A pointwise convergent numerical integration method for Guaranteed Lifelong Withdrawal Benefits with stochastic volatility

Yaowen Lu * Duy-Minh Dang†

February 1, 2023

Abstract

We develop an efficient and provably pointwise convergent Fourier-based numerical integration approach for Guaranteed Lifetime Withdrawal Benefit (GLWB) contracts in a realistic modeling setting with discrete withdrawals and the Cox-Ingersoll-Ross dynamics for the instantaneous variance of the sub-account' balance. Over each withdrawal interval, we formulate the GLWB no-arbitrage pricing problem as a double integral. The inner of this double integral takes the form of a convolution integral involving a conditional density of the (log) of the sub-account's balance, while the outer one involves a conditional density of the instantaneous variance. We develop a numerical Fourier-based integration method, which is stable, pointwise consistent (with respect to the double integral GLWB formulation), and ϵ-monotone, where $\epsilon > 0$ is an user-defined monotonicity tolerance. We mathematically demonstrate pointwise convergence of the proposed numerical integration scheme to the unique solution of the GLWB pricing problem as $\epsilon \to 0$. Numerical experiments demonstrate an excellent agreement with benchmark no-arbitrage prices and fair insurance fees of GLWBs obtained by Monte Carlo simulation. A study of the impact of stochastic variance of the sub-account's balance on the holder's optimal withdrawal behaviors is also presented.

Keywords: guaranteed lifelong withdrawal benefit, stochastic volatility, ϵ-monotonicity, pointwise convergence

AMS Classification 65T40, 60E10, 62P05, 91B30

1 Introduction

Variable annuities are a class of insurance products that offer various types of guaranteed benefit riders. These products provide policyholders with benefits similar to traditional life insurance or annuities, while, through guaranteed benefit riders, allowing policyholders to enjoy potentially favorable market conditions from their participation in equity investment. Guaranteed Lifetime Withdrawal Benefit (GLWB) provides unique features that combine longevity protection of an income benefit and periodic withdrawal benefit in which a certain percentage (often based on the holder's age) of the initial investment can be withdrawn as long as the policyholder lives, even if the balance of the sub-account drops to zero [5, 6, 18, 22, 25, 33, 34, 35, 51]. In principle, a GLWB is a lifelong Guaranteed Minimum Withdrawal Benefit (GMWB) with more complex withdrawal features, such as ratchet or bonus events (prompting holders not to withdraw). For the past few years, GLWBs continue to dominate living benefit riders due to their contract flexibility, income sustainability and potential market growth, particularly given the international rapid trend towards deprecation of Defined Benefit pension plans in favour of Defined Contribution plans. It is noted in [15] that GLWBs are offered by almost fourteen top variable annuity providers in North America that roughly accounted for over $65 billion in reported sales in 2017.

*School of Mathematics and Physics, The University of Queensland, St Lucia, Brisbane 4072, Australia, email: yaowen.lu@uq.edu.au

†School of Mathematics and Physics, The University of Queensland, St Lucia, Brisbane 4072, Australia, email: duyminh.dang@uq.edu.au

Research findings indicate that volatility of the balance of the sub-account exhibits strong impact on the no-arbitrage price and the fair insurance fee of a GLWB contract [10, 37, 42]. Therefore, various stochastic models for the sub-account's instantaneous variance, such as the Cox-Ingersoll-Ross (CIR) dynamics [16, 32] or 3/2 model [14], have been proposed for the GLWB no-arbitrage pricing problems (see, for example, [34, 35]). In light of the increased market volatility and the high sale volume of variable annuities, it is of significant importance to incorporate realistic modeling of volatility of the balance of the sub-account of GLWBs. In addition, it is also equally important to develop mathematically reliable numerical methods for those products, alleviating mispricing and enabling realistic and practically useful conclusions to be drawn from the numerical results. In this work, instead of considering continuous withdrawals, we primarily focus on a discrete withdrawal setting, which is adopted in practice[1]. In this setting, time-advancement between withdrawal dates typically involves a linear pricing problem, whereas, across withdrawal dates, an optimization problem needs to be solved to determine the optimal withdrawal amount.

The GLWB no-arbitrage pricing problem is an example of stochastic optimal control problems. In general, it is often the case that a closed-form expression for solutions to stochastic optimal control problems is not known to exist, and hence, these control problems must be solved numerically. In addition, since solutions to stochastic optimal control problems are often non-smooth, convergence issues of numerical methods, especially monotonicity considerations, are of primary importance. To illustrate this point further, consider a generic time-advancement scheme from time-(m-1) to time-m of the form $u_n^m = \sum_{\ell \in \mathcal{L}_n} \omega_{n,\ell}\ u_\ell^{m-1}$. Here, $\omega_{n,\ell}$ are the weights and and $\mathcal{L}_n$ is an index set typically capturing the computational stencil associated with the n-th spatial partition point. This time-advancement scheme is monotone if, for any spatial partition point, we have $\omega_{n,\ell} \geq 0$, $\forall \ell \in \mathcal{L}_n$. Optimal controls at time-m are determined typically by comparing candidates numerically computed from applying intervention on time-advancement results u_n^m. Therefore, these candidates need to be approximated using a monotone scheme as well. If interpolation is needed in this step, linear interpolation is commonly chosen, due to its monotonicity[2]. Non-monotone schemes could produce numerical solutions that fail to converge to financially relevant solution, i.e. a violation of the no-arbitrage principle [43, 45, 57]. For example, loss of monotonicity occurring in the time-advancement may result in $u_n^m < 0$ even $u_\ell^{m-1} \geq 0$ for all $\ell \in \mathcal{L}_n$.

For stochastic optimal control problems with a small number of stochastic factors, the partial differential equation (PDE) approach is a natural choice. To the best of our knowledge, finite differences methods are the only existing provably (pointwise) convergent methods for GLWBs [5, 6, 25], in which monotonicity in time-advancement is achieved via a positive coefficient discretization method (for the partial derivatives) combined with an implicit timestepping. This results in the discretization matrix being an M-matrix, and therefore, the tine-advancement scheme is monotone [23, 56]. Nonetheless, in a multi-dimensional setting, such as with stochastic variance, due to cross derivative terms in the pricing PDE, to ensure monotonicity through a positive coefficient discretization method, a wide-stencil method based on a local coordinate rotation is needed. However, this is very computationally expensive [17, 41]. In addition, in a discrete withdrawal setting, such as that for GLWBs, finite difference methods also require timestepping between (yearly) withdrawal dates. This incurs timestepping error and further increases the computational cost of the methods.

For GLWBs with stochastic variance, such as the CIR dynamics [16, 32], regression-based Monte-Carlo simulation and Fourier-based methods have been proposed (see, for example, [7, 34, 35]). Fourier-based methods often depend on the availability of an analytical expression of the Fourier transform of the underlying transition density function [2, 9, 35, 36, 38, 39, 40, 50][3]. If applicable, Fourier-based

[1]This is also similar to asset allocation, where in practice, rebalancing of portfolios is carried out discretely, see [53, 54, 55].

[2]Other non-monotone interpolation schemes are discussed in [26, 47].

[3]The well-known Fourier cosine series expansion method [19, 48] can achieve high-order convergence for piecewise smooth problems. However, optimal control problems are often non-smooth, and hence high order convergence cannot be expected.

methods offer several important advantages over finite differences and Monte-Carlo simulation, such as no timestepping error between withdrawals dates, and the capability of straightforward handling of realistic underlying dynamics, such as jump diffusion, regime-switching, and stochastic variance. However, for the GLWB no-arbitrage pricing problem, a major drawback of existing Fourier-based methods and Monte-Carlo simulation is their lack of monotonicity. In a context of Fourier-based methods for GLWBs with stochastic variance for the sub-account balance, it is remarked in [35] that the Fourier series expansion of the value function proposed therein might result in potential loss of monotonicity. A similar issue is also raised in [34] in which regression-based Monte-Carlo simulation is proposed for GLWBs with stochastic variance: the expansion of the value function in terms of a finite number of basis functions may lose monotonicity as well.

This paper aims to close the afore-mentioned research gap through the development of an efficient and provably convergent weakly monotone Fourier method for GLWB contracts with discrete withdrawals and the CIR dynamics for the instantaneous variance of the sub-account balance. Specifically, in our approach, the monotonicity requirement for time-advancement $\omega_{n,\ell} \geq 0$, $\forall \ell \in \mathcal{L}_n$ is relaxed to $\sum_{\ell \in \mathcal{L}_n} |\min(\omega_{n,\ell}, 0)| \leq \epsilon$, where $\epsilon > 0$ is an user-defined monotonicity tolerance [38, 39]. The significance of this approach lies in a full control of potential loss of monotonicity via the tolerance $\epsilon > 0$: potential loss of monotonicity is quantified and is constrained to $\mathcal{O}(\epsilon)$, allowing (pointwise) convergence to be established as $\epsilon \to 0$. In addition, no timestepping error is incurred between withdrawals dates, which is a substantial advantage over existing finite differences and Monte Carlo simulation.

We emphasize that we do not advocate for any specific stochastic variance models, but rather to (i) address an outstanding significant mathematical challenge in the no-arbitrage pricing of GLWBs with stochastic variance, as evidenced by the the above-mentioned research gap; and (ii) study the impact of stochastic volatility on variable annuities with a GLWB rider. In principle, our approach is applicable if (i) the Fourier transform of an associated conditional density of the (log) of the sub-account's balance is known in closed form, and (ii) this conditional density can be shown to satisfy very mild regularity conditions. For ease of presentation, we focus on the GLWB pricing problem with basic contract features.

The main contributions of the paper are as follows.

(i) We present a recursive and localized formulation of the GLWB no-arbitrage pricing problem in a discrete withdrawal setting, where the instantaneous variance of the sub-account's balance is given by the CIR dynamics. Over each withdrawal interval, the formulation involves a double integral. The inner integral takes the form of a convolution integral involving a conditional density of the balance of the sub-account taking the form of a convolution kernel, while the outer one is a definite integral that involves a conditional density of the variance.

(ii) We develop a numerical integration method built upon ϵ-monotone Fourier techniques for the inner integral, where $\epsilon > 0$ is an user-defined monotonicity tolerance, while the outer one is handled by Gauss-Legendre quadrature. We then propose an efficient implementation of the scheme via Fast Fourier Transform for the inner integral, including proper handling of boundary conditions and padding techniques.

(iii) We study the regularity of the conditional density in the inner integral. We mathematically demonstrate that the proposed scheme is stable, pointwise consistent with the double integral formulation, and ϵ-monotone. We rigourously prove the pointwise convergence of the scheme as the discretization parameter and the monotonicity tolerance ϵ approach zero.

(iv) Numerical experiments demonstrate an agreement with benchmark results obtained by finite difference method and Monte Carlo simulation in [25, 29], as well as the robustness of the proposed numerical methods. We shed more light on the impact of stochastic volatilities on optimal withdrawal behaviours under various scenarios.

Although we focus specifically on integration methods for GLWBs, our comprehensive and systematic approach could serve as a numerical and convergence analysis framework for the development of similar weakly monotone integration methods for control problems in finance.

In Section 2, we describe the underlying dynamics and introduce the contractual features of variable annuities embedded with GLWBs. A formulation of the no-arbitrage GLWB pricing problem in the form of a double integral together with appropriate boundary conditions is presented in Section 3, and its localization is discussed in Section 4. A numerical integration method is described in Section 5. The convergence proof of the proposed ϵ-monotone integration method is conducted in Section 6. Numerical results with respect to varying contractual parameters are given in Section 7. We also investigate the impact of stochastic volatility on fair insurance fees and optimal withdrawal behaviours. Section 8 concludes the paper and outlines possible future work.

2 Modelling

2.1 Underlying processes

We consider a probability space $(\mathcal{S}, \mathcal{F}, \{\mathcal{F}_t\}_{0\leq t\leq T}, \mathcal{Q})$ with sample space $\mathcal{S}$, sigma-algebra $\mathcal{F}$, filtration $\{\mathcal{F}_t\}_{0\leq t\leq T}$, where $T > 0$ is the fixed maturity of the GLWB contract, which generally reflects the time when the holder passes away, and a risk-neutral measure $\mathcal{Q}$ defined on $\mathcal{F}$. We respectively denote by $Z(t)$ and $A(t)$, $t \in [0, T]$, the time-t balance of the sub-account and of the guarantee account. The inception date of the GLWB contract is $t = 0$, and the balances of the sub-account and the guarantee account at inception are set to be $Z(0) = A(0) = z_0$, with z_0 be the premium paid upfront.

In practice, GLWB contracts typically allow the policy holder to withdraw at discrete, pre-determined, and often equally spaced, withdrawal times. We denote by $\mathcal{T}$ this set of withdrawal times in $[0, T]$, which is defined as follows

$$\mathcal{T} \quad = \quad \{t_m | t_m = m\Delta t,\ m = 1, \ldots, M-1\},\ \text{where } \Delta t = T/M. \tag{2.1}$$

We adopt the convention that $t_0 = 0$ is the inception date of the contract, and $t_M = T$ is the maturity date of the policy. We emphasize that $\mathcal{T}$ does not contain $t_0 = 0$ and $t_M = T$, i.e. no withdrawal is allowed at the inception date or the maturity date. Each withdrawal time $t_m \in \mathcal{T}$, $m = 1, \ldots, M-1$, is also referred to as an intervention time hereafter. For subsequent use, for any functional f, we let $f(t^-) := \lim_{\epsilon\to 0^+} f(t-\epsilon)$ and $f(t^+) := \lim_{\epsilon\to 0^+} f(t+\epsilon)$. Informally, t^- (resp. t^+) denotes the instant of time immediately before (resp. after) the forward time $t \in [0, T]$.

The evolution of the balances of the sub-account and the guarantee account in each interval $[t_{m-1}, t_m]$, $t_m \in \mathcal{T}$, can be viewed as consisting of two steps as follows. Over the time period $[t_{m-1}, t_m^-]$, no withdrawal is allowed, and the balance of the sub-account is uncontrolled, and is assumed to follow some risk-neutral dynamics; in addition, as contractually stipulated, the balance of the guarantee account remains unchanged during this time period. Over the time period $[t_m^-, t_m]$, the balances of both accounts change according to contractual features, such as withdrawals, bonuses, ratchets. In the following, we first discuss stochastic modeling of the balance of the sub-account, and then describe contractual features, and a jump condition in the balances of the two accounts due to these features.

For realistic modelling of $Z(t)$, we consider a diffusion model with a constant interest rate and stochastic variance. The instantaneous variance of $Z(t)$, denoted by $V(t)$, is uncontrolled at all $t \in [0, T]$, and is assumed to follow the well-know Cox-Ingersoll-Ros (CIR) dynamics [16]. Under the risk-neutral measure $\mathcal{Q}$, the CIR dynamics are given as follows

$$dV(t) = \lambda\left(\theta - V(t)\right)dt + \xi\sqrt{V(t)}dB_1(t), \quad t \in [0, T], \quad V(0) = \nu_0 \text{ given}, \tag{2.2}$$

where $\{B_1(t)\}_{t\in[0,T]}$ is a standard Brownian motion; λ, θ, and ξ are positive constants representing the mean-reversion rate, the long-term mean, and the instantaneous volatility, respectively.

Over the time interval $[t_{m-1}, t_m^-]$, $t_m \in \mathcal{T}$, the balance of the sub-account $Z(t)$ is assumed to follow the risk-neutral dynamics

$$dZ(t) = (r - \beta)\, Z(t)dt + \sqrt{V(t)}Z(t)\left(\rho dB_1(t) + \sqrt{1-\rho^2}dB_2(t)\right),\ t \in [t_{m-1}, t_m^-],\ t_m \in \mathcal{T}, \tag{2.3}$$

subject to (2.2) for $V(t)$. Here, in (2.3), $r > 0$ is the constant risk-free rate, $\beta > 0$ is the proportional insurance fee paid by the holder, B_1 and B_2 are uncorrelated standard Brownian motions under the risk-neutral $\mathcal{Q}$-measure, and ρ is the correlation coefficient between the processes of $Z(t)$ and $V(t)$. It is stipulated by (2.3) that, if $Z(t)$ reaches zero, it will remain at zero. Also, as noted previously, over the time interval $[t_{m-1}, t_m^-]$, $t_m \in \mathcal{T}$, the balance of the guarantee account remains $A(t)$ unchanged, i.e.

$$A(t) = A(t_{m-1}), \quad t \in [t_{m-1}, t_m^-], \quad t_m \in \mathcal{T}. \tag{2.4}$$

Remark 2.1 (Feller' condition). *For subsequent use, we define the constant $v = \frac{2\lambda\theta}{\xi^2} - 1$ which is directly related to the Feller's condition for the instantaneous variance process $\{V(t)\}$: if $v \geq 0$, the Feller's condition is satisfied, i.e. $2\lambda\theta \geq \xi^2$; otherwise, when $v < 0$, the Feller's condition is not satisfied ($2\lambda\theta > \xi^2$). When $v < 0$, as shown in [4, 27], zero is an attainable, but strongly reflecting, boundary for process* (2.2)*. For any value of v, ∞ is an unattainable boundary in this process. As discussed in [21], the definition interval for v is* $[-1, \infty)$.

We remark that the provable point-wise convergence of the proposed ϵ-monotone Fourier method to the value function does not depend on whether the Feller's condition is satisfied. However, as demonstrated theoretically and experimentally, the cost of achieving weak monotonicity does increase if the Feller's condition is not satisfied. We illustrate this through numerical experiments in Section 7.

2.2 Mortality risk

As commonly adopted in the GLWB literature, we assume that the mortality risk is well diversified across a large number of holders in this work.[4] We use the usual mortality table given in terms of integer ages of the policy holder available in the literature (see, for example, [25]).

We let m_0 be the age of the policy holder at the inception time $t = 0$ of the policy, which is a given positive integer constant. We denote by $p_m \equiv p_m^{m_0}$, $m = 0, \ldots, M$, the probability that the policy holder, who is m_0-years of age at the inception date, survives in the next m years, i.e. the holder lives past his/her $(m_0 + m)$th birthday. We also denote by $q_m \equiv q_m^{m_0}$, $m = 0, \ldots, M - 1$, the probability that the policy holder, who is m_0-years of age at the inception date, will pass away during the time interval (t_m, t_{m+1}). Here, it is understood that $p_0 = 1$, i.e. the policy holder is alive at the inception date of the policy; in addition, $p_M = q_{M-1} = 0$, i.e. the maximum longevity of the policy holder is capped at $m_0 + T$ years. For $m = 0, \ldots, M - 1$, the quantity $p_m q_m$ is the probability that the holder, who is m_0-years of age at the inception date, will pass away during the time interval (t_m, t_{m+1}). These probabilities will be included in the value function as shown subsequently.

2.3 Contractual features

The balances of the sub-account and the guarantee account change across each intervention time $t_m \in \mathcal{T}$ according to contractual features. In this work, we consider several popular features, namely withdrawal, ratchet, bonus, and death benefit.

2.3.1 Withdrawals

At each intervention time $t_m \in \mathcal{T}$, the holder is allowed to withdraw a finite amount. We denote by γ_m the withdrawal amount at each intervention time $t_m \in \mathcal{T}$. In addition, due to a penalty charge if the withdrawal amount exceeds a contractual amount (to be explained below), the net revenue cash flow may not be the same as the withdrawal amount γ_m. To this end, we denote by $f(\gamma_m)$ the function representing the net cash flow received by the holder at the withdrawal time t_m. This function is defined subsequently.

The range of γ_m is typically determined by (i) the balance of the sub-account at time t_m^-, i.e. $Z(t_m^-)$ and (ii) a pre-specified contractual amount proportional to the balance of the guarantee account at time t_m^-, i.e. proportional to $A(t_m^-)$. Specifically, this contractual amount is typically given by $C_r A(t_m^-)$,

[4] In the case that this assumption is not justified, then the risk-neutral value of the contract can be adjusted using an actuarial premium principle [28].

where $C_r > 0$ is a pre-determined contractual rate. Then, the withdrawal amount γ_m must satisfy $\gamma_m \in [0, \max(Z(t_m^-), C_r A(t_m^-))]$. Withdrawal events of a GLWB are specified according to different cases for $\max(Z(t_m^-), C_r A(t_m^-))$, as discussed below.

- If $\max(Z(t_m^-), C_r A(t_m^-)) = C_r A(t_m^-)$, then, in this case, since $\gamma_m \leq C_r A(t_m^-)$, we have

$$Z(t_m) = \max\left(Z(t_m^-) - \gamma_m, 0\right), \; A(t_m) = A(t_m^-), \; f(\gamma_m) = \gamma_m, \quad \text{if } \gamma_m \in [0, C_r A(t_m^-)]. \tag{2.5}$$

  That is, if the withdrawal amount γ_m does not exceed the pre-specified contractual amount $C_r A(t_m^-)$, then the balance of the guarantee account does not change as a result of the withdrawal; no penalty is applied to the withdrawal amount γ_m in this case.

- If $\max(Z(t_m^-), C_r A(t_m^-)) = Z(t_m^-)$, we consider two cases, namely $\gamma_m \in [0, C_r A(t_m^-)]$ and $\gamma_m \in (C_r A(t_m^-) Z(t_m^-)]$.

  If $\gamma_m \in [0, C_r A(t_m^-)]$, then $Z(t_m)$, $A(t_m)$ and $f(\gamma_m)$ are given by (2.5).

  If $\gamma_m \in (C_r A(t_m^-) Z(t_m^-)]$, it is stipulated that the excessive withdrawal amount beyond the contractual amount $C_r A(t_m^-)$, i.e. the amount $\gamma_m - C_r A(t_m^-)$, is subject to a penalty charge at a rate denoted by μ_m. Due to this penalty charge, the net revenue cash flow provided to the policy holder at time t_m is $f(\gamma_m) \equiv f(\gamma_m; A(t_m^-)) = \gamma_m - \mu_m(\gamma_m - C_r A(t_m^-))$, or equivalently,

$$f(\gamma_m) = C_r A(t_m^-) + (1 - \mu_m)(\gamma_m - C_r A(t_m^-)), \; \text{if } \gamma_m \in (C_r A(t_m^-), Z(t_m^-)]. \tag{2.6}$$

  In addition to the aforementioned penalty charge, the guarantee account balance is also reduced proportionately by the factor given by $\dfrac{Z(t_m^-) - \gamma_m}{Z(t_m^-) - C_r A(t_m^-)}$. To recap, in this case, we have

$$\begin{cases} Z(t_m) = \max(Z(t_m^-) - \gamma_m, 0), \; A(t_m) = \dfrac{A(t_m^-)(Z(t_m^-) - \gamma_m)}{Z(t_m^-) - C_r A(t_m^-)}, \\ f(\gamma_m) = C_r A(t_m^-) + (1 - \mu_m)(\gamma_m - C_r A(t_m^-)), \qquad \text{if } \gamma_m \in (C_r A(t_m^-), Z(t_m^-)]. \end{cases} \tag{2.7}$$

2.3.2 Ratchets

In a nutshell, under a ratchet provision, at an intervention time, the balance of the guarantee account, if less than that of the sub-account, can be increased to that balance; however, no withdrawal is allowed at that intervention time. The pre-determined set of contractual ratchet event times can be a subset of the set of intervention times $\mathcal{T}$. To this end, we denote by $\mathcal{T}_r \subseteq \mathcal{T}$ a pre-determined set of contractual ratchet event times. Mathematically, for $t_m \in \mathcal{T}_r$, under a ratchet provision, we have

$$Z(t_m) = Z(t_m^-), \; A(t_m) = \max\left(A(t_m^-), Z(t_m^-)\right), \; f(\gamma_m) = 0, \quad \text{if } t_m \in \mathcal{T}_r \subseteq \mathcal{T}. \tag{2.8}$$

2.3.3 Bonuses

Under a bonus provision, if the holder chooses not to withdraw at an intervention time $t_m \in \mathcal{T}$, i.e. $\gamma_m = 0$, the balance of the guarantee account can be increased proportionally by a pre-specified bonus rate, hereinafter denoted by b. Mathematically, in this case, we have

$$Z(t_m) = Z(t_m^-), \; A(t_m) = A(t_m^-)(1 + b), \; f(\gamma_m) = 0, \quad \text{if } \gamma_m = 0. \tag{2.9}$$

2.3.4 Death benefit

The death of the policy holder will terminate the contract, and the remaining balance of the sub-account will be passed to a beneficiary. Precisely, if the holder passes away during the time period (t_{m-1}, t_m), then the time-t_{m-1} balance of the sub-account, i.e. $Z(t_{m-1})$ is the death benefit which will be paid at the next event time t_m. That is, death benefits are assumed to be paid at the next event time, rather than continuously as in [25].

For subsequent references, we combine all the contractual events (2.5)-(2.9) into a case-defined jump condition for the balances of the two accounts over $[t_m^-, t_m]$, $t_m \in \mathcal{T}$, as follows.

$$(Z(t_m), A(t_m)) = \begin{cases} \left(Z(t_m^-), \max(A(t_m^-)(1+b), Z(t_m^-)\mathbf{1}_{\{t_m^- \in \mathcal{T}_r\}})\right), \\ \qquad\qquad\qquad\qquad \text{if } \gamma_m = 0, \\ \left(\max\left(Z(t_m^-) - \gamma_m, 0\right), \max\left(A(t_m^-), \max\left(Z(t_m^-) - \gamma_m, 0\right) \mathbf{1}_{\{t_m^- \in \mathcal{T}_r\}}\right)\right), \\ \qquad\qquad\qquad\qquad \text{if } \gamma_m \in (0, C_r A(t_m^-)], \\ \left(Z(t_m^-) - \gamma_m, \max\left(\dfrac{A(t_m^-)(Z(t_m^-) - \gamma_m)}{Z(t_m^-) - C_r A(t_m^-)}, (Z(t_m^-) - \gamma_m)\mathbf{1}_{\{t_m^- \in \mathcal{T}_r\}}\right)\right), \\ \qquad\qquad\qquad\qquad \text{if } \gamma_m \in (C_r A(t_m^-), Z(t_m^-)]. \end{cases} \tag{2.10}$$

Here, $\mathbf{1}_E$ is an indicator function of event E, with $\mathbf{1}_E = 1$ if E is true, and $\mathbf{1}_E = 0$ otherwise. From (2.5) and (2.6), the net revenue cash flow received by the holder at time $t_m \in \mathcal{T}$ is given by

$$f(\gamma_m) \equiv f\left(\gamma_m; A(t_m^-)\right) = \begin{cases} \gamma_m & \gamma_m \in [0, C_r A(t_m^-)], \\ C_r A(t_m^-) + (1 - \mu_m)(\gamma_m - C_r A(t_m^-)) & \gamma_m \in (C_r A(t_m^-), Z(t_m^-)]. \end{cases} \tag{2.11}$$

We emphasize that the jump condition (2.10) and the net revenue cash flow (2.11) are only realized if the policy holder does not pass away during $[t_{m-1}, t_m^+]$. If the policy holder passes away during $[t_{m-1}, t_m^+]$, the only cash flow paid at time t_m is the dead benefit, which is $Z(t_{m-1})$.

3 Formulation

For the multi-dimensional controlled underlying process $(Z(t), A(t), V(t))$, $t \in [0, T]$, we let (z, a, ν) be the state of the system. Given state (z, a, ν) at time t_m^-, $m = M-1, \ldots, 1$, the time-t_m state of the system at is denoted by $(\hat{z}, \hat{a}, \nu)$, where $(\hat{z}, \hat{a})$, as dictated by the jump condition (2.10), is given by

$$(\hat{z}, \hat{a}) \equiv \begin{cases} \left(z, \max\left(a(1+b), \quad z\mathbf{1}_{\{t_m \in \mathcal{T}_r\}}\right)\right) & \text{if } \gamma = 0, \\ \left(\max(z - \gamma, 0), \quad \max\left(a, \max(z - \gamma, 0)\mathbf{1}_{\{t_m \in \mathcal{T}_r\}}\right)\right) & \text{if } 0 < \gamma \leq C_r a, \\ \left(z - \gamma, \max\left(\dfrac{z - \gamma}{z - C_r a} a, \quad (z - \gamma)\mathbf{1}_{\{t_m \in \mathcal{T}_r\}}\right)\right) & \text{if } C_r a < \gamma \leq z. \end{cases} \tag{3.1}$$

The net revenue cash flow function $f(\gamma_m; a)$ is given by

$$f(\gamma; a) = \begin{cases} \gamma & \text{if } 0 \leq \gamma \leq C_r a, \\ \gamma(1 - \mu_{m+1}) + \mu_{m+1} C_r a & \text{if } C_r a < \gamma. \end{cases} \tag{3.2}$$

For subsequent discussions, we introduce a change of variables via the (natural) logarithmic transformation. As noted in Remark 2.1, zero is an attainable boundary for process $\{V(t)\}_{t \in [0,T]}$, while ∞ is not. Therefore, the state (z, a, ν) takes values in $[0, \infty) \times [0, \infty) \times [0, \infty)$. For $z > 0$ and $\nu > 0$, we introduce variables $w = \ln(z) \in (-\infty, \infty)$, and $\sigma = \ln(\nu) \in (-\infty, \infty)$. We denote by $u(w, a, \sigma, t_m)$, $m = M, \ldots, 0$, the time-t_m no-arbitrage price of the GLWB contract when the state of the system is $(Z(t_m), A(t_m), V(t_m)) = (e^w, a, e^\sigma)$.

Since $\log(\cdot)$ is undefined at zero, in (3.1), under the log-transformation in $w = \ln(z)$, the term $\max(z - \gamma, 0)$ becomes $\max(e^w - \gamma, e^{w_{-\infty}})$, for a finite $w_{-\infty} \ll 0$. With this in mind, the jump condition (3.1) for $(z, a) \to (\hat{z}, \hat{a})$ becomes the jump-condition for $(w, a) \to (\hat{w}, \hat{a})$ given by

$$(\hat{w}, \hat{a}) \equiv \begin{cases} \left(w, \max\left(a(1+b), e^w \mathbf{1}_{\{t_m \in \mathcal{T}_r\}}\right)\right) & \text{if } \gamma = 0, \\ \left(\ln\left(\max\left(e^w - \gamma, e^{w_{-\infty}}\right)\right), \max\left(a, \max\left(e^w - \gamma, e^{w_{-\infty}}\right)\mathbf{1}_{\{t_m \in \mathcal{T}_r\}}\right)\right) & \text{if } 0 < \gamma \leq C_r a, \\ \left(\ln\left(\max\left(e^w - \gamma, e^{w_{-\infty}}\right)\right), \max\left(\frac{e^w - \gamma}{e^w - C_r a} a, (e^w - \gamma)\mathbf{1}_{\{t_m \in \mathcal{T}_r\}}\right)\right) & \text{if } C_r a < \gamma \leq e^w. \end{cases} \tag{3.3}$$

For subsequent use, we define the intervention operator

$$\mathcal{M}(\gamma) u(w, a, \sigma, t_m^-) = u(\hat{w}, \hat{a}, \sigma, t_m) + p_{m-1} q_{m-1} e^w, \tag{3.4}$$

where $(\hat{w}, \hat{a})$ is given by (3.3).

For $(z, a, \nu) \in (0, \infty) \times [0, \infty) \times (0, \infty)$, with $w = \ln(z)$, and $\sigma = \ln(\nu)$, by dynamic programming arguments (e.g. [44, 46]), $u(w, a, \sigma, t_m)$, $m = M, \ldots, 0$, can be shown to satisfy

$$\begin{cases} u(w, a, \sigma, t_m^-) \;=\; = p_{m-1} e^w, \quad m = M, & \text{(3.5a)} \\ u(w, a, \sigma, t_m^-) \;=\; \sup\limits_{\gamma_m \in [0, e^w \vee C_r a]} \left(\mathcal{M}(\gamma_m) u(w, a, \sigma, t_m^-) \;+\; p_m f(\gamma_m; a)\right), \; \mathcal{M}(\cdot) \text{ given in (3.4)}, & \text{(3.5b)} \\ u(w, \cdot, \sigma, t_{m-1}) = \displaystyle\iint_{\mathbb{R}^2} u(w', \cdot, \sigma', t_m^-) g(w - w', \sigma, \sigma'; \Delta t)\, dw' \varrho(\cdot; \Delta t)\, d\sigma', \; m = M, \ldots, 1. & \text{(3.5c)} \end{cases}$$

Here, in (3.5a), we use the convention that, at $m = M$, $u(w, a, \sigma, t_M^-) = u(w, a, \sigma, t_M)$; In (3.5b), the intervention operator $\mathcal{M}(\cdot)$ is defined in (3.4). The double integral (3.5c) is obtained following the arguments in [21, Section 2.4], where the functions $g(\cdot; \Delta t)$ and $\varrho(\cdot; \cdot)$ are defined as follows.

- $g(\cdot; \Delta t) \equiv g_{\ln(Z)|\ln(V)}(w, w', \sigma, \sigma'; \Delta t)$ denotes the probability density of the logarithm of the balance of the sub-account at a future time (t_m^-), given the natural logarithm of the variance value as well as the information known at the current time (t_{m-1});

- $\varrho(\cdot; \cdot) \equiv p_{\ln(V)}(\sigma, \sigma'; \Delta t)$ denotes the probability density of the natural logarithm of the variance at a future time given the information at the current time; Δt is the length of the time increment.

It can be shown that $g(w, w', \sigma, \sigma'; \Delta t)$ has the form $g(w - w', \sigma, \sigma'; \Delta t)$, and therefore, in (3.5c), the inner integral takes the form of the convolution of $g(\cdot)$ and $u(\cdot, t_m^-)$. Although a closed-form expression for $g(w, w', \sigma, \sigma'; \Delta t)$ is not known to exist, its Fourier transform, denoted by $G(\cdot, \cdot; \Delta t)$, is known in closed-form. Specifically, we recall the Fourier transform and inverse Fourier transform

$$\mathfrak{F}[g(w; \cdot)] = G(\eta; \cdot) = \int_{-\infty}^{\infty} e^{-2\pi i \eta w} g(w; \cdot) dw, \quad \mathfrak{F}^{-1}[G(\eta; \cdot)] = g(w; \cdot) = \int_{-\infty}^{\infty} e^{2\pi i \eta w} G(\eta; \cdot) d\eta. \tag{3.6}$$

By [20, Equation (7)], we have

$$\begin{aligned} G(\eta; \sigma, \sigma', \Delta t) = \exp &\left(2\pi i \eta \left[(r - \beta)\Delta t + \frac{\rho}{\xi}(e^{\sigma'} - e^{\sigma} - \lambda\theta\Delta t) \right] - r\Delta t \right) \\ &\cdot \Phi\left(2\pi\eta \left(\frac{\lambda\rho}{\xi} - \frac{1}{2} \right) + \frac{1}{2} i (2\pi\eta)^2 (1 - \rho^2); e^{\sigma}, e^{\sigma'} \right). \end{aligned} \tag{3.7}$$

Here, $\Phi(\cdot)$ is defined by

$$\begin{aligned} \Phi(\varsigma; \nu, \nu') = &\frac{I_q\left[\sqrt{\nu\nu'}\frac{4c(\varsigma)e^{-\frac{1}{2}c(\varsigma)\Delta t}}{\xi^2(1-e^{-c(\varsigma)\Delta t})}\right]}{I_q\left[\sqrt{\nu\nu'}\frac{4\lambda e^{-\frac{1}{2}\lambda\Delta t}}{\xi^2(1-e^{-\lambda\Delta t})}\right]} \cdot \frac{c(\varsigma)e^{-\frac{1}{2}(c(\varsigma)-\lambda)\Delta t}(1 - e^{-\lambda\Delta t})}{\lambda\left(1 - e^{-c(\varsigma)\Delta t}\right)} \\ &\cdot \exp\left(\frac{\nu + \nu'}{\xi^2} \left[\frac{\lambda(1 + e^{-\lambda\Delta t})}{1 - e^{-\lambda\Delta t}} - \frac{c(\varsigma)(1 + e^{-c(\varsigma)\Delta t})}{1 - e^{-c(\varsigma)\Delta t}} \right] \right), \text{ with } c(\varsigma) := \sqrt{\lambda^2 - 2i\xi^2\varsigma}. \end{aligned} \tag{3.8}$$

Here, we follow the convention that the square root of a complex number is its principal one (with a positive real-part). A closed-form expression for $\varrho(\sigma, \sigma'; \Delta t)$ is given by [20, Equation (9)]

$$\begin{gathered} \varrho(\sigma, \sigma'; \Delta t) = \zeta e^{-\zeta(e^{\sigma}e^{-\lambda\Delta t} + e^{\sigma'})} \left(\frac{e^{\sigma'}}{e^{\sigma}e^{-\lambda\Delta t}} \right)^{\frac{q}{2}} e^{\sigma'} I_q\left(2\zeta e^{-\frac{1}{2}\lambda\Delta t}\sqrt{e^{\sigma}e^{\sigma'}} \right), \\ \text{with } v = 2\lambda\theta/\xi^2 - 1, \ \zeta = 2\lambda/\left((1 - e^{-\lambda\Delta t})\xi^2 \right), \end{gathered} \tag{3.9}$$

where $I_v(\cdot)$ is the modified Bessel function of the first kind with order v.

Remark 3.1 (Bang-bang control). *Under the risk-neutral dynamics* (2.2) - (2.4), $\{(Z(t), A(t), V(t))\}_{t\in[0,T]}$ *is a Markovian process and the stochastic volatility model preserves convexity, which then ensures the bang-bang analysis of the strategy space of GLWBs [5, 34]. Specifically, given the state $x = (z, a, \nu)$ at time t_m^-, $t_m \in \mathcal{T}$, the time-t_m admissible withdrawal values are: (i) zero withdrawal ($\gamma_m = 0$), (ii) contractual withdrawal amount ($\gamma_m = C_r a$), and (iii) full surrender ($\gamma_m = z$).*

For subsequent use, we introduce a result related to decay properties of function $G(\eta; \sigma, \sigma', \Delta t)$ as $|\eta| \to \infty$ for finite σ and σ'.

Lemma 3.1 (Decay properties of $|G(\eta; \cdot)|$ as $|\eta| \to \infty$). *For the function $G(\eta; \cdot)$ defined in* (3.7), *where σ and σ' are finite, we have*

$$\left| G(\eta; \sigma, \sigma', \Delta t) \right| \le C_1 \left(|\eta| e^{-\frac{1}{2}C_2|\eta|\Delta t} \right)^{v+1}, \ \textit{as } |\eta| \to \infty,$$

where $C_1, C_2 > 0$ are bounded constants independently of h, and $v = 2\lambda\theta/\xi^2 - 1$ as defined in (3.9).

Proof of Lemma 3.1. In this proof, we let $C_1 > 0$, $C_2 > 0$, and $C_3 > 0$ be generic finite constants independently of h, which may take different value from line to line. With $c(\varsigma)$ and $\Phi(\varsigma;\cdot)$ defined in (3.8), we define functions $\chi, \Psi : \mathbb{R} \to \mathbb{C}$, $\eta \mapsto \chi(\eta), \Psi(\eta)$, as follows

$$\chi(\eta) \coloneqq c(\varsigma) \equiv \sqrt{\lambda^2 - 2i\xi^2\varsigma} \text{ and } \quad \Psi(\eta) \coloneqq \Phi(\varsigma;\cdot), \quad \varsigma = 2\pi\eta\left(\frac{\lambda\rho}{\xi} - \frac{1}{2}\right) + \frac{1}{2}i(2\pi\eta)^2(1-\rho^2). \quad (3.10)$$

Recall the convention that $\chi(\eta)$ is the principal square root of a respective complex function of η, so $\mathrm{Re}(\chi(\eta)) > 0$.[5] With $\chi(\eta)$ given in (3.10), it is straightforward to obtain

$$\mathrm{Re}^2(\chi(\eta)) - \mathrm{Im}^2(\chi(\eta)) = \lambda^2 + \xi^2(2\pi)^2(1-\rho^2)\eta^2, \quad \mathrm{Re}(\chi(\eta))\,\mathrm{Im}(\chi(\eta)) = -2\pi\xi^2\left(\frac{\lambda\rho}{\xi} - \frac{1}{2}\right)\eta. \quad \forall\eta.$$

We then obtain $\mathrm{Re}^2(\chi(\eta)) \geq \lambda^2 + \xi^2(2\pi)^2(1-\rho^2)\eta^2$ which leads to $\mathrm{Im}^2(\chi(\eta)) \leq C_1$. Therefore, we have

$$\mathrm{Re}(\chi(\eta)) \geq \max\{\lambda, C_1|\eta|\}, \quad |\mathrm{Im}(\chi(\eta))| \leq C_2, \ \forall\eta\in\mathbb{R}, \text{ and } |\chi(\eta)| \leq C_3|\eta|, \text{ as } |\eta|\to\infty. \quad (3.11)$$

With a closed-form expression of $G(\eta;\cdot)$ given in (3.7), we have

$$|G(\eta;\cdot)| \overset{(3.7)}{=} C_1\,|\Phi(\varsigma;\cdot)| = C_1\,|\Psi(\eta;\cdot)| \overset{\text{(ii)}}{=} C_2\,|\Psi_1(\eta;\cdot)|\,|\Psi_2(\eta;\cdot)|\,|\Psi_3(\eta;\cdot)|,$$

where $\Psi_1(\eta) = I_q[\Psi_2(\eta)]$, $\Psi_2(\eta) = \frac{\chi(\eta)e^{-\frac{1}{2}\chi(\eta)\Delta t}}{\left(1-e^{-\chi(\eta)\Delta t}\right)}$ and $\Psi_3(\eta) = \exp\left(-\frac{\chi(\eta)(1+e^{-\chi(\eta)\Delta t})}{1-e^{-\chi(\eta)\Delta t}}\right)$. Here, (ii) is due to the fact that σ and σ' and all the model parameters bounded, and that $I_q[\cdot]$ in the denominator of $G(\eta;\cdot)$ is independent of η, and hence its modulus is also bounded by a positive constant.

We investigate $|\Psi_2(\eta;\cdot)| = \frac{|\chi(\eta)|\left|e^{-\frac{1}{2}\chi(\eta)\Delta t}\right|}{\left|1-e^{-\chi(\eta)\Delta t}\right|}$. As noted in (3.11), as $|\eta|\to\infty$, $|\chi(\eta)| \leq C_1|\eta|$, and $\left|e^{-\frac{1}{2}\chi(\eta)\Delta t}\right| = e^{-\frac{1}{2}\mathrm{Re}(\chi(\eta))\Delta t} \leq e^{-\frac{1}{2}C_2|\eta|\Delta t}$ as $|\eta|\to\infty$. In addition, noting $\mathrm{Re}(\chi(\eta)) \geq \lambda$ for all η gives $\frac{1}{\left|1-e^{-\chi(\eta)\Delta t}\right|} \leq \frac{1}{|1-e^{-\mathrm{Re}(\chi(\eta))\Delta t}|} \leq C_1$. Therefore,

$$|\Psi_2(\eta;\cdot)| \leq C_1|\eta|e^{-\frac{1}{2}C_2|\eta|\Delta t}, \text{ as } |\eta|\to\infty. \quad (3.12)$$

For the term $|\Psi_1(\eta)| = |I_q[\Psi_2(\eta)]|$, we recall a classical result on Bessel functions from [58, p. 51]: for complex numbers v and c, $I_v(c) = \frac{\left(\frac{1}{2}c\right)^v}{\Gamma(v+1)}(1+\varpi)$, where $|\varpi| < \exp\left(\frac{\frac{1}{4}|c|^2}{|v_0+1|}\right) - 1$ with $|v_0+1|$ is the smallest of the numbers $|v+1|$, $|v+2|$, ..., and $\Gamma(\cdot)$ is the Gamma function. Since our case, v is a bounded real number, applying this result on $I_v[\Psi_2(\eta)]$ gives

$$|\Psi_1(\eta)| = |I_v[\Psi_2(\eta)]| \leq C_1\,|\Psi_2(\eta)|^v\, e^{|\Psi_2(\eta)|^2} \overset{\text{(i)}}{\leq} C_1\left(|\eta|e^{-\frac{1}{2}C_2|\eta|\Delta t}\right)^v, \quad |\eta|\to\infty, \quad (3.13)$$

where (i) is due to (3.12), noting $e^{|\Psi_2(\eta)|^2} \to 1$ as $|\eta|\to\infty$.

For the term $|\Psi_3(\eta)|$, we have

$$|\Psi_3(\eta)| = \exp\left(-\mathrm{Re}\left(\frac{\chi(\eta)(1+e^{-\chi(\eta)\Delta t})}{1-e^{-\chi(\eta)\Delta t}}\right)\right) \to C_1e^{-\mathrm{Re}(\chi(\eta))} \leq C_1e^{-|\chi(\eta))|} = C_1e^0 \text{ as } |\eta|\to\infty. \quad (3.14)$$

Putting (3.12), (3.13), and (3.14) together gives the desired result. □

Remark 3.2 (Smoothness of $g(w;\cdot)$). *We recall a classical result, namely the Paley-Wiener theorem (see [13]): the decay of $|G(\eta;\cdot)|$ as $|\eta|\to\infty$ is also reflected in the smoothness of $g(w;\cdot)$. Specifically, since $g(w;\cdot)$ is integrable, if $|G(\eta;\cdot)| \leq C(1+|\eta|)^{-c}$, $c > 2$ as $|\eta|\to\infty$, then $g(w;\cdot)$ is continuous and has c continuous derivatives. From Lemma (3.1), if $v > -1$, then we can conclude that $g(w;\cdot)$ is at least in $\mathcal{C}^2(\mathbb{R})$. Although we note in Remark 2.1 that the definition range for v is $[-1,\infty)$, for typical market data available in the literature, $v > -1$ [3, 21]. Nonetheless, as we illustrate subsequently, the cost of achieving weak monotonicity (within a user-defined tolerance) does increase as $v \to -1$ (as more terms in an associated Fourier series are required).*

[5] If $\mathrm{Re}(\chi(\eta)) < 0$, similar steps of this proof can be followed to arrive at the same asymptotic result.

4 Localization

Under the log-transformation, the GLWB formulation (3.5) is posed on the infinite domain. For the problem statement and convergence analysis of numerical schemes, we define a localized GLWB formulation. To this end, with $w_{\min} < 0 < w_{\max}$, $\sigma_{\min} < 0 < \sigma_{\max}$, and $|w_{\min}|$, $w_{\max}$, $|\sigma_{\min}|$ and $\sigma_{\max}$ are sufficiently large, and $a_{\max} = e^{w_{\max}}$, we define the following spatial sub-domains:

$$\begin{aligned} &\Omega^{\infty} = (-\infty, \infty) \times [0, \infty] \times [\sigma_{\min}, \sigma_{\max}], && \Omega^{\infty}_{w_{\min}} = (-\infty, w_{\min}] \times [0, a_{\max}] \times [\sigma_{\min}, \sigma_{\max}], \\ &\Omega_{\text{in}} = (w_{\min}, w_{\max}) \times [0, a_{\max}] \times [\sigma_{\min}, \sigma_{\max}], && \Omega^{\infty}_{w_{\max}} = [w_{\max}, \infty) \times [0, a_{\max}] \times [\sigma_{\min}, \sigma_{\max}], \end{aligned} \tag{4.1}$$

We now present equations for spatial sub-domains defined in (4.1). We note that boundary conditions for $w \to -\infty$ and $w \to \infty$ are obtained by relevant asymptotic forms of the no-arbitrage price of GLWB when $z \to 0$ and $z \to \infty$, respectively, similar to [25, 34]. We also note that the terminal and boundary solutions in $\Omega^{\infty} \times \{T\}$ and $\Omega^{\infty}_{w_{\max}} \times \{t\}$, $t \in (0, T)$, may grow unbounded as $w \to \infty$. Therefore, to ensure boundedness of numerical solutions in the interior sub-domains Ω_{in}, where pointwise convergence to the unique solution is studied, we require the terminal and boundary solutions in $\Omega^{\infty} \times \{T\}$ and $\Omega^{\infty}_{w_{\max}} \times \{t\}$ to be bounded as $w \to \infty$. This is detailed below.

- For $(w, a, \sigma, T) \in \Omega^{\infty} \times \{T\}$, we apply the terminal condition (3.5a)
$$u(w, a, \sigma, T^-) = p_{m-1}(e^w \wedge e^{w_\infty}), \text{ for a finite } w_\infty \gg w_{\max}. \tag{4.2}$$

- As $w \to -\infty$, $z = e^w \to 0$. In [34][Proposition 2], an analytical solution to the no-arbitrage price of the GLWB when the balance of the sub-account $z = 0$ is given. Therefore, for $(w, a, \sigma, t) \in \Omega^{\infty}_{w_{\min}} \times \{t\}$, $t \in (0, T)$, we impose the boundary condition given by the above-mentioned analytical solution as follows
$$u(w, a, \sigma, t) = \max_{i^* \in \{i_0, i_0+1, \ldots, T-1\}} \left[\prod_{i=i_0}^{i^*-1} (1 + b) \left(\sum_{\tau = i^*}^{T-1} e^{-r(\tau - t)} p_\tau \right) \right] C_r a, \tag{4.3}$$
for $i_0 \leq T - 2$, where $i_0 := \inf\{i : t_i > t\}$. For each time $t = t_m^-$, the equation (4.3) can be solved by exhaustive search.

- For $(w, a, \sigma, t) \in \Omega^{\infty}_{w_{\max}} \times \{t\}$, $t \in (0, T)$, using similarity reduction results from [25, 34], we impose the boundary solutions
$$u(w, a, \sigma, t) = \frac{a}{a^*} u \left(\ln \left(\frac{e^w a^*}{a} \right), a^*, \sigma, t \right) \wedge e^{w_\infty}, \tag{4.4}$$
where a^* is selected such that $(\ln \left(\frac{e^w a^*}{a} \right), a^*, \sigma) \in \Omega_{\text{in}}$, i.e. a^* can be different for different w. For $t = t_m$, the boundary condition (4.4) can be obtained using the solution in $\Omega_{\text{in}} \times \{t_m\}$.

  We note that the theoretical quantity w_∞ is needed to indicate that the solutions in $\Omega^{\infty} \times \{T\}$ and $\Omega_{\text{in}} \times \{t_m\}$ are bounded as $w \to \infty$, and it does not need to be numerically specified.

- For $(w, a, \sigma, t_m^-) \in \Omega_{\text{in}} \times \{t_m^-\}$, $t_m \in \mathcal{T}$, the intervention result $u(w, a, \sigma, t_m^-)$ is given by (3.5b), i.e.
$$u(w, a, \sigma, t_m^-) = \sup_{\gamma_m \in [0, e^w \vee C_r a]} \left(\mathcal{M}(\gamma_m) u(x, t_m) \; + \; p_m f(\gamma; a) \right), \tag{4.5}$$
where the intervention operator $\mathcal{M}(\cdot)$ is defined in (3.4).

  For $(w, a, \sigma, t_{m-1}) \in \Omega_{\text{in}} \times \{t_{m-1}\}$, $t_m \in \mathcal{T}$, $u(x, t_{m-1})$ is given by (3.5c), which, after taking into account the boundary conditions in $\Omega^{\infty}_{w_{\min}} \times \{t_m\}$ given by (4.3) and $\Omega^{\infty}_{w_{\max}} \times \{t_m\}$, becomes
$$u(w, a, \sigma, t_{m-1}) = \int_{\sigma_{\min}}^{\sigma_{\max}} \int_{\mathbb{R}} \hat{u}(w', a, \sigma', t_m^-) \; g(w - w', \sigma, \sigma'; \Delta t) \; dw' \varrho(\sigma, \sigma'; \Delta t) \; d\sigma', \tag{4.6}$$
where the terminal condition $\hat{u}(w', a, \sigma', t_m^-)$ is given by
$$\hat{u}(w', a, \sigma', t_m^-) = \begin{cases} u(w', a, \sigma', t_m) \text{ satisfies (4.3)} & (w', a, \sigma') \in \Omega^{\infty}_{w_{\min}}, \\ u(w', a, \sigma', t_m^-) \text{ satisfies (4.5)} & (w', a, \sigma') \in \Omega_{\text{in}}, \\ u(w', a, \sigma', t_m) \text{ satisfies (4.4)} & (w', a, \sigma') \in \Omega^{\infty}_{w_{\max}}. \end{cases} \tag{4.7}$$

In Definition 4.1 below, we formally define a GLWB pricing problem.

Definition 4.1 (Localized GLWB pricing problem). *The GLWB pricing problem under a discrete withdrawal setting with the set of withdrawal times being $\mathcal{T}$ defined in* (2.1)*, and dynamics* (2.2)-(2.3)-(2.4) *is defined in $\Omega^{\infty} \times \mathcal{T} \cup \{t_0, t_M\}$ as follows. At each t_{m-1}, $t_m \in \mathcal{T}$, the solution to the GLWB pricing problem $u(w, a, \sigma, t_{m-1})$ satisfies (i) the double integral* (4.6) *in $\Omega_{in} \times \{t_{m-1}\}$, and (ii) the boundary conditions* (4.3) *and* (4.4) *in $\left\{\Omega^{\infty}_{w_{\min}}, \Omega^{\infty}_{w_{\max}}\right\} \times \{t_m\}$, $t_m \in \mathcal{T} \cup t_0$, respectively, subject to the terminal condition* (4.2) *in $\Omega^{\infty} \times \{t_M\}$.*

We introduce a result on uniform continuity of the solution to the GLWB pricing problem.

Proposition 4.1. *The solution $u(w, a, \sigma, t_m)$ to the GLWB pricing problem in Definition 4.1 is uniformly continuous within each sub-domain $\Omega_{in} \times \{t_m\}$, $m = M-1, \ldots, 0$.*

Proof of Proposition 4.1. This proposition can be proved using mathematical induction on m. For brevity, we outline key details below. We first observe that if $u(x, t)$ is a uniformly continuous function, then $\sup_{\gamma_m \in [0, e^w \vee C_r a]} \mathcal{M}(\gamma) u(x, t)$, where $\mathcal{M}(\cdot)$ defined in (3.4), is also uniformly continuous [30, Lemma 2.2] (also see [34, Proposition 3]). Therefore, it follows that if $u(w, a, \sigma, t_m)$, $m = M-1, \ldots, 1$ is uniformly continuous then $u(w, a, \sigma, t_m^-)$ defined in (4.5) is also uniformly continuous. The other key step is to show that, if $u(w, a, \sigma, t_m^-)$, $m = M, \ldots, 1$, is uniformly continuous, then the solution $u(w, a, \sigma, t_{m-1})$ for $(w, a, \sigma) \in \Omega_{\text{in}}$ given by the double integral (4.6) is also uniformly continuous. Combining these above two steps with the fact that $u(w, a, \sigma, t_M)$ given in (4.2) is uniformly continuous in (w, a, σ) gives the desired result. □

We conclude this section by emphasizing that $\Omega_{\text{in}} \times \{t_m\}$, $m = M-1, \ldots, 0$, is the target region where provable pointwise convergence of the proposed numerical method is investigated, which relies on Proposition 4.1.

5 Numerical methods

The key challenge in the development of numerical schemes for the GLWB pricing problem given by Definition 4.1 is the numerical approximation of $u(w, \cdot, \sigma, t_{m-1})$, via the double integral (4.6), for $(w, \sigma) \in (w_{\min}, w_{\max}) \times [\sigma_{\min}, \sigma_{\max}]$. The solution $u(w, \cdot, \sigma, t_{m-1})$ for $w \notin (w_{\min}, w_{\max})$ and $\sigma \in [\sigma_{\min}, \sigma_{\max}]$ is given by the boundary conditions (4.3) and (4.4). For computational purposes, we truncate the infinite region of integration of the convolution (the inner integral) in (4.6) to $[w^{\dagger}_{\min}, w^{\dagger}_{\max}]$, where $w^{\dagger}_{\min} \ll w_{\min} < 0 < w^{\dagger}_{\max} \ll w^{\dagger}_{\max}$, and $|w^{\dagger}_{\min}|$ and $w^{\dagger}_{\max}$ are sufficiently large. This results in the approximation

$$u(w, \cdot, \sigma, t_{m-1}) \simeq \int_{\sigma_{\min}}^{\sigma_{\max}} \left(\int_{w^{\dagger}_{\min}}^{w^{\dagger}_{\max}} \hat{u}(w', \cdot, \sigma', t_m^-)\ g(w - w', \sigma, \sigma'; \Delta t)\ dw' \right) \varrho(\sigma, \sigma'; \Delta t)\ d\sigma', \quad (5.1)$$

where $(w, \sigma) \in (w_{\min}, w_{\max}) \times [\sigma_{\min}, \sigma_{\max}]$. The error arising from this truncation is discussed in Section 6.

With the above in mind, we define finite computational domain and sub-domains as follows

$$\begin{aligned} \Omega &= [w^{\dagger}_{\min}, w^{\dagger}_{\max}] \times [0, a_{\max}] \times [\sigma_{\min}, \sigma_{\max}], \quad \Omega_{\text{in}} \text{ defined in (4.1)}, \qquad (5.2) \\ \Omega_{w_{\min}} &= [w^{\dagger}_{\min}, w_{\min}] \times [0, a_{\max}] \times [\sigma_{\min}, \sigma_{\max}], \quad \Omega_{w_{\max}} = [w_{\max}, w^{\dagger}_{\max}] \times [0, a_{\max}] \times [\sigma_{\min}, \sigma_{\max}]. \end{aligned}$$

We stress that the region $(\Omega_{w_{\min}} \cup \Omega_{w_{\max}}) \times (0, T)$ plays an important role in the proposed numerical method. It is well-documented that wraparound error (due to periodic extension) is an important issue for Fourier methods, particularly in the case of control problems (see, for example, [39]). Therefore, in (5.2), the region $(\Omega_{w_{\min}} \cup \Omega_{w_{\max}}) \times (0, T)$ is also set up to serve as padding areas for nodes in Ω_{in}. For this purpose, we assume that $|w_{\min}|$ and $w_{\max}$ are chosen sufficiently large so that

$$w^{\dagger}_{\min} = w_{\min} - \frac{w_{\max} - w_{\min}}{2} \quad \text{and} \quad w^{\dagger}_{\max} = w_{\max} + \frac{w_{\max} - w_{\min}}{2}. \quad (5.3)$$

Due to withdrawals, the intervention action in (4.5) may require evaluating a candidate value at a point having $w = \ln(\max(e^w - \gamma, e^{w_{-\infty}}))$, which could be outside $[w^{\dagger}_{\min}, w^{\dagger}_{\max}]$ if $w_{-\infty} < w^{\dagger}_{\min}$. Without loss of generality, we assume $w_{\text{-}\infty} \geq w^{\dagger}_{\min}$. Therefore, in computing $(\hat{w}_n, \hat{a}_j)$ via (3.3), $w_{\text{-}\infty}$ is replaced by $w^{\dagger}_{\min}$ in (3.3).

Remark 5.1 (Minimizing the wrap-around errors)**.** *Recall that we apply the asymptotic form of value solution in the region* $(\Omega_{w_{\min}} \cup \Omega_{w_{\max}}) \times (0, T)$ *to minimize the wrap-around errors. As presented in Section 5.2.1 below, for each time advancement, we initially pre-set the approximation values for every node in this area using suitable asymptotic forms around the boundaries and then discard the Fast-Fourier Transform (FFT) values of these nodes after applying the Fourier method to these nodes.*

As discussed above, the numerical scheme for problem (3.5) consists of two main parts, namely intervention actions due to withdrawals (3.5b) and time advancement (3.5c). The key challenge here is the development of a (weakly) monotone scheme for time advancement, which allows us to prove the pointwise convergence of the numerical scheme.

5.1 Discretization

The computational grid is constructed as follows:

(i) We denote by N (resp. $N^{\dagger}$) the number of points of a uniform partition of $[w_{\min}, w_{\max}]$ (resp. $[w^{\dagger}_{\min}, w^{\dagger}_{\max}]$). For convenience, we typically choose $N^{\dagger} = 2N$ so that only one set of w-coordinates is needed. Also let $P = w_{\max} - w_{\min}$, and $P^{\dagger} = w^{\dagger}_{\max} - w^{\dagger}_{\min}$. We define $\Delta w = \frac{P}{N} = \frac{P^{\dagger}}{N^{\dagger}}$. We use an equally spaced partition in the w-direction, denoted by $\{w_n\}$, where

$$\begin{aligned} w_n &= \hat{w}_0 + n\Delta w; \quad n = -N^{\dagger}/2, \ldots, N^{\dagger}/2, \quad \text{where} \\ \Delta w &= P/N = P^{\dagger}/N^{\dagger}, \quad \text{and } \hat{w}_0 = (w_{\min} + w_{\max})/2 = (w^{\dagger}_{\min} + w^{\dagger}_{\max})/2. \end{aligned} \tag{5.4}$$

(ii) We use an unequally spaced partition in the a-direction, denoted by $\{a_j\}$, $j = 0, \ldots, J$, with $\Delta a_{\max} = \max_{0 \leq j \leq J-1} (a_{j+1} - a_j)$, $\Delta a_{\min} = \max_{0 \leq j \leq J-1} (a_{j+1} - a_j)$.

(iii) We use an unequally spaced partition in the σ-direction, denoted by $\{\sigma_k\}$, $k = 0, \ldots, K-1$, where the nodes are given by the Gauss-Legendre quadrature rule.

We emphasize that no timestepping is required for the interval $[t_{m-1}, t_m^-]$, $t_m \in \mathcal{T}$. As noted earlier, $\Delta t = T/M$ is kept constant.

Assumption 5.1.

$$\Delta w = C_1 h, \quad \Delta a_{\max} = C_2 h, \quad \Delta a_{\min} = C_2' h, \quad K = C_3/h, \quad P^{\dagger} = C_4/h, \tag{5.5}$$

where the positive constants C_1*,* C_2*,* C_2'*,* C_3*, and* C_4 *are independent of* h*.*

It is also straightforward to ensure the theoretical requirement $P^{\dagger} \to \infty$ as $h \to 0$. For example, with $C_4 = 1$ in (5.5), we can quadruple $N^{\dagger}$ as we halve h.

For convenience, we define sets of indices: $\mathbb{N}^{\dagger} = \left\{-N^{\dagger}/2, \ldots, N^{\dagger}/2 - 1\right\}$, $\mathbb{N} = \{-N/2 + 1, \ldots, N/2 - 1\}$, $\mathbb{J} = \{0, \ldots, J\}$, and $\mathbb{K} = \{0, \ldots, K - 1\}$. We occasionally use $\mathbf{x}^m_{n,j,k} \equiv (\mathbf{x}_{n,j,k}, t_m)$ to refer to the reference gridpoint (w_n, a_j, ν_k, t_m), $n \in \mathbb{N}^{\dagger}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, $m = M, \ldots, 0$. We denote by $u^m_{n,j,k}$ a numerical approximation to the exact solution to $u(w_n, a_j, \nu_k, t_m)$.

5.2 Numerical schemes

For $(w_n, a_j, \sigma_k, t_M) \in \Omega \times \{T\}$, we impose the terminal condition (4.2) by

$$u^{M-}_{n,j,k} = p_{M-1}(e^{w_n} \wedge e^{w_\infty}), \quad n \in \mathbb{N}^{\dagger}, \ j \in \mathbb{J}, \ k \in \mathbb{K}. \tag{5.6}$$

We impose the condition (4.3) for $(w_n, a_j, \sigma_k, t_m) \in \Omega_{w_{\min}} \times \{t_m\}$, $t_m \in \mathcal{T}$, by

$$\begin{aligned} u^m_{n,j,k} &= \max_{i^* \in \{i_0, i_0+1, \ldots, T-1\}} \left[\prod_{i=i_0}^{i^*-1} (1+b) \left(\sum_{\tau=i^*}^{T-1} e^{-r(\tau-m)} p_\tau \right) \right] C_r a_j, \\ & n = -N^{\dagger}/2, \ldots, -N/2, \ j \in \mathbb{J}, \ k \in \mathbb{K}. \end{aligned} \tag{5.7}$$

for $i_0 \leq T - 2$, where $i_0 := \inf\{i : t_i > t_m\}$.

We impose the condition (4.4) for $(w_n, a_j, \sigma_k, t_m) \in \Omega_{w_{\max}} \times \{t_m\}$, $t_m \in \mathcal{T}$, as follows

$$u^m_{n,j,k} = \frac{a_j}{a^*} u \left(\ln\left(\frac{e^{w_n} a^*}{a_j} \right), a^*, \sigma_k, t_m \right) \wedge e^{w_\infty}, \ n = N/2, \ldots, N^{\dagger}/2 - 1, \ j \in \mathbb{J}, \ k \in \mathbb{K}. \tag{5.8}$$

where a^* is selected such that $(\ln\left(\frac{e^w a^*}{a}\right), a^*, \sigma) \in \Omega_{\text{in}}$.

We now focus on the interior spatial domain Ω_{in}, which forms the region of convergence of the numerical scheme. As discussed previously, over the time interval $[t_{m-1}, t_m]$, there are two key components to the proposed numerical scheme, namely (i) the intervention action over $[t_m^-, t_m]$ as given in (4.5), and (ii) the time advancement from t_m^- to t_{m-1}, as captured by the double integral (5.1).

5.2.1 Ω_{in}: a monotone scheme

In subsequent discussions, we denote by $\gamma_{n,j,k}^m$ the control representing the withdrawal amount at node $(w_n, a_j, \sigma_k, t_m) \in \Omega_{\text{in}} \times \{t_m\}$, where $t_m \in \mathcal{T}$. We let $\hat{u}_{n,j,k}^m$ be an approximation to $u(\hat{w}_n, \hat{a}_j, \sigma_k, t_m)$, where $(\hat{w}_n, \hat{a}_j)$ is given by (3.3) computed by linear interpolation. To this end, for fixed $k \in \mathbb{K}$, we denote by $\mathcal{I}\{u_k^m\}(w, a)$ a two-dimensional linear interpolation operator acting on the time-t_m discrete solution $\left\{(w_n, a_j), u_{n,j,k}^m\right\}$, $(n, j) \in \mathbb{N}^\dagger \times \mathbb{J}$. Then, $\hat{u}_{n,j,k}^m$ is computed as follows

$$\hat{u}_{n,j,k}^m = \mathcal{I}\{u_k^m\}(\hat{w}_n, \hat{a}_j), \quad \text{where } (\hat{w}_n, \hat{a}_j) \text{ is given by (3.3)}, \ k \in \mathbb{K} \text{ fixed}. \tag{5.9}$$

Here, we note that in computing $(\hat{w}_n, \hat{a}_j)$ via (3.3), $w_{-\infty}$ is replaced by $w_{\min}^\dagger$ in (3.3). We then compute the intermediate result $u_{n,j,k}^{m-}$ by solving the optimization problem

$$u_{n,j,k}^{m-} = \sup_{\gamma_{n,j,k}^m \in [0, e^{w_n} \vee C_r a_j]} \left(\hat{u}_{n,j,k}^m + p_m f\left(\gamma_{n,j,k}^m; a_j\right)\right) + p_{m-1} q_{m-1} e^{w_n}, \ (n, j, k) \in \mathbb{N} \times \mathbb{J} \times \mathbb{K}, \tag{5.10}$$

where $\hat{u}_{n,j,k}^m$ is given by (5.9) and $f\left(\gamma_{n,j,k}^m; a_j\right)$ is defined in (3.2).

Remark 5.2 (Attainability of supremum). *Guaranteed by the existence of bang-bang control, established in [5, 34], we can expect that the supremum defined in* (5.10) *must be attainable. This can be also shown from the boundedness of nodal values used in* $\mathcal{I}\{u_k^m\}(\cdot)$ *(see Lemma 6.2 on stability).*

Next, we focus on the numerical approximation of the double integral (5.1). The outer integral of (5.1) is handled by a K-point Gauss-Legendre quadrature rule along the σ-direction [12, 31, 52], whereas for the convolution in the inner one, an ϵ-monotone Fourier method is utilized. Specifically, for the reference point $(w_n, a_j, \sigma_k, t_{m-1})$, $t_m \in \mathcal{T}$, approximating the outer integral of the rhs-of-(5.1) using a K-point Gauss-Legendre quadrature rule gives an intermediate result denoted by $\tilde{u}_{n,j,k}^{m-1}$, $(n, j, k) \in \mathbb{N} \times \mathbb{J} \times \mathbb{K}$, as follows

$$\text{rhs-of-(5.1)} \simeq \sum_{s \in \mathbb{K}} \vartheta_s \ \varrho(\sigma_k, \sigma_s; \Delta t) \left(\int_{w_{\min}^\dagger}^{w_{\max}^\dagger} \hat{u}(w, a_j, \sigma_s, t_m^-) \ g(w_n - w, \sigma_k, \sigma_s; \Delta t) \ dw \right). \tag{5.11}$$

Here, the terms of ϑ_s, $s \in \mathbb{K}$, are the weights at the quadrature node $\sigma_s \in [\sigma_{\min}, \sigma_{\max}]$ of the quadrature rule. More details about the quadrature nodes and weights can be referred to [12, 31, 52].

To prepare for the convolution integral term in (5.11), where $(j, s) \in \mathbb{J} \times \mathbb{K}$ is fixed, we combine the intervention results for $\Omega_{\text{in}} \times \{t_m^-\}$ given by (5.10) and the padding (previously computed) values in $(\Omega_{w_{\min}} \cup \Omega_{w_{\max}}) \times \{t_m\}$ (see Remark 5.1) as below (with a slight abuse of notation)

$$u_{l,j,s}^{m-} = \begin{cases} u_{l,j,s}^m & \text{in (4.3)}, \quad l = -N^\dagger/2, \ldots, -N/2, \\ u_{l,j,s}^{m-} & \text{in (5.10)}, \quad l = -N/2 + 1, \ldots, N/2 - 1, \\ u_{l,j,s}^m & \text{in (4.4)}, \quad l = N/2, \ldots, N^\dagger/2 - 1. \end{cases} \tag{5.12}$$

For the integral in (5.11), the terminal condition $\hat{u}(w, a_j, \sigma_s, t_m^-)$, where $(j, s) \in \mathbb{J} \times \mathbb{K}$ is fixed, is replaced by a linear combination of the discrete values in (5.12). That is, $\hat{u}(w, a_j, \sigma_s, t_m^-)$ is given by

$$\hat{u}(w, a_j, \sigma_s, t_m^-) = \sum_{l \in \mathbb{N}^\dagger} \varphi_l(w) \, u_{l,j,s}^{m-}, \quad w \in [w_{\min}^\dagger, w_{\max}^\dagger], \ (j, s) \in \mathbb{J} \times \mathbb{K} \text{ fixed}, \tag{5.13}$$

where $\{\varphi_l(w)\}_{l \in \mathbb{N}^\dagger}$ are (hat-shaped) piecewise linear basis functions defined by

$$\varphi_l(w) = \begin{cases} (w - w_{l-1}) / \Delta w, & w_{l-1} \le w \le w_l, \\ (w_{l+1} - w) / \Delta w, & w_l \le w \le w_{l+1}, \\ 0, & \text{otherwise}. \end{cases} \tag{5.14}$$

Substituting $\hat{u}(w, a_j, \sigma_s, t_m^-)$ from (5.13) into the convolution integral in (5.11) gives an approximation for the rhs-of-(5.11) that involves a discrete convolution as follows

$$\begin{aligned}\text{rhs-of-(5.11)} &\simeq \sum_{s\in\mathbb{K}} \vartheta_s\ \varrho(\sigma_k, \sigma_s; \Delta t)\ \left(\int_{w_{\min}^\dagger}^{w_{\max}^\dagger} g(w_n - w, \sigma_k, \sigma_s; \Delta t)\Big(\sum_{l\in\mathbb{N}^\dagger} \varphi_l\,(w)\, u_{l,j,s}^{m-}\Big)\, dw\right)\\ &\overset{(i)}{=} \sum_{s\in\mathbb{K}} \vartheta_s\ \varrho(\sigma_k, \sigma_s; \Delta t)\ \Big(\Delta w \sum_{l\in\mathbb{N}^\dagger} u_{l,j,s}^{m-}\ \tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t)\Big). \qquad (5.15)\end{aligned}$$

Here, in (i), $\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t) \equiv \tilde{g}(w_n - w_l, \sigma_k, \sigma_s; \Delta t)$ is given by the convolution of $g(\cdot)$ and the basis function $\varphi_l(\cdot)$ as follows

$$\begin{aligned}\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t) &\equiv \frac{1}{\Delta w}\int_{w_{\min}^\dagger}^{w_{\max}^\dagger} \varphi_l(w)\ g(w_n - w, \sigma_k, \sigma_s; \Delta t)\ dw\\ &= \frac{1}{\Delta w}\int_{w_n - w_l - \Delta w}^{w_n - w_l + \Delta w} \varphi_{n-l}\,(w)\, g(w, \sigma_k, \sigma_s; \Delta t)\ dw, \qquad (5.16)\end{aligned}$$

using a suitable change of variable and the property $\varphi_l(w_n - w) = \varphi_{n-l}(w)$.

To summarize, approximations (5.11) and (5.15) for the outer and inner integrals, respectively, gives a scheme for the double integral (5.1) as follows

$$u_{n,j,k}^{m-1} = \sum_{s\in\mathbb{K}} \vartheta_s\ \varrho(\sigma_k, \sigma_s; \Delta t)\ \tilde{u}_{n,j,k,s}^{m-1}, \quad (n, j, k) \in \mathbb{N}\times\mathbb{J}\times\mathbb{K},\ t_m \in \mathcal{T}, \qquad (5.17)$$

$$\text{where } \tilde{u}_{n,j,k,s}^{m-1} = \Delta w \sum_{l\in\mathbb{N}^\dagger} \tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t)\ u_{l,j,s}^{m-}, \qquad (5.18)$$

with $u_{l,j,k}^{m-}$ are given in (5.12), and $\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t)$ is given by (5.16).

5.2.2 Ω_{in}: approximation of $\tilde{g}_{n-l}(\cdot)$

We note that, in (5.18), the exact weight $\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t)$, defined in (5.16), is not known in closed-form because an explicit formula for the function g is not known, and $\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t)$ needs to be approximated. To this end, following ideas in our papers [24, 38, 39], we replace the function $g(w, \cdot)$ in (5.16) by its localized, periodic approximation $\hat{g}(w, \cdot)$, where

$$\hat{g}(w, \sigma_k, \sigma_s; \Delta t) = \frac{1}{P^\dagger}\sum_{q=-\infty}^{\infty} e^{2\pi i \eta_q w} G(\eta_q, \sigma_k, \sigma_s; \Delta t) \quad \text{with} \quad \eta_q = \frac{q}{P^\dagger},\ P^\dagger = w_{\max}^\dagger - w_{\min}^\dagger. \qquad (5.19)$$

That is, $\tilde{g}_{n-l}(\cdot)$, defined in (5.16), is approximated by

$$\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t) \simeq \frac{1}{\Delta w}\int_{w_n - w_l - \Delta w}^{w_n - w_l + \Delta w} \varphi_{n-l}\,(w)\, \hat{g}(w, \sigma_k, \sigma_s; \Delta t)\ dw, \qquad (5.20)$$

where $\hat{g}(\cdot)$ is given in (5.19). We then integrate the resulting finite integral (5.20) to obtain an approximation $\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t, \infty)$ to $\tilde{g}_{n-l}(\cdot)$ in the form of the infinite series as follows

$$\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t, \infty) = \frac{1}{P^\dagger}\sum_{q=-\infty}^{\infty} e^{2\pi i \eta_q (n-l)\Delta w} \frac{\sin^2 \pi\eta_q\Delta w}{(\pi\eta_q\Delta w)^2}\ G(\eta_q, \sigma_k, \sigma_s; \Delta t). \qquad (5.21)$$

For computational purposes, the infinite series (5.21) must be truncated. For this truncation, we do not need use the same number of terms as the number of linear basis functions, which is $N^\dagger$. Specifically, for fixed (σ_k, σ_s), we truncate the infinite series for $\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t, \infty)$ to a total of $(\alpha_{k,s}N^\dagger)$ terms, where $\alpha_{k,s} \in \{2, 4, 8, \ldots\}$ is the refinement parameter. For a fixed $\alpha_{k,s} \in \{2, 4, 8, \ldots\}$, the result of this truncation is

$$\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t, \alpha) = \frac{1}{P^\dagger}\sum_{q=-\alpha_{k,s}N^\dagger/2}^{\alpha_{k,s}N^\dagger/2-1} e^{2\pi i \eta_q (n-l)\Delta w} \frac{\sin^2 \pi\eta_q\Delta w}{(\pi\eta_q\Delta w)^2}\ G(\eta_q, \sigma_k, \sigma_s; \Delta t). \qquad (5.22)$$

Note that, in its exact form (5.16), the weight $\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t)$ is strictly positive, since $g(\cdot) \geq 0$ and $\varphi_{n-l}(w) \geq 0$. However, its approximation, given by the discrete convolution (5.22), may not have this property, and hence the strict monotonicity of the timestepping method is not guaranteed. To this end, given an $\epsilon > 0$, we enforce an ϵ-monotonicity on the approximate weight $\tilde{g}(\cdot)$ in (5.22) by the condition

$$\Delta w \sum_{l \in \mathbb{N}^\dagger} \left| \min\left(\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t, \alpha_{k,s}), 0\right) \right| \;<\; \epsilon \frac{\Delta t}{T}. \tag{5.23}$$

Furthermore, in determining a suitable value for the refinement parameter $\alpha_{k,s}$ to be used in (5.22), given $\epsilon_\alpha > 0$, we employ the stopping criterion

$$\Delta w \sum_{l \in \mathbb{N}^\dagger} \left| \tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t, \alpha_{k,s}) - \tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t, \alpha_{k,s}/2) \right| \;<\; \epsilon_\alpha. \tag{5.24}$$

Remark 5.3 (Notational convention for $\alpha_{k,s}$). *We emphasize that the refinement parameter is dependent on (σ_k, σ_s), $k, s \in \mathbb{K}$. That is, different (σ_k, σ_s) may require different number of terms in the truncated series. For brevity, hereafter, we write α instead of $\alpha_{k,s}$, with the dependence on (σ_k, σ_s) implicitly understood.*

5.2.3 Convergence of truncated Fourier series

We now investigate the Fourier truncation error arising from using a finite $\alpha \in \{2, 4, 8, \ldots\}$ in $\tilde{g}_{n-l}(\cdot)$. In order to focus on the role of α, in this subsection, we will write $\tilde{g}_{n-l}(\alpha)$. We also write $G(\eta)$ instead of $G(\eta; \cdot)$. Based on the theoretical result in Lemma 3.1, we assume that $v + 1 > 0$. We also let $C_1 > 0$ and $C_2 > 0$ be generic finite constants independently of h, which may take different value from line to line. We have $|\tilde{g}_{n-l}(\alpha) - \tilde{g}_{n-l}(\infty)| = \ldots$

$$\ldots = \left| \frac{1}{P^\dagger} \sum_{q=\alpha N^\dagger/2}^{\infty} e^{2\pi i \eta_q (n-l)\Delta w} \left(\frac{\sin^2 \pi \eta_q \Delta w}{(\pi \eta_q \Delta w)^2} \right) G(\eta_q) + \frac{1}{P^\dagger} \sum_{q=-\infty}^{-\alpha N^\dagger/2-1} e^{2\pi i \eta_q (n-l)\Delta w} \left(\frac{\sin^2 \pi \eta_q \Delta w}{(\pi \eta_q \Delta w)^2} \right) G(\eta_q) \right|$$

$$\leq \frac{2}{P^\dagger} \sum_{q=\alpha N^\dagger/2}^{\infty} \frac{1}{(\pi \eta_q \Delta w)^2} |G(\eta_q)| \overset{\text{(i)}}{\leq} \frac{2}{P^\dagger} \frac{4}{\pi^2 \alpha^2} \sum_{q=\alpha N^\dagger/2}^{\infty} |G(\eta_q)| \overset{\text{(ii)}}{\leq} \frac{8}{P^\dagger \pi^2 \alpha^2} \sum_{q=\alpha N^\dagger/2}^{\infty} C_1 \left(\eta_q e^{-\frac{1}{2} C_2 \eta_q \Delta t} \right)^{v+1}$$

$$\leq \frac{C_1}{\alpha^2} \sum_{q=\alpha N^\dagger/2}^{\infty} \frac{(\eta_q)^{v+1}}{P^\dagger} e^{-C_2(v+1)\eta_q}. \tag{5.25}$$

Here, (i) is due to $\frac{1}{(\pi \eta_q \Delta w)^2} \leq \frac{4}{\pi^2 \alpha^2}$, since $\eta_q = \frac{k}{P^\dagger}$, $\Delta w = \frac{P^\dagger}{N^\dagger}$, and $q \geq \alpha N^\dagger/2$; (ii) is due to Lemma 3.1. We consider the function $d(x) = x^{v+1} e^{-C_2(v+1)x}$, which is positive and monotonically decreasing for $x > 0$, noting $v + 1 > 0$. Therefore, using the usual Integral Test, we can bound the sum in (5.25) as follows

$$\sum_{q=\alpha N^\dagger/2}^{\infty} \frac{(\eta_q)^{v+1}}{P^\dagger} e^{-C_2(v+1)\eta_q} \leq \int_{\alpha/2h}^{\infty} x^{v+1} e^{-C_2(v+1)x} dx. \tag{5.26}$$

Recall the definition of the Incomplete Upper Gamma function $\Gamma(y, x)$ [1]: $\Gamma(y, x) = \int_x^\infty t^{y-1} e^{-t} dt$, $\mathrm{Re}(y) > 0$, and the result below from calculus

$$\frac{d}{dx} \Gamma(y + 1, cx) = \frac{d}{dx} \int_{cx}^{\infty} t^y e^{-t} dt = -c^{y+1} x^y e^{-cx}, \quad x > 0. \tag{5.27}$$

Using (5.27), the integral in (5.26) becomes $\int_{\alpha/2h}^{\infty} x^{v+1} e^{-C_2(v+1)x} dx = \ldots$

$$\ldots = -(C_2(v+1))^{-(v+1)-1} \Gamma(v + 2, C_2(v+1)x) \Big|_{\alpha/2h}^{\infty} = (C_2(v+1))^{-(v+1)-1} \Gamma\left(v + 2, \frac{\alpha C_2(v+1)}{2h}\right).$$

Therefore,

$$|\tilde{g}_{n-l}(\alpha) - \tilde{g}_{n-l}(\infty)| \leq (C_2(v+1))^{-(v+1)-1} \Gamma\left(v + 2, \frac{\alpha C_2(v+1)}{2h}\right). \tag{5.28}$$

Consider the monotonicity test in (5.23), from (5.28), noting $\tilde{g}_{n-l}(\infty) \geq 0$, and $\sum_{l \in \mathbb{N}^\dagger} \Delta w = P^\dagger = C/h$, we have

$$\Delta w \sum_{l \in \mathbb{N}^\dagger} \left| \min\left(\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t, \alpha), 0\right) \right| \leq \frac{C_1}{h} \Gamma\left(v + 2, \frac{\alpha C_2(v+1)}{2h}\right).$$

As such, for a given monotonicity tolerance $\epsilon > 0$, the monotonicity test in (5.23) is satisfied for sufficiently large α. However, how quickly this monotonicity test is satisfied depends on how positive/negative $v + 1$ is and how small the discretization parameter h is. Therefore, in a practical situation, the cost for achieving weak monotonicity in this case is higher than that in the case of jump-diffusion with a constant instantaneous variance (with and without stochastic interest rate) as demonstrated in [38, 39].

5.2.4 Efficient implementation via FFT and algorithms

We now discuss an efficient implementation of the scheme via FFT. Given a fixed $\alpha \in \{2, 4, 8, \ldots\}$, the sequence $\{\tilde{g}_{-N^\dagger/2}(\sigma_k, \sigma_s; \Delta t, \alpha), \ldots, \tilde{g}_{N^\dagger/2-1}(\sigma_k, \sigma_s; \Delta t, \alpha)\}$ is $N^\dagger$-periodic for fixed $s \in \mathbb{K}$ and $k \in \mathbb{K}$. With this in mind, we let $q = n - l$ in the discrete convolution (5.22), and, for a fixed α, the set of approximate weights in the physical domain to be determined is $\tilde{g}_q(\alpha)$, $q = -N^\dagger/2, \ldots, N^\dagger/2-1$. Using this notation, in (5.22), with $q = n - l$, we rewrite $e^{2\pi i \eta_k (n-l)\Delta w} = e^{2\pi i k \alpha q/(\alpha N^\dagger)}$, and obtain for $s \in \mathbb{K}$, $k \in \mathbb{K}$ fixed,

$$\begin{aligned} \tilde{g}_q(\sigma_k, \sigma_s; \Delta t, \alpha) &= \frac{1}{P^\dagger} \sum_{\jmath=-\alpha N^\dagger/2}^{\alpha N^\dagger/2-1} e^{2\pi i \jmath(\alpha q)/(\alpha N^\dagger)} \; y_\jmath, \qquad\qquad q = -N^\dagger/2, \ldots, N^\dagger/2 - 1, \\ \text{where } y_\jmath &= \left(\frac{\sin^2 \pi \eta_\jmath \Delta w}{(\pi \eta_\jmath \Delta w)^2} \right) G(\eta_\jmath, \sigma_k, \sigma_s; \Delta t), \quad \jmath = -\frac{\alpha N^\dagger}{2}, \ldots, \frac{\alpha N^\dagger}{2} - 1. \end{aligned} \tag{5.29}$$

It is observed from (5.29) that, given $\{y_\jmath\}$, $\{\tilde{g}_q(\sigma_k, \sigma_s; \Delta t, \alpha)\}$ can be computed efficiently via a single FFT of size $\alpha N^\dagger$. A suitable value for α, i.e. satisfying the ϵ-monotonicity condition (5.23), can be determined through an iterative procedure based on formula (5.29). Let this value be α_ϵ. We also observe that once α_ϵ is found, the discrete convolutions (5.18) can also be computed efficiently using an FFT. This suggests that we only need to compute the weights in the Fourier domain, i.e. the DFT of $\{\tilde{g}_q(\sigma_k, \sigma_s; \Delta t, \alpha_\epsilon)\}$, only once, and reuse them for all timesteps. We define $\{\tilde{G}(\eta_q, \sigma_k, \sigma_s; \Delta t, \alpha_\epsilon)\}$ to be the DFT of $\{\tilde{g}_q(\alpha_\epsilon)\}$ given by

$$\tilde{G}(\eta_\jmath, \sigma_k, \sigma_s; \Delta t, \alpha_\epsilon) = \frac{P^\dagger}{N^\dagger} \sum_{q=-N^\dagger/2}^{N^\dagger/2-1} e^{-2\pi i q \jmath/N^\dagger} \; \tilde{g}_q(\sigma_k, \sigma_s; \Delta t, \alpha_\epsilon), \quad \jmath \in \mathbb{N}^\dagger. \tag{5.30}$$

An iterative procedure for computing $\{\tilde{G}(\eta_q, \sigma_k, \sigma_s; \Delta t, \alpha_\epsilon)\}$ for each pair $(k, s) \in \mathbb{K} \times \mathbb{K}$ is given in Algorithm 5.1, where we also use the stopping criterion introduced in (5.24).

Algorithm 5.1 Computation of weights $\tilde{G}(\eta_q, \sigma_k, \sigma_s; \Delta t, \alpha_\epsilon)$, $q \in \mathbb{N}^\dagger$, in Fourier domain, for each pair $(k, s) \in \mathbb{K} \times \mathbb{K}$.

```
set α = 1 and compute g̃_q(σ_k, σ_s; Δt, α), q ∈ N† using (5.29);
for α = 2, 4, ... until convergence do
    compute g̃_q(σ_k, σ_s; Δt, α) q ∈ N†, using (5.29);
    compute test_1 = Δw Σ_{q∈N†} min(g̃_q(σ_k, σ_s; Δt, α), 0) for monotonicity test;
    compute test_2 = Δw Σ_{q∈N†} |g̃_q(σ_k, σ_s; Δt, α) − g̃_q(σ_k, σ_s; Δt, α/2)| for accuracy test;
    if |test_1| < ε(Δt/T) and test_2 < ε_α then
        α_ε = α;
        break from for loop;
    end if
end for
use (5.30) to compute and output weights G̃_q(η_q, σ_k, σ_s; Δt, α_ε), q ∈ N†, in Fourier domain.
```

Remark 5.4. *For simplicity, unless otherwise stated, we adopt the following notional convention: for fixed* $(k, s) \in \mathbb{K} \times \mathbb{K}$, $\tilde{g}_{n-l}(\sigma_k, \sigma_s) = \tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t, \alpha_\epsilon)$ *and* $\tilde{G}(\eta_q, \sigma_k, \sigma_s; \Delta t) \equiv \tilde{G}(\eta_q, \sigma_k, \sigma_s; \Delta t, \alpha_\epsilon)$, *where* α_ϵ *is selected by Algorithm 5.1. That is,* α_ϵ *satisfies the* ϵ*-monotonicity condition* (5.23). *That is,* $\Delta w \sum_{l \in \mathbb{N}^\dagger} \left| \min\left(\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t, \alpha), 0\right) \right| < \epsilon \frac{\Delta t}{T}$, $\epsilon > 0$, *for all* $n \in \mathbb{N}$ *and for* $s \in \mathbb{K}$, $k \in \mathbb{K}$ *fixed.*

The intermediate result (inner integral) $\tilde{u}^{m-1}_{n,j,k,s}$ defined (5.18) can then be implemented efficiently via an FFT as follows

$$\begin{aligned} \tilde{u}^{m-1}_{n,j,k,s} &\simeq \sum_{q \in \mathbb{N}^\dagger} e^{2\pi i q n/N^\dagger} \; \tilde{U}(\eta_q, a_j, \sigma_s, t_m^-) \; \tilde{G}(\eta_q, \sigma_k, \sigma_s; \Delta t), \\ \text{with } \tilde{U}\left(\eta_q, a_j, \sigma_s, t_m^-\right) &= \frac{1}{N^\dagger} \sum_{l \in \mathbb{N}^\dagger} e^{-2\pi i q l/N^\dagger} \; u^{m-}_{l,j,s}, \; q \in \mathbb{N}^\dagger, \end{aligned} \tag{5.31}$$

where $\tilde{G}(\eta_q, \sigma_k, \sigma_s; \Delta t)$ is given by (5.30). Putting everything together, an ϵ-monotone algorithm for the computational domain Ω is presented in Algorithm 5.2.

Algorithm 5.2 An ϵ-monotone Fourier algorithm for the GLWB pricing problem; $x \circ y$ is the Hadamard product of vectors x and y.

1: compute vector $\tilde{G}_{k,s} = \left[\tilde{G}(\eta_q, \sigma_k, \sigma_s; \Delta t, \alpha)\right]_{q \in \mathbb{N}^\dagger}$ using Algorithm 5.1, $k \in \mathbb{K}$, $s \in \mathbb{K}$;

2: initialize $u_{n,j,k}^{M-} = p_{M-1}(e^{w_n} \wedge e^{w_\infty})$, $n \in \mathbb{N}^\dagger$, $j \in \mathbb{J}$, $k \in \mathbb{K}$;

3: **for** $m = M, \ldots, 1$ **do**

4:     **if** $m < M$ **then**

5:         solve the optimization problem (5.10) to obtain $u_{l,j,s}^{m-}$, $l \in \mathbb{N}$, $j \in \mathbb{J}$, $s \in \mathbb{K}$;

6:         combine results in Line (5) with $u_{l,j,s}^{m}$ corresponding to $\Omega_{w_{\min}}$ and $\Omega_{w_{\max}}$ to obtain vectors $u_{j,s}^{m-} = \left[u_{l,j,s}^{m-}\right]_{l \in \mathbb{N}^\dagger}$, $j \in \mathbb{J}$, $s \in \mathbb{K}$, as per (5.12);

7:     **end if**

8:     compute vectors of intermediate values

$$\left[\tilde{u}_{n,j,k,s}^{m-1}\right]_{n \in \mathbb{N}^\dagger} \quad = \quad \text{IFFT}\left\{\text{FFT}\left\{u_{j,s}^{m-}\right\} \circ \tilde{G}_{k,s}\right\}, \quad j \in \mathbb{J}, \; k \in \mathbb{K}, \; s \in \mathbb{K};$$

9:     discard FFT values in padding areas $\Omega_{w_{\min}}$ and $\Omega_{w_{\max}}$, i.e. discard for each fixed $(j,k,s) \in \mathbb{J} \times \mathbb{K} \times \mathbb{K}$, $\tilde{u}_{n,j,k,s}^{m-1}$ with $n \in \mathbb{N}^\dagger \setminus \mathbb{N}$;

10:     compute $u_{n,j,k}^{m-1} = \sum_{s \in \mathbb{K}} \vartheta_s \; \varrho(\sigma_k, \sigma_s; \Delta t) \; \tilde{u}_{n,j,k,s}^{m-1}$, $(n,j,k) \in \mathbb{N} \times \mathbb{J} \times \mathbb{K}$, as per (5.17); Ω_{in}

11:     compute $u_{n,j,k}^{m-1}$, $n = -N^\dagger/2, \ldots, -N/2$, $(j,k) \in \mathbb{J} \times \mathbb{K}$ using (5.7); $\Omega_{w_{\min}}$

12:     compute $u_{n,j,k}^{m-1}$, $n = N/2, \ldots, N^\dagger/2 - 1$, $(j,k) \in \mathbb{J} \times \mathbb{K}$ using (5.8); $\Omega_{w_{\max}}$

13: **end for**

Remark 5.5 (Fair insurance fees). *With respect to the insurance fee β, let $u(\beta; w, a, \sigma, t)$ be the exact solution, i.e. $u(w, a, \sigma, t)$, be parameterised by the insurance fee β. Then, the fair insurance fee for $t = 0$, denoted by β_f, solves the equation $u\left(\beta_f; \ln(z_0), z_0, \ln(\nu_0), 0\right) \;=\; z_0$. In a numerical setting, with a slight abuse of notation, let $u^0_{\ln(z_0), z_0, \ln(\nu_0)}(\beta)$ be the numerical solution parameterized by β. Then Newton iteration can be employed to solve for β_f from the non-linear equation $u^0_{\ln(z_0), z_0, \ln(\nu_0)}(\beta_f) = z_0$, where $u^0_{\ln(z_0), z_0, \ln(\nu_0)}$ is obtained by Algorithm 5.2.*

6 Pointwise convergence

In this section, we establish pointwise convergence of the proposed numerical integration method. We start by verifying three properties: ℓ_∞-stability, ϵ-monotonicity (as opposed to monotonicity), and consistency (with respect to the double integral formulation (4.6)). We will then show that convergence of our scheme is ensured if the monotonicity tolerance $\epsilon \to 0$ as $h \to 0$.

For subsequent use, we present relevant properties of the weights $\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t)$, where σ_s and σ_k are fixed. in the scheme (5.17).

Lemma 6.1. *Suppose the function $\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t)$ is given by (5.22), we have, $n \in \mathbb{N}$, $s \in \mathbb{K}$, $k \in \mathbb{K}$,*

$$\Delta w \sum_{l \in \mathbb{N}^\dagger} \tilde{g}_{n-l} \;=\; e^{-r\Delta t}, \quad \Delta w \sum_{l \in \mathbb{N}^\dagger} \left(\max(\tilde{g}_{n-l}, 0) + \left|\min\left(\tilde{g}_{n-l}, 0\right)\right|\right) \le 1 + 2\epsilon \frac{\Delta \tau}{T} \le e^{2\epsilon \frac{\Delta \tau}{T}}. \tag{6.1}$$

Proof of Lemma 6.1. Letting $\ell = n - l$, we have $\Delta w \sum_{l \in \mathbb{N}^\dagger} \tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t) = \ldots$

$$
\begin{aligned}
\ldots &= \frac{P^\dagger}{N^\dagger}\sum_{\ell\in\mathbb{N}^\dagger}\tilde{g}_\ell(\sigma_k,\sigma_s;\Delta t) \overset{\text{(i)}}{=} \frac{P^\dagger}{N^\dagger}\sum_{\ell\in\mathbb{N}^\dagger}\frac{1}{P^\dagger}\sum_{q=-\alpha N^\dagger/2}^{\alpha N^\dagger/2-1} e^{2\pi i\eta_q\ell\Delta w}\left(\frac{\sin^2\pi\eta_q\Delta w}{(\pi\eta_q\Delta w)^2}\right) G(\eta_q,\sigma_k,\sigma_s;\Delta t)\\
&= \frac{1}{N^\dagger}\sum_{q=-\alpha N^\dagger/2}^{\alpha N^\dagger/2-1}\left(\frac{\sin^2\pi\eta_q\Delta w}{(\pi\eta_q\Delta w)^2}\right) G(\eta_q,\sigma_k,\sigma_s;\Delta t)\sum_{\ell\in\mathbb{N}^\dagger}\exp\left(\frac{2\pi i\ell q}{N^\dagger}\right)\\
&\overset{\text{(ii)}}{=} G(0,\sigma_k,\sigma_s;\Delta t) \overset{\text{(iii)}}{=} e^{-r\Delta t}.
\end{aligned}
$$

Here, in (i), we use the definition of (5.22), in (ii), we apply the properties of $N^\dagger$th roots of unity; in (iii), we use the closed-form expression of $G(\cdot)$ in (3.7). Finally, in (6.1), the second result follows from the first, noting $\tilde{g}_{n-l} = \max(\tilde{g}_{n-l},0) + \min(\tilde{g}_{n-l},0)$, and $e^{-r\Delta\tau} \leq 1$, together with the monotonicity condition (5.23). □

Remark 6.1. *Since $\varrho(\sigma,\sigma';\Delta t)$, given in* (3.9)*, is a (conditional) probability density function, for a fixed $\sigma_k \in [\sigma_{\min},\sigma_{\max}]$, we have $\int_{\mathbb{R}}\varrho(\sigma_k,\sigma;\Delta t)\ d\sigma = 1$, hence, $\int_{\sigma_{\min}}^{\sigma_{\max}}\varrho(\sigma_k,\sigma;\Delta t)\ d\sigma \leq 1$. Applying Gauss-Legendre quadrature rule to approximate $\int_{\sigma_{\min}}^{\sigma_{\max}}\varrho(\sigma_k,\sigma;\Delta t)\ d\sigma$ gives rise to an approximation error ϵ_p defined as follows $\epsilon_p := \left|\sum_{s\in\mathbb{K}}\vartheta_s\ \varrho(\sigma_k,\sigma_s;\Delta t) - \int_{\sigma_{\min}}^{\sigma_{\max}}\varrho(\sigma_k,\sigma;\Delta t)\ d\sigma\right|$. It is straightforward to see that $\epsilon_p \to 0$ as $K\to\infty$, i.e. as $h\to 0$. Using the above results, recalling the weights ϑ_s, $s\in\mathbb{K}$, are positive, we have*

$$
0 \ \leq\ \sum_{s\in\mathbb{K}}\vartheta_s\varrho(\sigma_k,\sigma_s;\Delta t)\ \leq\ 1+\epsilon_p < e^{\epsilon_p}. \tag{6.2}
$$

6.1 Stability

Our scheme consists of the following equations: (5.6) for $\Omega\times\{T\}$, (5.7) for $\Omega_{w_{\min}}$, (5.8) for $\Omega_{w_{\max}}$, and finally (5.17) for Ω_{in}. We start by verifying ℓ_∞-stability of our scheme.

Lemma 6.2 (ℓ_∞-stability)**.** *Suppose the discretization parameter h satisfies* (5.5)*. If linear interpolation is used for the intervention action* (5.10)*, then the scheme* (5.6), (5.7), (5.8)*, and* (5.17) *satisfies $\sup_{h>0}\|u^m\|_\infty < \infty$ for all $m = M,\ldots,0$, as the discretization parameter $h\to 0$. Here, we have $\|u^m\|_\infty = \max_{n,j,k}|u^m_{n,j,k}|$, $(n,j,k)\in\mathbb{N}^\dagger\times\mathbb{J}\times\mathbb{K}$.*

Proof of Lemma 6.2. First, we note that, for any fixed $h>0$, as given by (5.6), we have $\left\|u^{M-}\right\|_\infty < \infty$. Therefore, we have $\sup_{h>0}\left\|u^{M-}\right\|_\infty < \infty$. Motivated by this observation, to demonstrate ℓ_∞-stability of our scheme, we will show that, for a fixed $h>0$, at any (w_n,a_j,σ_k,t_m), $m=M,\ldots,0$, we have

$$
|u^m_{n,j,k}| < C\left(\left\|u^{M-}\right\|_\infty + e^{w_{\max}}\right),\qquad C>0 \text{ bounded above independently of } h. \tag{6.3}
$$

In our proof techniques, the constant $C>0$ is typically of the form

$$
(M-m+1)e^{2(M-m)\epsilon\frac{\Delta t}{T}}e^{(M-m)\epsilon_p},\qquad m=M,\ldots,0,
$$

where ϵ is the monotonicity tolerance used in (5.23) with $0<\epsilon\ll 1$, and ϵ_p is the error of the quadrature rule discussed in Remark (6.1). Since $m\Delta\tau\leq T$ and $\epsilon_p\ll 1$, C is bounded above by $e^{2(M+1)}$, where M is a fixed positive constant.

For the rest of the proof, we will show the key inequality (6.3) when $h>0$ is fixed. For clarity, we will address stability for the boundary and interior sub-domains (together with their respective initial conditions) separately, starting with the boundary sub-domains.

It is straightforward to show that (5.6) is ℓ_∞-stable, since $\max_{n,j,k}|u^{M-}_{n,j,k}|$ obviously satisfies (6.3) for $(n,j,k)\in\mathbb{N}^\dagger\times\mathbb{J}\times\mathbb{K}$. Next, recall that $a_j\in[0,e^{w_{\max}}]$, $j\in\mathbb{J}$, hence (5.7) is also ℓ_∞-stable since $|u^m_{n,j,k}|$, $n=-N^\dagger/2,\ldots,-N/2$, $(j,k)\in\mathbb{J}\times\mathbb{K}$, also satisfies (6.3). It is also straightforward to show that (5.8) is also ℓ_∞-stable since $\max_{n,j,k}|u^m_{n,j,k}|$ $n=N/2,\ldots,N^\dagger/2-1$ and $(j,k)\in\mathbb{J}\times\mathbb{K}$, also satisfies (6.3).

Now we focus on the main task which is to demonstrate ℓ_∞-stability for (5.17) (for Ω_{in}). Recall the monotonicity tolerance ϵ used in (5.23), where $0 < \epsilon \ll 1$, and the error ϵ_p is given in Remark (6.1). We typically use $\epsilon \le 1/2$ in the proof below.

To prove ℓ_∞-stability for (5.17) (for Ω_{in}), we show that, for $m = M, \ldots, 0$, we have

$$\|u^m\|_\infty \le (M - m + 1) e^{2(M-m)\epsilon\frac{\Delta t}{T}} e^{(M-m)\epsilon_p} \left(\left\|u^{M-}\right\|_\infty + e^{w_{\max}}\right), \tag{6.4}$$

which is bounded above by $e^{2(M+1)} \left(\left\|u^{M-}\right\|_\infty + e^{w_{\max}}\right)$ independently of h, since $m\Delta\tau \le T$, and $\epsilon_p \ll 1$.

For subsequent use, with $(n, j, k) \in \mathbb{N} \times \mathbb{J} \times \mathbb{K}$, and $m = M, \ldots, 1$, we define the measures

$$\left\|u_{j,k}^{m-}\right\|_\infty = \max_n \left|u_{n,j,k}^{m-}\right| \quad \text{and} \quad \left\|u_{j,k}^{m}\right\|_\infty = \max_n \left|u_{n,j,k}^{m}\right|,$$

$$\left[u_{j,k}^{m-}\right]_{\max} = \max_n \left\{u_{n,j,k}^{m-}\right\}, \ \left[u_{j,k}^{m}\right]_{\max} = \max_n \left\{u_{n,j,k}^{m}\right\}, \ \left[u_{j,k}^{m-}\right]_{\min} = \min_n \left\{u_{n,j,k}^{m-}\right\}, \ \left[u_{j,k}^{m}\right]_{\min} = \min_n \left\{u_{n,j,k}^{m}\right\}.$$

To show (6.4), using induction on m, $m = M, \ldots, 0$, we will show that for all $(j, k) \in \mathbb{J} \times \mathbb{K}$,

$$\left[u_{j,k}^{m}\right]_{\max} \le (M - m + 1) e^{2(M-m)\epsilon\frac{\Delta t}{T}} e^{(M-m)\epsilon_p} \left(\left\|u^{M-}\right\|_\infty + e^{w_{\max}}\right), \tag{6.5}$$

$$\left[u_{j,k}^{m}\right]_{\min} \ge -(M - m + 1) e^{2(M-m)\epsilon\frac{\Delta\tau}{T}} e^{(M-m)\epsilon_p} \left(\left\|u^{M-}\right\|_\infty + e^{w_{\max}}\right). \tag{6.6}$$

<u>Base case:</u> when $m = M$, and hence $M - m = 0$, it follows that (6.5) and (6.6) become

$$\left[u_{j,k}^{M}\right]_{\max} \le \left(\left\|u^{M-}\right\|_\infty + e^{w_{\max}}\right), \quad \left[u_{j,k}^{M}\right]_{\min} \ge 0,$$

which hold for all $(n, j, k) \in \mathbb{N} \times \mathbb{J} \times \mathbb{K}$, due to from terminal condition (5.6).

<u>Hypothesis:</u> we assume that (6.5)-(6.6) hold for $m = \hat{m}$, where $1 \le \hat{m} \le M$ and $(j, k) \in \mathbb{J} \times \mathbb{K}$.

<u>Induction:</u> we show that (6.5)-(6.6) also hold for $m = \hat{m} - 1$ and $(j, k) \in \mathbb{J} \times \mathbb{K}$. This is done in two steps as follows. In Step 1, we show, for $(j, k) \in \mathbb{J} \times \mathbb{K}$, $\left[u_{j,k}^{\hat{m}-}\right]_{\max}$ and $\left[u_{j,k}^{\hat{m}-}\right]_{\min}$ satisfy

$$\left[u_{j,k}^{\hat{m}-}\right]_{\max} \le (M - (\hat{m} - 1) + 1) e^{2(M-\hat{m})\epsilon\frac{\Delta\tau}{T}} e^{(M-\hat{m})\epsilon_p} \left(\left\|u^{M-}\right\|_\infty + e^{w_{\max}}\right), \tag{6.7}$$

$$\left[u_{j,k}^{\hat{m}-}\right]_{\min} \ge -(M - (\hat{m} - 1) + 1) e^{2(M-\hat{m})\epsilon\frac{\Delta\tau}{T}} e^{(M-\hat{m})\epsilon_p} \left(\left\|u^{M-}\right\|_\infty + e^{w_{\max}}\right). \tag{6.8}$$

In Step 2, we bound the timestepping result (5.17) at $m = \hat{m} - 1$ using (6.7)-(6.8).

<u>Step 1 - Bound (6.7)-(6.8) for $u_{n,j,k}^{\hat{m}-}$:</u> As noted in Remark 5.2, for the intervention result $u_{n,j,k}^{\hat{m}-}$, the supremum of (5.10) is achieved by an optimal control $\gamma^* \in [0, e^{w_n} \vee C_r a_j]$. Therefore, for $m = \hat{m}$, $u_{n,j,k}^{\hat{m}-}$ can be written as

$$u_{n,j,k}^{\hat{m}-} = \mathcal{I}\left\{u_k^{\hat{m}}\right\}(\hat{w}_n^*, \hat{a}_j^*) + p_{\hat{m}} f(\gamma^*; a_j) + p_{\hat{m}-1} q_{\hat{m}-1} e^{w_n}, \quad \gamma^* \in [0, e^{w_n} \vee C_r a_j], \tag{6.9}$$

where $(\hat{w}_n^*, \hat{a}_j^*)$ are given by (3.3) with $\gamma = \gamma^*$. We assume that $\hat{w}_n^* \in [e^{w_{n'}}, e^{w_{n'+1}}]$ and $\hat{a}_j^* \in [a_{j'}, a_{j'+1}]$, and nodes that are used for linear interpolation are $(x_{n',j',k}^{\hat{m}}, \ldots, x_{n'+1,j'+1,k}^{\hat{m}})$, where $x = (w, a, \sigma)$. We note that these nodes could be outside Ω_{in}, i.e. in $\Omega_{w_{\min}}$. However, as previously demonstrated, the numerical solutions at these nodes satisfy the same bounds (6.5)-(6.6). Computing $u_{n,j,k}^{\hat{m}-}$ using linear interpolation results in

$$u_{n,j,k}^{\hat{m}-} = x_a \left(x_w \ u_{n',j',k}^{\hat{m}} + (1 - x_w) \ u_{n'+1,j',k}^{\hat{m}}\right) + (1 - x_a) \left(x_w \ u_{n',j'+1,k}^{\hat{m}} + (1 - x_w) \ u_{n'+1,j'+1,k}^{\hat{m}}\right), \tag{6.10}$$

where $0 \le x_a \le 1$ and $0 \le x_w \le 1$ are interpolation weights. By the induction hypothesis for (6.5), we have the following bound for nodal values used in (6.10):

$$\left\{u_{n',j',k}^{\hat{m}}, \ldots \ u_{n'+1,j'+1,k}^{\hat{m}}\right\} \le (M - \hat{m} + 1) e^{2(M-\hat{m})\epsilon\frac{\Delta t}{T}} e^{(M-\hat{m})\epsilon_p} \left(\left\|u^{M-}\right\|_\infty + e^{w_{\max}}\right). \tag{6.11}$$

Taking into account the non-negative weights in linear interpolation, and upper bounds in (6.11), the interpolated result $\mathcal{I}\left\{u_k^{\hat{m}}\right\}\left(\hat{w}_n^*, \hat{a}_j^*\right)$ in (6.9) is bounded by

$$\mathcal{I}\left\{u_k^{\hat{m}}\right\}(\hat{w}_n^*, \hat{a}_j^*) \le (M - \hat{m} + 1) e^{2(M-\hat{m})\epsilon\frac{\Delta t}{T}} e^{(M-\hat{m})\epsilon_p} \left(\left\|u^{M-}\right\|_\infty + e^{w_{\max}}\right). \tag{6.12}$$

Hence, using (6.12), (6.9) becomes

$$\begin{aligned}
u_{n,j,k}^{\hat{m}-} &\overset{\text{(i)}}{\le} (M - \hat{m} + 1) e^{2(M-\hat{m})\epsilon\frac{\Delta t}{T}} e^{(M-\hat{m})\epsilon_p} \left(\left\|u^{M-}\right\|_\infty + e^{w_{\max}}\right) + p_{\hat{m}} e^{w_{\max}} + p_{\hat{m}-1} q_{\hat{m}-1} e^{w_{\max}} \\
&\overset{\text{(ii)}}{\le} (M - \hat{m} + 1) e^{2(M-\hat{m})\epsilon\frac{\Delta t}{T}} e^{(M-\hat{m})\epsilon_p} (\|u^{M-}\|_\infty + e^{w_{\max}}) + p_{\hat{m}-1} e^{w_{\max}} \\
&\overset{\text{(iii)}}{\le} (M - (\hat{m} - 1) + 1) e^{2(M-(\hat{m})\epsilon\frac{\Delta t}{T}} e^{(M-\hat{m})\epsilon_p} (\|u^{M-}\|_\infty + e^{w_{\max}}).
\end{aligned} \tag{6.13}$$

Here, in (i), by (3.2), $f(\gamma^*; a_j) \le \gamma^* \le e^{w_n} < e^{w_{\max}}$. In (ii), we use $p_{\hat{m}} + p_{\hat{m}-1} q_{\hat{m}-1} = p_{\hat{m}-1}$, and (iii) follows from $e^{2(M-\hat{m})\epsilon \frac{\Delta t}{T}} e^{(M-\hat{m})\epsilon_p} \ge 1$. This proves (6.7) at $m = \hat{m} - 1$.

Next, we show the lower bound (6.8) for $u^{\hat{m}-}_{n,j,k}$. By the induction hypothesis (6.6), we have

$$u^{\hat{m}}_{n,j,k} \ge -(M - \hat{m} + 1) e^{2(M-\hat{m})\epsilon \frac{\Delta \tau}{T}} e^{(M-\hat{m})\epsilon_p} \left(\left\| u^{M-} \right\|_\infty + e^{w_{\max}} \right). \tag{6.14}$$

Comparing $u^{\hat{m}-}_{n,j,k}$, which is given by the supremum of the rhs of (5.10) (when $m = \hat{m}$), with $u^{\hat{m}}_{n,j,k} + p_{\hat{m}-1} q_{\hat{m}-1}$, which is the rhs of (5.10) evaluated at $\gamma^{\hat{m}}_{n,j,k} = 0$, yields

$$u^{\hat{m}-}_{n,j,k} \ge u^{\hat{m}}_{n,j,k} + p_{\hat{m}-1} q_{\hat{m}-1} e^{w_n} \overset{\text{(i)}}{\ge} -(M - (\hat{m}-1) + 1) e^{2(M-\hat{m})\epsilon \frac{\Delta \tau}{T}} e^{(M-\hat{m})\epsilon_p} \left(\left\| u^{M-} \right\|_\infty + e^{w_{\max}} \right). \tag{6.15}$$

Here, in (i), we use (6.14). This proves (6.8) at $m = \hat{m}$. From (6.13)-(6.15), noting $\epsilon \le 1/2$, we have

$$|u^{\hat{m}-}_{n,j,k}| \le (M - (\hat{m}-1) + 1) e^{2(M-(\hat{m})\epsilon \frac{\Delta t}{T}} e^{(M-\hat{m})\epsilon_p} (\|u^{M-}\|_\infty + e^{w_{\max}}). \tag{6.16}$$

<u>Step 2 - Bound for $u^{\hat{m}-1}_{n,j,k}$:</u> We now show that (6.5)-(6.6) hold at $m = \hat{m}-1$. For fixed $(n, j, k) \in \mathbb{N} \times \mathbb{J} \times \mathbb{K}$, we first consider $\Delta w \sum_{l \in \mathbb{N}^\dagger} \tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t) u^{\hat{m}-}_{l,j,s} \le \Delta w \sum_{l \in \mathbb{N}^\dagger} \left| \tilde{g}_{n-l}(\cdot) \right| \left| u^{\hat{m}-}_{l,j,s} \right| \ldots$

$$\overset{\text{(i)}}{\le} (M - (\hat{m}-1) + 1) e^{2(M-\hat{m})\epsilon \frac{\Delta t}{T}} e^{(M-\hat{m})\epsilon_p} (\|u^{M-}\|_\infty + e^{w_{\max}}) \left(\Delta w \sum_{l \in \mathbb{N}^\dagger} (\max(\tilde{g}_{n-l}, 0) + |\min(\tilde{g}_{n-l}, 0)|) \right)$$

$$\overset{\text{(ii)}}{\le} (M - (\hat{m}-1) + 1) e^{2(M-(\hat{m}-1))\epsilon \frac{\Delta t}{T}} e^{(M-\hat{m})\epsilon_p} (\|u^{M}\|_\infty + e^{w_{\max}}). \tag{6.17}$$

Here, in (i), we use the bound (6.16); and (ii) comes from the second result of (6.1). Then, using (6.17) and (6.2)[Remark 6.1], $u^{\hat{m}-1}_{n,j,k}$ as given by (5.17) can be bounded by

$$\left| u^{\hat{m}-1}_{n,j,k} \right| \le \sum_{s \in \mathbb{K}} \vartheta_s p(\sigma_k, \sigma_s; \Delta t) \; \Delta w \sum_{l \in \mathbb{N}^\dagger} \left| \tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t) \right| \left| u^{\hat{m}-}_{l,j,s} \right| \tag{6.18}$$

$$\le (M - (\hat{m}-1) + 1) e^{2(M-(\hat{m}-1))\epsilon \frac{\Delta t}{T}} e^{(M-\hat{m})\epsilon_p} (\|u^{M-}\|_\infty + e^{w_{\max}}) \left(\sum_{s \in \mathbb{K}} \vartheta_s \; p(\sigma_k, \sigma_s; \Delta t) \right)$$

$$\le (M - (\hat{m}-1) + 1) e^{2(M-(\hat{m}-1))\epsilon \frac{\Delta t}{T}} e^{(M-(\hat{m}-1))\epsilon_p} \left(\|u^{M-}\|_\infty + e^{w_{\max}} \right).$$

This proves (6.5) at $m = \hat{m} - 1$.

To prove (6.6) at $m = \hat{m} - 1$, first, with (6.7)-(6.8), we have $\Delta w \sum_{l \in \mathbb{N}^\dagger} \tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t) \; u^{\hat{m}-}_{l,j,s} \ldots$

$$\ldots \ge -(M - (\hat{m}-1) + 1) e^{2(M-\hat{m})\epsilon \frac{\Delta \tau}{T}} e^{(M-\hat{m})\epsilon_p} \; \Delta w \sum_{l=-N^\dagger/2}^{N^\dagger/2-1} (\max(\tilde{g}_{n-l}, 0) + |\min(\tilde{g}_{n-l}, 0)|)$$

$$\ge -(M - (\hat{m}-1) + 1) e^{2(M-\hat{m})\epsilon \frac{\Delta \tau}{T}} e^{(M-\hat{m})\epsilon_p} \left(1 + 2\epsilon \frac{\Delta t}{T} \right)$$

$$\ge -(M - (\hat{m}-1) + 1) e^{2(M-(\hat{m}-1))\epsilon \frac{\Delta \tau}{T}} e^{(M-\hat{m})\epsilon_p}. \tag{6.19}$$

Here, in the last inequality of (6.19), we use (6.1). Using (6.19) gives

$$u^{\hat{m}-1}_{n,j,k} = \sum_{s \in \mathbb{K}} \vartheta_s p(\cdot; \Delta t) \left(\Delta w \sum_{l \in \mathbb{N}^\dagger} \tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t) u^{\hat{m}-}_{l,j,s} \right) \ge -(M - (\hat{m}-1) + 1) e^{2(M-(\hat{m}-1))\epsilon \frac{\Delta \tau}{T}} e^{(M-(\hat{m}-1))\epsilon_p},$$

where, in the last inequality, we use the second result of (6.1). This proves (6.6) at $m = \hat{m} - 1$ and concludes the proof. □

6.2 Consistency

We now turn to the pointwise consistency of our scheme. Since it is straightforward that (5.6) ($\Omega \times \{T\}$), (5.7) ($\Omega_{w_{\min}}$), and (5.8) ($\Omega_{w_{\max}}$) are consistent, we focus primarily on the consistency of (5.17) (for Ω_{in}). That is, we will show that (5.17) is (local) consistent with the double integral (4.6).

We start by introducing notational convention: we use $\mathbf{x} = (w, a, \sigma) \in \Omega^\infty$ and $\mathbf{x}^m \equiv (w, a, \sigma, t_m) \in \Omega^\infty \times \{t_m\}$, $m = M, \ldots, 0$; in addition, for brevity, we use $u^m(\mathbf{x})$ instead of $u(\mathbf{x}, t_m)$, $m = M, \ldots, 0$.

For subsequent discussions, we write (4.6) for an arbitrary, but fixed, point $\mathbf{x}^{m-1} = (w, a, \sigma, t_{m-1}) \in \Omega_{\text{in}} \times t_{m-1}$, $m = M, \ldots, 1$, in an equivalent form via an operator $\mathcal{D}(\cdot)$ as follows

$$u^{m-1}(\mathbf{x}) \;=\; \mathcal{D}\left(\mathbf{x}^{m-1}, u^m\right), \text{ where} \tag{6.20}$$
$$\mathcal{D}\left(\mathbf{x}^{m-1}, u^m\right) := \int_{\sigma_{\min}}^{\sigma_{\max}}\int_{-\infty}^{\infty} \hat{u}(w', a, \sigma', t_m^-) g(w - w', \sigma, \sigma'; \Delta t)\; dw\; \varrho(\sigma, \sigma'; \Delta t)\; d\sigma,$$

with $\hat{u}(w', \cdot, \sigma', t_m^-)$ is given by (4.7). We write the proposed numerical scheme (5.17) for $(\mathbf{x}_{n,j,k}, t_m) = (w_n, a_j, \sigma_k, t_m) \in \Omega_{\text{in}} \times t_m$, $t_m \in \mathcal{T}$, in an equivalent form via an operator $\mathcal{D}_h(\cdot)$ as follows

$$u_{n,j,k}^{m-1} = \mathcal{D}_h\left(\mathbf{x}_{n,j,k}^{m-1}, \{u_{l,j,s}^m\}_{\substack{l\in\mathbb{N}^\dagger\\ s\in\mathbb{K}}}\right) := \sum_{s\in\mathbb{K}} \vartheta_s \varrho(\sigma_k, \sigma_s; \Delta t)\left(\Delta w \sum_{l\in\mathbb{N}^\dagger} \tilde{g}_{n-l}(\sigma_k, \sigma_s) u_{l,j,s}^{m-}\right), \tag{6.21}$$

where $\tilde{g}_{n-l}(\sigma_k, \sigma_s)$ and $u_{l,j,s}^{m-}$ are respectively given by (5.16) and (5.12).

We denote by $\mathcal{B}$ the set of bounded functions defined by [8, 49]

$$\mathcal{B}(\Omega^\infty \times [0, T]) = \left\{ v : \Omega^\infty \times [0, T] \to \mathbb{R}, \quad \sup_{\mathbf{x}\in\Omega^\infty\times[0,T]} |v(\mathbf{x})| < \infty \right\}. \tag{6.22}$$

We first show an supplementary result on functions in $(\mathcal{B}\cup\mathcal{C})(\Omega^\infty \times [0, T])$, i.e. bounded and continuous functions.

Lemma 6.3 (Functions in $\mathcal{B}\cup\mathcal{C}$)**.** *Suppose the discretization parameter h satisfies* (5.5)*. Let $\psi(w, a, \sigma, t)$ be in $(\mathcal{B}\cup\mathcal{C})(\Omega^\infty \times [0, T])$. For any $\mathbf{x}_{n,j,k}^{m-1} = (w_n, a_j, \sigma_k, t_{m-1})$, $(n, j, k) \in \mathbb{N}\times\mathbb{J}\times\mathbb{K}$ and $m \in \{M, \ldots, 1\}$, we have*

$$\mathcal{D}_h\left(\mathbf{x}_{n,j,k}^{m-1}, \{\psi_{l,j,s}^m\}_{\substack{l\in\mathbb{N}^\dagger\\ s\in\mathbb{K}}}\right) = \mathcal{D}\left(\mathbf{x}_{n,j,k}^{m-1}, \psi^m\right) + \mathcal{O}(h^2) + \mathcal{E}(\mathbf{x}_{n,j,k}^{m-1}, h), \tag{6.23}$$

where $\mathcal{E}(\mathbf{x}_{n,j,k}^{m-1}, h) \to 0$ as $h \to 0$. Here, the operators $\mathcal{D}(\cdot)$ and $\mathcal{D}_h(\cdot)$ are defined in (6.20) *and* (6.21)*, respectively*

Proof of Lemma 6.3. We first consider the term $\Delta w \sum_{l\in\mathbb{N}^\dagger} \tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t)\; \psi_{l,j,s}^m$ on the lhs-of-(6.23). For brevity, we fix $a = a_j$, $\sigma = \sigma_s$, and $t = t_m$, and instead of writing $\psi(w, a_j, \sigma_s, t_m)$, we will write $\psi(w)$ which is a bounded and continuous function of $w \in \mathbb{R}$. We will also write ψ_l instead of $\psi_{l,j,s}^m$, $l \in \mathbb{N}^\dagger$. Without loss of generality, we assume that $\psi(w) \in L^1(\mathbb{R})$; otherwise we can employ a smooth cut-off function to modify $\psi(w)$ into a function in $L^1(\mathbb{R})$ that agrees with $\psi(w)$ in $[w_{\min}, w_{\max}]$, as in [39, Lemma 5.4].

We have $\Delta w \sum_{l\in\mathbb{N}^\dagger} \tilde{g}_{n-l}(\sigma_k, \sigma_s)\; \psi_l = \ldots$

$$\begin{aligned}
\ldots &\overset{\text{(i)}}{=} \Delta w \sum_{l\in\mathbb{N}^\dagger} \tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t)\; \psi_l + \mathcal{E}_f \overset{(5.20)}{=} \sum_{l\in\mathbb{N}^\dagger} \psi_l\left(\int_{w_n-w_l-\Delta w}^{w_n-w_l+\Delta w} \varphi_{n-l}(w)\; \hat{g}(w; \sigma_k, \sigma_s; \Delta t)\; dw\right) + \mathcal{E}_f \\
&\overset{\text{(ii)}}{=} \sum_{l\in\mathbb{N}^\dagger} \psi_l\left(\int_{w_n-w_l-\Delta w}^{w_n-w_l+\Delta w} \varphi_{n-l}(w)\; g(w; \sigma_k, \sigma_s; \Delta t)\; dw\right) + \mathcal{E}_f + \mathcal{E}_{\hat{g}} \\
&\overset{\text{(iii)}}{=} \Delta w \sum_{l\in\mathbb{N}^\dagger} \psi_l\; g(w_n - w_l; \sigma_k, \sigma_s; \Delta t) + \mathcal{E}_f + \mathcal{E}_{\hat{g}} + \mathcal{E}_o, \\
&\overset{\text{(iv)}}{=} \int_{w_{\min}^\dagger}^{w_{\max}^\dagger} \psi(w, a_j, \sigma_s, t_m)\; g(w_n - w; \sigma_k, \sigma_s; \Delta t)\; dw + \mathcal{E}_f + \mathcal{E}_{\hat{g}} + \mathcal{E}_o + \mathcal{E}_c \\
&\overset{\text{(v)}}{=} \int_{-\infty}^{\infty} \psi(w, a_j, \sigma_s, t_m)\; g(w_n - w; \sigma_k, \sigma_s; \Delta t)\; dw + \mathcal{E}_f + \mathcal{E}_{\hat{g}} + \mathcal{E}_o + \mathcal{E}_c + \mathcal{E}_b,
\end{aligned} \tag{6.24}$$

where the errors $\mathcal{E}_f$, $\mathcal{E}_{\hat{g}}$, $\mathcal{E}_o$, $\mathcal{E}_c$, and $\mathcal{E}_b$ are described below.

- In (i), $\mathcal{E}_f \equiv \mathcal{E}_f(\mathbf{x}_{n,j,k}^{m-1}, \sigma_s, h)$ is the Fourier series error arising from truncating $\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t)$, defined in (5.21), to $\tilde{g}_{n-l}(\sigma_k, \sigma_s; \Delta t, \alpha)$, for some $\alpha \in \{2, 4, 8, \ldots\}$, in (5.22). (Here, in particular, as noted in Remark 5.4, $\alpha = \alpha_\epsilon$.) We note that, by results in Subsection 5.2.3, it follows that $\mathcal{E}_f \to 0$ as $h \to 0$.

- In (ii), the error $\mathcal{E}_{\hat{g}} \equiv \mathcal{E}_{\hat{g}}(\mathbf{x}_{n,j,k}^{m-1}, \sigma_s, h)$ is due to approximating $g(w, \sigma_k, \sigma_s; \Delta t)$ by its localized, periodic approximation $\hat{g}(w, \sigma_k, \sigma_s; \Delta t)$, and is defined by
$$\mathcal{E}_{\hat{g}} = \sum_{l \in \mathbb{N}^\dagger} \psi_l \left(\int_{w_n - w_l - \Delta w}^{w_n - w_l + \Delta w} \varphi_{n-l}(w) \ \left(\hat{g}(w, \sigma_k, \sigma_s; \Delta t) - g(w, \sigma_k, \sigma_s; \Delta t)\right) \ dw \right). \tag{6.25}$$
We note that, for fixed $\sigma', \sigma \in [\sigma_{\min}, \sigma_{\max}]$, $\hat{g}(w, \sigma', \sigma; \Delta t) \neq g(w, \sigma', \sigma; \Delta t)$ for $w \in [w_{\min}^\dagger, w_{\max}^\dagger]$. However, since h satisfies (5.5), i.e. $P^\dagger = C_4/h$, then, as $h \to 0$, we have
$$\hat{g}(w, \sigma, \sigma'; \Delta t) = \int_{\mathbb{R}} e^{2\pi i \eta w} G(\eta, \sigma, \sigma'; \Delta t) d\eta + \mathcal{O}\left(1/\left(P^\dagger\right)^2\right) = g(w, \sigma, \sigma'; \Delta t) + \mathcal{O}(h^2). \tag{6.26}$$
Here, the error $\mathcal{O}\left(1/\left(P^\dagger\right)^2\right) \sim \mathcal{O}(h^2)$ is due to the trapezoidal rule approximation of the integral. Using (6.26), noting that $\psi(w) \in L^1(\mathbb{R})$, we obtain $\mathcal{E}_{\hat{g}} = \mathcal{O}\left(h^2\right)$ as $h \to 0$.

- In (iii), $\mathcal{E}_o \equiv \mathcal{E}_o(\mathbf{x}_{n,j,k}^{m-1}, \sigma_s, h)$ is due to the simple lhs/rhs numerical rule to approximate an integral and is given by
$$\mathcal{E}_o = \Delta w \ g(w_n - w_l, \sigma_k, \sigma_s; \Delta t) - \int_{w_n - w_l - \Delta w}^{w_n - w_l + \Delta w} \varphi_{n-l}(w) \ g(w, \sigma_k, \sigma_s; \Delta t) \ dw. \tag{6.27}$$
Since the probability density function $g(\cdot)$ is at least in $\mathcal{C}^2$, it follows that $\mathcal{E}_o \to 0$ as $h \to 0$.

- In (iv), $\mathcal{E}_c \equiv \mathcal{E}_c(\mathbf{x}_{n,j,k}^{m-1}, \sigma_s, h)$ is the error arising from the simple lhs numerical integration rule
$$\mathcal{E}_c = \Delta w \sum_{l \in \mathbb{N}^\dagger} \psi_l \ g(w_n - w_l, \sigma_k, \sigma_s; \Delta t) - \int_{w_{\min}^\dagger}^{w_{\max}^\dagger} \psi(w, a_j, \sigma_s, t_m) \ g(w_n - w, \sigma_k, \sigma_s; \Delta t) \ dw. \tag{6.28}$$
Due to continuity and boundedness of the integrand, we have $\mathcal{E}_c \to 0$ as $h \to 0$.

- In (v), $\mathcal{E}_b \equiv \mathcal{E}_b(\mathbf{x}_{n,j,k}^{m-1}, h)$ is the boundary truncation error
$$\mathcal{E}_b = \int_{\mathbb{R} \setminus [w_{\min}^\dagger, w_{\max}^\dagger]} \psi(w, a_j, \sigma_s, t_m) \ g(w_n - w, \sigma_k, \sigma_s; \Delta t) \ dw. \tag{6.29}$$
As $h \to 0$, $P^\dagger \to \infty$, therefore, it follows that $\mathcal{E}_b \to 0$ as $h \to 0$, noting $\psi(w) g(w_n - w)$ is bounded function for all $w \in \mathbb{R}$.

For brevity, from (6.24), we let $\psi'(\sigma; \mathbf{x}_{n,j,k}^{m-1})$ be a function of $\sigma \in [\sigma_{\min}, \sigma_{\max}]$, parameterized by $\mathbf{x}_{n,j,k}^{m-1} \in \Omega_{\text{in}}$ defined as follows
$$\psi'(\sigma; \mathbf{x}_{n,j,k}^{m-1}) = \int_{-\infty}^{\infty} \psi(w, a_j, \sigma, t_m) \ g(w_n - w; \sigma_k, \sigma; \Delta t) \ dw.$$
Then, using (6.24), the lhs-of-(6.23) $= \sum_{s \in \mathbb{K}} \vartheta_s \varrho(\sigma_k, \sigma_s; \Delta t) \left(\sum_{l \in \mathbb{N}^\dagger} \psi_{l,j,s}^m \ g(w_n - w_l, \sigma_k, \sigma_s; \Delta t) \right) = \ldots$
$$\ldots \overset{\text{(i)}}{=} \sum_{s \in \mathbb{K}} \vartheta_s \varrho(\sigma_k, \sigma_s; \Delta t) \psi'(\sigma; \mathbf{x}_{n,j,k}^m) + \mathcal{E}_1 + \mathcal{O}(h) \overset{\text{(ii)}}{=} \int_{\sigma_{\min}}^{\sigma_{\max}} \psi'(\sigma; \mathbf{x}_{n,j,k}^m) \ \varrho(\sigma_k, \sigma; \Delta t) \ d\sigma + \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{O}(h),$$
where the errors $\mathcal{E}_1 \equiv \mathcal{E}_1(\mathbf{x}_{n,j,k}^{m-1}, h)$, $\mathcal{E}_2 \equiv \mathcal{E}_2(\mathbf{x}_{n,j,k}^{m-1}, h)$ and $\mathcal{O}(h)$ are described as follows.

- In (i) $\mathcal{E}_1 \equiv \mathcal{E}_1(\mathbf{x}_{n,j,k}^{m-1}, h) = \sum_{s \in \mathbb{K}} \vartheta_s \varrho(\sigma_k, \sigma_s; \Delta t)(\mathcal{E}_f + \mathcal{E}_o + \mathcal{E}_c + \mathcal{E}_b)$. Due to (6.2)[Remark 6.1], it follows that $\mathcal{E}_1 \to 0$ as $h \to 0$. The $\mathcal{O}(h^2)$ error comes from $\sum_{s \in \mathbb{K}} \vartheta_s \varrho(\sigma_k, \sigma_s; \Delta t) \mathcal{E}_{\hat{g}} = \mathcal{O}(h)$, also due to the same reason, noting $\mathcal{E}_{\hat{g}} = \mathcal{O}(h^2)$.

- In (ii), $\mathcal{E}_2 \equiv \mathcal{E}_2(\mathbf{x}_{n,j,k}^{m-1}, h)$ is an approximation error arising from a quadrature rule
$$\mathcal{E}_2 = \sum_{s \in \mathbb{K}} \vartheta_s \varrho(\sigma_k, \sigma_s; \Delta t) \ \psi'(\sigma_s; \mathbf{x}_{n,j,k}^{m-1}) - \int_{\sigma_{\min}}^{\sigma_{\max}} \psi'(\sigma; \mathbf{x}_{n,j,k}^{m-1}) \ \varrho(\sigma_k, \sigma; \Delta t) \ d\sigma. \tag{6.30}$$
It is straightforward to see that $\psi'(\sigma; \mathbf{x}_{n,j,k}^{m-1})$ is continuous and bounded function of $\sigma \in [\sigma_{\min}, \sigma_{\max}]$, and hence $\mathcal{E}_2 \to 0$ as $h \to 0$.

Letting $\mathcal{E}(\mathbf{x}_{n,j,k}^{m-1}, h) = \mathcal{E}_1(\mathbf{x}_{n,j,k}^{m-1}, h) + \mathcal{E}_2(\mathbf{x}_{n,j,k}^{m-1}, h)$ concludes the proof.

□

We now introduce a lemma on local consistency of the scheme.

Lemma 6.4 (Local consistency)**.** *Suppose that (i) the discretization parameter h satisfies* (5.5), *(ii) linear interpolation is used for the intervention action* (5.10)*. For any smooth test function $\phi \in (\mathcal{B} \cup \mathcal{C}^{\infty})(\Omega^{\infty} \times [0,T])$, with $\phi^m_{n,j,k} \equiv \phi(\mathbf{x}^m_{n,j,k})$ and $\mathbf{x}^{m-1}_{n,j,k} \in \Omega_{in} \times \{t_{m-1}\}$, $m = M, \ldots, 1$, and for a sufficiently small h, we have*

$$\mathcal{D}_h\left(\mathbf{x}^{m-1}_{n,j,k}, \{\phi^m_{l,j,s} + \chi\}_{\substack{l\in\mathbb{N}^\dagger\\ s\in\mathbb{K}}}\right) = \mathcal{D}\left(\mathbf{x}^{m-1}_{n,j,k}, \phi^m\right) + c(\mathbf{x}^{m-1}_{n,j,k})\chi + \mathcal{O}(h^2) + \mathcal{E}(\mathbf{x}^{m-1}_{n,j,k}, h). \tag{6.31}$$

Here, χ is a constant and $c(\cdot)$ is a bounded function satisfying $|c(\mathbf{x}^m_{n,j,k})| \leq \max(r,1)$ for all $\mathbf{x} \in \Omega$, and $\mathcal{E}(\mathbf{x}^{m-1}_{n,j,k}, h) \to 0$ as $h \to 0$. The operators $\mathcal{D}(\cdot)$ and $\mathcal{D}^{m-1}_h(\cdot)$ are defined in (6.20) *and* (6.21)*, respectively.*

Proof of Lemma 6.4. In this case, the operator $\mathcal{D}_h(\cdot)$ is written as

$$\ldots \mathcal{D}_h(\cdot) = \sum_{s\in\mathbb{K}} \vartheta_s \varrho(\sigma_k,\sigma_s;\Delta t)\Bigg(\Delta w \sum_{l=-N/2+1}^{N/2-1} \tilde{g}_{n-l}(\sigma_k,\sigma_s) \sup_{\gamma\in[0,e^{w_l}\vee C_r a_j]} (\hat{\phi}^m_{l,j,s} + p_m f(\gamma;a_j)) + p_{m-1}q_{m-1}e^{w_l},$$

$$+ \Delta w \sum_{l=-N^\dagger/2}^{N/2} \tilde{g}_{n-l}(\sigma_k,\sigma_s)\ (\phi^m_{l,j,s} + \chi) + \Delta w \sum_{l=N/2}^{N^\dagger/2-1} \tilde{g}_{n-l}(\sigma_k,\sigma_s)\ (\phi^m_{l,j,s} + \chi)\Bigg), \tag{6.32}$$

where $\hat{\phi}^m_{l,j,s} \equiv \mathcal{I}\{\phi^m_s + \chi\}(\hat{w}_l, \hat{a}_j) = \phi(\hat{w}_l, \hat{a}_j, \sigma_s, t_m) + \chi + \mathcal{O}\left(h^2\right)$, with $(\hat{w}_l, \hat{a}_j)$ defined in (3.3). Here, the error of size $\mathcal{O}\left((\Delta w + \Delta a_{\max})^2\right) = \mathcal{O}\left(h^2\right)$ is due to linear interpolation, noting that we can completely separate χ from interpolated values.

Let $\phi'(\mathbf{x}')$ be a function of $\mathbf{x}' = (w', a', \sigma', t) \in \Omega^{\infty} \times [0,T)$ defined by

$$\phi'(\mathbf{x}') = \begin{cases} \sup\limits_{\gamma\in[0,e^{w'}\vee C_r a']} \mathcal{M}(\gamma)\phi(\mathbf{x}') + p_m f(\gamma;a') & \mathbf{x}' \in \Omega_{\text{in}} \times [0,T) \quad (6.33\text{a}) \\ \phi(\mathbf{x}') & \text{otherwise}, \quad (6.33\text{b}) \end{cases}$$

where $\mathcal{M}(\cdot)$ is defined in (3.4). It is straightforward to show that $\phi' \in (\mathcal{B} \cup \mathcal{C}(\Omega^{\infty} \times [0,T))$; Therefore, using (6.33), (6.32) can be written as

$$\begin{aligned} \mathcal{D}_h(\cdot) &= \sum_{s\in\mathbb{K}} \vartheta_s \varrho(\sigma_k,\sigma_s;\Delta t)\Bigg(\Delta w \sum_{l\in\mathbb{N}^\dagger} \tilde{g}_{n-l}(\sigma_k,\sigma_s)\ \phi'(\mathbf{x}^m_{l,j,s})\Bigg) \\ &\qquad + \chi \sum_{s\in\mathbb{K}} \vartheta_s \varrho(\sigma_k,\sigma_s;\Delta t)\Bigg(\Delta w \sum_{l\in\mathbb{N}^\dagger} \tilde{g}_{n-l}(\sigma_k,\sigma_s)\Bigg) + \mathcal{O}\left(h^2\right) \\ &\overset{\text{(i)}}{=} \int_{\sigma_{\min}}^{\sigma_{\max}} \int_{-\infty}^{\infty} \phi'(w, a_j, \sigma, t_m)\ g(w_n - w, \sigma_k, \sigma; \Delta t)\ dw\ \varrho(\sigma_k,\sigma;\Delta t)\ d\sigma + c(\mathbf{x}^{m-1}_{n,j,k})\chi + \mathcal{O}(h^2) + \mathcal{E}(\mathbf{x}^{m-1}_{n,j,k}, h). \end{aligned}$$

Here, (i) is due to Lemma 6.3, with $c(\mathbf{x}^{m-1}_{n,j,k}) = 1 - \sum_{s\in\mathbb{K}} \vartheta_s \varrho(\sigma_k,\sigma_s;\Delta t)\big(\Delta w \sum_{l\in\mathbb{N}^\dagger} \tilde{g}_{n-l}(\sigma_k,\sigma_s)\big)$. This concludes the proof, noting that the double integral appearing above is $\mathcal{D}\left(\mathbf{x}^{m-1}_{n,j,k}, \phi^m\right)$. □

Remark 6.2. *Lemma 6.4 indicates that second-order convergence is possible if $\mathcal{E}(\mathbf{x}^{m-1}_{n,j,k}, h) \sim \mathcal{O}(h^2)$ as $h \to 0$. Through extensive numerical experiments, second-order convergence is observed in the numerical results. This will be elaborated further in Section 7.*

Next, we present a result on the ϵ-monotonicity of the numerical scheme $\mathcal{D}^{m-1}_h(\cdot)$.

Lemma 6.5 (ϵ-monotonicity)**.** *If linear interpolation is used for the intervention action* (5.10)*, and $\tilde{g}_{n-l}(\sigma_k,\sigma_s)$ satisfies the monotonicity condition* (5.23)*, i.e. $\Delta w \sum_{l\in\mathbb{N}^\dagger} \big|\min(\tilde{g}_{n-l}(\sigma_k,\sigma_s),0)\big| < \epsilon\frac{\Delta\tau}{T}$, where $\epsilon > 0$, then scheme* (6.21) *satisfies*

$$\mathcal{D}_h\left(\mathbf{x}^{m-1}_{n,j,k}, \{x^m_{l,j,s}\}_{\substack{l\in\mathbb{N}^\dagger\\ s\in\mathbb{K}}}\right) \ \leq\ \mathcal{D}_h\left(\mathbf{x}^{m-1}_{n,j,k}, \{y^m_{l,j,s}\}\right) \ +\ C\epsilon,$$

for bounded $\{x^m_{l,j,s}\}$ and $\{y^m_{l,j,s}\}$ having $\{x^m_{l,j,s}\} \leq \{y^m_{l,j,s}\}$, where the inequality is understood in the component-wise sense, and C is a positive constant independent of h and ϵ.

Proof of Lemma 6.5. Recall the linear interpolation operator $\mathcal{I}\{\cdot\}(\cdot)$ in (5.9). For each fixed $s \in \mathbb{K}$, let $\hat{x}^m_{n,j,s}$ and $\hat{y}^m_{n,j,s}$ be the results of the linear operators $\mathcal{I}\{x^m_s\}(\hat{w}_l, \hat{a}_j)$ and $\mathcal{I}\{y^m_s\}(\hat{w}_l, \hat{a}_j)$ acting on $\left\{\left((w_l, a_j), x^m_{l,j,s}\right)\right\}$, and $\left\{\left((w_l, a_j), y^m_{l,j,s}\right)\right\}$, $(l,j) \in \mathbb{N}^\dagger \times \mathbb{J}$, respectively. We also define for $x^{m-}_{l,j,s}$, $y^{m-}_{l,j,s}$, in a similar way that $u^{m-}_{l,j,s}$ in (5.9).

For the rest of the proof, let C be a generic positive constant independent of h and ϵ, which may take different values from line to line. From the boundedness of $\{x^m_{l,j,s}\}$ and $\{y^m_{l,j,s}\}$, and $\{x^m_{l,j,s}\} \leq \{y^m_{l,j,s}\}$, noting $\mathcal{I}\{x^m_s\}(\cdot)$ and $\mathcal{I}\{y^m_s\}(\cdot)$ are linear interpolation operators, we have, for each fixed $j \in \mathbb{J}$ and $s \in \mathbb{K}$,

$$x^{m-}_{l,j,s} \;\leq\; y^{m-}_{l,j,s} \text{ and } \left|x^{m-}_{l,j,s} - y^{m-}_{l,j,s}\right| \;\leq\; C, \quad l = -N^\dagger/2, \ldots, N^\dagger/2 - 1. \tag{6.34}$$

Next, using (6.34) together with the definition of the operator $\mathcal{D}_h(\cdot)$ in (6.21), we have

$$\begin{aligned}
&\mathcal{D}_h\left(\cdot, \{x^m_{l,j,s}\}_{\substack{l\in\mathbb{N}^\dagger \\ s\in\mathbb{K}}}\right) - \mathcal{D}_h\left(\cdot, \{y^m_{l,j,s}\}_{\substack{l\in\mathbb{N}^\dagger \\ s\in\mathbb{K}}}\right) = \sum_{s\in\mathbb{K}} \vartheta_s p(\sigma_k, \sigma_s; \Delta t)\big(\Delta w \sum_{l\in\mathbb{N}^\dagger} \tilde{g}_{n-l}(\sigma_k, \sigma_s)\big(x^{m-}_{l,j,s} - y^{m-}_{l,j,s}\big)\big) \\
&\leq \sum_{s\in\mathbb{K}} \vartheta_s p(\sigma_k, \sigma_s; \Delta t)\big(\Delta w \sum_{l\in\mathbb{N}^\dagger} \left|\min\left(\tilde{g}_{n-l}(\sigma_k, \sigma_s), 0\right)\right| \, \left|x^{m-}_{l,j,s} - y^{m-}_{l,j,s}\right|\big) \\
&\leq C \sum_{s\in\mathbb{K}} \vartheta_s p(\sigma_k, \sigma_s; \Delta t)\big(\Delta w \sum_{l\in\mathbb{N}^\dagger} \left|\min\left(\tilde{g}_{n-l}(\sigma_k, \sigma_s), 0\right)\right|\big) \\
&\overset{(i)}{\leq} \; C \sum_{s\in\mathbb{K}} \vartheta_s p_\sigma(\sigma_k, \sigma_s; \Delta t)\epsilon\frac{\Delta t}{T} \;\overset{(ii)}{\leq}\; C(1+\epsilon_p)\epsilon\frac{\Delta t}{T},
\end{aligned}$$

where in (i), we use the property (6.1), whereas in (ii), we use Remark 6.1. This concludes the proof. $\square$

With stability, consistency, and ϵ-monotonicity established in Lemmas 6.2, 6.4 and 6.5, and Proposition 4.1, we now establish the pointwise convergence of the proposed numerical scheme in $\Omega_{\text{in}} \times \{t_{m-1}\}$, $m = M, \ldots, 1$, as $h \to 0$. We first need to recall/introduce relevant notation.

We denote by Ω^h the computational grid parameterized by h, noting that $\Omega^h \to \Omega^\infty$ as $h \to 0$. We also have the respective Ω^h_{in}. In general, a generic gridpoint in $\Omega^h_{\text{in}} \times \{t_m\}$, $m = M, \ldots, 0$, is denoted by $\mathbf{x}^m_h = (\mathbf{x}_h, t_m)$, whereas an arbitrary point in $\Omega_{\text{in}} \times \{t_m\}$ is denoted by $\mathbf{x}^m = (\mathbf{x}, t_m)$. Numerical solutions at $(\mathbf{x}_h, t_{m-1})$, $m = M, \ldots, 1$, is denoted by $u^{m-1}_h(\mathbf{x}_h, u^m_h)$, where it is emphasized that u^m_h, which is the time-t_m numerical solution at gridpoints is used for the computation of u^{m-1}_h. The exact solution at an arbitrary point in $\mathbf{x}^{m-1} = (\mathbf{x}, t_{m-1}) \in \Omega_{\text{in}} \times \{t_{m-1}\}$, $m = M, \ldots, 1$, is denoted by $u^{m-1}(\mathbf{x}, u^m)$, where it is emphasized that u^m, which is the time-t_m exact solution in Ω^∞ is used. More specifically, $u^{m-1}_h(\mathbf{x}_h; u^m_h)$ and $u^{m-1}(\mathbf{x}; u^m)$ are defined via operators $\mathcal{D}_h(\cdot)$ and $\mathcal{D}(\cdot)$ as follows

$$u^{m-1}_h\big(\mathbf{x}_h, u^m_h\big) \coloneqq \mathcal{D}_h\big(\mathbf{x}^{m-1}_h, \{u^m_{l,j,s}\}\big), \qquad u^{m-1}\big(\mathbf{x}, u^m\big) \coloneqq \mathcal{D}\big(\mathbf{x}^{m-1}, u^m\big), \quad m = M, \ldots, 1. \tag{6.35}$$

Here, our convention is that $u_h\big(\mathbf{x}^{M-1}, u^M_h\big) = u_h\big(\mathbf{x}^{M-1}, u^M\big)$.

The pointwise convergence of the proposed scheme is stated in the main theorem below.

Theorem 6.1 (Pointwise convergence)**.** *Suppose that all the conditions for Lemma 6.2, 6.4, and 6.5 are satisfied. Under the assumption that the monotonicity tolerance $\epsilon \to 0$ as $h \to 0$, scheme* (6.21) *converges pointwise in $\Omega_{in} \times \{t_{m-1}\}$, $m \in \{M, \ldots, 1\}$, to the unique bounded solution of the GLWB pricing problem in Definition 4.1, i.e. for any $m \in \{M, \ldots, 1\}$, we have*

$$u^{m-1}(\mathbf{x}, u^m) \;=\; \lim_{\substack{h\to 0 \\ \mathbf{x}_h \to \mathbf{x}}} u^{m-1}_h(\mathbf{x}_h; u^m_h), \quad \text{for } \mathbf{x}_h \in \Omega^h_{in}, \quad \mathbf{x} \in \Omega_{in}. \tag{6.36}$$

Proof of Theorem 6.1. By Proposition 4.1, there exists $\phi \in (\mathcal{B} \cup \mathcal{C}^\infty)(\Omega^\infty \times [0,T])$ such that, for any $h > 0$,

$$u \leq \phi \leq u + h, \quad \text{in } \Omega^\infty \times \{t_m\}, \quad m = M, \ldots, 0. \tag{6.37}$$

We then define

$$u^{m-1}_h\big(\mathbf{x}_h, \phi^m\big) \coloneqq \mathcal{D}_h\big(\mathbf{x}^{m-1}_h, \{\phi^m_{l,j,s}\}\big), \quad u^{m-1}\big(\mathbf{x}, \phi^m\big) \coloneqq \mathcal{D}\big(\mathbf{x}^{m-1}, \phi^m\big),$$

noting our convention that $\phi^m_{l,j,s} = \phi(\mathbf{x}_{l,j,s}, t_m)$. To show (6.36), we will prove by mathematical induction on m the following result: for any $m \in \{M, \ldots, 1\}$, and for sequence $\{\mathbf{x}_h\}_{h>0}$ such that $\mathbf{x}_h \to \mathbf{x}$ as $h \to 0$,

$$\left|u^{m-1}_h(\mathbf{x}_h, u^m) - u^{m-1}(\mathbf{x}, u^m)\right| \leq \chi^{m-1}_h, \quad \chi^{m-1}_h \text{ is bounded } \forall h > 0 \text{ and } \chi^{m-1}_h \to 0 \text{ as } h \to 0. \tag{6.38}$$

In the following proof, we let K_1, K_2, and K_3 be generic positive constants independent of h and ϵ, which may take different values from line to line.

Base case $m = M$: by (6.37), we can write $u^M \le \phi^M \le u^M + h$. Therefore,

$$u_h^{M-1}\left(\mathbf{x}_h; u^M\right) \overset{(i)}{\le} u_h^{M-1}\left(\mathbf{x}_h; \phi^M\right) + K_1\epsilon, \tag{6.39}$$

$$u_h^{M-1}\left(\mathbf{x}_h; \phi^M\right) \overset{(ii)}{\le} u_h^{M-1}\left(\mathbf{x}_h; u^M + h\right) + K_2\epsilon \overset{(iii)}{\le} u_h^{M-1}\left(\mathbf{x}_h; u^M\right) + K_1\epsilon + K_2\epsilon h.$$

Here, (6.39) is due to the ϵ-monotonicity of the numerical scheme, noting the boundedness of u^M by Lemma 6.2 on stability, as demonstrated in Lemmas 6.4 combined with $u^M \le \phi^M$ for (i), $\phi^M \le u^M + h$ for (ii), and $|\phi^M - u^M| \le h$ for (iii). Therefore,

$$\left|u_h^{M-1}\left(\mathbf{x}_h; u^M\right) - u^{M-1}\left(\mathbf{x}; u^M\right)\right| \overset{(6.39)}{\le} \left|u_h^{M-1}(\mathbf{x}_h; \phi^M) - u^{M-1}(\mathbf{x}; \phi^M)\right| + K_1\epsilon + K_2\epsilon h$$

$$\overset{(i)}{\le} \left|u_h^{M-1}\left(\mathbf{x}_h; \phi^M\right) - u^{M-1}\left(\mathbf{x}_h; \phi^M\right)\right| + \left|u^{M-1}\left(\mathbf{x}_h; \phi^M\right) - u^{M-1}\left(\mathbf{x}; \phi^M\right)\right| + K_1\epsilon + K_2\epsilon h, \tag{6.40}$$

where (i) is due to the triangle inequality. By Lemma 6.3, we have

$$u_h^{M-1}\left(\mathbf{x}_h; \phi^M\right) - u^{M-1}\left(\mathbf{x}_h; \phi^M\right) = \mathcal{O}(h) + \mathcal{E}(\mathbf{x}_h^{M-1}, h). \tag{6.41}$$

Due to smoothness of $\phi(\cdot)$ and regularity of $g(\cdot)$ (see Remark 3.2), we have

$$\left|u^{M-1}\left(\mathbf{x}_h; \phi^M\right) - u^{M-1}\left(\mathbf{x}; \phi^M\right)\right| \le K_1\|\mathbf{x}_h - \mathbf{x}\|. \tag{6.42}$$

Therefore, using (6.40), (6.41), (6.42), we can show that

$$\left|u_h^{M-1}\left(\mathbf{x}_h; u^M\right) - u^{M-1}\left(\mathbf{x}; u^M\right)\right| \le \chi_h^{M-1}, \tag{6.43}$$

$$\chi_h^{M-1} = K_1\epsilon + K_2\epsilon h + \mathcal{O}(h) + |\mathcal{E}(\mathbf{x}_h^{M-1}, h)| + K_3\|\mathbf{x}_h - \mathbf{x}\| \longrightarrow 0, \text{ as } h \to 0,$$

noting $\mathbf{x}_h \to \mathbf{x}$ as $h \to 0$, and χ_h^{M-1} is bounded for all $h > 0$.

Induction hypothesis: assume that, for some $m \in \{M-1, \ldots, 2\}$, we have

$$\left|u_h^{m-1}\left(\mathbf{x}_h; u_h^m\right) - u^{m-1}\left(\mathbf{x}; u^m\right)\right| \le \chi_h^{m-1}, \text{ where } \chi_h^{m-1} \text{ is bounded, } \chi_h^{m-1} \to 0 \text{ as } h \to 0. \tag{6.44}$$

Induction step: By the triangle inequality, we have $\left|u_h^{m-2}\left(\mathbf{x}_h; u_h^{m-1}\right) - u^{m-2}\left(\mathbf{x}; u^{m-1}\right)\right| \le \ldots$

$$\ldots \le \left|u_h^{m-2}\left(\mathbf{x}_h; u_h^{m-1}\right) - u_h^{m-2}\left(\mathbf{x}_h; u^{m-1}\right)\right| + \left|u_h^{m-2}\left(\mathbf{x}_h; u^{m-1}\right) - u^{m-2}\left(\mathbf{x}; u^{m-1}\right)\right|. \tag{6.45}$$

By the induction hypothesis (6.44), $|u_h^{m-1} - u^{m-1}| \le \chi_h^{m-1}$, where $\chi_h^{m-1} \to 0$ as $h \to 0$. Therefore, the first term in (6.45) can be written as

$$\left|u_h\left(\mathbf{x}_h^{m-2}; u_h^{m-1}\right) - u_h\left(\mathbf{x}_h^{m-2}; u^{m-1}\right)\right| \overset{\text{(i)}}{\le} \chi_h' = K_1\chi_h^{m-1} + \mathcal{O}(h) + |\mathcal{E}(\mathbf{x}_h^{m-2}, h)| \to 0 \text{ as } h \to 0. \tag{6.46}$$

Here, (i) follows from the local consistency of the numerical scheme established in Lemma 6.4. Next, using the same arguments for the base case $m = M$ (see (6.43), with M being replaced by m), the second term in (6.45) can be bounded by χ_h'', where $\chi_h'' \to 0$ as $h \to 0$. Here, we note that u_h^{m-1} is bounded for all $h > 0$ by Lemma 6.2 on stability. Combining this with (6.46), we have $\left|u_h^{m-2}\left(\mathbf{x}_h; u_h^{m-1}\right) - u^{m-2}\left(\mathbf{x}; u^{m-1}\right)\right| \le \chi_h^{m-2}$, where χ_h^{m-2} is bounded for all $h > 0$, and $\chi_h^{m-2} \to 0$ as $h \to 0$. This concludes the proof. □

7 Numerical examples

This section presents selected numerical results for the GLWB no-arbitrage pricing problem under the dynamics (2.2)-(2.3)-(2.4), which, for convenience, will be referred to as the Heston model. In addition to validation examples, we particularly focus on investigating the impact of jump-diffusion dynamics and stochastic interest rates on the prices/the fair insurance fees, as well as on the holder's optimal withdrawal behaviors.

A set of GLWB parameters commonly used for subsequent experiments is given in Table 7.1. These include expiry time T, the annual withdrawal rate C_r, the premium z_0 which is also the initial balance of the guarantee account and the personal sub-account, mortality table/payments (assuming that the holder is a 65-year-old male at inception), and event times.

For experiments in this section, the computational domain is constructed with $w_{\min} = \ln(z_0) - 5$, $w_{\max} = \ln(z_0) + 10$, together with $w_{\min}^\dagger$, $w_{\max}^\dagger$ computed as discussed in Section 5. We also set $a_{\min} = 0$ and $a_{\max} = e^{w_{\max}}$. Regarding the choice of $\sigma_{\min}$ and $\sigma_{\max}$, we follow the numerical procedures presented in [11] to minimize the truncation error, which is outlined below for convenience.

- We set initial guess of $[\sigma_{\min}^{(0)}, \sigma_{\max}^{(0)}]$ using the mean and the variance of $V(T)$, which is assumed to follow the dynamics (2.2), i.e.

$$[\sigma_{\min}^{(0)}, \sigma_{\max}^{(0)}] := \left[\ln\left(\mathbb{E}[V(T)]\right) - 7\frac{\mathbb{V}[V(T)]}{\mathbb{E}[V(T)]^2}, \ln\left(\mathbb{E}[V(T)]\right) + 3\frac{\mathbb{V}[V(T)]}{\mathbb{E}[V(T)]^2}\right],$$

  where the mean $\mathbb{E}[V(T)]$ and the variance $\mathbb{V}[V(T)]$ can be calculated as

$$\mathbb{E}[V(T)] = \nu_0 e^{-\lambda T} + \theta\left(1 - e^{-\lambda T}\right), \ \mathbb{V}[V(T)] = \nu_0 \frac{\xi^2}{\lambda} e^{-\lambda T} - e^{-2\lambda T} + \theta \frac{\xi^2}{2\lambda}\left(1 - e^{-\lambda T}\right)^2. \quad (7.1)$$

- Given the initial guess $[\sigma_{\min}^{(0)}, \sigma_{\max}^{(0)}]$, we have two methods for finding the final interval $[\sigma_{\min}^{(n)}, \sigma_{\max}^{(n)}]$ with n be the index of iterative steps.
  - When the Feller's condition for the variance process is satisfied, we update the $[\sigma_{\min}^{(n)}, \sigma_{\max}^{(n)}]$ by subtracting and adding the approximated value for $\mathbb{V}[V(T)]$ given in (7.1).
  - When the Feller's condition for the variance process is not satisfied, we apply the Newton iteration, for which we need the first derivative of $p(\cdot)$ given in (3.9).

Unless otherwise stated, relevant details about the refinement levels are given in Table 7.2. Here, the timestep $M = 57$ corresponds to the case of $T = 57$ in Table 7.1. Since we apply the similarity reduction results from [25, 34], there is no need to discretize along a-dimension in practice. Based on the choices of N, we have $N^\dagger = 2N$ as in (5.4). We emphasize that, increasing $|w_{\min}|$, $w_{\max}$, $|\sigma_{\min}|$, or $\sigma_{\max}$ virtually does not change the no-arbitrage prices/fair insurance fees. Therefore, for practical purposes, with $P^\dagger \equiv w_{\max}^\dagger - w_{\min}^\dagger$ chosen sufficiently large as above, it can be kept constant for all refinement levels (as we let $h \to 0$).

For user-defined tolerances ϵ and ϵ_α in Algorithm 5.1, we use $\epsilon = \epsilon_\alpha = 10^{-6}$ for all experiments and all refinement levels. We note that using smaller ϵ or ϵ_α produces virtually identical numerical results.

Parameter	Value
Expiry time (T)	57 years
Annual withdrawal rate (C_r)	0.05
Init. lump-sum premium (z_0)	100
Mortality	DAV 2004R
Mortality payments	At year end
Event times	Yearly

TABLE 7.1: *GLWB parameters for numerical experiments.*

Refinement level	N (w)	J (a)	K (ν)	M (t)
0	2^9	2^9	2^4	57
1	2^{10}	2^{10}	2^5	57
2	2^{11}	2^{11}	2^6	57
3	2^{12}	2^{12}	2^7	57
4	2^{13}	2^{13}	2^8	57

TABLE 7.2: *Grid and timestep refinement levels for numerical experiments.*

7.1 Validation

We compare our numerical results with those obtained in the literature. When reference prices or insurance fees are not available for the dynamics considered in this work, for validation purposes, we compare no-arbitrage prices obtained by the proposed numerical method, hereafter referred to as "ϵ-mF", with those obtained by MC simulation. It is straightforward to carry out Monte Carlo validation for the case of deterministic withdrawal, which means that the holder withdraws at the pre-determined contractual rate at each event time.

To carry out Monte Carlo validation for the case of optimal withdrawal, we proceed in two steps as follows. In Step 1, we solve the GLWB pricing problem using the "ϵ-mF" method on a relatively fine computational grid (Refinement Level 2 in Table 7.2). During this step, the optimal control $\gamma_{n,j,k}^m$ is stored for each computational grid point. In Step 2, we carry out Monte Carlo simulation of dynamics (2.3), (2.4), and (2.2), for $Z(t)$, $A(t)$, and $V(t)$, respectively, following the stored optimal strategies $\{(\mathbf{x}_{n,j,k}^m, \gamma_{n,j,k}^m)\}$ obtained in Step 1. For Step 2, a total of 10^5 paths and a timestep size $\Delta t/20$ is used between event dates for the simulation of the dynamics. The antithetic variate technique is also employed to reduce the variance of MC simulation.

We consider the GLWB pricing problem under the Heston model with both the deterministic and optimal withdrawal strategies. The set of input parameters is summarized in Table 7.3, which is also used for numerical tests in [29]. As presented in Table 7.4, all the numerical results have a good agreement with the benchmark results obtained from a finite difference method in [29], as well as those obtained by MC simulation. We observe that second-order convergence is attained for the "ϵ-mF" method, even when the Feller's condition is not satisfied (compared with Monte Carlo simulation results). This indicates that the finite computational domain is chosen sufficiently large. Also see Remark 6.2.

Parameter	Deterministic withdrawal	Optimal withdrawal
Constant interest rate (r)	0.04	0.04
Speed of mean reversion (λ)	1.0	1.0
Volatility of variance (ξ)	0.2	0.2
Correlation (ρ)	-0.5	-0.5
Long time mean of variance (θ)	0.0225	0.0225
Initial variance (ν_0)	0.0225	0.0225
Penalty rate ($\mu(t)$)	N/A	$0 < t \leq 1 : 5\%$, $1 < t \leq 2 : 4\%$, $2 < t \leq 3 : 3\%$, $3 < t \leq 4 : 2\%$, $4 < t \leq 5 : 1\%$, $t > 5 : 0\%$
Bonus rate (b)	N/A	0.05
Ratchet cycle (f_r)	N/A or every year	N/A or every three years
Withdrawal strategy	$C_r A(t)$	Optimal

TABLE 7.3: *Input parameters used for the case of deterministic withdrawal in [29, Table 2], and the case of optimal withdrawal in [29, Table 2].*

Level	Deterministic withdrawal				Optimal withdrawal			
	Ratchet		No ratchet		Ratchet		No Ratchet	
	Value	Ratio	Value	Ratio	Value	Ratio	Value	Ratio
0	100.1485		100.1357		100.1178		100.1079	
1	100.0334		100.0284		100.0216		100.0193	
2	100.0083	4.57	100.0073	5.08	100.0052	5.87	100.0047	6.04
3	100.0026	4.47	100.0020	4.04	100.0013	4.20	100.0011	4.11
4	100.0012	3.95	100.0007	4.00	100.0004	4.20	100.0002	4.10
95%-CI	[99.9995, 100.1225]		[99.9483, 100.0883]		[99.9679, 100.1073]		[99.9557, 100.0791]	

TABLE 7.4: *Results from validation test under the Heston model. Regarding the case of deterministic withdrawal, we set insurance fee $\beta = 61.77$ bps for ratchet case, while $\beta = 37.06$ bps for no ratchet case. Regarding the case of optimal withdrawal, we set insurance fee $\beta = 70.89$ bps for ratchet case, while $\beta = 64.28$ bps for no ratchet case.*

7.2 Modeling impact

In this subsection, we investigate the impact of stochastic volatility dynamics on quantities of central importance to GLWBs, namely no-arbitrage prices and fair insurance fees. The set of input parameters for the Heston model is given in Table 7.5. For comparison purposes, a GBM model is used with the comparable constant instantaneous volatility, denoted by σ_c, being square root of the long time mean θ of the variance process, and the constant interest rate being the same as the Heston dynamics. Since $\theta = 0.0225$ (Table 7.3), $\sigma_c = 0.15$. The numerical results presented below are obtained using the level 2 grid. Unless otherwise stated, the difference between the numerical results under the GBM and Heston model, denoted by $\Delta\%$, is computed by $\Delta\% = (x_H - x_G)/x_G$, where x_H and x_G are numerical results under the GBM and Heston model, respectively.

Parameter	GBM	Heston
Constant interest rate (r)	0.04	0.04
Constant volatility (σ_c)	0.2	N/A
Speed of mean reversion (λ)	N/A	1.0
Volatility of variance (ξ)	N/A	0.2
Correlation (ρ)	N/A	-0.5
Long time mean of variance (θ)	N/A	0.04
Initial variance (ν_0)	N/A	0.04
Penalty rate ($\mu(t)$)	$0 < t \leq 1 : 3\%$, $1 < t \leq 2 : 2\%$, $2 < t \leq 3 : 1\%$, $t > 3 : 0\%$	Same with GBM model
Withdrawal rate (C_r)	0.05	0.05
Bonus rate (b)	0.06	0.06
Ratchet cycle (f_r)	Every three years	Every three years

Table 7.5: *Input parameters of the base case used for the comparison between constant volatility and stochastic volatility.*

7.2.1 Contractual parameters

We present numerical results on contractual parameters, namely the annual withdrawal rate C_r (Table 7.6), the annual bonus rate b (Table 7.7), ratchet cycle f_r (Table 7.8), and the penalty rate $\mu(t)$ (Table 7.9). We make the observations below.

- Overall, although the differences in no-arbitrage prices of GLWBs tend to be negligible between the models, the differences in fair insurance fees can be relatively significant in certain cases, e.g. with a large annual withdrawal rate (Table 7.6, compare $C_r = 0.04$ vs $C_r = 0.07$) or bonus rate (Table 7.7, compare $b = 0.0$ vs $b = 0.09$). We now primarily focus on fair insurance fees.

- For both GBM and Heston models, the fair insurance fees increase as the annual withdrawal rate C_r increases. This is because of the additional cost of larger contractual withdrawal amounts. To measure sensitivity, we compute the sensitivity of the fair insurance fee with respect to C_r using finite differencing, i.e. $(\beta_f^{C_r=0.07} - \beta_f^{C_r=0.04})/(0.07 - 0.04)$. For the GBM model, the sensitivity quantity is around 2.58, which is higher than that of the Heston model (around 2.28). It is interesting to observe from Table 7.6 that, when C_r increases the fair insurance fees under the Heston model become significantly less than those under the GBM model.

- For both GBM and Heston models, the fair insurance fees go up as the annual bonus rate b goes up, which is due to extra bonus for zero withdrawals. To measure sensitivity, we approximate the sensitivity of the fair insurance fee with respect to b using finite differencing, i.e. $(\beta_f^{b=0.09} - \beta_f^{b=0.0})/(0.09 - 0.0)$. For the GBM model, the sensitivity quantity is around 0.13, which is higher than that of the Heston model (around 0.11).

- The differences in insurance fees between the GBM and Heston models seem relatively insensitive to the ratchet cycle and the penalty rate (Table 7.8 and Table 7.9).

7.2.2 Dynamics parameters

We present numerical results on dynamics parameters, namely speed of mean reversion rate λ (Table 7.10), volatility of the variance ξ (Table 7.11), correlation coefficient ρ (Table 7.12), and the long term mean of the variance θ (Table 7.13). We have the following observations.

- When the speed of mean reversion λ goes up, the difference in fair insurance fees between the GBM and Heston models becomes more and more insignificant.

Cases	No-arbitrage price (u)			Fair insurance fee (β_f)		
	GBM	Heston	$\Delta\%$	GBM	Heston	$\Delta\%$
$C_r = 0.04$	106.2254	106.5504	0.31%	0.0070	0.0072	3.31%
$C_r = 0.06$	117.9638	117.7306	-0.20%	0.0329	0.0305	-7.30%
$C_r = 0.07$	126.0965	125.4281	-0.53%	0.0843	0.0756	-10.27%

TABLE 7.6: *Comparison of no-arbitrage prices u and fair insurance fees β_f between the GBM and Heston models under different annual withdrawal rates C_r. No-arbitrage prices u are computed with the insurance fee $\beta = 0$ bps. Other input parameters are given in Table 7.5.*

Cases	No-arbitrage price (u)			Fair insurance fee (β_f)		
	GBM	Heston	$\Delta\%$	GBM	Heston	$\Delta\%$
$b = 0.0$	108.7413	108.8670	0.12%	0.0142	0.0139	-1.74%
$b = 0.03$	109.0018	109.1995	0.18%	0.0142	0.0139	-1.74%
$b = 0.09$	119.0236	118.5356	-0.41%	0.0254	0.0236	-7.19%

TABLE 7.7: *Comparison of no-arbitrage prices u and fair insurance fees β_f between the GBM and Heston models under different annual bonus rates b. No-arbitrage prices u are computed with the insurance fee $\beta = 0$ bps. Other input parameters are given in Table 7.5.*

Cases	No-arbitrage price (u)			Fair insurance fee (β_f)		
	GBM	Heston	$\Delta\%$	GBM	Heston	$\Delta\%$
$f_r = \infty$	108.5294	108.7951	0.24%	0.0143	0.0140	-2.23%
$f_r = 3$	111.2943	111.4136	0.11%	0.0149	0.0146	-2.21%
$f_r = 1$	113.2096	113.0171	-0.17%	0.0173	0.0165	-4.33%

TABLE 7.8: *Comparison of no-arbitrage prices u and fair insurance fees β_f between the GBM and Heston models under different ratchet cycles f_r. For instance, $f_r = \infty$ implies $\mathcal{T}_r = \emptyset$. No-arbitrage prices u are computed with the insurance fee $\beta = 0$ bps. Other input parameters are given in Table 7.5.*

Cases	No-arbitrage price (u)			Fair insurance fee (β_f)		
	GBM	Heston	$\Delta\%$	GBM	Heston	$\Delta\%$
$\mu(t) = \mu_1(t)$	111.2943	111.4136	0.11%	0.0156	0.0153	-1.96%
$\mu(t) = \mu_2(t)$	111.2943	111.4136	0.11%	0.0149	0.0146	-2.21%
$\mu(t) = \mu_3(t)$	111.2943	111.4136	0.11%	0.0144	0.0140	-2.59%
$\mu(t) = \mu_4(t)$	111.2943	111.4136	0.11%	0.0138	0.0134	-2.73%

TABLE 7.9: *Comparison of no-arbitrage prices u and fair insurance fees β_f between the GBM and Heston models under different penalty rate $\mu(t)$, where $\mu_1(t)$ is defined to be 2% if $0 < t \le 1$, 1% if $1 < t \le 2$, otherwise 0%; $\mu_2(t)$ is defined to be 3% if $0 < t \le 1$, 2% if $1 < t \le 2$, 1% if $2 < t \le 3$, otherwise 0%; $\mu_3(t)$ is defined to be 4% if $0 < t \le 1$, 3% if $1 < t \le 2$, 2% if $2 < t \le 3$, 1% if $3 < t \le 4$, otherwise 0%; $\mu_4(t)$ is defined to be 5% if $0 < t \le 1$, 4% if $1 < t \le 2$, 3% if $2 < t \le 3$, 2% if $3 < t \le 4$, 1% if $4 < t \le 5$, otherwise 0%. No-arbitrage prices u are computed with the insurance fee $\beta = 0$ bps. Other input parameters are given in Table 7.5.*

- The difference in insurance fees between the GBM and Heston models seems a bit insensitive to the coefficient of correlation.
- When the volatility of variance goes up, the difference in fair insurance fees between the GBM and Heston models becomes more and more significant.
- The difference in insurance fees between the GBM and Heston models seems quite sensitive to the level of the long-term mean of variance, specially when the level of the long-term mean is relatively small.

Cases	No-arbitrage price (u)			Fair insurance fee (β_f)		
	GBM	Heston	$\Delta\%$	GBM	Heston	$\Delta\%$
$\lambda = 0.5$	111.2943	111.0473	-0.22%	0.0149	0.0142	-4.80%
$\lambda = 1.0$		111.4136	0.11%		0.0146	-2.21%
$\lambda = 1.5$		111.4639	0.15%		0.0147	-1.17%
$\lambda = 2.0$		111.4614	0.15%		0.0148	-0.70%

Table 7.10: *Comparison of no-arbitrage prices u and fair insurance fees β_f between the GBM and Heston models under different speeds of mean reversion rate λ. No-arbitrage prices u are computed with the insurance fee $\beta = 0$ bps. Other input parameters are given in Table 7.5.*

Cases	No-arbitrage price (u)			Fair insurance fee (β_f)		
	GBM	Heston	$\Delta\%$	GBM	Heston	$\Delta\%$
$\xi = 0.1$	111.2943	111.4068	0.10%	0.0149	0.0148	-0.45%
$\xi = 0.2$		111.4136	0.11%		0.0146	-2.21%
$\xi = 0.3$		111.3190	0.02%		0.0142	-4.95%
$\xi = 0.4$		111.1416	-0.14%		0.0137	-8.35%

Table 7.11: *Comparison of no-arbitrage prices u and fair insurance fees β_f between the GBM and Heston models under different volatilities of the variance ξ. No-arbitrage prices u are computed with the insurance fee $\beta = 0$ bps. Other input parameters are given in Table 7.5.*

Cases	No-arbitrage price (u)			Fair insurance fee (β_f)		
	GBM	Heston	$\Delta\%$	GBM	Heston	$\Delta\%$
$\rho = -0.1$	111.2943	111.2398	-0.05%	0.0149	0.0146	-2.09%
$\rho = -0.3$		111.3343	0.04%		0.0146	-2.08%
$\rho = -0.5$		111.4136	0.11%		0.0146	-2.21%
$\rho = -0.7$		111.4782	0.17%		0.0145	-2.46%

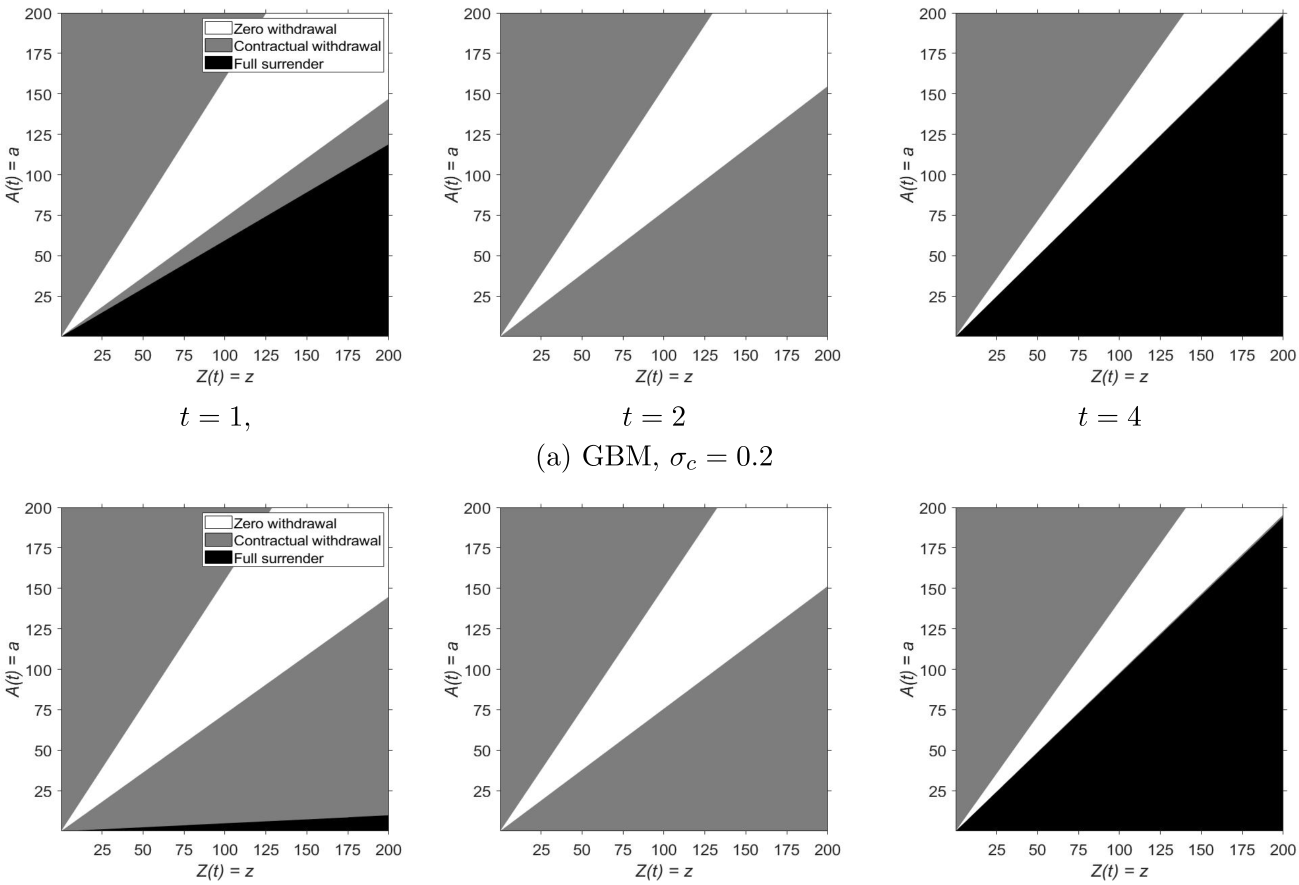
Table 7.12: *Comparison of no-arbitrage prices u and fair insurance fees β between the GBM and Heston models under different coefficients of correlation ρ. No-arbitrage prices u are computed with the insurance fee $\beta = 0$ bps. Other input parameters are given in Table 7.5.*

Cases	No-arbitrage price (u)			Fair insurance fee (β_f)		
	GBM	Heston	$\Delta\%$	GBM	Heston	$\Delta\%$
$\theta = 0.01$	102.8228	103.1047	0.27%	0.0033	0.0036	7.58%
$\theta = 0.04$	111.2943	111.4136	0.11%	0.0149	0.0146	-2.08%
$\theta = 0.09$	123.1994	122.9584	-0.20%	0.344	0.0329	-4.21%
$\theta = 0.16$	137.2857	136.5932	-0.50%	0.0607	0.0583	-3.93%

Table 7.13: *Comparison of no-arbitrage prices u and fair insurance fees β between the GBM and Heston models under different long time mean of the variance θ, with $\sigma_c^2 = \nu_0 = \theta$. No-arbitrage prices u are computed with the insurance fee $\beta = 0$ bps. Other input parameters are given in Table 7.5.*

7.3 Optimal withdrawals

We now turn our attention to optimal withdrawal strategies. In this study, we use the fair insurance fees for the GBM and Heston models, respectively denoted by β_f^G and β_f^H. We again use the input parameters reported in Table 7.5, and $\beta_f^G = 0.0149$ and $\beta_f^H = 0.0146$. In Figure 7.1, we present plots of optimal withdrawals for (calendar) time $t \in \{1, 2, 4\}$ (years) under the GBM model and the Heston model. Here, to ensure a fair comparison between the two models, since $\sigma_c = \sqrt{\theta} = \sqrt{0.04} = 0.2$, for



$t = 1,$ $t = 2$ $t = 4$

(a) GBM, $\sigma_c = 0.2$

$t = 1$ $t = 2$ $t = 4$

(b) Heston, $\sigma_k = 0.196$, $k \in \mathbb{K}$

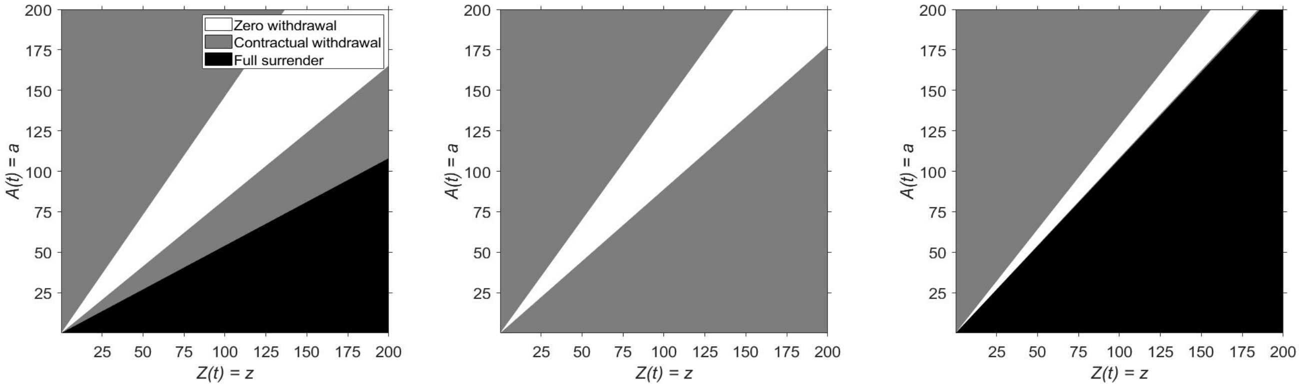
FIGURE 7.1: *The holder's optimal withdrawals at (calendar year) time $t \in \{1, 2, 4\}$; parameters specified in Table 7.5; fair insurance fees are $\beta_f^G = 0.0149$ and $\beta_f^H = 0.0146$; refinement level 2.*

the Heston model, we use the optimal withdrawals corresponding to the discretized σ closest to 0.2. In this case, the closest one is $\sigma_k = 0.196$, for some $k \in \mathbb{K}$. It is observed from Figure 7.1 that the optimal withdrawals at $t = 1$ under the two models are quite different in the full surrender decision; however, they are similar for $t = 2$ and $t = 4$. The holder should be more cautious about full surrender under stochastic volatility early year, possibly due to an early crash of the sub-account balance might significantly reduce the present value of the whole cash flow.

In Figure 7.2, we present plots of optimal withdrawals for (calendar) time $t \in \{1, 2, 4\}$ (years) under the Heston model corresponding to different discretized σ values, in particular for two extreme volatility scenarios, namely very low in which case $\sigma_k = 0.0013$, and very high in which case $\sigma_k = 0.6324$, for some $k \in \mathcal{K}$. Comparing across Figure 7.1 and Figure 7.2 for the same year, it observed that the optimal withdrawals tend to dictate that the policy holder should be more cautious under stochastic variance dynamics than he/she is under GBM dynamics where the volatility is constant, especially with the full surrender decision. We also observe that early in the policy $t = 1$, the holder should be extremely careful with the full surrender decision. Compare Figure 7.1 GBM, Figure 7.1 Heston, Figure 7.2 Heston $\sigma_k = 0.0013$ (low volatility) and Figure 7.2 Heston $\sigma_k = 0.6324$ (high volatility), all at $t = 1$.

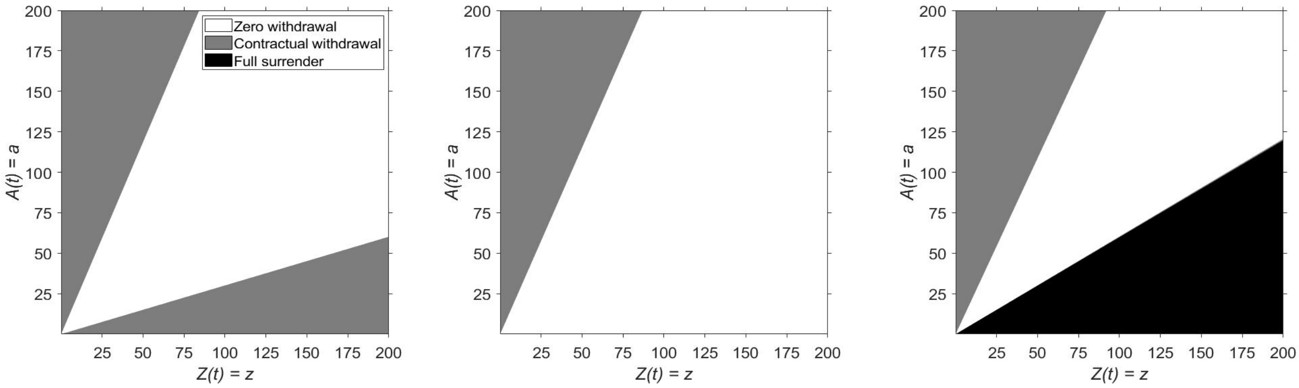
8 Conclusion

In this paper, we develop an ϵ-monotone numerical integration method for the no-arbitrage price of GLWB contracts with discrete withdrawals and the CIR dynamics for the variance of the personal sub-account. The pricing problem is formulated as a double integral, the inner of which is a convolution



$t = 1$ $t = 2$ $t = 4$

(a) Heston, $\sigma_k = 0.013$, $k \in \mathbb{K}$



$t = 1$ $t = 2$ $t = 4$

(b) Heston, $\sigma_k = 0.6324$, $k \in \mathbb{K}$

FIGURE 7.2: *The holder's optimal withdrawals at (calendar year) time* $t \in \{1, 2, 4\}$*; parameters specified in Table 7.5; fair insurance fee* $\beta_f^H = 0.0146$*; refinement level 2.*

integral involving a conditional density of the balance of the sub-account taking the form of a convolution kernel, while the outer one is a definite integral that involves a conditional density of the variance. We propose an efficient implementation of the inner integral via FFT, including proper handling of boundary conditions and padding techniques. We rigorously prove the convergence of the proposed scheme to the unique solution to the GLWB pricing problem as the discretization parameter and the monotonicity tolerance ϵ approach zero. Although we focus specifically on GLWB, our comprehensive and systematic approach could serve as a numerical and convergence analysis framework for the development of similar weakly monotone integration methods for control problems in finance.

References

[1] Milton Abramowitz, Irene A Stegun, and Robert H Romer. *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* American Association of Physics Teachers, 1988.

[2] J. Alonso-Garcca, O. Wood, and J. Ziveyi. Pricing and hedging guaranteed minimum withdrawal benefits under a general Lévy framework using the COS method. *Quantitative Finance*, 18:1049–1075, 2018.

[3] Leif BG Andersen. Efficient simulation of the heston stochastic volatility model. *Available at SSRN 946405*, 2007.

[4] Leif BG Andersen and Vladimir V Piterbarg. Moment explosions in stochastic volatility models. *Finance and Stochastics*, 11(1):29–50, 2007.

[5] P. Azimzadeh and P. Forsyth. The existence of optimal bang-bang controls for GMxB contracts. *SIAM Journal on Financial Mathematics*, pages 117–139, 02 2015.

[6] P. Azimzadeh, P.A. Forsyth, and K. Vetzal. *Hedging Costs for Variable Annuities Under Regime-Switching*, volume 209, chapter 6, pages 133–166. Springer International Series in Operations Research and Management, 04 2014.

[7] A.R. Bacinello, P. Millossovich, A. Olivieri, and E. Pitacco. Variable annuities: A unifying valuation approach. *Insurance: Mathematics and Economics*, 49(3):285–297, 2011.

[8] G. Barles and C. Imbert. Second-order elliptic integro-differential equations: viscosity solutions' theory revisited. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 25(3):567–585, 2008.

[9] D. Bauer, A. Kling, and J. Russ. A universal pricing framework for guaranteed minimum benefits in variable annuities. *ASTIN Bulletin*, 38:621–651, 2008.

[10] Eric Benhamou and Pierre Gauthier. Impact of stochastic interest rates and stochastic volatility on variable annuities. In *Paris December 2009 Finance International Meeting AFFI-EUROFIDAI*. Citeseer, 2009.

[11] Edouard Berthe, Duy-Minh Dang, and Luis Ortiz-Gracia. A Shannon wavelet method for pricing foreign exchange options under the Heston multi-factor CIR model. *Applied Numerical Mathematics*, 136:1–22, 2019.

[12] Ignace Bogaert. Iteration-free computation of Gauss–Legendre quadrature nodes and weights. *SIAM Journal on Scientific Computing*, 36(3):A1008–A1026, 2014.

[13] Hans Bremermann. *Distributions, complex variables, and Fourier transforms*. Addison-Wesley Publishing Company, 1965.

[14] Peter Christoffersen, Kris Jacobs, and Karim Mimouni. Volatility dynamics for the s&p500: Evidence from realized volatility, daily returns, and option prices. *The Review of Financial Studies*, 23(8):3141–3189, 2010.

[15] Matthew Coleman, Rick Hayes, Kendrick Lombardo, and Amber Ruiz. Variable annuities: Market pressures push the case for model sophistication. `https://www.wtwco.com/en-US/insights/2018/12/variable-annuities-market-pressures-push-the-case-for-model-sophistication`, 2019.

[16] John C Cox, Jonathan E Ingersoll Jr, and Stephen A Ross. A theory of the term structure of interest rates. In *Theory of valuation*, pages 129–164. World Scientific, 2005.

[17] K. Debrabant and E.R. Jakobsen. Semi-Lagrangian schemes for linear and fully non-linear diffusion equations. *Mathematics of Computation*, 82:1433–1462, 2013.

[18] M Drinkwater, I Iqbal, and J Montiminy. Variable annuity guaranteed living benefits utilization: 2012 experience. *A joint study sponsored by the Society of Actuaries and LIMRA*, 2014.

[19] F. Fang and C.W. Oosterlee. A novel pricing method for European options based on Fourier-Cosine series expansions. *SIAM Journal on Scientific Computing*, 31:826–848, 2008.

[20] F. Fang and C.W. Oosterlee. A Fourier-based valuation method for Bermudan and barrier options under Heston's model. *SIAM J. Financial Math.*, 2:439–463, 01 2011.

[21] Fang Fang and Cornelis W Oosterlee. A Fourier-based valuation method for Bermudan and barrier options under Heston's model. *SIAM Journal on Financial Mathematics*, 2(1):439–463, 2011.

[22] R. Feng and X. Jing. Analytical valuation and hedging of variable annuity guaranteed lifetime withdrawal benefits. *Insurance: Mathematics and Economics*, 72:36–48, 2017.

[23] P. Forsyth and G. Labahn. Numerical methods for controlled Hamilton-Jacobi-Bellman PDEs in finance. *Journal of Computational Finance*, 11 (Winter):1–44, 2008.

[24] P. Forsyth and G. Labahn. ϵ-monotone Fourier methods for optimal stochastic control in finance. *Journal of Computational Finance*, 22(4):25–71, 2019.

[25] P. Forsyth and K.R. Vetzal. An optimal stochastic control framework for determining the cost of hedging of variable annuities. *Journal of Economic Dynamics and Control*, 44:29–53, 07 2014.

[26] F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17:238–246, 1980.

[27] Peter Friz and Martin Keller-Ressel. Moment explosions in stochastic volatility models. *Encyclopedia of quantitative finance*, pages 1247–1253, 2010.

[28] Patrice Gaillardetz and Joe Youssef Lakhmiri. A new premium principle for equity-indexed annuities. *Journal of Risk and Insurance*, 78(1):245–265, 2011.

[29] L. Goudenége, A. Molent, and A. Zanette. Pricing and hedging GLWB in the Heston and in the Black-Scholes with stochastic interest rate models. *Insurance: Mathematics and Economics*, 70:38–57, 2016.

[30] X. Guo and G. Wu. Smooth fit principle for impulse control of multidimensional diffusion processes. *SIAM Journal on Control and Optimization*, 48(2):594–617, 2009.

[31] Michael T Heath. *Scientific computing: an introductory survey, revised second edition*. SIAM, 2018.

[32] S. Heston. A closed form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

[33] Daniela Holz, Alexander Kling, and Jochen Russ. Gmwb for life an analysis of lifelong withdrawal guarantees. *Zeitschrift für die gesamte Versicherungswissenschaft*, 101:305–325, 2012.

[34] Y.T. Huang and Y.K. Kwok. Regression-based Monte Carlo methods for stochastic control models: variable annuities with lifelong guarantees. *Quantitative Finance*, 16(6):905–928, 2016.

[35] Y.T. Huang, P. Zeng, and Y.K. Kwok. Optimal initiation of guaranteed lifelong withdrawal benefit with dynamic withdrawals. *SIAM Journal on Financial Mathematics*, 8:804–840, 2017.

[36] K. Ignatieva, A. Song, and J. Ziveyi. Fourier space time-stepping algorithm for valuing guaranteed minimum withdrawal benefits in variable annuities under regime-switching and stochastic mortality. *ASTIN Bulletin*, 48(1):139—-169, 2018.

[37] Zongxia Liang and Wenlong Sheng. Valuing inflation-linked death benefits under a stochastic volatility framework. *Insurance: Mathematics and Economics*, 69:45–58, 2016.

[38] Y. Lu and D.M. Dang. A semi-Lagrangian ϵ-monotone Fourier method for continuous withdrawal GMWBs under jump-diffusion with stochastic interest rate. `https://people.smp.uq.edu.au/Duy-MinhDang/papers/epsilon_GMWB_interest_rate.pdf`, 10 2022. Submitted.

[39] Y. Lu, D.M. Dang, P.A. Forsyth, and G. Labahn. An ϵ-monotone Fourier method for Guaranteed Minimum Withdrawal Benefit (GMWB) as a continuous impulse control problem. `https://people.smp.uq.edu.au/Duy-MinhDang/papers/epsilon_GMWB.pdf`, 06 2022. Submitted.

[40] X. Luo and P.V. Shevchenko. Valuation of variable annuities with guaranteed minimum withdrawal and death benefits via stochastic control optimization. *Insurance: Mathematics and Economics*, 62(3):5–15, 2015.

[41] K. Ma and P.A. Forsyth. An unconditionally monotone numerical scheme for the two-factor uncertain volatility model. *IMA Journal of Numerical Analysis*, 37(2):905–944, 2017.

[42] Patrick Kandege Mwanakatwe, Lixin Song, Emmanuel Hagenimana, and Xiaoguang Wang. Management strategies for a defined contribution pension fund under the hybrid stochastic volatility model. *Computational and Applied Mathematics*, 38:1–19, 2019.

[43] A.M. Oberman. Convergent difference schemes for degenerate elliptic and parabolic equations: Hamilton–Jacobi Equations and free boundary problems. *SIAM Journal Numerical Analysis*, 44(2):879–895, 2006.

[44] H. Pham. On some recent aspects of stochastic control and their applications. *Probability Surveys*, 2:506–549, 2005.

[45] D.M. Pooley, P.A. Forsyth, and K.R. Vetzal. Numerical convergence properties of option pricing PDEs with uncertain volatility. *IMA Journal of Numerical Analysis*, 23:241–267, 2003.

[46] M. Puterman. *Markov Decison Processes: Discrete Stochastic Dynamic Programming*. Wiley, New York, 1994.

[47] C. Reisinger and P.A. Forsyth. Piecewise constant Policy approximations to Hamilton-Jacobi-Bellman equations. *Applied Numerical Mathematics*, 103:27–47, 2016.

[48] M.J. Ruijter, C.W. Oosterlee, and R.F.T. Aalbers. On the Fourier cosine series expansion (COS) method for stochastic control problems. *Numerical Linear Algebra with Applications*, 20:598–625, 2013.

[49] R.C. Seydel. Existence and uniqueness of viscosity solutions for QVI associated with impulse control of jump-diffusions. *Stochastic Processes and Their Applications*, 119:3719–3748, 2009.

[50] P.V. Shevchenko and X. Luo. A unified pricing of variable annuity guarantees under the optimal stochastic control framework. *Risks*, 4(3):1–31, 2016.

[51] P. Steinorth and O.S. Mitchell. Valuing variable annuities with guaranteed minimum life time withdrawal benefits. *Insurance: Mathematics and Economics*, 64:246–258, 2015.

[52] Paul N Swarztrauber. On computing the points and weights for Gauss–Legendre quadrature. *SIAM Journal on Scientific Computing*, 24(3):945–954, 2003.

[53] P. M. Van Staden, D.M. Dang, and P.A. Forsyth. Time-consistent mean-variance portfolio optimization: a numerical impulse control approach. *Insurance: Mathematics and Economics*, 83(C):9–28, 2018.

[54] Pieter M Van Staden, Duy-Minh Dang, and Peter A Forsyth. Mean-quadratic variation portfolio optimization: A desirable alternative to time-consistent mean-variance optimization? *SIAM Journal on Financial Mathematics*, 10(3):815–856, 2019.

[55] Pieter M van Staden, Duy-Minh Dang, and Peter A Forsyth. On the distribution of terminal wealth under dynamic mean-variance optimal investment strategies. *SIAM Journal on Financial Mathematics*, 12(2):566–603, 2021.

[56] J. Wang and P.A. Forsyth. Maximal use of central differencing for Hamilton-Jacobi-Bellman PDEs in finance. *SIAM Journal on Numerical Analysis*, 46:1580–1601, 2008.

[57] X. Warin. Some non-monotone schemes for time dependent Hamilton-Jacobi-Bellman equations in stochastic control. *Journal of Scientific Computing*, 66:1122–1147, 03 2016.

[58] George Neville Watson. *A treatise on the theory of Bessel functions*, volume 3. The University Press, 1922.