Analysis of A Penalty Method for Pricing a Guaranteed Minimum Withdrawal Benefit (GMWB) *

Y. Huang † P.A. Forsyth ‡

July 21, 2010

Abstract

The no arbitrage pricing of Guaranteed Minimum Withdrawal Benefits (GMWB) contracts results in a singular stochastic control problem which can be formulated as a Hamilton Jacobi Bellman (HJB) Variational Inequality (VI). Recently, a penalty method has been suggested for solution of this HJB variational inequality (Dai et al., 2008). This method is very simple to implement. In this article, we present a rigorous proof of convergence of the penalty method to the viscosity solution of the HJB VI. Numerical tests of the penalty method are presented which show the experimental rates of convergence, and a discussion of the choice of the penalty parameter is also included.

Keywords: Singular stochastic control, HJB equation, viscosity solution, penalty method

AMS Classification 65N06, 93C20

1 Introduction

Stochastic control problems arise in many financial applications. For a survey of the literature on this topic, we refer to Pham (2005). When the set of possible admissible controls becomes unbounded, the control problem is said to be singular. A classical singular control problem in finance concerns optimal investment, where an infinite control corresponds to an instantaneous reallocation between a risky and risk-free asset (Tourin and Zariphopoulou, 1997). A survey of numerical methods for stochastic control is given in Kushner (2001) and Pham (2009). In this article, we focus on a singular stochastic control problem arising in the insurance industry, the Guaranteed Minimum Withdrawal Benefit (GMWB). Although we specifically consider the GMWB pricing problem, the methods we analyze here can be easily applied to many other singular stochastic control problems in finance.

In general, the solutions of singular stochastic control problems in finance are not smooth (Pham, 2005). Hence, we seek the viscosity solution of such problems (Barles, 1997).

*This work was supported by the Natural Sciences and Engineering Research Council of Canada

†Department of Electrical and Computer Engineering, University of Waterloo, Waterloo ON, Canada N2L 3G1, yqhuang@ecemail.uwaterloo.ca

‡David R. Cheriton School of Computer Science, University of Waterloo, Waterloo ON, Canada N2L 3G1 paforsyt@uwaterloo.ca

The pricing problem for the GMWB guarantee was originally formulated as a singular stochastic control problem in Milevsky and Salisbury (2006), which results in a Hamilton Jacobi Bellman (HJB) Variational Inequality (VI). Chen and Forsyth (2008) develop a method to solve an impulse control formulation of this problem. Methods for cases where withdrawals are only allowed at discrete times are given in Bauer et al. (2008) and Chen et al. (2008). Recently, Dai et al. (2008) have suggested a penalty method for solution of the HJB variational inequality for this problem, which is a generalization of the penalty method used for American options (Forsyth and Vetzal, 2002). The penalty method has also been applied to a singular stochastic control formulation of the continuous time portfolio selection problem (Dai and Zhong, 2010). In (Dai et al., 2008; Dai and Zhong, 2010), numerical examples were given by the authors to show the convergence of the proposed penalty method. However no formal proof of convergence was given. The penalty method is extremely simple to implement, and hence merits further analysis. For a discussion of the advantages of the penalty method compared with other numerical methods for singular control problems, we refer the reader to Dai et al. (2008) and Dai and Zhong (2010).

The main contributions of this article are

- We carry out a rigorous analysis of the penalty method in the context of the GMWB HJB variational inequality. Assuming that the GMWB problem satisfies a strong comparison principle, we verify that the penalty method is consistent, stable and monotone. Hence from the results in (Barles and Souganidis, 1991; Barles, 1997) we deduce convergence to the viscosity solution of the GMWB HJB variational inequality.

- We use the method described in Wang and Forsyth (2008), where central differencing is used as much as possible, yet still results in a monotone scheme. This results in noticeably faster convergence (as the mesh is refined) compared to use of pure upwinding schemes.

- Based on financial reasoning, we suggest an estimate for the size of the constant in the penalty term. Numerical tests show that the solution is insensitive to the value of this constant over several orders of magnitude.

- We discuss the advantages and disadvantages, from a computational point of view, of the singular control formulation compared to the impulse control formulation of this problem.

2 The GMWB Pricing Problem

2.1 Motivation

It is conventional wisdom that the long term investor is better off investing in equities as opposed to risk free bonds, hence the advice to retirees to invest a significant portion of their savings in equities. However, as discussed in Milevsky and Salisbury (2006), investing in equities can be very risky, once retirees begin to draw down their savings. This is because the order of random returns in this case becomes significant. Losses during the early years of retirement, coupled with withdrawals, will have a very different end result compared with losses which occur during the later years of retirement.

In order to mitigate this risk, insurance companies have developed guaranteed minimum withdrawal benefit (GMWB) guarantees. This contract consists of a lump sum payment to an insurance company. This initial sum is invested in risky assets. The holder can withdraw a specified amount each year of the contract, regardless of the performance of the risky asset. The holder can also withdraw more than the contract amount, subject to a penalty. At expiry of the contract, the holder is entitled to the value of the investment amount remaining. This contract allows the holder to participate in market gains, while providing a certain minimum cash flow. In return for providing this guarantee, the insurance company receives a proportional fee.

2.2 Formulation

This section briefly reviews the singular control model formulated in Dai et al. (2008) and introduces the notation to be used in this article. Let $W \equiv W(t)$ be the amount in the variable annuity account and $A \equiv A(t)$ be the guarantee account balance. We assume that the risky asset S which underlies the variable annuity account (before the deduction of any proportional fees) follows a standard Brownian Motion under the risk neutral measure. To be more precise, S satisfies the following stochastic differential equation

$$dS = rSdt + \sigma SdZ, \tag{2.1}$$

where dZ is an increment of the standard Gauss-Wiener process, σ is the volatility, and r is the risk free rate.

The major feature of the GMWB is the guarantee on the return of the entire premium via withdrawal. The insurance company charges the policy holder a proportional annual insurance fee η, in return for providing this guarantee. Consequently, we have the following stochastic differential equation for W:

$$dW = \begin{cases} (r-\eta)Wdt + \sigma WdZ + dA & \text{if } W > 0, \\ 0 & \text{if } W = 0. \end{cases} \tag{2.2}$$

Let $\gamma \equiv \gamma(t)$ denote the withdrawal rate at time t and assume $\gamma \in [0, \infty)$. An infinite withdrawal rate corresponds to an instantaneous withdrawal of a finite amount. The policy guarantees that the sum of withdrawals throughout the policy's life is equal to the premium paid up front, which is denoted by ω_0. As a result, we have $A(0) = \omega_0$, and

$$A(t) = \omega_0 - \int_0^t \gamma(u)du, \quad A(t) \geq 0 \ . \tag{2.3}$$

In addition, almost all policies with a GMWB have a cap on the maximum allowed withdrawal rate without penalty. Let G be such a contractual withdrawal rate, and $\kappa < 1$ be the proportional penalty charge applied on the portion of the withdrawal exceeding G. The net withdrawal rate $f(\gamma)$ received by the policy holder is then

$$f(\gamma) = \begin{cases} \gamma & 0 \leq \gamma \leq G, \\ G + (1-\kappa)(\gamma - G) & \gamma > G. \end{cases} \tag{2.4}$$

The no arbitrage value $V(W, A, t)$ of the variable annuity with GMWB is therefore given by (Dai et al., 2008)

$$V(W, A, t) = \max_{\gamma \in [0,\infty)} E_t \left[e^{-r(T-t)} \max((1-\kappa)A(T), W(T)) + \int_t^T e^{-r(u-t)} f(\gamma(u))du \right] \ , \tag{2.5}$$

where T is the policy maturity time and the expectation E_t is taken under the risk neutral measure. The withdrawal rate γ is the control variable chosen to maximize the value of $V(W, A, t)$. Equation

(2.5) represents the expected, discounted risk neutral cash flows from the guarantee, as discussed in Dai et al. (2008).

With an abuse of notation, we now (and in the rest of this article) let $V = V(W, A, \tau = T - t)$. It is shown in Dai et al. (2008) that the variable annuity value $V(W, A, \tau)$ is given by the following Hamilton-Jacobi-Bellman (HJB) Variational Inequality (VI)

$$\min\Big[V_\tau - \mathcal{L}V - G\max(\mathcal{F}V, 0), \kappa - \mathcal{F}V \Big] = 0 \ , \tag{2.6}$$

where the operators $\mathcal{L}, \mathcal{F}$ are defined as

$$\begin{aligned} \mathcal{L}V &= \frac{\sigma^2}{2} W^2 V_{WW} + (r - \eta) W V_W - rV \\ \mathcal{F}V &= 1 - V_W - V_A \ . \end{aligned} \tag{2.7}$$

Equation (2.6) or the equivalent form (2.5) are commonly used by insurance firms to determine the no-arbitrage value of the GMWB contract. The solution is also used to determine a hedging strategy for the contract (Milevsky and Salisbury, 2006; Bauer et al., 2008; Chen et al., 2008; Gilbert et al., 2007; Fenton and Czernicki, 2010). Historically, it has also been argued that equation (2.6) assumes optimal behaviour of consumers, which is unlikely in practice. However, it is now considered prudent to price these contracts assuming optimal behaviour, so that a worst case hedge can be constructed (Cramer et al., 2007). For an extension of these models to cases involving sub-optimal consumer behaviour, see Chen et al. (2008).

2.3 Informal Derivation of the HJB VI

We repeat here the informal derivation of equation (2.6) given in Dai et al. (2008). We will use this to give some intuition for our numerical scheme. Suppose that we restrict the maximum withdrawal range to be in $\gamma \in [0, \lambda]$ with $\lambda > G$ finite. Let $\lambda = 1/\varepsilon$. Then it is shown in Dai et al. (2008) that the variable annuity value parameterized by ε, denoted by $V^\varepsilon(W, A, \tau)$ is given from the solution to the following Hamilton-Jacobi-Bellman (HJB) equation

$$V^\varepsilon_\tau = \mathcal{L}V^\varepsilon + \max_{\gamma \in [0,\lambda]} h(\gamma), \tag{2.8}$$

where $h(\gamma)$ is given by

$$\begin{aligned} h(\gamma) &= f(\gamma) - \gamma V^\varepsilon_W - \gamma V^\varepsilon_A \\ &= \begin{cases} (1 - V^\varepsilon_W - V^\varepsilon_A)\gamma & \text{if } 0 \le \gamma \le G, \\ (1 - V^\varepsilon_W - V^\varepsilon_A - \kappa)\gamma + \kappa G & \text{if } \gamma > G. \end{cases} \end{aligned} \tag{2.9}$$

An informal derivation of equation (2.8) using a hedging argument is given in Appendix A. The function $h(\gamma)$ is piecewise linear, so its maximum value is achieved when γ is 0, G, or λ. Assuming $\lambda > G$, we then have

$$\max_{\gamma \in [0,\lambda]} h(\gamma) = \begin{cases} 0 & \text{if } \mathcal{F}V^\varepsilon \le 0, \\ G\mathcal{F}V^\varepsilon & \text{if } 0 < \mathcal{F}V^\varepsilon < \kappa, \\ \lambda(\mathcal{F}V^\varepsilon - \kappa) + \kappa G & \text{if } \mathcal{F}V^\varepsilon \ge \kappa. \end{cases} \tag{2.10}$$

The first two cases for $\max_{\gamma \in [0,\lambda]} h(\gamma)$ in (2.10) are identical to $G \max(0, \mathcal{F}V^\varepsilon)$. Substituting (2.10) into (2.8), we obtain (with $\lambda = 1/\varepsilon$)

$$-V_\tau^\varepsilon + \mathcal{L}V^\varepsilon + \max\left[G \max(0, \mathcal{F}V^\varepsilon), \frac{(\mathcal{F}V^\varepsilon - \kappa)}{\varepsilon} + \kappa G\right] = 0 \ . \tag{2.11}$$

The value function $V^\varepsilon(W, A, \tau)$ is then the solution of

$$\min\left[V_\tau^\varepsilon - \mathcal{L}V^\varepsilon - G \max(0, \mathcal{F}V^\varepsilon), V_\tau^\varepsilon - \mathcal{L}V^\varepsilon - \kappa G + \frac{(\kappa - \mathcal{F}V^\varepsilon)}{\varepsilon}\right] = 0 \ . \tag{2.12}$$

We can rewrite (2.12) (since $\varepsilon > 0$) equivalently

$$\min\left[V_\tau^\varepsilon - \mathcal{L}V^\varepsilon - G \max(0, \mathcal{F}V^\varepsilon), \kappa - \mathcal{F}V^\varepsilon + \varepsilon\left(V_\tau^\varepsilon - \mathcal{L}V^\varepsilon - \kappa G\right)\right] = 0 \ . \tag{2.13}$$

Taking the limit $\varepsilon \to 0$ (which corresponds to an instantaneous withdrawal of a finite amount) gives the following HJB variational inequality

$$\min\left[V_\tau - \mathcal{L}V - G \max(\mathcal{F}V, 0), \kappa - \mathcal{F}V\right] = 0 \ . \tag{2.14}$$

Consequently, we can see, at least intuitively, that

$$\lim_{\varepsilon \to 0}\left\{V_\tau^\varepsilon - \mathcal{L}V^\varepsilon - \max\left[G \max(0, \mathcal{F}V^\varepsilon), \frac{(\mathcal{F}V^\varepsilon - \kappa)}{\varepsilon} + \kappa G\right]\right\} = 0 \tag{2.15}$$

is equivalent to equation (2.6). Keeping ε finite, we can rewrite equation (2.15) in *control form*

$$V_\tau^\varepsilon \;\;=\;\; \mathcal{L}V^\varepsilon + \max_{\substack{\varphi \in \{0,1\}, \psi \in \{0,1\} \\ \varphi\psi = 0}}\left[\varphi G \mathcal{F}V^\varepsilon + \psi\left(\frac{(\mathcal{F}V^\varepsilon - \kappa)}{\varepsilon} + \kappa G\right)\right] \ . \tag{2.16}$$

The basic idea of the penalty method is to discretize equation (2.16), and let $\varepsilon \to 0$ as the mesh and timesteps tend to zero. In a subsequent section, we will give a rigorous proof that this algorithm converges to the viscosity solution of equation (2.6), provided that equation (2.6) satisfies a strong comparison principle.

3 Boundary Conditions

The original GMWB problem is posed on the domain Ω^∞

$$(W, A, \tau) \in [0, \infty) \times [0, \omega_0] \times [0, T] \ . \tag{3.1}$$

For computational purposes, we define the GMWB problem on a finite computational domain, as in Dai et al. (2008),

$$\Omega^L = [0, W_{\max}] \times [0, \omega_0] \times [0, T] \ . \tag{3.2}$$

We will analyse the convergence of the numerical scheme to the problem defined on Ω^L. Later, we will show that by solving the GMWB problem on successively larger domains, we converge to a unique limiting solution as $W_{\max} \to \infty$. We will also confirm this from some numerical experiments.

3.1 The terminal and boundary conditions

Define the following sets of points $(W, A, \tau) \in \Omega^L$

$$\begin{aligned}
\Omega_{\tau^0} &= [0, W_{\max}] \times [0, \omega_0] \times \{0\} \ , \\
\Omega_{W_0} &= \{0\} \times (0, \omega_0] \times (0, T] \\
\Omega_{W_{\max}} &= \{W_{\max}\} \times [0, \omega_0] \times (0, T] \\
\Omega_{A_0} &= [0, W_{\max}) \times \{0\} \times (0, T] \\
\Omega_{in} &= \Omega^L \backslash \Omega_{\tau^0} \backslash \Omega_{W_0} \backslash \Omega_{W_{\max}} \backslash \Omega_{A_0} && (3.3) \\
\partial\Omega_{in} &= \Omega_{\tau^0} \cup \Omega_{W_0} \cup \Omega_{W_{\max}} \cup \Omega_{A_0} \ . && (3.4)
\end{aligned}$$

For $(W, A, \tau) \in \Omega_{in}$, we solve

$$\begin{aligned}
\min\Big[V_\tau - \mathcal{L}V - G \max(\mathcal{F}V, 0), \kappa - \mathcal{F}V \Big] &= 0 \\
& (W, A, \tau) \in \Omega_{in} \ . && (3.5)
\end{aligned}$$

As discussed in Dai et al. (2008), at maturity, the policy holder takes the remaining guarantee withdrawal net of penalty charge or the remaining balance of the personal account, whichever is greater. Therefore at $\tau = 0$, the terminal condition is

$$\begin{aligned}
V(W, A, \tau = 0) &= \max\Big[W, (1-\kappa)A \Big] \\
& (W, A, \tau) \in \Omega_{\tau^0} \ . && (3.6)
\end{aligned}$$

As $W \to 0$, $V_W \to 0$ (Dai et al., 2008) (since W must be nonnegative). Thus, at $W = 0$, equation (2.16) becomes

$$\begin{aligned}
\min\Big[V_\tau - rV - G \max(1 - V_A, 0), \kappa - (1 - V_A) \Big] &= 0 \\
& (W, A, \tau) \in \Omega_{W_0} \ . && (3.7)
\end{aligned}$$

As $W \to \infty$, according to Dai et al. (2008), the withdrawal guarantee becomes insignificant for W sufficiently large. More precisely, a straightforward financial argument shows that the exact boundary condition at $W_{\max}$ is

$$V(W_{\max}, A, \tau) = e^{-\eta\tau} W_{\max} \left(1 + O\left(\frac{\omega_0}{W_{\max}} \right) \right) \ . \qquad (3.8)$$

Therefore as in Dai et al. (2008), we impose the following condition at $W_{\max}$

$$\begin{aligned}
V(W_{\max}, A, \tau) &= e^{-\eta\tau} W_{\max} \ , \\
& (W, A, \tau) \in \Omega_{W_{\max}} \ . && (3.9)
\end{aligned}$$

As $A \to 0$, no withdrawal is possible, so the PDE becomes the following linear PDE (Chen and Forsyth, 2008)

$$\begin{aligned}
V_\tau &= \mathcal{L}V \\
& (W, A, \tau) \in \Omega_{A_0} \ . && (3.10)
\end{aligned}$$

Note that as discussed in (Dai et al., 2008), no boundary condition is required at $A = \omega_0$ due to hyperbolic nature of the variable A. Since equations (3.7), (3.10) can be solved without any knowledge of the solution in the interior of Ω^L, they are essentially Dirichlet conditions.

3.2 Compact Representation

We now write the GMWB problem in a compact form, which includes the terminal and boundary conditions as a single equation. Define vector $\mathbf{x} = (W, A, \tau)$, and let $DV(\mathbf{x}) = (V_W, V_A, V_\tau)$ and $D^2V(\mathbf{x}) = V_{WW}$, and the equation

$$F_{\Omega^L} V \equiv F(D^2V(\mathbf{x}), DV(\mathbf{x}), V(\mathbf{x}), \mathbf{x}) = 0, \mathbf{x} \in \Omega^L, \tag{3.11}$$

where operator $F_{\Omega^L} V$ is defined by

$$F_{\Omega^L} V = \begin{cases} F_{\text{in}} V \equiv F_{\text{in}}(D^2V(\mathbf{x}), DV(\mathbf{x}), V(\mathbf{x}), \mathbf{x}), & \mathbf{x} \in \Omega_{in}, \\ F_{A_0} V \equiv F_{A_0}(D^2V(\mathbf{x}), DV(\mathbf{x}), V(\mathbf{x}), \mathbf{x}), & \mathbf{x} \in \Omega_{A_0}, \\ F_{W_0} V \equiv F_{W_0}(DV(\mathbf{x}), V(\mathbf{x}), \mathbf{x}), & \mathbf{x} \in \Omega_{W_0}, \\ F_{W_{\max}} V \equiv F_{W_{\max}}(V(\mathbf{x}), \mathbf{x}), & \mathbf{x} \in \Omega_{W_{\max}}, \\ F_{\tau_0} V \equiv F_{\tau_0}(V(\mathbf{x}), \mathbf{x}), & \mathbf{x} \in \Omega_{\tau_0}, \end{cases} \tag{3.12}$$

with operators

$$F_{in} V = \min\left[V_\tau - \mathcal{L}V - G \max(\mathcal{F}V, 0), \kappa - \mathcal{F}V\right], \tag{3.13}$$

$$F_{A_0} V = V_\tau - \mathcal{L}V, \tag{3.14}$$

$$F_{W_0} V = \min\left[V_\tau + rV - G \max(1 - V_A, 0), \kappa - 1 + V_A\right], \tag{3.15}$$

$$F_{W_{\max}} V = V - e^{-\eta\tau} W, \tag{3.16}$$

$$F_{\tau^0} V = V - \max\left[W, (1-\kappa)A\right]. \tag{3.17}$$

Definition 3.1 (Singular Control GMWB Pricing Problem)**.** *The pricing problem for the GMWB guarantee using a singular control formulation is defined as*

$$F_{\Omega^L}(D^2V(\mathbf{x}), DV(\mathbf{x}), V(\mathbf{x}), \mathbf{x}) = 0 \ . \tag{3.18}$$

F_{Ω^L} is proper and degenerate elliptic (Jakobsen, 2010)

$$F_{\Omega^L}(D^2V(\mathbf{x}) + \delta, DV(\mathbf{x}), V(\mathbf{x}) + \rho, \mathbf{x}) \quad \le \quad F_{\Omega^L}(D^2V(\mathbf{x}), DV(\mathbf{x}), V(\mathbf{x}), \mathbf{x}) \ \ ; \ \ \forall \delta \ge 0, \rho \le 0 \tag{3.19}$$

since the coefficient of $D^2V(\mathbf{x})$ in F_{Ω^L} is non-positive, and the coefficient of $V(\mathbf{x})$ is non-negative. Note that F_{Ω^L} is discontinuous (Barles and Souganidis, 1991; Barles, 1997), since we include the boundary equations in F_{Ω^L}, which are in general not the limit of the equations from the interior.

In the following, let u^* (u_*) denote the upper (lower) semi-continuous envelope of the function $u : X \to \mathbb{R}$, where X is a closed subset of $\mathbb{R}^N$, such that

$$u^*(\hat{x}) = \limsup_{\substack{x \to \hat{x} \\ \hat{x}, x \in X}} u(x), \qquad u_*(\hat{x}) = \liminf_{\substack{x \to \hat{x} \\ \hat{x}, x \in X}} u(x). \tag{3.20}$$

In general, the solution to singular stochastic control problems are non-smooth, and we seek the viscosity solution.

Definition 3.2 (Viscosity Solution)**.** *A locally bounded function $V : \Omega^L \to \mathbb{R}$ is a viscosity subsolution (respectively supersolution) of* (3.18) *if and only if for all smooth test functions $\phi(\mathbf{x}) \in C^2$, and for all maximum (respectively minimum) points $\mathbf{x}$ of $V^* - \phi$ (respectively $V_* - \phi$), one has*

$$\begin{aligned} &(F_{\Omega^L})_*(D^2\phi(\mathbf{x}), D\phi(\mathbf{x}), V^*(\mathbf{x}), \mathbf{x}) \leq 0 \\ \Bigg(\textit{respectively} \quad &(F_{\Omega^L})^*(D^2\phi(\mathbf{x}), D\phi(\mathbf{x}), V_*(\mathbf{x}), \mathbf{x}) \geq 0 \Bigg) . \end{aligned} \tag{3.21}$$

A locally bounded function V is a viscosity solution if it is both a viscosity subsolution and a viscosity supersolution.

In Seydel (2009), it is shown that an impulse control formulation of the GMWB pricing problem satisfies a strong comparison principle. However, there does not seem to be a proof of this result for the singular control formulation of this problem. Dai et al. (2008) state but do not prove the comparison principle for equation (3.18). Let $\Gamma \subset \partial\Omega_{in}$, which is unspecified for the moment. We make the following assumption.

Assumption 3.1 (Strong Comparison)**.** *The GMWB singular control problem as given in Definition 3.1 satisfies a strong comparison result in the domain $\Omega_{in} \cup \Gamma$, $\Gamma \subset \partial\Omega_{in}$. Hence a unique continuous viscosity solution exists in $\Omega_{in} \cup \Gamma$.*

Remark 3.1. *We cannot in general hope for a continuous solution over the whole of Ω^L. It is possible that loss of boundary data can occur over parts of $\partial\Omega_{in}$. For example, for points near $\Omega_{W_{\max}}$, if it is optimal to withdraw a finite amount instantaneously, then the HJB equation degenerates to a first order equation, with outgoing characteristices. Hence the the boundary condition at some points in $\Omega_{W_{\max}}$ may be irrelevant, in the sense that the boundary condition at these points does not influence the interior solution.*

Pham (2005) discusses another case where singular control problems cannot be continuous over the entire closed solution domain. It may be the case that the terminal condition at Ω_{τ^0} is not compatible with the control problem in the sense that it may be optimal to immediately make a transaction the instant after $\tau = 0$. This would result in a discontinuity in the solution as $\tau \to 0$, from points in $\Omega^L \backslash \Omega_{\tau^0}$. However, this does not occur in our case, since it is never optimal to make an instantaneous withdrawal at $\tau = 0^+$, with the particular initial condition (3.6).

All these issues need to be addressed in proving a strong comparison property, in order to define precisely those regions in Γ we can expect a continuous, unique viscosity solution.

However, the location of Γ has little impact on the computational algorithm. The boundary data is either used or irrelevant. In all cases we can consider the computed solution as the limiting value approaching $\partial\Omega_{in}$ from the interior.

Remark 3.2. *Note that in the case that an asymptotic form of the solution as $W_{\max} \to \infty$ is not available, it is possible to impose an arbitrary boundary condition (satisfying certain growth conditions) and take the limit as $W_{\max} \to \infty$. This will converge to the viscosity solution in the unbounded domain, as shown in Barles et al. (1995).*

4 Discretized Equations

4.1 Penalty Form

We will discretize the penalty form of the equations (2.16) and show that the discrete equations converge to the viscosity solution of the problem in Definition 3.2. Using the notation $\mathcal{D}_{WW}V =$

V_{WW}, $\mathcal{D}_W V = V_W$ and $\mathcal{D}_A V = V_A$, in $(W, A, \tau) \in \Omega_{in} \cup \Omega_{A_0}$ we will discretize

$$\begin{aligned} V^{\varepsilon}_{\tau} &= \mathcal{L}V^{\varepsilon} + \max_{\substack{\varphi\in\{0,1\},\psi\in\{0,1\}\\ \varphi\psi=0}} \left[\varphi G \mathcal{F} V^{\varepsilon} + \psi \left(\frac{(\mathcal{F}V^{\varepsilon} - \kappa)}{\varepsilon} + \kappa G \right) \right] \\ & \quad (W, A, \tau) \in \Omega_{in} \cup \Omega_{A_0} \ . \end{aligned} \tag{4.1}$$

where

$$\mathcal{L}V^{\varepsilon} = \frac{\sigma^2}{2} W^2 \mathcal{D}_{WW} V^{\varepsilon} + (r - \eta) W \mathcal{D}_W V^{\varepsilon} - r V^{\varepsilon} \ , \tag{4.2}$$

$$\mathcal{F}V^{\varepsilon} = 1 - \mathcal{D}_W V^{\varepsilon} - \mathcal{D}_A V^{\varepsilon} \ , \tag{4.3}$$

and we understand that $\phi = \psi = 0$ in Ω_{A_0}. At $W = 0$, we discretize

$$\begin{aligned} V^{\varepsilon}_{\tau} &= -rV^{\varepsilon} + \max_{\substack{\varphi\in\{0,1\},\psi\in\{0,1\}\\ \varphi\psi=0}} \left[\varphi G (1 - \mathcal{D}_A V^{\varepsilon}) + \psi \left(\frac{(1 - \mathcal{D}_A V^{\varepsilon} - \kappa)}{\varepsilon} + \kappa G \right) \right] \\ & \quad (W, A, \tau) \in \Omega_{W_0} \ . \end{aligned} \tag{4.4}$$

4.2 Discretization of the Penalized Equations

We will discretize equation (4.1) and equation (4.4) in the domain $\Omega_{\text{in}} \cup \Omega_{A_0} \cup \Omega_{W_0}$. We use an unequally spaced grid in the W direction, given by $\{W_0, \ldots, W_i, \ldots, W_{i_{\max}}\}$. The nodes in the A direction are denoted by $\{A_0, \ldots, A_j, \ldots, A_{j_{\max}}\}$, where $W_0 = A_0 = 0$, $W_{i_{\max}} = W_{\max}$ and $A_{j_{\max}} = \omega_0$. We denote the n^{th} time-step by $\tau^n = n\Delta\tau$, with $N = T/\Delta\tau$. We will always assume that $W_{i_{\max}} \gg A_{j_{\max}}$.

Denote the approximate solution at (W_i, A_j, τ^n) by $V^n_{i,j}$. We use a standard three point finite difference method to approximate the $\mathcal{D}_{WW}V$ derivative. This approximation is second order for smoothly varying grid spacing. The $\mathcal{D}_A V$ derivative is approximated by a first order backward differencing method. The $\mathcal{D}_W V$ derivative is approximated by second order central differencing or first order forward/backward differencing. Let $\mathcal{D}^h_A$, $\mathcal{D}^h_W$ and $\mathcal{D}^h_{WW}$ (defined in Appendix B) denote the discretized first and second order partial differential operators. The discretized $\mathcal{L}$ and $\mathcal{F}$ operators can then be written as

$$\mathcal{L}^h V^n_{i,j} = \begin{cases} \frac{\sigma^2}{2} W_i^2 \mathcal{D}^h_{WW} V^n_{i,j} + (r - \eta) W_i \mathcal{D}^h_W V^n_{i,j} - r V^n_{i,j}, & (W_i, A_j, \tau^n) \in \Omega_{in} \cup \Omega_{A_0} \\ -rV^n_{i,j}, & (W_i, A_j, \tau^n) \in \Omega_{W_0} \end{cases} , \tag{4.5}$$

$$\mathcal{F}^h V^n_{i,j} = \begin{cases} 1 - \mathcal{D}^h_W V^n_{i,j} - \mathcal{D}^h_A V^n_{i,j}, & (W_i, A_j, \tau^n) \in \Omega_{in} \\ 1 - \mathcal{D}^h_A V^n_{i,j}, & (W_i, A_j, \tau^n) \in \Omega_{W_0} \\ 0, & (W_i, A_j, \tau^n) \in \Omega_{A_0} \end{cases} . \tag{4.6}$$

Using fully implicit time-stepping, equation (4.1) has the following discretized form

$$\begin{aligned} \frac{V^{n+1}_{i,j} - V^n_{i,j}}{\Delta\tau} &= \mathcal{L}^h V^{n+1}_{i,j} + \max_{\substack{\varphi\in\{0,1\},\psi\in\{0,1\}\\ \varphi\psi=0}} \left[\varphi G \mathcal{F}^h V^{n+1}_{i,j} + \psi \left(\frac{(\mathcal{F}^h V^{n+1}_{i,j} - \kappa)}{\varepsilon} + \kappa G \right) \right] \\ & \quad i = 0, 1, 2, \ldots, i_{\max} - 1, \ j = 0, 1, 2, \ldots, j_{\max}, \ n = 1, 2, \ldots, N \ , \end{aligned} \tag{4.7}$$

or equivalently

$$\frac{V_{i,j}^{n+1} - V_{i,j}^{n}}{\Delta\tau} = \max_{\substack{\varphi\in\{0,1\},\psi\in\{0,1\}\\ \varphi\psi=0}} \left[\mathcal{L}^h V_{i,j}^{n+1} + \varphi G \mathcal{F}^h V_{i,j}^{n+1} + \psi \left(\frac{(\mathcal{F}^h V_{i,j}^{n+1} - \kappa)}{\varepsilon} + \kappa G \right) \right]$$
$$i = 0, 1, 2, \ldots, i_{\max} - 1,\ j = 0, 1, 2, \ldots, j_{\max},\ n = 1, 2, \ldots, N \ , \qquad (4.8)$$

and finally (by expanding $\mathcal{L}^h$, $\mathcal{F}^h$ and $\mathcal{D}_A^h$ operators)

$$\frac{V_{i,j}^{n+1} - V_{i,j}^{n}}{\Delta\tau} = \max_{\substack{\varphi\in\{0,1\},\psi\in\{0,1\}\\ \varphi\psi=0}} \left[\mathcal{A}_{\varphi,\psi}^h V_{i,j}^{n+1} + p_{i,j}^{n+1}(\varphi,\psi) V_{i,j-1} + q_{i,j}^{n+1}(\varphi,\psi) \right] ,$$
$$i = 1, 2, \ldots, i_{\max} - 1,\ j = 1, 2, \ldots, j_{\max},\ n = 1, 2, \ldots, N \ , \qquad (4.9)$$

where

$$\mathcal{A}_{\varphi,\psi}^h V_{i,j}^n \equiv a_{i,j}^n(\varphi,\psi) \mathcal{D}_{WW}^h V_{i,j}^n + b_{i,j}^n(\varphi,\psi) \mathcal{D}_W^h V_{i,j}^n - c_{i,j}^n(\varphi,\psi) V_{i,j}^n \qquad (4.10)$$

and

$$\begin{aligned} a_{i,j}^n(\varphi,\psi) &= \tfrac{\sigma^2}{2} W_i^2 \ , & p_{i,j}^n(\varphi,\psi) &= \tfrac{(\varphi G + \frac{\psi}{\varepsilon})}{\Delta A_j^-} \ , \\ b_{i,j}^n(\varphi,\psi) &= (r-\eta) W_i - (\varphi G + \tfrac{\psi}{\varepsilon}) \ , & q_{i,j}^n(\varphi,\psi) &= \varphi G + \psi(\tfrac{1-\kappa}{\varepsilon} + \kappa G) \ , \\ c_{i,j}^n(\varphi,\psi) &= r + \tfrac{(\varphi G + \frac{\psi}{\varepsilon})}{\Delta A_j^-} \ , & \Delta A_j^- &= A_j - A_{j-1} \ . \end{aligned} \qquad (4.11)$$

Let

$$\{\varphi_{i,j}^n, \psi_{i,j}^n\} = \operatorname*{arg\,max}_{\substack{\varphi\in\{0,1\},\psi\in\{0,1\}\\ \varphi\psi=0}} \left[\mathcal{A}_{\varphi,\psi}^h V_{i,j}^n + p_{i,j}^n(\varphi,\psi) V_{i,j-1}^n + q_{i,j}^n(\varphi,\psi) \right] . \qquad (4.12)$$

Equation (4.9) becomes

$$\frac{V_{i,j}^{n+1} - V_{i,j}^{n}}{\Delta\tau} = \mathcal{A}_{\varphi_{i,j}^{n+1},\psi_{i,j}^{n+1}}^h V_{i,j}^{n+1} + p_{i,j}^{n+1}(\varphi_{i,j}^{n+1}, \psi_{i,j}^{n+1}) V_{i,j-1}^{n+1} + q_{i,j}^{n+1}(\varphi_{i,j}^{n+1}, \psi_{i,j}^{n+1}) \ ,$$
$$i = 1, 2, \ldots, i_{\max} - 1,\ j = 1, 2, \ldots, j_{\max},\ n = 1, 2, \ldots, N \ . \qquad (4.13)$$

The discretized $\mathcal{D}_W^h V_{i,j}^n$ term in $\mathcal{A}_{\varphi,\psi}^h V_{i,j}^n$ can be obtained by applying central, forward, or backward differencing to the $\mathcal{D}_W V^\varepsilon$ term. A few steps of algebra show that the $\mathcal{A}_{\varphi,\psi}^h$ operator can also be written equivalently as

$$\mathcal{A}_{\varphi_{i,j}^n,\psi_{i,j}^n}^h V_{i,j}^n = \alpha_{i,j}^n V_{i-1,j}^n - (\alpha_{i,j}^n + \beta_{i,j}^n + c_{i,j}^n) V_{i,j}^n + \beta_{i,j}^n V_{i+1,j}^n,$$
$$i = 1, 2, \ldots, i_{\max} - 1,\ j = 1, 2, \ldots, j_{\max},\ n = 1, 2, \ldots, N \ . \qquad (4.14)$$

The $\alpha_{i,j}^n$ and $\beta_{i,j}^n$ in (4.14) are determined by the differencing method used in W direction, $\alpha_{i,j}^n \in \{\alpha_{i,j,cent}^n, \alpha_{i,j,for/back}^n\}$, $\beta_{i,j}^n \in \{\beta_{i,j,cent}^n, \beta_{i,j,for/back}^n\}$, which are defined in Appendix C. We use central differencing as much as possible in the W direction to ensure that the positive coefficient condition is satisfied (see Pooley et al. (2003))

$$\alpha_{i,j}^n \geq 0 \ \ ; \ \ \beta_{i,j}^n \geq 0 \ . \qquad (4.15)$$

Because $c_{i,j}^n \geq 0$ always holds, condition (4.15) is a sufficient condition to ensure a positive coefficient discretization scheme. Note that different nodes may use different differencing schemes.

By applying forward or backward differencing to $\mathcal{D}_W V^\varepsilon$ in the equation (4.1), the positive coefficient condition is guaranteed. In Dai et al. (2008), central differencing is used on V_W term in $\mathcal{L}V$ and backward differencing is used on V_W term in $\mathcal{F}V$. This requires a grid spacing condition in order to satisfy the positive coefficient condition. Because backward differencing in $\mathcal{F}V$ gives a first order truncation error in the W direction, whereas central differencing is second order correct (for smooth functions), we would like to use central differencing as much as possible on the V_W term both in $\mathcal{L}V$ and $\mathcal{F}V$. However, we must ensure that the positive coefficient condition (4.15) is satisfied. To use central differencing on the $\mathcal{D}_W V^\varepsilon$ term and maintain a positive coefficient condition at the same time, we require

$$\frac{1}{W_i - W_{i-1}} \quad \geq \quad \frac{(r_i - \eta) - \frac{(\varphi_{i,j}^{n+1} G + \frac{\psi_{i,j}^{n+1}}{\varepsilon})}{W_i}}{\sigma^2 W_i} \; ; \tag{4.16}$$

$$\frac{1}{W_{i+1} - W_i} \quad \geq \quad -\frac{(r_i - \eta) - \frac{(\varphi_{i,j}^{n+1} G + \frac{\psi_{i,j}^{n+1}}{\varepsilon})}{W_i}}{\sigma^2 W_i} \; . \tag{4.17}$$

In Wang and Forsyth (2008), the authors discussed maximal use of central differencing for HJB PDEs. Note that the differencing method to be used at a given node depends on the value of control parameters. At a given node, for a given control parameter value, we first try to discretize the $\mathcal{D}_W V^\varepsilon$ term by using central differencing. If this gives positive coefficients as described in (4.15), central differencing will be used for the node for this given control parameter value. Otherwise, either forward or backward differencing will be used for the node given this control parameter value. In our case, since we have three possible control parameter values, at each node, we determine the differencing method for each one of the three control parameter values. The local optimization criterion in (4.12) subsequently determines which control parameter value is the optimal value. The differencing method corresponding to this optimal control parameter value is then chosen to discretize the equation for the given node. Note that it is shown in Appendix C that at least one of central, forward or backward differencing must result in a positive coefficient scheme.

Equation (4.13) holds for $(W_i, A_j, \tau^{n+1}) \in \Omega_{in}$. The discrete forms of equations (3.6), (3.9), (3.10) and (4.4) are as follows. For $(W_i, A_j, \tau^{n+1}) \in \Omega_{\tau^0}$, $(\tau^n = 0)$ we have simply

$$V_{i,j}^0 = \max[W_i, (1-\kappa)A_j] \; . \tag{4.18}$$

In the region $(W_i, A_j, \tau^{n+1}) \in \Omega_{W_0}$ condition (4.4) is imposed by using equation (4.13) with

$$\alpha_{0,j}^{n+1} = \beta_{0,j}^{n+1} = 0, j = 1, \ldots, j_{\max}. \tag{4.19}$$

For $(W_i, A_j, \tau^{n+1}) \in \Omega_{A_0}$, condition (3.10) is imposed by using equation (4.13) with

$$\begin{aligned} \varphi_{i,0}^{n+1} &= \psi_{i,0}^{n+1} = 0 \;\; ; \;\; i = 0, 1, \ldots, i_{\max} - 1. \\ \alpha_{i,0}^{n+1} &= \beta_{i,0}^{n+1} = 0 \;\; ; \;\; i = 0 \; . \end{aligned} \tag{4.20}$$

At $W = W_{i_{\max}}$, or $(W_i, A_j, \tau^{n+1}) \in \Omega_{W_{\max}}$, we have (from equation (3.9))

$$V_{i_{\max},j}^{n+1} e^{\eta \Delta\tau} = V_{i_{\max},j}^n, \tag{4.21}$$

assuming $V^0_{i_{\max},j} = W_{\max}$. By setting

$$c^{n+1}_{i_{\max},j} = \eta \ \ ; \ \ \alpha^{n+1}_{i_{\max},j} = \beta^{n+1}_{i_{\max},j} = \varphi^{n}_{i_{\max},j} = \psi^{n}_{i_{\max},j} = 0; \quad j = 0, 1, \ldots, j_{\max}, \tag{4.22}$$

in equation (4.13) we obtain

$$V^{n+1}_{i_{\max},j}(1 + \eta \Delta\tau) = V^{n}_{i_{\max},j} \tag{4.23}$$

which is a second order approximation to equation (4.21). Consequently, at all points $(W_i, A_j, \tau^{n+1}) \in \Omega^L \backslash \Omega_{\tau^0}$, an equation of the form (4.13) holds, if we define $V_{-1,j} = V_{i_{\max}+1,j} = V_{i,-1} = 0$.

It will also prove useful to write equation (4.13) in the form

$$(1 - \Delta\tau \mathcal{A}^h_{\varphi^{n+1}_{i,j},\psi^{n+1}_{i,j}})V^{n+1}_{i,j} - \Delta\tau \ p^{n+1}_{i,j}(\varphi^{n+1}_{i,j}, \psi^{n+1}_{i,j})V^{n+1}_{i,j-1} = V^n_{i,j} + \Delta\tau q^{n+1}_{i,j}(\varphi^{n+1}_{i,j}, \psi^{n+1}_{i,j}) \ , \tag{4.24}$$

where, for future reference, we note from equation (4.11) that

$$p^{n+1}_{i,j}(\varphi^{n+1}_{i,j}, \psi^{n+1}_{i,j}) \geq 0 \ ; \ q^{n+1}_{i,j}((\varphi^{n+1}_{i,j}, \psi^{n+1}_{i,j}) \geq 0 \ . \tag{4.25}$$

4.3 The matrix form of the discrete equations

It is convenient to use a matrix form to represent the discretized equations. In this section we define a number of matrices and vectors to represent the discretized PDE in (4.13). Define vectors

$$\begin{aligned} \mathbf{v}^n_j &= \left(V^n_{0,j}, V^n_{1,j}, \ldots, V^n_{i_{\max},j}\right) \\ \mathbf{v}^n &= \left(\mathbf{v}^n_0, \mathbf{v}^n_1, \ldots, \mathbf{v}^n_{j_{\max}}\right) \ . \end{aligned} \tag{4.26}$$

Define the $(i_{\max}+1) \times (i_{max}+1)$ tridiagonal matrix $\mathbf{A}^n_j$ so that the element on the i^{th} row and k^{th} column is defined as

$$[\mathbf{A}^n_j]_{i,k} = \begin{cases} \alpha^n_{i,j} & \text{if } k = i-1, i = 1, \ldots, i_{\max} \\ \beta^n_{i,j} & \text{if } k = i+1, i = 0, \ldots, i_{\max}-1 \\ -(\alpha^n_{i,j} + \beta^n_{i,j} + c^n_{i,j}) & \text{if } k = i, i = 0, \ldots, i_{\max} \\ 0 & \text{otherwise .} \end{cases} \tag{4.27}$$

Define a diagonal $(i_{\max}+1) \times (i_{max}+1)$ matrix $\mathbf{P}^n_j$ so that elements on the diagonal are defined as

$$[\mathbf{P}^n_j]_{i,i} = \begin{cases} p^n_{i,j} & \text{if } i \leq i_{max} - 1, \\ 0 & \text{if } i = i_{max}. \end{cases} \tag{4.28}$$

Let vectors $\mathbf{q}^n_j$ and $\mathbf{q}^n$ be defined by

$$\begin{aligned} \mathbf{q}^n_j &= \left(q^n_{0,j}, q^n_{1,j}, \ldots, q^n_{i_{\max}-1,j}, 0\right) \ ; \\ \mathbf{q}^n &= \left(\mathbf{q}^n_0, \mathbf{q}^n_1, \ldots, \mathbf{q}^n_{j_{\max}}\right) \ . \end{aligned} \tag{4.29}$$

We can write equation (4.13) as

$$\left[\mathbf{I} - \Delta\tau \mathbf{A}^{n+1}_j\right] \mathbf{v}^{n+1}_j = \mathbf{v}^n_j + \Delta\tau \mathbf{P}^{n+1}_j \mathbf{v}^{n+1}_{j-1} + \Delta\tau \mathbf{q}^{n+1}_j \ , \tag{4.30}$$

where $\{\varphi_{i,j}^{n+1}, \psi_{i,j}^{n+1}\} =$

$$\underset{\substack{\varphi\in\{0,1\},\psi\in\{0,1\}\\ \varphi\psi=0}}{\arg\max} \left[\mathbf{A}_j^{n+1}(\varphi,\psi)\mathbf{v}_j^{n+1} + \mathbf{P}_j^{n+1}(\varphi,\psi)\mathbf{v}_{j-1}^{n+1} + \mathbf{q}_j^{n+1}(\varphi,\psi)\right]_i . \tag{4.31}$$

For notational completeness, we adopt the convention that $\mathbf{v}_{-1}^{n+1} = 0$. Note that $\mathbf{A}_j^{n+1} = \mathbf{A}_j^{n+1}(\varphi,\psi)$, $\mathbf{P}_j^{n+1} = \mathbf{P}_j^{n+1}(\varphi,\psi)$, $\mathbf{q}_j^{n+1} = \mathbf{q}_j^{n+1}(\varphi,\psi)$, through the local optimization problem (4.31). An exception occurs at $j = 0$, where $\mathbf{P}_0^n$ is a zero matrix and $\mathbf{q}_0^n$ is a zero vector. $\mathbf{A}_0$ no longer depends on the value of the control variables $\{\varphi,\psi\}$ or time $n\Delta\tau$ due to the boundary condition at $A = 0$. The matrix form of the degenerate equations becomes

$$[\mathbf{I} - \Delta\tau\mathbf{A}_0]\,\mathbf{v}_0^{n+1} = \mathbf{v}_0^n \tag{4.32}$$

on the boundary $j = 0$ (i.e. $A = 0$).

4.4 The algorithm

We use a policy iteration (Forsyth and Labahn, 2008) to solve the discretized PDE in (4.30). Let $(\mathbf{v}_j^{n+1})^m$ be the m^{th} estimate for $\mathbf{v}_j^{n+1}$ and the initial value of $(\mathbf{v}_j^{n+1})^0 = \mathbf{v}_j^n$. Algorithm 1 gives the details of the iterative technique. It is worthy to elaborate here that when $j \neq 0$, $\mathbf{A}_j^{n+1}$, $\mathbf{P}_j^{n+1}$ and $\mathbf{q}_j^{n+1}$ are dependent on the values of control variables $\{\varphi_{i,j}^{n+1}, \psi_{i,j}^{n+1}\}$. The approach we take is that when computing the values of $(\mathbf{v}_{i,j}^{n+1})^{m+1}$ at the m^{th} iteration, we construct $(\mathbf{A}_j^{n+1})^m$, $(\mathbf{P}_j^{n+1})^m$ and $(\mathbf{q}_j^{n+1})^m$ by using $\{(\varphi_{i,j}^{n+1})^m, (\psi_{i,j}^{n+1})^m\}$, the m^{th} estimate of $\{\varphi_{i,j}^{n+1}, \psi_{i,j}^{n+1}\}$, which is computed by using the value of $(\mathbf{v}_j^{n+1})^m$.

Note that in order to compute $\{(\varphi_{i,j}^{n+1})^m, (\psi_{i,j}^{n+1})^m\}$ in line 5 of Algorithm 1, We need to evaluate the local optimization objective function with all possible control parameter values. In our case, we only have three possible values because $\varphi,\psi \in \{0,1\}$ and $\varphi\psi = 0$.

The scale factor *scale* in Algorithm 1 should be of the same magnitude as the value of the annuity to be priced to avoid unrealistic levels of accuracy. If the annuity is to be priced in dollars, $scale = 1$ is a reasonable choice.

Theorem 4.1 (Convergence of the Policy Iteration)**.** *If the positive coefficient condition is satisfied, then the policy iteration in Algorithm 1 converges to the unique solution of (4.30) for any initial estimate* $(\mathbf{v}_j^{n+1})^0$.

Proof. This follows using similar steps as in Forsyth and Labahn (2008) and Wang and Forsyth (2008). Since we will refer to some of the properties of this proof in the numerical results section, we give a brief sketch here. Let

$$\mathbf{d}_j^m = (\mathbf{P}_j^{n+1})^m\mathbf{v}_{j-1}^{n+1} + (\mathbf{q}_j^{n+1})^m , \tag{4.33}$$

then we can write the basic iteration in Algorithm 1 as

$$\left[\mathbf{I} - \Delta\tau(\mathbf{A}_j^{n+1})^m\right](\mathbf{v}_j^{n+1})^{m+1} = \mathbf{v}_j^n + \Delta\tau\mathbf{d}_j^m . \tag{4.34}$$

Algorithm 1 Policy Iteration: compute $\mathbf{v}_j^{n+1}$ given $\mathbf{v}_j^n$

Require: $\mathbf{v}_j^n$ as defined in (4.26)

Solve $\mathbf{v}_0^{n+1}$ from $\left[\mathbf{I} - \Delta\tau \mathbf{A}_0\right] \mathbf{v}_0^{n+1} = \mathbf{v}_0^n$

for $j = 1, 2, \ldots, j_{\max}$ **do**

  Initialize $(\mathbf{v}_j^{n+1})^m$ with $\mathbf{v}_j^n$ for $m = 0$

  for $m = 0, 1, \ldots$ until converge **do**

$$\{(\varphi_{i,j}^{n+1})^m, (\psi_{i,j}^{n+1})^m\} \Leftarrow \underset{\{\varphi,\psi\} \in \{\{0,0\},\{0,1\},\{1,0\}\}}{\arg\max} \left[\mathbf{A}_j^{n+1}(\varphi,\psi)(\mathbf{v}_j^{n+1})^m + \mathbf{P}_j^{n+1}(\varphi,\psi)\mathbf{v}_{j-1}^{n+1} + \mathbf{q}_j^{n+1}(\varphi,\psi)\right]_i$$

    Construct $(\mathbf{A}_j^{n+1})^m$, $(\mathbf{P}_j^{n+1})^m$ and $(\mathbf{q}_j^{n+1})^m$ by using $\{(\varphi_{i,j}^{n+1})^m, (\psi_{i,j}^{n+1})^m\}$

    Solve $(\mathbf{v}_j^{n+1})^{m+1}$ from

$$\left[\mathbf{I} - \Delta\tau(\mathbf{A}_j^{n+1})^m\right](\mathbf{v}_j^{n+1})^{m+1} = \mathbf{v}_j^n + \Delta\tau(\mathbf{P}_j^{n+1})^m \mathbf{v}_{j-1}^{n+1} + \Delta\tau(\mathbf{q}_j^{n+1})^m$$

    if $\max\limits_i \dfrac{|(V_{i,j}^{n+1})^{m+1} - (V_{i,j}^{n+1})^m|}{\max\left[scale, (V_{i,j}^{n+1})^{m+1}\right]} < tolerance$ **then**

      break from the iteration

    end if

  end for

end for

Manipulation of equation (4.34) gives

$$\left[\mathbf{I} - \Delta\tau(\mathbf{A}_j^{n+1})^m\right]\left((\mathbf{v}_j^{n+1})^{m+1} - (\mathbf{v}_j^{n+1})^m\right)$$
$$= \Delta\tau\left(\mathbf{d}_j^m + (\mathbf{A}_j^{n+1})^m(\mathbf{v}_j^{n+1})^m\right) - \Delta\tau\left(\mathbf{d}_j^{m-1} + (\mathbf{A}_j^{n+1})^{m-1}(\mathbf{v}_j^{n+1})^m\right) . \qquad (4.35)$$

The proof proceeds by noting that the right hand side of equation (4.35) is always nonnegative (Wang and Forsyth, 2008), and since $\left[\mathbf{I} - \Delta\tau(\mathbf{A}_j^{n+1})^m\right]$ is an M-matrix, then the iterates form a bounded non-decreasing sequence. □

For some recent work on policy iteration and the relation to Newton iteration, we refer the reader to Santos and Rust (2004) and Bokanowski et al. (2009).

5 Convergence of the penalty discretization

From (Barles and Souganidis, 1991; Barles, 1997) we find that any scheme which is monotone, consistent (in the viscosity sense) and l_∞ stable converges to the viscosity solution. In the following sections, we will verify each of these properties in turn for the penalty scheme.

It will be convenient at this point to introduce the following definitions

$$\Delta W_{\max} = \max_i (W_{i+1} - W_i) \qquad \Delta W_{min} = \min_i (W_{i+1} - W_i)$$
$$\Delta A_{\max} = \max_j (A_{j+1} - A_j) \qquad \Delta A_{\min} = \min_j (A_{j+1} - A_j).$$

5.1 Stability

The stability of scheme (4.13), (4.18)-(4.23), is a direct result of the following Lemma:

Lemma 5.1 (Stability)**.** *If the discretized equation* (4.13) *satisfies the positive coefficient condition* (4.15)*, then scheme* (4.13), (4.18)-(4.23)*, satisfies*

$$e^{-\eta\tau^n} W_i \leq V_{i,j}^n \leq W_i + A_j \qquad (5.1)$$

for $0 \leq n \leq N$ *as* $\Delta\tau \to 0$*,* $\Delta W_{\min} \to 0$*,* $\Delta A_{\min} \to 0$*.*

Proof. Define a discrete bounding function $B_{i,j}^n$ such that

$$B_{i,j}^n = W_i + A_j \ . \qquad (5.2)$$

Consider the matrix

$$[\mathbf{Z}^{n+1}(\mathbf{v}^{n+1})\mathbf{v}^{n+1}]_{i,j} = -\alpha_{i,j}^{n+1} V_{i-1,j}^{n+1} + (\frac{1}{\Delta\tau} + \alpha_{i,j}^{n+1} + \beta_{i,j}^{n+1} + c_{i,j}^{n+1}) V_{i,j}^{n+1}$$
$$-\beta_{i,j}^{n+1} V_{i+1,j}^{n+1} - p_{i,j}^{n+1} V_{i,j-1}^{n+1} \ . \qquad (5.3)$$

Define vectors

$$\mathbf{b}_j^n = \left[B_{0,j}^n, B_{1,j}^n, \ldots, B_{i_{\max},j}\right] \quad ; \quad \mathbf{b}^n = \left[(\mathbf{b}_0^n), (\mathbf{b}_1^n), \ldots, (\mathbf{b}_{j_{\max}}^n)\right]' \ . \qquad (5.4)$$

Then, some straightforward (but lengthy) algebra shows that

$$\mathbf{Z}^{n+1}(\mathbf{v}^{n+1})(\mathbf{b}^{n+1} - \mathbf{v}^{n+1}) \quad = \quad \frac{1}{\Delta\tau}[\mathbf{b}^n - \mathbf{v}^n] + \mathbf{h}^{n+1}(\mathbf{v}^{n+1}) \ , \tag{5.5}$$

where

$$[\mathbf{h}^{n+1}]_{i,j} \quad = \quad \begin{cases} \eta W_i + rA_j + (\varphi_{i,j}^{n+1} G + \frac{\psi_{i,j}^{n+1}}{\varepsilon})(1-\delta_{i,0}) + \psi_{i,j}^{n+1}\kappa(1/\varepsilon - G) & i < i_{\max}, j > 0 \ , \\ \eta(W_i + A_j) & \text{otherwise} \ , \end{cases} \tag{5.6}$$

where $\delta_{i,j}$ is the Kronecker delta. Since $1/\varepsilon > G$, then $\mathbf{h}^{n+1} \geq 0$. Assume $\mathbf{b}^n - \mathbf{v}^n \geq 0$, then, since $\mathbf{Z}^{n+1}$ is an $M-$matrix, $\mathbf{b}^{n+1} - \mathbf{v}^{n+1} \geq 0$. Note from the initial condition (4.18), we have $\mathbf{b}^0 - \mathbf{v}^0 \geq 0$. Hence

$$V_{i,j}^n \quad \leq \quad W_i + A_j \quad , \quad \forall n \ . \tag{5.7}$$

For the lower bound, define the lower bounding grid function

$$L_{i,j}^n = \frac{W_i}{(1+\eta\Delta\tau)^n} \ . \tag{5.8}$$

Following a similar approach as used for the upper bound, we find that

$$V_{i,j}^n \geq \frac{W_i}{(1+\eta\Delta\tau)^n} > e^{-\eta\tau^n} W_i \ . \tag{5.9}$$

□

Remark 5.1. *For a given finite domain Ω^L, bound (5.1) clearly implies that $\|V^n\|_\infty$ is bounded. However, note that for fixed (W, A, τ), bound (5.1) is independent of $W_{\max}$, which is an important property if we solve the problem in Definition (3.1) on a sequence of larger domains.*

5.2 Consistency

This section shows that the discretization scheme (4.13), (4.18)-(4.21) is consistent with the singular control GMWB pricing problem as defined in Definition 3.2.

Consider the discretized equation (4.13), and the associated discretized boundary conditions (4.18)-(4.23). We make the following assumption regarding the mesh/time-step size.

Assumption 5.1. *There exists a mesh/time-step size parameter h such that*

$$h = \frac{\Delta W_{\max}}{C_1} = \frac{\Delta A_{\max}}{C_2} = \frac{\Delta\tau}{C_3} = \frac{\varepsilon}{C_4}, \tag{5.10}$$

where C_i $(i = 1, 2, 3, 4)$ are positive constants independent of h.

Equation (4.13) is equivalent to equation (4.7), which can be re-written as

$$\frac{V_{i,j}^{n+1} - V_{i,j}^n}{\Delta\tau} - \mathcal{L}^h V_{i,j}^{n+1} - \max\left(G\max(\mathcal{F}^h V_{i,j}^{n+1}, 0), \frac{(\mathcal{F}^h V_{i,j}^{n+1} - \kappa)}{\varepsilon} + \kappa G\right) = 0 \ , \tag{5.11}$$

or equivalently

$$\min\Bigg[\frac{V_{i,j}^{n+1}-V_{i,j}^{n}}{\Delta\tau}-\mathcal{L}^hV_{i,j}^{n+1}-\kappa G-\frac{1}{\varepsilon}(\mathcal{F}^hV_{i,j}^{n+1}-\kappa),$$
$$\frac{V_{i,j}^{n+1}-V_{i,j}^{n}}{\Delta\tau}-\mathcal{L}^hV_{i,j}^{n+1}-G\max(\mathcal{F}^hV_{i,j}^{n+1},0)\Bigg]=0\ . \tag{5.12}$$

Equation (5.12) implies that one of the following holds with equality:

$$\frac{V_{i,j}^{n+1}-V_{i,j}^{n}}{\Delta\tau}-\mathcal{L}^hV_{i,j}^{n+1}-\kappa G-\frac{1}{\varepsilon}(\mathcal{F}^hV_{i,j}^{n+1}-\kappa)\geq 0, \tag{5.13}$$
$$\frac{V_{i,j}^{n+1}-V_{i,j}^{n}}{\Delta\tau}-\mathcal{L}^hV_{i,j}^{n+1}-G\max(\mathcal{F}^hV_{i,j}^{n+1},0)\geq 0. \tag{5.14}$$

Since $\varepsilon>0$, equation (5.13) is equivalent to

$$\varepsilon\bigg(\frac{V_{i,j}^{n+1}-V_{i,j}^{n}}{\Delta\tau}-\mathcal{L}^hV_{i,j}^{n+1}-\kappa G\bigg)+(\kappa-\mathcal{F}^hV_{i,j}^{n+1})\geq 0. \tag{5.15}$$

As a result, equations (5.14) and (5.15) can be combined to give

$$\begin{aligned}\mathcal{H}_{i,j}^{n+1}\equiv&\mathcal{H}_{i,j}^{n+1}\bigg(h,V_{i,j}^{n+1},\Big\{V_{a,b}^{n+1}\Big\}_{\substack{a\neq i\\ \text{or } b\neq j}},V_{i,j}^{n}\bigg)\\ =&\min\Bigg[\frac{V_{i,j}^{n+1}-V_{i,j}^{n}}{\Delta\tau}-\mathcal{L}^hV_{i,j}^{n+1}-G\max(\mathcal{F}^hV_{i,j}^{n+1},0),\\ &\varepsilon\left(\frac{V_{i,j}^{n+1}-V_{i,j}^{n}}{\Delta\tau}-\mathcal{L}^hV_{i,j}^{n+1}-\kappa G\right)+(\kappa-\mathcal{F}^hV_{i,j}^{n+1})\Bigg]=0,\end{aligned} \tag{5.16}$$

where $\Big\{V_{a,b}^{n+1}\Big\}_{\substack{a\neq i\\ \text{or } b\neq j}}$ is the set of values $V_{a,b}^{n+1}$, $a=0,1,\ldots,i_{\max}$ and $b=0,1,\ldots,j_{\max}$, $(a,b)\neq(i,j)$. We can re-formulate the discretization scheme (4.13), (4.18)-(4.23) at node (W_i,A_j,τ^{n+1}) into one equation:

$$\begin{aligned}&\mathcal{G}_{i,j}^{n+1}\bigg(h,V_{i,j}^{n+1},\Big\{V_{a,b}^{n+1}\Big\}_{\substack{a\neq i\\ \text{or } b\neq j}},V_{i,j}^{n}\bigg)\\ =&\begin{cases}\mathcal{H}_{i,j}^{n+1}, & (W_i,A_j,\tau^{n+1})\in\Omega_{\text{in}}\cup\Omega_{W_0}\cup\Omega_{A_0},\\ V_{i,j}^{n+1}(1+\eta\Delta\tau)-V_{i,j}^{n}, & (W_i,A_j,\tau^{n+1})\in\Omega_{W_{\max}}\\ V_{i,j}^{n+1}-\max\left[W_i,(1-\kappa)A_j\right], & (W_i,A_j,\tau^{n+1})\in\Omega_{\tau^0}.\end{cases}\\ =&0.\end{aligned} \tag{5.17}$$

We follow here the definition of consistency in the viscosity sense (Barles, 1997). For an excellent overview of this topic, we refer the reader to (Jakobsen, 2010).

Definition 5.1 (Consistency)**.** *For any smooth test function* $\phi(W, A, \tau)$ *with* $\phi_{i,j}^{n+1} = \phi(W_i, A_j, \tau^{n+1})$, *having bounded derivatives of all orders with respect to* W, A, *and* τ, *assuming the mesh/time-step size parameter* h *satisfies Assumption 5.1, the numerical scheme* $\mathcal{G}_{i,j}^{n+1}\left(h, \phi_{i,j}^{n+1}, \left\{\phi_{a,b}^{n+1}\right\}_{\substack{a\neq i\\ or\ b\neq j}}, \left\{\phi_{i,j}^{n}\right\}\right)$ *is consistent if* $\forall \hat{\mathbf{x}} = (\hat{W}, \hat{A}, \hat{\tau}) \in \Omega^L$, $\forall \mathbf{x}_{i,j}^{n+1} = (W_i, A_j, \tau^{n+1}) \in \Omega^L$, *the following two inequalities hold.*

$$\limsup_{\substack{\mathbf{x}_{i,j}^{n+1}\to\hat{\mathbf{x}}\\ h\to 0\\ \xi\to 0}} \mathcal{G}_{i,j}^{n+1}\left(h, \phi_{i,j}^{n+1}+\xi, \left\{\phi_{a,b}^{n+1}+\xi\right\}_{\substack{a\neq i\\ or\ b\neq j}}, \left\{\phi_{i,j}^{n}+\xi\right\}\right) \leq (F_{\Omega^L})^*(\phi(\hat{x})), \tag{5.18}$$

$$\liminf_{\substack{\mathbf{x}_{i,j}^{n+1}\to\hat{\mathbf{x}}\\ h\to 0\\ \xi\to 0}} \mathcal{G}_{i,j}^{n+1}\left(h, \phi_{i,j}^{n+1}+\xi, \left\{\phi_{a,b}^{n+1}+\xi\right\}_{\substack{a\neq i\\ or\ b\neq j}}, \left\{\phi_{i,j}^{n}+\xi\right\}\right) \geq (F_{\Omega^L})_*(\phi(\hat{x})). \tag{5.19}$$

where $(F_{\Omega^L})^*$ and $(F_{\Omega^L})_*$ are the upper and lower semicontinuous envelopes of F_{Ω^L}. Before proving consistency, we shall need an intermediate result, which is given in the following Lemma.

Lemma 5.2 (Local consistency)**.** *Suppose the mesh size and the time-step parameter satisfy Assumption 5.1, then for any smooth function* $\phi(W, A, \tau)$ *having bounded derivatives of all orders in* $(W, A, \tau) \in \Omega^L$, *with* $\phi_{i,j}^{n+1} = \phi(W_i, A_j, \tau^{n+1})$, *and for* h, ξ *sufficiently small, we have that*

$$\begin{aligned}
&\mathcal{G}_{i,j}^{n+1}\left(h, \phi_{i,j}^{n+1}+\xi, \left\{\phi_{a,b}^{n+1}+\xi\right\}_{\substack{a\neq i\\ i\,or\ b\neq j}}, \left\{\phi_{i,j}^{n}+\xi\right\}\right)\\
&= \begin{cases}
F_{in}\phi_{i,j}^{n+1} + O(h) + O(\xi), & (W_i, A_j, \tau^{n+1}) \in \Omega_{in},\\
F_{W_0}\phi_{i,j}^{n+1} + O(h) + O(\xi,) & (W_i, A_j, \tau^{n+1}) \in \Omega_{W_0},\\
F_{A_0}\phi_{i,j}^{n+1} + O(h) + O(\xi), & (W_i, A_j, \tau^{n+1}) \in \Omega_{A_0},\\
F_{W_{\max}}\phi_{i,j}^{n+1} + O(h) + O(\xi), & (W_i, A_j, \tau^{n+1}) \in \Omega_{W_{\max}}\\
F_{\tau^0}\phi_{i,j}^{n+1} + O(\xi), & (W_i, A_j, \tau^{n+1}) \in \Omega_{\tau^0},
\end{cases}
\end{aligned} \tag{5.20}$$

where ξ *is a constant independent of* $\mathbf{x}_{i,j}^{n+1}$.

Proof. Before proving the Lemma, we first define the following notations for the operators applied to test functions, evaluated at node (W_i, A_j, τ^{n+1}).

$$\begin{aligned}
\mathcal{L}\phi_{i,j}^{n+1} &\equiv \mathcal{L}\phi(W_i, A_j, \tau^{n+1}), & \mathcal{F}\phi_{i,j}^{n+1} &\equiv \mathcal{F}\phi(W_i, A_j, \tau^{n+1}),\\
(\phi_W)_{i,j}^{n+1} &\equiv \phi_W(W_i, A_j, \tau^{n+1}), & (\phi_A)_{i,j}^{n+1} &\equiv \phi_A(W_i, A_j, \tau^{n+1}),\\
(\phi_\tau)_{i,j}^{n+1} &\equiv \phi_\tau(W_i, A_j, \tau^{n+1}).
\end{aligned}$$

By definitions of discrete operators $\mathcal{L}^h$ and $\mathcal{F}^h$ in (4.6), it can be easily verified that

$$\mathcal{L}^h(\phi_{i,j}^{n+1}+\xi) = \mathcal{L}^h\phi_{i,j}^{n+1} - r\xi \tag{5.21}$$

$$\mathcal{F}^h(\phi_{i,j}^{n+1}+\xi) = \mathcal{F}^h\phi_{i,j}^{n+1}. \tag{5.22}$$

From Taylor series expansions and the last two equations above, we have that

$$\mathcal{L}^h(\phi_{i,j}^{n+1} + \xi) = \mathcal{L}\phi_{i,j}^{n+1} - r\xi + O(\Delta W_{\max}), \tag{5.23}$$

$$\mathcal{F}^h(\phi_{i,j}^{n+1} + \xi) = \mathcal{F}\phi_{i,j}^{n+1} + O(\Delta W_{\max}) + O(\Delta A_{\max}), \tag{5.24}$$

$$\frac{\phi_{i,j}^{n+1} - \phi_{i,j}^{n}}{\Delta\tau} = (\phi_\tau)_{i,j}^{n+1} + O(\Delta\tau). \tag{5.25}$$

By using equation (5.16) together with the discretization error estimation in the last three equations above, and the inequality $|\min(x,y) - \min(a,b)| \le \max(|x-a|, |y-b|)$, we can see for nodes $(W_i, A_j, \tau^{n+1}) \in \Omega_{\text{in}}$:

$$\begin{aligned}
&\left|\mathcal{G}_{i,j}^{n+1}\left(h, \phi_{i,j}^{n+1} + \xi, \left\{\phi_{a,b}^{n+1} + \xi\right\}_{\substack{a\neq i \\ \text{or } b\neq j}}, \left\{\phi_{i,j}^{n} + \xi\right\}\right) - F_{\text{in}}\phi_{i,j}^{n+1}\right| \\
&\le \max\Bigg[\Bigg|\frac{\phi_{i,j}^{n+1} - \phi_{i,j}^{n}}{\Delta\tau} - \mathcal{L}^h(\phi_{i,j}^{n+1} + \xi) - G\max\left[\mathcal{F}^h(\phi_{i,j}^{n+1} + \xi), 0\right] \\
&\qquad - \left((\phi_\tau)_{i,j}^{n+1} - \mathcal{L}\phi_{i,j}^{n+1} - G\max\left[\mathcal{F}\phi_{i,j}^{n+1}, 0\right]\right)\Bigg|, \\
&\qquad \Bigg|\varepsilon\left(\frac{\phi_{i,j}^{n+1} - \phi_{i,j}^{n}}{\Delta\tau} - \mathcal{L}^h(\phi_{i,j}^{n+1} + \xi) - \kappa G\right) \\
&\qquad + \left(\mathcal{F}\phi_{i,j}^{n+1} - \mathcal{F}^h(\phi_{i,j}^{n+1} + \xi)\right)\Bigg|\Bigg] \\
&\le \max\Bigg[\Bigg|O(\Delta\tau) + O(\Delta W_{\max}) + r\xi + G\left|\mathcal{F}^h(\phi_{i,j}^{n+1} + \xi) - \mathcal{F}\phi_{i,j}^{n+1}\right|\Bigg|, \\
&\qquad \Bigg|O(\Delta W_{\max}) + O(\Delta A_{\max}) \\
&\qquad + \varepsilon\left(\frac{\phi_{i,j}^{n+1} - \phi_{i,j}^{n}}{\Delta\tau} - \mathcal{L}^h(\phi_{i,j}^{n+1} + \xi) - \kappa G\right)\Bigg|\Bigg] \\
&= \max\Bigg[\Bigg|O(\Delta\tau) + O(\Delta W_{\max}) + O(\Delta W_{\max} + \Delta A_{\max}) + r\xi\Bigg|, \\
&\qquad \Bigg|O(\Delta W_{\max}) + O(\Delta A_{\max}) \\
&\qquad + \varepsilon\left((\phi_\tau)_{i,j}^{n+1} - \mathcal{L}\phi_{i,j}^{n+1} + r\xi - \kappa G + O(\Delta\tau)) + O(\Delta W_{\max})\right)\Bigg|\Bigg]
\end{aligned} \tag{5.26}$$

By Assumption 5.1 and the inequality (5.26), we obtain

$$\mathcal{G}_{i,j}^{n+1}\left(h, \phi_{i,j}^{n+1} + \xi, \left\{\phi_{a,b}^{n+1} + \xi\right\}_{\substack{a\neq i \\ \text{or } b\neq j}}, \left\{\phi_{i,j}^{n} + \xi\right\}\right) = F_{\text{in}}\phi_{i,j}^{n+1} + O(h) + O(\xi). \tag{5.27}$$

This proves the first equation in (5.20). The rest of the equations in (5.20) are proved by following similar arguments. □

Lemma 5.3 (Consistency)**.** *Assume that all conditions in Lemma 5.2 are satisfied, then scheme* (5.17) *is consistent according to Definition 5.1.*

Remark 5.2 (Consistency in the viscosity sense)**.** *Given the local consistency result in Lemma 5.2, it is straightforward to show that scheme* (5.17) *is consistent in the sense of Definition 5.1. We will include these steps here for the convenience of the reader, although this is mainly an exercise in notational manipulation. In general, however, we may not be able to get local consistency everywhere. As an example, in Chen and Forsyth (2008), there are nodes in strips near the domain boundaries where local consistency is not achieved. In this case, the more relaxed definition of consistency in the viscosity sense is particularly useful, and the final steps required to prove consistency are non-trivial.*

Proof. First we prove that the inequality (5.18) holds. From the definition of lim sup, there exists sequences i_k, j_k, n_k, ξ_k and h_k such that

$$\text{as}\quad k \to \infty,\ \mathbf{x}_{i_k,j_k}^{n_k+1} \to \hat{\mathbf{x}},\ \xi_k \to 0,\ h_k \to 0, \tag{5.28}$$

and

$$\begin{aligned}
&\limsup_{k\to\infty} \mathcal{G}_{i,j}^{n+1}\left(h_k, \phi_{i_k,j_k}^{n_k+1} + \xi_k, \left\{\phi_{a_k,b_k}^{n_k+1} + \xi_k\right\}_{\substack{a_k\neq i_k \\ \text{or } b_k\neq j_k}}, \left\{\phi_{i_k,j_k}^{n_k} + \xi_k\right\}\right) \\
=&\limsup_{\substack{\mathbf{x}_{i,j}^{n+1}\to\hat{\mathbf{x}} \\ h\to 0 \\ \xi\to 0}} \mathcal{G}_{i,j}^{n+1}\left(h, \phi_{i,j}^{n+1} + \xi, \left\{\phi_{a,b}^{n+1} + \xi\right\}_{\substack{a\neq i \\ \text{or } b\neq j}}, \left\{\phi_{i,j}^{n} + \xi\right\}\right)
\end{aligned} \tag{5.29}$$

From Lemma 5.2, we have for k sufficiently large, there exist positive constants C_1, C_2 independent of k such that

$$\begin{aligned}
&\left|\mathcal{G}_{i,j}^{n+1}\left(h_k, \phi_{i_k,j_k}^{n_k+1} + \xi_k, \left\{\phi_{a_k,b_k}^{n_k+1} + \xi_k\right\}_{\substack{a_k\neq i_k \\ \text{or } b_k\neq j_k}}, \left\{\phi_{i_k,j_k}^{n_k} + \xi_k\right\}\right) - F_{\Omega^L}\phi_{i_k,j_k}^{n_k+1}\right| \\
&\leq C_1 h_k + C_2 \xi_k \quad ; \quad (W_{i_k}, A_{j_k}, \tau^{n_k+1}) \in \Omega^L\ .
\end{aligned} \tag{5.30}$$

Remark 5.3. *Suppose, for example, that $\hat{\mathbf{x}} \in \Omega_{W_0}$. Note that for k sufficiently large, $\mathbf{x}_{i_k,j_k}^{n_k+1}$ can be in either Ω_{W_0} or Ω_{in}. However, in each case, from Lemma 5.2, we have that inequality (5.30) holds. This is a consequence of the definition of F_{Ω^L}.*

From equations (5.29) and (5.30), we obtain

$$\begin{aligned}
&\limsup_{\substack{\mathbf{x}_{i,j}^{n+1}\to\hat{\mathbf{x}} \\ h\to 0 \\ \xi\to 0}} \mathcal{G}_{i,j}^{n+1}\left(h, \phi_{i,j}^{n+1} + \xi, \left\{\phi_{a,b}^{n+1} + \xi\right\}_{\substack{a\neq i \\ \text{or } b\neq j}}, \left\{\phi_{i,j}^{n} + \xi\right\}\right) \\
\leq&\limsup_{k\to\infty} F_{\Omega^L}\phi_{i_k,j_k}^{n_k+1} + \limsup_{k\to\infty}\left[C_1 h_k + C_2 \xi_k\right] \\
\leq&(F_{\Omega^L})^*(\phi(\hat{\mathbf{x}})),
\end{aligned} \tag{5.31}$$

Similarly,

$$\begin{aligned}
&\liminf_{\substack{\mathbf{x}_{i,j}^{n+1}\to\hat{\mathbf{x}}\\ h\to 0\\ \xi\to 0}} \mathcal{G}_{i,j}^{n+1}\left(h,\phi_{i,j}^{n+1}+\xi,\left\{\phi_{a,b}^{n+1}+\xi\right\}_{\substack{a\neq i\\ \text{or } b\neq j}},\left\{\phi_{i,j}^{n}+\xi\right\}\right)\\
\geq &\liminf_{k\to\infty} F_{\Omega^L}\phi_{i_k,j_k}^{n_k+1}+\liminf_{k\to\infty}\left[-C_1h_k-C_2\xi_k\right]\\
\geq &(F_{\Omega^L})_*(\phi(\hat{\mathbf{x}})). \qquad (5.32)
\end{aligned}$$

□

5.3 Monotonicity

Definition 5.2 (Monotonicity)**.** *The numerical scheme* $\mathcal{G}_{i,j}^{n+1}\left(h,V_{i,j}^{n+1},\left\{V_{a,b}^{n+1}\right\}_{\substack{a\neq i\\ or\ b\neq j}},V_{i,j}^{n}\right)$ *in* (5.17) *is monotone if for all* $Y_{i,j}^n\geq X_{i,j}^n,\forall i,j,n$

$$\mathcal{G}_{i,j}^{n+1}\left(h,V_{i,j}^{n+1},\left\{Y_{a,b}^{n+1}\right\}_{\substack{a\neq i\\ or\ b\neq j}},Y_{i,j}^{n}\right)\leq \mathcal{G}_{i,j}^{n+1}\left(h,V_{i,j}^{n+1},\left\{X_{a,b}^{n+1}\right\}_{\substack{a\neq i\\ or\ b\neq j}},X_{i,j}^{n}\right). \qquad (5.33)$$

Lemma 5.4 (Monotonicity)**.** *If scheme (5.17) satisfies the positive coefficient condition (4.15) then it is monotone according to Definition 5.2.*

Proof. This is easily done using the same steps as in (Forsyth and Labahn, 2008). □

5.4 Convergence in Ω^L

Theorem 5.1 (Convergence to the viscosity solution)**.** *Assume that scheme (5.17) satisfies all the conditions required for Lemmas 5.1, 5.3, and 5.4, and that Assumption 3.1 holds, then the scheme (5.17) converges to the unique, continuous viscosity solution of the GMWB problem given in Definition 3.2, at any point in* $\Omega_{in}\cup\Gamma$ *(see Definition of* Γ *in Assumption 3.1).*

Proof. Since the scheme is monotone, consistent and pointwise stable, this follows from the results in Barles and Souganidis (1991). □

Remark 5.4. *Note that since we have assumed that strong comparison holds only in* $\Omega_{in}\cup\Gamma$*, then we can guarantee uniqueness and continuity only in* $\Omega_{in}\cup\Gamma$*.*

5.5 Convergence in Ω^∞

The asymptotic form of the solution for $W\to\infty$ is given in Dai et al. (2008), which we impose at finite $W_{\max}$ through boundary condition (3.9). This, of course, causes an error due to finite $W_{\max}$ (see equation (3.8)).

Consider a sequence of converged viscosity solutions $(V(W,A,\tau))^k$, which satisfy Definition 3.2 on the sequence of grids $(\Omega^L)^k$, $k\to\infty$, with $W_{\max}^k>W_{\max}^{k-1}$. In Barles et al. (1995), the limiting problem of convergence to the viscosity solution on unbounded domains with quadratic growth in the solution is discussed. It is possible to appeal to the results in (Barles et al., 1995) to show convergence as $(\Omega^L)^k\to\Omega^\infty$. However we can use a simpler approach for problem at hand.

For simplicity, and to avoid notational complexity, we consider only points in $(\Omega_{in})^k$ in the following, since from Theorem 5.1 we are ensured of convergence at least to points in $(\Omega_{in})^k$.

We will use the following elementary Lemmas.

Lemma 5.5 (Bounds on solution on $(\Omega_{in})^k$)**.** *The converged viscosity solution on each domain $(\Omega_{in})^k$ has the bounds*

$$e^{-\eta\tau} W \leq (V(W, A, \tau))^k \leq W + A \quad . \tag{5.34}$$

Proof. Since the discrete solution satisfies the bounds in Lemma 5.1, independent of h, $W_{\max}$, we take the limit as $h \to 0$, and hence the viscosity solution satisfies these same bounds. □

Lemma 5.6. *The following bound holds*

$$(V(W, A, \tau))^{k+1} \quad \geq \quad (V(W, A, \tau))^k \quad ; \quad (W, A, \tau) \in (\Omega_{in})^k \quad . \tag{5.35}$$

Proof. We can regard $(V(W, A, \tau))^{k+1}$ on domain $(\Omega^L)^k$, as the solution to the GMWB pricing problem on $(\Omega^L)^k$, but with a known boundary condition at $W = W^k_{\max}$, which in general is not the same boundary condition as used for $V(W, A, \tau))^k$. From Lemma 5.5, we have that

$$(V(W^k_{\max}, A, \tau))^{k+1} \geq e^{-\eta\tau} W^k_{\max} = (V(W^k_{\max}, A, \tau))^k \; . \tag{5.36}$$

Hence $(V(W^k_{\max}, A, \tau))^{k+1}$ and $(V(W^k_{\max}, A, \tau))^k$ are solutions to the same PDE and boundary conditions, with the exception of the boundary condition at $W = W^k_{\max}$, which satisfies equation (5.36). Consider two discrete solutions $(V(W, A, \tau))^k_h, (V(W, A, \tau))^{k+1}_h$, defined on the same set of nodes in $(\Omega^L)^k$, and assume that the discretization satisfies all the conditions required for Theorem 5.1. Then, from Theorem 5.2 in (Forsyth and Labahn, 2008), we have that $(V(W, A, \tau))^{k+1}_h \geq (V(W, A, \tau))^k_h$ at all the nodes. Take the limit as $h \to 0$, and noting that $(V(W, A, \tau))^{k+1}_h \to (V(W, A, \tau))^{k+1}$ and $(V(W, A, \tau))^k_h \to V(W, A, \tau))^k$, and we obtain result (5.35).

□

Theorem 5.2 (Convergence in Ω^∞)**.** *Consider the sequence of grids $(\Omega^L)^k$, with $W^{k+1}_{\max} > W^k_{\max}$ and*

$$\lim_{k \to \infty} (\Omega^L)^k \quad = \quad \Omega^\infty \; . \tag{5.37}$$

For any fixed point $(W, A, \tau) \in (\Omega_{in})^\infty$ we have that the sequence $(V(W, A, \tau))^k$ converges to a unique value $(V(W, A, \tau))^\infty$ as $k \to \infty$.

Proof. Given a fixed point (W, A, τ), from Lemma 5.6 we have that the solution is a non-decreasing function of the domain index k. But from Lemma 5.5, the solution is locally upper bounded independent of the domain index k. Hence the sequence $(V(W, A, \tau))^k, k \to \infty$ is bounded and non-decreasing, and thus converges to a limit $(V(W, A, \tau))^\infty$. Consider another set of increasing domains $(\hat{\Omega}^L)^k$. Suppose this set of domains converges to a value

$$(\hat{V}(W, A, \tau))^\infty > (V(W, A, \tau))^\infty \quad . \tag{5.38}$$

But, applying Lemma 5.6 to subsequences of $(\Omega^L)^k$ and $(\hat{\Omega}^L)^k$ leads to a contradiction, hence the limit $(V(W, A, \tau))^\infty$ is unique. □

Remark 5.5. *We apply scheme (5.17) to a sequence of problems with smaller h, for fixed $W_{\max}$. We then increase $W_{\max}$ and repeat the process. Since we use unequally spaced grids, it is computationally inexpensive to choose a large $W_{\max}$, hence the process of determining the limit $W_{\max} \to \infty$ is rapidly convergent, in practice.*

6 Numerical Examples

In this section, numerical experiments are presented using the scheme discussed in previous sections to price the GMWB variable annuities.

6.1 No-arbitrage fee

Since no fee is paid up-front, the insurance company needs to charge a proportional fee η (see equation (2.2)), such that the value of the contract is equal to the initial premium ω_0. Let $V(\eta; W, A, \tau)$ be the value of the contract as a function of η. The no-arbitrage fee is the solution to the equation

$$V(\eta; W = \omega_0, A = \omega_0, \tau = T) \quad = \quad \omega_0 \ . \tag{6.1}$$

We solve equation (6.1) using Newton iteration with convergence tolerance

$$|\eta^{k+1} - \eta^k| < 10^{-8} \ , \tag{6.2}$$

with η^k being the $k'th$ iterate.

6.2 Computational Parameters

In the localized computational domain $\Omega = [0, W_{\max}] \times [0, \omega_0] \times [0, T]$, we set $W_{\max} = 100\omega_0$. Tests with $W_{\max} = 1000\omega_0$ showed no effect on the computed solution to twelve digits. The policy iteration error control parameter *tolerance* in Algorithm 1 is set to 10^{-8}.

From the analysis in the previous sections, we will obtain convergence if

$$\varepsilon \quad = \quad C\Delta\tau \tag{6.3}$$

for any $C > 0$. However, in order to obtain reasonable results for coarse grids/timesteps, we can estimate a suitable constant C as follows. Recall that the maximum withdrawal rate in equation (2.5) is $1/\varepsilon$. If $1/\varepsilon = \omega_0/(\Delta\tau)$, then the entire guarantee amount can be withdrawn in a single timestep. This would suggest that a reasonable value for ε would be

$$\varepsilon = \Delta\tau C^*/\omega_0, \tag{6.4}$$

with $C^* < 1$ a dimensionless constant. We also want to make the term

$$\varepsilon \left(V_\tau - \mathcal{L}V - \kappa G\right) \tag{6.5}$$

small in equation (2.13) for coarse grids. Hence, we choose $C^* = 10^{-2}$ in equation (6.4). In a later section, we will present the results for a series of tests with different values of C^* in equation (6.4), which show that the results are insensitive to C^* for values ranging over several orders of magnitude.

Our numerical experiments are performed on the example GMWB contract used in Chen and Forsyth (2008). The parameters for this contract are given in Table 6.1. Table 6.2 gives the mesh size and timestep parameters.

Parameter	Value
Expiry time T	10.0 years
Interest rate r	0.05
Maximum no penalty withdrawal rate G	10/year
Withdrawal penalty κ	0.10
Initial lump-sum premium ω_0	100
Initial guarantee account balance $A(0)$	100
Initial personal annuity account balance $W(0)$	100
Penalty parameter ε	$\Delta\tau 10^{-2}/\omega_0$

TABLE 6.1: *A sample GMWB contract parameters used in the numerical experiments*

Level	W Nodes	A Nodes	Time steps
1	117	111	120
2	233	221	240
3	465	441	480
4	929	881	960
5	1857	1761	1920

TABLE 6.2: *Grid and timestep data for convergence experiments*

Refinement	Central Differencing First			For/Backward Differencing Only		
Level	Value	Itns/step	Ratio	Value	Itns/step	Ratio
$\sigma = 0.2, \eta = 0.013886$						Fully Implicit Method
1	101.3114	3.51	N/A	101.6030	3.47	N/A
2	100.4488	3.62	N/A	100.6914	3.55	2.02
3	100.1267	3.70	2.68	100.2816	3.66	2.22
4	100.0270	3.77	3.23	100.1082	3.74	2.36
5	99.9999	3.89	3.69	100.0346	3.88	2.36
$\sigma = 0.2, \eta = 0.013886$						Crank Nicolson Method
1	101.3085	3.39	N/A	101.6017	3.35	N/A
2	100.4474	3.49	N/A	100.6909	3.42	N/A
3	100.1261	3.55	2.68	100.2815	3.52	2.22
4	100.0262	3.55	3.22	100.1082	3.52	2.36
5	99.9995	3.57	3.75	100.0343	3.55	2.35
$\sigma = 0.3, \eta = 0.031286$						Fully Implicit Method
1	100.5946	4.19	N/A	100.8998	4.10	N/A
2	100.1488	4.31	N/A	100.3363	4.26	N/A
3	100.0357	4.33	3.94	100.1173	4.32	2.57
4	100.0081	4.39	4.09	100.0435	4.38	2.97
5	100.0000	4.38	3.40	100.0167	4.37	2.76
$\sigma = 0.3, \eta = 0.031286$						Crank Nicolson Method
1	100.5882	4.01	N/A	100.8949	3.93	N/A
2	100.1448	4.12	N/A	100.3342	4.08	2.33
3	100.0338	4.16	4.00	100.1154	4.14	2.56
4	100.0072	4.18	4.17	100.0426	4.17	3.00
5	99.9996	4.17	3.48	100.0163	4.16	2.78

TABLE 6.3: *Convergence experiments for the GMWB guarantee value at $t = 0$ and $W = A = \omega_0 = 100$ using a fully implicit and Crank Nicolson method . Contract parameters are given in Table 6.1. The column "Central Differencing First" uses central differencing as much as possible for the V_W term in the equation. The column "For/Backward Differencing Only" uses forward or backward differencing for the V_W term in the equation. Itns/step refers to the average number of iterations per timestep for the lines $4 - 11$ in Algorithm 1. Ratio is the ratio of successive changes in the solution as the mesh/timesteps are refined. Since the no-arbitrage fee is imposed, the numerical solution should converge to $Value = \omega_0 = 100$.*

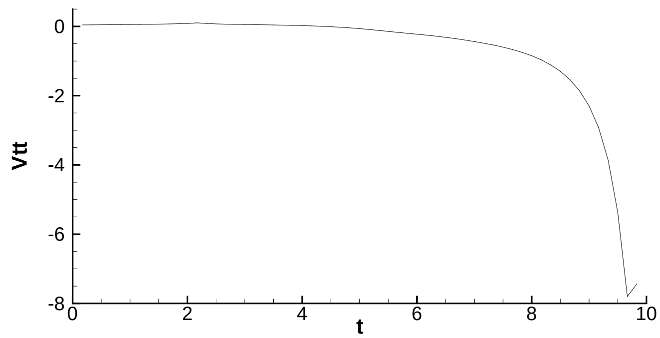

FIGURE 6.1: *V_{tt} versus t for node ($W = 100, A = 100$). $\sigma = 0.3$. Fair insurance fee (i.e. $\eta = 0.031286$) is imposed. Contract parameters are given in Table 6.1.*

6.3 Results

Table 6.3 presents the convergence results for the GMWB value with respect to two volatility values, assuming the no-arbitrage insurance fee is imposed.

Aside from fully implicit timestepping, we have also carried out some tests using Crank Nicolson timestepping, using an obvious modification of equation (4.13). Note that convergence has only been proven for the fully implicit method since Crank Nicolson timestepping is not monotone in general. The differencing method for the V_W term, which uses central differencing as much as possible, is also compared with forward or backward differencing only for the V_W term.

The Itns/step column in Table 6.3 shows the average number of iterations in each timestep required for lines $4-11$ in Algorithm 1. About $3-4$ non-linear iterations per timestep are required for the $\sigma = .2$ case, and about $4-5$ iterations per timestep are required in the $\sigma = .3$ case. The convergence ratio in the table is the ratio of successive changes in the solution, as the timestep and mesh size are reduced by a factor of two.

The number of iterations per timestep appears to be fairly insensitive to the grid size in Table 6.3. Note that since the timestep is reduced as the grid spacing is reduced, we have an excellent initial solution estimate at each timestep. This is consistent with the results for time dependent problems as reported in Bokanowski et al. (2009). For steady state problems, Santos and Rust (2004) and Bokanowski et al. (2009) report grid dependent number of iterations for policy iteration.

It can be seen that using central differencing as much as possible for the V_W term leads to more rapid convergence (as the mesh is refined) compared to pure forward or backward differencing for this term. Rather unexpectedly, the convergence ratios for both Crank Nicolson and fully implicit timestepping are similar. Figure 6.1 shows a plot of V_{tt} versus (forward) time, at the node ($W = 100, A = 100$). At $t = 0$ ($\tau = T$), we can see that $V_{tt} \simeq 0$, which would result in similar time truncation error for both Crank Nicolson and fully implicit timestepping.

Although the first column in Table 6.3 uses central differencing as much as possible, there are large regions in the solution domain where the optimal strategy is to withdraw a finite amount (an infinite rate), as shown in Figure 6.2. In these regions, forward or backward differencing is used in both the W and A directions, which should result in first order errors. However, in the finite

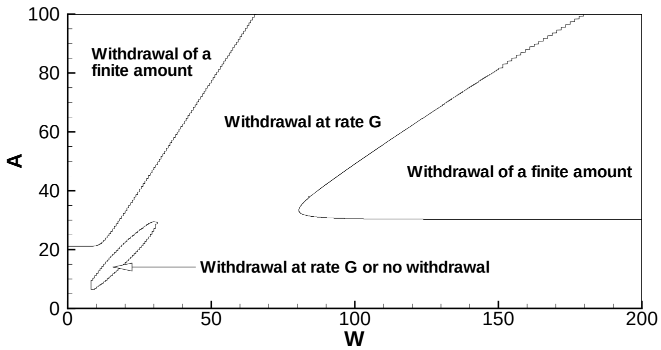

FIGURE 6.2: *The contour plot of optimal withdrawal strategy of the GMWB guarantee at* $t = \Delta\tau$ *in the* (W, A)*-plane.* $\sigma = 0.3$*. fair fee* $\eta = .031286$ *is imposed. Contract parameters are given in Table 6.1. This plot is similar to the results in Chen and Forsyth (2008).*

withdrawal amount (infinite withdrawal rate) regions, we essentially solve the PDE

$$1 - V_W - V_A = \kappa \ . \tag{6.6}$$

Noting that V is linear in A at $W = 0$, and linear in W as $W \to \infty$, then the solution of this PDE in the finite withdrawal region (assuming that this region is connected to $W = 0$ or $W \to \infty$) will be a linear function of (W, A), hence the use of forward or backward differencing is exact.

It is also interesting to see a region labeled *Withdrawal at rate G or no withdrawal.* Recall that in the finite withdrawal region, the solution satisfies

$$V_\tau = \mathcal{L}V + \max_{\gamma \in [0,G]} \left(\gamma(1 - V_W - V_A) \right) \ . \tag{6.7}$$

The solution in this region appears to converge to a value having $(1 - V_W - V_A) \simeq 0$. This suggests that the optimal control is a finite rate, but not unique, since either a rate of zero or G is optimal. The value function is, however, unique. This is consistent with the results in Chen and Forsyth (2008).

Since it appears (at least for this example) that fully implicit timestepping converges at a similar rate compared to Crank Nicolson, and that convergence can only be proven for fully implicit timestepping, it would appear that fully implicit timestepping is preferable to Crank Nicolson.

Recall that the no-arbitrage fee is determined by solving equation (6.1). Table 6.4 shows the convergence results in terms of the no-arbitrage fee for two different volatilities. The results are close to those reported in Chen and Forsyth (2008). Using central differencing first on the V_W term leads to faster convergence compared to using forward or backward differencing only on the V_W term.

It is also interesting to study the convergence of the penalty method for nodes near (or at) the finite withdrawal boundary. Figure 6.3 shows the location of the withdrawal boundaries at $A = 100$ versus t, when no insurance fee ($\eta = 0$) is imposed. Note that the node $(100, 100)$ is very near (or at) the boundary between a finite withdrawal rate and no withdrawal at $t = T$.

Examination of the solution near maturity (which is near the start of the numerical solution since we solve backwards in time) shows that the numerical solution changes between being in

Refinement	Central Differencing First		For/Backward Differencing Only	
Level	Fair Fee	Ratio	Fair Fee	Ratio
$\sigma = 0.2$	0.013891	(Value from (Chen and Forsyth, 2008))		
2	0.015207	N/A	0.015920	N/A
3	0.014245	N/A	0.014686	N/A
4	0.013961	3.38	0.014190	2.48
5	0.013886	3.80	0.013982	2.39
$\sigma = 0.3$	0.031258	(Value from (Chen and Forsyth, 2008))		
2	0.031904	N/A	0.032692	N/A
3	0.031431	N/A	0.031770	N/A
4	0.031319	4.22	0.031462	2.99
5	0.031286	3.43	0.031354	2.82

TABLE 6.4: *Convergence study for the fair insurance fee η value. Contract parameters are given in Table 6.1. Note that the results in Chen and Forsyth (2008) appear to be correct to about three (rounded) digits. The column "Central Differencing First" uses central differencing as much as possible for the V_W term in the equation. The column "For/Backward Differencing Only" uses forward or backward differencing for the V_W term in the equation.*

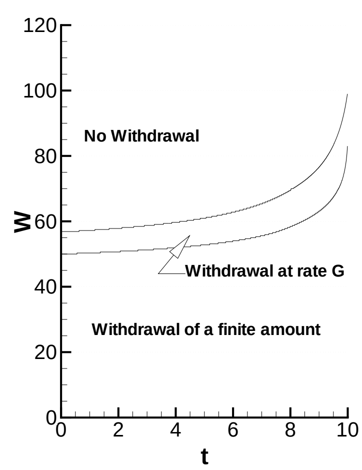

FIGURE 6.3: *The contour plot for the withdrawal boundary versus time t at $A = 100$, $\sigma = 0.3$. No insurance fee (i.e. $\eta = 0$) is imposed. Contract parameters are given in Table 6.1. Maximal use of central differencing on V_W term is applied.*

Fully Implicit Method						
Refinement	Central Differencing First			For/Backward Differencing Only		
	Value	Itns/step	Ratio	Value	Itns/step	Ratio
1	116.0354	2.88	N/A	116.2730	2.88	N/A
2	115.9134	2.89	N/A	116.0339	2.91	N/A
3	115.8879	2.97	4.78	115.9477	3.00	2.77
4	115.8845	3.10	7.52	115.9143	3.12	2.59
5	115.8859	3.25	-2.40	115.9008	3.26	2.47
6	115.8876	3.38	0.86	115.8950	3.39	2.33
extrapolated value from (Chen and Forsyth, 2008)				115.8897		

TABLE 6.5: *Convergence experiments for the GMWB guarantee value at $t = 0$ and $W = A = \omega_0 = 100$ by using the fully implicit method. $\sigma = 0.3$. No insurance fee ($\eta = 0$) is imposed. Contract parameters are given in Table 6.1. The column "Central Differencing First" use central differencing as much as possible for the V_W term. The column "For/Backward Differencing Only" uses forward or backward differencing for the V_W term. Itns/step refers to the average number of iterations per timestep for the lines $4-11$ in Algorithm 1. Ratio is the ratio of successive changes in the solution as the refinement is increased.*

the region of withdrawal at rate G to being in a region of zero withdrawal at refinement level 4 and above. This occurs when central differencing is used as much as possible. Table 6.5 gives the convergence results for this case ($\eta = 0$). We have proven that this method is convergent, but clearly convergence can be erratic at some exceptional nodes. Convergence (at this node) is smoother if the V_W term is discretized using a forward or backward differencing only.

In Section 6.2 we noted that the penalty method is convergent for any $C > 0$ such that $\varepsilon = C\Delta\tau$. We argued, based on financial reasoning that a good choice for ε is

$$\varepsilon \quad = \quad \frac{C^* \Delta\tau}{\omega_0} \ , \tag{6.8}$$

with C^* being a dimensionless constant. All the tests reported thus far use $C^* = 10^{-2}$. Table 6.6 shows the results at $W = A = \omega_0$, $t = 0$, with no insurance fee being imposed, for values of $C^* \in [1, 10^{-9}]$. The choice of C^* affects the solution only in the seventh digit for $C^* \in [10^{-2}, 10^{-7}]$.

In our initial tests varying C^*, we noticed convergence problems for $C^* < 10^{-7}$. Recall that in infinite precision arithmetic, the right hand side of equation (4.35) must always be non-negative. Analysis of the numerical experiments showed that at points near the withdrawal boundaries, for $C^* < 10^{-7}$, the right hand side of equation (4.35) was negative (at some iterations) at the level of machine precision. After solving equation (4.35), this caused changes in the solution in the eighth digit, which violated the convergence criteria (*tolerance* $= 10^{-8}$ in Algorithm 1). The iterations would then oscillate between two states, with positive and negative right hand sides of equation (4.35). This problem was eliminated by simply forcing the right hand side of equation (4.35) to be always non-negative. This, of course, would always be true in infinite precision arithmetic. Consequently, if C^* is selected too small, then this generates problems due to numerical precision issues. However, this is only a difficulty for very small C^*, very fine grids, and perhaps unrealistic convergence criteria. This issue is also discussed in Forsyth and Vetzal (2002).

	$\sigma = 0.2$		$\sigma = 0.3$	
ε	Value	Itns/step	Value	Itns/step
$\Delta\tau/\omega_0$	107.7315	3.24	115.8828	3.25
$10^{-1}\Delta\tau/\omega_0$	107.7336	3.24	115.8856	3.25
$10^{-2}\Delta\tau/\omega_0$	107.7338	3.24	115.8859	3.25
$10^{-3}\Delta\tau/\omega_0$	107.7339	3.24	115.8860	3.25
$10^{-4}\Delta\tau/\omega_0$	107.7339	3.34	115.8860	3.25
$10^{-5}\Delta\tau/\omega_0$	107.7339	3.24	115.8860	3.25
$10^{-6}\Delta\tau/\omega_0$	107.7339	3.24	115.8860	3.25
$10^{-7}\Delta\tau/\omega_0$	107.7338	3.38	115.8860	3.31
$10^{-8}\Delta\tau/\omega_0$	107.7338	3.46	115.8859	3.45
$10^{-9}\Delta\tau/\omega_0$	107.7294	4.82	115.8776	5.25

TABLE 6.6: *The effect of the penalty parameter at refinement level 5.* $W = A = 100$ *and* $t = 0$*. No insurance fee (i.e.* $\eta = 0$*) is imposed. Contract parameters are given in Table 6.1. Itns/step refers to the average number of iterations per timestep for the lines* $4 - 11$ *in Algorithm 1.*

6.4 Comparison: Penalty Method (Singular Control) and Impulse Control

As outlined in (Zakamouline, 2005), it is almost always possible to formulate a singular control problem as an impulse control problem, with arbitrarily small error. It is therefore interesting to consider the computational issues for both formulations.

If h is the discretization parameter (as in Assumption 5.1), then the computational complexity of the penalty method, singular control formulation is

$$\text{Complexity: Penalty method} \quad = \quad C' h^{-3} \tag{6.9}$$

where C' is the average number of iterations per step. Since it appears that C' is independent of h, then the complexity of the penalty method is $O(h^{-3})$.

In the impulse control formulation, the numerical method described by Chen and Forsyth (2008) has a complexity of $O(h^{-4})$. This is due to the linear search required in the local optimization step of the algorithm in Chen and Forsyth (2008). The linear search guarantees location of the global maximum with $O(h)$ error for smooth test functions.

On the basis of complexity, it would appear that the penalty method is a clear winner. However, as noted in Chen and Forsyth (2008), it is trivial to handle discrete withdrawal times and complex contract features using an impulse control formulation. These generalizations may be very difficult to handle with a singular control formulation. Zakamouline (2005) suggests that an impulse control formulation is preferred in general. In addition, the experimental convergence rate in Chen and Forsyth (2008) is smooth as the mesh is refined. This contrasts with the sometimes erratic convergence of the penalty method for nodes near the withdrawal boundaries. As well, the impulse control formulation does not require an estimate of the constant for the penalty parameter. There also appears to be a limit on the solution accuracy, due to numerical precision problems, with the penalty method. However, this limit is probably at a level of accuracy which is far beyond what would be required in practice.

7 Conclusions

In this paper, we study the penalty algorithm proposed by Dai et al. (2008) to price GMWB variable annuities. Provided the original problem satisfies a strong comparison property, we prove that the penalty algorithm converges to the unique viscosity solution of the HJB variational inequality corresponding to the singular control model formed in Dai et al. (2008).

We find that using central differencing as much as possible (Wang and Forsyth, 2008) results in noticeably faster convergence (as the grid/timesteps are refined) compared to forward or backward differencing only discretization.

Our experimental results show that the penalty method has some limitations in determining the withdrawal boundaries to high accuracy. For nodes near the withdrawal boundaries, convergence is somewhat erratic.

However the penalty method is very easy to implement, and convergence is fast to a level of accuracy probably far beyond what would be required in practice. This method has a lower complexity than the impulse control approach in Chen and Forsyth (2008), but at the expense of some loss of generality.

The penalty method can be easily applied to a wide variety of singular stochastic control problems.

Appendix

A Hedging Argument for (2.8)

In this Appendix, we give an informal hedging argument for deriving equation (2.8). Consider the following scenario. The underlying asset W (a mutual fund) in the investor's account follows the process

$$dW = (\mu - \eta)W dt + W\sigma dZ \ , \tag{A.1}$$

where μ is the drift rate, η is the fee for the guarantee, and dZ is the increment of a Wiener process.

We assume that the mutual fund tracks an index $\hat{W}$ which follows the process

$$d\hat{W} = \mu\hat{W} dt + \hat{W}\sigma dZ \ . \tag{A.2}$$

We assume that it is not possible to short the mutual fund, so that the obvious arbitrage opportunity cannot be exploited. (This is typically a fiduciary requirement). We further assume that it is possible to track the index $\hat{W}$ without basis risk.

Now, consider the writer of the GMWB contract, with no-arbitrage value $V(W, A, t)$. The writer sets up the hedging portfolio

$$\Pi(W, \hat{W}, t) = -V(W, t) + x\hat{W} \ , \tag{A.3}$$

where x is the number of units of the index $\hat{W}$.

Over the time interval $t \to t + dt$, assuming that Ito's Lemma can be used, we obtain

$$\begin{aligned} d\Pi = &-\left[\left(V_t + (\mu-\eta)WV_W + \frac{1}{2}\sigma^2 W^2 V_{WW} + f(\gamma) - \gamma V_W - \gamma V_A\right) dt + \sigma W V_W dZ\right] \\ &+ x[\mu\hat{W} dt + \sigma\hat{W} dZ] \ , \end{aligned} \tag{A.4}$$

where γ is the (finite) rate of withdrawal by the contract holder.

Choose

$$x = \frac{W}{\hat{W}} V_W, \tag{A.5}$$

so that equation (A.4) becomes

$$d\Pi = -\left[\left(V_t - \eta W V_W + \frac{1}{2}\sigma^2 W^2 V_{WW} + f(\gamma) - \gamma V_W - \gamma V_A\right) dt\right]. \tag{A.6}$$

The worst case for the hedger will be when the contract holder chooses an action to minimize the value of the hedging portfolio (this of course corresponds to the contract holder maximizing her no-arbitrage long position), so that

$$d\Pi = \min_{\gamma}\left[-\left(V_t - \eta W V_W + \frac{1}{2}\sigma^2 W^2 V_{WW} + f(\gamma) - \gamma V_W - \gamma V_A\right) dt\right]. \tag{A.7}$$

Let r be the risk free rate, and so setting $d\Pi = r\Pi\ dt$ (since the portfolio is now riskless) gives

$$\begin{aligned} r\left(-V + V_W W\right) &= -\max_{\gamma}\left[\left(-V_\tau - \eta W V_W + \frac{1}{2}\sigma^2 W^2 V_{WW} + f(\gamma) - \gamma V_W - \gamma V_A\right)\right] \\ &= V_\tau + \eta W V_W - \frac{1}{2}\sigma^2 W^2 V_{WW} - \max_{\gamma}\left[f(\gamma) - \gamma V_W - \gamma V_A\right], \end{aligned} \tag{A.8}$$

which is equation (2.8).

Another way to verify this equation is the following. Imagine that the hedger replicates the cash flows associated with the total GMWB contract. In this case, the underlying mutual fund can be regarded as a purely virtual instrument, following process (A.1). The actual hedging instrument on the other hand follows process (A.2). Having elminated the random term by delta hedging, the hedger then assumes the worst case which occurs when the contact holder maximizes (deterministically) the no-arbitrage value of the contract. In this case, $V = U + W$, where V is the value of the entire contract, and U is the value of the guarantee. We can obtain an equation for the guarantee portion U by substituting $V = U + W$ into equation (A.8).

Chen et al. (2008) use a similar argument to value the guarantee portion of the GMWB using the impulse control formulation.

Of course, the above arguments assume that the rate of withdrawal is finite, and that the solution is sufficiently smooth so that Ito's Lemma can be applied. These assumptions are not in general valid (i.e. we take the limit as the maximum withdrawal rate becomes infinite), and a much more careful analysis is required to derive the singular control problem in rigorous fashion. Delta hedging strategies for GMWB contacts are commonly used in the insurance industry (Bauer et al., 2008; Gilbert et al., 2007), although usually based on the impulse control formulation.

B Finite Difference Approximation

In this appendix, we use standard finite difference method to approximate the first and second partial derivatives in the PDE. The discretized differential operators $\mathcal{D}_A^h$, $\mathcal{D}_W^h$ and $\mathcal{D}_{WW}^h$ are given by

$$\mathcal{D}_A^h V_{i,j}^n = \frac{V_{i,j}^n - V_{i,j-1}^n}{\Delta A_j^-}, \quad \text{backward differencing,} \tag{B.1}$$

$$\mathcal{D}_W^h V_{i,j}^n = \begin{cases} \frac{V_{i,j}^n - V_{i-1,j}^n}{\Delta W_i^-} & \text{backward differencing,} \\ \frac{V_{i+1,j}^n - V_{i,j}}{\Delta W_i^+} & \text{forward differencing,} \\ \frac{V_{i+1,j}^n - V_{i-1,j}^n}{\Delta W_i^\pm} & \text{central differencing,} \end{cases} \tag{B.2}$$

$$\mathcal{D}_{WW}^h V_{i,j}^n = \frac{\frac{V_{i-1,j}^n - V_{i,j}^n}{\Delta W_i^-} + \frac{V_{i+1,j}^n - V_{i,j}^n}{\Delta W_i^+}}{\frac{\Delta W_i^\pm}{2}}, \tag{B.3}$$

where

$$\Delta A_j^- = A_j - A_{j-1}, \quad \Delta W_i^- = W_i - W_{i-1}, \quad \Delta W_i^+ = W_{i+1} - W_i, \quad \text{and} \quad \Delta W_i^\pm = W_{i+1} - W_{i-1}.$$

C Discrete Equation Coefficients

Let $\{\varphi_{i,j}^n, \psi_{i,j}^n\}$ denote the optimal local control parameter value for node (W_i, A_j, τ^n).

$$\begin{aligned} & \mathcal{A}_{\varphi_{i,j}^n, \psi_{i,j}^n}^h V_{i,j}^n \\ = \; & a_{i,j}^n(\varphi_{i,j}^n, \psi_{i,j}^n) \mathcal{D}_{WW}^h V_{i,j}^n + b_{i,j}^n(\varphi_{i,j}^n, \psi_{i,j}^n) \mathcal{D}_W^h V_{i,j}^n - c_{i,j}^n(\varphi_{i,j}^n, \psi_{i,j}^n) V_{i,j}^n \\ = \; & \alpha_{i,j}^n V_{i-1,j}^n - (\alpha_{i,j}^n + \beta_{i,j}^n + c_{i,j}^n) V_{i,j}^n + \beta_{i,j}^n V_{i+1,j}^n. \end{aligned}$$

If central differencing is used for the $\mathcal{D}_W^h V_{i,j}^n$ term, then

$$\begin{aligned} \alpha_{i,j,cent}^n &= \frac{2a_{i,j}^n(\varphi_{i,j}^n, \psi_{i,j}^n)}{\Delta W_i^\pm \Delta W_i^-} - \frac{b_{i,j}^n(\varphi_{i,j}^n, \psi_{i,j}^n)}{\Delta W_i^\pm}, \\ \beta_{i,j,cent}^n &= \frac{2a_{i,j}^n(\varphi_{i,j}^n, \psi_{i,j}^n)}{\Delta W_i^\pm \Delta W_i^+} + \frac{b_{i,j}^n(\varphi_{i,j}^n, \psi_{i,j}^n)}{\Delta W_i^\pm}. \end{aligned} \tag{C.1}$$

When a forward/backward differencing is used for the $\mathcal{D}_W^h V_{i,j}^n$ term, we obtain

$$\begin{aligned} \alpha_{i,j,for/back}^n &= \frac{2a_{i,j}^n(\varphi_{i,j}^n, \psi_{i,j}^n)}{\Delta W_i^\pm \Delta W_i^-} + \max\left[0, \frac{-b_{i,j}^n(\varphi_{i,j}^n, \psi_{i,j}^n)}{\Delta W_i^-}\right], \\ \beta_{i,j,for/back}^n &= \frac{2a_{i,j}^n(\varphi_{i,j}^n, \psi_{i,j}^n)}{\Delta W_i^\pm \Delta W_i^+} + \max\left[0, \frac{b_{i,j}^n(\varphi_{i,j}^n, \psi_{i,j}^n)}{\Delta W_i^+}\right]. \end{aligned} \tag{C.2}$$

where

$$\Delta W_i^- = W_i - W_{i-1}, \quad \Delta W_i^+ = W_{i+1} - W_i, \quad \text{and} \quad \Delta W_i^\pm = W_{i+1} - W_{i-1}.$$

References

Barles, G. (1997). Convergence of numerical schemes for degenerate parabolic equations arising in finance. In L. C. G. Rogers and D. Talay, eds., *Numerical Methods in Finance*, pp. 1–21. Cambridge University Press, Cambridge.

Barles, G., C. Daher, and M. Romano (1995). Convergence of numerical shemes for parabolic eqations arising in finance theory. *Mathematical Models and Methods in Applied Sciences* 5, 125–143.

Barles, G. and P. Souganidis (1991). Convergence of approximation schemes for fully nonlinear equations. *Asymptotic Analysis* 4, 271–283.

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

Bokanowski, O., S. Maroso, and H. Zidani (2009). Some convergence results for Howard's algorithm. *SIAM Journal on Numerical Analysis* 47, 3001–3026.

Chen, Z. and P. A. Forsyth (2008). A numerical scheme for the impulse control formulation for pricing variable annuities with a guaranteed minimum withdrawal benefit (GMWB). *Numerische Mathematik* 109, 535–569.

Chen, Z., K. Vetzal, and P. Forsyth (2008). The effect of modelling parameters on the value of gmwb guarantees. *Insurance: Mathematics and Economics* 43, 165–173.

Cramer, E., P. Matson, and L. Rubin (2007). Common practices relating to FASB statement 133, Accounting for Derivative Instruments and Hedging Activities as it Relates to Variable Annuities with Guaranteed Benefits. Practice Note, American Academy of Actuaries, available at www.actuary.org.

Dai, M., Y. K. Kwok, and J. Zong (2008). Guaranteed minimum withdrawal benefit in variable annuities. *Mathematical Finance* 18, 595–611.

Dai, M. and Y. Zhong (2010). Penalty methods for continuous time portfolio selection with transaction costs. To appear, Journal of Computational Finance.

Fenton, J. and D. Czernicki (2010). Pricing GMWB features in a challenging market. Towers Watson Report, available at www.towerswatson.com.

Forsyth, P. and K. Vetzal (2002). Quadratic convergence of a penalty method for valuing American options. *SIAM Journal on Scientific Computing* 23, 2095–2122.

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

Gilbert, C., K. Ravindran, and R. Reitano (2007). Results on the survey on variable annuity hedging programs for life insurance companies. Society of Actuaries Report, available at www.soa.org.

Jakobsen, E. (2010). Monotone schemes. In R. Cont, ed., *Ecyclopedia of Quantitative Finance*, pp. 1253–1263. Wiley, New York.

Kushner, H. (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time.* Springer.

Milevsky, M. A. and T. S. Salisbury (2006). Financial valuation of guaranteed minimum withdrawal benefits. *Insurance: Mathematics and Economics* 38, 21–38.

Pham, H. (2005). On some recent aspects of stochastic control and their applications. *Probability Surveys* 2, 506–549.

Pham, H. (2009). *Continuous-time Stochastic Control and Optimization with Financial Applications.* Springer.

Pooley, D., P. Forsyth, and K. Vetzal (2003). Numerical convergence properties of option pricing PDEs with uncertain volatility. *IMA Journal of Numerical Analysis* 23, 241–267.

Santos, M. and J. Rust (2004). Convergence properties of policy iteration. *SIAM Journal on Control and Optimization* 42, 2094–2115.

Seydel, R. (2009). Existence and uniqueness of viscosity solutions for QVI associated with impulse control of jump-diffusions. *Stochastic Processes and their Applications* 119, 3719–3748.

Tourin, A. and T. Zariphopoulou (1997). Viscosity solutions and numerical schemes for investment/consumption models with transaction costs. In *Numerical Methods in Finance*, editors: L.C.G. Rogers and D. Talay, Cambridge University Press.

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

Zakamouline, V. I. (2005). A unified approach to portfolio optimization with linear transaction costs. *Mathematical Methods of Operations Research* 62, 319–343.