Robustness of Dynamical Quantities of Interest via Goal-Oriented Information Theory

Jeremiah Birrell · Markos Katsoulakis · Luc Rey-Bellet

the date of receipt and acceptance should be inserted later

Abstract Variational-principle-based methods that relate expectations of a quantity of interest to information-theoretic divergences have proven to be effective tools for obtaining distributional robustness bounds under both parametric and non-parametric model-form uncertainty. Here, we extend these ideas to utilize information divergences that are specifically targeted at the quantity-of-interest being investigated. Due to their goal-oriented nature, and when combined with the data processing inequality, the resulting robustness bounds are tighter and a wider class of problems can be treated. We find the method especially useful for problems involving unbounded time-horizons, a case where established information-theoretic methods typically result in trivial (infinite) bounds. Problem types that can be treated within this framework include robustness bounds on the expected value and distribution of a stopping time, time averages, and exponentially discounted quantities. We illustrate these results with several examples, including option pricing, stochastic control, semi-Markov queueing models, and expectations and distributions of hitting times.

Keywords distributional robustness; uncertainty quantification; relative entropy; semi-Markov queueing models; stochastic control; option pricing; hitting times

Mathematics Subject Classification (2010) 62F35, 62B10, 60G40, 60J60, 93E20, 91G20

J. Birrell
Department of Mathematics and Statistics
University of Massachusetts Amherst
Amherst, MA 01003, USA
E-mail: birrell@math.umass.edu

M. Katsoulakis
Department of Mathematics and Statistics
University of Massachusetts Amherst
Amherst, MA 01003, USA

L. Rey-Bellet
Department of Mathematics and Statistics
University of Massachusetts Amherst
Amherst, MA 01003, USA
1 Introduction

Variational-principle-based methods, relating expectations of a quantity of interest to information-theoretic divergences, have proven to be effective tools for deriving distributional robustness bounds. These bounds, here referred to as uncertainty quantification (UQ) bounds, have been applied to both parametric and non-parametric model-form uncertainty [14,12,26,36,29], and have been extended to distributions on path-space [43,20,28,34,7] and risk-sensitive quantities [5,19]. Such robustness questions are also central to the problem of distributionally robust optimization [16,27,49,32,23,37,41,50,9], wherein optimal-transport (Wasserstein-metric) based methods are often used.

One known strength of the information-theoretic robustness methods is the scaling properties of relative entropy, both with system size and in the long-time regime [34], that result from the chain rule; this property facilitates the derivation of UQ bounds for ergodic averages that are well behaved in the long-time limit [34,7]. However, time-averages are somewhat of a special case within these established methods; for other quantities-of-interest (QoI’s) with an unbounded time-horizon, such as stopping times, current relative-entropy based methods result in trivial (infinite) bounds. The primary innovation of the present work is the introduction of so-called goal-oriented relative entropy and Fisher information; these objects are defined so as to target the specific QoI under investigation. When combined with the data processing inequality, Eq. (3), these information-theoretic quantities allow for the derivation of tighter UQ bounds and yield nontrivial, computable bounds in many new cases.

In this paper, UQ will refer to the following general problem: One has a baseline model, described by a probability measure \( P \), and a family of alternative models, \( \tilde{P} \in \tilde{P} \), also known as the ambiguity set. The base model is typically the ‘simpler’ model, meaning one can calculate QoI’s exactly, or it is relatively inexpensive to simulate, but it may contain many sources of error and uncertainty; it may depend on parameters with uncertain values (obtained from experiment, Monte-Carlo simulation, variational inference, etc.) or is obtained via some approximation procedure (dimension reduction, neglecting memory terms, linearization, asymptotic approximation, etc.). Any quantity of interest computed from \( P \) therefore has significant uncertainty associated with it. In contrast, the family of alternative models is generally thought of as containing the ‘true’ (or at least, a more precise) model, but due to its complexity or a lack of knowledge, it is considered intractable. Our UQ goal is then to, for a given QoI (real-valued function), \( F \), bound the error incurred from using \( P \) in place of \( \tilde{P} \), i.e.

\[
\text{Bound the bias: } \quad E_{\tilde{P}}[F] - E_P[F], \quad \tilde{P} \in \tilde{P}.
\]  

Here, and in the following, \( E_Q \) will denote the expectation under the probability measure \( Q \). We will also refer to \( E_P[F] \) (resp. \( E_{\tilde{P}}[F] \)) as the base model (resp. alternative model) QoI.

As examples, \( F \) could be the failure-time of a system (or a more general stopping time), the value of some asset, or a cost-functional. A bound on Eq. (1) is then a performance guarantee, given the modeling uncertainty expressed by the family \( \tilde{P} \). See Section 5 for more specific examples.

The primary philosophy behind the methods derived below is that the bound Eq. (1) should be specialized to the QoI, \( F \), and the family \( \tilde{P} \) as much as possi-
the base model, while still involving quantities that can be effectively computed/estimated. In short, one does not necessarily care about all differences between, \( \tilde{P} \) and \( P \), but only the differences insofar as they impact the expectation of the QoI.

1.1 Summary of Results

Our new results all rest upon an improvement to the Gibbs-Variational-Principle-based UQ bounds that were developed and employed in \([14,12,26,36,29,43,20,28,34,7]\); this improvement is the content of Section 3. Specifically, Theorem 1 bounds (under suitable assumptions) the difference in expectation of a QoI, \( F \), with respect to a base model, \( P \), and an alternative model, \( \tilde{P} \), as follows:

\[
\pm (E_{\tilde{P}}[F] - E_P[F]) \leq \inf_{c>0} \left\{ \frac{1}{c} A^\tilde{P}_F(\pm c) + \frac{1}{c} R(F, \tilde{P}||F, P) \right\}.
\]  

Here \( \tilde{P} = F - E_P[F] \) is the QoI centered under \( P \), \( A^\tilde{P}_F(c) = \log E_P[e^{c\tilde{F}}] \), is the cumulant generating function of \( \tilde{F} \) under \( P \), \( F, P \) denotes the distribution of \( F \) on \( \mathbb{R} \) under \( P \), and similarly for \( F, \tilde{P} \) and \( R(\cdot||\cdot) \) denotes the relative entropy (i.e. Kullback-Leibler divergence). See Section 2 for detailed definitions and further background.

\( P \) is typically taken to be a tractable model, so that the cumulant generating function can be computed or bounded. As for the second term, \( R(F, \tilde{P}||F, P) \) is generally unobtainable, as it requires detailed knowledge of the distribution of \( \tau \), the filtration up to the stopping time. In many cases, \( R(F, \tilde{P}||F, P) \) can be computed or bounded. As for the second term, \( R(F, \tilde{P}||F, P) \) is generally unobtainable, as it requires detailed knowledge of the distribution of \( \tau \), the filtration up to the stopping time. In many cases, \( R(F, \tilde{P}||F, P) \) can be computed or bounded.

In scenarios like this, Eq. (2) becomes a practical tool for UQ when it is combined with the data processing inequality (again, see Theorem 1). Specifically, if \( F \) is \( \mathcal{G} \)-measurable, where \( \mathcal{G} \) is a sub sigma-algebra of the full sigma algebra of the probability space, then

\[
R(F, \tilde{P}||F, P) \leq R(\tilde{P}||\mathcal{G}) \leq R(\tilde{P}||P).
\]  

Here, \( \mathcal{G} \) denotes the restriction of the probability measure \( Q \) to the sub sigma-algebra. For example, many of our applications will involve some \( \mathcal{F}_\tau \)-stopping time, \( \tau \); in such cases, a useful choice is typically \( \mathcal{G} = \mathcal{F}_\tau \), the filtration up to the stopping time.

Eq. (3) gives one a great deal of freedom when weakening the robustness bound Eq. (2): ideally, this freedom can be exploited to obtain bounds that depend on an explicitly computable relative entropy, while still maintaining as much information regarding \( F \) as possible (i.e. enlarging \( \mathcal{G} \) as little as possible). With this aim in mind, we call \( R(F, \tilde{P}||F, P) \) and \( R(\tilde{P}||\mathcal{G}) \) goal-oriented relative entropies, to contrast them with the non-goal-oriented quantity \( R(\tilde{P}||P) \). In many cases, \( R(\tilde{P}||P) \) is infinite (see, for example, Remark 11), and so the use of goal-oriented quantities is necessary if one is to obtain non-trivial bounds. Even when \( R(\tilde{P}||P) \) is
finite, the use of goal-oriented quantities results in tighter bounds. These points will be illustrated throughout our examples; for a particularly simple and clear demonstration, see Section 5.1.1, especially Figure 1.

Clearly, it is only useful to weaken the robustness bound Eq. (2) via the data processing inequality when \( R(\tilde{P}|\mathcal{G}||P|\mathcal{G}) \) is computable in some sense. In practice, we will require a bound of the following form to hold:

\[
R(\tilde{P}|\mathcal{G}||P|\mathcal{G}) \leq E_{\tilde{P}}[G] \text{ for some } \mathcal{G}\text{-measurable } G \in L^1(\tilde{P}). \tag{4}
\]

We will prove that Eq. (2) combined with Eq. (3) and Eq. (4) imply the following UQ bounds, which involve expectations only under the (tractable) base model, \( P \) (see Theorem 2):

\[
\pm E_{\tilde{P}}[F] \leq \inf_{c>0} \left\{ \frac{1}{c} \log E_P \left[ \exp \left( \pm c F + G \right) \right] \right\}. \tag{5}
\]

In fact, (under appropriate assumptions) taking \( G = \log \left( \frac{d\tilde{P}|\mathcal{G}}{dP|\mathcal{G}} \right) \) makes Eq. (5) an equality. Of course, this choice is typically impractical to work with, but it indicates the potential utility of the bound Eq. (5). In practice, one should aim to select appropriate \( G \) and \( G \) so that the right-hand-side of Eq. (5) is comparatively simple to compute/estimate, while at the same time keeping the inequalities as tight as possible. Apart from Eq. (5), other techniques for eliminating the \( \tilde{P} \)-dependence of the bounds are also useful; see Remark 6 and the example in Section 5.5.1.

Eq. (2) and Eq. (5) are non-perturbative UQ-bounds. For parametric families, one can similarly obtain perturbative UQ-bounds (i.e. sensitivity bounds) in terms of so-called goal-oriented Fisher information; see Section 3.1.

Our primary application of these results will be to QoI’s up to a stopping time. In this setting, one has a time-dependent process, \( F_t \), adapted to a filtration \( \mathcal{F}_t \), and a \( \mathcal{F}_t \)-stopping time \( \tau \), and the objective is to bound the expectation of the stopped process, \( F_{\tau} \). This general framework is developed in Section 4, where we also discuss several classes of problems that fit this description. These are listed in Table 1.1 below:

| Problem types | QoI |
|---------------|-----|
| 1. Relative entropy up to a stopping time. | \( R(\tilde{P}|\mathcal{F}_\tau||P|\mathcal{F}_\tau) \) |
| 2. Distribution of a stopping time. | \( \tilde{P}(\tau < T) \) |
| 3. Expected value of a stopping time. | \( E_{\tilde{P}}[\tau] \) |
| 4. Expectation of exponentially discounted QoI’s. | \( \int_0^\infty f_s \lambda e^{-\lambda s} ds \) |
| 5. Expectation of time averages. | \( \frac{1}{T} \int_0^T f(X_t) dt \) |

Relative entropy has many properties that make it convenient for addressing such path-space problems; in particular, when \( \mathcal{G} = \mathcal{F}_\tau \), bounds of the form Eq. (4) are often computable via standard tools. For example, in Corollary 6 we are able to use Eq. (4) and Eq. (5) to obtain the following bound on the expectation of \( \tau \):

\[
- \inf_{c>0} \left\{ (c + K)^{-1} (A_{\tilde{P}}(-c) + \eta_0) \right\} \leq E_{\tilde{P}}[\tau] \leq \inf_{c>0} \left\{ (c - K)^{-1} (A_{\tilde{P}}(c) + \eta_0) \right\}, \tag{6}
\]
where the ambiguity set consists of all alternative models, $\tilde{P}$, satisfying a relative entropy bound

$$R(\tilde{P} | \mathcal{F}_{\tau,n} \| P | \mathcal{F}_{\tau,n}) \leq \eta_0 + K E_{\tilde{P}}[\tau \land n]$$

(7)

for some fixed $\eta_0, K \in [0, \infty)$ and all $n \in \mathbb{Z}^+$. As we will show, the bound Eq. (6) does not require a-priori knowledge that $E_{\tilde{P}}[\tau]$ is finite; it provides a method for proving finiteness. In contrast, the non-goal-oriented relative entropy $R(\tilde{P} | \mathcal{F}_{n} \| P | \mathcal{F}_{n})$ typically grows like $n$, and hence results only in trivial bounds for problems with an unbounded time-horizon. When the base and alternative models come from Markov processes, one can often obtain a bound of the form Eq. (7) (e.g. by Girsanov’s theorem); see the online supplement for details.

Finally, we illustrate the use of these results in Section 5 through several examples:

1. The distribution and expectation of hitting times of Brownian motion with constant drift, as compared to Brownian motion perturbed by a non-constant drift; see Section 5.1.
2. Expected hitting times of an Ornstein-Uhlenbeck process perturbed by a non-constant drift; see Section 5.2.
3. Robustness of linear-quadratic stochastic control under non-linear perturbations; see Section 5.3.
4. Robustness of continuous-time queueing models under non-exponential waiting-time perturbations (i.e. semi-Markov perturbations); see Section 5.4.
5. The value of American put options in a variable-interest-rate environment (Vasicek model); see Section 5.5 and Appendix B.

2 Background on Uncertainty Quantification via Information-Theoretic Variational Principles

Here we record important background information on the information-theoretic-variational-principle approach to UQ; the new methods presented in this paper (starting in Section 3) will be built upon these foundations.

First, we fix some notation. Let $P$ be a probability measure on a measurable space $(\Omega, \mathcal{F})$. $\mathbb{R}$ will denote the extended reals and we will refer to a random variable $F : \Omega \rightarrow \mathbb{R}$ as a quantity-of-interest (QoI). The cumulant generating function of $F$ is defined by

$$\Lambda^F_P(c) = \log E_P[e^{cF}],$$

(8)

where $\log$ denotes the natural logarithm. We use the continuous extensions of $\exp(x)$ to $\mathbb{R}$ and of $\log$ to $[0, \infty]$.

The following notation will be used for the set of real-valued QoIs with a well-defined and finite moment generating function (MGF) under $P$:

$$\mathcal{E}(P) = \left\{ F : \Omega \rightarrow \mathbb{R} : E_P[e^{\pm \alpha F}] < \infty \text{ for some } \alpha > 0 \right\}.$$  

(9)

It is not difficult to prove (see e.g. [17]) that for $F \in \mathcal{E}(P)$, the cumulant generating function is a convex function, finite and infinitely differentiable in some interval
(c_, c+) with \(-\infty \leq c_- < 0 < c_+ \leq \infty\), equal to \(+\infty\) outside of \([c_-, c_+]\), and such \(F\) have moments of all orders.

If \(F \in L^1(P)\) we write
\[
\hat{P} = F - E_P[F]
\]
for the centered quantity of mean 0. We will often use the cumulant generating function for a centered QoI \(\hat{F}\):
\[
A_P^\hat{F}(c) = \log E_P[e^{c(F-E_P[F])}] = A_P^F(c) - cE_P[F].
\]
Recall that \(A_P^\hat{F} \geq 0\).

Recall also that the relative entropy (or Kullback-Leibler divergence) of another probability measure, \(\check{P}\), with respect to \(P\) is defined by
\[
R(\check{P}||P) = \begin{cases} E_{\check{P}} \left[ \log \left( \frac{d\check{P}}{dP} \right) \right] & \text{if } \check{P} \ll P \\ +\infty & \text{otherwise.} \end{cases}
\]
It has the property of a divergence, that is \(R(\check{P}||P) \geq 0\) and \(R(\check{P}||P) = 0\) if and only if \(\check{P} = P\); see, for example, [18] for this and further properties.

The starting point of our approach is a non-goal-oriented UQ bound, derived in [14,20] (a similar inequality is used in the context of concentration inequalities, see e.g. [10], and was also used independently in [12,26]). We summarize the proof here for completeness.

**Proposition 1 (Gibbs information inequality)** Let \(F : \Omega \rightarrow \mathbb{R}\), \(F \in L^1(P) \cap L^1(\hat{P})\). Then
\[
\pm \left( E_{\hat{P}}[F] - E_P[F] \right) \leq \inf_{c > 0} \left\{ \frac{1}{c} A_P^\hat{F}(\pm c) + \frac{1}{c} R(\hat{P}||P) \right\}.
\]

**Proof** The starting point is the Gibbs variational principle, which relates the cumulant generating function and relative entropy as follows (see [18]): If \(G : \Omega \rightarrow \mathbb{R}\) is measurable and satisfies \(E_P[e^{cG}] < \infty\) then
\[
\log E_P[e^{cG}] = \sup_{Q: R(Q||P) < \infty} \left\{ E_Q[G] - R(Q||P) \right\}.
\]

When \(F \in \mathcal{E}(P)\) and \(R(\hat{P}||P) < \infty\), the result Eq. (13) follows from applying Eq. (14) to \(G = \pm c\hat{F}\) and bounding the supremum below by the value at \(Q = \hat{P}\). One can relax the assumptions to \(F \in L^1(P) \cap L^1(\hat{P})\) by truncating \(F\), using the result for bounded \(F\), and then taking limits. The case \(R(\hat{P}||P) = \infty\) is trivial. \(\Box\)

The objects
\[
\Xi(\hat{P}||P; F) \equiv \inf_{c > 0} \left\{ \frac{1}{c} A_P^\hat{F}(c) + \frac{1}{c} R(\hat{P}||P) \right\}
\]
appearing in the Gibbs information inequality, Eq. (13), have many interesting properties; we recall several of them below:

**Proposition 2** Assume \(R(\hat{P}||P) < \infty\) and \(F \in \mathcal{E}(P)\). We have:

1. **(Divergence)** \(\Xi(\hat{P}||P; F)\) is a divergence, i.e. \(\Xi(\hat{P}||P; F) \geq 0\) and \(\Xi(\hat{P}||P; F) = 0\) if and only if either \(P = \hat{P}\) or \(F\) is constant \(\hat{P}\)-almost-surely (a.s.).
2. (Linearization) If $R(\tilde{P}||P)$ is sufficiently small we have

$$\Xi(\tilde{P}||P; F) = \sqrt{2\text{Var}_P[F]R(\tilde{P}||P) + O(R(\tilde{P}||P))}. \quad (16)$$

3. (Tightness) For $\eta > 0$ consider $U_\eta = \{\tilde{P} ; R(\tilde{P}||P) \leq \eta\}$. There exists $\eta^*$ with $0 < \eta^* \leq \infty$ such that for any $\eta < \eta^*$ there exists a measure $P^\eta$ with

$$\sup_{\tilde{P} \in U_\eta} \{E_{\tilde{P}}[F] - E_P[F]\} = E_{P^\eta}[F] - E_P[F] = \Xi(P^\eta||P; F). \quad (17)$$

The measure $P^\eta$ has the form $dP^\eta = e^{cF - \Lambda_F^\eta(c)}dP$, where $c = c(\eta)$ is the unique non-negative solution of $R(\tilde{P}||P) = \eta$.

**Proof** Items 1 and 2 are proved in [20]; see also [36] for item 2. Various versions of the proof of item 3 can be found in in [14] or [20]. See Proposition 3 in [19] for a more detailed statement of the result; see also similar results in [12,11]. ⊓⊔

3 Robustness of QoIs via Goal Oriented Information Theory

Much of the utility of the Gibbs information inequality, as expressed in Eq. (13), comes from the manner in which it splits the problem of obtaining UQ bounds into two sub-problems:

1. Bound the cumulant generating function of the QoI with respect to the base model, $\Lambda_F^\circ(c)$. (Note that there is no dependence on the alternative model, $\tilde{P}$.)
2. Bound the relative entropy of the alternative model with respect to the base model: $R(\tilde{P}||P)$.

Despite the tightness property given in Theorem 2, when one is interested in a specific QoI and a specific (class of) alternative model(s), the Gibbs information inequality, Eq. (13), is not quite ideal. Requiring a bound on the full relative entropy of the alternative model with respect to the base, $R(\tilde{P}||P)$, is too restrictive in general and can lead to poor or useless bounds, e.g. when $\tilde{P}$ and $P$ are the distributions of time-dependent processes on path space with infinite time-horizon.

In the following theorem, we show how a simple modification results in a relative-entropy term that is targeted at the information present in the QoI:

**Theorem 1 (Goal-Oriented Information Inequality)** Let $F : \Omega \rightarrow \overline{\mathbb{R}}$.

1. If $F \in L^1(\tilde{P}) \cap L^1(P)$. Then

$$\pm (E_{\tilde{P}}[F] - E_P[F]) \leq \inf_{c > \theta} \left\{ \frac{1}{c} \Lambda_F^\circ(\pm c) + \frac{1}{c} R(F, \tilde{P}||F, \tilde{P}) \right\}, \quad (18)$$

where $F, P$ is the distribution of $F$ on $\overline{\mathbb{R}}$ under $P$, and similarly for $F, \tilde{P}$.

2. If $F$ is $G$-measurable, where $G \subset F$ is a sub sigma-algebra, then

$$R(F, \tilde{P}||F, \tilde{P}) \leq R(\tilde{P}||P|G) \leq R(\tilde{P}||P). \quad (19)$$
Remark 1 Eq. (19) is one of the key properties that makes the bound Eq. (18) useful in practice. The relative entropy term \( R(F, \mathcal{G} || F, \mathcal{G}) \) in Eq. (18) is generally inaccessible, but Eq. (19) allows one to relax the UQ bound, with the intent of finding a computable upper bound. For UQ purposes, the objective is then to find the smallest such \( \mathcal{G} \) for which \( R(F, \mathcal{G} || F, \mathcal{G}) \) is ‘easy’ to compute/bound. This will be made more concrete in Theorem 2 below.

Due to the way Eq. (18) and Eq. (19) allow for bounds to be tailored to the QoI, \( F \), we call any such \( R(F, \mathcal{G} || F, \mathcal{G}) \) a goal-oriented relative entropy associated with \( F \) and refer to this general approach, including the use of goal-oriented Fisher information that we introduce in Section 3.1, as goal-oriented information theory.

Proof First write
\[
\pm (E_{\mathcal{G}}[F] - E_P[F]) = \pm \left( \int_{\mathbb{R}} z(F, \tilde{P})(dz) - \int_{\mathbb{R}} z(F, P)(dz) \right)
\]
and then apply Proposition 1 to the right hand side of Eq. (20):
\[
\pm (E_{\mathcal{G}}[F] - E_P[F]) \leq \inf_{c>0} \left\{ \frac{1}{c} A^{\mathcal{G}}_{F, P}(\pm c) + \frac{1}{c} R(F, \tilde{P} || F, P) \right\}.
\]
Here, \( \text{id} \) denotes the identity function on \( \mathbb{R} \).

Recalling the definition of \( \Lambda \), Eq. (11), we find
\[
A^{\mathcal{G}}_{F, P}(\pm c) = \log E_{F, P}[e^{\pm c(\text{id} - E_{F, P}[\mathcal{G}])}] = \log E_{F, P}[e^{\pm c(F - E_{F, P}[F])}] = A^{\mathcal{G}}_{F, P}(\pm c).
\]
If \( F \) is \( \mathcal{G} \)-measurable for some sub sigma-algebra, \( \mathcal{G} \subset F \), then \( F, P = F, (P|\mathcal{G}) \), and similarly for \( \tilde{P} \). Eq. (19) then follows from the data processing inequality (see Theorem 14 in [38]). \( \square \)

Remark 2 The data processing inequality holds in much greater generality than the form we used above. In particular, it holds for any \( f \)-divergence [38]. For relative entropy, the data-processing inequality can be obtained from the chain rule, together with the fact that marginalization reduces the relative entropy.

It is also useful to have an uncentered version of Theorem 1:

Corollary 1 Let \( F : \Omega \to \mathbb{R}, F \in L^1(\mathcal{P}) \). Then
\[
\pm E_{\mathcal{P}}[F] \leq \inf_{c>0} \left\{ \frac{1}{c} A^F_{\mathcal{P}}(\pm c) + \frac{1}{c} R(F, \tilde{P} || F, P) \right\}.
\]

Remark 3 In many cases, Corollary 1 can be used to prove that \( F \in L^1(\mathcal{P}) \), by first truncating \( F \), so that Eq. (23) can be applied to a bounded QoI, and then taking limits using the dominated and/or monotone convergence theorems. Such a truncation argument is also how one obtains Corollary 1, in its full generality, from Theorem 1.

Eq. (23) is also useful when the desired QoI for the alternative model differs from that of the base model; in such cases, working with the \( P \)-centered QoI, \( \tilde{F} \), is inconvenient. This situation arises in the example in Section 5.5.
As stated above, relaxing the bounds via the data processing inequality, Eq. (19), is useful only when $R(\bar{P}|_\mathcal{G}| P|_\mathcal{G})$ is computable in some sense. More specifically, our goal will be to find (the smallest) $\mathcal{G}$ for which we have a 'sufficiently simple' $\mathcal{G}$-measurable function, $G$, that satisfies $R(\bar{P}|_\mathcal{G}| P|_\mathcal{G}) \leq E_{\bar{P}}[G]$. Given this, the following theorem provides robustness bounds on the QoI that are expressed solely in terms of expectations under the base model, $P$. As we are thinking of $P$ as the tractable model, these robustness bounds are often computable.

**Theorem 2** Let $F : \Omega \to \mathbb{R}$, $F \in L^1(\bar{P})$ be $\mathcal{G}$-measurable, where $\mathcal{G} \subset \mathcal{F}$ is a sub sigma-algebra, and suppose

$$R(\bar{P}|_\mathcal{G}| P|_\mathcal{G}) \leq E_{\bar{P}}[G]$$

(24)

for some $\mathcal{G}$-measurable $G \in L^1(\bar{P})$. Then

$$\pm E_{\bar{P}}[F] \leq \inf_{c > 0} \left\{ \frac{1}{c} \log E_{\bar{P}}[\exp(\pm c F + G)] \right\}. \tag{25}$$

**Proof** Given $c_0 > 0$, apply Corollary 1, the data processing inequality, and the assumption Eq. (24) to the $\mathcal{G}$-measurable QoI, $F_{c_0} \equiv F \pm c_0^{-1} G$:

$$\pm E_{\bar{P}}[F \pm c_0^{-1} G] = \inf_{c > 0} \left\{ \frac{1}{c} A_{\bar{P}}^{F_{c_0}}(\pm c) + \frac{1}{c} R(\bar{P}|_\mathcal{G}| P|_\mathcal{G}) \right\} \tag{26}$$

$$\leq \inf_{c > 0} \left\{ \frac{1}{c} A_{\bar{P}}^{F_{c_0}}(\pm c) + \frac{1}{c} E_{\bar{P}}[G] \right\}.$$

In particular, bounding the right-hand-side by the value at $c = c_0$, we can cancel the $E_{\bar{P}}[G]$ term (which is assumed to be finite) to find

$$\pm E_{\bar{P}}[F] \leq \frac{1}{c_0} A_{\bar{P}}^{F_{c_0}}(\pm c_0). \tag{27}$$

Taking the infimum over all $c_0 > 0$ yields the result. □

The bounds Eq. (25) are also tight, in the following sense:

**Corollary 2** Assume that $R(\bar{P}|_\mathcal{G}| P|_\mathcal{G}) < \infty$ and $c \to \log E_{\bar{P}}[\exp(c F)]$ is finite on a neighborhood of 0. Then there exists a $\mathcal{G}$-measurable $G$ such that Eq. (25) is an equality.

**Proof** Let $G = \log(d\bar{P}|_\mathcal{G}| dP|_\mathcal{G})$. The assumptions, together with Eq. (25), imply $G \in L^1(\bar{P})$ and we have

$$\pm E_{\bar{P}}[F] \leq \inf_{c > 0} \left\{ \frac{1}{c} \log E_{\bar{P}}[\exp(\pm c F + G)] \right\} = \lim_{c \searrow 0} \frac{1}{c} \log E_{\bar{P}}[\exp(\pm c F)] \tag{28}$$

$$= \frac{d}{dc} \Big|_{c=0} \log E_{\bar{P}}[\exp(\pm c F)] = \pm E_{\bar{P}}[F].$$

Note that in the first line, we used the fact that $c \to A_{\bar{P}}^{F_{c_0}}(\pm c)/c$ are increasing on $c > 0$. □

**Remark 4** The choice $G = \log(d\bar{P}|_\mathcal{G}| dP|_\mathcal{G})$ is rarely practical, but the fact that Eq. (25) is tight for appropriately chosen $\mathcal{G}$ and $G$ suggests that Theorem 2 has promise as a tool for finding bounds that are both tractable and not overly pessimistic.
Remark 5 Once a $G$ satisfying the hypotheses of Theorem 2 has been identified, one way to improve the bounds is to take any $G$-measurable, $\tilde{P}$-mean zero function, $H$, replace $G$ with $G + \alpha H$ in Eq. (25), and minimize over $\alpha \in \mathbb{R}$.

Another technique for improving the bounds is available when $P$ is part of a tractable parametric family, $\{P_\theta\}_{\theta \in U}$. If one can carry out the procedure of Theorem 2 for each $\theta$ then one obtains

$$\pm E_{\tilde{P}}[F] \leq \inf_{c > 0} \left\{ \frac{1}{c} \inf_{\theta \in U} \log E_{P_\theta} [\exp (\pm c F + G_\theta)] \right\}. \quad (29)$$

See Figure 7 in Appendix B for an example that employs this idea.

3.1 Parametric Models, Goal-Oriented Fisher Information, and Linearized UQ Bounds

Theorem 1, Corollary 1, or Theorem 2 describe methods for deriving non-perturbative UQ bounds, i.e. they are exact bounds and do not require the base and alternative model to be part of a ‘nice’ parametric family or to be ‘close’. Non-perturbative bounds are our primary focus in this work, however, linearized bounds are a useful tool when one does have a parametric family. Here we give a short discussion regarding how the above ideas can be adapted to derive goal-oriented sensitivity bounds for parametric models.

Consider a parametric model $\{P_\theta\}_{\theta \in U}$, where $U$ is an open subset of $\mathbb{R}^n$. For $\theta \in U$ we consider the base model $P \equiv P_\theta$, and an alternative model $\tilde{P} \equiv P_\theta + h$. $E_\theta$ will denote the expectation with respect to $P_\theta$.

By combining the usual notion of Fisher information (see, for example, [45]) with the same ideas that motivated the use of goal-oriented relative entropy in Theorem 1, we arrive at sensitivity bounds that involve goal-oriented variants of Fisher information. More specifically, if the QoI, $F$, is measurable with respect to a sub sigma-algebra $G \subset \mathcal{F}$, then one can rewrite $E_\theta[F]$ as an expectation with respect to $P_\theta|_G$. This leads to:

**Theorem 3** Assume the existence of a density $dP_\theta|_G = p^G_\theta \, d\mu$ with respect to some fixed $\mu$. Under appropriate assumptions on $p^G_\theta$ (detailed in Section 1.5 of the online supplement) one has the following:

Define the goal-oriented score,

$$V^G_\theta \equiv 1_{\mu^G_\theta \neq 0} \nabla p^G_\theta / p^G_\theta = 1_{\mu^G_\theta \neq 0} \nabla \log(p^G_\theta), \quad (30)$$

and the goal-oriented Fisher information in the direction $v \in \mathbb{R}^n$,

$$I^G_\theta (v) \equiv E_\theta [(v \cdot V^G_\theta)^2] = E_\theta [1_{\mu^G_\theta \neq 0} (\nabla_v p^G_\theta / \mu^G_\theta)^2]. \quad (31)$$

Then for all $\theta \in U$ we have:

1. $E_\theta [V^G_\theta] = 0. \quad (32)$
Eq. (190) results in the linearized UQ bound:

$$|E_{\theta} + h[F] - E_{\theta}[F]| \leq \sqrt{\text{Var}_{P_{\theta}}[F]} I_{\theta}^G(h) + o(\|h\|).$$

Again, for detailed assumptions, and the proof, see Section 1.5 of the online supplement.

3.2 A Guide to Using the Goal-Oriented UQ Bounds

All further UQ bounds in this paper will be based upon the goal-oriented information inequality, as expressed by one of Theorem 1, Corollary 1, or Theorem 2. A guide to using these results is given below.

**Workflow Outline:**

1. Given a QoI, $F$, a base model, $P$, and an alternative model, $\tilde{P}$, first identify a sub sigma-algebra, $G$, with respect to which $F$ is measurable. The sigma algebra $G$ should be as small as possible, while keeping in mind that the subsequent steps of this procedure must be feasible as well.

2. Identify a $G$-measurable function, $G$, that satisfies $R(\tilde{P}|G||P|G) \leq E_{\tilde{P}}[G]$. Again, one wants a tight bound, but $G$ must also be simple enough to make following steps tractable.

3. Obtain robustness bounds that do not involve $\tilde{P}$, either by using Theorem 2, or by separately bounding $E_{\tilde{P}}[G]$ (see Section 4.1), and then substituting the result into the goal-oriented information inequality, Theorem 1.

4. Bound the resulting $P$-expectations. Depending on the choices made above, this may or may not be tractable.

   In the examples we treat in Section 5, these (moment generating functions) will be expressible in terms of explicit, low-dimensional integrals, which are then computed numerically, but Monte Carlo or other methods can also be used. Another option for bounding the MGF of a bounded QoI is to utilize general-purpose inequalities, such as the Hoeffding, Bernstein, or Bennett inequalities (see [17, 10]). These yield bounds on the MGF in terms of the $\infty$-norm, mean, and variance of the QoI under $P$. For use of these in UQ, see [29, 7].

5. Solve the resulting optimization problem; the optimization is over a 1-D parameter, and is typically simple to perform numerically.

Even with the above outline, it is not always transparent how each step is to be accomplished in any given problem. A general framework for time-dependent QoI’s up to a stopping time, our primary area of application, will be given in the following section.

4 Goal-Oriented UQ for QoI’s up to a Stopping Time

The applications of the above general theory that we treat in Section 5 can all be thought of as involving time-dependent QoI’s up to a stopping time. Many important problem types fit under this umbrella, such as time averages of observables,
discounted observables, and the expectation and distribution of stopping times; we will study each of these, and more, starting in Section 4.1. The general setting in which we work is as follows:

**Assumption 1** Suppose:

1. \((\Omega, \mathcal{F}_\infty, \{\mathcal{F}_t\}_{t \in T})\) is a filtered probability space, where \(T = [0, \infty)\), or \(T = \mathbb{Z}_0\). We define \(T \equiv T \cup \{\infty\}\).
2. \(P\) and \(\tilde{P}\) are probability measures on \((\Omega, \mathcal{F}_\infty)\).
3. \(F: \mathbb{R} \times \Omega \to \mathbb{R}\), written \(F_t(\omega)\), is progressively measurable (progressive), i.e. \(F\) is measurable and \(F|[0,t] \times \Omega\) is \(\mathcal{B}([0, t]) \otimes \mathcal{F}_t\)-measurable for all \(t \in T\) (intervals refer to subsets of \(T\)).
4. \(\tau: \Omega \to \mathbb{T}\) is a \(\mathcal{F}_t\)-stopping time.

The quantity-of-interest we consider is then the stopped process, \(F_\tau\). First recall:

**Lemma 1** \(F_\tau\) is \(\mathcal{F}_\tau\)-measurable, where \(\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : \forall t \in \mathcal{T}, A \cap \{\tau \leq t\} \in \mathcal{F}_t\}\) (35) is the filtration up to the stopping-time \(\tau\), a sub sigma-algebra of \(\mathcal{F}_\infty\).

For a general stopped QoI, \(F_\tau\), we have the following (uncentered) UQ bounds, following from Corollary 1 with the choice \(\mathcal{G} = \mathcal{F}_\tau\):

**Corollary 3** Suppose \(F_\tau \in L^1(\tilde{P})\). Then

\[
\pm E_{\tilde{P}}[F_\tau] \leq \inf_{c > 0} \left\{ \frac{1}{c} A_{\tilde{P}}(\pm c) + \frac{1}{c} R((F_\tau)_\ast \tilde{P}|(F_\tau)_\ast P) \right\} 
\leq \inf_{c > 0} \left\{ \frac{1}{c} A_{\tilde{P}}(\pm c) + \frac{1}{c} R(\tilde{P}|\mathcal{F}_\tau)_\ast (P|\mathcal{F}_\tau) \right\}. 
\] (36)

One similarly has a centered variant (we omit the details), following from Theorem 1. From these results, one can see that for stopped QoI’s, \(R(\tilde{P}|\mathcal{F}_\tau)_\ast (P|\mathcal{F}_\tau)\) is an appropriate goal-oriented relative entropy.

Next, we use the above framework to derive more specialized UQ bounds for several classes of problems, essentially carrying out the steps of the ‘user’s guide’ from Section 3 for the QoI’s in Table 1.1. Not all of our examples will fit neatly into one of these categories, but rather, they illustrate several important use cases; the general tools that were developed in Section 3 are significantly more general and flexible. Consideration of specific models can be found in Section 5.

4.1 Bounding the Relative Entropy

First, we show how the above results can be used to bound a goal-oriented relative entropy. First consider the general case of a QoI, \(F\), that is \(\mathcal{G}\)-measurable for some sub sigma-algebra \(\mathcal{G} \subset \mathcal{F}\). We obtain the following as a consequence of Theorem 2 (after reparameterizing the infimum):
Corollary 4 Suppose $G : \Omega \to \mathbb{R}$, $G \in L^1(\tilde{P})$ is $\mathcal{G}$-measurable for some sub sigma-algebra $\mathcal{G} \subset \mathcal{F}$, and $R(\tilde{P}|_{\mathcal{G}}||P|_{\mathcal{G}}) \leq E_{\tilde{P}}[|G|]$. Then

$$R(\tilde{P}|_{\mathcal{G}}||P|_{\mathcal{G}}) \leq \inf_{c > 1} \left\{ (c - 1)^{-1} A_{\tilde{F}}^G(c) \right\}. \tag{37}$$

Note that requirement on the relative entropy is the same as in Theorem 2 (see Eq. (24)). Often, a useful function $G$ can be obtained from $\log(d\tilde{P}|_{\mathcal{G}}/dP|_{\mathcal{G}})$ by eliminating a $\tilde{P}$-mean-zero component; see the online supplement for examples. Also, note that the bound on $R(\tilde{P}|_{\mathcal{G}}||P|_{\mathcal{G}})$ in Eq. (37) only involves expectations under the (tractable) base model, $P$. This is in contrast to the assumed bound $R(\tilde{P}|_{\mathcal{G}}||P|_{\mathcal{G}}) \leq E_{\tilde{P}}[|G|]$, which involves the intractable model $\tilde{P}$.

When the QoI is a stopped process, $F_\tau$, and one takes $\mathcal{G} = \mathcal{F}_\tau$, then applying Corollary 4 requires one to find a progressive process, $G_t$, such that

$$R(\tilde{P}|_{\mathcal{F}_\tau}||P|_{\mathcal{F}_\tau}) \leq E_{\tilde{P}}[G_T]. \tag{38}$$

This can often be done for Markov processes when $\tau$ is bounded; again, see the online supplement. For unbounded $\tau$, one can often obtain results by working with the bounded stopping times $\tau \wedge T$ and then taking $T \to \infty$; see Lemma 3 and the examples in Section 5.

Remark 6 The utility of the above result is primarily in situations when the UQ bound Eq. (25) from Theorem 2 is still difficult to compute. In such cases, it can be useful to first bound $R(\tilde{P}|_{\mathcal{G}}||P|_{\mathcal{G}})$ via Eq. (37) and then use the UQ bound from Theorem 1 directly.

4.2 Distribution of a Stopping-Time

Next we show how to bound the distribution of a stopping time. First, consider the general case of a QoI of the form $F = 1_A$, where $A$ is some event. The cumulant generating function is

$$A_{\tilde{P}}^\Lambda(c) = \log \left( P(A)e^{c(1-P(A))} + (1-P(A))e^{-cP(A)} \right), \quad c \in \mathbb{R}. \tag{39}$$

Theorem 1 then gives the following UQ bound:

Corollary 5 Let $\mathcal{G}$ be a sub sigma-algebra of $\mathcal{F}$ and $A \in \mathcal{G}$. Then

$$\pm \left( \tilde{P}(A) - P(A) \right) \leq \inf_{c > 0} \left\{ \frac{1}{c} \log \left( P(A)e^{\pm c(1-P(A))} + (1-P(A))e^{-cP(A)} \right) + \frac{1}{c} R(\tilde{P}|_{\mathcal{G}}||P|_{\mathcal{G}}) \right\}. \tag{40}$$

In particular, given a stopping time, $\tau$, and $T > 0$, Corollary 5 can be applied to $A = \{\tau < T\}$ and $\mathcal{G} = \mathcal{F}_{\tau \wedge T}$, resulting in a bound on the distribution of $\tau$; for a concrete example, see Section 5.1.1.

Remark 7 Given (a numerical estimate of) the single number $P(A)$ (i.e. base-model probability) and a bound on the relative entropy (see the methods of Appendix A and the online supplement), the 1-D optimization problem in Eq. (5) is trivial to solve numerically. This can make Eq. (5) useful even when the base model is intractable analytically, as long as it is inexpensive to compute $P(A)$ via simulation.
Remark 8 UQ bounds based on Corollary 5 are far from optimal for very rare events. See [5,19] for a related approach to UQ using Rényi divergences instead of relative entropy.

4.3 Expectation of a Stopping Time

Here we consider the expectation of a stopping time, τ. For τ ∈ L¹(𝑃) ∩ L¹(̂𝑃), (the centered variant of) Corollary 3 applied to 𝐹ₜ = 𝑡 yields

\[ ±(E_{\tilde{P}}[\tau] - E_{P}[\tau]) \leq \inf_{c > 0} \left\{ \frac{1}{c} \Lambda_{\tilde{P}}^{c}(±c) + \frac{1}{c} R(\tilde{P}|_{F_{\tau}}||P|_{F_{\tau}}) \right\}. \]  (41)

For unbounded τ ∈ L¹(𝑃) ∩ L¹(̂𝑃), it is often useful to use Lemma 3 from Appendix A to obtain:

\[ ±(E_{\tilde{P}}[\tau] - E_{P}[\tau]) \leq \inf_{c > 0} \left\{ \frac{1}{c} \Lambda_{\tilde{P}}^{c}(±c) + \frac{1}{c} \liminf_{n \to \infty} R(\tilde{P}|_{F_{\tau \wedge n}}||P|_{F_{\tau \wedge n}}) \right\}. \]  (42)

In addition, if a relative entropy bound of the form

\[ R(\tilde{P}|_{F_{\tau \wedge n}}||P|_{F_{\tau \wedge n}}) \leq \eta_0 + KE_{\tilde{P}}[\tau \wedge n], \quad \eta_0, K \in [0, \infty) \]  (43)

holds for all n (this is not uncommon for Markov processes; see the online supplement), then one can use Theorem 2 with \( F = \tau \wedge n \) and \( G = \eta_0 + K(\tau \wedge n) \). Taking \( n \to \infty \) via the monotone and dominated convergence theorems, and reparameterizing the infimum, results in:

Corollary 6 Assume the relative entropy satisfies a bound of the form Eq. (43) for all \( n \in \mathbb{Z}^+ \). Then

\[ -\inf_{c > 0} \left\{ (c + K)^{-1} (\Lambda_{\tilde{P}}^{c}(-c) + \eta_0) \right\} \leq E_{\tilde{P}}[\tau] \leq \inf_{c > K} \left\{ (c - K)^{-1} (\Lambda_{\tilde{P}}^{c}(c) + \eta_0) \right\}. \]  (44)

We illustrate this technique on several concrete examples in Sections 5.1 and 5.2.

Remark 9 One does not need to assume \( E_{\tilde{P}}[\tau] < \infty \) to derive Eq. (44); if the given upper bound is proven to be finite, then that is sufficient to prove that \( E_{\tilde{P}}[\tau] \) is finite.

4.4 Time-Integral/Discounted QoI’s

A particular sub-case of interest is that of time-dependent QoI’s, \( F_t \), that are themselves the time-integral of some other process; here we derive an alternative (uncentered) UQ bound that is applicable in such circumstances. A centered version also holds, with obvious modifications.
Theorem 4 Let \( \pi \) be a sigma-finite positive measure on \( T \) and \( f : T \times \Omega \to \mathbb{R} \) be progressively measurable such that \( \int_T f_t(\omega) \pi(\,dt) \) exists (in \( \mathbb{R} \)) for all \( \omega \in \Omega \) and \( \int_{[0,t]} |f_s|\pi(\,ds) < \infty \) for all \( \omega \in \Omega \), \( t \in T \). For \( t \in T \) define the right-continuous progressive process

\[
F_t = \int_{[0,t]} f_s \pi(\,ds). \tag{45}
\]

Suppose we also have one of the following two conditions:

1. \( \pi \) is a finite measure and \( f_s \geq 0 \).
2. \( E^{\tilde{P}} \left[ \int_{[0,\tau]} |f_s| \pi(\,ds) \right] < \infty \). \tag{46}

Then

\[
\pm E^{\tilde{P}} [F_\tau] \leq \int_T \inf_{c>0} \left\{ \frac{1}{c} A^{1_{s\leq \tau} f_s}_P(\pm c) + \frac{1}{c} R((1_{s\leq \tau} f_s)_P, \tilde{P} || (1_{s\leq \tau} f_s)_P P) \right\} \pi(\,ds) \tag{47}
\]

\[
\leq \int_T \inf_{c>0} \left\{ \frac{1}{c} A^{1_{s\leq \tau} f_s}_P(\pm c) + \frac{1}{c} R(\tilde{P} | \mathcal{F}_{s\wedge \tau}, |P| | \mathcal{F}_{s\wedge \tau}) \right\} \pi(\,ds).
\]

Proof First suppose Eq. (46) holds. This allows for use of Fubini’s theorem:

\[
E^{\tilde{P}} [F_\tau] = \int_T E^{\tilde{P}} [1_{s\leq \tau} f_s] \pi(\,ds). \tag{48}
\]

Then

\[
\pm E^{\tilde{P}} [1_{s\leq \tau} f_s] \leq \int_T \inf_{c>0} \left\{ \frac{1}{c} A^{1_{s\leq \tau} f_s}_P(\pm c) + \frac{1}{c} R((1_{s\leq \tau} f_s)_P, \tilde{P} || (1_{s\leq \tau} f_s)_P P) \right\} \pi(\,ds).
\]

Combining these gives the result.

Remark 10 If \( \pi \) is a probability measure then one could alternatively apply Theorem 1 to the product measures \( \pi \times P \) and \( \pi \times \tilde{P} \). However, Jensen’s inequality implies that the bound Eq. (47) is tighter than the bound obtained by this alternative procedure.

Remark 11 When \( \tau = \infty \), if one attempts to use Corollary 3 in place of Theorem 4 then one generally finds that the relative-entropy term, \( R(\tilde{P} | \mathcal{F}_\infty || P | \mathcal{F}_\infty) \), is infinite and so the corresponding UQ bound is useless. On the other hand, the bound Eq. (47) often yields something nontrivial; see the example in Section 5.3 below.
Theorem 4 is especially useful for studying exponentially discounted QoI’s, where \( d\pi = \lambda e^{-\lambda t} dt \) for some \( \lambda > 0 \) and \( \tau = \infty \). Such QoIs are often used in control theory (see page 38 in [35]) and economics (see page 64 in [42] and page 147 in [44]). In this case, the bound becomes

\[
\pm E_P \left[ \int_0^\infty f_s \lambda e^{-\lambda s} ds \right] \leq \int_0^\infty \inf_{c > 0} \left\{ \frac{1}{c} A_P^F (\pm c) + \frac{1}{c T} R(P_{[\mathcal{F}_T]} || P_{[\mathcal{F}_T]}) \right\} \lambda e^{-\lambda s} ds. \tag{50}
\]

One can weaken either of the bounds Eq. (47) or Eq. (50) by pulling the infimum over \( c \) outside the integral. The bound then consists of two terms, an integrated moment generating function, and an integrated relative entropy. In the exponentially discounted case, the latter can be written as

\[
\eta \equiv \int_0^\infty R(P_{[\mathcal{F}_T]} || P_{[\mathcal{F}_T]}) \lambda e^{-\lambda s} ds. \tag{51}
\]

Remark 12 Eq. (51) is the same information-theoretic quantity, an exponentially discounted relative entropy, that was defined in [44] (see page 147), where it was proposed as a measure of model uncertainty for control problems. Theorem 4 provides a rigorous justification for its use in UQ for exponentially discounted QoI’s.

4.5 Time-Averages

Finally, we consider the case of a time-average of some observable: \( T = [0, \infty) \), \( T > 0 \), \( \pi = \frac{1}{T} 1_{[0,T]} dt \), \( \tau = T \), and

\[
F_T = F_T = \frac{1}{T} \int_0^T f_s ds. \tag{52}
\]

(One can similarly treat the discrete-time case.) UQ bounds for such QoI’s were treated in [34, 7] using non-goal-oriented relative entropy. Here we show how they fit into the general framework of Section 4, and that we recover the same bounds.

Supposing that \( F_T \in L^1(P) \cap L^1(\tilde{P}) \), Corollary 3 yields

\[
\pm \left( E_P [F_T] - E_P [F_T] \right) \leq \inf_{c > 0} \left\{ \frac{1}{c T} A_T^P (\pm c) + \frac{1}{c T} R(\tilde{P}_{[\mathcal{F}_T]} || P_{[\mathcal{F}_T]}) \right\}, \tag{53}
\]

where

\[
A_T^P (\pm c) = \log E_P \left[ \exp \left( \pm c \left( \int_0^T f_s ds - E_P \left[ \int_0^T f_s ds \right] \right) \right) \right]. \tag{54}
\]

and we reparameterized \( c \to cT \) in the infimum to better exhibit the behavior as \( T \to \infty \).

In particular, we see that an appropriate information-theoretic quantity for controlling the long-time behavior is the relative entropy rate,

\[
H(\tilde{P} || P) \equiv \limsup_{T \to \infty} \frac{1}{T} R(\tilde{P}_{[\mathcal{F}_T]} || P_{[\mathcal{F}_T]}), \tag{55}
\]

which is finite in many cases of interest. This is the same result as obtained in [20] by more specialized methods.
Remark 13 The time integral in the exponent of the expectation in Eq. (54) can yield a connection to the Feynman-Kac semigroup, and the possibility of bounding it via functional inequalities [7]. This is one reason why we use Corollary 3 here, rather than the time-integral method discussed in Section 4.4.

Combined with Remark 11 this demonstrates that, in cases where both methods are applicable, neither the general method of Corollary 3 nor the integral method of Section 4.4 is always superior.

5 Examples

We now apply the goal-oriented UQ methods developed above to several examples:

1. Hitting times for perturbations of Brownian motion in Section 5.1.
2. Hitting times for perturbations of an Ornstein-Uhlenbeck process in Section 5.2.
3. Expected cost in stochastic control systems in Section 5.3.
4. Long-time behavior of semi-Markov perturbations of an M/M/∞ queue in Section 5.4.
5. Pricing of American options with variable interest rate in Section 5.5.

5.1 Perturbed Brownian Motion

We first illustrate the use of our method, and several of its features, with a simple example wherein many computations can be done explicitly and exactly; the example consists of perturbations to Brownian-motion-with-constant-drift. Specifically, take the base model to be the distribution on path space, $C([0, \infty), \mathbb{R})$, of a $\mathbb{R}$-valued Wiener process (Brownian motion), $W_t$, with constant drift, $\mu > 0$, starting from 0, i.e. the distribution of the following process:

**Base Model:**

$$X_t = \mu t + W_t.$$  \hspace{1cm} (56)

The alternative model will be the distribution of the solution to a stochastic differential equation (SDE) of the form:

**Alternative Models:**

$$d\tilde{X}_t = (\mu + \beta(t, \tilde{X}_t, \tilde{Y}_t, \tilde{Z}))dt + d\tilde{W}_t, \quad \tilde{X}_0 = 0,$$  \hspace{1cm} (57)

where $\tilde{W}_t$ is also a Wiener process. We allow the drift perturbation, $\beta$, to depend on an additional $\mathbb{R}^k$-valued process, $\tilde{Y}_t$, and also on independent external data, $\tilde{Z}$ (see the online supplement for further discussion of the type of perturbations we have in mind). More specific assumptions on $\beta$ will be stated later.

In Section 5.1.1 we will study the distribution of $\tau_a$, the level-$a$ hitting time:

**QoI 1:**

$$\tilde{P}^0(\tau_a < T), \quad T > 0, \text{ where } \tau_a[y] = \inf\{t : y_t = a\}. \hspace{1cm} (58)$$

$\tau_a$ is a stopping time on path space, $C([0, \infty), \mathbb{R})$. $\tilde{P}^0$ denotes the distribution of $\tilde{X}_t$ on path space and similarly for $P^0$; the superscript 0 indicates that the initial condition is 0.

We will also derive UQ bounds for the expected level-$a$ hitting time,

**QoI 2:**

$$E_{\tilde{P}^0}[\tau_a], \quad a > 0. \hspace{1cm} (59)$$
Note that here, $\mu$ and $a$ must have the same sign in order to ensure the hitting time is a.s.-finite under the base model. We will obtain UQ bounds on Eq. (59) in Sections 5.1.2 and 5.1.3, under different assumptions on the alternative family of $\beta$’s.

5.1.1 Hitting Time Distribution: Benefits of Being Goal-Oriented

We use the QoI Eq. (58) to illustrate the benefits of using a goal-oriented relative entropy, versus a non-goal-oriented counterpart. This example uses the technique of Section 4.2.

Here we consider the following ambiguity set (i.e. the subset of models of the form Eq. (57) that will make up the alternative family, $\tilde{P}$):

**Ambiguity Set:** Perturbations by drifts, $\beta$, such that, for a specified $\alpha > 0$:

$$R(\tilde{P}_0|_{\mathcal{F}_{\tau_a \wedge T}}||\tilde{P}_0|_{\mathcal{F}_{\tau_a \wedge T}}) \leq \frac{1}{2} \alpha^2 E_{\tilde{P}_0}[\tau_a \wedge T]. \quad (60)$$

For example, Eq. (60) holds if $\|\beta\|_{\infty} \leq \alpha$ (assuming that the models satisfy the assumptions required to compute the relative entropy via Girsanov’s theorem; see the online supplement for details).

The QoI, $F \equiv 1_{\tau_a < T}$, is $\mathcal{F}_{\tau_a \wedge T}$-measurable, and so $R(\tilde{P}_0|_{\mathcal{F}_{\tau_a \wedge T}}||\tilde{P}_0|_{\mathcal{F}_{\tau_a \wedge T}})$ is an appropriate goal-oriented relative entropy; the bound Eq. (60) implies that we can take $\tilde{G} = \mathcal{F}_{\tau_a \wedge T}$ and $G = \frac{1}{2} \alpha^2(\tau_a \wedge T)$ in Theorem 2. This gives the following goal-oriented UQ bound:

$$\pm \tilde{P}_0(\tau_a < T) \leq \inf_{c>0} \left\{ \frac{1}{c} \log E_{\tilde{P}_0} \left[ \exp \left( \pm c1_{\tau_a < T} + \frac{1}{2} \alpha^2(\tau_a \wedge T) \right) \right] \right\}. \quad (61)$$

The QoI is a function of the path up to time $T$, so a non-goal-oriented bound can be obtained from Proposition 1 with the relative entropy $R(\tilde{P}_T||P^0_T)$ (the subscript $T$ denotes taking the distribution on path space $\mathcal{C}([0,T], \mathbb{R})$).

**Remark 14** One can also attempt to use the full distributions on $\mathcal{C}([0,\infty), \mathbb{R})$, but that only results in trivial bounds, as the maximum of $R(\tilde{P}_0||P^0)$ over the ambiguity set Eq. (60) is infinite; this can be seen by taking a constant perturbation $\beta = 0$.

The same reasoning that lead to Theorem 2 then gives the non-goal-oriented UQ bound (the same bound as obtained from the methods in [14,12,26,36,29,43,20,28,34,7])

$$\pm P^0(\tau_a < T) \leq \inf_{c>0} \left\{ \frac{1}{c} \log E_{P^0} \left[ \exp (\pm c1_{\tau_a < T}) \right] + \frac{1}{c} \alpha^2 T \right\}. \quad (62)$$

The above $P^0$-expectations can be computed using the known formula for the distribution of $\tau_a$ under the base model (see page 196 in [33] and also the online supplement). Assuming that $a > 0$ and $\mu > 0$, the goal-oriented bound is

$$\pm \tilde{P}^0(\tau_a < T) \leq \inf_{c>0} \left\{ \frac{1}{c} \log \left[ \frac{a}{\sqrt{2\pi}} \int_0^\infty e^{\pm c1_{z<T} + \frac{1}{2} \alpha^2(z \wedge T)} e^{-(a-\mu z)^2/(2z)z^{-3/2}dz} \right] \right\}. \quad (63)$$
Fig. 1 Bound on the distribution of $\tau_a$, the level-$a$ hitting time, for a perturbation of Brownian-motion-with-constant-drift. The solid black line shows the distribution under the base model. The red curves (dot-dashed, Eq. (62)) are obtained from a non-goal-oriented relative entropy, while the tighter bounds, in blue (dashed, Eq. (63)), were obtained by our method with an appropriate goal-oriented relative entropy.

Similarly, one can compute the non-goal-oriented bound, Eq. (62), and the probability under the base model.

Figure 1 shows numerical results comparing these bounds, with parameter values $\mu = 1$, $a = 2$, $\alpha = 0.2$ (here, and as a general rule, we solve the 1-D optimization problems numerically). The black curve shows the distribution of $\tau_a$ under the base model. The blue curves show the goal-oriented bounds Eq. (63); these constrain $\tilde{P}^0(\tau_a < T)$ to be within the gray region. The red curves are from the non-goal-oriented bounds Eq. (62); we see that the non-goal-oriented bounds significantly overestimate the uncertainty, especially as $\tilde{P}^0(\tau_a < T)$ approaches 1.

5.1.2 Expected Hitting Time: Optimality of UQ Bounds

As discussed in Corollary 2, our UQ method is capable of producing tight bounds, for appropriate choices of $G$ and $\mathcal{G}$ (but certainly not for all choices). We illustrate this fact with a simple example where optimal bounds can be obtained explicitly via our method.

Here, for a given $\alpha > 0$, we work with the ambiguity set consisting of

**Ambiguity Set:** perturbations by drifts, $\beta$, with $||\beta||_\infty \leq \alpha$, (64)

and consider the QoI Eq. (59) (the expected level-$a$ hitting time). Note that we must have $\alpha < \mu$ to ensure that the perturbed model has a.s.-finite hitting time. This example uses the technique of Section 4.3.

The optimal bounds can be obtained directly (i.e. without using the UQ methods developed above) by combining the known distribution of $\tau_a$ [33] with the comparison principle, i.e. with

$$(\mu - ||\beta||_\infty) t + \tilde{W}_t \leq \tilde{X}_t \leq (\mu + ||\beta||_\infty) t + \tilde{W}_t.$$ (65)
This yields
\[
\frac{a}{\mu + \alpha} \leq E_{\tilde{P}_0}[\tau_a] \leq \frac{a}{\mu - \alpha},
\] (66)
where upper and lower bounds are achieved in the cases of constant drift perturbations \(\alpha\) and \(-\alpha\) respectively.

We now compare Eq. (66) with goal-oriented UQ bounds. Girsanov’s theorem gives a bound on the relative entropy:
\[
R(\tilde{P}_0|_{\mathcal{F}_{\tau_a \land n}} || P_0|_{\mathcal{F}_{\tau_a \land n}}) \leq \frac{\alpha^2}{2} E_{\tilde{P}_0}[\tau_a \land n] (67)
\]
for all \(n\); in the language of Theorem 2, we can take \(G = \mathcal{F}_{\tau_a \land n}\) and \(G = \frac{\alpha^2}{2}(\tau_a \land n)\).

Using the known formula for the cumulant generating function of \(\tau_a\) (see Chapter 8 of [46]) and analytic continuation, the cumulant generating function in the base model can be computed:
\[
\Lambda_{\tau_a}^{P_0}(c) = a\mu - a\sqrt{\mu^2 - 2c}, \quad c < \mu^2/2. (68)
\]
Therefore Corollary 6 yields
\[
-\inf_{c>0} \left\{(c + K)^{-1} \left(a\mu - a\sqrt{\mu^2 + 2c}\right)\right\} \leq E_{\tilde{P}_0}[\tau_a] \leq \inf_{K < c < \mu^2/2} \left\{(c - K)^{-1} \left(a\mu - a\sqrt{\mu^2 - 2c}\right)\right\}, (69)
\]
where \(K \equiv \frac{1}{2} \alpha^2\).

The above optimization problems can be solved explicitly, with minimizers \(c^* = K + \sqrt{2K}\mu\) and \(c^* = -K + \sqrt{2K}\mu\) respectively. Computing the corresponding minimum values, one finds that the UQ bounds resulting from our method, Eq. (69), are the same as the optimal bounds, Eq. (66), that were obtained from the comparison principle. We also note that, while the comparison principle is only available in very specialized circumstances, our UQ method is quite general; in particular, our method is not restricted to 1-D systems. Finally, the non-goal-oriented UQ bounds, Eq. (13), are not effective here; the non-goal-oriented relative entropy over an infinite time-horizon is \(R(\tilde{P}_0| P_0)\), which is infinite for constant drift perturbations \(\beta = \pm \alpha\) (which are part of the ambiguity set being considered here, Eq. (64)).

5.1.3 Time-Dependent Envelope

Now we show that the goal-oriented information inequality improves on the comparison-principle bounds of Eq. (66) when one has further information on \(\beta\). Specifically, we work with the following ambiguity set:

**Ambiguity Set:** Perturbations by drifts, \(\beta\), with \(|\beta(t, \cdot, \cdot, \cdot)| \leq h(t), \quad t \geq 0\),

for a given \(h \in L^2_{loc}([0, \infty))\). (70)

This example will require a variant of the technique from Section 4.3.
Fig. 2 Bound on the expected level-$a$ hitting time, for a perturbation of Brownian-motion-with-constant-drift. The parameter $\beta$ denotes the drift perturbation and we are assuming $|\beta(t, x)| \leq \alpha |\cos(t)|$. The red lines are obtained from the comparison principle, Eq. (65), applied to $|\beta|_\infty = \alpha$. The tighter bounds, in blue, were obtained by our method, Eq. (73), using the assumed time-dependent envelope; they constrain the QoI to the gray region. The asymptote of the comparison-principle upper bound occurs at $\alpha = \mu$, at which point the hitting time for $(\mu - |\beta|_\infty)t + \tilde{W}_t - \tilde{W}_1$ has infinite expected value.

An appropriate sub sigma-algebra is $\mathcal{G} = \mathcal{F}_{\tau_a \wedge n}$, for which Girsanov’s theorem gives

$$ R(\tilde{P}_0|_{\mathcal{F}_{\tau_a \wedge n}}, |P^0|_{\mathcal{F}_{\tau_a \wedge n}}) \leq E_{\tilde{P}_0}[G], \quad G \equiv \frac{1}{2} \int_0^{\tau_a \wedge n} |h(s)|^2 ds. \quad (71) $$

Combining this with Theorem 2, we find

$$ \pm E_{\tilde{P}_0}[\tau_a \wedge n] \leq \inf_{c > 0} \left\{ \frac{1}{c} \log E_{P^0} \left[ \exp(\pm c\tau_a \wedge n + \frac{1}{2} \int_0^{\tau_a \wedge n} |h(s)|^2 ds) \right] \right\}. \quad (72) $$

Using the monotone convergence theorem on both sides of the upper bound, we can take $n \to \infty$. If the resulting upper bound is finite then this justifies the use of the dominated convergence theorem on the lower bound. The end result is

$$ \pm E_{\tilde{P}_0}[\tau_a] \leq \inf_{c > 0} \left\{ \frac{1}{c} \log E_{P^0} \left[ \exp \left( \pm c\tau_a + \frac{1}{2} \int_0^{\tau_a} |h(s)|^2 ds \right) \right] \right\}. \quad (73) $$

Note that the same technique can lead to UQ bounds on $g(\tau_a)$ for more general $g$.

The expectations in Eq. (73) can again be evaluated using the distribution of $\tau_a$ under $P$:

$$ E_{P^0} \left[ \exp \left( \pm c\tau_a + \frac{1}{2} \int_0^{\tau_a} |h(s)|^2 ds \right) \right] = \frac{a}{\sqrt{2\pi}} \int_0^{\infty} \exp \left( \pm ct + \frac{1}{2} \int_0^t |h(s)|^2 ds - (a - \mu t)^2/(2t) \right) t^{-3/2} dt. \quad (74) $$

In Figure 2 we show the bounds Eq. (73) for the case of an oscillating envelope of the form $h(t) = \alpha |\cos(t)|$, and with $a = 3$, $\mu = 2$. Note that our method (dashed
curves) is an improvement over the comparison-principle bounds of Eq. (66) (dot-dashed curves), and it works for large perturbations, far from the linear regime. Again, the corresponding non-goal-oriented bounds are generally trivial (infinite). Finally, note that our method is applicable in dimension greater than 1, while the comparison principle is generally not.

5.2 Ornstein-Uhlenbeck Hitting Times

Another base process that is amenable to exact hitting-time calculations is the solution to an Ornstein-Uhlenbeck equation:

**Base Model:**

\[ dX_t^x = -\lambda X_t^x dt + dW_t, \quad X_0^x = x \in \mathbb{R}, \quad \lambda > 0. \] (75)

We perturb it by a bounded drift

**Alternative Models:**

\[ d\tilde{X}_t^x = (-\lambda \tilde{X}_t^x + \beta(t, \tilde{X}_t^x, \tilde{Y}_t, \tilde{Z}_t)) dt + d\tilde{W}_t, \] (76)

**Ambiguity Set:** perturbations by drifts, \( \beta \), with \( \| \beta \|_\infty \leq \alpha \),

\[ \text{QoI: } \tau_a[y] = \inf\{t : y_t = a\}, \quad x < a. \] (78)

Again, this example uses the technique from Section 4.3.

Let \( P_x \) denote the distribution of \( X^x \) on path space. The moment generating function of \( \tau_a \) under \( P_x \) for \( c < 0 \) is (see [2])

\[ E_{P_x}[e^{c\tau_a}] = \frac{H_{c/\lambda}(-x\sqrt{\lambda})}{H_{c/\lambda}(-a\sqrt{\lambda})}, \] (79)

where \( H_\nu(z) \) are the Hermite functions. These can be written in terms of hypergeometric functions:

\[ H_\nu(z) = 2^{\nu} \sqrt{\pi} \left( \frac{1}{\Gamma\left(1 - \nu^2/4\right)} F_1\left(-\frac{\nu}{2}, \frac{3}{2}, z^2\right) - \frac{2z}{\Gamma\left(-\nu/2\right)} F_1\left(1 - \nu/2, \frac{3}{2}, z^2\right) \right). \] (80)

\( H_\nu(z) \) are entire functions in \( \nu \), hence if we let \( c^* \) be the first positive zero of \( c \to H_{c/\lambda}(-a\sqrt{\lambda}) \) then by analytic continuation, the equality Eq. (79) extends to \( c \in (-\infty, c^*) \).

Once again, the appropriate sub sigma-algebra is \( G = \mathcal{F}_{\tau_a \wedge n} \) and, using the relative entropy bound

\[ R(\tilde{P}^x|_{\mathcal{F}_{\tau_a \wedge n}}, \|P^x|_{\mathcal{F}_{\tau_a \wedge n}}) \leq E_{\tilde{P}_x}[G], \quad G \equiv \frac{1}{2} \alpha^2 \tau_a \wedge n, \] (81)

Corollary 6 yields

\[ -\inf_{c > 0} \left\{ (c + K)^{-1} A_{P_x}^c (-c) \right\} \leq E_{\tilde{P}_x}[\tau_a] \leq \inf_{K < c < c^*} \left\{ (c - K)^{-1} A_{P_x}^c (c) \right\}, \] (82)

where \( K \equiv \alpha^2 / 2 \); the moment generating function is given by Eq. (79).
Figure 3 shows the bound on the expected hitting time, using the base-model parameter values $\lambda = 1/2$, $x = 0$, and both $a = 1$ and $a = 2$. This is another test case where the comparison principle furnishes us with optimal bounds, based on a given value of $\alpha$ (red dot-dashed lines). Unlike in the Brownian motion example from Section 5.1.2, here our method does not reproduce the comparison-principle based bounds. In particular, our upper bound has an asymptote at a finite value of $\alpha$ that does not correspond to any singularity in $E_{\tilde{P}_x}[\tau_a]$.

The observed singularity can be traced to the strength of the perturbation, $K$, approaching $e^c$, the point at which the moment generating function of $\tau_a$ becomes infinite. As the MGF of $\tau_a$ at $K$ forms an essential part of our bounds (in the language of Theorem 2, here we are essentially using $G = K\tau_a$), such singularities in the MGF that are not reflected in the alternative-model-expectation of the QoI will result in overly pessimistic bounds when the perturbation becomes too large; this is certainly a drawback of our method. As a counterpoint, we again note that our method is applicable in dimension greater than 1, while the comparison principle is generally not.

Note that the UQ bound Eq. (82) remains the same if one enlarges to the ambiguity set defined by the relative entropy bound Eq. (81). It is plausible that one can perturb the drift in an unbounded manner (such as by perturbing $\lambda$) while maintaining a relative entropy bound of this form, and thereby obtain explosion in the expectation of $\tau_a$ at finite $\alpha$; pursuing this direction further is not our concern here.
5.3 Discounted Quadratic Observable for Non-Gaussian Perturbations of a Gaussian Process

Here we consider a $\mathbb{R}^n$-valued progressively measurable process, $X_t$, that has a Gaussian distribution for every $t$:

**Base Model:**

$$(X_t), P = N(\mu_t, \Sigma_t), \quad t \geq 0. \quad (83)$$

We assume $\Sigma_t$ is (strictly) positive-definite and study the discounted QoI:

**QoI:**

$$F \equiv \int_T \left( \frac{1}{2} \Sigma_t C_s \hat{X}_s + d_s \hat{X}_s \right) \pi(ds), \quad (84)$$

where $C_s$ is a positive-definite $n \times n$-matrix valued function, $d_s$ is $\mathbb{R}^n$-valued, and $\pi$ is a probability measure. Without loss of generality (i.e. neglecting an overall constant factor and after a redefinition of $d_s$), we have chosen to express $F$ in terms of the centered process $\hat{X}_s = X_s - \mu_s$. The techniques involved here are those of Section 4.4.

At this point, we are being purposefully vague about the nature of the alternative model and even the details of the base model, as a large portion of the following computation is independent of these details. We also emphasize that the alternative models do not need to be Gaussian. In Section 5.3.1 we will make both base and alternative models concrete, as we study an application to control theory.

Assuming that the conditions of Theorem 4 are met, we obtain the following UQ bound (note that the appropriate sub sigma-algebras are $\mathcal{G} = \mathcal{F}_s$):

$$\pm \mathbb{E}_P[F] \leq \int_{c > 0} \inf_{T < c} \left\{ \frac{1}{c} A_P^f(\pm c) + \frac{1}{c} R((f_s)_* P[[((f_s)_* P)]) \right\} \pi(ds), \quad (85)$$

$$f_s \equiv \frac{1}{2} \hat{X}_s^T C_s \hat{X}_s + d_s^T \hat{X}_s.$$

$X_t$ is Gaussian under $P$, so we can compute

$$A_P^f(c) = \log \left( (2\pi)^n |\det(\Sigma_s)| \right)^{-1/2} \int e^{c d_s^T z - \frac{1}{2} z^T (\Sigma_s^{-1} - c C_s) z} d_s, \quad (86)$$

$$\mathbb{E}_P[F] = \int \frac{1}{2} \text{tr}(\Sigma_s C_s) \pi(ds).$$

For $c$ small enough, $\Sigma_s^{-1} - c C_s$ is positive-definite. More specifically, $C_s$ is positive-definite, so we can compute a Cholesky factorization $C_s = M_s^T M_s$ with $M_s$ nonsingular. $\Sigma_s^{-1} - c C_s$ is positive-definite iff $y^T (M_s^{-1})^T \Sigma_s^{-1} M_s^{-1} y > c y^T y$ for all nonzero $y$ iff $c < 1/\|M_s \Sigma_s M_s^T\|_2$ ($\ell^2$-matrix norm).

For $c < 1/\|M_s \Sigma_s M_s^T\|_2 \equiv C_s^*$, the integral in Eq. (86) can be computed in terms of the moment generating function of a Gaussian with covariance $(\Sigma_s^{-1} - c C_s)^{-1}$, resulting in the UQ bound

$$-\int_{T < c} \inf_{c > 0} \left\{ \frac{1}{c} A_P^f(-c) + \frac{\eta_s}{c} \right\} \pi(ds) \leq \mathbb{E}_P[F] \leq \int_{0 < c < c^*} \inf_{c > 0} \left\{ \frac{1}{c} A_P^f(c) + \frac{\eta_s}{c} \right\} \pi(ds), \quad (87)$$
where $\eta_s \equiv R((f_s, \tilde{P})|(f_s), P)$ and for the above ranges of $c$ we have

$$A^p_f(\pm c) = -\frac{1}{2} \log (|\det (I \mp c\Sigma_s C_s)|) + \frac{1}{2} c^2 d^T_s (I \mp c\Sigma_s C_s)^{-1} \Sigma_s d_s.$$ 

To make these bounds usable, one needs a bound on the relative entropy; this requires more specificity regarding the nature of the family of alternative models. We study an application to control theory in the following subsection.

### 5.3.1 Linear-Quadratic Stochastic Control

Here, Eq. (87) will be used to study robustness for linear-quadratic stochastic control (robustness under nonlinear perturbations). Specifically, suppose one is interested in controlling some nonlinear system,

$$d\tilde{X}_t = (B\tilde{X}_t + \sigma\beta(t, \tilde{X}_t) + Du_t)dt + \sigma dW_t, \quad \tilde{X}_0 \sim N(0, \Sigma_0),$$

where $B, D, \sigma$ are $n \times n$ matrices, and $W$ is a $\mathbb{R}^n$-valued Wiener process. We write the non-linear term with an explicit factor of $\sigma$ to simplify the use of Girsanov’s theorem later on. The control variable is denoted by $u_t$ and $\tilde{X}_t$ is the state; we take $\tilde{X}_t$ to be an observable quantity that can be used for feedback.

$\beta$ may be unknown, and even when it is known, optimal control of Eq. (88) is a difficult problem, both analytically and numerically [35]. Therefore, one option is to consider the linear approximation

$$dX_t = (BX_t + Du_t)dt + \sigma dW_t, \quad X_0 \sim N(0, \Sigma_0),$$

obtain an explicit formula for the optimal feedback control (under a cost functional to be specified below) for Eq. (89), and use that same feedback function to (suboptimally) control the original system Eq. (88). To evaluate the performance, one must bound the cost functional when the control for the linearized system is used on the nonlinear system.

To make the above precise, we must first specify the cost functional, which we take to be an exponentially discounted quadratic cost:

$$C = E \left[ \int_0^\infty \frac{1}{2} \left( X_t^T QX_t + u_t^T Ru_t \right) e^{-\lambda t} dt \right],$$

(and similarly with $X_t$ replaced by $\tilde{X}_t$ for the nonlinear system). Here, $Q$ and $R$ are positive-definite $n \times n$-matrices.

The optimal feedback control for the the linear system Eq. (89) with cost Eq. (90) is given by $u_t = K\lambda X_t$ where $K\lambda$ is obtained as follows (see [3, 6] for details): Define $B\lambda \equiv B - \frac{1}{2} I$ and solve the following algebraic Riccati equation for $Y\lambda$:

$$B^T\lambda Y\lambda + Y\lambda B\lambda + Q - Y\lambda DR^{-1} D^T Y\lambda = 0.$$ 

The optimal control is then

$$u_t = K\lambda X_t, \quad K\lambda \equiv -R^{-1} D^T Y\lambda.$$
Having obtained Eq. (92), we can finally fit the above problem into our UQ framework:

**Base Model:**
\[
dX_t = A_\lambda X_t dt + \sigma dW_t, \quad A_\lambda \equiv B + DK_\lambda, \quad X_0 \sim N(0, \Sigma_0)
\]  
(93)

**Alternative Models:**
\[
d\tilde{X}_t = (A_\lambda \tilde{X}_t + \sigma \beta(t, \tilde{X}_t)) dt + \sigma dW_t, \quad \tilde{X}_0 \sim N(0, \Sigma_0)
\]  
(94)

**Ambiguity Set:** Perturbations by $\beta$ such that $R(\tilde{P}|F_s||P|F_s) \leq \alpha^2 s^2$ for all $s \geq 0$.

(95)

**QoI:**
\[
F \equiv \int_0^\infty \frac{1}{2} x^T C_\lambda x e^{-\lambda s} ds, \quad C_\lambda \equiv Q + K_\lambda^T R K_\lambda,
\]  
(96)

where $x \in C([0, \infty), \mathbb{R}^n)$ is the path of the state variable.

Note that, by Girsanov’s theorem, the above ambiguity set contains all bounded-drift-perturbations of the base model with $\|\beta\|_\infty \leq \alpha$.

The UQ bounds Eq. (87) applied to the above problem are bounds on the cost of controlling the nonlinear system Eq. (88) by using the feedback control function that was derived to optimally control the linear system Eq. (89). We can make the bounds explicit as follows:

Eq. (93) has the solution
\[
X_t = e^{tA_\lambda} \left( X_0 + \int_0^t e^{-sA_\lambda} \sigma dW_s \right).
\]  
(97)

In particular, $X_t$ is Gaussian with mean 0 and covariance
\[
\Sigma_t^{\alpha\beta} = (e^{tA_\lambda})^\alpha (e^{tA_\lambda})^\beta \Sigma_0 + (e^{tA_\lambda})^\alpha \int_0^t (e^{-sA_\lambda} \sigma)^i \delta^{ik} (e^{-sA_\lambda} \sigma)^k ds.
\]  
(98)

Next, compute a Cholesky factorization of $C_\lambda$, $C_\lambda = M_\lambda^T M_\lambda$. The integrand in Eq. (90) is non-negative, so the hypotheses of Theorem 4 are met and we are justified in using Eq. (87) to obtain
\[
-\int_0^\infty \inf_{c>0} \left\{ -\frac{1}{2c} \log(|\det(I + c\Sigma_s C_\lambda)|) + \frac{1}{2c} \alpha^2 s \right\} \pi(ds)
\leq E_\tilde{P} [F] \leq \int_0^\infty \inf_{0 < c < 1/\|M_\lambda \Sigma_s M_\lambda^T\|} \left\{ -\frac{1}{2c} \log(|\det(I - c\Sigma_s C_\lambda)|) + \frac{1}{2c} \alpha^2 s \right\} \pi(ds).
\]  
(99)

Figure 4 shows numerical results, corresponding to the following simple example system, with a 2-D state variable and a 1-D control variable:
\[
B = \begin{bmatrix} 2 & 0.1 \\ 0.1 & -1 \end{bmatrix}, \quad D = \begin{bmatrix} \kappa \\ 0 \end{bmatrix}.
\]  
(100)

$Q = I$, $R = 1$, $\Sigma_0 = 0$, $\lambda = 1/2$, $\sigma = I$, $\alpha = 1/2$. Note that the size of the perturbation, $\sigma \beta$, is not required to be ‘small’ for the method to produce non-trivial bounds, as this is a non-perturbative calculation. Finally, if one instead attempts to use the non-goal-oriented relative entropy to derive UQ bounds, the results are again trivial (infinite), due to the unbounded time-horizon.
Fig. 4  Bound on the expected quadratic cost, Eq. (96), for a nonlinear stochastic control system (alternative model) Eq. (88), (suboptimally) controlled via the optimal control for an approximating linear system (base model) Eq. (89). The blue dashed lines, computed from Eq. (99), constrain the QoI to the gray region. The parameter $\kappa$ (defined in Eq. (100)) governs the strength of the controller.

5.4 Semi-Markov Perturbations of a $M/M/\infty$ Queue

Continuous-time jump processes with non-exponential waiting times (i.e. waiting times with memory) are often approximated by a much simpler (though often unrealistic [13]) exponentially-distributed waiting time model. Here we derive robustness bounds for such approximations.

More specifically, in this example the base model will be a $M/M/\infty$ queue (i.e. continuous-time birth-death process) with arrival rate $\alpha > 0$ and service rate $\rho > 0$:

**Base Model:**

\[
P(X_{t+dt} = x + 1 | X_t = x) = \alpha dt, \]
\[
P(X_{t+dt} = x - 1 | X_t = x) = \rho x dt, \quad x \in \mathbb{Z}_0.
\]

The embedded jump Markov-chain has transition probabilities, $a(x, y)$, given by

\[a(x, x + 1) = \alpha / (\alpha + \rho x), \quad a(x, x - 1) = \rho x / (\alpha + \rho x),\]

and the waiting-times are exponentially distributed with jump rates

\[\lambda(x) = \alpha + \rho x.\]

We consider the stationary case, i.e. $X_0$ is Poisson with parameter $\alpha / \rho$.

The alternative models will be semi-Markov processes; these describe jump-processes with non-exponential waiting times (i.e. with memory). Mathematically, a semi-Markov model is a piecewise-constant continuous-time process defined by a jump chain, $X'_n$, and jump times, $J_n$, and waiting times (i.e. jump intervals), $\Delta_{n+1} \equiv J_{n+1} - J_n$, that satisfy

**Alternative Models:**

\[
P(X'_{n+1} = y, \Delta_{n+1} \leq t | X_1', ..., X_{n-1}' , X_n', J_1, ..., J_n) = P(X'_{n+1} = y, \Delta_{n+1} \leq t | X_n') \equiv \tilde{Q}_{X_n'}(t).
\]
$Q_{x,y}(t)$ is called the semi-Markov kernel (see [31, 39] for further background and applications). Note that the base process Eq. (101) is described by the semi-Markov kernel

$$Q_{x,y}(t) \equiv \int_0^t q_{x,y}(s)ds = a(x,y) \int_0^t \lambda(x)e^{-\lambda(x)s}ds. \quad (105)$$

The QoI we will consider is the average queue-length,

$$F_T \equiv \frac{1}{T} \int_0^T x_t dt. \quad (106)$$

More specifically, we will be concerned with the long-time behavior (limit of $E_{\tilde{P}}[F_T]$ as $T \to \infty$).

Time-averages such as Eq. (106) were considered in Section 4.5, where we gave the UQ bounds

$$\pm (E_{\tilde{P}}[F_T] - E_P[F_T]) \leq \inf_{c > 0} \left\{ \frac{1}{cT} \tilde{F}_P(cT) + \frac{1}{cT} R(\tilde{P}|_{\mathcal{F}_T}|P|_{\mathcal{F}_T}) \right\}. \quad (107)$$

$E_P[F_T] = \alpha/\rho$ is the average queue-length in the (stationary) base process. Note that one does not need to first prove $F_T \in L^1(\tilde{P})$ to obtain Eq. (107); non-negativity of the QoI allows one to first work with a truncated QoI and then take limits.

The cumulant generating function of the time-averaged queue-length has the following limit (see Section 4.3.4 in [7]):

$$\lim_{T \to \infty} T^{-1} A_{P^T}(cT) = \frac{\alpha c^2}{\rho^2 (1 - c/\rho)}, \quad c < \rho. \quad (108)$$

Therefore

$$\limsup_{T \to \infty} (E_{\tilde{P}}[F_T] - \alpha/\rho) \leq \inf_{0 < c < \rho} \left\{ \frac{1}{c \rho^2 (1 - c/\rho)} + \frac{1}{c} \eta \right\} \quad (109)$$

$$= \left(2\sqrt{\eta/\alpha + \eta/\alpha} \right)^2 \frac{\alpha}{\rho},$$

$$\liminf_{T \to \infty} (E_{\tilde{P}}[F_T] - \alpha/\rho) \geq - \inf_{c > 0} \left\{ \frac{1}{c \rho^2 (1 + c/\rho)} + \frac{1}{c} \eta \right\} \quad (107)$$

$$= - \left(2\sqrt{\eta/\alpha - \eta/\alpha} \right)^2 \frac{\alpha}{\rho} 1_{\eta < \alpha} - \frac{\alpha}{\rho} 1_{\eta \geq \alpha},$$

where $\eta$ is the relative entropy rate:

$$\eta \equiv \limsup_{T \to \infty} \frac{1}{T} R(\tilde{P}|_{\mathcal{F}_T}|P|_{\mathcal{F}_T}). \quad (110)$$

A formula for the relative entropy rate between semi-Markov processes was obtained in [24] under the appropriate ergodicity assumptions:

$$\eta = \frac{1}{\bar{m}_{\tilde{\pi}}} \sum_{x,y} \tilde{\pi}(x) \int_0^\infty \tilde{q}_{x,y}(t) \log(\tilde{q}_{x,y}(t)/q_{x,y}(t)) dt, \quad (111)$$

$$\bar{m}_{\tilde{\pi}} \equiv \sum_{x} \tilde{\pi}(x) \int_0^\infty (1 - \tilde{H}_x(t)) dt, \quad \tilde{H}_x(t) = \sum_y \tilde{Q}_{x,y}(t),$$
where $\tilde{\pi}$ is the invariant distribution for the embedded jump chain of the tilde-process.

In particular, an alternative process with the same jump-chain as the base process but a different waiting-time distribution, $\tilde{H}_x(t) = \int_0^t \tilde{h}_x(s)ds$, is described by the semi-Markov kernel

$$Q_{x,y}(t) = a(x,y)\tilde{H}_x(t). \tag{112}$$

In this case, $\tilde{\pi} = \pi$ and $\eta$ can be expressed in terms of the relative entropy of the waiting-time distributions:

$$\eta = \limsup_{T \to \infty} \frac{1}{T} R(\tilde{P}|\mathcal{F}_T)||P|\mathcal{F}_T) = \frac{1}{\tilde{m}_\pi} E_{\pi(dx)} \left[ R(\tilde{H}_x||H_x) \right], \tag{113}$$

$$\tilde{m}_\pi = \sum_x \pi(x) \int_0^\infty (1 - \tilde{H}_x(t))dt, \tag{114}$$

where $\pi$ is the invariant distribution for $a(x,y)$. In particular, the $M/M/\infty$-queue jump chain, Eq. (102), has invariant distribution

$$\pi(x) = \left( \frac{\alpha + \rho x}{\rho} \right) \left( \frac{\alpha}{\rho} \right)^x e^{-\alpha/\rho}. \tag{115}$$

**Remark 15** The quantity $\tilde{m}_\pi$ is the mean sojourn time under the invariant distribution, $\pi$, and $E_{\pi(dx)} \left[ R(\tilde{H}_x||H_x) \right]$ can be viewed as the mean relative entropy of a single jump (comparing the alternative and base model waiting-time distributions). The formula for $\eta$, Eq. (113), therefore has the clear intuitive meaning of an information loss per unit time.

We next consider robustness bounds on semi-Markov alternative processes of the form Eq. (112) in more detail; the ambiguity set will be expressed in terms of constraints on $\tilde{H}_x$, as compared to $H_x$. Other than $\tilde{H}_x$, the formulas Eq. (109) and Eq. (113) include only base model quantities. Therefore, the bounds we obtain will be computable.

### 5.4.1 Defining Ambiguity Sets via Phase Type Distributions

Phase-type distributions constitute a semi-parametric description of waiting-time distributions, going beyond the exponential case to describe systems with memory; see [22,4,13] for examples and information on fitting to such distributions. Probabilistically, they can be characterized in terms the time to absorption for a continuous-time Markov chain with a single absorbing state. The density and distribution function of a phase type-distribution, $PH_k(\nu, T)$, are characterized by a $k \times k$ matrix $T$, and a probability vector $\nu \in \mathbb{R}^k$ (see [8] for background):

$$\tilde{h}(t) = \nu e^{Tt}(-T1), \quad \tilde{H}(t) = \int_0^t \tilde{h}(s)ds = 1 - \nu e^{Tt}1, \tag{116}$$

where 1 is the vector of all 1’s and $T$ satisfies:
1. \( T_{j,j} < 0 \) for all \( j \),
2. \( T_{j,\ell} \geq 0 \) for all \( j \neq \ell \),
3. \(-T1\) has all non-negative entries.

**Remark 16** Phase-type distributions are not a (finite-dimensional) parametric family, as \( k \) can be any positive integer. Hence, a non-parametric approach to robustness is desirable here.

Combining Eq. (116) with Eq. (113), we obtain a formula for the relative entropy rate of a semi-Markov perturbation with state-\( x \) waiting-time distributed as PH\(_{k(x)}(\nu(x), T(x))\) (and with the same jump-chain, \( a(x, y) \), as the base process):

\[
\eta = \frac{1}{m_\pi} \sum_x \pi(x) \int_0^\infty \log \left( \lambda(x)^{-1} \nu(x) e^{(T(x) + \lambda(x) t)(-T(x)1)} \nu(x) e^{T(x) t} (-T(x)1) dt, \right)
\]

\[
\bar{m}_x = \sum_x \pi(x) (-\nu(x) T^{-1}(x)1).
\]

As a simple example, consider the case where the waiting-time at \( x \) is distributed as a convolution of \( \exp(\lambda(x)) \) and \( \exp(\gamma(x)) \)-distributions (convolutions of exponential distributions are examples of phase-type distributions; again, see [8]), for some choices of \( \gamma(x) > \lambda(x) \). Recall the baseline model rates were given in Eq. (103). The semi-Markov kernel for the alternative process is then

\[
\hat{Q}_{x,y}(t) = a(x, y) \int_0^t \lambda(x) e^{-\lambda(x)s} \frac{1 - e^{-(\gamma(x) - \lambda(x))s}}{1 - \lambda(x)/\gamma(x)} ds.
\]

We consider the alternative family consisting of processes of the form Eq. (118) with the following choice for the ambiguity set (of course, other choices are possible):

**Ambiguity Set:**

\[
0 \leq \delta \leq \lambda(x)/\gamma(x) \leq \epsilon < 1 \quad \text{for all } x,
\]

i.e. \( \frac{1}{\epsilon} \lambda(x) \leq \gamma(x) \leq \frac{1}{\delta} \lambda(x) \),

where the jump-rates for the base model are \( \lambda(x) = \alpha + \rho x \). As a function of \( x \), the ambiguity set Eq. (119) therefore constrains \( \gamma(x) \) to lie between two lines. This set describes waiting-times that are approximately \( \exp(\lambda(x)) \)-distributed when \( \epsilon \) is small (i.e. when \( \gamma(x) \gg \lambda(x) \)). The parameter \( \delta \) is a lower-bound on the perturbation.

Using Eq. (119) we can bound

\[
\eta = \frac{1}{m_\pi} \sum_x \pi(x) \int_0^\infty \lambda(x) e^{-\lambda(x)t} \frac{1 - e^{-(\gamma(x) - \lambda(x))t}}{1 - \lambda(x)/\gamma(x)} \log \left( \frac{1 - e^{-(\gamma(x) - \lambda(x))t}}{1 - \lambda(x)/\gamma(x)} \right) dt
\]

\[
\leq \alpha r(\delta, \epsilon),
\]

\[
r(\delta, \epsilon) = \frac{2}{1 + \delta} \int_0^\infty e^{-u} \left( \frac{1 - e^{-(\epsilon^{-1} - 1)u}}{1 - \delta} \right) \left( \frac{1 - e^{-(\delta^{-1} - 1)u}}{1 - \epsilon} \right) du.
\]
Fig. 5 Upper bound on the relative error for the expected time-averaged-queue-length in the long-time limit, Eq. (122), for a semi-Markov perturbation of a M/M/∞ queue. The parameters δ and ǫ quantify the size of the perturbation; see Eq. (119) for their definitions. As ǫ → 0, the semi-Markov waiting times approach exponential waiting times and the relative error approaches 0. Increasing δ reduces the size of the ambiguity set, and hence reduces the uncertainty; this is also reflected in the upper bound.

To obtain the above, we used the inequality

$$\tilde{m}_\pi = \sum_x \pi(x) \lambda(x)^{-1} (1 + \lambda(x)/\gamma(x)) \geq (1 + \delta) \sum_x \pi(x) \lambda(x)^{-1} = \frac{1 + \delta}{2\alpha}. \quad (121)$$

Combining Eq. (120) and Eq. (121) with Eq. (109) gives the relative-error bounds

$$\begin{align*}
\limsup_{T \to \infty} \left( \frac{E_{\tilde{P}}[F_T]}{\left(\frac{\alpha}{\rho}\right)} - 1 \right) &\leq 2\sqrt{r(\delta, \epsilon)} + r(\delta, \epsilon), \\
\liminf_{T \to \infty} \left( \frac{E_{\tilde{P}}[F_T]}{\left(\frac{\alpha}{\rho}\right)} - 1 \right) &\geq -\left(2\sqrt{r(\delta, \epsilon)} - r(\delta, \epsilon)\right) 1_{r(\delta, \epsilon) < 1} - 1_{r(\delta, \epsilon) \geq 1}.
\end{align*} \quad (122)$$

Note that these bounds depend only on the uncertainty parameters ε and δ, and not on the base-model parameters. See Figure 5 for contour plots of the logarithm of the upper bound Eq. (122).

5.5 Value of American Put Options with Variable Interest Rate

As a final example, we consider the value of an asset in a variable-interest-rate environment, as compared to a constant-interest-rate base model. For background see, for example, Chapter 8 of [46] or Chapter 8 of [25]. This example does not fit neatly into any of the problem categories from Section 4, though the technique of Section 4.1 will play a critical role. In particular, we will see that the natural QoI for the base and alternative models is different; see Eq. (128) and Eq. (130).

Specifically, the base model in this example is geometric Brownian motion:

**Base Model:**

$$dX_t^r = rX_t^r dt + \sigma X_t^r dW_t, \quad (123)$$

where $$r, \sigma > 0$$ are constants. This has the explicit solution

$$X_t^r = X_0^r \exp \left( \sigma W(t) + (r - \sigma^2/2)t \right). \quad (124)$$
The variable-interest-rate alternative model has the general form

**Alternative Models:** \[ \dot{X}_t = (r + \Delta r(t, X_t, Y_t)) X_t dt + \sigma \tilde{X}_t dW_t. \] (125)

\[ r \] will be thought of as a known, fixed value, while \( \Delta r \) will be considered an unknown perturbation. See [46] for discussion of various interest rate models.

The perturbation of the rate, \( \Delta r \), is allowed to depend on time, on the asset price, as well as on an additional \( \mathbb{R}^k \)-valued process, \( Y_t \), that solves some SDE (see the online supplement for further technical details). The initial values \( X_0^r = X_0 \equiv X_0 \) will be fixed (here, the superscript on \( X \) will always denote the constant interest rate, \( r \), that is being used).

The quantity of interest we consider is the value of a perpetual American put option: Let \( K > 0 \) be the option strike price. The payout if the option is exercised at time \( t \geq 0 \) is \( K - x_t \), where \( x_t \) is the asset price at time \( t \). The relevant QoI is then the value, discounted to the present time. In the base model, this takes the form \( V : [0, \infty) \times C([0, \infty), \mathbb{R}) \to \mathbb{R} \),

\[ V_t[x] = e^{-rt}(K - x_t) 1_{t<\infty}, \] (126)

while for the alternative model it is \( \tilde{V} : [0, \infty) \times C([0, \infty), \mathbb{R} \times \mathbb{R}^k) \to \mathbb{R} \),

\[ \tilde{V}_t[x, y] = e^{-\int_0^t r + \Delta r(s, x_s, y_s) ds}(K - x_t) 1_{t<\infty}. \] (127)

The option-holder’s strategy is to exercise the option when the stock price hits some level \( L \), assumed to satisfy \( L < X_0 \) and \( 0 < L < K \), i.e. we consider the stopping time \( \tau[x, y] = \tau[x] = \inf\{t \geq 0 : x_t \leq L\} \). Therefore the base model QoI is

**Base QoI:** \[ E[V_\tau[X^r]] = (K - L)(L/X_0)^{2r/\sigma^2}. \] (128)

To evaluate this, one uses the fact that

\[ \tau \circ X^r = \inf\{t : X^r_t = L\} = \inf\{t : W_t + (r/\sigma - \sigma/2) t = -\sigma^{-1} \log(X_0/L)\} \] (129)

is the level \( a \equiv -\sigma^{-1} \log(X_0/L) \) hitting time of a Brownian motion with constant drift \( \mu \equiv r/\sigma - \sigma/2 \), combined with the formula for the distribution of such hitting times (see [33] and also the online supplement).

The goal of our analysis is now to bound the expected option value in the alternative model:

**Alternative QoI:** \[ E[\tilde{V}_{\tau[X^r]}] = (K - L) E\left[1_{\tau[X^r] < \infty} e^{-\int_0^{\tau[X^r]} r + \Delta r(s, X_s, Y_s) ds} \right]. \] (130)

Note that both QoI’s have unbounded time-horizon; therefore, to obtain non-trivial bounds one must again use a goal-oriented relative entropy.

**Remark 17** The methods developed in this paper, applied to the goal oriented relative entropy \( R(\hat{P}|\mathcal{F}_t)||\hat{P}|\mathcal{F}_t) \), are still not capable of comparing the base model Eq. (123), with asset-price volatility \( \sigma \), to an alternative model with perturbed asset-price volatility, \( \sigma + \Delta \sigma \), due to the loss of absolute continuity. It is likely that one does have absolute continuity on a smaller sigma algebra than \( \mathcal{F}_\tau \), but we do
not currently know how to bound the relative entropy on any such smaller sigma algebra.

An alternative approach to robust option pricing under uncertainty in $\sigma$, utilizing $H_\infty$-control methods, was developed in [40].

To obtain UQ bounds for Eq. (130), it is useful to define the modified QoI

$$ F_t[x, y] = (K - L) \exp \left( - \int_0^t r + \Delta r(s, x_s, y_s) \, ds \right) 1_{t < \infty}, \quad (131) $$

and note that $E \left[ V_{\tau \circ \tilde{X}}[\tilde{X}, \tilde{Y}] \right] = E_{\tilde{X}, \tilde{Y}}[F_\tau]$; where $E_{\tilde{X}, \tilde{Y}}$ denotes the expectation with respect to the joint distribution of $(\tilde{X}, \tilde{Y})$ on path space. (The notation $E_{X, Y}$ will similarly be used below.) In the next subsection, we will bound $E_{\tilde{X}, \tilde{Y}}[F_\tau]$ for one class of rate perturbations.

5.5.1 Vasicek Interest Rate Model

Here we study a specific type of dynamical interest rate model, known as the Vasicek model (see, for example, page 150 in [46]):

**Alternative Models:**

$$ d\tilde{X}_t = (r + \Delta r_t) \tilde{X}_t \, dt + \sigma \tilde{X}_t \, dW_t, \quad (132) $$

$$ d\Delta r_t = -\gamma \Delta r_t \, dt + \tilde{\sigma} d\tilde{W}_t, \quad \Delta r_0 = 0, \quad \gamma > 0, \quad \tilde{\sigma} > 0, \quad (133) $$

where $W$ and $\tilde{W}$ are independent Brownian motions. The base model is still given by Eq. (123). Robustness bounds under such perturbations can be viewed as a model stress test for the type of financial instrument (QoI) studied here; see [21] for a detailed discussion of stress testing.

The SDE Eq. (133) defines a 2-dimensional parametric family of alternative models, parametrized by $\gamma > 0$ and $\tilde{\sigma} > 0$. Therefore, we need not specify an ambiguity set here (unlike the previous examples, where the alternative family was infinite dimensional). We will simply obtain UQ bounds as functions of these two parameters. Infinite-dimensional alternative families of $\Delta r$-models can also be considered; see Appendix B for one such class of examples.

The SDE Eq. (133) for $\Delta r_t$ has the exact solution

$$ \Delta r_t = \tilde{\sigma} e^{-\gamma t} \int_0^t e^{\gamma s} d\tilde{W}_s. \quad (134) $$

Note that $\Delta r_t$ is unbounded, and so a comparison-principle bound is is not possible here.

Under the model Eq. (132), $r + \Delta r$ can become negative. Here, for both modeling and mathematical reasons, we suppress this effect by conditioning on the event $\int_0^\tau X + \Delta r_s \, ds \geq 0$, i.e. we only consider those paths with non-negative average interest rate up to the stopping time. This amounts to using the modified QoI

$$ \tilde{F}_t[x, y] \equiv (K - L) \exp \left( - \int_0^t r + y_s \, ds \right) 1_{t < \infty} 1_{\int_0^t r + y_s \, ds \geq 0}. \quad (135) $$
and bounding

**Alternative QoI:**

\[
\tilde{E}[\tilde{F}_r] \equiv E_{\tilde{X}_r,\Delta r}[\tilde{F}_r]/P_{\tilde{X}_r,\Delta r}\left(\int_0^\tau r + y_s ds \geq 0\right). \tag{136}
\]

Other, more complex, interest rate models exist that enforce a positive rate via the dynamics; see, for example, page 275 in [46].

We also restrict to the parameter values \( r > \frac{\sigma^2}{2a} \), for which a negative average interest rate is sufficiently rare. More precisely, this assumption implies \( \lim_{n \to \infty} \int_0^n r + \Delta r_s ds = \infty \) a.s.; see the online supplement for details. As a consequence, we also obtain \( \tilde{F}_n \to \tilde{F}_\infty = 0 \) a.s. (under both the base and alternative models). Therefore, Corollary 3 and Lemma 3 (see Appendix A) can both be used, resulting in

\[
\pm E_{\tilde{X}_r,\Delta r}[\tilde{F}_r] \leq \inf_{c > 0} \left\{ \frac{1}{c} A_{P_{X^r,\Delta r}}(\pm c) + \frac{1}{c} \lim_{n \to \infty} R(P_{\tilde{X}_r,\Delta r},[X^r,\Delta r],[\tau,\Delta r]) \right\}. \tag{137}
\]

For this example, we do not utilize Theorem 2, as the resulting expectation is difficult to evaluate. Instead, we proceed as in Remark 6, working with the cumulant generating function and relative entropy terms separately.

First we compute the cumulant generating function. \( \tau \circ X^r \) and \( \Delta r \) are independent, hence we can use the result of [1] to find

\[
A_{P_{X^r,\Delta r}}(\pm c) \equiv \log \int \int_{-\infty}^{\infty} \left(2\pi \sigma^2_t\right)^{-1/2} \exp\left(\pm c(K-L)e^{-t}1_{t<\infty}1_{z \geq 0}\right) e^{-\frac{(z-\gamma^2t)^2}{2\gamma^2}} dz \rho_{\tau \circ X^r}(dt), \tag{138}
\]

\[
\sigma^2_t \equiv \frac{\sigma^2}{2\gamma} \left(2\gamma t + 4e^{-\gamma t} - e^{-2\gamma t} - 3\right).
\]

(See the online supplement for more details.) The integral over \( t \) in Eq. (138) can then be computed as discussed in the text surrounding Eq. (129), with \( a \equiv -\sigma^{-1} \log(X_0/L) \) and \( \mu \equiv r/\sigma - \sigma/2 \). (When \( t = \infty \), the inner integral should be interpreted as equaling 1.)

**Remark 18** Note that if the integrated rate (\( z \) in Eq. (138)) is allowed to be negative (i.e. we remove \( 1_{z \geq 0} \) from the exponential) then \( A_{P_{X^r,\Delta r}}(c) \) is infinite and the upper UQ bound is the trivial bound, \( +\infty \). This is the mathematical reason for conditioning on the event \( \int_0^\tau r + \Delta r_s ds \geq 0 \).

The relative entropy can be computed via Girsanov’s theorem (see the online supplement for a proof that the Girsanov factor is an exponential martingale)

\[
R(P_{\tilde{X}_r,\Delta r},[X^r,\Delta r],[\tau,\Delta r]) \leq E_{\tilde{X}_r,\Delta r}[G_n] , \quad G_n[x,y] \equiv \frac{1}{2} \int_0^\tau (x + y_s) \sigma^2 |y_s|^2 ds.
\]

Note that \( G_n \) is \( \tau \wedge \Delta r \)-measurable and has finite expectation, as one can bound \( \tau \wedge n \) by \( n \) and then use the fact that \( \Delta r_s \) is normal with mean 0 and variance \( \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma s}) \). Hence, Lemma 4 is applicable, resulting in

\[
R(P_{\tilde{X}_r,\Delta r},[X^r,\Delta r],[\tau,\Delta r]) \leq \inf_{c > 1} \left\{ (c - 1)^{-1} A_{P_{X^r,\Delta r}}(c) \right\}. \tag{139}
\]
Robustness via Goal-Oriented Information Theory

Figure 6: Bound on the expected option value under the Vasicek model. Black curves are for the base model (see Eq. (128)) and blue dashed curves are the bounds on the alternative model QoI, Eq. (136), that result from our method (see Eq. (137) - Eq. (143)); the QoI’s are constrained to the gray regions. Left: As a function of the volatility of the interest-rate perturbation, with $\gamma = 2$, $r = 5/4$, $K = 1$, $L = 1/2$, $X_0 = 2$, $\sigma = 4$. Right: As a function of the volatility of the asset price, with $\gamma = 2$, $r = 1$, $K = 1$, $L = 1/2$, $X_0 = 2$, $\tilde{\sigma} = 6$.

Again using the independence of $\tau \circ X^r$ and $\Delta r$, we have

$$A_{P_{X^r,\Delta r}}^G(c) = \log \int E \left[ e^{c \int_0^T \frac{|\Delta r_s|^2}{\sigma^2}} \right] P_{\tau \circ X^r}(dt). \quad (140)$$

The inner expectation can be evaluated using Theorem 1 (B) of [15]:

$$E \left[ e^{c \int_0^T \frac{|\Delta r_s|^2}{\sigma^2}} \right] \leq e^{\gamma t/2} \left[ \frac{1}{\sqrt{1 - \frac{c \sigma^2}{\sigma^2 \gamma}}} \sinh \left( \gamma t \sqrt{1 - \frac{c \sigma^2}{\sigma^2 \gamma}} \right) + \cosh \left( \gamma t \sqrt{1 - \frac{c \sigma^2}{\sigma^2 \gamma}} \right) \right]^{-1/2}. \quad (141)$$

By analytic continuation, this formula is valid for $c < \sigma^2 \gamma^2 / \tilde{\sigma}^2$. Also, note that for the $n \to \infty$ limit of Eq. (140) to be finite, we need $\tau \circ X^r \lesssim \infty$ $P$-a.s., i.e. $\mu$ and $a$ must have the same signs.

We can similarly bound $P_{X^r,\Delta r} \left( \int_0^T r + y_s ds \geq 0 \right)$:

$$\pm P_{X^r,\Delta r} \left( \int_0^T r + y_s ds \geq 0 \right) \leq \inf_{c > 0} \frac{1}{c} \log \left( 1 + (e^{\pm c} - 1) P_{X^r,\Delta r} \left( \int_0^T r + y_s ds \geq 0 \right) \right)$$

$$+ \frac{1}{c} \lim \inf_{n \to \infty} R(P_{X^r,\Delta r, X_{\tau,\Delta r}} \sup ||P_{X^r,\Delta r}|| P_{X^r,\Delta r, X_{\tau,\Delta r}}). \quad (142)$$
where we used Eq. (39) to compute the MGF, and
\[
P_{\bar{x}^r, \Delta r} \left( \int_0^\tau r + y_t ds \geq 0 \right) = \int \int_0^\infty (2\pi \tilde{\sigma}_t^2)^{-1/2} e^{-\frac{(\tau r + \xi - \tilde{\mu}_t r)^2}{2 \tilde{\sigma}_t^2}} dz P_{\tau \circ \bar{X}_r}(dt) \tag{143}
\]
\[
= \int \frac{1}{2} \left( \text{erf} \left( \frac{r t / \sqrt{2 \tilde{\sigma}_t^2} + 1}{1} \right) \right) P_{\tau \circ \bar{X}_r}(dt).
\]

Figure 6 shows numerical results for the bounds on Eq. (136) that result from combining Eq. (137) through Eq. (143).

Remark 19 The reason we are able to perturb the diffusion parameter, \( \tilde{\sigma} \), without losing absolute continuity is that, technically, we are taking the base model to be the joint distribution of \( (X^x_t, \Delta r_t) \). Of course, \( X^x_t \) is not coupled to \( \Delta r_t \), nor does the base QoI depend on it, but regardless, this makes it clear that there is no loss of absolute continuity between the distributions of \( (X^x_t, \Delta r_t) \) and \( (\bar{X}_t, \Delta r_t) \) when \( \tilde{\sigma} \) is changed.

A Tools for Computing the Goal-Oriented Relative Entropy

In this appendix, we present several lemmas that are generally useful when computing the goal-oriented relative entropy. First, we address the question of processes with different initial distributions:

Lemma 2 Suppose that the base and alternative models are of the form
\[
P = P^\mu = \int P^x(\cdot|\mu) dx, \quad \bar{P} = \bar{P}^\mu = \int \bar{P}^x(\cdot|\bar{\mu}) dx \tag{144}
\]
for some probability kernels, \( P^x \) and \( \bar{P}^x \), from \( (X, \mathcal{M}) \) to \( (\Omega, \mathcal{F}) \), and some probability measures (e.g. initial distributions), \( \mu \) and \( \bar{\mu} \), on \( (X, \mathcal{M}) \).

In this case, one can use the chain-rule for relative entropy (see, for example, [18]) to bound
\[
R(F, \bar{P}^x||F, P^x) \leq R(\bar{\mu}||\mu) + \int R(F, \bar{P}^x||F, P^x) \bar{\mu}(dx). \tag{145}
\]
As discussed in Theorem 1, if \( F \) is measurable with respect to a sub-sigma algebra, \( \mathcal{G} \), then one can weaken the bound via \( R(F, \bar{P}^x||F, P^x) \leq R(\bar{P}^x|\mathcal{G})||P^x|\mathcal{G}) \).

Remark 20 Note that without further assumptions, Eq. (145) is not an equality. For example, let \( \nu \) be a probability measure on \( \mathbb{R} \), \( F = \pi_2 \) (the projection onto the second component in \( \mathbb{R} \times \mathbb{R} \)), and for \( x \in \mathbb{R} \) define \( P^x = \bar{P}^x = \delta_x \times \nu \). Then for \( \mu \neq \bar{\mu} \) we have \( P^\mu = \mu \times \nu \), \( \bar{P}^\mu = \bar{\mu} \times \nu \) and
\[
R(F, \bar{P}^\mu||F, P^\mu) = R(\nu||\nu) = 0 < R(\bar{\mu}||\mu) + \int R(F, \bar{P}^x||F, P^x) \bar{\mu}(dx). \tag{146}
\]

Intuitively, inequality in Eq. (145) arises when the QoI ‘forgets’ some of the information regarding the discrepancy between the initial condition, as the right-hand-side in Eq. (145) incorporates the full discrepancy.

The main tool we use to compute the relative entropy up to a stopping time, Lemma 4 below, requires one to work with bounded stopping times. This limitation can often be overcome by taking limits. For example:
Lemma 3 Suppose that one is working within the framework of Assumption 1 and one of the following two conditions holds:
1. \( \tau \) is finite \( P \)-a.s. and \( \bar{P} \)-a.s.
2. \( F_n \to F_\infty \) as \( n \to \infty \), \( n \in \mathbb{Z}^+ \), both \( P \) and \( \bar{P} \)-a.s.

Then
\[
R((F_t)_t, \bar{P}((F_t)_t), P) \leq \liminf_{n \to \infty} R((F_{\tau \wedge n})_t, \bar{P}((F_{\tau \wedge n})_t, P)
\]
(147)
\[
\leq \liminf_{n \to \infty} R(\bar{P}|_{F_{\tau \wedge n}} || P|_{F_{\tau \wedge n}}).
\]

Proof Either condition implies that \( (F_{\tau \wedge n})_t, \bar{P} \to (F_t)_t, \bar{P} \) and \( (F_{\tau \wedge n})_t, P \to (F_t)_t, P \) weakly as \( n \to \infty \). The result then follows from lower-semicontinuity of relative entropy (see [18]). \( \square \)

The general tool we use to compute relative entropy up to a stopping time is the optional sampling theorem (see, for example, [33]). The technique is summarized by the following lemma:

Lemma 4 Let \( P, \bar{P} \) be probability measures on the filtered space \( (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in T}) \) such that \( \bar{P}|_{\mathcal{F}_t} \ll P|_{\mathcal{F}_t} \) for all \( t \).

Then \( \rho_t = \frac{d\bar{P}|_{\mathcal{F}_t}}{dP|_{\mathcal{F}_t}} \) is an \( (\mathcal{F}_T, P) \)-martingale. In the continuous-time case, we also assume that we have a right-continuous version of \( \rho_t \).

Given this, for any stopping time, \( \tau \), we have \( \frac{d\bar{P}|_{\mathcal{F}_\tau \wedge t}}{dP|_{\mathcal{F}_\tau \wedge t}} = E(\rho_{\tau \wedge t} | \mathcal{F}_{\tau \wedge t}) = \rho_{\tau \wedge t} \) and
\[
R(\bar{P}|_{\mathcal{F}_\tau \wedge t} || P|_{\mathcal{F}_\tau \wedge t}) = E_P[\log(\rho_{\tau \wedge t})].
\]
(148)

Using Lemma 4, the relative entropy up to a stopping time for various types of Markov processes can be computed via standard techniques; explicit formulas can be found in the online supplement.

B American Options: Bounded Rate Perturbations

In this appendix, we derive UQ bounds on the value of American options for another class of rate perturbations. This is a simpler example to analyze than the Vasichek model from Section 5.5.1 and provides a good benchmarking case, as we will also have access to comparison principle bounds. We use the terminology and notation introduced in Section 5.5.

Specifically, here we consider interest rate perturbations defined by the following ambiguity set.

Ambiguity Set:
\[
\Delta r_- \leq h_-(t) \leq \Delta r(t, \cdot) \leq h_+(t) \leq \Delta r_+, \quad t \geq 0,
\]
(149)
where \( \Delta r_\pm \in \mathbb{R}, \Delta r_+ > 0 \), and \( h_-, h_+: [0, \infty) \to \mathbb{R} \). The parameters \( \Delta r_- \) and \( \Delta r_+ \) define a fluctuation range, with \( h_\pm \) allowing for specification of a time-dependent envelope on the fluctuation.

An elementary bound on \( \overline{X}_t \) follows from the uniform bounds \( \Delta r_- \leq \Delta r(t, x, y) \leq \Delta r_+ \) together with the comparison principle:
\[
X_t^{\tau \wedge \Delta r_-} \leq \overline{X}_t \leq X_t^{\tau \wedge \Delta r_+}, \quad \text{and hence} \quad \tau \circ X^{\tau \wedge \Delta r_-} \leq \tau \circ \overline{X} \leq \tau \circ X^{\tau \wedge \Delta r_+}.
\]
(150)

Using only the information \( \Delta r_- \leq \Delta r(t) \leq \Delta r_+ \), the optimal UQ bounds can be obtained from Eq. (149) and Eq. (150), together with the exact value for the constant-rate processes, Eq. (128):
\[
(K - L)(L/X_0)^{2(r+\Delta r_+)/\sigma^2} \leq E_P[\overline{X}_{\tau \wedge \overline{X}} | \overline{X}, \overline{Y}] \leq (K - L)(L/X_0)^{2(r+\Delta r_-)/\sigma^2}.
\]
(151)
This provides a useful comparison for the UQ bounds derived via the methods developed above.
To use the goal-oriented method, first note the relative entropy bounds (obtained from Girsanov’s theorem):

\[ R(P_{\tilde{X},\tilde{Y}} \| P_{X,Y}) \leq E_{\tilde{X},\tilde{Y}}[G_n], \quad G_n[x,y] = \frac{\sigma^2}{2} \int_0^{T_{\inf}(\tau,n)} |\Delta r(s,x,y)|^2 ds. \]

\[ G_n \text{ is } F_{\tau,n}-\text{measurable, hence we can use Theorem 2 to obtain} \]

\[ \pm E_{\tilde{X},\tilde{Y}}[F_{\tau,n}] \leq \inf_{c > 0} \left\{ \frac{1}{c} \log E_{X,Y} \left[ \exp \left( \pm c F_{\tau(n)} + \frac{\sigma^2}{2} \int_0^{T_{\inf}(\tau,n)} |\Delta r(s,x,y)|^2 ds \right) \right] \right\}. \]

The above steps are justified by the bounds on \( \Delta r \).

The bounds on \( \Delta r \) also allow us to take \( n \to \infty \) in Eq. (153) to obtain

\[ \pm E \left[ \tilde{V}_{r} \big| \tilde{X},\tilde{Y} \right] = \pm E_{\tilde{X},\tilde{Y}} [F_r] = \lim_{n \to \infty} \pm E_{\tilde{X},\tilde{Y}} [F_{\tau,n}] \]

\[ \leq \inf_{c > 0} \left\{ \frac{1}{c} \log E_{X,Y} \left[ \exp \left( \pm c F_{\tau(n)} + \frac{\sigma^2}{2} \int_0^{T_{\inf}(\tau,n)} |\Delta r(s,x,y)|^2 ds \right) \right] \right\}. \]

Bounding \( \Delta r \) in terms of \( h \) then yields

\[ \pm E \left[ \tilde{V}_{r} \big| \tilde{X},\tilde{Y} \right] \leq \inf_{c > 0} \left\{ \frac{1}{c} \log E_{X,Y} [M] \right\}, \]

\[ M \equiv \exp \left( \frac{\sigma^2}{2} \int_0^{T_{\inf}(\tau)} |h(s)|^2 ds \pm c(K-L)e^{-\int_0^{T_{\inf}(\tau)} r+h(z)dz} I_{r<\infty} \right), \]
where $h(t) \equiv \max\{h_+(t), -h_-(t)\}$. The expectation can be evaluated using the known distribution of $\tau \circ X^r$ (see the discussion surrounding Eq. (129)):

$$
E_{X^r}[M^2_h] = \frac{|a|}{\sqrt{2\pi}} \int_0^\infty \exp\left(\frac{\sigma^2}{2} \int_0^t h^2_s ds \pm c(K - L)e^{-\frac{\sigma^2}{2} t} \frac{(a - \mu)^2}{2t}\right) t^{-3/2} dt
$$

$+$ $\exp\left(\frac{\sigma^2}{2} \int_0^\infty h^2_s ds\right)(1 - e^{\mu t - |a| t})$.

Note that all reference to the driving process $Y$, which is not coupled to $X^r$, has been eliminated.

Remark 21 One can obtain somewhat tighter bounds in Eq. (155) by optimizing over the parameter that defines the base model, similarly to Eq. (29); in other words, replace $r$ with a parameter, $\kappa$, on the right-hand side, find the corresponding $h^*_\kappa$ satisfying $h^*_\kappa \leq r - \kappa + \Delta r \leq h^*_+ \leq \kappa$, and minimize over $\kappa$; see Figure 7.

We show numerical results for the following scenario: Suppose the rate drops from $r + \Delta r_+$ to $r$, and is certain to have completed this drop by time $t_f$ (one could also study a similar scenario where $r$ increases), but the timing and profile of the drop is otherwise unknown, i.e. we consider $\Delta r_+$ and $t_f$ to be known, and the only constraint on $\Delta r$ is $0 \leq \Delta r \leq \Delta r_{+1 (0,t_f)}(t) = \Delta r_+ 1_{(0,t_f)}(t) = h(t)$.

UQ bounds are obtained by combining Eq. (155) with Eq. (156).

In Figure 7 we show the resulting bounds on the option value, as a function of $t_f$. Note that the bounds resulting from our UQ method are an improvement over the comparison-principle bound Eq. (151) for an initial range of $t_f$, and stabilizes at a value near the comparison-principle bound for large $t_f$. For large $t_f$ our method is sub-optimal, but still competitive. Minimizing over the parameters assigned to the base model (right plot) improves the bound. We find similar behavior when assuming other time-dependent envelopes of $\Delta r$. It should again be noted that the comparison principle is available only in very special circumstances and is used here for benchmarking purposes; see Section 5.5.1 for a more realistic model where the comparison principle is not an effective tool.

Acknowledgments

The research of L. R.-B. was partially supported by the National Science Foundation (NSF) under the grant DMS-1515712 and the Air Force Office of Scientific Research (AFOSR) under the grant FA-9550-18-1-0214.

The research of M. K. was partially supported by the National Science Foundation under the grant DMS-1515712 and the Air Force Office of Scientific Research (AFOSR) under the grant FA-9550-18-1-0214.

References

1. Mario Abundo, On the representation of an integrated Gauss-Markov process, Scientiae Mathematicae Japonicae 77 (2015), no. 3, 357–361.
2. L. Alili, P. Patie, and J. L. Pedersen, Representations of the first hitting time density of an Ornstein-Uhlenbeck process, Stochastic Models 21 (2005), no. 4, 967–980.
3. B.D.O. Anderson and J.B. Moore, Optimal control: Linear quadratic methods, Dover Books on Engineering, Dover Publications, 2007.
4. Sren Asmussen, Olle Nerman, and Marita Olsson, Fitting phase-type distributions via the EM algorithm, Scandinavian Journal of Statistics 23 (1996), no. 4, 419–441.
5. R. Atar, K. Chowdhary, and P. Dupuis, Robust bounds on risk-sensitive functionals via Rényi divergence, SIAM/ASA Journal on Uncertainty Quantification 3 (2015), no. 1, 18–33.
6. Hildo Bijl and Thomas B. Schn, *Optimal controller/observer gains of discounted-cost lqg systems*, Automatica 101 (2019), 471 – 474.

7. Jeremiah Birrell and Luc Rey-Bellet, *Uncertainty Quantification for Markov Processes via Variational Principles and Functional Inequalities*, arXiv:1812.05174 (2018).

8. M. Bladt and B.F. Nielsen, *Matrix-exponential distributions in applied probability*, Probability Theory and Stochastic Modelling, Springer US, 2017.

9. Jose Blanchet and Karthyek Murthy, *Quantifying distributional model risk via optimal transport*, Mathematics of Operations Research 44 (2019), no. 2, 565–600.

10. S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities*, Oxford University Press, Oxford, 2013. MR 3185195

11. T. Breuer and I. Csiszár, *Measuring distribution model risk*, Mathematical Finance 26 (2013), no. 2, 395–411.

12. T. Breuer and I. Csiszár, *Systematic stress tests with entropic plausibility constraints*, Journal of Banking & Finance 37 (2013), no. 5, 1552 – 1559.

13. Lawrence Brown, Noah Gans, Avishai Mandelbaum, Anat Sakov, Haipeng Shen, Sergey Zeltyn, and Linda Zhao, *Statistical analysis of a telephone call center*, Journal of the American Statistical Association 100 (2005), no. 469, 469–50.

14. K. Chowdhary and P. Dupuis, *Distinguishing and integrating aleatoric and epistemic variation in uncertainty quantification*, ESAIM: Mathematical Modelling and Numerical Analysis 47 (2013), no. 3, 635-662.

15. Thad Daniel, *On the distribution of the integrated square of the Ornstein-Uhlenbeck process*, SIAM Journal on Applied Mathematics 51 (1991), no. 2, 568–574.

16. Erick Delage and Yinyu Ye, *Distributionally robust optimization under moment uncertainty with application to data-driven problems*, Operations Research 58 (2010), no. 3, 595–612.

17. A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Stochastic Modelling and Applied Probability, Springer Berlin Heidelberg, 2009.

18. P. Dupuis and R.S. Ellis, *A weak convergence approach to the theory of large deviations*, Wiley Series in Probability and Statistics, Wiley, 2011.

19. P. Dupuis, M. A. Katsoulakis, Y. Pantazis, and L. Rey-Bellet, *Sensitivity analysis for rare events based on Rényi divergence*, arXiv:1805.06917 (2018).

20. P. Dupuis, M.A. Katsoulakis, Y. Pantazis, and P. Plecháč, *Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics*, SIAM/ASA Journal on Uncertainty Quantification 4 (2016), no. 1, 80–111.

21. B. Engelmann and R. Rauhmeier, *The Basel ii risk parameters: Estimation, validation, stress testing - with applications to loan risk management*, Springer Berlin Heidelberg, 2011.

22. M. J. Faddy, *Examples of fitting structured phase-type distributions*, Applied Stochastic Models and Data Analysis 10 (1994), no. 4, 247–255.

23. Rui Gao and Anton J. Kleywegt, *Distributionally Robust Stochastic Optimization with Wasserstein Distance*, arXiv e-prints (2016), arXiv:1604.02199.

24. V. Girardin and N. Limnios, *On the entropy for semi-Markov processes*, Journal of Applied Probability 40 (2003), no. 4, 1060–1068.

25. P. Glasserman, *Monte Carlo methods in financial engineering*, Stochastic Modelling and Applied Probability, Springer New York, 2013.

26. P. Glasserman and X. Xu, *Robust risk measurement and model risk*, Quantitative Finance 14 (2014), no. 1, 29–58.

27. Joel Goh and Melvyn Sim, *Distributionally robust optimization and its tractable approximations*, Operations Research 58 (2010), no. 4-part-1, 902–917.

28. K. Gourgoulias, M.A. Katsoulakis, and L. Rey-Bellet, *Information Metrics For Long-Term Errors in Splitting Schemes For Stochastic Dynamics and Parallel Kinetic Monte Carlo*, SIAM Journal on Scientific Computing 38 (2016), no. 6, A3808–A3832.

29. K. Gourgoulias, M.A. Katsoulakis, L. Rey-Bellet, and J. Wang, *How biased is your model? Concentration inequalities, information and model bias*, arXiv:1706.10260 (2017).

30. S. Heinz and H. Bessaih, *Stochastic equations for complex systems: Theoretical and computational topics*, Mathematical Engineering, Springer International Publishing, 2015.

31. J. Janssen and R. Manca, *Applied semi-markov processes*, Springer e-Books, Springer US, 2006.

32. Ruwei Jiang and Yongpei Guan, *Data-driven chance constrained stochastic program*, Mathematical Programming 158 (2016), no. 1, 291–327.
A.1 Hitting Times of Brownian Motion with Constant Drift

For the reader’s convenience, here we recall the moment generating function and distribution of hitting times for Brownian motion with constant drift; see, for example, Chapter 8 of [46] and page 196 in [33].

**Lemma 5** Let $W_t$ be a $\mathbb{R}$-valued Wiener process (i.e. Brownian motion) on $(\Omega, \mathcal{F}, P)$, $\mu \in \mathbb{R}$, $a \in \mathbb{R} \setminus \{0\}$. Let $X_t = \mu t + W_t$ be Brownian motion with constant drift $\mu$. Define the level-$a$ hitting time $\tau_a = \inf\{t \geq 0 : X_t = a\}$. For any measurable $A \subset (0, \infty)$ we have

$$P(\tau_a \in A) = \frac{|a|}{\sqrt{2\pi}} \int_{A/\sqrt{2\mu}}^{\infty} e^{-y^2/2} dt + \left(1 - e^{\mu a - |a|} + 1\right) 1_{\infty \in A}.$$ (157)

In addition, for $\lambda > 0$ the MGF is

$$E[e^{-\lambda \tau_a}] = \exp\left(\mu a - |a| \sqrt{\mu^2 + 2\lambda}\right).$$ (158)
A.2 Relative Entropy up to a Stopping Time for Markov Processes

Below, we use Lemma 4 to give explicit formulas for the relative entropy up to a stopping time for several classes of Markov processes in both discrete and continuous time.

A.2.1 Discrete-Time Markov Processes

Let $\mathcal{X}$ be a Polish space, $p(x, dy)$ and $\tilde{p}(x, dy)$ be transition probabilities on $\mathcal{X}$ and $\mu$, $\tilde{\mu}$ be probability measures on $\mathcal{X}$. Let $P^n$ and $\tilde{P}^n$ be the induced probability measures on $\Omega \equiv \prod_{i=0}^{\infty} \mathcal{X}$ with transition probabilities $p$ and $\tilde{p}$ respectively, and initial distribution $\mu$ and $\tilde{\mu}$. Let $\pi_n : \mathcal{X} \rightarrow \mathcal{X}$ be the natural projections, $\mathcal{F}_n = \sigma(\pi_n : m \leq n)$, and $\tau$ be a $\mathcal{F}_n$-stopping time.

We have the following bound on the relative entropy, via the optional sampling theorem:

**Lemma 6** Suppose that $\tilde{p}(x, dy) = g(x, y)p(x, dy)$ for some measurable $g : \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty)$ and let $N \geq 0$. Then

$$R(\tilde{P}^\tau |_{\mathcal{F}_{\tau \wedge N}} || P^\tau |_{\mathcal{F}_{\tau \wedge N}}) = E^\pi \left[ \log \left( \prod_{t=1}^{\tau \wedge N} g(\pi_{t-1}, \pi_t) \right) \right]. \tag{159}$$

A.2.2 Continuous-Time Markov Chains

Let $\mathcal{X}$ be a countable set, $P$ and $\tilde{P}$ be probability measures on $(\Omega, \mathcal{F})$, and $X_t : \Omega \rightarrow \mathcal{X}$, $t \in [0, \infty)$, such that $(\Omega, \mathcal{F}, P, X_1)$ and $(\Omega, \mathcal{F}, \tilde{P}, X_1)$ are continuous-time Markov chains (CTMCs) with transition probabilities $a(x, y)$, $\tilde{a}(x, y)$, jump rates $\lambda(x)$, $\tilde{\lambda}(x)$, and initial distributions $\mu$, $\tilde{\mu}$ respectively. Let $\mathcal{F}_t$ be the natural filtration for $X_t$, and $X^n_t$ be the embedded jump chain with jump times $J_n \in [0, \infty]$. We assume $J_n(\omega) \rightarrow \infty$ as $n \rightarrow \infty$ for all $\omega$.

Suppose $\tilde{\mu} \ll \mu$ and that whenever $a(x, y) = 0$ we also have $\tilde{a}(x, y) = 0$ (note that this also implies that $\lambda(x) = 0$ whenever $\tilde{\lambda}(x) = 0$). Then $\tilde{P}^t |_{\mathcal{F}_t} \ll P^t |_{\mathcal{F}_t}$ for all $t \geq 0$ and

$$\frac{d\tilde{P}^t |_{\mathcal{F}_t}}{dP^t |_{\mathcal{F}_t}} = \tilde{\mu}(X_0) \exp \left( \int_0^t \log \left( \frac{\tilde{\lambda}(X_s) - \tilde{\lambda}(X_s)}{\lambda(X_s)} \right) N(ds) - \int_0^t \tilde{\lambda}(X_s) - \lambda(X_s) ds \right), \tag{160}$$

where $N_t = \max\{n : J_n \leq t\}$ is the number of jumps up to time $t$ (see, for example, [30]). Note that in defining the functions $\tilde{\mu}(x)/\mu(x)$ and $\tilde{\lambda}(x)\tilde{a}(x, y)/(\lambda(x)a(x, y))$, we use the convention $0/0 \equiv 1$.

Lemma 4 then yields the following:

**Lemma 7**

$$R(\tilde{P} |_{\mathcal{F}_{\tau \wedge 1}} || P |_{\mathcal{F}_{\tau \wedge 1}}) = R(\tilde{\mu} || \mu) + E^\pi \left[ \int_0^{\tau \wedge 1} \log \left( \frac{\tilde{\lambda}(X_s) - \tilde{\lambda}(X_s)}{\lambda(X_s)} \right) N(ds) - \int_0^{\tau \wedge 1} \tilde{\lambda}(X_s) - \lambda(X_s) ds \right]. \tag{161}$$

A.2.3 Change of Drift for SDEs

Here we consider the case where $P^x$ and $\tilde{P}^x$ are the distributions on path space, $C([0, \infty), \mathbb{R}^n)$, of the solution flows $X^x_t$ and $\tilde{X}^x_t$ of a pair of SDEs starting at $x \in \mathbb{R}^n$. More precisely:

**Assumption 2** Assume:

1. We have filtered probability spaces, $(M, \mathcal{G}_\infty, \mathcal{G}_t, P)$, $(\tilde{M}, \tilde{\mathcal{G}}_\infty, \tilde{G}_t, \tilde{P})$ that satisfy the usual conditions [33] and are equipped with $\mathbb{R}^m$-valued Wiener processes $W_t$ and $\tilde{W}_t$ respectively (Wiener processes with respect to the respective filtrations).
2. We have another probability space $(\mathcal{N}, \mathcal{N}', \nu)$, a $\mathcal{G}_0$-measurable random quantity $Z : M \rightarrow \mathcal{N}$, and a $\tilde{\mathcal{G}}_0$-measurable random quantity $\tilde{Z} : \tilde{M} \rightarrow \mathcal{N}'$, and they are both distributed as $\nu$. 

Remark 22 Intuitively, we are thinking of $Z$ as representing some outside data, independent of $\{W_t\}_{t\geq 0}$, that is informing the model. $\tilde{G}_t$ can then be taken to be the completed sigma algebra generated by $Z$ and the Wiener process up to time $t$. Since $Z$ and the Wiener process are independent, $W_t$ is still a Wiener process with respect to the resulting filtration.

3. We have $x \in \mathbb{R}^n$ and $\mathbb{R}^n$-valued processes $X^x_t$ and $\tilde{X}^x_t$ that are adapted to $\mathcal{G}_t$ and $\tilde{\mathcal{G}}_t$ respectively, and satisfy the SDEs

$$dX^x_t = b(t, X^x_t, Z)dt + \sigma(t, X^x_t, Z)dW^x_t, \quad X^x_0 = x,$$

$$d\tilde{X}^x_t = \tilde{b}(t, \tilde{X}^x_t, \tilde{Z})dt + \sigma(t, \tilde{X}^x_t, \tilde{Z})d\tilde{W}^x_t, \quad \tilde{X}^x_0 = x,$$

where $b : [0, \infty) \times \mathbb{R}^n \times N \to \mathbb{R}^n$, $\sigma : [0, \infty) \times \mathbb{R}^n \times N \to \mathbb{R}^n \times m$, and $\beta : [0, \infty) \times \mathbb{R}^n \times N \to \mathbb{R}^m$ are measurable and the modified drift is

$$\tilde{b} = b + \sigma \beta.$$

4. $\beta(s, X^x_s) \in L^2_{\text{loc}}(W)$ and the following process is a martingale under $P$:

$$\rho_T^x \equiv \exp \left( \int_0^T \beta(s, X^x_s, Z) \cdot dW^x_s - \frac{1}{2} \int_0^T \|\beta(s, X^x_s, Z)\|^2 ds \right).$$

Recall that this holds if the Novikov condition is satisfied. See, for example, page 199 of [23].

5. The SDE Eq. (163) satisfies the following uniqueness in law property for all $T > 0$:

Suppose $X^x_t$ is a continuous, adapted process on another filtered probability space satisfying the usual conditions, $(\mathcal{M}, \mathcal{G}_n, \mathcal{G}_t, P')$, that is also equipped with $\mathcal{G}_t$-Wiener process $W^x_t$ and a $N$-valued $\mathcal{G}_0$-measurable random quantity, $Z'$, distributed as $\nu$. Finally, suppose also that $X^x_t$ satisfies

$$X^x_t = x + \int_0^t b(s, X^x_s, Z') ds + \int_0^t \sigma(s, X^x_s, Z') dW^x_t,$$

for $0 \leq t \leq T$. Then the joint distribution of $(X^x|_{[0,T]}, Z')$ equals the joint distribution of $(X^x|_{[0,T]}, Z)$.

Given this, we define $P^x = (X^x, P)$ and $\tilde{P}^x = (\tilde{X}^x, \tilde{P})$, i.e. the distributions on path space,

$$(\Omega, \mathcal{F}_\infty, \mathcal{F}_t) \equiv (C([0,\infty), \mathbb{R}^n), \mathcal{B}(C([0,\infty), \mathbb{R}^n)), \sigma(\pi_s, s \leq t),$$

where $\pi_t$ denotes evaluation at time $t$ and $C([0,\infty), \mathbb{R}^n)$ is given the topology of uniform convergence on compact sets. For $T \geq 0$, we let $P^x_T \equiv P^x|_{\mathcal{F}_T}$ and $\tilde{P}^x_T \equiv \tilde{P}^x|_{\mathcal{F}_T}$.

The relative entropy can be computed via Girsanov’s theorem, together with Lemma 4:

**Lemma 8** Let $x \in \mathbb{R}^n$ and suppose Assumption 2 holds. For all $T > 0$ we have

$$R(\tilde{P}^x_{\tau \wedge T}||P^x_{\tau \wedge T}) \leq \frac{1}{2} E_P \left[ \int_0^{\tau_T} \|\beta(s, \tilde{X}^x_s, \tilde{Z})\|^2 ds \right],$$

where $\tau_T = (\tau \circ \tilde{X}^x) \wedge T$. In terms of the base process, we have

$$R(\tilde{P}^x_{\tau \wedge T}||P^x_{\tau \wedge T}) \leq \frac{1}{2} E_P \left[ \int_0^{\tau_T} \|\beta(s, X^x_s, Z)\|^2 ds \rho^x_T \right],$$

where $\sigma^x_T = (\tau \circ X^x) \wedge T$ and

$$\rho^x_T = \exp \left( \int_0^{\tau_T} \beta(s, X^x_s, Z) \cdot dW^x_s - \frac{1}{2} \int_0^{\tau_T} \|\beta(s, X^x_s, Z)\|^2 ds \right).$$
A class of alternative models that are of particular case of interest is covered by the following corollary:

**Corollary 7** Suppose the base model is the solution to the following SDE on $\mathbb{R}^n$,
\[
dX_t^b = b(t, X_t^b)dt + \sigma(t, X_t^b)dW_t, \quad X_0^b = x, \tag{171}
\]
and the alternative model is given by
\[
d\tilde{X}_t^a = b(t, \tilde{X}_t^a)dt + \sigma(t, \tilde{X}_t^a)\beta(t, \tilde{X}_t^a, \tilde{Y}_t^a, \tilde{Z}_t)dt + \sigma(t, \tilde{X}_t^a)d\tilde{W}_t^1, \quad \tilde{X}_0^a = x, \tag{172}
\]
which includes a modified drift that depends on external data $\tilde{Z}$, and is also coupled to another SDE
\[
d\tilde{Y}_t^a = f(t, \tilde{X}_t^a, \tilde{Y}_t^a, \tilde{Z}_t)dt + \eta(t, \tilde{X}_t^a, \tilde{Y}_t^a, \tilde{Z}_t)d\tilde{W}_t^2, \tag{173}
\]
on $\mathbb{R}^k$, where $\tilde{W}_t^1$ and $\tilde{W}_t^2$ are independent Wiener processes.
Suppose also we have $Y_t^b$ such that
\[
dY_t^b = f(t, X_t^b, Y_t^b, Z_t)dt + \eta(t, X_t^b, Y_t^b, Z_t)dW_t', \quad Y_0^b = y, \tag{174}
\]
where $W'$ is a Wiener process independent from $W$.

If the base, $(X_t^b, Y_t^b)$, and alternative, $(\tilde{X}_t^a, \tilde{Y}_t^a)$, systems satisfy Assumption 2, we have
\[
R((X_t^b)^P | x_{\tau \wedge T}, (X_t^b)^P | x_{\tau \wedge T}) \leq R((X_t^b, Y_t^b)^P | x_{\tau \wedge T}, (X_t^b, Y_t^b)^P | x_{\tau \wedge T})
\]
\[
\leq \frac{1}{2} \mathbb{E}_P \left[ \int_0^{r \wedge T} \|\beta(s, \tilde{X}_s^a, \tilde{Y}_s^a, \tilde{Z}_s)\|^2 ds \right] \tag{175}
\]
for any stopping time $\tau$ on $C([0, \infty), \mathbb{R}^n)$. Here $\pi_1$ denotes the map that takes a path in $C([0, \infty), \mathbb{R}^n \times \mathbb{R}^k)$ and returns the path of the $\mathbb{R}^n$-valued component.

### A.3 Time-Integral Processes with a Smooth Density

Here we discuss an alternative approach to time-integral QoIs (see Section 4.4) that is applicable when the measure defining the integral QoI has a smooth density and the QoI is stopped at $\tau = \infty$.

Suppose that $dt = h(t)dt$ with $h \in C^1$, $h' \leq 0$, $th'(t) \in L^1([0, \infty))$, and $th(t) \to 0$ as $t \to \infty$. The quantity of interest is then
\[
F \equiv F_\tau = \int_0^\infty f_s h(t)dt. \tag{176}
\]

For simplicity, we will assume here that $f$ is uniformly bounded.

Integrating by parts, we can rewrite $F$ as
\[
F = \int_0^\infty \int_0^t f_s ds (-h'(t))dt = h(0) \int_0^\infty \int_0^t f_s dp(dt), \tag{177}
\]
where we define the probability measure $dp = -\frac{h'(t)}{h(0)} dt$ and note that $h(0) > 0$, since $h$ is assumed to be non-increasing and integrate to 1.

Using Fatou’s theorem and Eq. (33) we obtain
\[
\pm (E_P[F] - E_P[F]) \leq h(0) \int_0^\infty \inf_{c > 0} \left\{ \frac{1}{c} A_{G_t^P}^c(\pm ct) + \frac{1}{c} R(P | x_t | P | x_t) \right\} \rho(dt), \tag{178}
\]
where $G_t \equiv \frac{1}{c} \int_0^t f_s ds$ and
\[
A_{G_t^P}^c(\pm ct) = \log E_P \left[ \exp \left( \pm c \left( \int_0^t f_s ds - E_P \left[ \int_0^t f_s ds \right] \right) \right) \right]. \tag{179}
\]
A.4 Calculations Required in the Analysis of the Vasicek Model

Here we record a pair of lemmas that are needed for the analysis of option values under the Vasicek interest rate model. See Section 5.5.1 for definitions of the notation.

**Lemma 9** Suppose $r > \frac{\gamma^2}{2\sigma^2}$. Then $\lim_{n \to \infty} \int_0^n r + \Delta r_s ds = \infty$ a.s. under both the base and the alternative models.

**Proof** The distribution of $\Delta r$ is the same under both the base and alternative models, so it does not matter which one we consider.

As shown in [1], $\int_0^n r + \Delta r_s ds$ is normally distributed for all $t \geq 0$, with mean $m_t \equiv rt$ and variance

$$\sigma_t^2 = \frac{\sigma^2}{2\gamma^2} (2\gamma t + 4e^{-\gamma t} - e^{-2\gamma t} - 3).$$

(180)

Using this, for any $R > 0$ we can compute

$$P \left( \liminf_{n \to \infty} \int_0^n r + \Delta r_s ds < R \right) \leq \lim_{N \to \infty} \sum_{n \geq N} P \left( \int_0^n r + \Delta r_s ds < R \right)$$

$$\leq \lim_{N \to \infty} \sum_{n \geq N} P \left( \exp \left( - \int_0^n r + \Delta r_s ds \right) > \exp(-R) \right)$$

$$\leq e^R \lim_{N \to \infty} \sum_{n \geq N} e^{-rn + \frac{3}{2}n}$$

$$\leq e^{R+\frac{3}{2}n^3} \lim_{N \to \infty} \sum_{n \geq N} e^{-\left(r - \frac{3}{2}\gamma^2\right)n} = 0.$$

Therefore

$$P \left( \liminf_{n \to \infty} \int_0^n r + \Delta r_s ds < \infty \right) = \lim_{R \to \infty} P \left( \liminf_{n \to \infty} \int_0^n r + \Delta r_s ds < R \right) = 0.$$

(182)

**Lemma 10**

$$\rho_t \equiv \exp \left( \int_0^t \sigma^{-1} \Delta r_s dW_s - \frac{1}{2} \int_0^t \sigma^{-2} |\Delta r_s|^2 ds \right).$$

(183)

is a $P$-martingale.

**Proof** As shown in [33] (see Corollary 5.14), it suffices to prove that there exists $\{t_k\}_{k=0}^\infty$ with $0 = t_0 < t_1 < \ldots < t_n \to \infty$ such that

$$E \left[ \exp \left( \frac{1}{2} \int_{t_{k-1}}^{t_k} \sigma^{-2} |\Delta r_s|^2 ds \right) \right] < \infty \text{ for all } n.$$  

(184)

Letting $\Delta t_n = t_n - t_{n-1}$ and using Jensen’s inequality, we have

$$E \left[ \exp \left( \frac{1}{2} \int_{t_{k-1}}^{t_k} \sigma^{-2} |\Delta r_s|^2 ds \right) \right] \leq E \left[ \Delta t_n^{-1} \int_{t_{k-1}}^{t_k} \exp \left( \frac{\Delta t_n}{2} \sigma^{-2} |\Delta r_s|^2 \right) ds \right]$$

$$\leq \sup_{t_{k-1} \leq s \leq t_k} E \left[ \exp \left( \frac{\Delta t_n}{2} \sigma^{-2} |\Delta r_s|^2 \right) \right].$$

(185)

$\Delta r_s$ is normal with mean 0 and variance $\sigma^2(1 - e^{-2\gamma t})/(2\gamma)$. Hence, if we fix $\Delta t_n = \epsilon$ small enough then the upper bound in Eq. (185) is finite for all $n$. □
A.5 Parametric Models, Goal-Oriented Fisher Information, and Linearized UQ Bounds

Here, we provide additional detail regarding the notion of goal-oriented Fisher information that was introduced in Section 3.1. The main theorem is:

**Theorem 5** Suppose the following:

1. \( P_\theta, \theta \in U \) (an open subset of \( \mathbb{R}^n \)) are probability measures.
2. \( F \in L^2(P_\theta) \) for all \( \theta \) and is \( \mathcal{G} \)-measurable for some sub sigma-algebra \( \mathcal{G} \subset \mathcal{F} \).
3. \( \mu \) is a sigma-finite positive measure, \( P_\theta|_\mathcal{G} \ll \mu \) for all \( \theta \), and we have a version \( p_\theta|_\mathcal{G} \equiv \frac{dP_\theta|_\mathcal{G}}{d\mu} \geq 0 \) that is \( C^1 \) in \( \theta \).
4. There exists \( h \in L^1(\mu) \) such that \( \| \nabla p_\theta|_\mathcal{G} \| \leq h \) on \( U \times \Omega \).
5. There exists \( g \in L^1(\mu) \) such that \( \| F \nabla p_\theta|_\mathcal{G} \| \leq g \) on \( U \times \Omega \).

Define the goal-oriented score,

\[
V_\theta^G \equiv 1_{p_\theta|_\mathcal{G} \neq 0} \nabla p_\theta|_\mathcal{G} / p_\theta|_\mathcal{G} = 1_{p_\theta|_\mathcal{G} \neq 0} \nabla \log(p_\theta|_\mathcal{G}),
\]

and the goal-oriented Fisher information in the direction \( v \in \mathbb{R}^n \),

\[
I_\theta^G(v) \equiv E_\theta[\nabla v \cdot V_\theta^G] = E_\theta[1_{p_\theta|_\mathcal{G} \neq 0} (\nabla v p_\theta|_\mathcal{G} / p_\theta|_\mathcal{G})^2].
\]

(Note that, without further assumptions, this expectation is possibly \(+\infty\).)

Then for all \( \theta \in U \) we have:

1. \( E_\theta[V_\theta^G] = 0 \).
2. \( \theta \to E_\theta[F] \) is \( C^1 \) on \( U \) and
   \[
   \nabla E_\theta[F] = E_\theta[(F - E_\theta[F]) V_\theta^G].
   \]
3. \( |\nabla_h E_\theta[F]| \leq \sqrt{\text{Var}_{P_\theta}[F] I_\theta^G(h)} \) for all \( h \in \mathbb{R}^n \).

Eq. (190) results in the linearized UQ bound:

\[
|E_{\theta+h}[F] - E_\theta[F]| \leq \sqrt{\text{Var}_{P_\theta}[F] I_\theta^G(h)} + o(h).
\]

Proof Use the dominated convergence theorem to differentiate \( 1 = \int p_\theta^G \, d\mu \) and \( E_\theta[F] = \int F p_\theta^G \, d\mu \) under the integral signs. The bound Eq. (190) then follows from the Cauchy-Schwarz inequality.

Given sufficient smoothness and integrability assumptions on \( p_\theta^G \) and its derivatives in order to justify differentiating under the integral (which we do not state explicitly here), one also obtains an expansion of the goal-oriented relative entropy in terms of the goal-oriented Fisher information:

**Lemma 11**

\[
R(P_{\theta+h}|_\mathcal{G} \| P_\theta|_\mathcal{G}) = \frac{1}{2} I_\theta^G(h) + O(||h||^3).
\]

**Remark 24** When \( F \) is bounded, linearizing the non-perturbative UQ bound from Theorem 1 (see the derivation of Eq. (16) in [20]) and combining it with Eq. (192) results in the same leading-order-term in \( h \) as Eq. (191).
A.5.1 Example: Linearization about Brownian Motion with Constant Drift

Here we illustrate the use of goal-oriented Fisher information by deriving linearized bounds on hitting-time probabilities for a perturbation to Brownian motion with constant drift (see Section 5.1 for more detail regarding this model). Specifically, we consider the probability that the level-$a$ hitting time, $\tau_a$, is less than $T$ in a perturbed model

$$d\tilde{X}_t = (\mu + \alpha \beta(t, \tilde{X}_t, Y_t, Z_t))dt + d\tilde{W}_t, \quad \tilde{X}_0 = 0,$$

(193)

where $\beta$ is bounded, and will linearize in the real parameter $\alpha$.

First use Girsanov’s theorem, together with Lemma 4, to write

$$\tilde{P}_0^\alpha(\tau_a < T) = E_P \left[ 1_{\tau_a \circ X < T} \exp \left( \alpha \int_0^{\tau_a \circ X < T} \beta(s, X_s, Y_s, Z_s) dW_s - \frac{\alpha^2}{2} \int_0^{\tau_a \circ X < T} |\beta(s, X_s, Y_s, Z_s)|^2 ds \right) \right],$$

(194)

where $\sigma_T \equiv (\tau_a \wedge T) \circ X$ and the subscript on $\tilde{P}_0^\alpha$ denotes the parameter value, $\alpha$, being used in the alternative model.

Eq. (194) is analytic in $\alpha$ and can be differentiated under the integral, so we obtain

$$\frac{d}{d\alpha} \bigg|_{\alpha=0} \tilde{P}_0^\alpha(\tau_a < T) = E_P \left[ (1_{\tau_a \circ X < T} - P^0(\tau_a < T)) \int_0^{\tau_a \circ X < T} \beta(s, X_s, Y_s, Z_s) dW_s \right].$$

(195)

Note that the quantities under the expectation are all up to the stopping-time i.e. we are using a goal-oriented score, as in Eq. (186).

Using the Cauchy-Schwarz inequality we obtain

$$\frac{d}{d\alpha} \bigg|_{\alpha=0} \tilde{P}_0^\alpha(\tau_a < T) \leq \sqrt{P^0(\tau_a < T) - P^0(\tau_a < T)^2} E_P \left[ \int_0^{\tau_a \circ X < T} |\beta(s, X_s, Y_s, Z_s)|^2 ds \right].$$

(196)

The first term under the root is the variance of the QoI and the second is the goal-oriented Fisher information; this is the general form of the sensitivity bound given in Theorem 5, see Eq. (190).

One can obtain a computable bound by, for example, bounding $\beta$:

$$\frac{d}{d\alpha} \bigg|_{\alpha=0} \tilde{P}_0^\alpha(\tau_a < T) \leq \|\beta\|_\infty \sqrt{P^0(\tau_a < T) - P^0(\tau_a < T)^2} E_P [\tau_a \wedge T]^{1/2}.$$  

(197)

These quantities can be evaluated by using the known distribution of $\tau_a$ [33]. We omit the details.