Bayesian gradient sensing in the presence of rotational diffusion

Maja Novak and Benjamin M Friedrich
TU Dresden, Dresden, Germany
* Author to whom any correspondence should be addressed.
E-mail: benjamin.m.friedrich@tu-dresden.de

Keywords: chemotaxis, physical limits of chemosensation, rotational diffusion, sequential Bayesian estimation, bearing tracking, time-varying environment

Abstract

Biological cells estimate concentration gradients of signaling molecules with a precision that is limited not only by sensing noise, but additionally by the cell’s own stochastic motion. We ask for the theoretical limits of gradient estimation in the presence of both motility and sensing noise. We introduce a minimal model of a stationary chemotactic agent in the plane subject to rotational diffusion with rotational diffusion coefficient $D$. The agent uses Bayesian estimation to optimally infer the gradient direction relative to itself from noisy concentration measurements. Meanwhile, this direction changes on a time-scale $1/D$. We show that the optimal effective measurement time, which characterizes the time interval over which past gradient measurements should be averaged to reduce sensing noise, does not scale with the rotational diffusion time $1/D$, but with the square root $(rD)^{-1/2}$, where $r$ is a rate of information gain defined as a signal-to-noise ratio normalized per unit time. This result for gradient sensing parallels a recent result by Mora et al (2019 Phys. Rev. Lett.) for sensing absolute concentration in time-varying environments.

1. Introduction

Many motile biological cells navigate in concentration gradients of signaling molecules in a process termed chemotaxis [1–6]. At cellular scales, the stochastic binding of signaling molecules results in molecular shot noise and renders concentration measurements inherently noisy [7]. This sensing noise imposes physical limits on the precision of chemotaxis [7–13]. Experimental work suggests that biological cells indeed operate at these limits in shallow concentration gradients [14–18]. Temporal averaging over extended measurement intervals is a common strategy to reduce sensing noise [7, 19–21]. Yet, temporal averaging may be limited in time-varying environments [18, 22–25], or more directly by the stochastic motion of chemotactic agents themselves [26]. How should a chemotactic agent integrate previous and more recent measurements to most accurately estimate the relative direction of a concentration gradient if this direction changes stochastically in time?

Classical work by Berg and Purcell showed that the relative accuracy $(\langle \delta c / c \rangle^2)$ for sensing absolute concentration $c$ in the presence of measurement noise scales as $1/T$ with total measurement time $T$ [7]. Intriguingly, Mora et al recently showed that in time-varying environments, this accuracy scales instead as $(\tau r)^{-1/2}$, where $\tau$ is the time-scale on which the concentration changes, and $r$ is a rate of information gain [25]. Correspondingly, the optimal effective measurement time for temporal averaging reads $T_{eff} = (\tau/r)^{1/2}$ [25]. To the best of our knowledge, it is not known if similar statements hold also for the nonlinear problem of estimating the direction of a concentration gradient. Here, we develop a theory of optimal estimation of gradient direction in the presence of rotational diffusion with rotational diffusion coefficient $D$, and hence rotational diffusion time $\tau = (2D)^{-1}$. We derive results that are analogous to Mora et al, provided $r$ is now interpreted as a signal-to-noise ratio of gradient sensing normalized by unit time.

Review of previous results. Previous work suggested that bacteria use Kalman filters to track a time-dependent concentration signal [27], providing an optimal weighting of past and recent
measurements, see also [26]. Yet, Kalman filters address linear problems [28], while sensing a direction is a nonlinear problem of circular statistics, which prompts Bayesian updating of an angular likelihood distribution at each time step [29]. Bayesian estimation had been successfully applied, e.g. for monitoring time-varying environments with two states [30], estimating absolute concentration [25, 31–33], or temporal changes thereof [24, 34]. In contrast, previous theories of Bayesian sensing of direction [11, 12, 15, 35], or target position [36], did not account for motility noise as we do. In the engineering literature, estimation algorithms known as ‘bearing tracking’ do account for motility noise [37, 38], yet lack an analytical theory on the theoretical limits of gradient sensing.

Structure of manuscript. Our manuscript is structured as follows: we first introduce our minimal model of gradient sensing and formulate sequential Bayesian estimation, consisting of alternating update and prediction steps for a likelihood distribution of relative gradient direction. We use a suitably discretized time, and the calculus of von-Mises distributions; where von-Mises distributions can be considered the analogue of the normal distribution for direction angles. As our main result, we derive an analytic result for the ultimate precision of gradient estimation, see equation (19). We show that the estimated precision of an individual agent faithfully reflects the true accuracy of estimation errors in an ensemble of identical agents, as expected for optimal Bayesian estimation. Estimation becomes sub-optimal if the agents do not know their rotational diffusion coefficient exactly. If the agents, however, do not account for their rotational diffusion at all, gradient sensing would fail in a characteristic manner. We emphasize that our minimal model is representative of a larger class of problems, e.g. agents exploring time-dependent concentration gradients that likewise change stochastically on a characteristic time-scale $\tau$.

2. Minimal model of spatial gradient sensing

We consider a chemotactic agent that seeks to estimate the direction of an external concentration gradient by detecting binding events of signaling molecules that diffuse to its surface.

As a minimal model, we consider a stationary agent that is positioned at the origin of a two-dimensional plane. The agent is disklike with radius $a$. To leading order, the external concentration gradient will be linear

$$c(r) = c_0 + \nabla c \cdot r = c_0 (1 + \alpha \mathbf{e} \cdot r/a).$$

(1)

Here, $c_0$ denotes a base concentration, $\alpha = |\nabla c|a/c_0$ a normalized strength of the concentration gradient, and $\mathbf{e} = \nabla c/|\nabla c|$ is the unit vector pointing in the direction of the gradient. We introduce a material frame of the agent with orthonormal unit vectors $\mathbf{h}_1$ and $\mathbf{h}_2$, see figure 1(A). The agent seeks to estimate the direction angle $\psi^*$ of the gradient relative to the material frame, such that $\mathbf{e} = \cos \psi^* \mathbf{h}_1 + \sin \psi^* \mathbf{h}_2$. The circumference $r(\phi)$ of the agent is parameterized by $r(\phi) = a \cos \phi \mathbf{h}_1 + a \sin \phi \mathbf{h}_2$. The rate $\Lambda(\phi)$ at which molecules are detected at the circumference of the agent at position $r(\phi)$ is proportional to the local concentration

$$\Lambda(\phi) = (2\pi)^{-1} \lambda c[r(\phi)],$$

(2)

where $\lambda$ is a proportionality constant that depends on the diffusion constant of the signaling molecules.

The sequence of binding events at respective times $\theta_j$ and polar direction angles $\varphi_j$ can be conveniently encoded in a complex signal

$$s(t) = \sum_j \delta(t - \theta_j) e^{i \varphi_j}.$$  

(3)

By definition, the expectation value of this signal equals the Fourier transform of the angle-dependent binding rate, $\langle s(t) \rangle = \int d\varphi \Lambda(\varphi) \exp(i \varphi)$. Generally, it is much harder to estimate a gradient direction than absolute concentration $c_0$. Correspondingly, we focus on the problem of estimating the relative gradient direction angle $\psi^*$ and assume that the agent a priori knows $c_0$ as well as gradient strength $\alpha$.

We introduce the signal-to-noise ratio of gradient sensing for a given time interval $\Delta t$, following [39]

$$\text{SNR} = \alpha^2 \lambda c_0 \Delta t/2.$$  

(4)

This signal-to-noise ratio scales linearly with $\Delta t$ and defines an information rate with units of an inverse time

$$r = \frac{\text{SNR}}{\Delta t} = \alpha^2 \lambda c_0/2.$$  

(5)

The information rate defines a characteristic time scale of the problem as $1/r$. On times shorter than $1/r$, the agent cannot gather much useful information on the gradient direction.

We now turn to a time-dependent problem and consider rotational diffusion of the agent with rotational diffusion coefficient $D = D_{\text{rot}}$. The rotational diffusion coefficient defines a second time-scale $\tau = 1/(2D)$. 


Correspondingly, the relative gradient direction angle sensing. In the following, we consider the limit of weak rotational diffusion, for $t/W$ we will show that these two time-scales, $1/c$ characterizes the time-scale of convergence of the estimated precision of gradient sensing to its limit value (and, in fact, represent an optimal effective measurement time). Thus, the second inequality in equation (6)

We will show that these two time-scales, $1/r$ and $\tau$, together determine the ultimate precision of gradient sensing. In the following, we consider the limit of weak rotational diffusion, $D \ll r$, which is the biologically relevant case. In the opposite limit, $D \gg r$, the relative gradient direction would change so fast that the agent could simply not track it.

To make analytical progress, we assume a discrete time dynamics with equidistant time points $t_j = j \Delta t$. We choose the interval $\Delta t$ of time discretization such that

$$\frac{1}{\tau} \ll \Delta t \ll T_{\text{eff}},$$

where

$$T_{\text{eff}} = \sqrt{\tau}/r.$$  

Here, we introduced a time-scale $T_{\text{eff}}$ as the geometric mean of $1/r$ and $\tau$; the time-scale $T_{\text{eff}}$ will later characterize the time-scale of convergence of the estimated precision of gradient sensing to its limit value (and, in fact, represent an optimal effective measurement time). Thus, the second inequality in equation (6) ensures that the time discretization is finer than the dynamics of the problem. Our choice of $\Delta t$ formally corresponds to a large signal-to-noise ratio $\text{SNR} = r \Delta t \gg 1$ for a time interval of duration $\Delta t$. At the same time, our choice ensures that the effect of rotational diffusion is small compared to sensing noise during a time interval of duration $\Delta t$, i.e. $D \Delta t \ll 1/\text{SNR}$. Note that for a time step that is too coarse, alternating update and prediction steps would cause the estimated precision to oscillate; a small time step ensures that the amplitude of these oscillations is small.

At the times $t_j$, the agent shall undergo discrete rotational diffusion events, see figure 1(B). Correspondingly, the relative gradient direction angle $\psi^r$ will be constant in each time interval $\psi^r(t) = \psi^r_j$ for $t \in T_j$ with $T_j = (t_j, t_{j+1})$, but change by a discrete amount $\Delta_j = \psi^r_{j+1} - \psi^r_j$ at $t_j$. The stochastic increments $\Delta_j$ are uncorrelated and satisfy $\langle \Delta_j \Delta_k \rangle = 2D \Delta t \delta_{jk}$. While estimating the direction of the concentration gradient, the agent will account for its rotational diffusion, assuming that its rotational diffusion coefficient is $D$.

![Diagram](image)

**Figure 1.** Chemotactic agent subject to rotational diffusion. (A) In our minimal model, a chemotactic agent seeks to estimate the direction of an external concentration gradient $\nabla c$ of signaling molecules (relative to its material frame $h_1$ and $h_2$) by detecting binding events of signaling molecules on its circumference, corresponding to a strategy of spatial comparison. At discrete times $t_i$, the agent undergoes rotational diffusion with rotational diffusion coefficient $D$. Between these times, the agent integrates the position of binding events for a time interval $T_i$ of duration $\Delta t$ into a single, complex measurement $M_i$, see equation (8). This measurement represents a noisy measurement $\psi_i = \arg M_i$ of the true gradient direction angle $\psi^r(t)$ enclosed between gradient direction $\nabla c$ and material frame vector $h_i$. (B) Due to rotational diffusion, the true gradient direction angle $\psi^r(t)$ becomes a random walk with constant $\psi^r(t) = \psi^r_j$ for $t \in T_j$ and stochastic increments $\Delta_j$ at $t_j$. This motility noise limits the precision by which the agent can estimate the relative gradient direction angle $\psi$ from the measurements $M_1, \ldots, M_m$. (C) The agent performs sequential Bayesian estimation to estimate $\psi(t)$: The agent computes a likelihood distribution $\mathcal{L}^r(\psi)$ of possible gradient directions, iteratively executing a prediction step that accounts for its rotational diffusion (which flattens the distribution), and an update step that incorporates a new measurement $M_m$ (which usually sharpens the distribution).
During the time intervals $T_j = (t_{j-1}, t_j)$, the agent does not rotate and integrates the stochastic signal $s(t)$ for a time span $\Delta t$

$$\mathcal{M}_j = \int_{t_{j-1}}^{t_j} dt \, s(t).$$

(8)

The angle $\psi_j = \arg \mathcal{M}_j$ of this complex number $\mathcal{M}_j$ represents a noisy measurement of the true current gradient direction angle $\psi_j^\ast$.

3. Bayesian gradient sensing

3.1. Statistics of single measurements

What is the statistics of the measurements $\mathcal{M}_j$? In total, the agent will detect on average $\nu = \int d\varphi \, \lambda(\varphi) \Delta t = \lambda_0 \, \Delta t$ molecules during a time interval of duration $\Delta t$. A short calculation yields for the mean and variance of $\mathcal{M}_j$

$$\langle \mathcal{M}_j \rangle = \frac{\alpha \nu}{2} e^{\psi_j^\ast}$$

(9)

$$\langle |\mathcal{M}_j|^2 \rangle - |\langle \mathcal{M}_j \rangle|^2 = \nu.$$  

(10)

For this calculation, we considered all binding events in the time interval $T_j$, enumerated by an index $k$ with respective angles $\varphi_k$, where the total number $\nu_j$ of these events is a Poisson random variable with mean $\nu$ and variance $\nu$. We then find $\langle \mathcal{M}_j \rangle = \langle \sum_k \exp i\varphi_k \rangle = \int d\varphi \, \lambda(\varphi) \exp(i\varphi) \Delta t$, and $\langle |\mathcal{M}_j|^2 \rangle = \langle \sum_k \exp i(\varphi_k - \varphi) \rangle + \langle \sum_{k \neq j} \exp i(\varphi_k - \varphi_j) \rangle = (\alpha/2)^2 \langle \nu(\nu_j - 1) \rangle + \langle \nu_j \rangle = (\alpha/2)^2 + \nu$, where we used $\langle \Delta t/\nu \rangle^2 \int d\varphi \, d\varphi' \lambda(\varphi) \lambda(\varphi') \exp i(\varphi - \varphi') = (\alpha/2)^2$. Similarly, $\langle \mathcal{M}_j^2 \rangle = \langle \mathcal{M}_j \rangle^2$, which implies that fluctuations of $\mathcal{M}_j$ in the complex plane are isotropic [40].

Our choice of $\Delta t$ in equation (6) implies in particular $\nu \gg 1$. In this limit of large molecule counts, we can approximate the probability distribution of the measurements $\mathcal{M}_j$ by a normal distribution with the same mean $\mu = \langle \mathcal{M}_j \rangle$ and the same variance $\Sigma = \nu/2$ of the real and imaginary part of $\mathcal{M}_j$ (diffusion approximation)

$$p(\mathcal{M}_j|\psi_j^\ast) = (2\pi \Sigma)^{-1} \exp \left( -\frac{|\mathcal{M}_j - \mu|^2}{\nu} \right)$$

$$\sim \exp \left( \frac{\alpha |\mathcal{M}_j|}{\nu} \cos(\psi_j - \psi_j^\ast) \right).$$

(11)

In the last step, we simply used the Pythagorean trigonometric identity. The factor of proportionality depends only on $|\mathcal{M}_j|$. If we interpret the last expression as a function of $\psi_j^\ast$, it provides a first example of a von-Mises distribution with so-called measure-of-concentration $\zeta_j$. von-Mises distributions can be regarded as a generalization of the normal distribution for angular variables and are further discussed in section 3.2 and appendix C. The expectation value $\langle \zeta_j \rangle$ of the measure-of-concentration $\zeta_j$ for a single measurement is just the signal-to-noise ratio defined in equation (4)

$$\langle \zeta_j \rangle \approx \text{SNR}.$$  

(12)

A cursory explanation for equation (12) reads $\langle \zeta_j \rangle = \alpha (|\mathcal{M}_j|) \frac{|\psi_j^\ast|}{\alpha} \lambda_0 |\mathcal{M}_j| = \alpha^2 \nu/2 = \text{SNR}$, where we used equation (6) in step (*); a more rigorous derivation can be found in appendix B.

3.2. Estimation of directions using von-Mises distributions

The agent uses the sequence of measurements $\mathcal{M}_j, j = 1, \ldots, m$, to compute a likelihood distribution $\mathcal{L}_m(\psi)$ for the unknown gradient direction angle $\psi_m^\ast$ at time $t_m$. In each time step, the agent performs a prediction step (accounting for rotational diffusion), and an update step (accounting for the new measurement $\mathcal{M}_m$), see figure 1(C). For the prediction step, the agents starts with a prior $\mathcal{L}_{m-1}(\psi)$ that incorporates all previous measurements, $\mathcal{M}_j$ with $j \leq m - 1$, and computes from it a modified likelihood distribution $\mathcal{L}_{m-1}^{\text{mod}}(\psi)$ that accounts for the loss of knowledge due to the rotational diffusion event at $t_{m-1}$. In the update step, the agent computes the new likelihood distribution $\mathcal{L}_m(\psi)$ using Bayes’ rule, incorporating the new measurement $\mathcal{M}_m$ with the predicted likelihood $\mathcal{L}_{m-1}^{\text{mod}}(\psi)$ as prior.

A fundamental difficulty of this estimating problem is that it concerns an angular variable, i.e. all likelihood distributions must be $2\pi$-periodic. This sets our problem apart from other problems previously considered in the literature [25, 30].
We will approximate the different likelihood distributions $\mathcal{L}(\psi)$ by von-Mises distributions
\begin{equation}
\mathcal{V}_\mathcal{M}(\psi; \mu, \kappa) \sim \exp[\kappa \cos(\psi - \mu)],
\end{equation}
where $\mu$ denotes the mean and $\kappa$ is the so-called measure-of-concentration. The von-Mises distribution can be regarded as a generalization of the normal distribution for angular variables, where $\kappa$ plays a role similar to an inverse variance. The von-Mises distribution is the maximum-entropy distribution $p(\psi)$ that maximizes information entropy $\int d\psi \, p(\psi) \ln p(\psi)$ if the first complex moment $\langle \exp(i\psi) \rangle$ is specified. Importantly, the product of two von-Mises distributions is again a von-Mises distribution (up to normalization), $\mathcal{V}_\mathcal{M}(\psi; \mu_1, \kappa_1) \mathcal{V}_\mathcal{M}(\psi; \mu_2, \kappa_2) \approx \mathcal{V}_\mathcal{M}(\psi; \mu, \kappa)$. Here, the new mean $\mu$ and new measure-of-concentration $\kappa$ satisfy $\kappa \exp(i\psi_1) = \kappa_1 \exp(i\psi_1) + \kappa_2 \exp(i\psi_2)$, i.e. the new mean is a weighted average of the old mean values. If the old mean values are close, we have $\kappa \approx \kappa_1 + \kappa_2$, which generalizes an analogous, exact statement for normal distributions. Likewise, $\mu \approx (\kappa_1/\kappa) \mu_1 + (\kappa_2/\kappa) \mu_2$.

The convolution of two von-Mises distributions is only approximately a von-Mises distribution, $\mathcal{V}_\mathcal{M}(\psi; \mu_1, \kappa_1) \ast \mathcal{V}_\mathcal{M}(\psi; \mu_2, \kappa_2) \approx \mathcal{V}_\mathcal{M}(\psi; \mu, \kappa)$, where the mean values of these von-Mises distributions represent the current maximum-likelihood estimates $\hat{\psi}_1$ of the gradient direction angle. The convolution of two von-Mises distributions is only approximately a von-Mises distribution, $\mathcal{V}_\mathcal{M}(\psi; \mu_1, \kappa_1) \ast \mathcal{V}_\mathcal{M}(\psi; \mu_2, \kappa_2) \approx \mathcal{V}_\mathcal{M}(\psi; \mu, \kappa)$, where the mean values of these von-Mises distributions represent the current maximum-likelihood estimates $\hat{\psi}_1$ of the gradient direction angle.

The agent accounts for this stochastic dynamics of the relative direction angle by time step $t_m$, the agent updates its prior according to Bayes’ rule
\begin{equation}
\mathcal{L}_{m-1}(\psi) = \mathcal{V}_\mathcal{M}(\psi; \hat{\psi}_{m-1}, \kappa'_{m-1}).
\end{equation}

In light of the new measurement $\mathcal{M}_m$, the agent updates this prior according to Bayes’ rule
\begin{equation}
\mathcal{L}_m(\psi) = \frac{P(\mathcal{M}_m | \psi) \mathcal{L}_{m-1}(\psi)}{P(\mathcal{M}_m | \mathcal{M}_{1:m-1})}.
\end{equation}

Here, $P(\mathcal{M}_m | \psi)$ is the probability to obtain the specific measurement $\mathcal{M}_m$ if the true gradient angle were $\psi$, i.e. the measurement model. Similarly, the normalization factor $P(\mathcal{M}_m | \mathcal{M}_{1:m-1})$ is the probability to obtain this specific measurement $\mathcal{M}_m$ if the true gradient direction were distributed according to the Bayesian prior $\mathcal{L}_{m-1}(\psi)$; explicitly, $P(\mathcal{M}_m | \mathcal{M}_{1:m-1}) = \int d\psi \, P(\mathcal{M}_m | \psi) \mathcal{L}_{m-1}(\psi)$. The right-hand side of equation (15) represents a product of von-Mises distributions, which yields again a von-Mises distribution, whose measure-of-concentration shall be called $\kappa_m$. Our choice of time step $\Delta t$ in equation (6) implies that both the mean $\mu_1 = \hat{\psi}_m = \arg\mathcal{M}_m$ of the first factor $P(\mathcal{M}_m | \psi)$ and the mean $\mu_2 = \hat{\psi}_{m-1}$ of the second factor $\mathcal{L}_{m-1}(\psi)$ are close to the true angle $\psi$, provided true and assumed rotational diffusion coefficient are equal, $D = \tilde{D}$, or at least of similar magnitude, see below. Consequently, also $|\mu_1 - \mu_2| \ll 1$. The calculus of von-Mises distributions thus yields the iteration rule $\kappa_m \approx \kappa'_{m-1} + \zeta_m$, or, using equation (12)
\begin{equation}
\kappa_m \approx \kappa'_{m-1} + \text{SNR}.
\end{equation}

Fluctuations of $\zeta_m$ cause small fluctuations of $\kappa_m$ not considered here.

### 3.4. Prediction step: rotational diffusion
For rotational diffusion, the stochastic increments $\Delta t_{\text{rot}}$ of the relative direction angle are distributed according to a wrapped normal distribution with variance parameter $2\tilde{D}\Delta t$. In the limit of weak rotational diffusion considered here, this distribution is well approximated by a von-Mises distribution with zero mean and measure-of-concentration $(2\tilde{D}\Delta t)^{-1}$. The agent accounts for this stochastic dynamics of the relative direction angle in a prediction step, assuming its rotational diffusion coefficient to be $\tilde{D}$ (which may or may not be equal to the true value $D$). The prediction step thus corresponds to a convolution of two von-Mises distributions
\begin{equation}
\mathcal{L}_m' \approx \mathcal{V}_\mathcal{M}(\Delta; 0, (2\tilde{D}\Delta t)^{-1}).
\end{equation}

Equation (17) is a special case of a Chapman–Kolmogorov equation, $\mathcal{L}_m' = \mathcal{L}_m'(\psi' | \mathcal{M}_{1:m}) = \int d\psi \, P(\psi' | \psi) \mathcal{L}(\psi | \mathcal{M}_{1:m}) d\psi$ with assumed transition probability $P(\psi' | \psi)$ from $\psi$ to $\psi'$ given by $P(\psi' | \psi) \approx \mathcal{V}_\mathcal{M}(\psi'; \psi, (2\tilde{D}\Delta t)^{-1})$. 

\[\]
By the calculus of von-Mises distributions, we obtain again an iteration rule, now for the measure-of-concentration \( \kappa'_m \) of the likelihood distribution after the prediction step

\[
\text{prediction rule: } \frac{1}{\kappa'_m} \approx \frac{1}{\kappa_m} + 2\tilde{D}\Delta t. \tag{18}
\]

The prediction step does not change the maximum-likelihood angle, i.e. \( \hat{\psi}'_m = \hat{\psi}_m \), since the diffusion kernel is centered at zero.

### 3.5. Theoretical limit of gradient sensing in the presence of rotational diffusion

Together, the nested iteration rules equations (16) and (18) define a monotonically increasing sequence of measures-of-concentration \( \kappa_m \), which converges to a limit value \( \kappa_\infty \) for \( D = \tilde{D} \). The measure-of-concentration \( \kappa_\infty \) represents a theoretical limit for the estimated precision of gradient sensing at steady-state. We compute \( \kappa_\infty \) as a fixed point of the nested iteration rules equations (16) and (18).

Specifically, the limit values \( \kappa_\infty = \lim_{m \to \infty} \kappa_m \) and \( \kappa'_\infty = \lim_{m \to \infty} \kappa'_m \) satisfy \( \kappa_\infty = \kappa'_\infty + \text{SNR} \) and \( \kappa'_\infty = (\kappa_\infty - \text{SNR})^{-1} + 2\tilde{D}\Delta t \) eliminating \( \kappa'_\infty \) yields \( \kappa_\infty = [((\kappa_\infty - \text{SNR})^{-1} + 2\tilde{D}\Delta t)^{-1} + \text{SNR} \), or \( (\kappa_\infty - \text{SNR})^{-1} = (\kappa_\infty)^{-1} + 2\tilde{D}\Delta t \). Multiplying with the common denominator and rearranging yields the quadratic equation \( \kappa_\infty (\kappa_\infty - \text{SNR}) = \text{SNR} / (2\tilde{D}\Delta t) \), hence \( \kappa_\infty \approx [\text{SNR} / (2\tilde{D}\Delta t)]^{1/2} \) since \( \kappa_\infty \gg \text{SNR} \) by equation (6). We thus have

\[
\text{Case } D = \tilde{D}: \kappa_\infty = \sqrt{\frac{\tau}{2\tilde{D}}} = \sqrt{\frac{\lambda}{4e_0\tilde{D}}} |\nabla e|.
\]

This result for \( \kappa_\infty \) highlights the competition between information gain with rate \( \tau = \text{SNR} / \Delta t \) and information loss by rotational diffusion with rotational diffusion coefficient \( \tilde{D} \). We can rewrite equation (19) as \( \kappa_\infty = \tau T_{\text{eff}}, \) where we introduced the effective measurement time \( T_{\text{eff}} = (\tau / \tilde{D})^{1/2} \).

This theoretical limit is strict and faithful, i.e. it reflects the distribution of estimation errors in an ensemble of identical agents, as we show next.

### 3.6. Distribution of estimation errors

The mean \( \hat{\psi}_m \) of \( \mathcal{L}_m(\psi) \) represents the maximum-likelihood estimate of the agent at time \( t_m \). By the update step, equation (15), which expresses \( \mathcal{L}_m(\psi) \) as a product of von-Mises distributions [and the rules for products of von-Mises distributions, see appendix C], \( \hat{\psi}_m \) is a weighted circular average of the previous estimate \( \hat{\psi}_{m-1} \) and the current measurement \( \check{\psi}_m \)

\[
\kappa_m e^{\hat{\psi}_m} = \kappa'_{m-1} e^{\hat{\psi}_{m-1}} + \text{SNR} e^{\check{\psi}_m}. \tag{20}
\]

We consider the distribution \( p(\epsilon_m) \) of estimation errors in an ensemble of agents with \( \epsilon_m = \hat{\psi}_m - \check{\psi}_m \). We are interested in the true accuracy of estimation, defined as the circular variance of this distribution, \( \text{CV}[p(\epsilon_m)] = 1 - \langle |\epsilon_m| \rangle \). We have \( \text{CV}[p(\epsilon_m)] = 1 - \langle \epsilon_m \rangle \approx \langle \epsilon_m^2 \rangle / 2 \). By multiplying equation (20) with \( \exp(-\check{\psi}_m) \), and linearizing the imaginary part, we obtain an iteration rule for the estimation error \( \epsilon_m \) as an affine weighted sum

\[
\epsilon_m \approx \kappa'_m \kappa_m^{-1} (\hat{\psi}_m - \check{\psi}_m) + \text{SNR} \kappa_m \langle \hat{\psi}_m - \check{\psi}_m \rangle. \tag{21}
\]

Recall that \( \Delta_m = \hat{\psi}_m - \check{\psi}_m \) is the random change of the true gradient direction angle \( \psi^*(t) \) at time \( t_{m-1} \).

Case \( D = \tilde{D} \). We first assume that the agent knows the exact value of its own rotational diffusion coefficient, i.e. \( D = \tilde{D} \). Equation (21) represents a linear superposition of two independent random angles. In the limit of sufficiently sharp distributions, valid by equation (6), the circular variance of the superposition is a weighted sum of the circular variances of these random angles, with coefficients squared, analogous to the case of normal distributions. We thus conclude from equation (21) an iteration rule for the true accuracy

\[
\text{CV}[p(\epsilon_m)] \approx \left( \frac{\kappa'_m \kappa_m^{-1}}{\kappa_m} \right)^2 \left( \text{CV}[p(\epsilon_{m-1})] + D\Delta t \right) + \left( \frac{\text{SNR}}{\kappa_m} \right)^2 \text{CV}[p(\hat{\psi}_m - \check{\psi}_m)]. \tag{22}
\]

Note that \( \text{CV}[p(\hat{\psi}_m - \check{\psi}_m)] = (2\text{SNR})^{-1} \) is just the sensing noise of a single measurement. For agents that are aware of their rotational diffusion with \( D = \tilde{D} \), this iteration rule for the true accuracy \( \text{CV}[p(\epsilon_m)] \) in an
ensemble of identical agents, equation (22), is equivalent to the combination of update and prediction rule for the estimated precision \( \kappa_m \) of an individual agent, equations (16) and (18). Indeed, if \( CV[p(\varepsilon_m)] \approx 1/(2\kappa_m) \) at \( t = t_{m-1} \), then \( CV[p(\varepsilon_m)] \approx 1/(2\kappa_m) \) at \( t = t_m \) by a short calculation. Specifically, from the initial assumption it follows that \( CV[p(\varepsilon_m-1)] + D\Delta t \approx 1/(2\kappa_m) \), hence the right-hand side of equation (22) becomes 
\[
(\kappa_m + \kappa_m^{-1} + \text{SNR})^2 (2\kappa_m^{-1})^{-1} + \text{SNR}/(\kappa_m^{-1} + \text{SNR})^2 
\]
\( (2\text{SNR})^{-1} = (\kappa_m^{-1} + \text{SNR})/[2(\kappa_m^{-1} + \text{SNR})^2] = 1/[2(\kappa_m^{-1} + \text{SNR})] = 1/(2\kappa_m) \). This confirms that Bayesian estimation of gradient direction is faithful, i.e. the estimated precision of an individual agent follows, on average, the true accuracy in an ensemble of agents, as expected for optimal Bayesian estimation. As a corollary, the true accuracy converges to the same limit value as the (circular variance corresponding to the) estimated precision
\[
\lim_{m \to \infty} CV[p(\varepsilon_m)] \approx (2\kappa_m)^{-1}. \tag{23}
\]
Figure 2(A) shows simulation results for a case where estimated precision and true accuracy are initially different, but quickly converge to a common limit value. For details on numerical methods, see appendix A.

Case \( \hat{D} \neq D \). More generally, we can consider agents that assume a value \( \hat{D} \) for their rotational diffusion coefficient when performing the prediction step equation (18), while the true rotational diffusion coefficient is \( D \). In this case, the estimated precision \( \kappa_m \) converges to \( \kappa_\infty \approx \text{SNR}/(2\hat{D} \Delta t)^{1/2} \), i.e. equation (19) with \( D \) replaced by \( \hat{D} \), while the true accuracy \( CV[p(\varepsilon_m)] \) converges to a limit value \( CV_\infty \) different from \( 1/(2\kappa_\infty) \).

To compute the true accuracy \( CV_\infty \) in the long term limit for the case \( \hat{D} \neq D \), \( CV_\infty = \lim_{m \to \infty} CV[p(\varepsilon_m)] \), we note that the iteration rule for \( CV[p(\varepsilon_m)] \), equation (22), still holds. Taking the limit \( m \to \infty \) yields a
new self-consistency equation for $CV_\infty$. A short calculation yields the unique solution, see appendix D

\[ CV_\infty \approx \frac{1}{2} \sqrt{\frac{2 D \Delta t}{\text{SNR}}} \cdot \frac{D + \hat{D}}{2 \sqrt{D \hat{D}}}. \]  

This expression attains the minimal value $CV_\infty = \left[D \Delta t/(2\text{SNR})\right]^{1/2}$ exactly for $\hat{D} = D$. Equation (24) compares favorably to simulation results, see figure 3. For details on numerical methods, see appendix A.

We conclude that Bayesian gradient estimation can still operate close to its optimum even if the agent has only inaccurate knowledge of its true rotational diffusion coefficient, provided the true and the assumed rotational diffusion coefficient, $D$ and $\hat{D}$, are of similar magnitude.

However, the estimation of gradient direction fails if the agent would completely neglect its rotational diffusion, corresponding to $\hat{D} = 0$, as we show next.

**Agents unaware of their rotational diffusion:** $D > 0$, $\hat{D} = 0$. If agents were unaware of their own rotational diffusion with $\hat{D} = 0$, they would perform only the update step, equation (15). Correspondingly, the likelihood distribution $L_m(\psi)$ for the gradient direction angle at time $t_m$ becomes a product of $m + 1$ von-Mises distributions. The calculus of von-Mises distribution yields an expression for the measure-of-concentration $\kappa_m$ of this likelihood distribution as a circular average of all previous measurements

\[ \kappa_m e^{i\tilde{\psi}_m} = \kappa_0' e^{i\tilde{\psi}_0} + \sum_{j=1}^{m} \zeta_j e^{i\tilde{\psi}_j}. \]  

Here, the first term represents the contribution from the Bayesian prior, whereas the sum represents the contribution from the sequential measurements. For simplicity, we assume a flat prior with $\kappa_0' = 0$. We can rewrite equation (25) in terms of the complex measurement $M_j$ as $\kappa_m \exp(i\tilde{\psi}_m) = \alpha \sum_j M_j$. To evaluate this sum, we note the correlations between subsequent measurements, stemming from the rotational diffusion of the agent

\[ \langle M_j M_k \rangle = \left(\frac{\alpha'\nu'}{2}\right)^2 e^{-D|k-j|} \quad \text{for} \ j \neq k. \]  

We can now compute $\langle \kappa_m^2 \rangle$ in a straight-forward manner, noting that the term $\sum_{j \neq k} \langle M_j M_k \rangle$ yields a double-geometric series, see appendix E for details and the (lengthy) result. Using the approximation $\kappa_m \approx (\kappa_m^2)^{1/2}$, we obtain the asymptotic scaling

\[ \kappa_m \approx \begin{cases} rt_m \text{ for } t_m \ll \tau \\ 2r \sqrt{t_m/\tau} \text{ for } t_m \gg \tau \end{cases}. \]  

Hence, for $D > 0$ but $\hat{D} = 0$, the estimated precision $\kappa_m$ will initially increase linearly with $m$, but then cross-over to the asymptotic scaling $\kappa_m \sim m^{1/2}$ after a characteristic cross-over time on the order of $\tau = (2D)^{-1}$. Simultaneously, $CV[p(\epsilon_m)] \rightarrow 1$ for $t_m \gg \tau$. These simple scaling relations are corroborated by numerical simulations, see figure 2(B). For details on numerical methods, see appendix A.
Thus, agents not aware of their own rotational diffusion will arrive at erroneous gradient estimates, because past measurements will have become partially invalidated by rotational diffusion, yet are nonetheless incorporated by the agent in its gradient estimates with full weight. Concomitantly, the estimated precisions $\kappa_m$ that individual agents assign to their estimates of gradient direction do not reflect the true accuracy, i.e. the dispersion of estimation errors in an ensemble of agents. Individual agents are ‘over-confident’ of their own estimates.

Case $D = \tilde{D} = 0$. Finally, in the absence of rotational diffusion with $D = \tilde{D} = 0$, the estimated precision $\kappa_m$ grows linearly with time as $(\kappa_m) \approx mSNR = rt_m$, as expected. Correspondingly, the true accuracy of estimation errors in an ensemble of agents decreases inversely with total measurement time, $CV_m \approx 1/(2rt_m)$. This is Berg-and-Purcell’s classical result for the estimation of absolute concentration [7], generalized to the problem of estimating gradient direction.

4. Discussion

Summary of results. We considered a minimal model of gradient sensing in the presence of both sensing and motility noise. We analytically derived the theoretical limit of gradient sensing for a chemotactic agent that undergoes rotational diffusion. The chemotactic agent is placed in an external concentration gradient and detects stochastic binding events of signaling molecules on its circumference, while it undergoes rotational diffusion with rotational diffusion coefficient $D$.

We find that the accuracy by which this agent can estimate the time-varying relative direction of the gradient depends on two fundamental time-scales: the rotational diffusion time $\tau = 1/(2D)$ on which the relative gradient direction changes, and a second time-scale set by a rate of information gain $r = SNR/\Delta t$ defined as a signal-to-noise ratio of gradient sensing normalized by unit time. The error $\varepsilon$ between true and estimated gradient direction is distributed at steady-state with a circular variance $CV_m[p(\varepsilon)] \approx (rT_{\text{eff}})^{-1}$ that scales inversely with an effective measurement time $T_{\text{eff}} = \sqrt{r/\tau}$ given by the geometric mean of the two time-scales of the problem. Numerical simulations corroborate this scaling relation, see figure 4. For details on numerical methods, see appendix A. Intuitively, the agent averages noisy measurements only over the recent past defined by a time interval of duration $T_{\text{eff}}$. The effective measurement time $T_{\text{eff}}$ reflects a trade-off between the requirement to average noisy measurements over long time intervals to reduce sensing noise, and a second requirement to use only recent measurements because the relative gradients direction already changes during the measurement interval due to motility noise.

For the case of optimal estimation, we assume that the agent uses sequential Bayesian estimation and has perfect knowledge of its own rotational diffusion coefficient. In this case, the precision that individual agents assign to their own gradient estimates matches the accuracy of gradient estimates in an ensemble of identical agents, i.e. gradient estimation is faithful. If the agent assumes a rotational diffusion coefficient $\tilde{D}$ that is either too small or too large, this is no longer the case, and the circular variance $CV_m[p(\varepsilon)]$ of estimation errors will be larger than in the optimal case $D = \tilde{D}$. If, however, the agent would fully ignore its own rotational diffusion, corresponding to $\tilde{D} = 0$, gradient sensing will fail in a characteristic way: agents not aware of their rotational diffusion will extend their temporal averaging infinitely into the past. Concomitantly, the estimated gradient direction decorrelates from the true direction on the time-scale $\tau$ of rotational diffusion, while agents erroneously believe that their estimates become more and more precise as function of total measurement time $t$. We find an abnormal asymptotic scaling of the variance of gradient estimates, which decreases with time as $CV[L(\psi)] \sim t^{-1/2}$ (estimated precision), while the errors between true and estimated gradient direction converges to one $CV[p(\varepsilon)] \rightarrow 1$ (true accuracy). The abnormal scaling of the estimated precision represents a signature of erroneous state estimation, and could be tested in real-world applications, e.g. of bearing tracking.

Cause of the square-root scaling of optimal measurement time. Our theory shows that the optimal measurement time for the estimation of a time-varying gradient direction scales as $T_{\text{eff}} = (\tau/\tau)^{1/2}$, while the variance of the expected estimation error scales as $1/(rT_{\text{eff}}) = (\tau r)^{-1/2}$. This result parallels a recent result by Mora et al on the optimal estimation of an absolute concentration that varies in time [25]. In both Mora et al and our work, $\tau$ and $r$ denote a characteristic fluctuation time-scale, and a rate of information gain of the problem, respectively.

We propose that the inverse square-root scaling of the optimal measurement time could be a general feature of Bayesian estimation of time-varying signals. The following heuristic argument highlights the common idea: if an agent averages past measurements of a time-varying signal $s(t)$ with an effective measurement time $T_{\text{eff}}$, this agent effectively estimates $\hat{s}(t) = T_{\text{eff}}^{-1} \int_{t-T_{\text{eff}}}^{t} dt' s(t') \approx s(t - T_{\text{eff}}/2)$, i.e. the estimate lags behind the current value of the signal. The variance $\sigma^2$ of the expected estimation error $\varepsilon(t) = \hat{s}(t) - s(t)$ between the estimated value $\hat{s}(t)$ of the signal and its true value $s(t)$ can be
span set by the rotational diffusion time, i.e. motility in the presence of sensing and motility noise. This filter extended temporal comparison over a time bacterial chemotaxis, i.e. the optimal sensori-motor transfer function for bacteria with run-and-tumble effective measurement time for estimation of time-varying signals are different from previous work [26].

Previously, Strong approximately decomposed into two contributions: $\sigma^2_{\text{sensing}} = \langle (\hat{s} - \overline{s})^2 \rangle$ and $\sigma^2_{\text{time}} = \langle (\hat{s} - \overline{s})^2 \rangle$. The first contribution predominantly reflects sensing noise, and approximately scales as $\sigma^2_{\text{sensing}} \sim 1/(\tau T_{\text{eff}})$ as function of effective measurement time $T_{\text{eff}}$. The second contribution accounts for the stochastic change of the signal during the time $T_{\text{eff}}$ and approximately scales as $\sigma^2_{\text{time}} \sim T_{\text{eff}}/\tau$. While the first contribution decreases with $T_{\text{eff}}$, the second summand increases, thus posing an optimization problem for the optimal $T_{\text{eff}}$ that minimizes the total variance of the error $\sigma \approx \sigma^2_{\text{sensing}} + \sigma^2_{\text{time}}$. Solving for $d\sigma^2/dT_{\text{eff}} = 0$ yields $T_{\text{eff}} \sim (\tau/\tau)^{1/2}$. A detailed version of this argument can be found in appendix E.

**Difference to bacterial chemotaxis.** We emphasize that our and Mora’s recent results on the optimal effective measurement time for estimation of time-varying signals are different from previous work [26].

Previously, Strong et al extended the seminal work by Berg and Purcell [7] to compute the optimal filter for bacterial chemotaxis, i.e. the optimal sensori-motor transfer function for bacteria with run-and-tumble motility in the presence of sensing and motility noise. This filter extended temporal comparison over a time span set by the rotational diffusion time, i.e. $T_{\text{eff}} \sim D^{-1}$, (see equation (73) in loc. cit.), and additionally had an extremely long tail (see equation (70) in loc. cit.). This difference may be related to the fact that bacteria like *E. coli* use a chemotaxis strategy based on temporal comparison, which is different from the strategy of spatial comparison considered here; these bacteria estimate a time derivative of a concentration signal traced along their swimming path [19]. In contrast, most eukaryotic cells use spatial comparison, i.e. estimate spatial concentration gradients across the cell diameter [4, 5, 41–43], as considered here. A third chemotaxis strategy represented by marine sperm cells navigating along helical paths [3], can be mapped on the case of spatial comparison: while moving along helical paths, these cells ‘visit’ different sensor positions during one helical turn.

---

**Figure 4.** Square-root scaling of optimal gradient sensing. To test the analytical result for the estimated precision at steady state, equation (19), we simulated an ensemble of agents and determined an average precision $\kappa_{\infty}$, numerically for different signal-to-noise ratios SNR by varying base concentration (A) and relative gradient strength $\alpha$ (B). Shown are mean estimated precision of individual agents (red), mean true accuracy of estimation errors in an ensemble of agents (blue), analytical result equation (19) (gray dashed). Note that the assumption equation (6) concerning the time step $\Delta t$ of time discretization is not fulfilled for the large SNR considered; correspondingly the estimated precision after update and prediction step, $\kappa_{\text{eff}}$, and $\alpha_{\text{eff}}$, respectively, oscillate around a limit value $\kappa_{\infty} = \lim_{\nu \to \infty} \sqrt{\kappa_{\text{eff}}\nu}$, which is therefore shown in the figure. Parameters: different symbols correspond to different signal-to-noise ratios: For panel A, $\text{SNR} = 0.1, 0.2, 0.5, 1, 2.5, 5, 10$ (circle, square, diamond, left triangle, right triangle, up triangle, asterisk), obtained by varying the expected number of binding events per time step $\nu$ with $\alpha = 0.03$; for panel B, $\text{SNR} = 0.25, 2.25, 25$ (circle, square, diamond), obtained by varying the relative gradient strength $\alpha = 0.01, 0.03, 0.1$ with $\nu = 5000$. For each SNR, we tested different values of the rotational diffusion coefficient with $D \Delta t = 0.01, 0.025, 0.05, 0.1, 0.25, 0.5$ and $D = D$ (symbols connected by thin line, direction of increasing $D$ indicated by arrow).

Approximately decomposed into two contributions: $\sigma^2_{\text{sensing}} = \langle (\hat{s} - \overline{s})^2 \rangle$ and $\sigma^2_{\text{time}} = \langle (\hat{s} - \overline{s})^2 \rangle$. The first contribution predominantly reflects sensing noise, and approximately scales as $\sigma^2_{\text{sensing}} \sim 1/(\tau T_{\text{eff}})$ as function of effective measurement time $T_{\text{eff}}$. The second contribution accounts for the stochastic change of the signal during the time $T_{\text{eff}}$ and approximately scales as $\sigma^2_{\text{time}} \sim T_{\text{eff}}/\tau$. While the first contribution decreases with $T_{\text{eff}}$, the second summand increases, thus posing an optimization problem for the optimal $T_{\text{eff}}$ that minimizes the total variance of the error $\sigma \approx \sigma^2_{\text{sensing}} + \sigma^2_{\text{time}}$. Solving for $d\sigma^2/dT_{\text{eff}} = 0$ yields $T_{\text{eff}} \sim (\tau/\tau)^{1/2}$. A detailed version of this argument can be found in appendix E.

**Difference to bacterial chemotaxis.** We emphasize that our and Mora’s recent results on the optimal effective measurement time for estimation of time-varying signals are different from previous work [26].

Previously, Strong et al extended the seminal work by Berg and Purcell [7] to compute the optimal filter for bacterial chemotaxis, i.e. the optimal sensori-motor transfer function for bacteria with run-and-tumble motility in the presence of sensing and motility noise. This filter extended temporal comparison over a time span set by the rotational diffusion time, i.e. $T_{\text{eff}} \sim D^{-1}$, (see equation (73) in loc. cit.), and additionally had an extremely long tail (see equation (70) in loc. cit.). This difference may be related to the fact that bacteria like *E. coli* use a chemotaxis strategy based on temporal comparison, which is different from the strategy of spatial comparison considered here; these bacteria estimate a time derivative of a concentration signal traced along their swimming path [19]. In contrast, most eukaryotic cells use spatial comparison, i.e. estimate spatial concentration gradients across the cell diameter [4, 5, 41–43], as considered here. A third chemotaxis strategy represented by marine sperm cells navigating along helical paths [3], can be mapped on the case of spatial comparison: while moving along helical paths, these cells ‘visit’ different sensor positions during one helical turn.

---

**Figure 4.** Square-root scaling of optimal gradient sensing. To test the analytical result for the estimated precision at steady state, equation (19), we simulated an ensemble of agents and determined an average precision $\kappa_{\infty}$, numerically for different signal-to-noise ratios SNR by varying base concentration (A) and relative gradient strength $\alpha$ (B). Shown are mean estimated precision of individual agents (red), mean true accuracy of estimation errors in an ensemble of agents (blue), analytical result equation (19) (gray dashed). Note that the assumption equation (6) concerning the time step $\Delta t$ of time discretization is not fulfilled for the large SNR considered; correspondingly the estimated precision after update and prediction step, $\kappa_{\text{eff}}$, and $\alpha_{\text{eff}}$, respectively, oscillate around a limit value $\kappa_{\infty} = \lim_{\nu \to \infty} \sqrt{\kappa_{\text{eff}}\nu}$, which is therefore shown in the figure. Parameters: different symbols correspond to different signal-to-noise ratios: For panel A, $\text{SNR} = 0.1, 0.2, 0.5, 1, 2.5, 5, 10$ (circle, square, diamond, left triangle, right triangle, up triangle, asterisk), obtained by varying the expected number of binding events per time step $\nu$ with $\alpha = 0.03$; for panel B, $\text{SNR} = 0.25, 2.25, 25$ (circle, square, diamond), obtained by varying the relative gradient strength $\alpha = 0.01, 0.03, 0.1$ with $\nu = 5000$. For each SNR, we tested different values of the rotational diffusion coefficient with $D \Delta t = 0.01, 0.025, 0.05, 0.1, 0.25, 0.5$ and $D = D$ (symbols connected by thin line, direction of increasing $D$ indicated by arrow).

Approximately decomposed into two contributions: $\sigma^2_{\text{sensing}} = \langle (\hat{s} - \overline{s})^2 \rangle$ and $\sigma^2_{\text{time}} = \langle (\hat{s} - \overline{s})^2 \rangle$. The first contribution predominantly reflects sensing noise, and approximately scales as $\sigma^2_{\text{sensing}} \sim 1/(\tau T_{\text{eff}})$ as function of effective measurement time $T_{\text{eff}}$. The second contribution accounts for the stochastic change of the signal during the time $T_{\text{eff}}$ and approximately scales as $\sigma^2_{\text{time}} \sim T_{\text{eff}}/\tau$. While the first contribution decreases with $T_{\text{eff}}$, the second summand increases, thus posing an optimization problem for the optimal $T_{\text{eff}}$ that minimizes the total variance of the error $\sigma \approx \sigma^2_{\text{sensing}} + \sigma^2_{\text{time}}$. Solving for $d\sigma^2/dT_{\text{eff}} = 0$ yields $T_{\text{eff}} \sim (\tau/\tau)^{1/2}$. A detailed version of this argument can be found in appendix E.

**Difference to bacterial chemotaxis.** We emphasize that our and Mora’s recent results on the optimal effective measurement time for estimation of time-varying signals are different from previous work [26].

Previously, Strong et al extended the seminal work by Berg and Purcell [7] to compute the optimal filter for bacterial chemotaxis, i.e. the optimal sensori-motor transfer function for bacteria with run-and-tumble motility in the presence of sensing and motility noise. This filter extended temporal comparison over a time span set by the rotational diffusion time, i.e. $T_{\text{eff}} \sim D^{-1}$, (see equation (73) in loc. cit.), and additionally had an extremely long tail (see equation (70) in loc. cit.). This difference may be related to the fact that bacteria like *E. coli* use a chemotaxis strategy based on temporal comparison, which is different from the strategy of spatial comparison considered here; these bacteria estimate a time derivative of a concentration signal traced along their swimming path [19]. In contrast, most eukaryotic cells use spatial comparison, i.e. estimate spatial concentration gradients across the cell diameter [4, 5, 41–43], as considered here. A third chemotaxis strategy represented by marine sperm cells navigating along helical paths [3], can be mapped on the case of spatial comparison: while moving along helical paths, these cells ‘visit’ different sensor positions during one helical turn.
**Directional memory and rotational diffusion.** Biological cells performing chemotaxis possess internal memory of recent measurements of gradient direction. In eukaryotic cells with crawling cell motility, this memory manifests itself, e.g. in a persistent polarization of cell shape and the acto-myosin cytoskeleton, as well as anisotropic distributions of chemotactic effector molecules at the cell boundary [44, 45]. Stochastic remodeling of the cytoskeleton, as well as stochastic formation of pseudopods, would cause a decorrelation of cell polarization in the absence of chemotactic guidance cues. For marine sperm cells, the axis of helical swimming paths represents a consolidated memory of previous noisy steering responses [46]. In our minimal, coarse-grained model, we capture all effects that randomize direction in terms of a single effective rotational diffusion coefficient.

**Biophysical implementation.** Storing the likelihood distribution of estimated gradient directions, or just a proxy thereof, requires internal memory. Recent mathematical models of cellular gradient sensing such as LEGI (local excitation, global inhibition) [47], or balanced-inactivation models [48], consider distributions of chemotactic effector molecules on the cell boundary, which could serve as a proxy for a likelihood distribution of gradient direction: while the position of a concentration peak in such a distribution can represent a maximum likelihood estimate, its amplitude could encode a level of certainty. Slow surface diffusion of effector molecules would relax their distribution toward a uniform distribution in a way that is formally equivalent to a prediction step that accounts for the loss of knowledge due to rotational diffusion.

**Typical parameters.** Typical rotational diffusion coefficients for the bacterium *E. coli* are \( D \sim 0.1 \, \text{s}^{-1} \), close to the theoretical lower limit of a passive particle of same size and shape. For ten-fold larger sperm cells, active fluctuations dominate [49], resulting in an estimate \( D \sim 0.01 - 0.1 \, \text{s}^{-1} \) [40, 46]. The motility of crawling *Dictyostelium* cells was characterized by a persistence time of \( \sim 5 \) min in the absence of chemoattractant [50], which sets an effective rotational diffusion coefficient \( D \sim 0.003 \, \text{s}^{-1} \). For immune T cells in three-dimensional tissue, a persistence time of \( \sim 1 \) min was found (displaying a characteristic speed dependence) [51]. If binding of signaling molecules to receptors on the cell surface is diffusion-limited, we can estimate the rate of binding by \( \lambda_0 = 4\pi D_c a_0 \) for a perfectly absorbing spherical cell of radius \( a \), where \( D_c \) denotes the translational diffusion coefficient of signaling molecules [52]. For typical values \( (D_c \sim 300 \, \mu \text{m}^2 \, \text{s}^{-1}, a \sim 10 \, \mu \text{m}, c_0 \sim 1 \, \text{nM}) \), we estimate \( \lambda_0 \sim 10^4 \, \text{s}^{-1} \). Thus, for a concentration gradient of either \( \alpha_0 \approx 1\% \), \( 0.5\% \), or \( 0.1\% \) across the diameter of a cell, we estimate a rate of information gain of \( r \approx 0.5 \, \text{s}^{-1} \), \( 0.12 \, \text{s}^{-1} \), and \( 0.005 \, \text{s}^{-1} \), respectively, or, equivalently, a signal-to-noise ratio of gradient sensing SNR \( \approx 0.5, \approx 0.12, \approx 0.005 \) for a time interval \( \Delta t = 1 \) s. Assuming \( D \sim 0.003 \, \text{s}^{-1} \) [50], our main result, equation (19), predicts for the ultimate precision of gradient sensing \( CV_{\infty} \approx 0.06, \approx 0.11 \), and \( \approx 0.55 \) for these three gradient strengths, respectively. Reversible binding of signaling molecules effectively increases sensing noise, thus decreasing the signal-to-noise ratio by a constant prefactor [8, 9, 12]. Some cells, such as sperm cells, respond chemotactically even at pico-molar concentrations [53], corresponding to even lower rates of information gain [20, 39]. For the slime mold *dictyostelium*, the precision of chemotactic migration is set by sensing noise at low rates of information gain (corresponding to low signal-to-noise ratios), while noise of downstream intracellular signaling [54] becomes relevant at high rates of information gain [12, 16, 17]. Our theory of Bayesian gradient estimation provides a theoretical limit for the precision of gradient sensing, even if some cells do not achieve this limit, e.g. because these cells use shorter effective measurement times.

**Time-varying gradients.** The minimal model of a stationary agent subject to rotational diffusion considered here represents a prototypical example for a broader class of problems: if the agent actively moves in a concentration landscape, also the absolute concentration of chemoattractant and the strength of the gradient will change as the agent moves. Thus, the agent faces two estimation problems at the same time: estimating absolute concentration, which is comparatively easy, and estimating the direction of a weak concentration gradient, which is considerably more difficult. Our minimal model decouples these two problems and studies the second, more difficult problem in isolation. Similarly, in a dynamic concentration landscape, the direction of the concentration gradient will change in time with a characteristic fluctuation time-scale \( \tau \). For an idealized case of random changes of gradient direction, this case of time-varying concentration gradients is formally equivalent to the case of rotational diffusion with rotational diffusion time \( \tau \) considered here.

**Data availability statement**

No new data were created or analysed in this study.
Acknowledgments

MN and BMF are supported by the German National Science Foundation (DFG) through the Excellence Initiative by the German Federal and State Governments (Clusters of Excellence cfaed EXC-1056 and PoL EXC-2068), as well as DFG grant FR3429/3-1 and FR3429/4-1 (Heisenberg program) to BMF. We thank all members of the 'Biological Algorithms Group', in particular Andrea Auconi and Steffen Lange, for stimulating discussions.

Appendix A. Numerical methods

For numerical simulations, we represented likelihood distributions \( L_m(\psi) \) using an equidistant grid for \( \psi \) with step size of \( \Delta \psi = 0.01 \). For the case of an agent unaware of its own rotational diffusion \( (D > 0, \hat{D} = 0) \), where likelihood distributions become very sharp, we used a locally refined grid: we first identified a region of interest (ROI), where the likelihood distribution exceeded a relative threshold of \( 10^{-7} \max(L_m(\psi)) \), and used two equidistant grids, inside and outside of this ROI, with 3000 grid points each.

In each time step, the likelihood distributions were updated according to the update step, equation (15), and the prediction step, equation (17). We used the diffusion approximation, equation (11), to sample stochastic measurements \( M_j \). Drawing stochastic measurements from a Poisson distribution instead gave virtually identical results in initial simulations (same circular variance and Shannon information of measurements even for small SNR). Likewise, we confirmed that results do not change if we use a finer grid spacing for \( \psi \).

To numerically compute the true accuracy \( \text{CV}_{\infty} \) at steady state shown in figure 3, we simulated agents for \( m = 100 \) time steps and averaged \( \text{CV}[p(\varepsilon_m)] \) over the last 25 time steps. In all cases, we confirmed that \( \text{CV}[p(\varepsilon_m)] \) had converged and was fluctuating around a limit value.

To compute the average estimated precision \( \pi_\infty \) in figure 4, we first determined the circular variances \( \text{CV}[L_m(\psi)] \) and \( \text{CV}[L'_m(\psi)] \) of simulated likelihood distributions after update and prediction step, respectively, and then converted these circular variances to equivalent measures-of-concentration \( \kappa_m \) and \( \kappa'_m \) by inverting \( \text{CV} = 1 - I_1(\kappa)/I_0(\kappa) \). Finally, \( \pi_\infty \) is obtained as limit value of the geometric mean \( \sqrt[\text{CV}_{\infty} \kappa'] \) (determined as average over the last 25 of a total of 100 time steps, when all measures-of-concentration have converged). Similarly, to compute the true accuracy, we first determined the circular variances \( \text{CV}[p(\varepsilon_m)] \) and \( \text{CV}[p(\varepsilon'_m)] \) of the distribution of estimation errors in an ensemble of agents after update and prediction step, respectively, converted these again to equivalent measures-of-concentration, and finally determined the limit value of their geometric mean.

Error bars determined by bootstrapping are smaller than symbol size.

Appendix B. Expected precision of a single measurement \( \langle \zeta_j \rangle \)

To prove equation (12) stating \( \langle \zeta_j \rangle = \alpha(|M_j|) \approx \text{SNR} \), we compute the expectation value \( \langle |M_j| \rangle \) using the law of cosines. We write \( \rho = |M_j| - \langle |M_j| \rangle \) for the deviation of \( |M_j| \) from its mean \( \mu = \langle |M_j| \rangle = \omega \nu / 2 \exp i \psi_j^* \), where \( \rho = b \exp i (\psi_j^* + \gamma) \). We thus have a triangle in the complex plane with vertices \( 0, \mu, M_j \in \mathbb{C} \) and side lengths \( a = |\langle |M_j| \rangle|, b = |\rho|, c = |M_j| \), where \( \gamma \) is the angle enclosed by the sides of length \( a \) and \( b \). The law of cosines for this triangle reads

\[
c = \sqrt{a^2 + b^2 - 2ab \cos \gamma}.
\]

According to equation (11), the deviations \( \rho \) are isotropic, hence the angle \( \gamma \) is a uniformly distributed with \( p(\gamma) = 1/(2\pi) \). The squared magnitude of the deviation \( b^2 = |\rho|^2 \) follows a \( \chi^2 \)-distribution for 2 degrees of freedom (namely, real and imaginary part of \( \rho \), respectively); hence, \( p(b) = 2b/\nu \exp(-b^2/\nu) \). Thus,

\[
\langle |M_j| \rangle = \langle \zeta \rangle = \int_0^\infty db \ p(b) \int_0^{2\pi} d\varphi \ p(\varphi) \ c(b, \varphi).
\]

The first integration, \( \mathcal{I}(b) = \int_0^{2\pi} \ d\varphi \ p(\varphi) \ c(b, \varphi) \), results in an elliptic integral (explicitly, \( 2\pi \mathcal{I}(b) = \pm(\alpha \nu - 2b)E[-8\alpha b/(\alpha \nu - 2b^2)] + (\alpha \nu + 2b)E[8\alpha b/(\alpha \nu + 2b^2)] \), where \( E \) denotes the complete elliptic integral of the second kind, and the ‘+’ sign applies for \( b < \alpha \nu/2 \) and ‘−’ otherwise). The second integration in equation (S2) with respect to \( b \) cannot be performed analytically. However, \( \mathcal{I}(b) \) can
be well approximated by
\[
I(b) \approx \begin{cases} \frac{\alpha \nu}{2} \left[ 1 + \left( \frac{b}{\alpha \nu} \right)^2 \right] & b \leq \alpha \nu / 2 \\ b \left[ 1 + \left( \frac{\alpha \nu}{4 b} \right)^2 \right] & b > \alpha \nu / 2 \end{cases}
\]  
(S3)

The second integration with respect to \( b \) can now be easily done, yielding
\[
\alpha(J_M) = \alpha(c) \approx \text{SNR} + \frac{1}{2} + \cdots ,
\]  
(S4)

where the ellipses represent terms that decay exponentially fast for \( \text{SNR} \gg 1 \) (explicitly, \( \alpha(c) = \text{SNR} + (1/2 - [(2 + \text{SNR})/4] \exp(-\text{SNR}/2) + \sqrt{2\pi\text{SNR}}([4 + \text{SNR})/8]\text{erfc}(\sqrt{\text{SNR}}))\).

**Appendix C. Basic properties of circular distributions**

A probability distribution \( p(\psi) \) of angles should be \( 2\pi \)-periodic, i.e. \( p(\psi) = p(\psi + 2\pi) \), and normalized to one on the unit circle, i.e. \( \oint_0^{2\pi} d\psi p(\psi) = 1 \). The circular variance of such a circular distribution is defined as
\[
\text{CV}[p(\psi)] = 1 - \left| \oint_0^{2\pi} e^{i\psi} p(\psi) \right| .
\]  
(S5)

An important circular distribution is the **wrapped normal distribution**
\[
\mathcal{WN}(\psi; \mu, \sigma^2) = \sum_{l=-\infty}^{\infty} (2\pi\sigma^2)^{-1/2} \exp \left( -\frac{(\psi - \mu + 2\pi l)^2}{2\sigma^2} \right)
\]  
(S6)

with variance parameter \( \sigma^2 \), whose circular variance reads \( \text{CV} = 1 - e^{-\sigma^2/2} \). In the limit \( \sigma^2 \ll 1 \), \( \text{CV} \approx \sigma^2/2 \). The wrapped normal distribution is closely approximated by the von-Mises distribution, which is commonly used in directional statistics due to its mathematical tractability [55]
\[
\mathcal{VM}(\psi; \mu, \kappa) = (2\pi I_0(\kappa))^{-1} \exp \left[ \kappa \cos(\psi - \mu) \right] ,
\]  
(S7)

where \( \kappa \) is the so-called measure-of-concentration or precision, and \( I_0(\kappa) \) is the modified Bessel function of order \( n \). The circular variance reads \( \text{CV} = 1 - I_1(\kappa)/I_0(\kappa) = 1/(2\kappa) + 1/(8\kappa^2) + \mathcal{O}(\kappa^{-3}) \).

The normalized product of two von-Mises distributions, say \( \mathcal{VM}(\psi; \mu_1, \kappa_1) \) and \( \mathcal{VM}(\psi; \mu_2, \kappa_2) \), is again a von-Mises distribution \( \mathcal{VM}(\psi; \mu, \kappa) \). Such normalized product appears, e.g. in Bayes formula, equation (15). Specifically, the mean \( \mu \) and measure-of-concentration \( \kappa \) of the normalized product satisfy
\[
\kappa e^{i\mu} = \kappa_1 e^{i\mu_1} + \kappa_2 e^{i\mu_2}.
\]  
(S8)

If mean values are close, \( |\mu_1 - \mu_2| \ll 1 \), we have the approximate sum rule \( \kappa \approx \kappa_1 + \kappa_2 \).

In contrast, the circular convolution of two von-Mises distributions \( \mathcal{VM}(\psi; \mu_1, \kappa_1) \) and \( \mathcal{VM}(\psi; \mu_2, \kappa_2) \) is only approximately a von-Mises distribution [55]. To find such approximation, one can first map the two von-Mises distributions onto wrapped normal distributions of same respective mean and circular variance, and compute the convolution of these wrapped normal distributions [38]. The convolution of two wrapped normal distributions, say with variance parameters \( \sigma^2_1 \) and \( \sigma^2_2 \), is again a wrapped normal distribution with new variance parameter \( \sigma^2 = \sigma^2_1 + \sigma^2_2 \). Finally, this new wrapped normal distribution \( \mathcal{WN}(\psi; \mu_1 + \mu_2, \sigma^2) \) is mapped back onto a von-Mises distribution \( \mathcal{VM}(\psi; \mu, \kappa) \). Thus, \( \mathcal{VM}(\psi; \mu, \kappa) \approx \mathcal{VM}(\psi; \mu_1, \kappa_1) \mathcal{VM}(\psi; \mu_2, \kappa_2) \) with \( \mu = \mu_1 + \mu_2 \) and
\[
I_1(\kappa)/I_0(\kappa) = \exp(-\sigma^2/2) = \exp(-\sigma^2_1) \exp(-\sigma^2_2) = [I_1(\kappa_1)/I_0(\kappa_1)] \cdot [I_1(\kappa_2)/I_0(\kappa_2)] .
\]  
In the limit \( \kappa_1, \kappa_2 \gg 1 \), we have \( \kappa^{-1} = 1 + \kappa_1^{-1} + \kappa_2^{-1} \).

For ease of reference, we highlight the sum rules for the measure-of-concentration of either a normalized product or convolution of two von-Mises distributions
\[
\text{convolution: } \frac{1}{\kappa} \approx \frac{1}{\kappa_1} + \frac{1}{\kappa_2} ,
\]  
(S9)

\[
\text{multiplication: } \kappa \approx \kappa_1 + \kappa_2 ,
\]  
(S10)

While equation (S9) is valid for \( \kappa_1, \kappa_2 \gg 1 \), equation (S10) is valid for \( |\mu_1 - \mu_2| \ll 1 \). Table C presents a summary of this section. Figure S1 graphically relates von-Mises distributions and wrapped normal distributions of equal circular variance.
Appendix D. Agents assuming wrong rotational diffusion coefficient

We provide details on the calculation of the true accuracy of gradient sensing in the long time limit, CV_{\infty} = \lim_{m \to \infty} CV[p(\varphi_m)]$, for the case that agents assume a wrong rotational diffusion coefficient $\hat{D} \neq D$.

First, we note that the iteration rule for $CV_m$, equation (22), still holds in the case $\hat{D} \neq D$. We take the limit $m \to \infty$ on both sides an obtain a self-consistency equation for $CV_{\infty}$

$$CV_{\infty} \approx \left( \frac{\kappa_{\infty}'}{\kappa_{\infty}} \right)^2 (CV_{\infty} + D\Delta t) + \left( \frac{\text{SNR}}{\kappa_{\infty}} \right)^2 \frac{1}{2 \text{SNR}}$$

$$\approx \frac{\kappa_{\infty}^2}{2 \text{SNR} \kappa_{\infty}} (CV_{\infty} + D\Delta t) + \left( \frac{\text{SNR}}{\kappa_{\infty}} \right)^2 \frac{1}{2 \text{SNR}}.$$  \hspace{1cm} (S11)

where we used $\kappa_{\infty}' = \lim_{m \to \infty} \kappa_{\infty}' \approx \kappa_{\infty} - \text{SNR}$. Note that $\kappa_{\infty}, \kappa_{\infty}' \gg \text{SNR}$ by equation (6). In the main text, we had shown for the estimated precision $\kappa_{\infty}$ in the long time limit, $\kappa_{\infty} = [\text{SNR}/(2D\Delta t)]^{1/2}$, hence $\text{SNR}/\kappa_{\infty} = (\text{SNR} \cdot 2D\Delta t)^{1/2}$. Inserting this into equation (S11) gives

$$CV_{\infty} \approx \left( 1 - 2\sqrt{\text{SNR} \cdot 2D\Delta t} \right) (CV_{\infty} + D\Delta t) + \hat{D}\Delta t.$$  \hspace{1cm} (S12)
or after rearranging
\[
1 - 2 \sqrt{\text{SNR} \cdot 2D\Delta t} = \frac{CV_\infty - \hat{D}\Delta t}{CV_\infty + D\Delta t} \approx 1 - \frac{\hat{D} + D}{CV_\infty} \Delta t.
\]
(S13)

Solving for \(CV_\infty\) yields the result stated in equation (24), \(CV_\infty \approx \frac{1}{2} \sqrt{2D\Delta t/\text{SNR} \cdot (D + \hat{D})/\sqrt{4\hat{D}}.}

**Appendix E. Agents unaware of their rotational diffusion**

We compute the second moment \(\langle \kappa_m^2 \rangle\) of the measure-of-concentration \(\kappa_m\) of the likelihood distribution \(L_m(\psi)\) for the gradient angle \(\psi\) after \(m\) measurements.

Using the calculus of von-Mises distributions, we can express \(\kappa_m\) as a circular average of all previous measurements, see equation (25), which reads \(\kappa_m \exp(i\tilde{\psi}_m) = \kappa_0 \exp(i\tilde{\psi}_0) + \sum_{j=1}^m \zeta_j \exp(i\tilde{\psi}_j)\). For simplicity, we assume a flat initial prior with \(\kappa_0 = 0\). Using \(\zeta_j = \alpha |M_j|\), we can rewrite this equation in terms of the complex measurement \(M_j\) as

\[
\kappa_m e^{i\tilde{\psi}_m} = \alpha \sum_{j=1}^m M_j.
\]
(S14)

Note

\[
\alpha^2 (|M_j|^2) = \alpha^2 \left( \frac{\alpha^2}{2} \right)^2 = \text{SNR}^2 + 2 \text{SNR},
\]
(S15)

and

\[
\alpha^2 (M_j M_k^*) = \alpha^2 \left( \frac{\alpha^2}{2} \right)^2 e^{-D|j-k|\Delta t}.
\]
(S16)

We also note the expression for the double-geometric series with \(0 < \xi < 1\)

\[
\sum_{j=1}^m \xi^{j-k} = 2 \sum_{j=1}^{m-1} \sum_{k=j+1}^m \xi^{k-j} = 2 \sum_{j=1}^{m-1} \frac{\xi^{m-j} - 1}{1 - \xi} = 2(m-1) \frac{\xi}{1-\xi} - 2 \frac{\xi^2}{(1-\xi)^2} (1 - \xi^{m-1}).
\]
(S17)

Thus,

\[
\langle \kappa_m^2 \rangle = \alpha^2 \sum_{j=1}^m (|M_j|^2) + \sum_{j \neq k} (M_j M_k^*)
\]

\[
= m (\text{SNR}^2 + 2 \text{SNR}) + \text{SNR}^2 \sum_{j \neq k} \xi^{j-k}, \quad \xi = e^{-D\Delta t}
\]

\[
= 2m \text{SNR} + \left( m e^{\Delta t} + 1 \right) \left( e^{\Delta t} - 1 \right)^{-1} - 2 e^{\Delta t} \left( e^{\Delta t} - 1 \right)^{-2} \text{SNR}^2.
\]
(S18)

For long times, \(t_m \gg \tau\), this expression scales with \(m\), as \(\langle \kappa_m^2 \rangle \approx 2m^2 \tau^2\). This follows from \(\exp(D\Delta t) \approx 1 + D\Delta t\) and \(\exp(-mD\Delta t) \approx 0\) to leading order in \(D\Delta t\). For short times, \(t_m \ll \tau\), however, \(\langle \kappa_m^2 \rangle \approx m^2 \text{SNR}^2\). Specifically, a Fourier expansion of \(\exp(-mD\Delta t) \approx 1 - mD\Delta t + (mD\Delta t)^2/2\) to second order yields \(\langle \kappa_m^2 \rangle = 2m \text{SNR} + [m^2 + (m - m^3)D\Delta t/3] \text{SNR}^2 + O(D\Delta t)^2\).

**Appendix F. Optimal effective measurement time**

We consider a minimal example for the estimation of a time-varying signal to illustrate the origin of the square-root scaling of the optimal effective measurement time with the fluctuation time-scale of the signal. Suppose a signal \(s'(t)\) varies stochastically with characteristic fluctuation time-scale \(\tau\), e.g. as a diffusion process with \(ds'/dt = \xi(t)\), where \(\xi(t)\) denotes Gaussian white noise with \(\langle \xi(t)\xi(t') \rangle = \tau^{-1}\delta(t-t')\). The agent estimates \(\tilde{s}(t) = T_{\text{eff}}^{-1} \int_{t'-T_{\text{eff}}}^{t-T_{\text{eff}}} ds'(t') + \eta(t')\) by averaging noisy measurements over a time interval \(T_{\text{eff}}\) with measurement noise \(\eta(t)\) satisfying \(\langle \eta(t)\eta(t') \rangle = \tau^2\delta(t-t')\). The variance \(\sigma^2\) of the expected
estimation error $\delta(t) = \tilde{s}(t) - s'(t)$ between the estimated value $\tilde{s}(t)$ of the signal and its true value $s'(t)$ can then be composed into two parts, where $\overline{\tilde{s}} = T^{-1}_{\text{eff}} \int_0^t \mathrm{d}t' s'(t')$

$$\sigma^2 = \left\langle \left[ \tilde{s}(t) - s'(t) \right]^2 \right\rangle \approx \frac{\left\langle \tilde{s}(t) - \overline{\tilde{s}} \right\rangle^2}{\sigma^2} + \frac{\left\langle s'(t) - \overline{\tilde{s}} \right\rangle^2}{\sigma^2_{\text{time}}}.$$  \hspace{1cm} (S19)

An explicit calculation yields

$$\sigma^2 = (r T_{\text{eff}})^{-1} + T_{\text{eff}}/(3 \tau).$$ \hspace{1cm} (S20)

Solving for $d \sigma^2 / d T_{\text{eff}} = 0$ yields $T_{\text{eff}} = (3 \tau / r)^{1/2}$.

**Appendix G. List of symbols used**

| Symbol | Meaning |
|--------|---------|
| $\alpha$ | Normalized gradient strength $\alpha = |\nabla c|/c_0$ |
| $\nu$ | Sensing length scale, radius of agent |
| $\epsilon_0$ | Base concentration |
| c.c. | In equations: add complex conjugate of last term |
| CV | Circular variance of an angular distribution $p(\psi)$, $CV = 1 - |\exp(\mathrm{i} \psi)| |$ |
| $\epsilon_m$ | Estimation error of individual agent at time $t_m$, $\epsilon_m = \psi_{\text{meas}} - \psi_{\text{true}}$ |
| $\Delta_t$ | Random rotation of agent at time $t_j$ |
| $\Delta t$ | Time step, cf equation (6) |
| $D$ | Rotational diffusion coefficient of agent (with units of inverse time) |
| $D'$ | Value of $D$ assumed by agent |
| $e_i$ | Gradient direction vector $e_i = \nabla c / |\nabla c|$ |
| $h_1, h_2$ | Orthonormal material frame vectors of agent |
| $\kappa_m$ | Measure-of-concentration of likelihood distribution $L_m(\psi)$ after $m$ measurements |
| $\kappa_\infty$ | Limit value of $\kappa_m$, representing ultimate precision |
| $\kappa_m'$ | Measure-of-concentration of likelihood distribution $L_m'(\psi)$ after prediction step |
| $\lambda$ | Proportionality factor between rate of molecule detection and local concentration |
| $\Lambda(\psi)$ | Rate of molecule detection at polar angle position $\psi$ |
| $L_m(\psi)$ | Likelihood distribution of estimated gradient direction at time $t_m$ after Bayesian updating |
| $L_m'(\psi)$ | Likelihood distribution of estimated gradient direction at time $t_m$ before Bayesian updating |
| $\mu$ | Expectation value of measurements $\mu = \langle M \rangle$ |
| $M$ | Measurement for time interval $T_j$, cf equation (8) |
| $\nu$ | Expected total molecule count during time $\Delta t$ |
| $\psi'(t)$ | True gradient direction angle relative to material frame of agent |
| $\psi_m'$ | Constant value of $\psi'(t)$ during time interval $T_j$ |
| $P(M|\psi)$ | Probability for measurement $M$ given direction angle $\psi$ |
| $r$ | Rate of information gain, $r = \text{SNR} / \Delta t$, cf equation (5) |
| $s(t)$ | Complex signal of directional binding events $s(t) = \sum \delta(t - \theta) \psi^{*\to \psi}$, cf equation (3) |
| $\Sigma$ | Covariance of measurement |
| $\text{SNR}$ | Signal-to-noise ratio of gradient measurements, cf equation (4) |
| $\Delta t$ | Time step (duration of measurement intervals) |
| $t_j$ | Discrete time points $t_j = j \Delta t$ |
| $T_j$ | Measurement intervals $T_j = (t_{j-1}, t_j)$ |
| $\tau$ | Rotational diffusion time $\tau = 1/(2 D')$ |
| $\mathcal{V}_M(\psi; \mu, \kappa)$ | von-Mises distribution for variable $\psi$ with mean $\mu$ and measure-of-concentration $\kappa$ |
| $\mathcal{W}_N(\psi; \mu, \sigma^2)$ | Wrapped normal distribution for $\psi$ with mean $\mu$ and variance parameter $\sigma^2$ |
| $\zeta_i$ | Measure-of-concentration of update kernel $L(M|\psi); \text{note } \langle \zeta_i \rangle = \text{SNR}$ |

**ORCID IDs**

Maja Novak https://orcid.org/0000-0002-6493-8597

Benjamin M Friedrich https://orcid.org/0000-0002-9742-6555

**References**

[1] Berg H C and Brown D A 1972 Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking Nature 239 500–4

[2] Eisenbach M and Giojalas L C 2006 Sperm guidance in mammals—an unpaved road to the egg Nat. Rev. Mol. Cell Biol. 7 276–85

[3] Alvarez L, Friedrich B M, Gompper G and Kaupp U B 2014 The computational sperm cell Trends Cell Biol. 24 198–207
[49] Ma R, Klint G S, Riedel-Kruse I H, Jülicher F and Friedrich B M 2014 Active phase and amplitude fluctuations of flagellar beating Phys. Rev. Lett. 113 048101
[50] Van Haastert P J M 2010 A model for a correlated random walk based on the ordered extension of pseudopodia PLoS Comp. Biol. 6 e1000874
[51] Jerison E R and Quake S R 2020 Heterogeneous T cell motility behaviors emerge from a coupling between speed and turning in vivo eLife 9 e53933
[52] Berg H C 1993 Random Walks in Biology (Princeton, NJ: Princeton University Press)
[53] Strünker T, Alvarez L and Kaupp U 2015 At the physical limit—chemosensation in sperm Curr. Opin. Neurobiol. 34 110–6
[54] Ueda M and Shibata T 2007 Stochastic signal processing and transduction in chemotactic response of eukaryotic cells Biophys. J. 93 11–20
[55] Mardia K V and Jupp P E 2000 Directional Statistics (New York: Wiley)