal odorant concentration for the ol-factory neuron. The solid line describes the dependency between the odorant log-con-centration, s (horizontal axis), and the response, the count of activated receptors (ver-tical axis). (a) Deterministic approach is based on the slope of the input-output func-tion $C(s)$ (solid line). (b) Stochastic approach is based on the statistical properties of the response; the mean response $\mathsf{E}(C(s))$ (solid line) accompanied by standard deviations $\mathsf{E}(C(s)) \pm \mathsf{SD}(C(s))$ (dashed curves) are plotted. Dotted lines denotes the connection be-tween change of response ε and corresponding change Δ of the signal. For details see text.

odorant. This is done either by investigating the classical deterministic models or by using the statistical properties of the responses to the different concentrations of the odorant. Under the statistical approach the behaviour of the number of activated receptors is represented as a stochastic process with the odorant con-centration as a parameter. Such a description was used for example by Arányi and Tóth [8], Gurewich et al. [23], Kurtz [38], Lam and Lampard [39], Lánský and Rospars [40], McNeil and Schach [54] and Tuckwell [67].

In the models, as well as in reality, neurons do not code all the concentrations of odorant equally well. Here we compare the classical approach to the optimum signal determination with the definition based on the application of Fisher in-formation measure. Application of Fisher information in studies on the neural computation has become a common tool recently; for example see Brunel and Nadal [14], Johnson and Ray [28], Lánský et al. [45, 43], Sanchez-Montanes and Pearce [64], Wilke and Eurich [68].

Both the deterministic and the stochastic concepts are described schematically in Figure 3.1. The classical deterministic approach is depicted in Figure 3.1a, where a unique response, $C(s)$, is plotted against the odorant log-concentrati-

on, s. The term "log-concentration" is meant as a quantity of molecules related to a volume unit and expressed in natural logarithmic scale. Two equal changes ε in response, the number of activated receptors $C(s)$, are caused by different changes in corresponding odorant concentrations, $\Delta_1 < \Delta_2$, because of varying slope of the input-output function. Therefore, the changes in the odorant concentration in region around Δ_1 can be determined from the knowledge of the response $C(s)$ more precisely than in region around Δ_2.

In the statistical approach, depicted in Figure 3.1b, not only the shape of the mean response function but also fluctuations (deviations) of the response are taken into account and plotted versus the odorant log-concentration, s. Hence the changes Δ_1 and Δ_2 in odorant log-concentration are different from the situation in Figure 3.1a. Due to the larger variability of the responses in the central part of the transfer function we obtain $\Delta_1 \approx \Delta_2$.

Describing the behaviour of interaction with stochastic processes we gain the probability distribution of the response $C(s)$ and it is used to search for the optimal signal. In models constructed below, the asymptotic mean of the process always coincides with the deterministic input-output function. However, this does not need to be a general property in dependency on complexity of the model.

3.2 Models

Presented models consider interaction between odorant molecules and receptors on the surface of olfactory receptor neurons. We assume that there is only one odorant substance, that each receptor molecule possesses only one binding site and that the total number of the receptors on the surface of the membrane is fixed and equal to N. Let A denote the odorant molecule in perireceptor space and A_E in external space, with concentration $A = e^s$ (in concentration detector) or $A_E = e^s$ (in flux detector), which is assumed to be fixed until the olfactory system achieves the steady state. We distinguish three states in which the receptors can appear: unbound (free) state, R, bound inactive state, C^*, and bound activated state C. Only activated receptors trigger the response. The models assume that the response, the count of activated receptors, $C(s)$, in steady state is completely determined by the signal, which is fixed log-concentration, s, of odorant. Thus, in the models investigated here the count is a dependent variable with the odorant log-concentration, s, as a parameter. Three models of concentration detectors and one flux detector model as stochastic processes, $\{C_t(s); \ t \geq 0\}$, are described now. Being interested in the steady state of these processes, we study the random variable $C(s) = C_\infty(s) = \lim_{t \to \infty} C_t(s)$.

3.2.1 Basic model

At first we consider the simplest model in which each occupied receptor becomes activated instantaneously with its occupation. Thus, the number of bound but not activated receptors is always zero. It is assumed that each receptor is occupied and released independently of others in accordance with stochastic reaction schema

$$A + R \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C \; , \tag{3.1}$$

where A represents an unbound molecule of odorant, R unoccupied receptor and C stands for bound activated receptor (complex of the odorant molecule and the receptor), k_1 and k_{-1} are fixed reaction rates coefficients of association and dissociation of the odorant molecules. The ratio $K_1 = k_{-1}/k_1$ is commonly called the dissociation constant. A receptor unbound at time t becomes bound (and activated) in time interval $(t, t+\Delta t]$ with probability $\lambda \Delta t + o(\Delta t)$, a receptor activated at time t becomes unbound in time interval $(t, t+\Delta t]$ with probability $\mu \Delta t + o(\Delta t)$. The parameter λ is an increasing function of the concentration, A, of the odorant. We consider this dependency to be linear, $\lambda = k_1 A = k_1 e^s$, with the constant of proportionality k_1 equal to the rate of association. The parameter μ is considered to be equal to the rate of dissociation, $\mu = k_{-1}$. As the total number of receptor sites on the surface of the membrane is equal to N, relation $R_t + C_t = N$ is satisfied for all $t \geq 0$.

In accordance with these assumptions the continuous-time stochastic process $\{C_t; \; t \geq 0\}$ giving the count of bound activated receptors at time t can be described as a birth and death process (see [38, 39, 40]) with birth rates λ_i and death rates μ_i,

$$\lambda_i = \lambda(N - i) = k_1(N - i)e^s \; , \tag{3.2}$$

$$\mu_i = \mu i = k_{-1}i \tag{3.3}$$

for $i \in \{0, 1, \ldots, N\}$. It means that the transition probabilities are

$$\begin{aligned}
\mathsf{P}\left\{(i) \xrightarrow{\Delta t} (i+1)\right\} &= k_1(N-i)e^s\Delta t + o(\Delta t) \; , \\
\mathsf{P}\left\{(i) \xrightarrow{\Delta t} (i-1)\right\} &= k_{-1}i\Delta t \qquad\quad + o(\Delta t) \; .
\end{aligned} \tag{3.4}$$

This process, independently of the initial condition, achieves a stationary state with probability distribution with mass function,

$$\pi_i = \mathsf{P}\{C(s) = i\} = \left(1 + \frac{e^s}{K_1}\right)^{-N} \binom{N}{i} \left(\frac{e^s}{K_1}\right)^i \; , \quad i \in \{0, 1, \ldots, N\} \; , \tag{3.5}$$

see for example [33]. Using this stationary distribution to derive the mean and variance of the count of activated receptors in steady state, $C(s)$, we obtain

$$\mathsf{E}(C(s)) = \frac{N}{1 + K_1 \mathrm{e}^{-s}} \quad , \tag{3.6}$$

$$\mathsf{Var}(C(s)) = \frac{N K_1 \mathrm{e}^{-s}}{(1 + K_1 \mathrm{e}^{-s})^2} \quad , \tag{3.7}$$

$$\mathsf{E}(C^2(s)) = \frac{N^2 + N K_1 \mathrm{e}^{-s}}{(1 + K_1 \mathrm{e}^{-s})^2} \quad , \tag{3.8}$$

$$\mathsf{E}(C^4(s)) = \frac{N (K_1 \mathrm{e}^{-s})^{N-1}}{(1 + K_1 \mathrm{e}^{-s})^N} \; {}_4\mathrm{F}_3\left(2,2,2,1{-}N;1,1,1;-\frac{\mathrm{e}^s}{K_1}\right) \quad , \tag{3.9}$$

where

$$_p\mathrm{F}_q(\boldsymbol{a}, \boldsymbol{b}; x) = {}_p\mathrm{F}_q(a_1, \ldots, a_p; b_1, \ldots, b_q; x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \frac{\displaystyle\prod_{i=1}^{p} \frac{\Gamma(a_i + k)}{\Gamma(a_i)}}{\displaystyle\prod_{j=1}^{q} \frac{\Gamma(b_j + k)}{\Gamma(b_j)}} \tag{3.10}$$

stands for generalized hypergeometric function at x (for details see [5]) and $\Gamma(x)$ denotes Gamma function. Hence, we have

$$_4\mathrm{F}_3(2,2,2,1{-}N;1,1,1;x) = 1 + \sum_{k=1}^{\infty} \frac{x^k}{k!}(k+1)^3 \prod_{i=1}^{k}(i - N) \quad . \tag{3.11}$$

As a function of s, the steady-state mean given by equation (3.6) monotonically increases from 0 to N and it has typical sigmoidal shape with inflexion point located at $s = \ln K_1$. Variance (3.7) has unimodal shape and, because of satisfying of relation $\partial \mathsf{E}(C(s))/\partial s = \mathsf{Var}(C(s))$, its maximum value is achieved for the same odorant log-concentration $s = \ln K_1$. For extremely low and high odorant concentrations the variance tends to zero, $\mathsf{Var}(C(\pm\infty)) = 0$. The mean and standard deviation as functions of the log-concentration of odorant are plotted in Figure 3.2.

If the number of receptor sites, N, is sufficiently high, it is natural to consider continuous variant of the steady-state count of activated receptors $C(s)$ and we need to know the distribution of such continuous approximation. One possibility how to do it is to use the central limit theorem (for details see for example [49]). The other legitimate approach is to use the diffusion approximation of the birth and death process (3.3), as described for example in [64, 67]. Following Tuckwell [67], the stochastic process $\{C_t; \; t \geq 0\}$ can be approximated by the process $\{Y_t; \; t \geq 0\}$,

$$C_t \approx Y_t = \frac{\lambda N}{\lambda + \mu} + \sqrt{N}\, U_t = \frac{N}{1 + K_1 \mathrm{e}^{-s}} + \sqrt{N}\, U_t \quad , \tag{3.12}$$
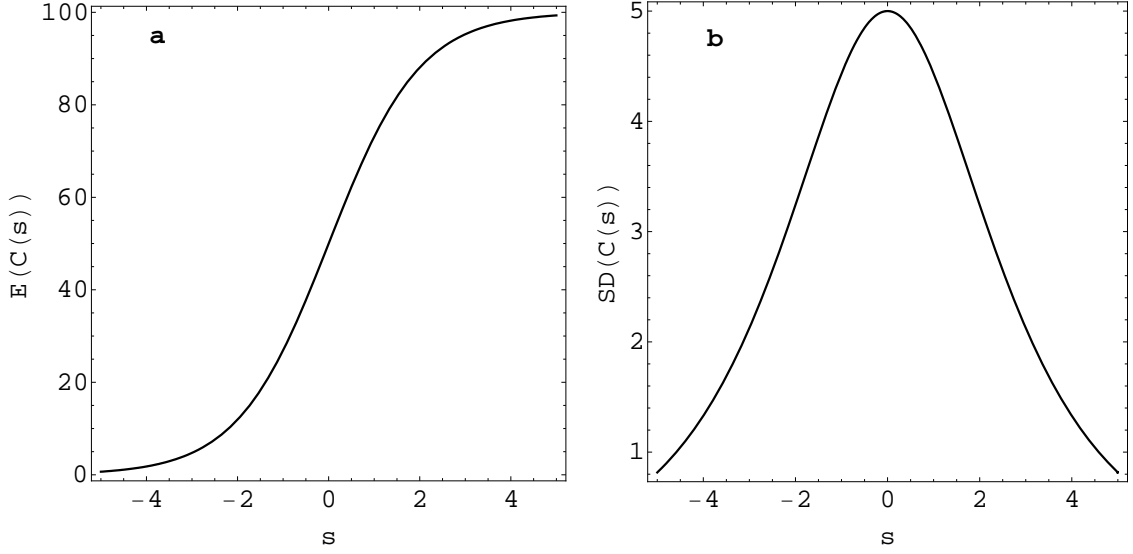
Figure 3.2:   (a) Mean $\mathsf{E}(C(s))$ and (b) standard deviation $\mathsf{SD}(C(s))$ of the number of activated receptors in the basic model, as functions of the odorant log-concentration, s, in perireceptor space. Parameters are $K_1 = 1$ and $N = 100$. Both the inflexion point of the mean and the point of maximal standard deviation are located at $s = \ln K_1 = 0$.

where U_t is an Ornstein-Uhlenbeck stochastic process. Process U_t is given by the stochastic differential equation

$$\mathrm{d}U_t = -\left(k_1 \mathrm{e}^s + k_{-1}\right) U_t \,\mathrm{d}t + \sqrt{\frac{2k_1 k_{-1} \mathrm{e}^s}{k_1 \mathrm{e}^s + k_{-1}}} \,\mathrm{d}W_t \ , \tag{3.13}$$

where $\{W_t; \ t \geq 0\}$ is the standard Wiener process. The properties of Y_t follow directly from the fact that U_t is the Ornstein-Uhlenbeck process with drift $-(k_1 \mathrm{e}^s + k_{-1})$ (see [33]). The process Y_t is Gaussian and its steady-state moments are given by equations (3.6) and (3.7). Therefore, the discrete stationary distribution of the steady-state count $C(s)$ with mass function π_i can be approximated by a normal random variable with probability density function

$$g(x; s) = \frac{1 + K_1 \mathrm{e}^{-s}}{\sqrt{2\pi N K_1 \mathrm{e}^{-s}}} \exp\left\{ -\frac{(N - x - xK_1 \mathrm{e}^{-s})^2}{2N K_1 \mathrm{e}{-s}} \right\} \tag{3.14}$$

with s as a parameter. Note that due to the diffusion approximation the state space of the continuous variant of the random variable $C(s)$ is, at least formally, the whole range, $(-\infty, \infty)$. However, its validity holds only if π_0 and π_N are very small.

Hereafter, we use the notation $C(s)$ for both discrete as well as continuous variant of the steady-state number of activated receptors. In case of eventual misunderstanding we explicitly emphasize the type of variant.

3.2.2 Model of simple activation

Consider now the model where not every bound receptor (complex) is activated immediately. The receptors really appear in three states: unbound, R, occupied but not activated, C^*, and occupied activated, C. Model described by Lánský and Rospars [40] supposes that each occupied receptor can either become activated, C, with probability $p \in (0, 1)$, or stay inactive, C^*, with probability $1 - p$, independently of its past behavior and of the behavior of other receptors. Such an interaction corresponds to the following reaction schema,

$$
C^* \; \underset{k_{1N}}{\overset{k_{-1}}{\rightleftharpoons}} \; A + R \; \underset{k_{-1}}{\overset{k_{1A}}{\rightleftharpoons}} \; C \; , \tag{3.15}
$$
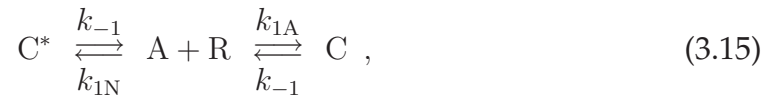
where $k_{1A} = pk_1$ and $k_{1N} = (1 - p)k_1$ are association rate coefficients for the activated and inactive state and k_1, k_{-1} have the same meaning as in basic model (3.1).

Denoting by B_t the number of bound receptors, B, at time t, regardless of their activation, the relation $B_t = C_t + C_t^*$ is satisfied. Then, because of the independence of behaviour of the receptor sites the binding process follows reaction schema

$$
A + R \; \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \; B \tag{3.16}
$$

(compare with (3.1)) and provided $B_t = b$ the random variable C_t has binomial distribution $C_t \sim \mathsf{Bi}(b, p)$ for all $t \geq 0$. Then, the conditional steady-state distribution is binomial, too, $(C_\infty | B_\infty = b) \sim \mathsf{Bi}(b, p)$. From the knowledge of the mass function (3.5) of the random variable B_∞, the unconditional probability distribution of $C(s)$ can be derived,

$$
\pi_i = \mathsf{P}\{C(s) = i\} = \binom{N}{i} q(s)^i \left(1 - q(s)\right)^{N-i} \; ; \quad i \in \{0, 1, \ldots, N\} \; , \tag{3.17}
$$

where $q(s) = p/(1 + K_1 \mathrm{e}^{-s})$. Hence, the steady-state number of activated receptors has binomial distribution $C(s) \sim \mathsf{Bi}(N, q(s))$ and its mean and variance can be directly derived,

$$
\mathsf{E}(C(s)) = \frac{Np}{1 + K_1 \mathrm{e}^{-s}} \; , \tag{3.18}
$$

$$
\mathsf{Var}(C(s)) = \frac{Np K_1 \mathrm{e}^{-s}}{(1 + K_1 \mathrm{e}^{-s})^2} + \frac{Np(1 - p)}{(1 + K_1 \mathrm{e}^{-s})^2} \; , \tag{3.19}
$$

$$
\mathsf{E}(C^2(s)) = \frac{Np\left(1 + p(N - 1) + K_1 \mathrm{e}^{-s}\right)}{(1 + K_1 \mathrm{e}^{-s})^2} \; , \tag{3.20}
$$

$$
\mathsf{E}(C^4(s)) = \frac{Np\left(1 - \frac{p\,\mathrm{e}^s}{K_1 + \mathrm{e}^s}\right)^N {}_4\mathrm{F}_3\left(2, 2, 2, 1 - N; 1, 1, 1; \frac{p\,\mathrm{e}^s}{\mathrm{e}^s(p-1) - K_1}\right)}{K_1 \mathrm{e}^{-s} - (p - 1)} \; . \tag{3.21}
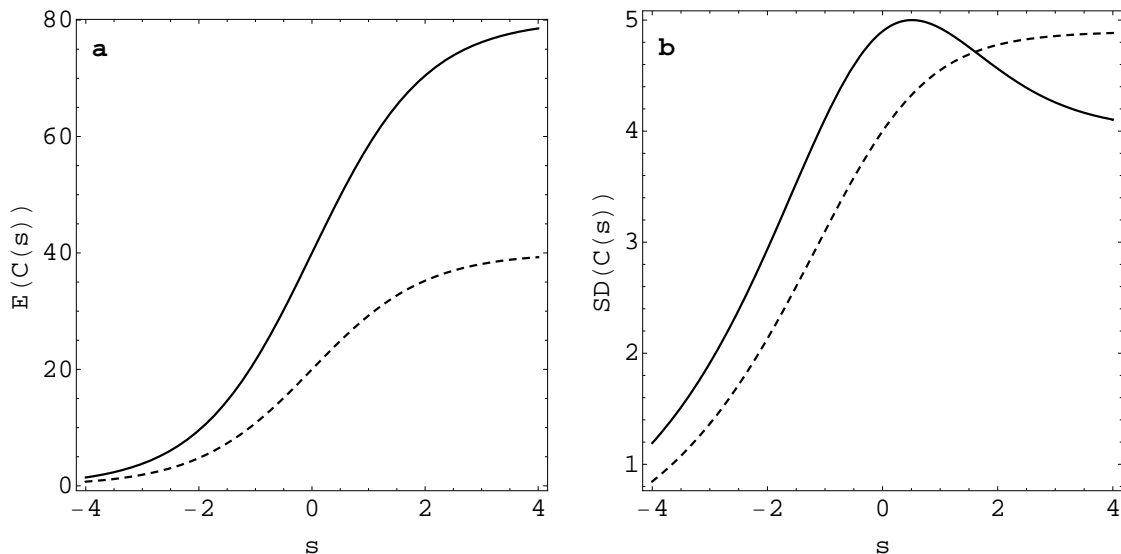$$

Figure 3.3:   (a) Mean $\mathsf{E}(C(s))$ and (b) standard deviation $\mathsf{SD}(C(s))$ of the number of activated receptors in model of simple activation, as functions of the odorant log-concentration, s, in perireceptor space. Parameters are $K_1 = 1$ and $N = 100$. Activation probability p is set to 0.8 (solid curves) and 0.4 (dashed curves). Inflexion points of the mean curves are located at $s = \ln K_1 = 0$ (independently on the activation probability p). The variance is either monotonically increasing (for $p = 0.4$) or has a maximum located at $s = \ln K_1 - \ln(2p - 1) \approx 0.511$ (for $p = 0.8$).

steady-state mean (3.18) monotonically increases from 0 to Np and has sigmoidal shape. Its inflexion point is located at $s = \ln K_1$, independently of the value of activation probability p. For $p \in (0, 0.5]$, variance (3.19) is monotonically increasing from zero to the limit value $\mathsf{Var}(C(\infty)) = Np(1 - p)$. For $p \in (0.5, 1)$, it increases from zero to maximal value $N/4$ achieved at $s = \ln K_1 - \ln(2p - 1)$ and then decreases to the limit value $\mathsf{Var}(C(\infty)) = Np(1 - p)$. Both the mean and the standard deviation as functions of the log-concentration of odorant are plotted in Figure 3.3. Of course, model (3.1) is a limit case of model (3.15) for p converging to 1. The continuous approximation of $C(s)$ can be derived analogously as in Section 3.2.1.

3.2.3   Double-step model

This model has often been used for describing odorant-receptor interaction (see Kaissling [29], Gurewich et al. [23], Maurin [51], Rospars et al. [59, 60]). As in the previous model, the receptors may appear in three different states. The interaction between unbound, R, bound not activated, C*, and bound activated

receptors, C, is formed by the transitions via the stochastic reaction schema

$$
\mathrm{A} + \mathrm{R} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{C}^* \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{C} \ ,
\tag{3.22}
$$

where rate coefficients k_2 and k_{-2} characterize the activation-deactivation process. In contrast with the model of interaction with simple activation (3.15), in the double-step interaction it is assumed that occupied receptor can become activated only with a delay after the binding. Analogously as in the basic model, the stochastic process $\{C_t; \ t \geq 0\}$ giving the count of activated receptors at time t can be described as a homogenous Markov chain with $(N+1)(N+2)/2$ states $\{(i,j); 0 \leq i+j \leq N\}$ and transition probabilities (for time interval Δt)

$$
\begin{aligned}
\mathsf{P}\Big\{(i,j) \xrightarrow{\Delta t} (i+1,j)\Big\} &= k_1(N-i-j)\mathrm{e}^s\Delta t + \mathrm{o}(\Delta t) \ , \\
\mathsf{P}\Big\{(i,j) \xrightarrow{\Delta t} (i-1,j)\Big\} &= k_{-1}i\Delta t && + \mathrm{o}(\Delta t) \ , \\
\mathsf{P}\Big\{(i,j) \xrightarrow{\Delta t} (i-1,j+1)\Big\} &= k_2 i\Delta t && + \mathrm{o}(\Delta t) \ , \\
\mathsf{P}\Big\{(i,j) \xrightarrow{\Delta t} (i+1,j-1)\Big\} &= k_{-2}j\Delta t && + \mathrm{o}(\Delta t) \ ,
\end{aligned}
\tag{3.23}
$$

where the first coordinate denotes the count of bound not activated receptors and the second one denotes the count of activated receptors.

The stationary distribution of $C(s)$ is multinomial; see [20] for general formulae. The steady-state mean number of activated receptors, $C(s)$, is

$$
\mathsf{E}(C(s)) = \frac{N}{1 + K_2\left(1 + K_1\mathrm{e}^{-s}\right)} \ ,
\tag{3.24}
$$

where $K_2 = k_{-2}/k_2$. We have not been able to evaluate the steady-state variance analytically. Nevertheless, it can be computed numerically and as its good approximation we found the function

$$
\mathsf{Var}(C(s)) \approx \frac{a + b\mathrm{e}^{-s}}{1 + c\mathrm{e}^{-s} + d\mathrm{e}^{-2s}}
\tag{3.25}
$$

with general parameters a, b, c, d. The steady-state mean given by equation (3.24) monotonically increases from zero to $N/(1+K_2)$ and has (in general) asymmetric sigmoidal shape with inflexion point located at $s = \ln K_1 + \ln K_2 - \ln(1+K_2)$. The steady-state variance fulfills relations $\mathsf{Var}(C(-\infty)) = 0$ and $\mathsf{Var}(C(\infty)) = a > 0$. Both the mean and the standard deviation as functions of the log-concentration of odorant are depicted in Figure 3.4. The continuous approximation of $C(s)$ can be derived analogously as in Section 3.2.1.
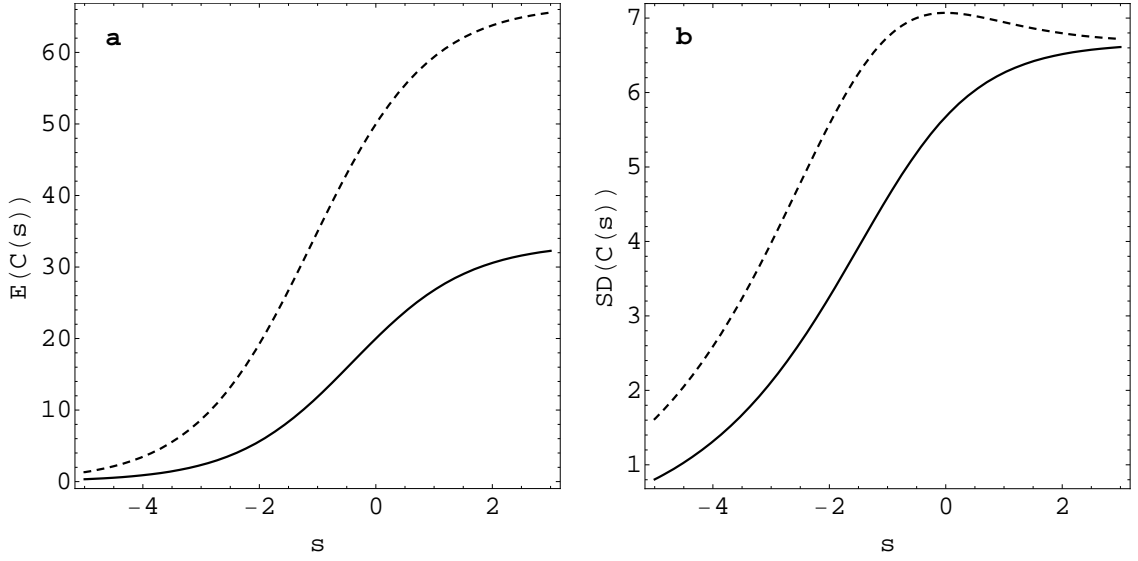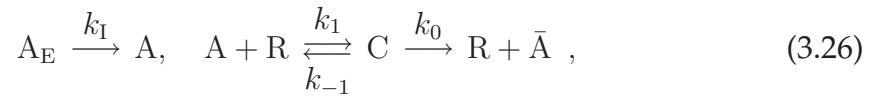
Figure 3.4: (a) Mean $\mathsf{E}(C(s))$ and (b) standard deviation $\mathsf{SD}(C(s))$ of the number of activated receptors in the double-step model, as functions of the odorant log-concentration, s, in perireceptor space. Parameters are $N = 100$; $K_1 = 1$ and $K_2 = 2$ (solid curves), $K_1 = 1$ and $K_2 = 0.5$ (dashed curves). Inflexion points of the mean curves are located at $s \approx -0.405$ (solid) and $s \approx -1.099$ (dashed). The standard deviation is either monotonically increasing (solid) or has a local maximum (dashed).

3.2.4 Flux detector

In contrast with the concentration detector, in the flux detector model, the rate of influx of odorant from the external to the perireceptor space is quantitatively taken into account. This scenario has been introduced by Kaissling [30] and further analyzed by Lánský and Rospars [41], Rospars et al. [60]. The transfer of odorant molecules between the external and perireceptor spaces is relatively slow and irreversible (no outflux is permitted). The concentration of the odorant in perireceptor space can be substantially higher than outside. Stochastic schema of the single-step reaction is

$$\mathrm{A_E} \xrightarrow{k_\mathrm{I}} \mathrm{A}, \quad \mathrm{A + R} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{C} \xrightarrow{k_0} \mathrm{R + \bar{A}} \ , \tag{3.26}$$

where $\mathrm{A_E}$ denotes the odorant of fixed concentration $\mathrm{A_E} = \mathrm{e}^s$ in the external space, A odorant in the perireceptor space and $\bar{\mathrm{A}}$ denotes degraded form of the odorant (cannot interact with receptors anymore). The parameter k_I is the rate of influx of the odorant molecules from the external to the perireceptor space and k_0 is the rate of degradation of molecules of the odorant. Because of irreversible reaction, the Michaelis-Menten constant $K_\mathrm{M} = (k_{-1} + k_0)/k_1$ should be used for

description of reaction rate instead of the dissociation constant K_1. Further, let us denote the ratio of the influx and the degradation rates by $K = k_I/k_0$.

Continuous-time stochastic process $\{C_t;\ t \geq 0\}$ giving the count of activated receptors at time t can be described by inhomogeneous birth and death process with transition rates

$$\lambda_{i,t} = k_1(N - i)A_t,\ \mu_i = \mu i = (k_{-1} + k_0)i;\ i \in \{0, 1, \ldots, N\}\ , \tag{3.27}$$

where A_t is the odorant concentration in the perireceptor space at time t. The birth rates $\lambda_{i,t}$ are time-dependent (inhomogeneous) because of the stable influx of the odorant from the external to the perireceptor space. The process has a stationary state $C(s)$ if the relation $s \leq \ln N - \ln K$ is fulfilled. In this case the concentration of the odorant in the perireceptor space reaches an equilibrium $A_\infty(s) < \infty$. For $s > \ln N - \ln K$ the process C_t converges to the value $C(s)$ whereas the odorant concentration A_t in the perireceptor space grows without bounds.

Using the law of mass action, the deterministic behaviour of the flux detector model (see schema (3.26)) can be described by two independent differential equations

$$\frac{dC(t)}{dt} = k_1 N A(t) - (k_{-1} + k_0 + k_1 A(t))\, C(t)\ , \tag{3.28}$$

$$\frac{dA(t)}{dt} = k_I e^s - k_1 N A(t) + (k_{-1} + k_1 A(t))\, C(t)\ , \tag{3.29}$$

where $A(t)$ denotes concentration of the odorant in the perireceptor space and $C(t)$ the number of activated receptors at time $t \geq 0$. Computing steady-state solution of equations (3.28) and (3.29), formally letting

$$\frac{dC(t)}{dt} = 0 \quad \text{and} \quad \frac{dA(t)}{dt} = 0\ , \tag{3.30}$$

under the condition $s \leq \ln N - \ln K$ we obtain the steady-state odorant concentration in the perireceptor space,

$$A_\infty(s) = \frac{K_M K e^s}{(N - K e^s)}\ . \tag{3.31}$$

Now, we can replace the inhomogeneous birth and death process of the reaction (3.26) with the homogenous one according to reaction (3.1). That means, we consider the basic model of concentration detector with fixed concentration of odorant in the perireceptor space, $A_\infty(s)$. Both processes yield the same stationary probability distribution because of use of birth and death processes which are markovian (they do not depend on the initial conditions). Corresponding time-independent birth and death rates are

$$\lambda_i = k_1(N - i)A_\infty(s), \quad \mu_i = k_{-1}i; \quad i \in \{0, 1, \ldots, N\}\ . \tag{3.32}$$

Analogously as in Section 3.2.1 we derive the stationary probability distribution $\{\pi_i\}$ of the birth and death process $\{C_t; \ t \geq 0\}$ with rates (3.32). Then, the mean and the variance of the count of activated receptors in steady state, $C(s)$, are

$$\mathsf{E}(C(s)) = \frac{k_1 N A_\infty(s)}{k_{-1} + k_1 A_\infty(s)} \ , \tag{3.33}$$

$$\mathsf{Var}(C(s)) = \frac{k_1^2 N A_\infty(s)}{(k_{-1} + k_1 A_\infty(s))^2} \ . \tag{3.34}$$

Assuming $s \leq \ln N - \ln K$ and substituting (3.31) into equations (3.33) and (3.34) the moments of of the count of activated receptors can be computed,

$$\mathsf{E}(C(s)) = K\mathrm{e}^s \ , \tag{3.35}$$

$$\mathsf{Var}(C(s)) = K\mathrm{e}^s \left(1 - \frac{K\mathrm{e}^s}{N} \right) \ , \tag{3.36}$$

$$\mathsf{E}(C^2(s)) = K\mathrm{e}^s \left(1 + \frac{(N-1)K\mathrm{e}^s}{N} \right) \ , \tag{3.37}$$

$$\mathsf{E}(C^4(s)) = K\mathrm{e}^s \left(1 - \frac{K\mathrm{e}^s}{N} \right)^{N-1} {}_4\mathrm{F}_3 \left(2, 2, 2, 1{-}N; 1, 1, 1; \frac{K\mathrm{e}^s}{K\mathrm{e}^s - N} \right) \ . \tag{3.38}$$

Note that the Michaelis-Menten constant K_{M} does not play any role in the behaviour of $C(s)$.

The steady-state mean given by equation (3.35) is increasing function of the odorant concentration. The variance (as a function of the odorant concentration) given by equation (3.36) has unimodal asymmetric shape with maximum value $N/4$ achieved for $s = \ln N - \ln K - \ln 2$. For low as well as possible high odorant log-concentrations it becomes practically zero, because

$$\lim_{s \to -\infty} \mathsf{Var}(C(s)) = \mathsf{Var}(C(\ln N - \ln K)) = 0 \ . \tag{3.39}$$

The mean and standard deviation of $C(s)$ are plotted in Figure 3.5. The continuous approximation of $C(s)$ can be derived analogously as in Section 3.2.1. It is normally distributed and has the probability density function

$$g(x; s) = \sqrt{\frac{N\mathrm{e}^{-s}}{2\pi K(N - K\mathrm{e}^s)}} \exp \left\{ \frac{-N(K\mathrm{e}^s - x)^2}{2K\mathrm{e}^s(N - K\mathrm{e}^s)} \right\} \ . \tag{3.40}$$

3.3 Methods

We aim to investigate how precisely the odorant concentration, s, can be determined from a knowledge of the response, $C(s)$, and which concentration levels
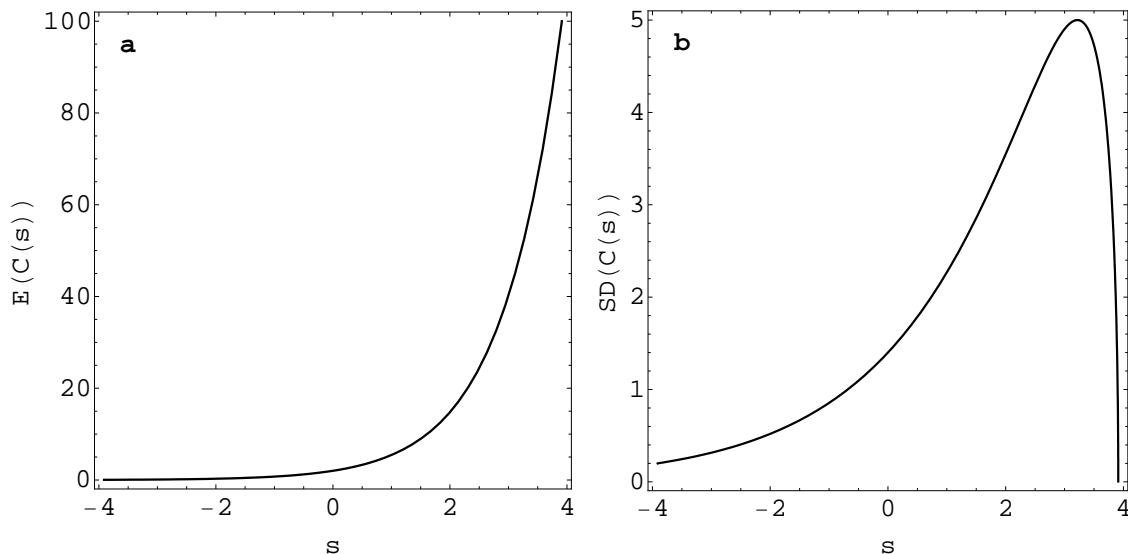
Figure 3.5: (a) Mean $\mathsf{E}(C(s))$ and (b) standard deviation $\mathsf{SD}(C(s))$ of the number of activated receptors in the flux detector model, as functions of the odorant log-concentration, s, in perireceptor space. Parameters are $K = 1$ and $N = 100$.

are optimal, that means can be well determined from the knowledge of a random sample of $C(s)$. In other words, we consider an experiment in which a fixed concentration is applied and steady-state responses of the system are observed. These are independent (it is the random sample) realizations of random variable $C(s)$ from which we wish to determine s.

There are two properties limiting the optimal signal determination. First, the minimal resolution ability. The system cannot distinguish two response values that are near one to another and the corresponding signal values are determined as equal. Moreover, there are two bounds, minimal and maximal response, between them the system can code the information. Second, the fluctuation of realizations of the response. On the same signal level, observed responses are not equal.

The classical, deterministic, approach to determine the optimal concentration is based on shape of the input-output function, $f(s)$. This approach is commonly used analyzing of various signal-response models using differential equations or fitted functions, see for example Cleland and Linster [17], Lánský and Getz [42], Nizami [57], Rospars et al. [60]. As an optimality measure the first derivative of the input-output function with respect to the concentration of odorant is used, $J_1(s) = \partial f(s)/\partial s$, which measures the slope of the function $f(s)$. Because in models proposed here the relation $\mathsf{E}(C(s)) = f(s)$ is fulfilled, we can define J_1

criterion as follows,

$$J_1(s) = \frac{\partial \mathsf{E}(C(s))}{\partial s} \quad . \tag{3.41}$$

The optimal concentration of odorant, s_1, is that for which the measure J_1 is maximal,

$$J_1(s_1) = \max_s \{ J_1(s) \} \quad . \tag{3.42}$$

An alternative approach used here, is based on statistical properties of $C(s)$. This approach in search for optimal signal was used in various applications of signal-response dependencies, for example by Brunel and Nadal [14], Johnson and Ray [28], Lánský et al. [45, 43], Sanchez-Montanes and Pearce [64], Wilke and Eurich [68]. Models described in the previous section suppose that the observed number of activated receptors in steady state, $C(s)$, is a continuous random variable with a distribution belonging to the regular parametric family of probability density functions, $C(s) \sim g(x; s)$, with the odorant concentration, s, as a parameter.

Determination of the concentration, s, from sampling responses $\{x_1, \ldots, x_n\}$ of $C(s)$ corresponds to its estimation, $\hat{s}$, in chosen family of probability density functions. As a measure of optimality, the Fisher information

$$J(s) = \mathsf{E}\left(\left(\frac{\partial \ln g(s)}{\partial s} \right)^2 \right) = \int \frac{1}{g(x; s)} \left(\frac{\partial g(x; s)}{\partial s} \right)^2 \mathrm{d}x \tag{3.43}$$

with respect to the odorant log-concentration, s, is used for sample of size equal to one. See Definition 3 on page 13 for detailed description. For the sample of size n independent measurements, the Fisher information about the parameter s is equal to $nJ(s)$. Since this multiplication cannot change the behaviour of the Fisher information as a function of s, we study characteristics of $J(s)$ only.

It should be mentioned that optimality measures based on Fisher information assume that the responses are recorded from different sensory neurons or separated by a sufficient time period to ensure independence of the measurements. Otherwise, the Fisher information would not be scaled with the sample size due to autocorrelation in time-averaging measurements and the optimality criteria should be modified.

The Cramér-Rao inequality (2.8) gives relation between the Fisher information $J(s)$ and quality of the estimator $\hat{s}$,

$$\mathsf{Var}(\hat{s}) \geq J(s)^{-1} \quad , \tag{3.44}$$

hence, the Fisher information is the inverse asymptotic variance of the best unbiased estimator of s. Therefore, the higher the Fisher information $J(s)$ is the better

estimation of s can be achieved and we define the optimal odorant concentration, s_0, as the concentration for which the measure J is maximal,

$$J(s_0) = \max_s \left\{ J(s) \right\} \quad . \tag{3.45}$$

In some cases the analytical expression of the Fisher information is very complicated. Then we can use the lower bound, $J_2(s) \le J(s)$, of the Fisher information (see Definition 7 on page 16 and [43]),

$$J_2(s) = \frac{1}{\mathsf{Var}(C(s))} \left(\frac{\partial \mathsf{E}(C(s))}{\partial s} \right)^2 = \frac{J_1(s)^2}{\mathsf{Var}(C(s))} \quad , \tag{3.46}$$

which only requires the knowledge of first two moments of the distribution. Then we can define the optimal odorant concentration, s_2, in terms of $J_2(s)$ instead of $J(s)$ analogously. Moreover, we have equality $J_2(s) = J(s)$ if the family of probability density functions $g(x; s)$ fulfills the equation

$$\frac{1}{g(x;s)} \frac{\partial g(x;s)}{\partial s} = a_s \left(x - \mathsf{E}(C(s)) \right) \tag{3.47}$$

for a_s independent on x (see Theorem 11 on page 19 and [43]).

Similarly, recalling Definition 7 on page 16, lower bound $J_3(s)$ of the Fisher information can also be used as an optimality criterion,

$$J_3(s) = \frac{1}{\mathsf{Var}(C^2(s))} \left(\frac{\partial \mathsf{E}(C^2(s))}{\partial s} \right)^2 \quad . \tag{3.48}$$

3.4 Results

In this section we compute the optimality criteria $J_1(s)$, $J_2(s)$, $J_3(s)$ and $J(s)$ for the models of odorant-receptor interaction mentioned above and compare the results in searching for optimal concentration of odorant. In all this section $C(s)$ denotes the continuous random variable which gives the number of activated receptors in steady state as a function of the odorant concentration, s. This is justified by using the diffusion approximation (3.12) and its analogous variants. In accordance with the diffusion approximation, $C(s)$ is assumed to have the normal probability distribution with moments specified by the particular models,

$$C(s) \sim \mathsf{N}\Big(\mathsf{E}(C(s)), \mathsf{Var}(C(s)) \Big) \quad . \tag{3.49}$$

3.4.1 Basic model

In the basic model based on reaction (3.1) the number of activated receptors in steady state, $C(s)$, has moments specified by equations (3.6) and (3.7) with respect to the odorant log-concentration, s, in the perireceptor space. Assuming the normal distribution of $C(s)$, its probability density function is equal to (3.14). All three criteria of optimality are directly derived,

$$J_2(s) = J_1(s) = \frac{NK_1 \mathrm{e}^{-s}}{(1 + K_1 \mathrm{e}^{-s})^2} \ , \tag{3.50}$$

$$J(s) = \frac{1}{2} + \frac{(N-2)K_1 \mathrm{e}^{-s}}{(1 + K_1 \mathrm{e}^{-s})^2} = \frac{1}{2} + \frac{N-2}{N} J_2(s) \ , \tag{3.51}$$

the analytical expression of $J_3(s)$ can be easily expressed from the knowledge of (3.8) and (3.9).

The shapes of optimality criteria are plotted in Figure 3.6. The criteria $J_1(s)$ and $J_2(s)$ are equal and have unimodal shape. For $N > 2$ (which is natural in reality), the Fisher information $J(s)$ is also unimodal and it is very close to $J_1(s)$. All these criteria attain maximum value $N/4$ for the odorant log-concentration

$$s_0 = s_1 = s_2 = \ln K_1 \ . \tag{3.52}$$

The criterion $J_3(s)$ has also unimodal shape, but its maximum is slightly shifted from s_1 to higher signal intensities. This shift, however, is very small and depends only on N (the shift rises with increasing N). For extremely low as well as high odorant concentrations all the criteria decrease, $J_1(\pm\infty) = J_2(\pm\infty) = 0$, $J(\pm\infty) = 1/2$. Both deterministic and statistical approach give the same result and locate the optimal concentration of odorant in the region around the concentration s_1 (see Figure 3.6).

3.4.2 Model of simple activation

In the model of simple activation, where $C(s)$ is normally distributed with moments given by equations (3.18) and (3.19), the criteria J_1 and J_2 are derived analytically,

$$J_1(s) = \frac{pNK_1 \mathrm{e}^{-s}}{(1 + K_1 \mathrm{e}^s)^2} \ , \tag{3.53}$$

$$J_2(s) = \frac{pNK_1^2 \mathrm{e}^{-s}}{(1 + K_1 \mathrm{e}^{-s})^2 (K_1 + (1-p)\mathrm{e}^s)} \ , \tag{3.54}$$

$J_3(s)$ and the Fisher information $J(s)$ are evaluated numerically.

As well as in basic model (3.1), maximum value of the criterion $J_1(s)$ is located at odorant log-concentration $s_1 = \ln K_1$, independently on the activation
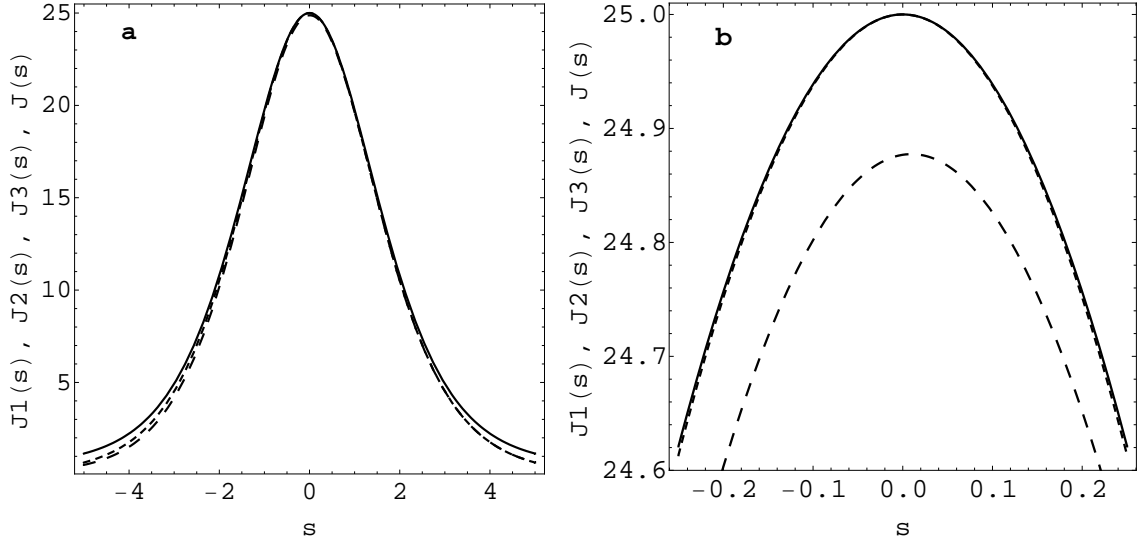
Figure 3.6: (a) Optimality criteria in the basic model: the criteria $J_1(s) = J_2(s)$ (dotted curve), $J_3(s)$ (dashed curve) and the Fisher information $J(s)$ (solid curve). Parameters are $K_1 = 1$ and $N = 100$. Criteria J, J_1, J_s attain maximum value $N/4 = 25$ for the odorant log-concentration $s = \ln K_1 = 0$. (b) Detail of (a); note, that the maximum of J_3 criterion is shifted.

probability p. The criterion $J_2(s)$ achieves its maximum for the odorant log-concentration

$$s_2 = \ln K_1 - \ln \frac{4(1-p)}{\sqrt{9 - 8p} - 1} \quad . \tag{3.55}$$

For $p \in (0, 1)$, the relation

$$s_1 - \ln 2 < s_2 < s_1 \tag{3.56}$$

holds. For lower activation probabilities p the location of maximum of $J_2(s)$ is shifted to lower concentrations of odorant. As shown in Figure 3.7 and Figure 3.8, the criteria $J_3(s)$ and $J(s)$ behaves similarly as $J_2(s)$ criterion. Note, that $s_0 < s_2 < s_1$, $s_0 < s_3 < s_1$ holds regardless of the value of p. The deterministic and statistical approaches can give different results, the optimum from statistical point of view is located at lower concentrations of odorant than that obtained with the approach based on the slope of the input-output function. Limiting behaviour of the optimality criteria is similar to the previous model (3.1), except for $J(\infty) = 0$.

3.4.3 Double-step model

Let us assume that the binding and activation process has two steps. It means that the number of activated receptors has normal distribution with mean given by
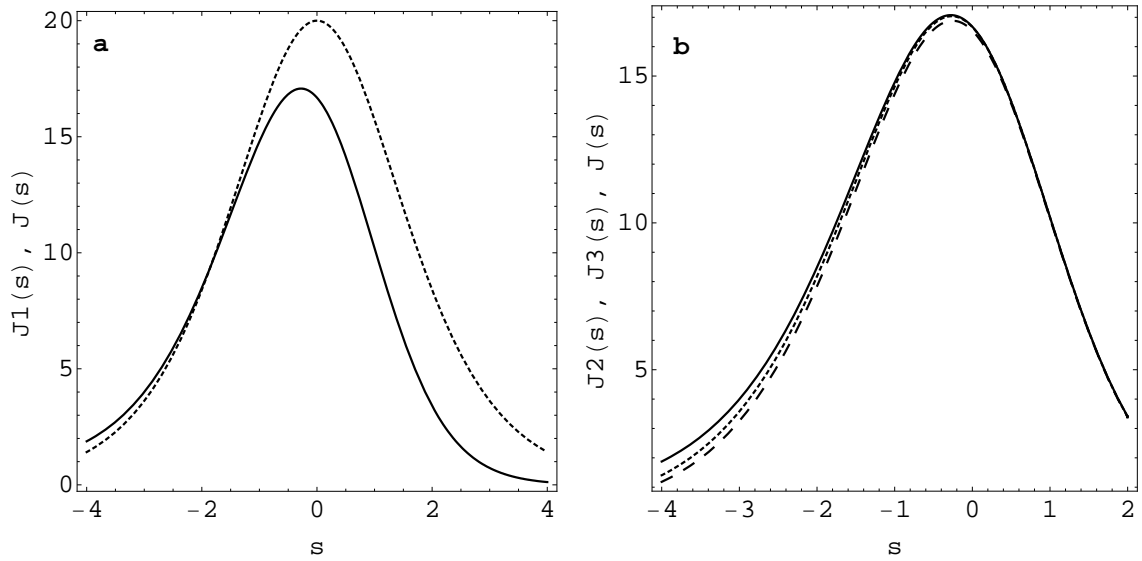
Figure 3.7: Optimality criteria in the model with simple activation: (a) first derivative of the input-output function $J_1(s)$ (dotted curve) and Fisher information $J(s)$ (solid), (b) criteria $J(s)$ (solid curve), $J_2(s)$ (dotted) and $J_3(s)$ as functions of the odorant log-concentration, s, in the perireceptor space. Maximum of $J_1(s)$ is located at $s_1 = 0$. Maximum of $J_2(s)$ is located at $s_2 \approx -0.267$ and maximum of $J(s)$ at $s \approx -0.277$. Parameters are $K_1 = 1$, $N = 100$ and $p = 0.4$.
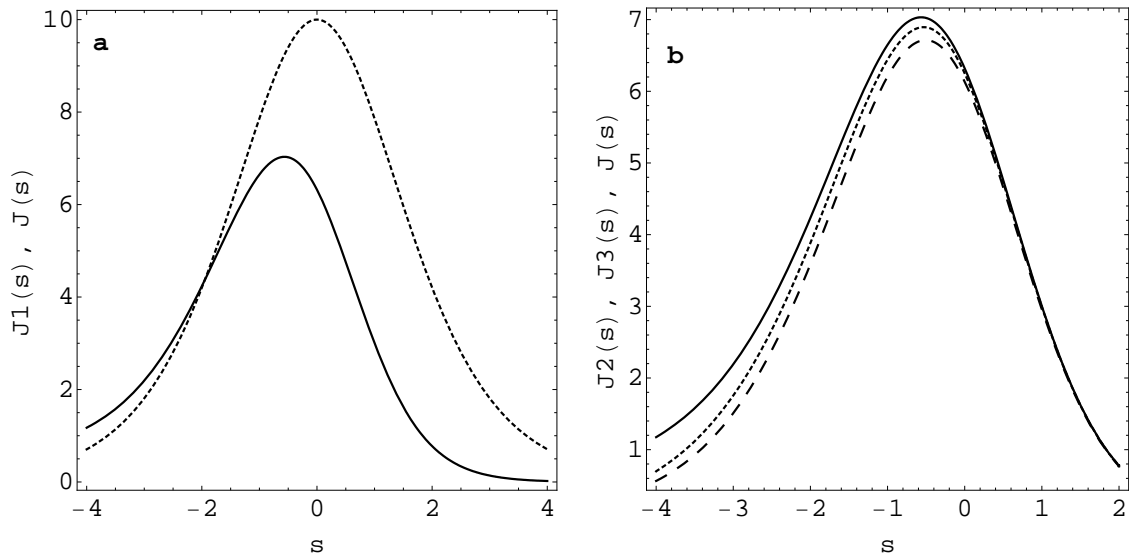
Figure 3.8: Optimality criteria in the model with simple activation: (a) first derivative of the input-output function $J_1(s)$ (dotted curve) and Fisher information $J(s)$ (solid), (b) criteria $J(s)$ (solid curve), $J_2(s)$ (dotted) and $J_3(s)$ as functions of the odorant log-concentration, s, in the perireceptor space. Maximum of $J_1(s)$ is located at $s_1 = 0$. Maximum of $J_2(s)$ is located at $s_2 \approx -0.533$ and maximum of $J(s)$ at $s \approx -0.565$. Parameters are $K_1 = 1$, $N = 100$ and $p = 0.8$.
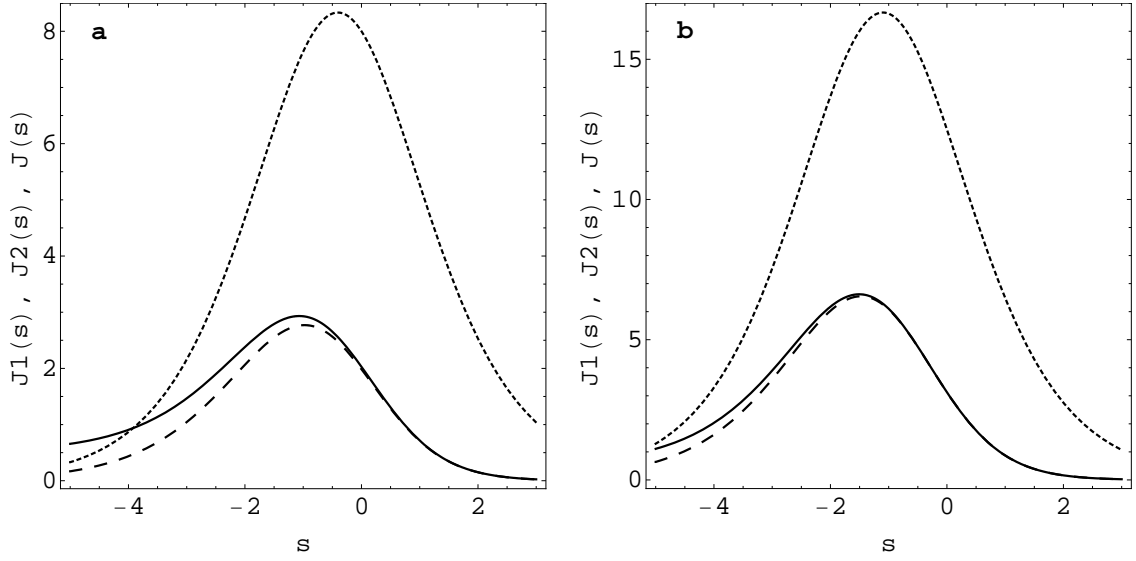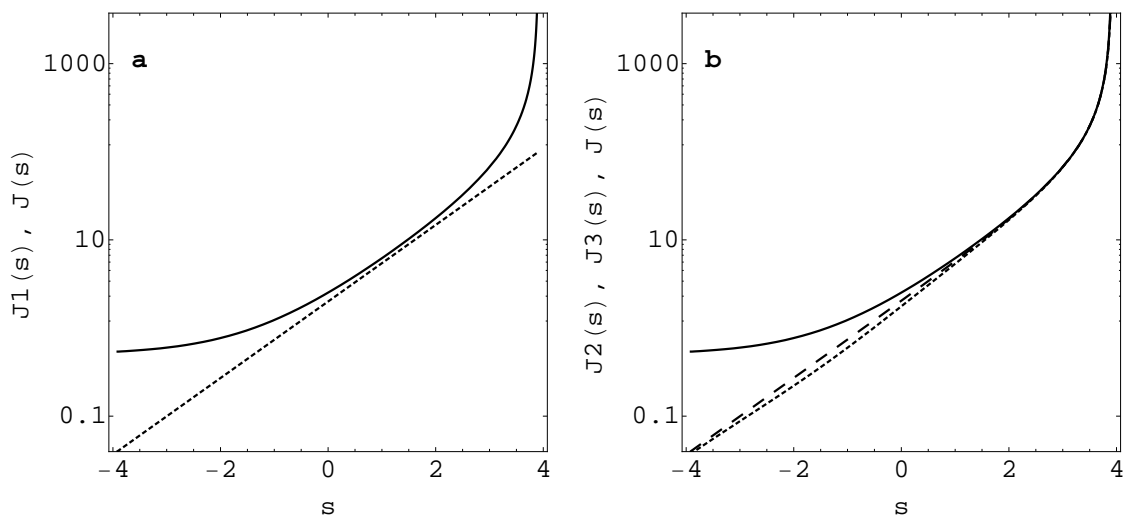
Figure 3.9: Optimality criteria in the double-step model: first derivative of the input-output function $J_1(s)$ (dotted curve), Fisher information $J(s)$ (solid) and its lower bound $J_2(s)$ (dashed). (a) Parameters are $N = 100$, $K_1 = 1$ and $K_2 = 2$. Maximum of $J_1(s)$ is located at $s_1 \approx -0.405$, maximum of $J_s(s)$ at $s \approx -0.977$ and maximum of $J(s)$ at $s \approx -1.068$. (b) Parameters are $N = 100$, $K_1 = 1$ and $K_2 = 0.5$. Maximum of $J_2(s)$ is located at $s_1 \approx -1.099$, maximum of $J_s(s)$ at $s \approx -1.475$ and maximum of $J(s)$ at $s \approx -1.506$. Note, that $s_0 < s_2 < s_1$ holds.

equation (3.24) and known variance, for example of the form (3.25). The criterion J_1 is equal to

$$J_1(s) = \frac{N K_1 K_2 \mathrm{e}^{-s}}{\left(1 + K_2 \left(1 + K_1 \mathrm{e}^{-s}\right)\right)^2} \tag{3.57}$$

and its maximum is located at

$$s_1 = \ln K_1 + \ln K_2 - \ln(1 + K_2) \ . \tag{3.58}$$

The criteria $J_2(s)$ and $J(s)$ are computed numerically. Their maximum values appear at lower levels of odorant log-concentrations than s_1, as depicted in Figure 3.9. Deterministic and statistical approaches give different results, in general. Limit values of the optimality criteria are equal as in the model with simple activation (3.15).

3.4.4 Flux detector

In the flux detector model (schema 3.26), the number of activated receptors in steady state, $C(s)$, has moments specified by equations (3.35) and (3.36) with respect to the odorant log-concentration, s, in the external space. Assuming the nor-

mal distribution of $C(s)$, its probability density function is equal to (3.40). The optimality criteria J_1 and J_2 are derived analytically,

$$J_1(s) = \mathsf{E}(C(s)) = K\mathrm{e}^s \ , \tag{3.59}$$

$$J_2(s) = \frac{NK\mathrm{e}^s}{N - K\mathrm{e}^s} \ , \tag{3.60}$$

$J_3(s)$ and the Fisher information $J(s)$ are evaluated numerically.

All the criteria are monotonically increasing and satisfy the relation

$$J_1(s) \leq J_2(s) \leq J(s) \quad \text{for } s \leq \ln N - \ln K \ . \tag{3.61}$$

The $J_1(s)$ criterion is equal to the steady-state mean, $\mathsf{E}(C(s))$, and grows as the odorant log-concentration, s, increases. For low odorant concentrations s, the criteria J_1 and J_2 became equal. For maximum possible odorant log-concentrations $s \to \ln N - \ln K$ the criteria J_2 and J become equal and grow substantially more faster than the criterion J_1. Limit values of the optimality criteria are equal as in the model with simple activation (schema 3.15). Nevertheless, both deterministic and stochastic approach in search for the optimal signal give the same result and locate the optimal odorant log-concentration at highest possible value

$$s_1 = s_2 = s_0 = \ln N - \ln K \ . \tag{3.62}$$

The shapes of optimality criteria $J(s)$, $J_1(s)$, $J_2(s)$ and $J_3(s)$ as functions of odorant log-concentrations s are depicted in Figure 3.10, where, for transparence, the vertical axis has logarithmic scale.

Figure 3.10: Optimality criteria in the flux detector model: (a) first derivative of the input-output function $J_1(s)$ (dotted curve) and Fisher information $J(s)$ (solid), (b) criteria $J(s)$ (solid curve), $J_2(s)$ (dotted) and $J_3(s)$ as functions of the odorant log-concentration, s, in the external space. Maximum (limiting) values of all three criteria are located at $s_1 = \ln N - \ln K$. The vertical axis has logarithmic scaling. Parameters are $K = 1$ and $N = 100$.

Classification of stimuli

<div align="right">

4

</div>

This chapter contains mainly extended text of article [4]. Some earlier ideas can be found in article [3], too. A theoretical model of olfactory neuron with different types of noise, an biophysical and an empirical model are analysed both from point of view of statistical inference and from point of view of information theory.

4.1   Introduction

To determine how neural responses vary with different stimuli is a key question. Characterization of the input-output properties of sensory neurons is commonly done by using the so called input-output transfer function, $R(s)$, in which the output (e.g. firing frequency) is plotted against the input s (stimulus intensity; the vocabulary may depend on the sensory modality considered). Although the output is usually the spiking frequency, it can also be the activity or level of any intermediate variable, e.g. effector enzyme concentration, ionic channel conductance or receptor potential. The response curves are usually monotonously increasing functions, most often of sigmoid shape, assigning a unique response to each input signal, as shown in different sensory systems ([16, 52, 57, 61] and others). The response functions are usually presented as a single curve, sometimes accompanied by standard deviations (see [12, Fig. 1] and [66, Fig. 2]), which indicates that the relation between stimulus and response is influenced by random factors. The complete distributions of response variability are not, as far as we know, given in experimental reports, despite some attempts at understanding and describing the response variability ([16]). Taking into account the variability of responses, the published curves should be called more appropriately mean response functions ([56]). As shown in this chapter, such a variability of responses may substantially influence the determination of the stimulus intensities considered as *optimal*.

The intuitive concept of *just noticeable difference* has been deeply studied in psychophysics ([22]). It is indirectly involved in the determination of the *optimal* signal as derived from the transfer function $R(s)$. If there is a minimum detectable increment ε of the response, the just noticeable difference in the signal can be cal-

culated for every stimulus intensity. Obviously, this just noticeable difference in the signal is smallest, i.e. the system is most sensitive, where the transfer function is the steepest. Therefore the optimal stimulus intensity, that is the best detected intensity, is where the slope of the transfer function is the highest. This measure of signal optimality is based on the deterministic (mean) transformation of a signal into a response. In other words it applies to situations in which to each signal corresponds a deterministically unique response.

Variability is considered as a consequence of noise acting upon the transfer function. The presence of noise complicates the concept of signal optimality based on the *just noticeable difference*. To include the effects of noise, alternative measures of signal optimality are required. The aim of this paper is to summarize and investigate these measures that can be applied to noisy transfer functions in sensory neurons. The measures have to take into account two factors – the shape of the mean transfer function and the amplitude or complete distribution of the responses. Two points of view of optimality are employed and both are related to the role of the sensory neurons – to transfer maximum information and to do it reliably. The first measure is based on the question "which signal intensity brings the highest information". To answer this question, an additional knowledge, the distribution of stimulus, is required. This approach stems from information theory. The second approach is based on a different question "which signal intensity can be most precisely deduced from a response" and it is based on the theory of statistical inference.

4.2 Response function and optimality criteria

4.2.1 Response function and its variability

To illustrate the main concepts involved we consider a generic sigmoid transfer function $m(s) = (1 + e^{-s})^{-1}$, to which a suitable fluctuation is added, taking the form

$$R(s) = m(s) + \xi(s) \ , \tag{4.1}$$

where $m(s)$ is the firing frequency at stimulus intensity s and $\xi(s)$ is the random component of the firing frequency with mean equal to zero, $\mathsf{E}(\xi(s)) = 0$. In other words, for stimulation at level s, the response is a random variable $R(s)$ given by right-hand side of equation (4.1). As pointed out in Introduction (see p. 44), for the present treatment only the shape of the curve is important, not its scale and position. Thus, $m(s)$ contains no free parameters. Function $m(s)$ above is the so-called Hill function with exponent $n = 1$ (noncooperative reactions), in logarithmic scale, so the effect of the Hill's coefficient n, which is to modify the slope of

the sigmoid curve, is left out. Depending on the distribution of $\xi(s)$, model (4.1) can result in negative values of the response, which is physically not acceptable. Actually, the transfer function (4.1) is already normalized as its interval of variation is on $[0, 1]$, which means that we assume there is no spontaneous activity (at $s = -\infty$, i.e. in absence of signal) and no response greater than one (asymptotic maximum). In practice, a simple shift (adding a constant to $R(s)$) solves the problem of negative responses and nothing is changed in the following treatment.

We introduce three simple examples of noise distributions to show that the choice of type of the distribution can substantially change the results in searching for optimal signal. The first two examples are symmetrical and differ by the speed of decay of the noise amplitude. The last one illustrates the effect of noise asymmetry on the results. One of the simplest and most natural way to introduce the noise in the transfer function is to assume that it has Gaussian distribution, then

$$R(s) \sim \mathsf{N}\left(m(s), \sigma^2(s)\right) \ , \tag{4.2}$$

where variance $\sigma^2(s)$ also depends on the signal s. This dependency of the variance of the response on s substantially influences the signal identification as will be seen later. We investigate two different examples of model (4.2). In the third example another type of noise is introduced to see the effect of asymmetry in the noise amplitude. In the first example we consider equation (4.2) specified by

$$\sigma^2(s) = \sigma_0^2(s)\, \mathrm{e}^{-c\,s^2} \ , \tag{4.3}$$

where $c \geq 0$. If $c > 0$, the variance tends to zero at the endpoints of the range of the signal and parameter c controls how fast this tendency is. If $c = 0$, the variance is a constant independent of signal intensity.

In the second example the variance of the response function depends on s in the following way,

$$\sigma^2(s) = \frac{4\,\sigma_0^2\, \mathrm{e}^{c\,s}}{(1 + \mathrm{e}^{c\,s})^2} \ , \tag{4.4}$$

where $c \geq 0$. The difference between (4.3) and (4.4) is that the variance given by (4.4) decreases more slowly with increasing absolute value of s. Both models are illustrated in Figure 4.1.

In both examples (4.3) and (4.4) the Gaussian character of the responses is preserved which does not need to be the case especially when extreme values of the signal are presented. Thus we present another example in which the distribution of responses is not symmetrical,

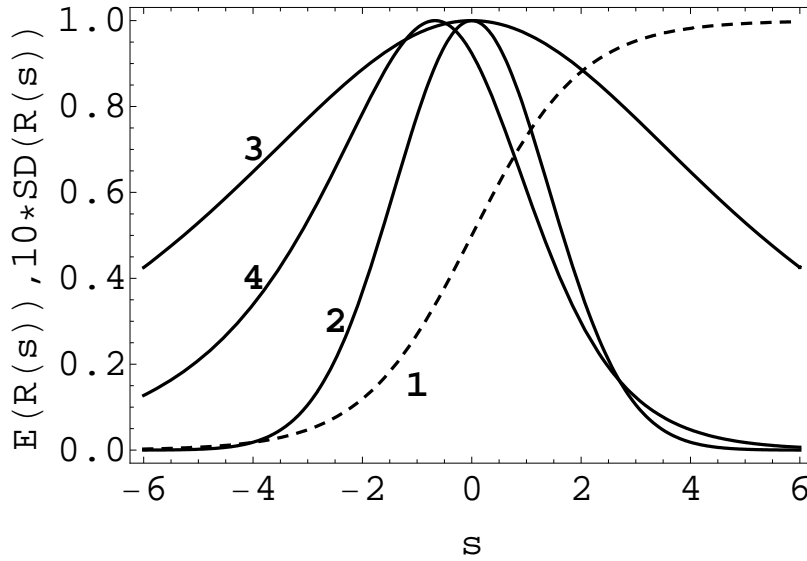$$R(s) \sim \mathsf{B}\left(\beta\, \mathrm{e}^{c\,s},\ \beta\right) \ , \tag{4.5}$$

Figure 4.1: Mean transfer function $m(s)$ (curve 1) and standard deviations $\mathsf{SD}(R(s))$ of the response function for models with Gaussian distributed response with quickly decreasing noise given by (4.3) (curve 2), slowly decreasing noise given by (4.4) (curve 3), and Beta distributed response given by (4.5) (curve 4). Parameters are $\sigma_0^2 = 0.01$ and $c = 0.5$ for the Gaussian distribution and $\beta = 14.15$ for the Beta distribution to ensure the same maximum value of the variance for all types of distribution.

where B denotes the Beta distribution with probability density function

$$B\left(x;a,b\right) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \, x^{a-1}(1-x)^{b-1} \quad \text{for } x \in (0,1) \tag{4.6}$$

and $\beta > 0$ is a free parameter controlling the amplitude of the noise as σ_0 in (4.3) and (4.4). This choice ensures that the mean response is again given by (4.1) but the noise is not symmetrically distributed along the response axis, see Figure 4.1. It is commonly reported that firing rates have standard deviations which are roughly comparable with their mean value (see [16, and references there]). This is in contradiction with our concept of decreasing variability for extreme values of the signal. However, models can be easily modified in this direction by choosing appropriate dependency of the noise on the signal. Examples of similar kind of dependencies can be seen in the empirical and the biophysical model studied in this paper.

Under the deterministic scenario, $\sigma(s) \equiv 0$, which is an idealization, as seen from the experimental data, there is a one-to-one correspondence between signal and response. In such a case any signal has a unique response and from the detectability point of view all signals are equal. The same is true from the point of view of transferred information. Nevertheless, it may be proposed that the most suitable signal is that for which the difference between the responses, $m(s_i)$ and
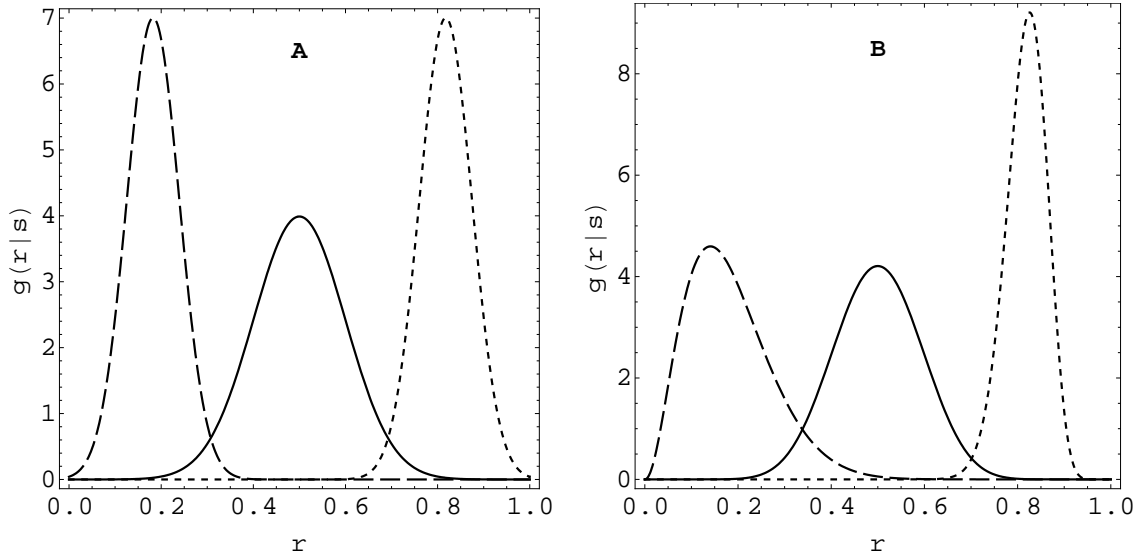
Figure 4.2: Examples of response distributions $g(r|s)$. Probability density functions for $s = -1.5$ (dashed lines), $s = 0$ (solid lines) and $s = 1.5$ (dotted lines) are plotted. (A) Model with Gaussian distribution of the response with quickly decreasing noise – equation (4.3). In model with slowly decreasing noise of the response – equation (4.4) – the distribution behaves qualitatively in the same way, but maximum of the density varies slowly. (B) Model with Beta distribution of the response – equation (4.5). Parameters are the same as in Figure 4.1, $\sigma_0^2 = 0.01$, $c = 0.5$ and $\beta = 14.15$.

$m(s_j)$, to pair of different signal values, s_i and s_j, keeping the same distance, $|s_i - s_j| = \text{const.}$, is the highest. This has been pointed out by Borst and Theunissen [12]: "the information in the response could be measured as the difference between $m(s_i)$ and $m(s_j)$". Obviously, such a signal for transfer functions (4.2) and (4.5) is at $s = 0$, where the derivative of the transfer function is maximal.

4.2.2 Signal bringing the highest information

Information processing is the role of neuronal systems, and thus it is not surprising that classification of signals by their information content is very common in computational neuroscience (see [58]). Under the stochastic scenario, for a given signal s, there is not a unique response but a family of responses. Each of the responses appears with a predefined probability. From this point of view, equations (4.2), (4.5) can be seen as determining the conditional probabilities of responses given specific stimuli. The conditional probability density functions $g(r|s)$ given by the theoretical models defined above are illustrated in Figure 4.2.

Application of information theory is one possible extension of the criteria based on the derivative of the mean transfer function mentioned in the previous subsection. The use of information concept is based on calculating "a distance"

between the conditional distribution of response, $g(r|s)$, and the unconditional distribution of response, $g(r)$. For that purpose, however, the stimulus has to be considered as a random variable, S, with probability distribution $f(s)$, instead of simply considering the deterministic stimulus. Then, by using Bayes formula, the unconditional distribution of the response is calculated by

$$g(r) = \int_M g(r|s)f(s)\mathrm{d}s \ , \tag{4.7}$$

where the range of integration M depends on the distribution of stimuli. Alternatively, instead of assuming the distribution $f(s)$, this approach can be used to predict the stimulus distribution in natural conditions. This concept was introduced in the study of the visual sensory system by Laughlin [47] and extended recently to the auditory (see e.g. [48]) and olfactory systems (see e.g. [35]) by taking into account the specificity of these modalities.

Several examples of stimulus distribution, $f(s)$, can be considered. The simplest example is that of a uniform distribution, $S \sim \mathsf{U}\,(-a/2, a/2)$, in which all stimuli are equally likely. If we take into account that inflection point of the response function (4.1) is located at zero, formally, the stimuli take values in the interval $(-a/2, a/2)$ and $f(s) = 1/a$. Another, maybe more realistic assumption is that the stimuli are normally distributed, $S \sim \mathsf{N}\,(0, a^2/12)$, avoiding the extremal values. For the sake of comparison we took variance equal to $a^2/12$ which ensures in both cases the same variance of the signal.

Now there are several options to define the distance between $g(r|s)$ and $g(r)$. One possibility is to use the formula

$$I_1(R|s) = \int g(r|s) \ln \frac{g(r|s)}{g(r)} \mathrm{d}r \ . \tag{4.8}$$

Then the optimum signal is determined by searching the maximum of the criteria $I_1(R|s)$. From (4.8) follows that $I_1(R|s) = 0$ if the distribution of responses does not depend on the signal, $g(r|s) = g(r)$. Criterion (4.8) was called *specific surprise* by DeWeese and Meister [19] and mentioned by several authors ([12, 11], originally introduced by Fano [21]). It is Kullback-Leibler divergence between marginal distribution gr and conditional distribution $g(r|s)$ (see [34, 36, 55]). It is "a distance" in some sense but not a true metric because of lack of symmetry between the two distributions.

DeWeese and Meister [19] argued that $I_1(R|s)$ does not provide a good measure of stimulus significance and proposed a *stimulus-specific* information

$$I_2(R|s) = -\int g(r) \ln g(r)\mathrm{d}r + \int g(r|s) \ln g(r|s)\mathrm{d}r \ , \tag{4.9}$$

which is again equal to zero for $g(r|s) = g(r)$. The method was further discussed by Butts [15]. Finally Bezzi [11] added two other measures, which are not mentioned in the present chapter because it is not primarily oriented at comparing information-based measures to one another.

4.2.3 Concept based on best identifiable signal

In this section we investigate which deterministic signal, s, can be identified from the transfer function, $R(s)$, with greater precision. As a general measure of signal optimality from the point of view of its identification we propose to use Fisher information. Fisher information has become a common tool in computational neuroscience, see [65, 14, 24, 27, 68, 10, 69, 28, 6, 25, 26, 43, 44, 3, 70]. Now we see the conditional distribution of the response for a specific signal, appearing in the previous section, as a probability density function belonging to a parametric family $g(r; s)$, where s is an unknown parameter which should be identified from the observation of responses r.

Recalling Definition 3 on page 13, Fisher information with respect to parameter s is defined by equation

$$J(s) = \int \frac{1}{g(r;s)} \left(\frac{\partial\, g(r;s)}{\partial\, s} \right)^2 \mathrm{d}r \ .$$

(4.10)

The use of Fisher information as a tool to locate the optimal signal for information transfer is theoretically motivated by Cramér-Rao inequality. It says that the variance of an unbiased estimate of the signal cannot be smaller than the inverse of Fisher information, $\mathsf{Var}(\hat{s}) \geq J(s)^{-1}$, see [18]. In other words, the largest the Fisher information, the best the estimate of s can be achieved. This conclusion is important to know how well one can expect to identify the signal.

Criterion (4.10) requires a complete knowledge of the distribution $g(r; s)$, but an approximation of the Fisher information can also be used. It is a lower bound of J based on the first two moments of the random variable $R(s)$,

$$J_2(s) = \frac{1}{\mathsf{Var}(R(s))} \left(\frac{\partial\, \mathsf{E}(R(s))}{\partial\, s} \right)^2 \ ,$$

(4.11)

see Definition 7 on page 16 and [43] for details. This is of exceptional importance in an experimental context. It is common to measure (estimate) the moments (mean and variance) of the responses, but their complete distribution is rarely evaluated. In addition, for a large class of distributions there is an equality $J_2 = J$ (see Theorem 11 on page 19 and [43]), and obviously J_2 is computationally much simpler to obtain as it requires only the first two moments but not the complete probability distribution.

4.3 Results

4.3.1 Statistical model

The model is characterized by response function (4.1) and the chosen distribution of the noise. For the normal distribution of the response we have

$$g(r|s) = \frac{1}{\sigma(s)\sqrt{2\pi}} \exp\left\{-\frac{(r-m(s))^2}{2\sigma^2(s)}\right\} . \tag{4.12}$$

For Beta distribution, an analogous formula can be written. Assuming the signal to be a random variable, S, with the probability distribution $f(s)$ and having the conditional distribution of the response, $g(r|s)$, the unconditional distribution $g(r)$ of the response can be calculated by (4.7). It is shown in Figure 4.3 for both the uniform as well as Gaussian distribution of the signal. The shapes of $g(r)$ for the Gaussian distribution of the responses are similar (Figure 4.3A); they differ only quantitatively. The distribution is bimodal, that means without knowledge about realization of the signal, s, there is a high probability to observe the response near 0 as well as near 1. For the Beta distribution (Figure 4.3B), $g(r)$ becomes asymmetric.

Now all the quantities required to evaluate the signal yielding maximal information based on measures I_1 (4.8) and I_2 (4.9) are available and their profiles can be calculated, see Figure 4.4. It shows that the most informative signals are those which are extreme, only the speed at which these values are achieved is different. The only exception is illustrated in Figure 4.4B, where there is a local maximum in the center of the coding range. The results are not qualitatively different for the Beta distribution, only the shapes are not symmetrical.

For determining the best identifiable signal we return to the criteria based on Fisher information. For the model with variance (4.3), the optimality criterion J_2 is evaluated directly from (4.11),

$$J_2(s) = \frac{e^{c\,s^2+2s}}{\sigma_0^2 \left(1+e^s\right)^4} , \tag{4.13}$$

and the criterion J is calculated directly from (4.10),

$$J(s) = 2c^2 s^2 + \frac{e^{c\,s^2+2s}}{\sigma_0^2 \left(1+e^s\right)^4} . \tag{4.14}$$

For the model with variance given by (4.4), we can derive

$$J_2(s) = \frac{\left(1+e^{c\,s}\right)^2 e^{2s-c\,s}}{4\sigma_0^2 \left(1+e^s\right)^4} , \tag{4.15}$$
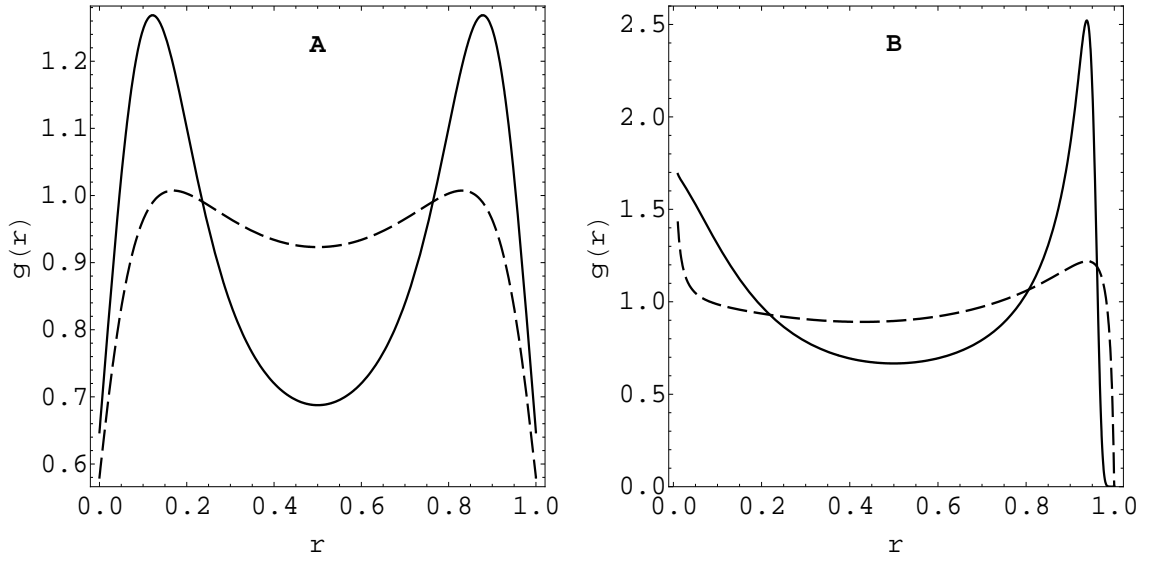
Figure 4.3: Examples of probability density functions $g(r)$ of response distributions under the condition of uniform distribution of the signal (solid lines) and under the condition of Gaussian distribution of the signal (dashed lines). (A) Model with Gaussian distribution of the response with slowly decreasing noise – equation (4.4). In model with quickly decreasing noise – equation (4.3) – the distribution behaves qualitatively in the same way. (B) Model with Beta distribution of the response – equation (4.5). Parameters are the same as in Figure 4.1, $\sigma_0^2 = 0.01$, $c = 0.5$ and $\beta = 14.15$.

the criterion J is evaluated numerically. In model (4.5) with Beta distributed noise we obtain

$$J_2(s) = \frac{(1 + e^s)\beta + 1}{(1 + e^s)^2} e^s \quad , \qquad (4.16)$$

the criterion J is evaluated numerically. The results are illustrated in Figure 4.5 which shows that the type of noise entirely determines the position of the best detectable signal. For Gaussian distributed response with quickly decreasing noise the best identifiable signal is located at low as well as high stimulus intensities. For that one with slowly decreasing noise it is located around $s = 0$. In the case of Beta distributed noise the best identifiable signal is at high stimulus intensities.

4.3.2 Empirical model

The criteria summarized above were also applied to data on olfactory receptor neurons presented by Tomaru and Kurahashi [66] where the transfer function (firing frequency) and its standard deviation are plotted against the injected current. It gives evidence that the noise is not symmetrically distributed along the stimulus axis. This example, which is taken from a study on the olfactory receptor
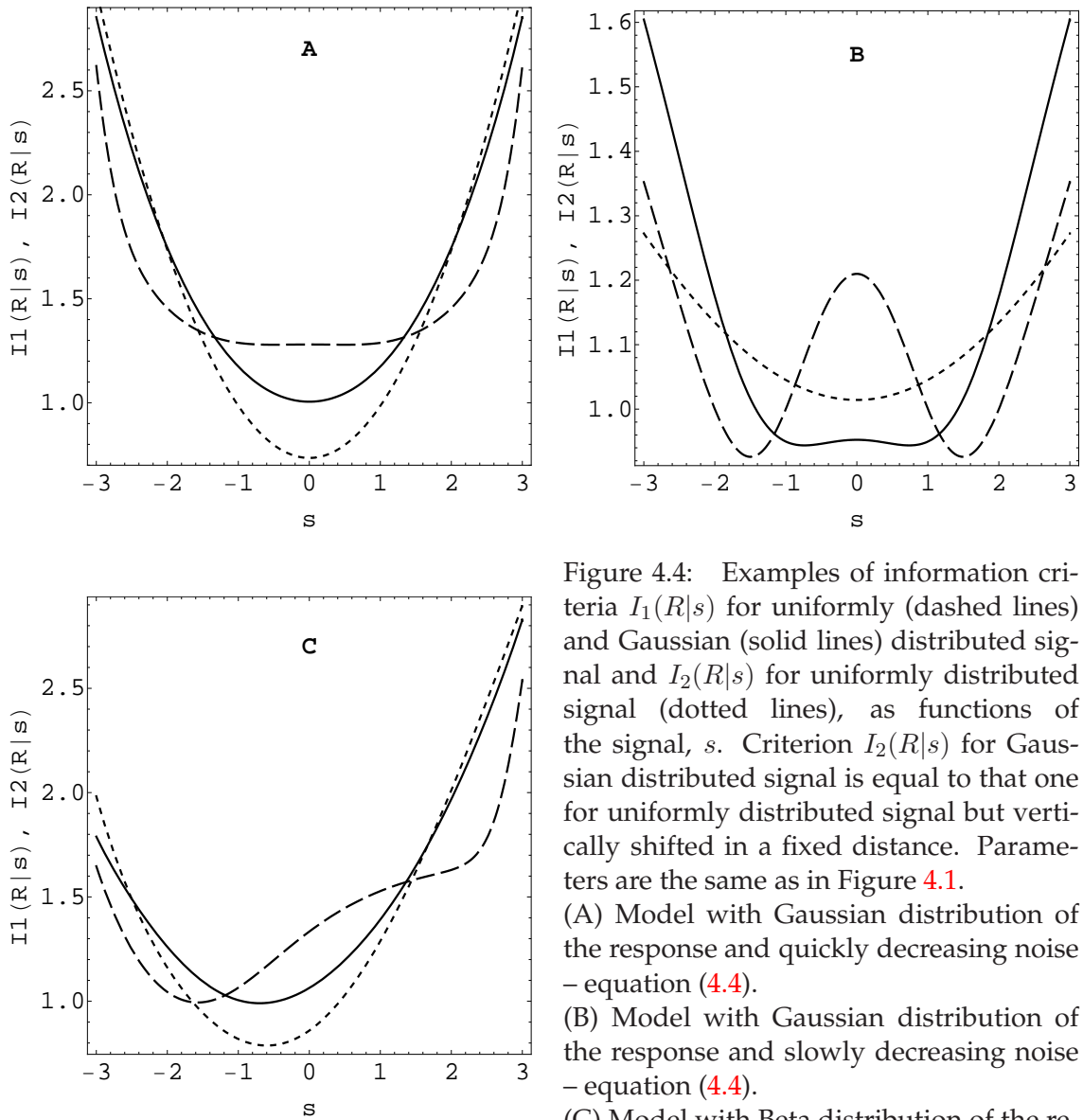
Figure 4.4: Examples of information criteria $I_1(R|s)$ for uniformly (dashed lines) and Gaussian (solid lines) distributed signal and $I_2(R|s)$ for uniformly distributed signal (dotted lines), as functions of the signal, s. Criterion $I_2(R|s)$ for Gaussian distributed signal is equal to that one for uniformly distributed signal but vertically shifted in a fixed distance. Parameters are the same as in Figure 4.1.

(A) Model with Gaussian distribution of the response and quickly decreasing noise – equation (4.4).

(B) Model with Gaussian distribution of the response and slowly decreasing noise – equation (4.4).

(C) Model with Beta distribution of the response – equation (4.5).
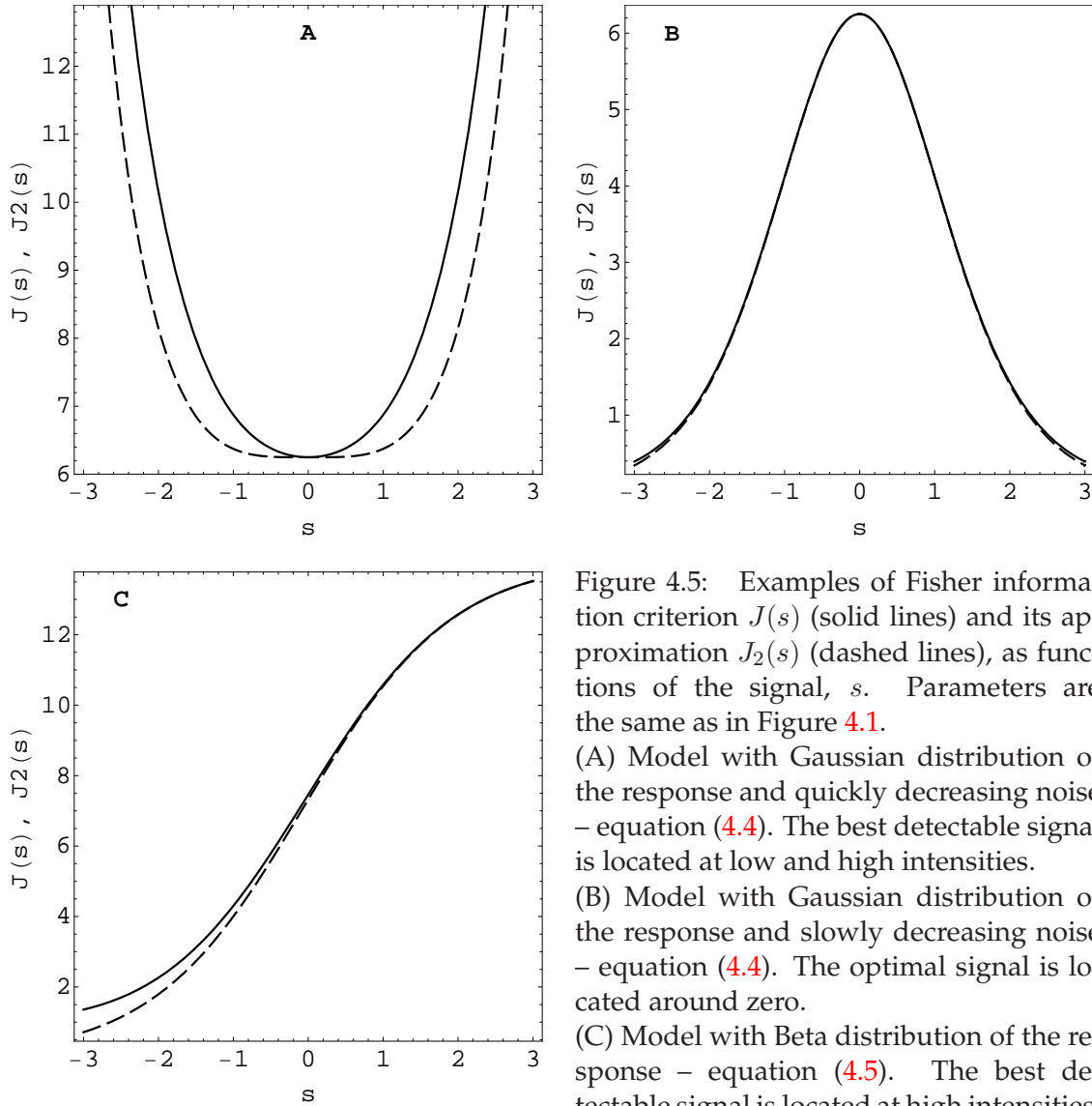
Figure 4.5: Examples of Fisher information criterion $J(s)$ (solid lines) and its approximation $J_2(s)$ (dashed lines), as functions of the signal, s. Parameters are the same as in Figure 4.1.
(A) Model with Gaussian distribution of the response and quickly decreasing noise – equation (4.4). The best detectable signal is located at low and high intensities.
(B) Model with Gaussian distribution of the response and slowly decreasing noise – equation (4.4). The optimal signal is located around zero.
(C) Model with Beta distribution of the response – equation (4.5). The best detectable signal is located at high intensities.

neuron, has its parallel in the visual system. Figure 8 of [16] illustrates the relationship between the firing rate (mean and standard deviation) and the injected current. The author concludes that it should be fairly common for the firing-rate variance to saturate at high stimulus intensity, possibly showing a plateau or even a decrease.

The data from Figure 2 of [66] has been reconstructed and used as a model for establishing the optimum signal. The properties of the experimental data analogous to the theoretical models illustrated in Figure 4.1 are given in Figure 4.6. Comparing these two figures shows that the empirical data cover only the left part of the transfer function and the rest has to be hypothesized. Furthermore, the response in Figure 4.6 is not normalized, in contrast to Figure 4.1.
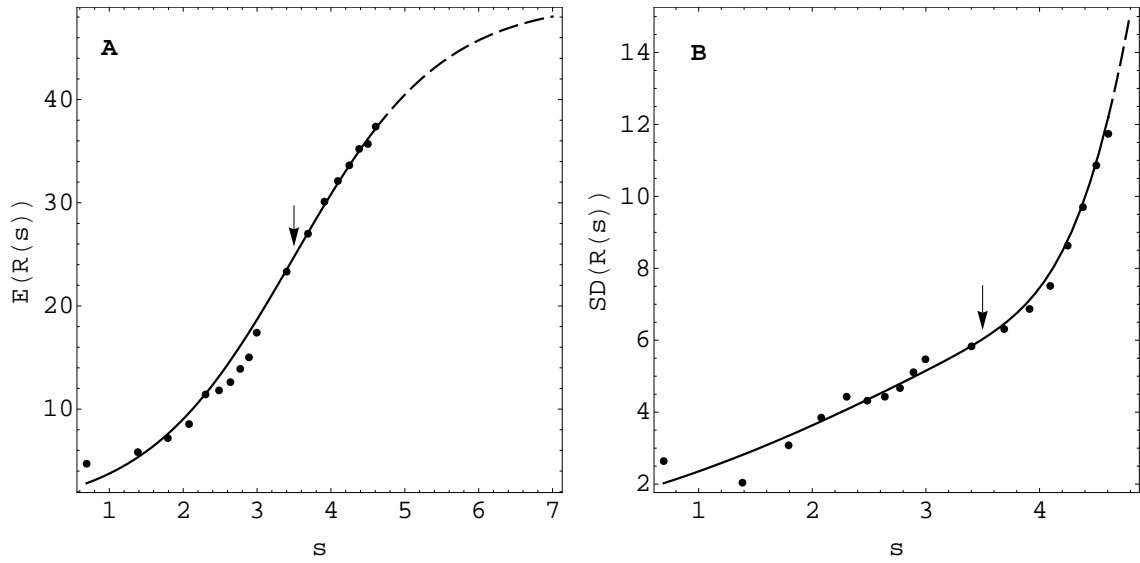
Figure 4.6: Empirical data – neuron firing frequency [Hz] as a function of logarithm of injected current (originally [pA]) – from [66] and their properties. Original data are marked as small discs. (A) Mean response function $m(s)$ and (B) standard deviation $\text{SD}(R(s))$ are plotted as functions of injected current, s. Dashed parts of curves denote theoretical continuation of the dependencies on the signal. Arrows denote the position of the inflection point of the mean response function $m(s)$ at $s = 3.5$.

The theoretical mean response function is

$$m(s) = \frac{49.5}{1 + \exp\{3.5 - s\}} \tag{4.17}$$

with inflection point at $s = 3.5$. Thus from the deterministic point of view, signal at this level is the optimal one. The standard deviation was fitted by function

$$\text{SD}(R(s)) = 8.75 \exp\left\{-\left(\frac{s - 6.5}{4.8}\right)^2\right\} + 29.9 \exp\left\{-\left(\frac{s - 6.1}{1.1}\right)^2\right\} . \tag{4.18}$$

Knowing the mean and the standard deviation, one can evaluate J_2 but to obtain the other measures, the distributions of the responses and stimuli have to be assumed. We investigated two of them, Gaussian and Beta, as in the theoretical models. There is no obvious difference between the distribution of response here and in the theoretical models (see Figure 4.2). The shapes of the conditional, $g(s|r)$, and unconditional, $g(r)$, probability density functions of the response for both the Gaussian and Beta distribution are plotted in Figure 4.7 and Figure 4.8.

The results for the criteria based on Fisher information are shown in Figure 4.9. For both models the Fisher information behaves practically in the same way, except in the rightmost part where the Beta distribution creates a different pattern. The optimal signal is shifted to the left with respect to that derived
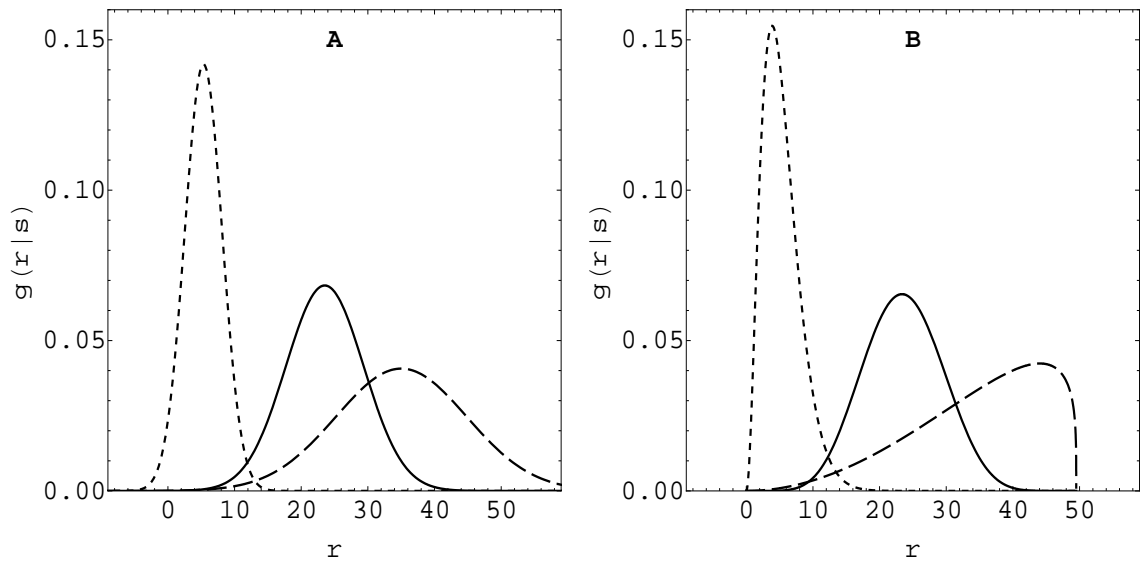
Figure 4.7: Conditional probability density functions of the response for empirical data shown in Figure 4.6: $s = 1.386$ (dotted lines; corresponds to injected current 4 pA), $s = 3.401$ (solid lines; 30 pA), $s = 4.382$ (dashed lines; 80 pA), (A) for Gaussian distribution, (B) for Beta distribution. Note, that the for values of signal, s, in the center of range of used current the shapes of probability density function for both Gaussian and Beta distribution are almost identical, whereas for low or high current the shapes differ.
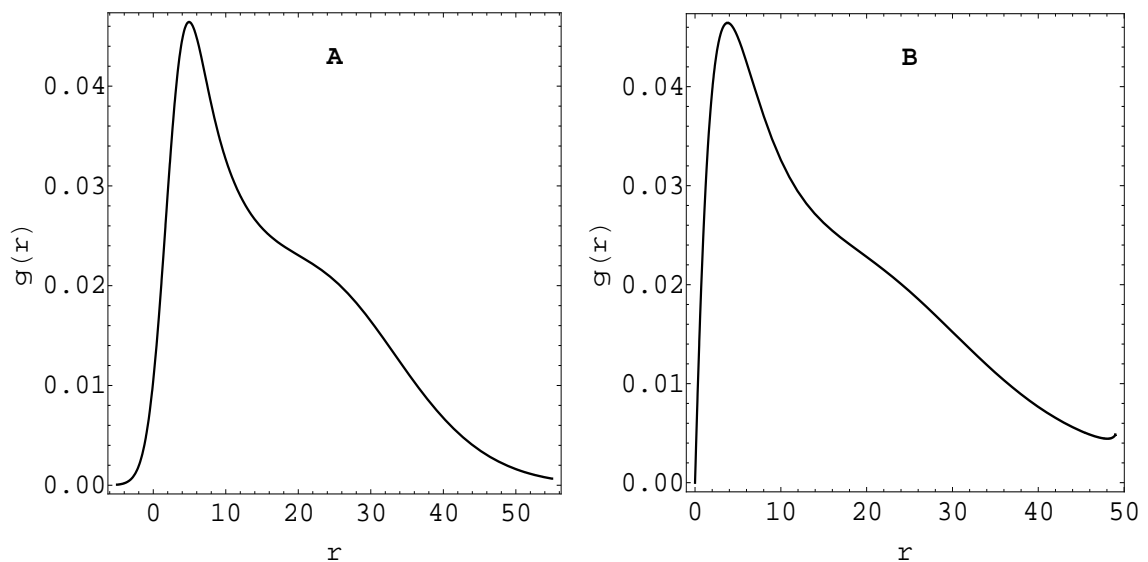
Figure 4.8: Unconditional probability density functions of the response for empirical data shown in Figure 4.6, (A) for Gaussian distribution, (B) for Beta distribution. The theoretical most probable firing frequency has value around 5 Hz (see A) or 4 Hz (see B).
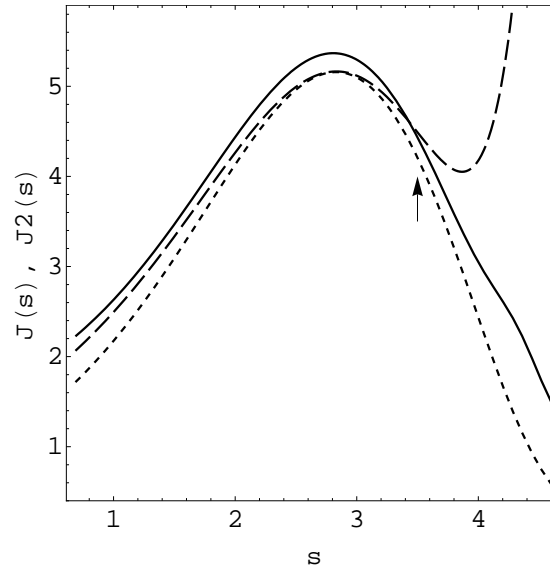
Figure 4.9:   Fisher information results for empirical data shown in Figure 4.6. Fisher information $J(s)$ for Gaussian distributed response (solid line), for Beta distributed response (dashed line) and their lower bound $J_2(s)$ (dotted line) are plotted as functions of the injected current, s. The arrow denotes the position of the inflection point of the mean response curve $m(s)$ at $s = 3.5$.

from the deterministic model. The optimal current has value around 17 pA (corresponds to $s \approx 2.833$ in Figure 4.9). Further, as mentioned, for application of information measures a prior distribution of signals must be assumed. We take a uniform distribution on the interval $(\ln 2, \ln 100)$, which is the interval of the applied stimuli. The information measures are given in Figure 4.10. The criterion I_1 is almost the same for both models, with high values for low intensity signals and with lowest values for signals between 2 and 3. The local maximum of I_1 which appears for Gaussian distribution of the response near $s = 4$ vanishes for Beta distributed response. In both models, the criterion I_1 is decreasing function of the signal.

4.3.3   Biophysical model

For illustrating the proposed measures of signal optimality on a realistic model of the responses we study olfactory receptor neurons located in the nasal olfactory epithelium. When stimulated during, say, one second, odorant molecules interact with receptor proteins embedded at the membrane surface of these neurons. Binding of odorants to receptors triggers biochemical events that result in the opening of ionic channels. The ensuing current generates a locally spreading receptor potential, which in turn initiates a spike train. The relations between
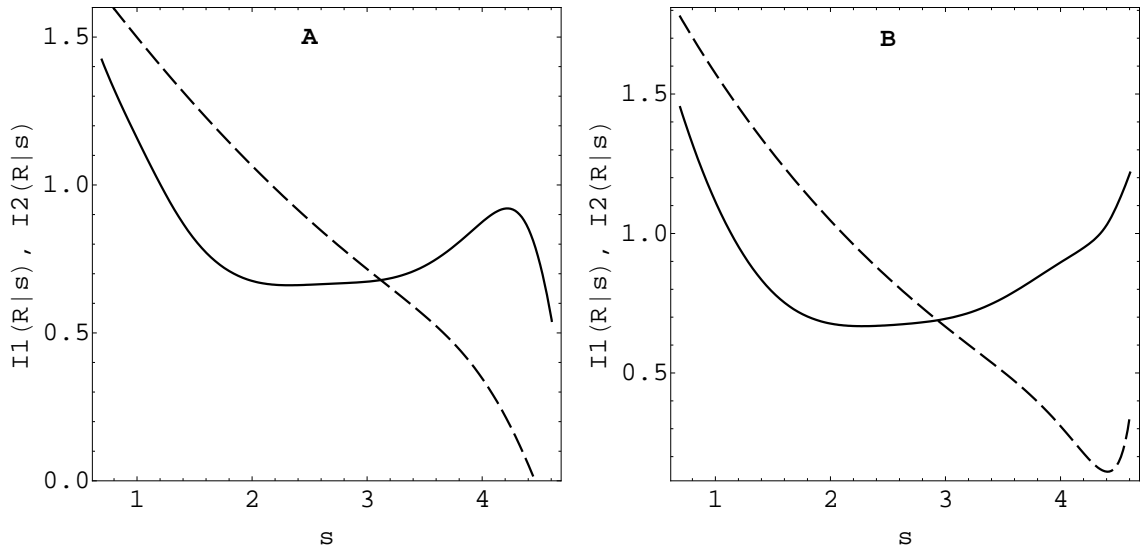
Figure 4.10: Information criteria $I_1(R|s)$ (solid line) and $I_2(R|s)$ (dashed line) for empirical data shown in Figure 4.6, (A) for Gaussian distribution of the response, (B) for Beta distribution of the response.

the concentration s of odorant molecules and the density of activated receptors, or the neuron depolarization, or its firing rates, are examples of transfer functions. We investigated a stochastic variant of a model describing odorant transport through the mucus, enzymatic deactivation of odorant molecules in the mucus, and receptor binding then activation (see [30, 31, 32, 42, 37, 60, 61, 63]) with parameter values given in [32]. The originally deterministic model was randomized by adding two types of noise (Gaussian and Beta) to the number of occupied and activated receptors. The properties of the steady-state distribution was estimated from repeated simulations of the model.

The statistical properties of the model are given in Figure 4.11. The mean number of activated receptors is saturated at value $0.24\ \mu$M for the odorant log-concentration equal to -6.88 (corresponds to odorant concentration $0.001\,03\ \mu$M). The standard deviation of the number of activated receptors is increasing. The position on horizontal scale depends on the flux constant, $k_{\mathrm{I}} = 29\,000\ \mathrm{s}^{-1}$. Decreasing k_{I} would result in a shift of the curves to higher concentrations. Note the similarity between Figure 4.11 and Figure 4.6.

Results for the criteria based on Fisher information of the best identifiable stimulus – Figure 4.13 – show that the model behaves similarly whatever the type of noise (see Figure 4.12). The optimal signal can be found in the range of odorant concentrations where a peak occurs. In the range $(-9.5, -6.88)$ Fisher information $J(s)$ is rather flat with a Gaussian noise and presents a peak with a Beta noise. The lower bound J_2 shows only a peak at relatively high odorant con-
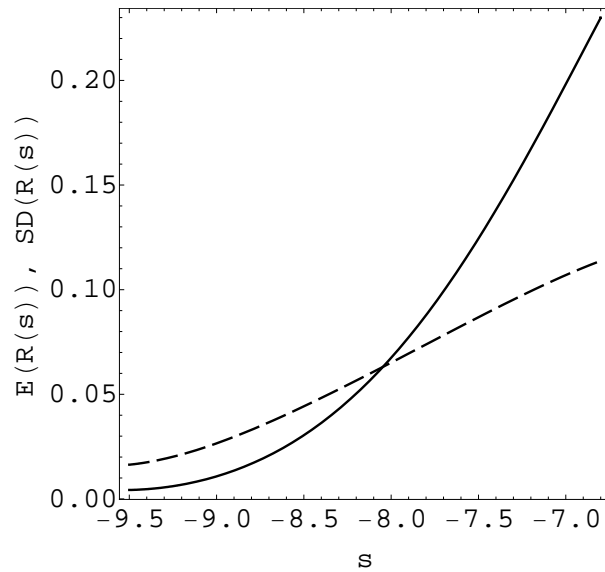
Figure 4.11: Biophysical model – concentration of activated receptors [μM] as a function of odorant log-concentration – and its properties. Mean response function $m(s)$ (solid line) and standard deviation $\mathsf{SD}(R(s))$ (dashed line) are plotted as functions of the natural logarithm, s, of the odorant concentration.
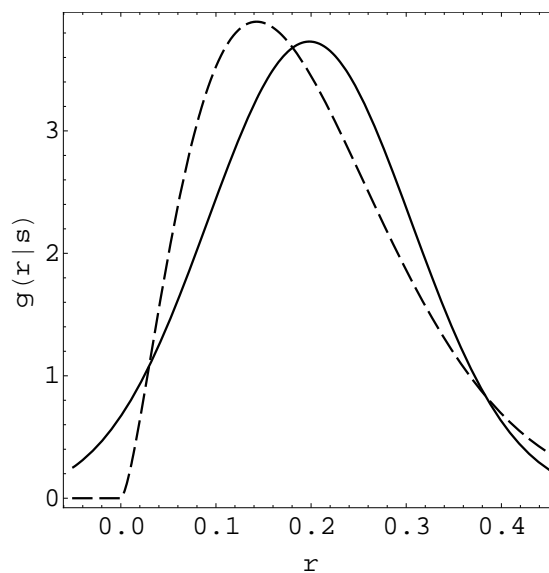

Figure 4.12: Conditional probability density functions of the response concentration [μM] for the biophysical model characterized in Figure 4.11 and $s = -7$ (corresponds to odorant concentration 0.00091 μM) for Gaussian (solid line) and Beta distribution (dashed line).
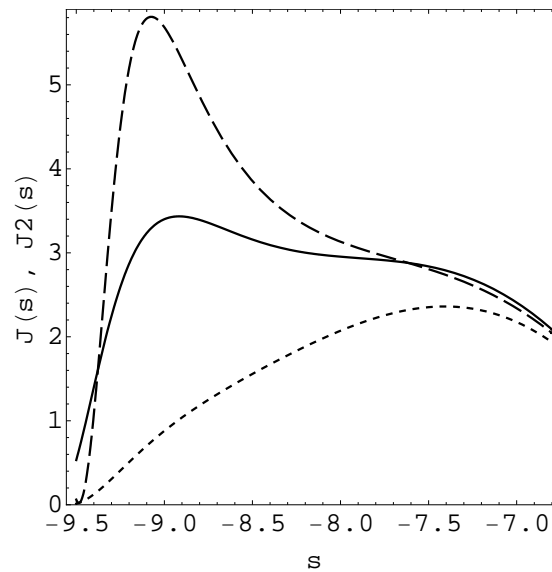
Figure 4.13:   Fisher information results for the biophysical model characterized in Figure 4.11. Fisher information $J(s)$ for Gaussian distributed response (solid line), for Beta distributed response (dashed line) and their lower bound $J_2(s)$ (dotted line) are plotted as functions of the natural logarithm, s, of the odorant concentration.

centration. The Fisher information criteria give high value also for extremely low odorant concentrations. One possible explanation of this phenomenon lies in incorrectness of used continuous stochastic process for extremely low odorant concentrations. In such case the concentration changes in relatively small steps rather than continuously.

For application of information measures we have to assume a prior distribution of signal. The information measures are given in Figure 4.14 for a uniformly distributed signal on interval $(-9.5, -6.88)$. The criterion $I_1(R|s)$ forms qualitatively the same pattern for both models. It has a local minimum approximately in the center of studied range of the signals. In both models, the criterion $I_2(R|s)$ is decreasing function of the signal.
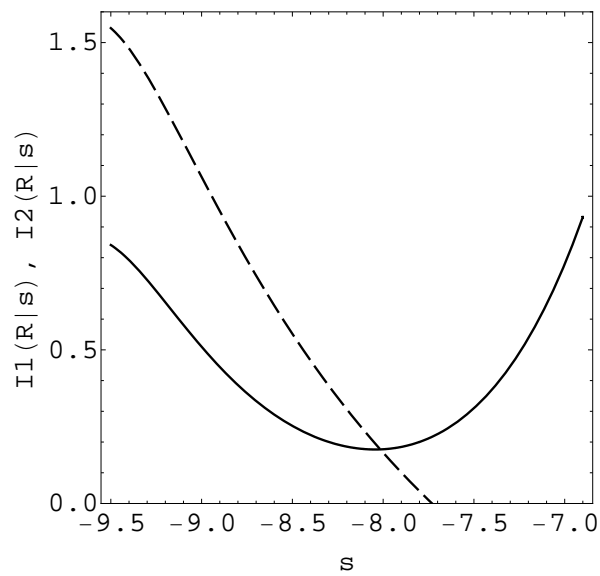
Figure 4.14: Information criteria $I_1(R|s)$ (solid line) and $I_2(R|s)$ (dashed line) for Gaussian distributed response of the biophysical model characterized in Figure 4.11 are plotted as functions of the natural logarithm, s, of the odorant concentration. For Beta distributed response both criteria behave qualitatively in the same way.

Conclusions 5

Three theoretical models of the concentration detector and one of the flux detector were searched for the optimal signal as defined by the application of the Fisher information measure and its approximations. The result for the model of basic interaction coincides with the optimum obtained via maximum slope of the input-output function, which is located at its inflexion point. In models of the concentration detector with activation stage (simple activation or double-step interaction), the statistically determined optimal signal differs from the inflexion point. In the flux detector model both the deterministic and stochastic approach give the same result, too. In general, models with separated binding and activation yield the variances of count of activated receptors which can change the position of the optimal odorant concentration.

Further, criteria based on measures of transferred information was computed for a theoretical model with different types of added noise and for two realistic models. Adding noise to the transfer function has significant consequences. In the studied models, the variances are similar, except that they decrease faster one case or slower in the other. The optimal values of the signal are different. The shapes of the measures in dependency on the signal are not only different, but also the values are reached with different speeds. This fact is important for identifying the dynamical range.

The CD-ROM which is a supplement of this dissertation includes few M-files for Matlab. These scripts provide functions for numerical computation of steady--state probability distribution of the number of activated receptors $C(s)$,

DS1.m   in double-step interaction (3.22),
FD1.m, **FD2.m**   in flux detector model (3.26),
K1.m, **K2.m**   in biophysical model (see Section 4.3.3).

Other group of scripts provide functions for simulation of progress of the number of activated receptors $C_t(s)$ in time,

sB1.m   in basic interaction (3.1),
sB2.m   in basic interaction with multiple types of receptors,
sFD.m   in flux detector model (3.26).

References

[1] Pokora, O. – Lánský, P. steady-state Properties of Coding of Odor Intensity in Olfactory Sensory Neurons. *Advances in Brain, Vision, and Artificial Intelligence, LNCS* **4729**: 360–367, 2007.

[2] Pokora, O. – Lánský, P. Statistical approach in search for optimal signal in simple olfactory neuronal models. *Mathematical Biosciences* **214**: 100–108, 2008.

[3] Lánský, P. – Pokora, O. – Rospars, J.-P. Stimulus-Response Curves in Sensory Neurons: How to Find the Stimulus Measurable with the Highest Precision. *Advances in Brain, Vision, and Artificial Intelligence, LNCS* **4729**: 360–367, 2007.

[4] Lánský, P. – Pokora, O. Classification of stimuli based on stimulus-response curves and their variability. *Brain Research* **122**: 57–66, 2008.

[5] Abramowitz, M. – Stegun, I. A. *Handbook of Mathematical Functions*. Dover Publications, New York, 1972.

[6] Amari, S. – Nakahara, H. Difficulty of singularity in population coding. *Neural Computation* **17**: 839–858, 2005.

[7] Anděl, J. *Základy matematické statistiky*. MATFYZPRESS, Praha, 2005.

[8] Arányi, P. – Tóth, J. A Full Stochastic description of the Michaelis-Menten reaction for small systems. *Acta Biochimica et Biophysica Academiae Scientificarum Hungariae* **12**: 375–388, 1977.

[9] Beidler, B. W. Taste receptor stimulation. *Prog. Biophys. Chem.* **12**: 107–151, 1962.

[10] Bethge, M. – Rotermund, D. – Pawelzik, K. Optimal short-term population coding: When Fisher information fails. *Neural Computation* **14**: 2317–2351, 2002.

[11] Bezzi, M. Quantifying the information transmitted in a single stimulus. *BioSystems* **89**: 4–9, 2007.

[12] Borst, A. – Theunissen, F. E. Information theory and neural coding. *Nature Neuroscience* **2**: 947–957, 1999.

[13] Brown, D. – Rothery, P. *Models in Biology: Mathematics, Statistics and Computing*. Wiley, Chichester, 1993.

[14] Brunel, N. – Nadal, J. P. Mutual information, Fisher information and popu-

lation coding. *Neural Computation* **10**: 1731–1757, 1998.

[15] Butts, D. A. How much information is associated with a particular stimulus?. *Network: Computational Neural Systems* **14**: 177–187, 2003.

[16] Carandini, M. Amplification of trial-to-trial response variability by neurons in visual cortex. *PLoS Biology* **2**: 1483–1493, 2004.

[17] Cleland, T. A. – Linster, C. Concentration tuning by spare receptor capacity in olfactory sensory neurons: a theoretical study. *Neural Computation* **11**: 1673–1690, 1999.

[18] Cramér, H. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, 1946.

[19] DeWeese, M. R. – Meister, M. How to measure the information gained from one symbol. *Network* **10**: 325–340, 1999.

[20] Érdi, P. – Tóth, J. *Mathematical Models of Chemical Reactions*. Princeton University Press, Princeton, 1989.

[21] Fano, R. M. *Transmission of Information: A Statistical theory of Communication*. MIT Press, New York, 1961.

[22] Green, D. M. – Swets, J. A. *Signal detection theory and psychophysics*. Wiley, New York, 1966.

[23] Gurewich, K. G. – Agutter, P. S. – Wheatley, D. N. Stochastic description of the ligand-receptor interaction of biologically active substances at extremely low doses. *Cellular Signalling* **15**: 447–453, 2003.

[24] Greenwood, P. E. – Ward, L. M. – Russel, D. F. – Neiman, A. – Moss, F. Stochastic resonance enhances the electrosensory information available to paddlefish for prey capture. *Physical Review Letters* **84**: 4773–4776, 2000.

[25] Greenwood, P. E. – Lánský, P. Optimum signal in a simple neuronal model with signal-dependent noise. *Biological Cybernetics* **92**: 199-205, 2005.

[26] Greenwood, P. E. – Lánský, P. Information content in threshold data with non-Gaussian noise. *Fluctuation Noise Letters* **7**: L79–L89, 2007.

[27] Jenison, R. L. Decoding first-spike latency: A likelihood approach. *Neurocomputation* **38**: 239–248, 2001.

[28] Johnson, D. H. – Ray, W. Optimal stimulus coding by neural populations using rate codes. *Journal of Computational Neuroscience* **16**: 129–138, 2004.

[29] Kaissling, K. E. Kinetics of olfactory receptor potentials. *Olfaction and Taste III*. Rockefeller University Press, New York, 1969.

[30] Kaissling, K. E. Flux detectors vs. concentration detectors: two types of chemoreceptors. *Chemical Senses* **23**: 99–111, 1998.

[31] Kaissling, K. E. Olfactory perireceptor and receptor events in moths: a kinetic model. *Chemical Senses* **26**: 125–150, 2001.

[32] Kaissling, K. E. – Rospars J.-P. Dose-response relationships in an olfactory flux detector model revisited. *Chemical Senses* **29**: 529–531, 2004.

[33] Karlin, S. – Taylor, H. M. *A First Course in Stochastic Processes, Second Edition*. Academic Press, San Diego, London, 1975.

[34] Košťál, L. – Lánský, P. – Rospars J.-P. Neuronal coding and spiking randomness. *European Journal of Neuroscience* **6**: 2693–2701, 2007.

[35] Košťál, L. – Lánský, P. – Rospars J.-P. Efficient olfactory coding in the pheromone receptor neuron of a moth. *PLoS Computational Biology* **4**. , 2008.

[36] Koyama, S. – Shinomoto, S. Inference of intrinsic spiking irregularity based on the Kullback-Leibler information. *BioSystems* **89**: 69–73, 2007.

[37] Křivan, V. – Lánský, P. – Rospars J.-P. Coding of periodic pulse stimulations in chemoreceptors. *BioSystems* **67**: 121–128, 2002.

[38] Kurtz, T. G. The relationship between stochastic and deterministic models for chemical reactions. *Journal of Chemical Physics* **57**: 2976–2978, 1972.

[39] Lam, H. S. – Lampard, D. G. Modelling of drug receptor interaction with birth and death processes. *Journal of Mathematical Biology* **12**: 153–172, 1981.

[40] Lansky, L. – Rospars J.-P. Coding of odor intensity. *BioSystems* **31**: 15–38, 1993.

[41] Lánský, P. – Rospars J.-P. Odorant concentration and receptor potential in olfactory sensory neurons. *BioSystems* **48**: 131–138, 1998.

[42] Lansky, L. – Getz, W. M. Receptor heterogenity and its effect on sensitivity and coding range in olfactory sensory neurons. *Bulletin of Mathematical Biology* **63**: 885–908, 2001.

[43] Lánský, P. – Greenwood, P. E. Optimal signal estimation in neuronal models. *Neural Computation* **17**: 2240–2257, 2005.

[44] Lánský, P. – Greenwood, P. E. Optimal signal in sensory neurons under extended rate coding concept. *BioSystems* **89**: 10–15, 2007.

[45] Lansky, L. – Sacerdote, L. – Zucca, C. Optimum signal in a diffusion leaky integrate-and-fire neuronal model. *Mathematical Biosciences* **207**: 261–274, 2007.

[46] Lauffenburger, D. A. – Linderman, J. *Receptors: Models for Binding, Trafficking and Signaling*. Oxford University, Oxford, 1993.

[47] Laughlin, S. B. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch* **36**: 910–912, 1981.

[48] Lewicki, M. Efficient coding of natural sounds. *Nature Neuroscience* **5**: 356–363, 2002.

[49] Lehmann, E. L. – Casella, G. *Theory of Point Estimation, 2nd edition*. Springer, New York, 1998.

[50] Mankin, R. W. – Mayer, M. S. A phenomenological model of the perceived intensity of single odorants. *Journal of Theoretical Biology* **100**: 123–138, 1983.

[51] Maurin, F. A mathematical model of chemoreception for odours and taste. *Journal of Theoretical Biology* **215**: 297–303, 2002.

[52] McKeegan, D. E. F. Spontaneous and odour evoked activity in single avian olfactory bulb neurones. *Brain Research* **929**: 48–58, 2002.

[53] McCullagh, P. – Nelder, J. A. *Generalized linear models, 2nd edition*. Chapman & Hall, London, 1989.

[54] McNeil, D. R. – Schach, S. Central limit analogues for Markov population processes. *Journal of the Royal Statistical Society, Series B* **35**: 1–23, 1973.

[55] Nakahara, H. – Amari, S. Information-geometric measure for neural spikes. *Neural Computation* **14**: 2269–2316, 2002.

[56] Nizami, L. – Schneider, B. Auditory dynamic range derived from the mean rate-intensity function in the cat. *Mathematical Biosciences* **141**: 1–28, 1997.

[57] Nizami, L. Estimating auditory neuronal dynamic range using a fitted function. *Hearing Research* **167**: 13–27, 2002.

[58] Rieke, F. – Warland, D. – de Ruyter Van Steveninck, R. – Bialek, W. *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, 1999.

[59] Rospars, J.-P. – Lánský, P. – Tuckwell, H. C. – Vermeulen, A. Coding of odor intensity in a steady-state deterministic model of an olfactory receptor neuron. *Journal of Computational Neuroscience* **3**: 51–72, 1996.

[60] Rospars, J.-P. – Krivan, V. – Lánský, P. Perireceptor and receptor events in olfaction. Comparison of concentration and flux detectors: a modeling study. *Chemical Senses* **25**: 293–311, 2000.

[61] Rospars J.-P., Lansky P, Duchamp-Viret, P. – Duchamp, A. Relation between stimulus intensity and response in frog olfactory receptor neurons in vivo. *European Journal of Neuroscience* **18**: 1135–1154, 2003.

[62] Rospars, J.-P. – Lánský, P. Stochastic pulse stimulation in chemoreceptors and its properties. *Mathematical Biosciences* **188**: 133–145, 2004.

[63] Rospars, J.-P. – Lucas, P. – Coppey, M. Modelling the early steps of transduction in insect olfactory receptor neurons. *BioSystems* **89**: 101–109, 2007.

[64] Sanchez-Montanes, M. A. – Pearce, T. C. Fisher information and optimal odor sensors. *Neurocomputing* **38**: 335–341, 2001.

[65] Stemmler, M. A single spike suffices: The simplest form of stochastic resonance in model neurons. *Network: Computation in Neural Systems* **7**: 687–716, 1996.

[66] Tomaru, A. – Kurahashi, T. Mechanisms determining the dynamic range of the bullfrog olfactory receptor cell. *Journal of Neurophysiology* **93**: 1880–1888, 2005.

[67] Tuckwell, H. C. Diffusion approximations to channel noise. *Journal of Theoretical Biology* **127**: 427–438, 1987.

[68] Wilke, S. D. – Eurich, C. W. Representational accuracy of stochastic neural populations. *Neural Computation* **14**: 155–189, 2002.

[69] Wu, S. – Amari, S. – Nakahara, H. Information processing in a neuron

ensemble with the multiplicative correlation structure. *Neural Networks* **17**: 205–214, 2004.

[70] Zhang, K. C. – Sejnowski, T. J. Neuronal tuning: To sharpen or broaden?. *Neural Computation* **11**: 75–84, 1999.

List of figures