Risk-Averse Optimal Control

Alois Pichler∗ Ruben Schlotter†
November 4, 2019

Abstract

In the context of multistage stochastic optimization, it is natural to consider nested risk measures, which originate by repeatedly composing risk measures, conditioned on realized observations. Starting from this discrete time setting, we extend the notion of nested risk measures to continuous time by adapting the risk levels in a time dependent manner. This time dependent modification is necessary for a risk measure to be non-degenerate in continuous time. Moreover, the Entropic Value-at-Risk turns out to be the natural and universal choice for a coherent nested risk measure, in the context of optimal control.

Consequently, we construct the risk-averse analogue of the infinitesimal generator based on risk measures and obtain risk-averse Hamilton–Jacobi–Bellman equations.

Keywords: Risk measures, Optimal control, Stochastic processes
Classification: 90C15, 60B05, 62P05

1 Introduction

Recently, increased attention is paid to optimal control problems, which are risk-averse in addition. In a general stochastic programming framework, these problems are large-scale and computationally expensive. Specific problems allow dynamic characterizations, which can be solved much faster and more efficiently. Most prominently, dynamic equations appear in optimal control problems.

This paper extends dynamic control problems, which are risk-neutral, to a risk-averse environment. For this purpose, it is crucial to compose (or nest) risk measures, as exactly nested risk measures allow quantifying risk via dynamic programming equations.

More specifically, this paper formulates risk-averse optimal control problems by extending the discrete time setting to a continuous time framework. Here, a decision maker incurs an uncertain stream of costs $c(\cdot)$ over time and his goal is to manage and minimize the total costs incurred. The risk-averse decision maker intends to guard against undesired scenarios in particular. The nested risk measure $\rho_{0:T}$ governs the risk over the finite time horizon $[0, T]$ and the optimization problem in consideration thus is

$$\inf_{u(\cdot)} \rho_{0:T} \left( \int_0^T c(t, u(t)) \, dt \right),$$

where $u(\cdot)$ are feasible control policies. Here, the nested risk measure $\rho_{0:T}$ allows to guard against risk in every instant of time, which is the typical objective of a risk manager permanently hedging against risk.

∗Both authors: Technische Universität Chemnitz, 09126 Chemnitz, Germany
†Contact: ruben.schlotter@math.tu-chemnitz.de
Classical, non-nested risk measures only assess the accumulated position at terminal time $T$, so that the risk manager can never intervene.

Nested risk measures were introduced in Ruszczyński and Shapiro (2006), for which Philpott et al. (2013) provide an economic interpretation as an insurance premium on a rolling horizon basis. Pichler and Shapiro (2018) elaborate that exactly these risk measures allow expanding the associated control problem to a risk-averse framework. New dynamic programming equations, derived below, then reflect the risk-averse character (cf. De Lara and Leclère (2016) for a discussion on risk measures and dynamic optimization).

The risk-neutral stochastic optimal control problem is extensively studied in the literature, see, e.g., Yong and Zhou (1999) and Fleming and Soner (2006). However, the risk-averse case has not received much attention yet. Çavuş and Ruszczyński (2014) and Fan and Ruszczyński (2014) study risk-averse control problems in discrete time, Dentcheva and Ruszczyński (2018) analyze the value functions in a continuous time Markov chain model and derive dynamic programming equations. In a very different setting, Kouri and Surowiec (2018) study a risk-averse PDE-constrained problem. Additionally, Ruszczyński and Yao (2015) discuss a risk-averse optimal control problem in continuous time using $g$-expectations and derive an Hamilton–Jacobi–Bellman equation, whereas Guigues and Römisch (2012) consider a class of multiperiod risk measures and provide dynamic programming equations in this setting. This paper, in contrast, describes the continuous time evolution equations based on a classical approach using nested risk measures and in this way provides a much simpler understanding of the governing evolution equations.

For the purpose of this study, special focus is given to nested risk measures based on the Entropic Value-at-Risk defined by

$\text{EV}_{\beta}(Y) := \sup \{ EYZ : Z \geq 0, E Z = 1, E Z \log Z \leq \beta \}, \quad (1)$

where $\beta \geq 0$ is the coefficient of risk aversion (or risk level) and $E Z \log Z$ is the Shannon entropy (see also Dupuis et al. (2016) for a discussion of entropy and stochastic processes). Its nested counterpart allows an explicit formula for the Brownian motion, provided that $\beta$ reflects the risk of the time horizon spanned. We give explicit formulas for the nested Entropic Value-at-Risk for Brownian motion and investigate more general diffusion processes. Other risk measures can be nested as well, but their composite counterpart is often degenerate. It is of independent interest that the Average Value-at-Risk (also known as Conditional Value-at-Risk), the most prominent risk measure, degenerates as well when nested naively.

The infinitesimal generator is the essential tool for classical stochastic optimal control in continuous time, see Fleming and Soner (2006), e.g. In this classical, risk-neutral setting, the infinitesimal generator of an Itô process is a linear second order differential operator. We introduce the risk-averse analogue, called risk generator and derive explicit expressions for Itô processes. Here, the risk generator is a second order operator, which is nonlinear in the first derivative. We then derive risk-averse Hamilton–Jacobi–Bellman equations, which extend the classical, risk-neutral dynamic programming equations to the risk-averse setting.

Based on the derivation of the risk generator we conclude that the Entropic Value-at-Risk is the natural choice for a nested risk measure in continuous time by adapting the risk levels in a natural and intuitive way. Furthermore, this approach provides an interpretation of the Hamilton–Jacobi–Bellman equation given in Ruszczyński and Yao (2015).

Outline of the paper. Section 2 introduces conditional and nested risk measures and provides the general mathematical setup in a discrete time setting. Section 3 deals with the extension to continuous time. Section 4 provides risk evaluations for Itô processes and introduces the risk generator of a stochastic process. We close with a discussion on other risk measures besides the Entropic Value-at-Risk. Section 5 introduces the risk-averse stochastic control problem and derives the corresponding risk-averse Hamilton–Jacobi–Bellman partial differential equation.
2 Notation and preliminaries

We consider a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathcal{T}}, P)\) and associate \(t \in \mathcal{T}\) with stage or time. The entropic space is

\[ E := \left\{ Y : \Omega \to \mathbb{R} \mid E e^{\ell Y} < \infty \text{ for all } \ell \in \mathbb{R} \right\} \]

and we shall assume throughout the paper that the process \(X = (X_t)_{t \in \mathcal{T}}\) is adapted to the filtration \((\mathcal{F}_t)_{t \in \mathcal{T}}\) with marginals \(X_t \in E\) for all \(t \in \mathcal{T}\). We shall write \(X_t \mathcal{F}_t\) to indicate that \(X_t\) is measurable with respect to \(\mathcal{F}_t\).

2.1 Conditional risk measures and EV@R

We recall the definition of law invariant, coherent risk measures \(\rho : L \to \mathbb{R}\) defined on some vector space \(L\) of \(\mathbb{R}\)-valued random variables first. They satisfy the following axioms introduced by Artzner et al. (1999).

A1 Monotonicity: \(\rho(Y) \leq \rho(Y')\), provided that \(Y \leq Y'\) almost surely;
A2 Translation equivariance: \(\rho(Y + c) = \rho(Y) + c\) for \(c \in \mathbb{R}\);
A3 Convexity: \(\rho(1 - \lambda) Y + \lambda Y' \leq (1 - \lambda) \rho(Y) + \lambda \rho(Y')\) for \(\lambda \in [0, 1]\);
A4 Positive homogeneity: \(\rho(\lambda Y) = \lambda \rho(Y)\) for \(\lambda \geq 0\);
A5 Law invariance: \(\rho(Y) = \rho(Y')\), whenever \(Y\) and \(Y'\) have the same law, i.e., \(P(Y \leq y) = P(Y' \leq y)\) for all \(y \in \mathbb{R}\).

Remark 1. Any functional \(\rho : L \to \mathbb{R}\) satisfying the Axioms A1–A4 can be represented by

\[ \rho(Y) = \sup_{Q \in \mathcal{Q}} E_Q Y \]

for a convex set of probability measures \(\mathcal{Q}\) absolutely continuous with respect to \(P\) (cf. Delbaen (2002)). We consider the conditional risk measures \(\rho'\) with respect to the sigma algebra \(\mathcal{F}_t\) defined by

\[ \rho'(Y | \mathcal{F}_t) := \text{ess sup}_{Q \in \mathcal{Q}} E_Q [Y | \mathcal{F}_t] . \]  

(2)

Note that \(\rho'\) satisfies conditional versions of the Axioms A1–A5. For further details, we refer the interested reader to Ruszczynski and Shapiro (2006) and Riedel (2004). For the essential supremum of a set of random variables as in (2) we refer to Karatzas and Shreve (1998, Appendix A).

Definition 2 (Entropic Value-at-Risk). The Entropic Value-at-Risk of a random variable \(Y \in E\) at risk level \(\beta \geq 0\) is (cf. (1))

\[ \text{EV@R}_\beta(Y) = \sup \{ EYZ : Z \geq 0, EZ = 1, E Z \log Z \leq \beta \} . \]

(3)

Similarly, for the risk level \(0 \leq \beta < \mathcal{F}_t\), we define the conditional Entropic Value-at-Risk \(\text{EV@R}_\beta(Y | \mathcal{F}_t)\) as

\[ \text{EV@R}_\beta(Y | \mathcal{F}_t) := \text{ess sup} \{ E[ZY | \mathcal{F}_t] : 0 \leq Z, E[Z | \mathcal{F}_t] = 1, E[Z \log Z | \mathcal{F}_t] \leq \beta \} . \]

(4)

Remark 3. For a random variable \(Y \mathcal{F}_t\), the density \(Z\) in the defining equation (4) may be chosen to satisfy \(Z \mathcal{F}_t\). Indeed, the density \(Z_t := E[Z | \mathcal{F}_t]\) satisfies all constraints as well, as can be seen by applying the conditional Jensen inequality

\[ E[Z_t \log Z_t | \mathcal{F}_t] \leq E[E[Z \log Z | \mathcal{F}_t] | \mathcal{F}_t] = E[Z \log Z | \mathcal{F}_t] . \]

The tower property of the expectation finally insures the assertion.
For future reference we provide a closed form for the Entropic Value-at-Risk for Gaussian random variables.

**Proposition 4 (EV@R of Gaussians).** For a normally distributed random variable \( Y \sim N(\mu, \sigma^2) \) with \( \sigma \geq 0 \) and \( a, b \in \mathbb{R} \) it holds that

\[
\text{EV@R}_b(a + bY) = a + b\mu + \sigma |b| \sqrt{2\beta}.
\]

**Proof.** Let \( Y \sim N(\mu, \sigma^2), \beta \in [0, \infty) \) and consider the alternative representation of the Entropic Value-at-Risk (cf. Ahmadi-Javid (2012))

\[
\text{EV@R}_b(Y) = \inf_{\ell > 0} \frac{1}{\ell} \left( \beta + \log E e^{\ell Y} \right).
\]

It holds that \( E e^{\ell Y} = \exp \left( \mu\ell + \frac{1}{2} \ell^2\sigma^2 \right) \) and thus

\[
\frac{1}{\ell} \beta + \frac{1}{\ell} \log \left( \exp^{\mu\ell + \frac{1}{2} \ell^2\sigma^2} \right) = \frac{1}{\ell} \beta + \mu + \frac{1}{2} \ell \sigma^2,
\]

which attains its infimum at \( \ell^* = \frac{1}{\sigma} \sqrt{2\beta} \). The Entropic Value-at-Risk thus is

\[
\text{EV@R}_b(Y) = \frac{1}{\ell^*} \beta + \mu + \frac{1}{2} \ell^* \sigma^2 = \mu + \sigma \sqrt{2\beta}.
\]

Finally notice that \( a + b Y \sim N(a + b\mu, b^2 \sigma^2) \) and hence

\[
\text{EV@R}_b(a + b Y) = a + b\mu + |b| \sigma \sqrt{2\beta}.
\]

the assertion. \( \square \)

### 2.2 Nested risk measures

Nested risk measures are compositions of conditional risk measures. This section elaborates general properties of nested risk measures. The results are then discussed in more detail for the Entropic Value-at-Risk. Throughout, we will always consider risk on the time interval \([0, T]\), \( T > 0 \), and denote by \( \mathcal{P} := \{0 = t_0 < t_1 < \cdots < t_n = T\} \) a finite partition of the interval \([0, T]\) including its endpoints. With \( \Delta t_i := t_{i+1} - t_i \) we denote the time step and \( \|\mathcal{P}\| := \max_{0 \leq i < n} \Delta t_i \) is the mesh size of \( \mathcal{P} \).

**Definition 5 (Nested risk measures).** Let \( \mathcal{P} \) be a partition of the interval \([0, T]\) and let \( Y \triangleleft \mathcal{F}_T \). For a collection of conditional risk measures \( (\rho^t)_{t \in \mathcal{P}} \) and \( i < n \), the nested risk measure is

\[
\rho^{t_{i+1}:tn} (Y \mid \mathcal{F}_{t_i}) := \rho^t (\rho^{t_{i+1}:t_{i-1}} (Y \mid \mathcal{F}_{t_{i-1}}) \ldots \mid \mathcal{F}_{t_i}) \mid \mathcal{F}_{t_i}.
\]

**Remark 6 (Risk martingales).** Nested risk measures naturally follow a martingale like pattern. Indeed, the stochastic process \( Y_t := \rho^{t_{i+1}:tn} (Y \mid \mathcal{F}_t) \) satisfies

\[
\rho^{t_{i+1}:tn} (Y_{t_{i+1}} \mid \mathcal{F}_{t_i}) = \rho^{t_{i+1}:t_{i+1}} (\rho^{t_{i+1}:tn} (Y \mid \mathcal{F}_{t_{i+1}}) \mid \mathcal{F}_{t_i})
\]

\[
= \rho^{t_{i+1}:t_{i+1}} (\rho^{t_{i+1}:t_{i+2}} (\rho^{t_{i+2}:t_{i+3}} (Y \mid \mathcal{F}_{t_{i+3}}) \ldots \mid \mathcal{F}_{t_{i+1}}) \mid \mathcal{F}_{t_i})
\]

\[
= Y_{t_i}.
\]

We call this process \( (Y_t)_{t \in \mathcal{P}} \) a risk martingale with respect to the family of risk measures \( (\rho^t)_{t \in \mathcal{P}} \).
Often, the risk evaluation $\rho^{0:T}(X_T)$ of the terminal value $X_T$ of some stochastic process $X$ is of interest. The terminal value $X_T$ can then be represented as the sum of its increments $\Delta X_{t_j} := X_{t_{j+1}} - X_{t_j}$ as

$$X_T = X_0 + \sum_{j=1}^{n-1} \Delta X_{t_j}.$$  

As a consequence of translation equivariance we have the following useful proposition.

**Proposition 7.** Suppose that $(X_t)_{t \in \mathcal{P}}$ is a discrete time stochastic process adapted to the filtration $\mathcal{F}_t, t \in \mathcal{P}$. The nested risk measure (5) is

$$\rho^{\beta, T}(X_T | \mathcal{F}_t) = X_t + \rho^{\beta_1}(\Delta X_t) + \rho^{\beta_{i+1}}(\Delta X_{t_{i+1}} | \mathcal{F}_{t_{i+1}}) \cdots \rho^{\beta_{n-1}}(\Delta X_{t_{n-1}} | \mathcal{F}_{t_{n-1}}) | \mathcal{F}_t).$$

It is therefore sufficient to study conditional risk evaluations of increments. The next section exploits this observation by giving explicit formulas in important cases.

### 2.3 The nested Entropic Value-at-Risk

In what follows we consider the nested Entropic Value-at-Risk, but we adjust the risk level to span the respective time interval. For convenience of the reader and future reference we reemphasize these details in the following definition.

**Definition 8** (Nested Entropic Value-at-Risk). Let $\mathcal{P}$ be a partition of $[0,T]$ and $Y \in \mathcal{F}_T$. For a vector of risk levels $\beta := (\beta_0, \Delta t_0, \ldots, \beta_{n-1}, \Delta t_{n-1})$, the nested Entropic Value-at-Risk is

$$nEV@R_{\beta}^{0:T}(Y | \mathcal{F}_t) := EV@R_{\beta_1}^{\Delta t_0}(Y | \mathcal{F}_0) \cdots EV@R_{\beta_{n-1}}^{\Delta t_{n-1}}(Y | \mathcal{F}_{t_{n-1}}) | \mathcal{F}_t).$$

To emphasize the dependence on the partition we will also write $nEV@R_{\beta}^{0:T}(Y | \mathcal{F}_t)$ for (7). Furthermore, for the trivial sigma algebra $\mathcal{F}_0 = \{\emptyset, \Omega\}$, we simply write $nEV@R_{\beta}^{0:T}(Y)$.

As a corollary to Proposition 4, we provide an explicit formula for the $nEV@R$ for a Wiener process evaluated at discrete time points.

**Proposition 9** (Nested EV@R for the Gaussian random walk). Let $W = (W_t)_{t \in \mathcal{P}}$ be a Wiener process evaluated on the partition $\mathcal{P}$. Furthermore, let $\beta := (\beta_0, \Delta t_0, \ldots, \beta_{n-1}, \Delta t_{n-1})$ be a vector of risk levels. Then the nested Entropic Value-at-Risk is

$$nEV@R_{\beta}^{0:T}(W_T) = \sum_{i=0}^{n-1} \Delta t_i \sqrt{2 \beta_i}. $$

**Proof.** Note that $W_{t_{i+1}} - W_{t_i} \sim N(0, \Delta t_{i+1} - t_i)$ and by Proposition 4, the conditional Entropic Value-at-Risk is

$$EV@R_{\beta_1}^{\Delta t_0}(W_{t_{i+1}} | W_{t_i}) = W_{t_i} + \sqrt{\Delta t_i} \sqrt{2 \beta_i} \Delta t_i.$$  

Iterating as in Proposition 7 shows

$$nEV@R_{\beta}^{0:T}(W_T) = \sum_{i=0}^{n-1} \sqrt{\Delta t_i} \sqrt{2 \beta_i} \Delta t_i = \sum_{i=0}^{n-1} \Delta t_i \sqrt{2 \beta_i},$$

the assertion.  \(\square\)
Statement: Comparing the explicit formula (8) with Proposition 4 we observe the surprising consistency property

\[
n\text{EV@} R_{\beta}^\nu(W_T) = \text{EV@} R_{\beta_0}\beta T(W_T)
\]

for the risk levels \( \beta = (\beta_0 \Delta t_0, \ldots, \beta_n \Delta t_{n-1}) \) with constant risk level \( \beta_0 \). This is a consequence of the parametrization chosen in (3), the definition of the Entropic Value-at-Risk.

Nested risk measures are coherent (Axioms A1–A4) if each conditional risk measure \( \rho' \) is coherent. Furthermore, for fixed \( Y \in E \), the mapping \( \beta \mapsto \text{EV@} R_{\beta}(Y \mid F_t) \) is monotone increasing with

\[
\text{EV@} R_\beta(Y \mid F_t) = E[Y \mid F_t]
\]

and hence (7) gives the lower bound

\[
n\text{EV@} R_{\beta}^\nu T(Y \mid F_t) \geq E[Y \mid F_t].
\]

As nested risk measures are convex they admit a general dual representation.

\textbf{Remark 10 (Parametrization).} Comparing the explicit formula (8) with Proposition 4 we observe the surprising consistency property

\[
n\text{EV@} R_{\beta}^\nu(W_T) = \text{EV@} R_{\beta_0}\beta T(W_T)
\]

for the risk levels \( \beta = (\beta_0 \Delta t_0, \ldots, \beta_n \Delta t_{n-1}) \) with constant risk level \( \beta_0 \). This is a consequence of the parametrization chosen in (3), the definition of the Entropic Value-at-Risk.

Nested risk measures are coherent (Axioms A1–A4) if each conditional risk measure \( \rho' \) is coherent. Furthermore, for fixed \( Y \in E \), the mapping \( \beta \mapsto \text{EV@} R_{\beta}(Y \mid F_t) \) is monotone increasing with

\[
\text{EV@} R_\beta(Y \mid F_t) = E[Y \mid F_t]
\]

and hence (7) gives the lower bound

\[
n\text{EV@} R_{\beta}^\nu T(Y \mid F_t) \geq E[Y \mid F_t].
\]

As nested risk measures are convex they admit a general dual representation.

\textbf{Lemma 11 (Dual representation).} Let \( \beta = (\beta_0 \Delta t_0, \ldots, \beta_n \Delta t_{n-1}) \) be a vector of risk levels. The nested Entropic Value-at-Risk for \( Y \in E \) has the dual representation

\[
n\text{EV@} R_{\beta}^{0,T}(Y) = \sup \left\{ E \left[ Y \mid Z_{t_1} \cdots Z_{t_n} \right] \mid E \left[ Z_{t_i} \log Z_{t_i} \mid F_{t_{i-1}} \right] \leq \beta_{i-1} \Delta t_{i-1}, 0 \leq Z_{t_i}, \text{ for } i = 1, \ldots, n \right\}.
\]

Furthermore, the dual representation can be reformulated as a supremum over risk-neutral martingales \( Z = (Z_t)_{t \in T} \). More precisely,

\[
n\text{EV@} R_{\beta}^{0,T}(Y) = \sup \left\{ E \left[ Y \mid Z_{t_1} \cdots Z_{t_n} \right] \mid E \left[ Z_{t_i} \log Z_{t_i} \mid F_{t_{i-1}} \right] \leq \beta_{i-1} \Delta t_{i-1}, Z_{t_i} = Z_{t_{i-1}} + \log Z_{t_{i-1}}, 0 \leq Z_{t_i} < F_{t_i}, \text{ for } i = 1, \ldots, n \right\}.
\]

\textit{Proof.} We demonstrate (10) by induction and it is sufficient to discuss the two-stage case with stages \( 0 = t_0 < t_1 < t_2 \). Let \( Z_{t_1} \) (\( Z_{t_2} \), resp.) satisfy the constraints in \( \text{EV@} R_{\beta_0}^{0,T}(\cdot \mid F_{t_0}) \) (in \( \text{EV@} R_{\beta_1}^{0,T}(\cdot \mid F_{t_1}) \), resp.). Then \( Z_{t_1} \) and \( Z_{t_2} \) also satisfy the constraints in

\[
\sup \left\{ E \left[ Y Z_{t_1} Z_{t_2} \right] \mid E \left[ Z_{t_i} \log Z_{t_i} \mid F_{t_{i-1}} \right] \leq \beta_{i-1} \Delta t_{i-1}, 0 \leq Z_{t_i}, \text{ for } i = 1, 2 \right\}
\]

and hence ‘\( \geq \)’ in (10).

For the converse inequality, recall the definition of the conditional Entropic Value-at-Risk in (4).

\[
\text{EV@} R_{\beta_1}^{0,T}(Y \mid F_{t_1}) = \text{ess sup} \left\{ E \left[ Y Z_{t_2} \mid F_{t_1} \right] \mid E \left[ Z_{t_2} \log Z_{t_2} \mid F_{t_1} \right] \leq \beta_1 \Delta t_1, E \left[ Z_{t_2} \mid F_{t_1} \right] = 1, 0 \leq Z_{t_2} \right\},
\]

and hence we may express the nested Entropic Value-at-Risk \( \text{EV@} R_{\beta_1}^{0,T}(\text{EV@} R_{\beta_1}^{0,T}(Y \mid F_{t_1})) \) by

\[
\sup \left\{ E \left[ \text{ess sup} \left\{ E \left[ Y Z_{t_1} Z_{t_2} \mid F_{t_1} \right] \mid E \left[ Z_{t_1} \log Z_{t_1} \mid F_{t_1} \right] \leq \beta_1 \Delta t_1, E \left[ Z_{t_1} \mid F_{t_1} \right] = 1, 0 \leq Z_{t_2} \right\} \right\} \right\}.
\]

As the objective function in (12), for fixed \( Z_{t_1} \), is always larger than \( E Y Z_{t_1} Z_{t_2} \) for all feasible \( Z_{t_2} \), we conclude the inequality ‘\( \geq \)’, i.e.,

\[
n\text{EV@} R_{\beta}(Y) \geq \sup \left\{ E \left[ Y Z_{t_1} Z_{t_2} \right] \mid E \left[ Z_{t_i} \log Z_{t_i} \mid F_{t_{i-1}} \right] \leq \beta_{i-1} \Delta t_{i-1}, 0 \leq Z_{t_i}, \text{ for } i = 1, 2 \right\}
\]
and thus equality in (10). Using induction shows the first part of the assertion.

For the second part (11) we again consider the two stage case. Let $Z_{t_1}$ and $Z_{t_2}$ be feasible in (10) and define a new variable $Z_{t_2} := Z_{t_1} Z_{t_2}$, where, without loss of generality, $Z_{t_1} \prec \mathcal{F}_{t_1}$. The optimization problem (10) rewrites as

$$\begin{align*}
\text{maximize} \quad & E[Z_{t_2}] \\
\text{subject to} \quad & E[Z_{t_2} \mid \mathcal{F}_{t_1}] = Z_{t_1}, \ E[Z_{t_2}] = 1, \\
& E \left[ \frac{Z_{t_2}}{Z_{t_1}} \log \frac{Z_{t_2}}{Z_{t_1}} \mid \mathcal{F}_{t_1} \right] \leq \beta_1 \Delta t_1, \\
& E[Z_{t_1}] \log Z_{t_1} \leq \beta_0 \Delta t_0 \\
& Z_{t_1}, Z_{t_2} \geq 0.
\end{align*}$$

(13)

We can further rewrite the constraint (13) as

$$E \left[ Z_{t_2} \log Z_{t_2} \mid \mathcal{F}_{t_1} \right] - Z_{t_1} \log Z_{t_1} \leq \beta_1 \Delta t_1 Z_{t_1}.$$ 

This shows the assertion in a two stage setting. The general case follows again by considering the product $Z_{t_i} = Z_{t_0} \cdot Z_{t_1} \cdots Z_{t_i}$ together with the first part of the proof. □

3 Quantification of risk in continuous time

The previous section considers nested risk measures in discrete time on partitions $\mathcal{P} = \{0 = t_0 < \cdots < t_n = T\}$ of the interval $\mathcal{T} := [0, T]$. In what follows, we refine the partitions to obtain a limit of the nested risk measures in continuous time.

To this end, we first extend the vector of risk levels to continuous time by introducing a function $\beta: \mathcal{T} \rightarrow [0, \infty)$ called the risk rate and extend the definition of the nested Entropic Value-at-Risk to risk rates $\beta(\cdot)$.

**Definition 12.** For a given finite partition $\mathcal{P}$ and Riemann integrable risk rate $\beta(\cdot)$ the nested Entropic Value-at-Risk is defined as

$$nEV@R^P_{\beta(\cdot)}(Y) := \inf_{\tilde{\beta}} nEV@R^P_{\tilde{\beta}}(Y),$$

where the vector of risk levels is $\tilde{\beta} := (\beta(t_0) \Delta t_0, \ldots, \beta(t_{n-1}) \Delta t_{n-1})$.

Given a risk rate $\beta(\cdot)$ we now investigate the relationship of $nEV@R$ for different partitions. We will assume throughout that the risk rate $\beta(\cdot)$ is Riemann integrable.

**Theorem 13.** Let $Y \prec \mathcal{F}_T$ be a random variable and let $\beta: \mathcal{T} \rightarrow [0, \infty)$ be a piecewise constant risk rate. For a partition $\mathcal{P}_1$ of $\mathcal{T}$, fine enough to contain all points of discontinuity of $\beta$ and for every refinement $\mathcal{P}$ of $\mathcal{P}_1$, the inequality

$$nEV@R^P_{\beta(\cdot)}(Y) \leq nEV@R^P_{\beta(\cdot)}(Y)$$

holds true.

**Proof.** It is enough to consider the following partitions $\mathcal{P}_1 = \{0 = t_0 < t_2 = T\}$ and $\mathcal{P} = \{0 = t_0 < t_1 < t_2 = T\}$ of $[0, T]$ and the constant risk rate $\beta(\cdot)$. Consider the infimum representation

$$EV@R^{\beta(\cdot) \Delta t_1}(Y \mid \mathcal{F}_{t_1}) = \inf_{\ell} \frac{1}{\ell} \left( \beta(t_1) \cdot \Delta t_1 + \log E \left[ e^{\ell Y} \mid \mathcal{F}_{t_1} \right] \right).$$

7
of the conditional Entropic Value-at-Risk. By nesting we obtain
\[
\text{nEV@R}^{P}_{\beta(t)}(Y) = \inf_{\mathcal{P}} \frac{1}{x} \left( \beta(t_0) \cdot \Delta t_0 + \log E \left[ \exp \left( x \left( \inf_{\ell} \frac{1}{\ell} \left( \beta(t_1) \cdot \Delta t_1 + \log E \left[ e^{\ell Y} \, | \mathcal{F}_{t_1} \right] \right) \right) \right] \right) \cdot \mathcal{F}_{t_0}.
\]
Choosing \( \ell = x \) gives the upper bound
\[
\text{nEV@R}^{P}_{\beta(t)}(Y) \leq \frac{1}{x} \left( \beta(t_0) \cdot \Delta t_0 + \beta(t_1) \cdot \Delta t_1 + \log E \left( E \left[ e^{xY} \, | \mathcal{F}_{t_1} \right] \right) \right) = \text{EV@R}^{P}_{\beta(t_0) \cdot \Delta t_0 + \beta(t_1) \cdot \Delta t_1}(Y).
\]
Because \( \beta(t) \) is constant, it follows that
\[
\text{nEV@R}^{P}_{\beta(t)}(Y) \leq \text{EV@R}^{P}_{\beta(t_0)}(Y) = \text{nEV@R}^{P}_{\beta(t)}(Y).
\]
The general case follows by induction. \( \square \)

We are now able to extend the nested Entropic Value-at-Risk to continuous time and demonstrate that the extension is well-defined.

**Definition 14** (Nested Entropic Value-at-Risk in continuous time). Let \( T > 0 \), \( t \in [0, T] \) and \( Y \in E \). The nested Entropic Value-at-Risk in continuous time for the risk rate \( \beta : [t, T] \to [0, \infty) \) is
\[
\text{nEV@R}^{\beta(t)}_{\beta(u)}(Y | \mathcal{F}_t) := \inf_{\mathcal{P}, \beta(u) \geq \beta(t)} \text{nEV@R}^{P}_{\beta(u)}(Y | \mathcal{F}_t),
\]
where the infimum is among all partitions \( \mathcal{P} \subset [t, T] \) and piecewise constant functions \( \tilde{\beta}(\cdot) \geq \beta(\cdot) \).

**Remark 15.** Theorem 13 shows that the essential infimum in (14) can be replaced by the limit of a nonincreasing sequence. Furthermore, we have the lower bound
\[
\text{nEV@R}^{\beta(t)}_{\beta(u)}(Y | \mathcal{F}_t) \geq E [Y | \mathcal{F}_t]
\]
and hence the essential infimum is well defined and an element of \( E \).

The next proposition extends the dual representation of \( \text{nEV@R} \) to the continuous time setting (cf. Lemma 11).

**Proposition 16** (Duality of the nested EV@R). Let the risk rate \( \beta : \mathcal{T} \to [0, \infty) \) be Riemann integrable. For \( Y \in E \) the nested Entropic Value-at-Risk in continuous time has the representation
\[
\text{nEV@R}^{\beta(t)}_{\beta(u)}(Y | \mathcal{F}_t) = \sup \left\{ E [Y Z_T | \mathcal{F}_t] \left| \begin{array}{l}
E [Z_s \log Z_s | \mathcal{F}_u] \leq Z_u \int_s^T \beta(r) \, dr + Z_u \log Z_u, \\
E [Z_s | \mathcal{F}_u] = Z_u, Z_u = 1
\end{array} \right. \right\}.
\]

**Proof.** The proof follows the lines of the proof of Lemma 11. Let \( \mathcal{P} = \{ t = t_0 < \cdots < t_n = T \} \) be a partition of the interval \([t, T]\). Then the martingale condition in Lemma 11,
\[
E [Z_s | \mathcal{F}_{t_{u-1}}] = Z_{t_{u-1}},
\]
extends, for \( n \to \infty \), to the martingale condition
\[
E [Z_s | \mathcal{F}_{t_{u}}} = Z_{t_{u}} \quad (t \leq u \leq s \leq T)
\]
in continuous time. Furthermore, note that the entropic condition in (11) can be nested and thus

\[ E \left[ Z_t \log Z_t \mid \mathcal{F}_{t-1} \right] \leq E \left[ Z_{t,-1} \beta(t_{-1}) \cdot \Delta t_{-1} + Z_{t,-1} \log Z_{t,-1} \mid \mathcal{F}_{t-2} \right] \]

\[ = Z_{t,-2} \beta(t_{-1}) \cdot \Delta t_{-1} + \beta(t_{-2}) \cdot \Delta t_{-2} Z_{t,-2} + Z_{t,-2} \log Z_{t,-2} \]

We can now take the limit over all partitions such that the mesh size \( \|\mathcal{P}\| \) tends to zero, leading to the constraints

\[ E \left[ Z_s \log Z_s \mid \mathcal{T}_u \right] \leq Z_u \int_u^s \beta(r) \, dr + Z_u \log Z_u, \quad 0 \leq u \leq s \leq T. \]

The assertion now follows.

**Remark 17.** It is a consequence of Theorem 13 that

\[ E Y \leq n \text{EV@R}_{R^{0,T}}(Y) \leq \text{EV@R}_{\int_0^T \beta(s) \, ds}(Y), \]

which shows that the nested Entropic Value-at-Risk is a continuous function from \( E \) to \( \mathbb{R} \) in the norm topology of \( \text{EV@R} \) (cf. Ahmadi-Javid and Pichler (2017)).

### 4 Itô processes and nested risk measures

The preceding sections introduce nested risk measures in discrete time and subsequently extend the nested Entropic Value-at-Risk to continuous time using the monotonicity property of Theorem 13. Continuing the ideas of Proposition 7 we now consider nested risk measures on increments of a stochastic process in continuous time. A large class of such processes is given by Itô processes driven by Brownian motion. We now focus our attention on this important class.

As a motivating example, consider the Brownian motion, where Proposition 9 makes an explicit formula available.

**Proposition 18.** The nested Entropic Value-at-Risk of the Wiener process \( W \) on \( T = [0,T] \) for a risk rate \( \beta : T \to [0,\infty) \) is

\[ \text{nEV@R}_{R^{0,T}}(W_T) = \int_0^T \sqrt{2\beta(t)} \, dt. \]

**Proof.** Consider a partition \( \mathcal{P} \) of size \( n \) of \( T \) as in the previous section. By nesting it follows from (7) that

\[ \text{nEV@R}_{R^{P}}(W_T) = \sum_{i=0}^{n-1} \sqrt{\Delta t_i} \sqrt{2\beta(t_i)} \Delta t_i = \sum_{i=0}^{n-1} \Delta t_i \sqrt{2\beta(t_i)}. \]

Taking the limit \( n \to \infty \) shows that

\[ \text{nEV@R}_{R^{0,T}}(W_T) = \int_0^T \sqrt{2\beta(t)} \, dt, \]

the assertion.
In what follows, we demonstrate that the nested Entropic Value-at-Risk is well defined and finite in general cases of Itô processes. For important stochastic processes following a linear stochastic differential equation, we can give explicit formulas. To this end consider the general stochastic differential equation
\begin{equation}
\begin{aligned}
dX_t &= b(t, X_t) \, dt + \sigma(t, X_t) \, dW_t, \quad 0 < t \leq T, \\
X_0 &= x_0,
\end{aligned}
\end{equation}
where $b, \sigma : \mathcal{T} \times \mathbb{R} \to \mathbb{R}$ are measurable functions. The next lemma recalls conditions for the solution of the stochastic differential equation (16) to exist.

**Lemma 19** (cf. Øksendal (2003, Theorem 5.2.1)). *Let $t \in [0, T]$ and $x, y \in \mathbb{R}$ and suppose that $|b(t, x)| + |\sigma(t, x)| \leq C(1 + |x|)$ and $|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq C|x - y|$. Then the stochastic differential equation (16) has a unique solution for all initial values $x_0 \in \mathbb{R}$.*

Without loss of generality we assume throughout that the solution $X$ of (16) has continuous paths. In this case we may choose the canonical representation $\Omega = C(\mathcal{T})$ with $X(t, \omega) = \omega(t)$. The next theorem extends the explicit formula obtained in Proposition 18 to a large class of linear stochastic differential equations.

**Theorem 20.** *For $\mathcal{T} = [0, T]$ let $X = (X_t)_{t \in \mathcal{T}}$ be a linear diffusion process driven by the stochastic differential equation
\begin{equation}
\begin{aligned}
dX_t &= (A(t)X_t + a(t)) \, dt + \sigma(t) \, dW_t, \quad X_0 = x_0,
\end{aligned}
\end{equation}
where the functions $A, a, \sigma$ and $\beta : \mathcal{T} \to [0, \infty)$ are bounded. Then the nested Entropic Value-at-Risk is given explicitly by
\begin{equation}
nEV@R^{BT}_{\beta, t_\cdot} (X_T) = e^{\int_0^T A(s) \, ds} X_0 + \int_0^T a(u)e^{\int_u^T A(s) \, ds} \, du + \int_0^T e^{\int_u^T A(s) \, ds} \left| \sigma(u) \right| \sqrt{2\beta(u)} \, du.
\end{equation}
*Proof.* The solution of the linear stochastic differential equation (17) is given by (see, e.g., Karatzas and Shreve (1991, Section 5.6))
\begin{equation}
X_t = e^{\int_0^T A(s) \, ds} \left| X_0 + \int_0^T a(u)e^{-\int_u^T A(s) \, ds} \, du + \int_0^T \sigma(u)e^{-\int_u^T A(s) \, ds} \, dW_u \right|
\end{equation}
Set $\Phi(r, t) := \exp \left\{ \int_r^t A(s) \, ds \right\}$ and consider the partition $\mathcal{P} = \{ t_0, \ldots, t_n \}$ of $[0, T]$. By translation equivariance it follows that
\begin{equation}
\begin{aligned}
\text{EV@R}_{\beta(t_\cdot)} (X_{t_{i+1}} | \mathcal{F}_{t_i}) &= \Phi(t_i, t_{i+1}) X_{t_i} \\
&+ \int_{t_i}^{t_{i+1}} a(u) \Phi(u, t_{i+1}) \, du + \text{EV@R}_{\beta(t_\cdot)} \left( \int_{t_i}^{t_{i+1}} \Phi(u, t_{i+1}) \sigma(u) \, dW_u \bigg| \mathcal{F}_{t_i} \right).
\end{aligned}
\end{equation}
The random variable
\begin{equation}
\int_{t_i}^{t_{i+1}} \Phi(u, t_{i+1}) \sigma(u) \, dW_u
\end{equation}
is Gaussian with mean zero and variance
\[
\int_{t_i}^{t_{i+1}} \Phi^2(u, t_{i+1}) \sigma^2(u) \, du.
\] (19)

From Proposition 4 we conclude that
\[
\text{EV}_\beta \left( X_{t_{i+1}} | \mathcal{F}_{t_i} \right) = \Phi(t_i, t_{i+1}) X_0 + \int_{t_i}^{t_{i+1}} a(u) \Phi(u, t_{i+1}) \, du + \left( \int_{t_i}^{t_{i+1}} \Phi^2(u, t_{i+1}) \sigma^2(u) \, du \cdot 2\beta(t_i) \Delta t_i \right)^{\frac{1}{2}}.
\]

Repeating the same steps gives
\[
\text{EV}_\beta | \mathcal{F}_{t_{i-1}} \left( \text{EV}_\beta \left( X_{t_{i+1}} | \mathcal{F}_{t_i} \right) \right) = \Phi(t_{i-1}, t_{i+1}) X_{t_{i-1}} + \int_{t_{i-1}}^{t_{i+1}} a(u) \Phi(u, t_{i+1}) \, du + \left( \int_{t_{i-1}}^{t_{i+1}} \Phi^2(u, t_{i+1}) \sigma^2(u) \, du \cdot 2\beta(t_i) \Delta t_i \right)^{\frac{1}{2}}
\]
\[
+ \Phi(t_i, t_{i+1}) \cdot \left( \int_{t_{i-1}}^{t_i} \Phi^2(u, t_i) \sigma^2(u) \, du \cdot 2\beta(t_{i-1}) \Delta t_{i-1} \right)^{\frac{1}{2}}.
\]

Iterating this argument and nesting with respect to \( n \) stages we obtain the explicit formula
\[
\text{nEV}_\beta | \mathcal{F}_{t_{i-1}} (X_T) = \Phi(0, T) X_0 + \int_0^T a(u) \Phi(u, T) \, du + \sum_{i=0}^{n-1} e^{\int_{t_i}^{t_{i+1}} A(s) \, ds} \left( \int_{t_i}^{t_{i+1}} \Phi^2(u, t_{i+1}) \sigma^2(u) \, du \cdot 2\beta(t_i) \Delta t_i \right)^{\frac{1}{2}}
\]
\[
= \Phi(0, T) X_0 + \int_0^T a(u) \Phi(u, T) \, du + \sum_{i=0}^{n-1} \left( \int_{t_i}^{t_{i+1}} e^{2 \int_{t_u}^{t_{i+1}} A(s) \, ds} \sigma^2(u) \, du \cdot 2\beta(t_i) \Delta t_i \right)^{\frac{1}{2}}
\]
for the nested Entropic Value-at-Risk of a linear diffusion. Taking the limit over all partitions as in (14) and using the linear approximation
\[
\int_{t_i}^{t_{i+1}} e^{2 \int_{t_u}^{t_{i+1}} A(s) \, ds} \sigma^2(u) \, du = e^{2 \int_{t_i}^{t_{i+1}} A(s) \, ds} \sigma^2(t_i) \Delta t_i + o(\Delta t_i)
\]
we first have
\[
\lim_{n \to \infty} \sum_{i=0}^{n-1} \left( \int_{t_i}^{t_{i+1}} e^{2 \int_{t_u}^{t_{i+1}} A(s) \, ds} \sigma^2(u) \, du \cdot 2\beta(t_i) \Delta t_i \right)^{\frac{1}{2}} = \lim_{n \to \infty} \sum_{i=0}^{n-1} e^{\int_{t_u}^{t_{i+1}} A(s) \, ds} |\sigma(t_i)| \Delta t_i \cdot \sqrt{2\beta(t_i)}.
\]

In the limit we obtain
\[
\text{nEV}_\beta | \mathcal{F}_{t_{i-1}} (X_T) = e^{\int_0^T A(s) \, ds} X_0 + \int_0^T a(u) e^{\int_u^T A(s) \, ds} \, du + \int_0^T e^{\int_u^T A(s) \, ds} |\sigma(u)| \sqrt{2\beta(u)} \, du
\]
and thus the assertion.

The Ornstein–Uhlenbeck process is a well-known example of a process satisfying the assumptions of Theorem 20 above. For this process the nested Entropic Value-at-Risk simplifies further.
Example 21 (Ornstein–Uhlenbeck). Consider the Ornstein–Uhlenbeck process $X_t$ following the stochastic differential equation with constant coefficients
\[
d X_t = \theta(\mu - X_t) \, dt + \sigma \, dW_t, \quad X_0 = x_0.
\] (20)
The closed form solution of (20) is
\[
X_t = e^{-\theta t} x_0 + \mu (1 - e^{-\theta t}) + \int_0^t \sigma e^{-\theta(t-s)} \, dW_s.
\]

From Theorem 20 we obtain the explicit formula by setting $A(t) = -\theta$, $a(t) = \theta \mu$ and $\sigma(t) = \sigma$ in (17) and thus
\[
nEV@R^0_{\beta}(X_T) = e^{-\theta T} x_0 + \mu (1 - e^{-\theta T}) + \int_0^T e^{-\theta(T-t)} \sigma \sqrt{2\beta(t)} \, dt.
\]
In case $\beta(\cdot) = \beta_0$ the integral evaluates further so that
\[
nEV@R^0_{\beta}(X_T) = e^{-\theta T} x_0 + \mu (1 - e^{-\theta T}) + \frac{\sigma \sqrt{2\beta_0}}{\theta} \left(1 - e^{-\theta T}\right).
\]

Remark 22. It is essential for the proof of Theorem 20 that the diffusion coefficient $\sigma(\cdot)$ is independent of the state variable $x$. Otherwise, the stochastic integral in (18) is not Gaussian. However, the nested Entropic Value-at-Risk is well-defined for Itô processes where $\sigma(\cdot)$ depends on the state. Due to the strong integrability conditions for random variables in $E$ further assumptions besides those in Lemma 19 have to be imposed as the next example illustrates.

Example 23. Let $(X_t)_t$ be the Wald martingale
\[
X_t = \exp \left\{ -\frac{\sigma^2}{2} t + \sigma W_t \right\},
\]
then $X_t$ is log normally distributed for which the moment generating function
\[
m_Y(\ell) := E e^{\ell X_t}
\]
is not defined for $\ell > 0$ and $EV@R_{\beta}(X_t) = \infty$.

We conclude that for a general Itô process $(X_t)_t$ the diffusion coefficient $\sigma$ must not be unbounded, in general.

4.1 The risk generator

A fundamental tool in the classical theory of risk-neutral stochastic optimal control is the infinitesimal generator, a differential operator describing the evolution of the system. This subsection introduces the generator in the presence of risk, which extends the notion of the infinitesimal generator of Markov processes by replacing the expectation by a risk measure. This enables us to formulate and solve risk-averse control problems.

Throughout this section we consider the interval $T = [t, T]$ and the Itô process $(X_s)_{s \in T}$ given by the stochastic differential equation
\[
d X_s = b(s, X_s) \, ds + \sigma(s, X_s) \, dW_s, \quad s \in T,
\] (21)
\[
X_t = x.
\]

For such processes we can define the risk generator.
Definition 24 (Risk generator). Let $(X_t)_{t \in T}$ be the solution of (21) on $T$ with initial condition $X_t = x$. For the risk rate $\beta : T \to \mathbb{R}$, the risk generator based on the Entropic Value-at-Risk is

$$R_\beta \Phi(t, x) := \lim_{h \downarrow 0} \frac{1}{h} \left( \operatorname{EV@R}_{\beta(t)h} \Phi(t + h, X_{t+h}) | X_t = x \right) - \Phi(t, x),$$

(22)

for those functions $\Phi : T \times \mathbb{R} \to \mathbb{R}$ for which the limit exists.

For normally distributed random variables Proposition 4 gives explicit representations for EV@R and we may calculate the risk generator in the case where $X_t$ is given by (21). For the risk generator based on the Entropic Value-at-Risk we impose the following time regularity on the diffusion coefficient, which we assume from now on.

Assumption 25 (Hölder continuity). There exists a $\tilde{C} > 0$ and an $\alpha \in (0, 1/2)$ such that

$$|\sigma(u, y) - \sigma(s, x)| \leq \tilde{C}|u - s|^\alpha, \quad s, u \in T,$$

uniformly for all $x, y \in \mathbb{R}$.

Proposition 26 (Risk generator). Let $(X_t)_{t \in T}$ be the solution of (21) on $T$ with initial condition $X_t = x$. For $\Phi \in C^{1,2}(T \times \mathbb{R})$ such that $\frac{\partial \Phi}{\partial \alpha}$ is bounded, the risk generator based on EV@R satisfies

$$R_\beta \Phi(t, x) = \frac{\partial \Phi}{\partial t}(t, x) + b(t, x) \frac{\partial \Phi}{\partial x}(t, x) + \frac{\sigma^2(t, x)}{2} \frac{\partial^2 \Phi}{\partial x^2}(t, x) + \sqrt{2\beta(t)} \left( \frac{\partial \Phi}{\partial \alpha}(t, x) \right).$$

(23)

Remark 27. The risk generator $R_\beta$ can be decomposed as the sum of the classical generator and the nonlinear term $\sqrt{2\beta} \cdot |\sigma \frac{\partial \Phi}{\partial \alpha}|$. This additional risk term can be interpreted as a directed drift term where the uncertain drift $\frac{\partial \Phi}{\partial \alpha}(t, X_t)$ is scaled with volatility $\sigma$. The coefficient $\sqrt{2\beta(t)}$ expresses risk aversion. For $\beta = 0$ we obtain the classical risk-neutral infinitesimal generator. Furthermore, if $\sigma = 0$, i.e., no randomness occurs in the model, the generator reduces to a first order differential operator describing the dynamics of a deterministic system, where risk does not apply.

Proof. By assumption $\Phi \in C^{1,2}(T \times \mathbb{R})$ and hence we may apply Itô’s formula

$$\Phi(t + h, X_{t+h}) - \Phi(t, X_t) = \int_t^{t+h} \left( \frac{\partial \Phi}{\partial t} + b \frac{\partial \Phi}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 \Phi}{\partial x^2} \right)(s, X_s)ds + \int_t^{t+h} \left( \frac{\partial \Phi}{\partial \alpha} \right)(s, X_s)dW_s.$$

For convenience we set $f_1(t, x) := \left( \frac{\partial \Phi}{\partial t} + b \frac{\partial \Phi}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 \Phi}{\partial x^2} \right)(t, x)$ and $f_2(t, x) := \left( \frac{\partial \Phi}{\partial \alpha} \right)(t, x)$. In this setting $R_\beta$ rewrites as

$$R_\beta \Phi(t, x) = \lim_{h \downarrow 0} \frac{1}{h} \left[ \operatorname{EV@R}_{\beta(t)h} f_1(s, X_s)ds + \int_t^{t+h} f_2(s, X_s)dW_s \right] | X_t = x.$$

We need to show (23) for each fixed $(t, x)$, i.e., the inequality

$$\left| R_\beta \Phi(t, x) - f_1(t, x) - \sqrt{2\beta(t)} f_2(t, x) \right| \leq 0.$$
Using convexity of $\mathbb{E}@R$ and the triangle inequality we have

\[
0 \leq \lim_{h \downarrow 0} \mathbb{E}@R_{\beta(t) \cdot h} \left[ \frac{1}{h} \int_t^{t+h} f_1(s, X_s) ds - f_1(t, x) \right] X_t = x \bigg] + \\
+ \lim_{h \downarrow 0} \mathbb{E}@R_{\beta(t) \cdot h} \left[ \frac{1}{h} \int_t^{t+h} f_2(s, X_s) dW_s - \sqrt{2\beta(t)} f_2(t, x) \right] X_t = x \bigg] .
\]

We continue by looking at each term separately. Note that $s \mapsto f_1(s, X_s) - f_1(t, x)$ is continuous almost surely and hence the mean value theorem for definite integrals implies that there exists a $\xi \in [t, t+h]$ such that

\[
\frac{1}{h} \int_t^{t+h} f_1(s, X_s) ds - f_1(t, x) = f_1(\xi, X_\xi) - f_1(t, x), \quad \text{almost surely.}
\]

From convexity of $\mathbb{E}@R$ we may conclude

\[
\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}@R_{\beta(t) \cdot h} \left( \int_t^{t+h} |f_1(s, X_s) - f_1(t, x)| ds \right) X_t = x = 0.
\]

Note that the stochastic integral term in (24) can be bounded by

\[
\mathbb{E}@R_{\beta(t) \cdot h} \left[ \frac{1}{h} \int_t^{t+h} f_2(s, X_s) dW_s \bigg| X_t = x \right] \leq \mathbb{E}@R_{\beta(t) \cdot h} \left[ \frac{1}{h} \int_t^{t+h} f_2(s, X_s) - f_2(t, x) dW_s \bigg| X_t = x \right] \\
+ \mathbb{E}@R_{\beta(t) \cdot h} \left[ \frac{1}{h} \int_t^{t+h} f_2(t, x) dW_s \bigg| X_t = x \right],
\]

where $\mathbb{E}@R_{\beta(t) \cdot h} \left[ \frac{1}{h} \int_t^{t+h} f_2(s, X_s) - f_2(t, x) dW_s \bigg| X_t = x \right] = \sqrt{2\beta(t)} |f_2(t, x)|$ and hence

\[
(24) \leq \lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}@R_{\beta(t) \cdot h} \left[ \frac{1}{h} \int_t^{t+h} f_2(s, X_s) - f_2(t, x) dW_s \bigg| X_t = x \right].
\]

Furthermore, the stochastic integral $M_h := \int_t^{t+h} f_2(s, X_s) - f_2(t, x) dW_s$ is a continuous martingale with bounded quadratic variation,

\[
\langle M \rangle_h \leq \frac{C^2 h^{1+2\alpha}}{2\alpha + 1}.
\]

Recall the infimum representation of $\mathbb{E}@R$, then

\[
\frac{1}{h} \mathbb{E}@R_{\beta(t) \cdot h} (M_h | X_t = x) = \inf_{0 < h \epsilon} \frac{1}{\epsilon} (\beta(t) \cdot h + \log [\mathbb{E} \exp (\ell M_h) | X_t = x]),
\]

but $M_h$ satisfies Novikov’s condition and thus $1 = \mathbb{E} \exp \left( \ell M_h - \ell^2 \langle M \rangle_h \right)$ holds. Together with (25), this shows that

\[
\mathbb{E} e^{\ell M_h} \leq \exp \left( \frac{\ell^2}{2} \cdot \frac{C^2 h^{1+2\alpha}}{2\alpha + 1} \right).
\]

It follows similarly to Proposition 4 that

\[
\frac{1}{h} \mathbb{E}@R_{\beta(t) \cdot h} (M_h | X_t = x) \leq \inf_{\ell > 0} \frac{1}{\ell h} \left( \beta(t) \cdot h + \frac{\ell^2}{2} \cdot \frac{C^2 h^{1+2\alpha}}{2\alpha + 1} \right) \leq \inf_{\ell > 0} \frac{\beta(t) \cdot h}{\ell} + \frac{\ell}{2} \cdot \frac{C^2 h^{2\alpha}}{2\alpha + 1} \leq \frac{\sqrt{2\beta(t)} \cdot C h^\alpha}{\sqrt{2\alpha + 1}}.
\]
We will show that (28) equals zero. For ease of notation, we now omit the arguments whenever there is no ambiguity. Let \( \Phi \in C^{1,2}(\mathcal{T} \times \mathbb{R}) \) such that \( \frac{\partial \Phi}{\partial x} \) is bounded, the risk-averse Dynkin formula

\[
\text{nEV@R}^{t,r} (\Phi(r, X_r)) = \Phi(t, x) + \text{nEV@R}^{t,r} \left( \int_t^r \mathcal{R}_\beta(s, X_s) \, ds \right)
\]

holds for any \( t \leq r \leq T \).

**Proof.** The left side of (26) rewrites as

\[
\Phi(t, x) + \text{nEV@R}^{t,r} \left( \int_t^r \left( \frac{\partial \Phi}{\partial t} + b \frac{\partial \Phi}{\partial x} + \sigma^2 \frac{\partial^2 \Phi}{\partial x^2} \right) (s, X_s) \, ds + \int_t^r \left( \sigma \frac{\partial \Phi}{\partial x} \right) (s, X_s) \, dW_s \right)
\]

using Itô’s formula and hence the representation of the risk generator in Proposition 26 shows that

\[
\left| \text{nEV@R}^{t,r} (\Phi(r, X_r)) - \Phi(t, x) - \text{nEV@R}^{t,r} \left( \int_t^r \mathcal{R}_\beta(s, X_s) \, ds \right) \right|
\]

can be bounded by

\[
n\text{EV@R}^{t,r} \left( \left| \int_t^r \sigma(s, X_s) \frac{\partial \Phi}{\partial x} (s, X_s) \, dW_s - \int_t^r \sqrt{2}\beta \left| \sigma(s, X_s) \frac{\partial \Phi}{\partial x} (s, X_s) \right| \, ds \right| \right).
\]

We will show that (28) equals zero. For ease of notation, we now omit the arguments whenever there is no ambiguity. Let \( n \in \mathbb{N} \) and consider the partition \( \mathcal{P}_n \) of \([t, r]\). It follows from convexity that

\[
(28) \leq \sum_{i=1}^n \text{nEV@R}^{t,i} \left( \left| \int_{t_i}^{t_{i+1}} \sigma(t, X_t) \frac{\partial \Phi}{\partial x} (t, X_t) \, dW_t - \int_{t_i}^{t_{i+1}} \sqrt{2}\beta \left| \sigma(t, X_t) \frac{\partial \Phi}{\partial x} (t, X_t) \right| \, dt \right| \right).
\]

Moreover, monotonicity of \( \text{nEV@R} \) as well as Theorem 13 show that the summands of (29) are bounded by

\[
n\text{EV@R}^{t,i} \left( \text{EV@R}^{t,i} \text{d} \left| \int_{t_i}^{t_{i+1}} \sigma(t, X_t) \frac{\partial \Phi}{\partial x} (t, X_t) \, dW_t - h\sqrt{2}\beta \right| \sigma(t_i, X_{t_i}) \frac{\partial \Phi}{\partial x} (t_i, X_{t_i}) + o(\Delta t_i) \, \mathcal{F}_{t_i} \right),
\]

\[15\]
as $\int_{t_i}^{t_{i+1}} \sqrt{2\beta} |\sigma \frac{\partial \Phi}{\partial s}| \, ds$ is bounded by Assumption 25. We demonstrate that the inner conditional risk measure converges to zero fast enough. To this end, we argue similarly to the proof of Proposition 26 and split the stochastic integral in two parts as

$$
\int_{t_i}^{t_{i+1}} \sigma(s) \frac{\partial \Phi}{\partial x}(s, X_s) \, dW_s = \int_{t_i}^{t_{i+1}} \sigma(t_i) \frac{\partial \Phi}{\partial x}(t_i, X_{t_i}) \, dW_s + \int_{t_i}^{t_{i+1}} \sigma(s) \frac{\partial \Phi}{\partial x}(s, X_s) - \sigma(t_i) \frac{\partial \Phi}{\partial x}(t_i, X_{t_i}) \, dW_s.
$$

Thus (30) can be bounded by

$$
n_{\text{EVAR}}^t \left( \int_{\mathcal{F}_i} \left[ \int_{t_i}^{t_{i+1}} \sigma(t_i) \frac{\partial \Phi}{\partial x}(t_i, X_{t_i}) \, dW_s - h \sqrt{2\beta t_i} \left| \sigma(t_i, X_{t_i}) \frac{\partial \Phi}{\partial x}(t_i, X_{t_i}) \right| \, \mathbb{1} \left( \mathcal{F}_i \right) \right] + n_{\text{EVAR}}^t \left[ \int_{t_i}^{t_{i+1}} \sigma(s) \frac{\partial \Phi}{\partial x}(s, X_s) - \sigma(t_i) \frac{\partial \Phi}{\partial x}(t_i, X_{t_i}) \, dW_s + o(\Delta t_i) \, \mathbb{1} \left( \mathcal{F}_i \right) \right] \right) \leq n_{\text{EVAR}}^t (0 + o(\Delta t_i) \, \mathbb{1} \left( \mathcal{F}_i \right))
$$

The first part is equal to zero and the argument in the proof of Proposition 26 shows that the second part tends to zero faster than linearly. In conclusion we obtain

$$
n_{\text{EVAR}}^n \left( \int_{t_i}^{t_{i+1}} \sigma(s) \frac{\partial \Phi}{\partial x}(s, X_s) \, dW_s - \int_{t_i}^{t_{i+1}} \sqrt{2\beta} \left| \sigma \frac{\partial \Phi}{\partial x} \right| \, ds \, \mathbb{1} \left( \mathcal{F}_i \right) \right) \leq n_{\text{EVAR}}^n (0 + o(\Delta t_i) \, \mathbb{1} \left( \mathcal{F}_i \right))
$$

and hence for $n \to \infty$

$$
\sum_{\mathcal{F}_i \in \mathcal{P}_n} n_{\text{EVAR}}^n \left( \int_{t_i}^{t_{i+1}} \sigma(s) \frac{\partial \Phi}{\partial x}(s, X_s) \, dW_s - \int_{t_i}^{t_{i+1}} \sqrt{2\beta} \left| \sigma \frac{\partial \Phi}{\partial x} \right| \, ds \, \mathbb{1} \left( \mathcal{F}_i \right) \right) \to 0,
$$

which concludes the argument. $\square$

### 4.2 The nested version of the Average Value-at-Risk

The previous sections focus on developing a dynamic nested version of Entropic Value-at-Risk and a risk-averse analogue of the generator is introduced. The situation for the Average Value-at-Risk, the most prominent coherent risk measure, is considerably more involved. We demonstrate that the nested Average Value-at-Risk degenerates either to the expectation or the essential supremum provided that the limit over all partitions in (5) is taken without properly rescaling the risk levels. Furthermore, we provide the proper rescaling such that the nested Average Value-at-Risk does not degenerate in the continuous time setting.

The Average Value-at-Risk at risk level $\alpha$ is given by (cf. Rockafellar and Uryasev (2002))

$$\text{AVAR} = \min_{x \in \mathbb{R}} x + \frac{1}{1-\alpha} \mathbb{E} \left( Y - x \right)_+.$$

For normally distributed $Y \sim N(\mu, \sigma^2)$ we have the explicit formula $\text{AVAR} = \mu + \sigma \varphi^{-1}(\alpha)$, where $\varphi$ and $\Phi$ are the density and cumulative distribution function of the standard normal distribution, respectively. The conditional Average Value-at-Risk is given by (cf. Xin and Shapiro (2011))

$$\text{AVAR} (Y \mid \mathcal{F}_x) := \min_{x \in \mathbb{R}} x + \frac{1}{1-\alpha} \mathbb{E} \left[ (Y - x)_+ \mid \mathcal{F}_x \right].$$

Similarly to the Entropic Value-at-Risk, we now consider a risk rate $\alpha : \mathcal{T} \to [0, 1]$ and introduce the nested Average Value-at-Risk for a vector of risk levels $(\alpha(t_0), \ldots, \alpha(t_{n-1}))$ on a partition $\mathcal{P}$ by

$$\text{nAVAR}^t (Y \mid \mathcal{F}_t) := \text{AVAR} (\alpha(t_0) \ldots \text{AVAR} (\alpha(t_{n-1}) (Y \mid \mathcal{F}_{t_{n-1}}) \ldots \mathcal{F}_t).$$
Then the nested Average Value-at-Risk for the Brownian motion is given by

$$nAV\n\varpi^P_n(W_T) = \sum_{i=0}^{n-1} \sqrt{\Delta t_i} \frac{\varphi\left(\Phi^{-1}\left(\alpha(t_i)\right)\right)}{1 - \alpha(t_i)}.$$

The following example show that natural choices for risk rates $\alpha(\cdot)$ lead to a degenerate risk evaluation.

**Example 29.** Let us assume an equidistant partition of $[0, T]$ of size $n$ with step size $h := \frac{T}{n}$ as well as constant risk levels $\alpha(t_i) = \alpha_0 \in (0, 1)$. Then the limit of the right side of (32) above is

$$\lim_{n \to \infty} \sum_{i=0}^{n-1} \sqrt{\Delta t_i} \frac{\varphi\left(\Phi^{-1}\left(\alpha(t_i)\right)\right)}{1 - \alpha(t_i)} = \lim_{n \to \infty} n \sqrt{\frac{T}{n}} \frac{\varphi\left(\Phi^{-1}\left(\alpha_0\right)\right)}{1 - \alpha_0} = \infty.$$ 

On the other hand the analysis in Xin and Shapiro (2011) suggests to choose the risk rate $\alpha(t_i) := h \in (0, 1)$. Then

$$\sum_{i=0}^{n-1} \sqrt{\Delta t_i} \frac{\varphi\left(\Phi^{-1}\left(\alpha(t_i)\right)\right)}{1 - \alpha(t_i)} = \frac{T}{\sqrt{h}(1 - \frac{1}{n} h)} \cdot \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\Phi^{-1}\left(\frac{1}{T} h\right)\right)^2 \right\},$$

but for $p$ close to zero the asymptotic relation $\Phi^{-1}(p) \sim -\sqrt{-2 \log p}$ holds (see the stable reference http://dlmf.nist.gov/7.17.iii) and thus

$$\frac{T}{\sqrt{h}(1 - \frac{1}{n} h)} \cdot \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\Phi^{-1}\left(\frac{1}{T} h\right)\right)^2 \right\} = \frac{c\sqrt{h}}{(1 - \frac{1}{n} h)\sqrt{2\pi}}.$$

which tends to zero as $h \downarrow 0$.

We now derive the correct asymptotic behavior of $\alpha$, such that the nested Average-Value-at-Risk does not degenerate. To this end, we construct a modified risk rate $\alpha^\alpha(t, h)$ depending on time $t$ as well as the step size $h$.

**Theorem 30.** Consider a Brownian motion $W = (W_t)_{t \in [0, T]}$ on the interval $[0, T]$ and for a risk rate $\alpha : [0, T] \to [0, \infty)$ we set

$$\alpha^\alpha(t, h) := \Phi\left(-\sqrt{-\log(2\pi h \cdot \alpha(t))}\right).$$

Then the nested Average Value-at-Risk for the Brownian motion is given by

$$nAV\n\varpi^{\alpha^\alpha}_{\alpha^\alpha}(W_T) := \lim_{n \to \infty} \sum_{i=0}^{n-1} \sqrt{\Delta t_i} \frac{\varphi\left(\Phi^{-1}\left(\alpha^\alpha(t_i, \Delta t_i)\right)\right)}{1 - \alpha^\alpha(t_i, \Delta t_i)} = \int_0^T \sqrt{\alpha(s)} \, ds,$$

where the limit is taken over all partitions $\mathcal{P}$ with mesh size $\|\mathcal{P}\| \to 0$.

*The asymptotic expansion

$$\alpha^\alpha(t, h) \sim \sqrt{h} \left( \frac{1}{\sqrt{-\log(2\pi h \alpha(t))}} - \frac{1}{\sqrt{-\log(2\pi h \alpha(t))}} + \frac{3}{\sqrt{-\log(2\pi h \alpha(t))}} \cdots \right)$$

holds for $h \to 0$. 

17
Proof. Let $\varphi$ be the density of the standard normal distribution, and $\Phi$ be its cumulative distribution function. For the partition $\mathcal{P}$ we set $\Delta t = t_{i+1} - t_i$ and consider $A^\alpha(t,h)$ as in the theorem. Then we have

$$\varphi \left( \Phi^{-1}(A^\alpha(t_i, \Delta t)) \right) = \sqrt{\Delta t} \frac{\alpha(t_i)}{\varphi(\alpha(t_i))}.$$ 

Moreover, in the limit we obtain

$$\text{nAV@R}_{A^\alpha(\cdot)}^{0,T} \left( \int_0^T dW_s \right) = \lim_{\|P\| \to 0} \sum_i \sqrt{\Delta t} \frac{\varphi \left( \Phi^{-1}(A^\alpha(t_i, \Delta t)) \right)}{1 - A^\alpha(t_i, \Delta t)} = \int_0^T \sqrt{\alpha(s)} \, ds,$$

which is independent of the choice of the partition. $\square$

The following proposition shows that the risk generator with respect to the modified Average Value-at-Risk with risk rate $A^\alpha(\cdot)$ is equal to the risk generator with respect to the Entropic Value-at-Risk up to a scaling factor. The proof is analogous to the proof of Proposition 26.

**Proposition 31** (Risk generator for AV@R). Let $(X_s)_{s \in T}$ be the solution of (21) on $T = [t,T]$ with initial condition $X_t = x$ satisfying Assumption 25. For $\Phi \in \mathcal{C}^{1,2}(T \times \mathbb{R})$ such that $\frac{\partial \Phi}{\partial x}$ is bounded, the risk generator based on AV@R satisfies

$$\mathcal{R}\Phi(t,x) = \frac{\partial \Phi}{\partial t}(t,x) + \frac{\partial \Phi}{\partial x}(t,x)b(t,x) + \frac{1}{2} \frac{\partial^2 \Phi}{\partial x^2}(t,x)\sigma^2(t,x) \left[ \sigma(t,x) \frac{\partial \Phi}{\partial x}(t,x) \right].$$

**Remark 32.** From the viewpoint of dynamic programming equations it is not important whether the Entropic Value-at-Risk or the Average Value-at-Risk is considered as the risk generators are essentially the same. However, the adapted risk levels for the Entropic Value-at-Risk can be interpreted intuitively as a fixed risk level scaled with the length of the time horizon, whereas the modified risk levels $A^\alpha$ for the Average Value-at-Risk do not allow for such an interpretation and are thus unnatural. For this reason, we consider the nested Entropic Value-at-Risk the natural choice for a coherent risk measure in continuous time.

## 5 The risk-averse control problem

The preceding section develops a risk-averse extension of the infinitesimal generator. Moreover, a risk-averse Dynkin formula is shown. Using these results, we now formulate a risk-averse optimal control problem and derive associated Hamilton–Jacobi–Bellman equations.

Consider the set of controls

$$\mathcal{U}[0,T] := \{ u : T \times \Omega \to U \mid u \text{ is adapted} \}$$

and $U \subset \mathbb{R}$. For any initial condition $(t,x) \in [0,T) \times \mathbb{R}$ and control $u \in \mathcal{U}[t,T]$ we consider the controlled stochastic process $(X^{t,x,u}_s)_{s \in [t,T]}$ given by (compare to (21))

$$dX^{t,x,u}_s = b(s,X^{t,x,u}_s,u(s)) \, ds + \sigma(s,X^{t,x,u}_s,u(s)) \, dW_s, \quad s \in [t,T],$$

$$X^{t,x,u}_t = x.$$

The aim is to evaluate the risk of the accumulated cost over time, therefore we consider a cost rate $c : [0,T] \times \mathbb{R} \times U \to \mathbb{R}$ and a terminal cost $\Psi : U \to \mathbb{R}$ so that the total cost accumulated over $[t,T]$ is

$$\int_t^T c(s,X^{t,x,u}_s,u(s)) \, ds + \Psi(X^{t,x,u}_T).$$

18
For $u \in \mathcal{U}[t,T]$ and adapted $\beta: [0,T] \rightarrow [0,\infty)$ we define the controlled value function $V^u$ by

$$V^u(t,x) := \inf_{u \in \mathcal{U}[t,T]} \text{EV} @ \mathbb{P}^{t,T}_{\beta(\cdot)} \left( \int_t^T c(s,X_s^{t,x,u},u(s)) \, ds + \Psi(X_s^{t,x,u}) \bigg| X_t = x \right).$$  \hspace{1cm} (34)

For arbitrary $u \in \mathcal{U}[0,T]$ the controlled value function $V^u$ may not exist. We follow Fleming and Soner (2006, p. 141) and introduce the set of admissible controls.

**Definition 33** (Admissible control). $\mathcal{U}[t,T]$ is called an admissible control system if it satisfies the following conditions.

(i) For $u \in \mathcal{U}[t,T]$, the function $u: [t,T] \times \Omega \rightarrow U$ is an adapted process with respect to the Brownian filtration.

(ii) For any initial value $x \in \mathbb{R}$ and $u \in \mathcal{U}[t,T]$ the stochastic differential equation (33) admits a unique solution and $V^u(t,x)$ is well defined.

From now on $\mathcal{U}[t,T]$ always denotes an admissible control system and we define the optimal value function $V: [0,T] \times \mathbb{R} \rightarrow \mathbb{R}$ as

$$V(t,x) := \inf_{u \in \mathcal{U}[t,T]} V^u(t,x) = \inf_{u \in \mathcal{U}[t,T]} \text{EV} @ \mathbb{P}^{t,T}_{\beta(\cdot)} \left( \int_t^T c(s,X_s^{t,x,u},u(s)) \, ds + \Psi(X_s^{t,x,u}) \bigg| X_t = x \right). \hspace{1cm} (35)$$

The risk-averse control problem can now be formulated as:

given $(t,x) \in [0,T] \times \mathbb{R}$, find an admissible control $u^* \in \mathcal{U}[t,T]$ such that

$$V^{u^*}(t,x) = \inf_{u \in \mathcal{U}[t,T]} V^u(t,x). \hspace{1cm} (36)$$

The following proposition guarantees that condition (ii) in Definition 33 is satisfied. It is an extension of Lemma 19 as well as Lemma 26.

**Proposition 34.** Let $s \in [t,T]$ and $x_1, x_2 \in \mathbb{R}$ and $u_1, u_2 \in U$. Suppose there exists a constant $C > 0$ such that

$$|b(s,x_1,u_1)| + |\sigma(s,x_1,u_1)| + |c(s,x_1,u_1)| + |\Psi(x_1)| \leq C(1 + |x_1| + |u_1|)$$

and

$$|b(s,x_1,u_1) - b(s,x_2,u_2)| + |\sigma(s,x_1,u_1) - \sigma(s,x_2,u_2)| + |c(s,x_1,u_1) - c(s,x_2,u_2)|$$

$$\leq C \left( |x_1 - x_2| + |u_1 - u_2| \right)$$

hold. Then the stochastic differential equation (33) has a unique solution. Moreover, if the diffusion coefficient $\sigma(\cdot,u(\cdot))$ satisfies Assumption 25, the controlled value function $V^u(t,x)$ is well defined and deterministic.

### 5.1 Principle of dynamic programming

We show that the risk-averse optimal value function $V(\cdot, \cdot)$ defined in (35) satisfies an analogue of the dynamic programming principle. Furthermore, we introduce the risk-averse Hamilton–Jacobi–Bellman equations and show that the optimal value function solves these equations in the sense of viscosity solutions. Additionally, we provide a verification theorem, showing that a classical solution to the risk-averse Hamilton–Jacobi–Bellman equation is the optimal value function (35).
Lemma 35 (Dynamic programming principle). Let \((t, x) \in [0, T] \times \mathbb{R}\) and \(r \in [t, T]\) and suppose that \(\mathcal{U}[t, T]\) is an admissible control system, then it holds that

\[
V(t, x) = \inf_{u \in \mathcal{U}[t, r)} \text{nEV@R}_\beta^T \left( \int_t^r c(s, X_s^{t,x,u}, u_s) \, ds + V(r, X_r^{t,x,u}) \middle| X_t = x \right). \tag{37}
\]

**Proof.** For every \(\varepsilon > 0\) there exists a \(\bar{u}(\cdot) \in \mathcal{U}[t, T]\) such that \(V(t, x) + \varepsilon \geq V(t, x)\). Using the recursive property of the nested risk measures we obtain

\[
V(t, x) + \varepsilon \geq n\text{EV@R}_\beta^T \left( \int_t^r c(s, X_s^{t,x,\bar{u}}, \bar{u}_s) \, ds + \Psi(X_r^{t,x,\bar{u}}) \middle| \mathcal{F}_r \right) \geq V(t, x).
\]

For each \(r \in [t, T]\) the inequality

\[
\text{nEV@R}_\beta^T \left( \int_r^T c(s, X_s^{x,\bar{u}}) \, ds + \Psi(X_T^{x,\bar{u}}) \middle| \mathcal{F}_T \right) \geq V(r, X_r^{x,\bar{u}})
\]
holds almost surely and thus

\[
V(t, x) + \varepsilon \geq \inf_{u \in \mathcal{U}[t, r)} \text{nEV@R}_\beta^T \left( \int_t^r c(s, X_s^{t,x,u}, u_s) \, ds + V(r, X_r^{t,x,u}) \middle| X_t = x \right).
\]

As \(\varepsilon > 0\) can be chosen arbitrarily we have shown the inequality \(\geq\) in (37).

To see the converse inequality consider a fixed \(\varepsilon > 0\) and let \(\bar{u} \in \mathcal{U}[t, r)\) be an \(\varepsilon\)-optimal solution to (37), that is

\[
\inf_{u \in \mathcal{U}[t, r)} \text{nEV@R}_\beta^T \left( \int_t^r c(s, X_s^{t,x,u}, u_s) \, ds + V(r, X_r^{t,x,u}) \middle| X_t = x \right) + \varepsilon \\
\geq n\text{EV@R}_\beta^T \left( \int_t^r c(s, X_s^{t,x,\bar{u}}, \bar{u}_s) \, ds + V(r, X_r^{t,x,\bar{u}}) \middle| X_t = x \right).
\]

For every \(y \in \mathbb{R}\), let \(\bar{u}(y) \in \mathcal{U}[r, T]\) be such that \(V(r, y) + \varepsilon \geq V(t, y)\). We may assume that the mapping \(y \mapsto \bar{u}(y)\) is measurable (measurable selection theorem) and construct the control function

\[
\bar{u}_s = \begin{cases} 
\bar{u}_s & s \in [t, r) \\
\bar{u}_s(X_r^{t,x,\bar{u}}) & s \in [r, T].
\end{cases}
\]

Using monotonicity and the recursive property of the nested risk measure we get

\[
\text{nEV@R}_\beta^T \left( \int_t^r c(s, X_s^{t,x,\bar{u}}, \bar{u}_s) \, ds + V(r, X_r^{t,x,\bar{u}}) \middle| X_t = x \right) \\
\geq \text{nEV@R}_\beta^T \left( \int_t^r c(s, X_s^{t,x,\bar{u}}, \bar{u}_s) \, ds + V(\bar{u}_s(X_r^{t,x,\bar{u}}))(r, X_r^{t,x,\bar{u}}) \middle| X_t = x \right) - \varepsilon \\
= \text{nEV@R}_\beta^T \left( \int_t^T c(s, X_s^{t,x,u_0}, u_0) \, ds + \Psi(X_T^{t,x,u_0}) \middle| X_t = x \right) - \varepsilon \\
\geq V^{u_0}(t, x) - \varepsilon.
\]
Combining the last inequalities we get
\[
\inf_{u \in \mathcal{U}(t,r)} \text{nEV}\@ \mathbb{R}_{\mathbb{B}t}^t \left( \int_t^T c(s, X_s) ds + V(r, X_{t+h}) \bigg| X_t = x \right) + \epsilon \geq V^{u_0}(t,x) - \epsilon \geq V(t,x) - \epsilon
\]
and as \(\epsilon\) was arbitrary, the assertion follows. \(\square\)

5.2 Hamilton–Jacobi–Bellman equations

The dynamic programming principle (37) suggests to consider quantities of the form
\[
\text{nEV}\@ \mathbb{R}_{\mathbb{B}t}^t \int_t^T c(s, X_s) ds + V(T, X_T) - V(t,x) \bigg| X_t = x \right)
\]
where \(c(\cdot, \cdot)\) is a cost functional and \(V(\cdot, \cdot)\) a terminal cost functional. The next theorem extends Proposition 26 to this case.

**Theorem 36.** Let \((X_s)_{s \in T}\) be the solution of (21) on \(T\) with initial condition \(X_t = x\). Let \(c(\cdot, \cdot), V(\cdot, \cdot) \in C^{1,2}(T \times \mathbb{R})\) satisfy the growth conditions of Proposition 26, then it holds that
\[
\lim_{h \downarrow 0} \frac{1}{h} \text{nEV}\@ \mathbb{R}_{\mathbb{B}t}^{t+h} \int_t^{t+h} c(s, X_s) ds + V(t + h, X_{t+h}) - V(t,x) \bigg| X_t = x \right) = c(t,x) + \mathbb{R}_t V(t,x).
\]

**Proof.** From convexity of coherent risk measures it follows that the left side of (38) can be bounded from above by
\[
\lim_{h \downarrow 0} \frac{1}{h} \text{nEV}\@ \mathbb{R}_{\mathbb{B}t}^{t+h} \int_t^{t+h} c(s, X_s) ds \bigg| X_t = x \right) + \lim_{h \downarrow 0} \text{nEV}\@ \mathbb{R}_{\mathbb{B}t}^{t+h} \frac{V(t + h, X_{t+h}) - V(t,x)}{h} \bigg| X_t = x \right),
\]
where the first part converges to \(c(t,x)\) following the arguments in the proof of Proposition 26.

The second term can be rewritten as a limit over all possible partitions \(P\) of \([t, t+h]\) given by \(P = (t_0, \ldots, t_n)\), i.e.,
\[
\lim_{\|P\| \downarrow 0} \text{EV}\@ \mathbb{R}_{\mathbb{B}(t_0)} \left( \cdots EV\@ \mathbb{R}_{\mathbb{B}(t_{n-1})} \left( \frac{V(t + h, X_{t+h}) - V(t,x)}{h} \bigg| \mathcal{F}_{t_{n-1}} \right) \cdots \bigg| \mathcal{F}_{t_0} \right).
\]
Consequently, the iterated limit
\[
\lim_{h \downarrow 0} \lim_{\|P\| \downarrow 0} \text{EV}\@ \mathbb{R}_{\mathbb{B}(t_0)} \left( \cdots EV\@ \mathbb{R}_{\mathbb{B}(t_{n-1})} \left( \frac{V(t + h, X_{t+h}) - V(t,x)}{h} \bigg| \mathcal{F}_{t_{n-1}} \right) \cdots \bigg| \mathcal{F}_{t_0} \right)
\]
converges uniformly in \(h\) for fixed \(P\) and hence can be interchanged. By Definition 24 we then obtain
\[
\lim_{h \downarrow 0} \text{nEV}\@ \mathbb{R}_{\mathbb{B}t}^{t+h} \left( \frac{V(t + h, X_{t+h}) - V(t,x)}{h} \bigg| \mathcal{F}_{t_0} \right) = \lim_{h \downarrow 0} \text{EV}\@ \mathbb{R}_{\mathbb{B}(t_0)} \left( \frac{V(t + h, X_{t+h}) - V(t,x)}{h} \bigg| \mathcal{F}_{t_0} \right) = \mathbb{R}_t V(t,x).
\]

Now, applying the triangle inequality to
\[
\lim_{h \downarrow 0} \text{nEV}\@ \mathbb{R}_{\mathbb{B}t}^{t+h} \left( \frac{1}{h} \int_t^{t+h} c(s, X_s) ds + V(t + h, X_{t+h}) - V(t,x) \bigg| \mathcal{F}_t \right) - c(t,X_t) - \mathbb{R}_t V(t,x),
\]
the assertion follows immediately. \(\square\)
Formally taking the limit $r \to t$ in the dynamic programming principle

$$0 = \inf_{u \in U(t,r)} \frac{1}{r-t} \operatorname{nEV} \circ \mathcal{R}^T_{\beta} \left( \int_{t}^{r} c(s, X_{s}^{t,x,u}, u(s)) \, ds + V(r, X_{r}^{t,x,u}) - V(t, x) \right)_{X_t = x}$$

together with Theorem 36 shows that

$$0 = \inf_{u \in U(t,x)} c(t,x,u) + \mathcal{R}^T_{\beta} V(t,x)$$

$$= \inf_{u \in U(t,x)} \left\{ c(t,x,u) + \frac{\partial V}{\partial t}(t,x) + b(t,x,u) \frac{\partial V}{\partial x}(t,x) + \frac{\sigma^2(t,x,u)}{2} \frac{\partial^2 V}{\partial x^2}(t,x) + \sqrt{2\beta(t)} \sigma(t,x,u) \frac{\partial V}{\partial x}(t,x) \right\}.$$

Rewriting the last line as a supremum suggests to consider the partial differential equation on the space $C_b^{1,2}([t,T] \times \mathbb{R})$ for $(t,x) \in [0,T] \times \mathbb{R}$

$$\frac{\partial V}{\partial t}(t,x) = H \left( t, x, \frac{\partial V}{\partial x}, \frac{\partial^2 V}{\partial x^2} \right), \quad (39)$$

with terminal condition $v(T,x) = \Psi(x)$. Here, $H$ denotes the Hamiltonian $H: [0,T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ for the nested Entropic Value-at-Risk given by

$$H(t,x,g,H) = \sup_{u \in U} \left\{ -c(t,x,u) - g \cdot b(t,x,u) - H \cdot \frac{1}{2} \sigma^2(t,x,u) - |g| \cdot \sqrt{2\beta(t)} \sigma(t,x,u) \right\}.$$

The formal derivation suggests that the value function

$$V(t,x) = \inf_{u \in U(t,T)} \operatorname{nEV} \circ \mathcal{R}^{T}_{\beta} \left( \int_{t}^{T} c(s, X_{s}^{t,x,u}, u(s)) \, ds + \Psi(X_{T}^{t,x,u}) \right)_{X_t = x}$$

solves the nonlinear partial differential equation (39). However, the value function may not be regular enough and a more general concept of solutions for (39) is needed. Therefore, the concept of viscosity solution was introduced by Crandall and Lions (1983). We recall the definition in the next subsection and note that the classical theory of viscosity solutions developed for the risk-neutral setting is sufficient for the risk-averse case as well.

**Remark 37 (Discounted costs).** It is possible to consider a more general value function

$$V(t,x) = \inf_{u \in U(t,T)} \operatorname{nEV} \circ \mathcal{R}^{T}_{\beta} \left( \int_{t}^{T} e^{\int_{t}^{s} (\rho(r,X_{s}^{t,x,u}, u(s)) \, dr + \int_{t}^{s} \rho(r,X_{s}^{t,x,u}) \, dr) \Psi(X_{T}^{t,x,u}) \right)_{X_t = x}$$

(40)

with a discount function $\rho$. In this case the dynamic programming principle still holds and the same formal arguments as above show that $V$ defined in (40) satisfies the partial differential equation

$$-\frac{\partial V}{\partial t}(t,x) + \mathcal{H}_{\rho} \left( t, x, V, \frac{\partial V}{\partial x}, \frac{\partial^2 V}{\partial x^2} \right) = 0,$$

$$v(T,x) = \Psi(x).$$

Here, the Hamiltonian $\mathcal{H}_{\rho}$ is defined as

$$\mathcal{H}_{\rho}(t,x,V,g,H) = \sup_{u \in U} \left\{ -c(t,x,u) - g \cdot b(t,x,u) - \rho(t,x) \right\}.$$

(40)
5.3 Viscosity solutions

We show that the optimal value function $V(\cdot, \cdot)$ defined in (35) solves equation (39) and vice versa a solution of (39) is the optimal value function of problem (35). In order to discuss solutions of the partial differential equation (39) we recall the concept of viscosity solutions.

**Definition 38** (Viscosity solution). A function $v : [0, T] \times \mathbb{R} \to \mathbb{R}$ satisfying $v(T, x) = \Psi(x)$ for all $x \in \mathbb{R}$ is called a viscosity solution of (39) if the following two conditions are met:

- $v$ is a viscosity subsolution, i.e., for every $w \in C_b^{1,2}([0, T] \times \mathbb{R})$ such that $w \geq v$ on $[0, T] \times \mathbb{R}$ and $\min_{(t,x)} \{w(t, x) - v(t, x)\} = 0$, the inequality
  $$0 \geq -\frac{\partial w}{\partial t}(\bar{t}, \bar{x}) + \mathcal{H}(\bar{t}, \bar{x}, \frac{\partial w}{\partial x}, \frac{\partial^2 w}{\partial x^2})$$
  holds for every $(\bar{t}, \bar{x}) \in [0, T] \times \mathbb{R}$ such that $w(\bar{t}, \bar{x}) = v(\bar{t}, \bar{x})$.

- $v$ is a viscosity supersolution, i.e., for every $w \in C_b^{1,2}([0, T] \times \mathbb{R})$ such that $w \leq v$ on $[0, T] \times \mathbb{R}$ and $\min_{(t,x)} \{v(t, x) - w(t, x)\} = 0$, the inequality
  $$0 \leq -\frac{\partial w}{\partial t}(\bar{t}, \bar{x}) + \mathcal{H}(\bar{t}, \bar{x}, \frac{\partial w}{\partial x}, \frac{\partial^2 w}{\partial x^2})$$
  holds for every $(\bar{t}, \bar{x}) \in [0, T] \times \mathbb{R}$ such that $w(\bar{t}, \bar{x}) = v(\bar{t}, \bar{x})$.

The following theorems highlights the relation between the optimal value function $V(\cdot, \cdot)$ defined in (35) and the partial differential equation (39). We first show that the optimal value function $V(\cdot, \cdot)$ solves the Hamilton–Jacobi–Bellman partial differential equation (39) in the sense of viscosity solutions.

**Theorem 39.** Suppose the assumptions of Proposition 34 as well as Assumption 25 are satisfied and suppose the control set $U$ is compact. Then the optimal value function $V(\cdot, \cdot)$ is a viscosity solution of the equations (39).

**Proof.** Let $w \in C_b^{1,2}([0, T] \times \mathbb{R})$ be such that $w \geq V$ on $[0, T] \times \mathbb{R}$ and $\min_{(t,x)} \{w(t, x) - V(t, x)\} = 0$. Consider a point $(t', x')$ such that $w(t', x') = V(t', x')$, let $h > 0$ and consider a constant control $u_s = v$ on $[t', t' + h]$. From Lemma 35 it follows that

$$V(t', x') \leq \mathrm{neV} @ R_{\beta}^{t', t'+h} \left( \int_{t'}^{t'+h} c(s, X_s^{t', x', v}, v) \, ds + V(t' + h, X_{t'+h}^{t', x', v}) \bigg| X_t = x' \right)$$

$$\leq \mathrm{neV} @ R_{\beta}^{t', t'+h} \left( \int_{t'}^{t'+h} c(s, X_s^{t', x', v}, v) \, ds + w(t' + h, X_{t'+h}^{t', x', v}) \bigg| X_t = x' \right).$$

It follows from translation equivariance that

$$0 \leq \mathrm{neV} @ R_{\beta}^{t', t'+h} \left( \int_{t'}^{t'+h} c(s, X_s^{t', x', v}, v) \, ds + w(t' + h, X_{t'+h}^{t', x', v}) - w(t', x') \bigg| X_t = x' \right).$$

By assumption $w \in C_b^{1,2}$ and hence Itô’s formula holds. Furthermore Theorem 36 can be applied and hence

$$\frac{1}{h} \mathrm{neV} @ R_{\beta}^{t', t'+h} \left( \int_{t'}^{t'+h} c(s, X_s^{t', x', v}, v) \, ds + w(t' + h, X_{t'+h}^{t', x', v}) - w(t', x') \bigg| X_t = x' \right)$$
converges to
\[ 0 \leq c(t, x) + \frac{\partial v}{\partial t} (t, x) + b(t, x, v) \frac{\partial w}{\partial x} (t, x, v) + \frac{1}{2} \sigma^2(t, x, v) \cdot \frac{\partial^2 w}{\partial x^2} (t, x, v) + \sqrt{2\beta(t)\sigma(t, x, v)} \frac{\partial w}{\partial x} (t, x, v). \]

Since the constant control \( v \) was arbitrary it follows that
\[ 0 \leq \frac{\partial w}{\partial t} (t, x) - \mathcal{H} \left( t, x, \frac{\partial w}{\partial x}, \frac{\partial^2 w}{\partial x^2} \right) \]
and hence \( V \) is a viscosity subsolution. The same argument holds for viscosity supersolutions, which concludes the proof. \( \square \)

We finally demonstrate that a classical solution of (39) is the optimal value function of the optimal control problem (36). This provides a converse statement to Theorem 39.

**Theorem 40 (Verification theorem).** Suppose the assumptions of Proposition 34 as well as Assumption 25 are satisfied. Let \( L \in C^{1,2}_b ([0,T] \times \mathbb{R}) \) be bounded and satisfy the partial differential equation (39), then \( L(t, x) \leq V^u(t, x) \) for all \( u \in \mathcal{U}[t,T] \) and all \( (t, x) \in [0,T] \times \mathbb{R} \). Moreover, if a control \( u^* \in \mathcal{U}[0,T] \) exists such that for almost all \( (s, \omega) \in [0,T] \times \Omega \) the relation
\[
 u^*_s \in \arg \min_{v \in \mathcal{U}} \left\{ c(s, X_s^{l,x,u^*}, v) + b(s, X_s^{l,x,u^*}, v) \frac{\partial L}{\partial x} (t, X_s^{l,x,u^*}, v) + \frac{1}{2} \sigma^2(s, X_s^{l,x,u^*}, v) \frac{\partial^2 L}{\partial x^2} (t, X_s^{l,x,u^*}, v) \right\}
\]
holds, then \( L(t, x) = V(t, x) = V^u(t, x) \) for all \( (t, x) \in [0,T] \times \mathbb{R} \).

**Proof.** Let \( (t, x) \in [0,T] \times \mathbb{R} \) and consider a control \( \bar{u} \in \mathcal{U}[t,T] \). It holds by assumption that
\[ 0 = -\frac{\partial L}{\partial t} (t, x) + \mathcal{H}(t, x, \frac{\partial L}{\partial x}, \frac{\partial^2 L}{\partial x^2}) \] (42)
Equation (42) shows that for the fixed control \( \bar{u} \in \mathcal{U}[t,T] \) and all \( s \in [t,T] \)
\[ 0 \leq c(s, X_s^{l,x,\bar{u}}, \bar{u}_s) + \mathcal{R}_p L(s, X_s^{l,x,\bar{u}}). \] (43)
It now follows that
\[ L(t, x) - \int_t^T c(s, X_s^{l,x,\bar{u}}, \bar{u}_s) ds \leq L(t, x) + \int_t^T \mathcal{R}_p L(s, X_s^{l,x,\bar{u}}) ds \]
or
\[ L(t, x) \leq \text{nEV@R}^{\mathcal{U}}_{\bar{p}} \left( L(t, x) + \int_t^T \mathcal{R}_p L(s, X_s^{l,x,\bar{u}}) ds + \int_t^T c(s, X_s^{l,x,\bar{u}}, \bar{u}_s) ds \right). \]
Using Lemma 28 gives
\[ L(t, x) \leq \text{nEV@R}^{\mathcal{U}}_{\bar{p}} \left( \int_t^T c(s, X_s^{l,x,\bar{u}}, \bar{u}_s) ds + L(T, X_T^{l,x,\bar{u}}) \right) = V^\bar{u}(t, x), \]
which concludes the first part of the assertion. Now suppose a control \( u^* \in \mathcal{U}[0,T] \) exists such that (41) is satisfied, then the inequality (43) becomes an equality, where again Lemma 28 shows that
\[ L(t, x) = \text{nEV@R}^{\mathcal{U}}_{\bar{p}} \left( \int_t^T c(s, X_s^{l,x,u^*}, u^*_s) ds + \Psi(X_T^{l,x,u^*}) \right) = V^{u^*}(t, x), \]
concluding the proof. \( \square \)
6 Summary

This paper introduces nested risk measures in continuous time, which are constructed as suitable limits from discrete time risk measures. A natural choice of risk levels for the nested Entropic Value-at-Risk allows the extension to the nested Entropic Value-at-Risk in continuous time. In contrast, the popular Average Value-at-Risk does not allow for such a natural modification of its risk levels and we elaborate the correct asymptotic behavior of the risk levels.

We further discuss properties of these continuous time risk measures and provide explicit formulas for the nested Entropic Value-at-Risk in important cases. Building on this, we introduce the risk generator, an extension of the infinitesimal generator accounting for the presence of risk. Subsequently, we discuss a risk-averse optimal control problem in continuous time. Using the risk generator we derive dynamic programming equations, which extend the classical Hamilton–Jacobi–Bellman equations to the risk-averse setting. Here, the risk-averse Hamilton–Jacobi–Bellman equations involve an additional nonlinear term: the absolute value of the first derivative accounts for risk.

References


