https://doi.org/10.1016/j.automatica.2019.108668

Power Penalty Method For Solving HJB Equations Arising From Finance

Kai Zhang[1,*] Xiaoqi Yang[2] and Yaohua Hu[3]
[1] Shenzhen Audencia Business School,
[3] School of Mathematics and Statistics,
Shenzhen University, Shenzhen, China, 518060
[1] kaizhang@szu.edu.cn [3] mayhhu@szu.edu.cn
[2] Department of Applied Mathematics,
The Hong Kong Polytechnic University, Hong Kong
[2] mayangxq@polyu.edu.hk

Abstract

Nonlinear Hamilton-Jacobi-Bellman (HJB) equation commonly occurs in financial modeling. Implicit numerical scheme is usually applied to the discretization of the continuous HJB so as to find its numerical solution, since it is generally difficult to obtain its analytic viscosity solution. This type of discretization results in a nonlinear discrete HJB equation. We propose a power penalty method to approximate this discrete equation by a nonlinear algebraic equation containing a power penalty term. Under some mild conditions, we give the unique solvability of the penalized equation and show its convergence to the original discrete HJB equation. Moreover, we establish a sharp convergence rate of the power penalty method, which is of an exponential order with respect to the power of the penalty term. We further develop a damped Newton algorithm to iteratively solve the lower order penalized equation. Finally, we present a numerical experiment solving an incomplete market optimal investment problem to demonstrate the rates of convergence and effectiveness of the new method. We also numerically verify the efficiency of the power penalty method by comparing it with the widely used policy iteration method.

Keywords. HJB equation, penalty method, convergence, optimal portfolio, policy iteration.
AMS subject classifications. 65N12, 65K10, 91B28.

1 Introduction

There are a number of stochastic optimal control problems naturally linked to the financial modeling of real world problems ([9]). Examples include the derivatives pricing, risk management ([13]), and optimal portfolio ([23]). Applying the dynamic programming principle to the stochastic control problem yields a type of nonlinear and degenerate second order partial differential equations (PDEs), called HJB equations. Generally, it is very difficult to get the analytical solution of the HJB equation due to its nonlinear and non-smooth structure. Hence, numerical approximations are widely used to solve this problem. The basic idea of numerical approximations is to discretize the continuous HJB first, which results in a discrete HJB, then to apply some iterative solution algorithm to solving the discrete nonlinear algebraic system.

Implicit methods are usually chosen over explicit method in the discretization of the continuous HJB equations. In view of the nonlinearity and non-smoothness of the HJB equations, one need to consider the viscosity solution ([4]) when studying the numerical approximation. In financial context, the viscosity solution of the nonlinear HJB equation is a viable and financial relevant one. As pointed in [2, 8], a numerical discretization of the continuous HJB equation need to be consistent, stable and monotone to ensure its convergence in the sense of viscosity solution. Hence, implicit methods result in a system of nonlinear, non-smooth discrete HJB equations.

It is a key task to find an efficient solution method to solving these discrete HJB equations. In the literature there exist two main types of solution methods, i.e., value iteration method and policy iteration method, cf. [7]. The value iteration method is a relaxation method and cheap to be implemented. However, it suffers from the linear convergence rate and hence its computational efficiency is very low. On the contrary, the policy iteration method, also called Howard's algorithm ([3]), is very effective to solve the discrete HJB equations in most cases. It is in particular well behaved in the case of the discrete HJB equations with sparsity or good structure. The good performance of the policy iteration method owes to its equivalence to the generalized Newton method, which makes it process a superlinear convergence rate, see [3].

Nevertheless, there exist some cases where the policy iteration method loses its efficiency. For example, when the underlying asset follows a jump diffusion process ([16]) or a regime switching process ([21]), the resulted discrete HJB equations will include global or coupled operator and hence lose its sparsity or good structure. In these cases, it turns out that the policy iteration method also encounters computational inefficiency issue. Moreover, as stated in [3, 11, 18], the policy iteration method has some rough similarity to the simplex algorithm of linear programming, which makes its computational cost increasing exponentially as the size of the discrete HJB equations increases. This becomes computational bottleneck when high accuracy is concerned, which is usually a natural requirement in financial modeling.

Inspired by the penalty approach to complementarity problems in [5, 26, 27], a linear (l_1) penalty method was develop in [22] to approximate to solutions of the discrete HJB equations in finance. By virtue of the semi-smoothness of the l_1 penalty function, they showed that the convergence rate of the solution of the penalized equation to that of the original discrete HJB equation is of first order with respect to the penalty parameter $\lambda > 0$, i.e., $\mathcal{O}(1/\lambda)$. More importantly, they numerically verified that the computation cost only increases linearly as the size of the discrete HJB equations increases, which is a much more desirable result over the policy iteration method. One drawback of the l_1 penalty method is that its solutions only satisfy the original HJB equations approximately. Though this can be alleviated by enlarging the penalty parameters λ, it is well known that the too large penalty parameters can cause computational difficulties. To overcome these difficulties, the lower order $(l_{1/k}, k > 1)$ penalty methods have been used extensively for solving various complementarity problems arising in finance ([5, 25, 27]) in recent years. This is because the lower order penalty methods have the merit that they have exponential convergence rates with respect to the penalty parameters λ, i.e. $\mathcal{O}(1/\lambda^k)$. Therefore, a desirable accuracy in the approximate solution can be achieved by a much smaller value of the penalty parameter. Moreover, the power penalty method possesses the advantage that the resulting penalized equation is of a simple form that is easy to discretize in any dimensions on both structured and unstructured meshes, hence can easily handle various types of problems in finance.

Motivated by the lower order penalty approach to complementarity problems in [19] and the linear penalty approach to the discrete HJB equations in [22], we propose in this work a power $\left(l_{1/k}, k > 0\right)$ penalty method, which includes the linear penalty method $(k = 1)$ and the lower order penalty method $(k > 1)$, for solving the discrete HJB equation arising from finance modeling and study its mathematical properties such as its solvability and convergence. In this approach, we approximate the discrete HJB equations by a system of nonlinear algebraic equations containing a power penalty term. We will establish a sharp convergence rate for the approximate solution and design a solution method to solve the lower order penalized equation. To the best of our knowledge, there are no studies on power penalty methods for solving large scale discrete HJB equations arising particularly in finance, though such methods have been studied extensively for solving general linear and nonlinear programming problems ([15]).

The main contributions of this paper are summarized as follows:

- We first propose a power penalty approach to the discrete HJB equations in finance and show its solvability.

- We prove the convergence property of the power penalty method and establish an exponential convergence rate with respect to the penalty parameter λ, i.e., $\mathcal{O}(1/\lambda^k)$.

- We design a damped Newton's method to solve the lower order penalized equations and numerically demonstrate the computational efficiency by comparing with the classic policy iteration method.

The remainder of this paper is organized as follows. In the next section, we give some standard definitions and assumptions and state the discrete HJB equation. In Section 3, we propose the power penalty approach and show that the penalized equation is uniquely solvable under some assumptions. In Section 4, we propose a convergence theory for the power penalty approach and particularly establish an exponential convergence rate. We develop a damped Newton's method to solve the lower order penalized equation in Section 5. In Section 6, numerical experiments using an incomplete market optimal investment problem are designed to demonstrate the rates of convergence and the effectiveness of the method. We also, in this section, numerically verify the efficiency of the power penalty method by comparing to the widely used policy iteration method. Finally, some conclusions are drawn in the last section.

2 Problem formulation

Consider the following continuous HJB equation: Find $V(x,t) : \Omega \times (0,T] \mapsto \mathbb{R}$, such that

$$\inf_{q\in\mathcal{Q}} \{\mathcal{L}^q V(x,t)\} = 0, \tag{1}$$

for $x \in \Omega \subset \mathbb{R}^n$, $t \in (0,T]$ with appropriate boundary and initial/terminal conditions, where q is the control parameter, the set of feasible controls $\mathcal{Q} \subset \mathbb{R}$ is a nonempty compact set, and $\mathcal{L}^q$ $(q \in \mathcal{Q})$ is the linear differential operator of the form

$$\mathcal{L}^q V = V_\tau - \left[\sum_{i,j=1}^{n} \sigma_{ij}\left(x,\tau,q\right) V_{x_i x_j} + \sum_{i=1}^{n} \mu_i\left(x,\tau,q\right) V_{x_i} - r\left(x,\tau,q\right) V + f\left(x,\tau,q\right)\right] \tag{2}$$

with $\tau = T - t$. In a financial context it is usually assumed that $\sigma_{ii}\left(x,\tau,q\right) \geq 0$ and $r\left(x,\tau,q\right) \geq 0$ in (2), which corresponds to non-negative volatilities and interest rates.

We will make use of a rather general class of discretizations, which usually is consistent, stable and monotone to ensure the convergence to the viscosity solution of the continuous HJB equation (1), cf. for example [2, 25]. We aim to study the following discrete HJB equations resulted from these types of discretizations.

Problem 2.1. Find $x \in \mathbb{R}^N$, such that

$$\min_{Q\in\mathcal{Q}^N} \{A\left(Q\right)x - b\left(Q\right)\} = 0, \tag{3}$$

where for every $Q = (q_1, \cdots, q_N)^\top$, $A\left(Q\right) : \mathcal{Q}^N \to \mathbb{R}^{N\times N}$ and $b\left(Q\right) : \mathcal{Q}^N \to \mathbb{R}^N$ refer to a $N \times N$ matrix and a vector in $\mathbb{R}^N$ associated to the control Q, respectively defined by $A(Q) := (a_{ij}(q_i))$ and $b(Q) := (b_1(q_1), \cdots, b_N(q_N))^\top$.

For the sake of concreteness, we introduce the following notations. Denote by $\mathcal{M}$ the set of real-valued $N \times N$ matrices, and let $\mathbb{I}$ be the set of $\{1, \cdots, N\}$. Throughout this paper, for every $x, y \in \mathbb{R}^N$, the notation $y \geq x$ means that $y_i \geq x_i$, $\forall i \in \mathbb{I}$. We also denote by $\min\{x, y\}$ (resp. $\max\{x, y\}$) the vector with components $\min(x_i, y_i)$ (resp. $\max\{x_i, y_i\}$). The definitions extend trivially to other relational operators). With these notations, Equation (3) can also be stated as the following component form

$$\min_{Q \in \mathcal{Q}^N} \{A(Q)x - b(Q)\}_i = \min_{q_i \in \mathcal{Q}} \left\{ \sum_{j=1}^{N} a_{ij}(q_i) x_j - b_i(q_i) \right\} = 0, \quad i = 1, \cdots N.$$

It has been shown in [2] that when the HJB equation (1) satisfies the strong comparison principle, the maximum principle and the ellipticity condition, the discretized form $A(Q)$ of the operator $\mathcal{L}^q$ in (2), under a consistent, stable and monotone discretization scheme, will be always a strictly diagonally dominant M-matrix. As in most financial contexts the above conditions are satisfied ([2, 13]), throughout the paper we use an important assumption that $A(Q)$ is a strictly diagonally M-matrix. More precisely, we make the following assumptions on the system matrices $A(Q)$ and the vectors $b(Q)$ in Problem 2.1:

(A1) The matrix $A(Q)$ is a strictly diagonally dominant M-matrix for every $Q \in \mathcal{Q}^N$.

(A2) The functions $A(Q) : \mathcal{Q}^N \mapsto \mathcal{M}$ and $b(Q) : \mathcal{Q} \mapsto \mathbb{R}^N$ are continuous.

It is known that If $A = (a_{i,j})$ is a strictly diagonally dominant M-matrix, then we have $a_{ii} > 0$, $a_{ij} \leq 0, i \neq j$, for $i, j \in \mathbb{I}$, and $a_{ii} - \sum_{i \neq j} |a_{ij}| > 0$. Moreover, $A^{-1} > 0$, cf. ([20]).

It has been shown in ([22]) that under the assumptions (A1) and (A2), the discrete HJB Problem 2.1 is uniquely solvable.

3 Power penalty approach to the discrete HJB equation

The idea of power penalty approach to the discrete HJB equation originates from that to the complementarity problems in [17]. In the power penalty approach to complementarity problems, a power penalty term is used to penalize the violation of one of the constraints and then is added to the other constraints, which results in a system of nonlinear algebraic equation containing the power penalty term. As complementarity problems can be sometimes viewed as an HJB equation ([3]), we generalize the power penalty approach to the discrete HJB equation (3) and get the following penalized problem.

Problem 3.1. Find $x_\lambda \in \mathbb{R}^N$, such that

$$A(\bar{Q}) x_\lambda - b(\bar{Q}) - \lambda \max_{Q \in \mathcal{Q}^N} [b(Q) - A(Q) x_\lambda]_+^{1/k} = 0, \tag{4}$$

where $\bar{Q} = (\bar{q}, \cdots, \bar{q})^\top \in \mathcal{Q}^N$ is arbitrarily chosen, $\lambda > 0$ is the penalty parameter and for any $y \in \mathbb{R}^N$, $[y]_+^{1/k} \doteq [(\max\{y_1, 0\})^{1/k}, \cdots, (\max\{y_N, 0\})^{1/k}]^\top$ is the power penalty term $l_{1/k}$ with $1/k > 0$ being the power.

It is worth noting that in Problem 3.1 we penalize the maximum violation of the constraints. The essence is to enforce all the constraints to be satisfied by letting $\lambda \to \infty$. As the power penalty term in (4) is nondecreasing, Lemma 3.1 directly follows from Lemmas 3.3 and 3.4 in [22].

Lemma 3.1 (Uniqueness and boundedness)**.** *Suppose there exists a solution x_λ to Problem 3.1, then it is unique. Moreover, the solution is bounded, i.e., there exists a constant $C > 0$, independent of λ and k, such that*

$$\|x_\lambda\|_\infty \leq C. \tag{5}$$

Now, we will show that Problem 3.1 has at least one solution in the following Lemma.

Lemma 3.2 (Existence)**.** *For any $\lambda > 0$, there exists a solution x_λ to Problem 3.1.*

Proof. For clarity, we omit the subscript λ of x_λ in this proof. We show that Problem 3.1 has a solution in a bounded region

$$S := \{x \in \mathbb{R}^N : -\varepsilon^{-1} e < b\left(Q\right) - A\left(Q\right) x < \delta^{-1} e,\ \text{for all } Q \in \mathcal{Q}^N\},$$

where $e = (1, ..., 1)^\top$ and ε and δ are (sufficiently small) positive constants. Let $F(x) := A(\bar{Q})x - b(\bar{Q}) - \lambda \max_{Q \in \mathcal{Q}^N} [b(Q) - A(Q)x]_+^{1/k}$. Clearly, $F = (f_1, \cdots, f_n) : \bar{S} \subset \mathbb{R}^N \mapsto \mathbb{R}^N$ is continuous. To prove this theorem, it suffices to verify the conditions of Miranda's theorem[1]. We first show that $F(x) \neq 0$ for x on the boundary ∂S of S. More specifically, we will show that $0 \notin F(\partial S)$ when both $\varepsilon > 0$ and $\delta > 0$ are sufficiently small. To prove this, we assume that $0 \in F(\partial S)$, that is, there exists an $x \in \partial S$ such that $F(x) = 0$. Then, we show this is not possible when both δ and ε are sufficiently small in the following two cases:

Case 1. Suppose there exits $l \in \{1, ..., N\}$ such that the lth component of $b\left(Q\right) - A\left(Q\right) x$ is $(b(Q) - A(Q)x)_l = \delta^{-1}$. Then, we have

$$f_l(x) = \left(A\left(\bar{Q}\right) x - b\left(\bar{Q}\right)\right)_l - \lambda \max_{Q \in \mathcal{Q}^N} \left(\left[b\left(Q\right) - A\left(Q\right) x\right]_+^{1/k}\right)_l = \left(A\left(\bar{Q}\right) x - b\left(\bar{Q}\right)\right)_l - \lambda \delta^{-1/k}.$$

[1]Let $G = \{x \in R^n : |x_i| < L,\ \text{for } 1 \leq i \leq n\}$ and suppose the mapping $F = (f_1, \cdots, f_n) : \bar{G} \to R_n$ is continuous on the closure $\bar{G}$ of G such that $F(x) \neq 0$ for x on the boundary ∂G of G, and

1. $f_i(x_1, x_2, \cdots, x_{i-1}, -L, x_{i+1}, \cdots, x_n) \leq 0$, for $1 \leq i \leq n$, and
2. $f_i(x_1, x_2, \cdots, x_{i-1}, L, x_{i+1}, \cdots, x_n) \geq 0$, for $1 \leq i \leq n$.

Then, $F(x) = 0$ has a solution in G. See [1].

It follows from $f_l(x) = 0$ that

$$x = A\left(\bar{Q}\right)^{-1}(\cdot,\cdot,\cdots,\underbrace{b_l\left(\bar{Q}\right)+\lambda\delta^{-1/k}}_{l\text{th}},\cdots,\cdot)^\top.$$

Since $A(\bar{Q})$ is an M-matrix, $A(\bar{Q})^{-1} := (\bar{a}_{ij})$ is non-negative. Also, there must be at least one index m such that $\bar{a}_{ml} \neq 0$, as otherwise, $A(\bar{Q})^{-1}$ is singular. Thus, combining $\bar{a}_{ml} \neq 0$ and non-negativity of $A(\bar{Q})^{-1}$, we have $x_m = \mathcal{O}(\bar{a}_{ml}(b_l + \lambda\delta^{-1/k})) \to \infty$ as $\delta \to +0$. This violates (5) in Lemma 3.1, and thus we conclude that when $\delta > 0$ is sufficiently small, $0 \notin F(\partial S)$ with $(b(\bar{Q}) - A(\bar{Q})x)_l = \delta^{-1}$ for a feasible l.

Case 2. We now consider the case x on ∂S such that at least one component of $b(Q) - A(Q)x$ is equal to $\mathcal{O}(-1/\varepsilon)$, say $(b\left(Q\right) - A\left(Q\right)x)_l = -\varepsilon^{-1}$ for a feasible index l. In this case we have

$$x = A\left(Q\right)^{-1}(\cdot,\cdot,...,\underbrace{b_l\left(q_l\right)+\varepsilon^{-1}}_{l\text{th}},...,\cdot)^\top.$$

Using the similar argument as in Case 1, we also have that Lemma 3.1 is violated by this x and thus $0 \notin F(\partial S)$ with $(b\left(Q\right) - A\left(Q\right)x)_l = -\varepsilon^{-1}$.

Combining the above two cases we see that when $\varepsilon > 0$ and $\delta > 0$ are both sufficiently small, $0 \notin F(\partial S)$.

Now, we will check whether the conditions $f_i(x_1, x_2, \cdots, x_{i-1}, -C, x_{i+1}, \cdots, x_n) \leq 0$ and $f_i(x_1, x_2, \cdots, x_{i-1}, C, x_{i+1}, \cdots, x_n) \geq 0$ are satisfied, where C is the constant defined in (5). In fact, it follows from $A(\bar{Q})$ is a strictly diagonally dominant M-matrix that $a_{ii}(\bar{Q}) > 0$, $a_{ij}(\bar{Q}) \leq 0, i \neq j$, for $i, j \in \mathbb{I}$, and $a_{ii}(\bar{Q}) - \sum_{i\neq j}\left|a_{ij}(\bar{Q})\right| > 0$. Hence, combing this property and $\|x\|_\infty \leq C$, we have for $1 \leq i \leq n$

$$\begin{aligned}
&f_i(x_1, x_2, \cdots, x_{i-1}, -C, x_{i+1}, \cdots, x_n)\\
=&\left(A\left(\bar{Q}\right)x - b\left(\bar{Q}\right)\right)_i - \lambda\max_{Q\in\mathcal{Q}^N}\left([b\left(Q\right) - A\left(Q\right)x]_+^{1/k}\right)_i\\
=&a_{i,i-1}\left(\bar{Q}\right)x_{i-1} - a_{i,i}\left(\bar{Q}\right)C + a_{i,i+1}\left(\bar{Q}\right)x_{i+1} - b_i\left(\bar{Q}\right)\\
&\quad - \lambda[b_i\left(Q^\lambda\right) - a_{i,i-1}\left(Q^\lambda\right)x_{i-1} + a_{i,i}\left(Q^\lambda\right)C - a_{i,i+1}\left(Q^\lambda\right)x_{i+1}]_+^{1/k}\\
\leq&0,
\end{aligned}$$

when C is sufficiently large. In the same way, we also have that when C is sufficiently large

$$\begin{aligned}
&f_i(x_1, x_2, \cdots, x_{i-1}, C, x_{i+1}, \cdots, x_n)\\
=&\left(A\left(\bar{Q}\right)x - b\left(\bar{Q}\right)\right)_i - \lambda\max_{Q\in\mathcal{Q}^N}\left([b\left(Q\right) - A\left(Q\right)x]_+^{1/k}\right)_i\\
=&a_{i,i-1}\left(\bar{Q}\right)x_{i-1} + a_{i,i}\left(\bar{Q}\right)C + a_{i,i+1}\left(\bar{Q}\right)x_{i+1} - b_i\left(\bar{Q}\right)\\
&\quad - \lambda[b_i\left(Q^\lambda\right) - a_{i,i-1}\left(Q^\lambda\right)x_{i-1} - a_{i,i}\left(Q^\lambda\right)C - a_{i,i+1}\left(Q^\lambda\right)x_{i+1}]_+^{1/k}\\
\geq&0,
\end{aligned}$$

for $1 \leq i \leq n$ with $Q^\lambda = \arg\max_{Q\in\mathcal{Q}^N}([b(Q) - A(Q)x_\lambda]_+^{1/k})$, since $A(Q^\lambda)$ is a strictly diagonally dominant M-matrix as well.

From the above analysis, we see that all the conditions of Miranda's theorem are satisfied. Hence, the existence of the solution to the penalized Problem 3.1 is proved. □

4 Convergence analysis of the power penalty approach

In this section, we will examine the convergence property of the power penalty method. We first show that the power penalty method possesses a monotonic convergence property with respect to the penalty parameter in Subsection 4.1, and then in Subsection 4.2 we establish an exponential rate of convergence.

4.1 Monotonic convergence property

To show the convergence property of the power penalty approach, we first introduce the definition of a lower solution of the discrete HJB Problem 2.1 and that of the penalized Problem 3.1.

Definition 4.1 (Lower solution)**.** $\bar{x}$ is called a lower solution of the discrete HJB Problem 2.1 if

$$\min_{Q\in\mathcal{Q}^N} \{A(Q)\,\bar{x} - b(Q)\} \leq 0.$$

$\bar{x}_\lambda$ is called a lower solution of Problem 3.1 if

$$A\left(\bar{Q}\right)\bar{x}_\lambda - b\left(\bar{Q}\right) - \lambda \max_{Q\in\mathcal{Q}^N} [b(Q) - A(Q)\,\bar{x}_\lambda]_+^{1/k} \leq 0.$$

For the lower solution of Problem 3.1, we present several properties in the following lemmas.

Lemma 4.1. *Let $\lambda > 1$ and $k > 0$ and x_λ be the solution of Problem 3.1. If $\bar{x}_\lambda$ is a lower solution of Problem 3.1, then $\bar{x}_\lambda \leq x_\lambda$.*

Proof. Since x_λ is the solution of Problem 3.1 and $\bar{x}_\lambda$ is a lower solution of Problem 3.1 , it follows from Definition 4.1 that

$$\begin{aligned} &A\left(\bar{Q}\right)\bar{x}_\lambda - b\left(\bar{Q}\right) - \lambda \max_{Q\in\mathcal{Q}^N} [b(Q) - A(Q)\,\bar{x}_\lambda]_+^{1/k} \\ \leq& 0 = A\left(\bar{Q}\right)x_\lambda - b\left(\bar{Q}\right) - \lambda \max_{Q\in\mathcal{Q}^N} [b(Q) - A(Q)\,x_\lambda]_+^{1/k}. \end{aligned}$$

This implies that

$$
\begin{aligned}
A\left(\bar{Q}\right)\left(\bar{x}_\lambda - x_\lambda\right) &\le \lambda \max_{Q\in\mathcal{Q}^N}\left[b\left(Q\right) - A\left(Q\right)\bar{x}_\lambda\right]_+^{1/k} - \lambda \max_{Q\in\mathcal{Q}^N}\left[b\left(Q\right) - A\left(Q\right)x_\lambda\right]_+^{1/k} \\
&\le \lambda \max_{Q\in\mathcal{Q}^N}\left\{\left[b\left(Q\right) - A\left(Q\right)\bar{x}_\lambda\right]_+^{1/k} - \left[b\left(Q\right) - A\left(Q\right)x_\lambda\right]_+^{1/k}\right\}.
\end{aligned}
$$

By defining $Q^\star = (q_1^\star, \cdots, q_N^\star)^\top$ to be such that

$$
Q^\star := \arg\max_{Q\in\mathcal{Q}^N}\left\{\left[b\left(Q\right) - A\left(Q\right)\bar{x}_\lambda\right]_+^{1/k} - \left[b\left(Q\right) - A\left(Q\right)x_\lambda\right]_+^{1/k}\right\},
$$

the above inequality becomes

$$
A\left(\bar{Q}\right)\left(\bar{x}_\lambda - x_\lambda\right) \le \lambda\left\{\left[b\left(Q^\star\right) - A\left(Q^\star\right)\bar{x}_\lambda\right]_+^{1/k} - \left[b\left(Q^\star\right) - A\left(Q^\star\right)x_\lambda\right]_+^{1/k}\right\}. \tag{6}
$$

Define two disjoint nonempty index subsets I_1 and I_2 of $\mathbb{I}$ as follows

$$
I_1 = \left\{i \,\middle|\, \left(\left[b\left(Q^\star\right) - A\left(Q^\star\right)\bar{x}_\lambda\right]_+^{1/k}\right)_i \le \left(\left[b\left(Q^\star\right) - A\left(Q^\star\right)x_\lambda\right]_+^{1/k}\right)_i\right\}, \tag{7}
$$

$$
I_2 = \left\{i \,\middle|\, \left(\left[b\left(Q^\star\right) - A\left(Q^\star\right)\bar{x}_\lambda\right]_+^{1/k}\right)_i > \left(\left[b\left(Q^\star\right) - A\left(Q^\star\right)x_\lambda\right]_+^{1/k}\right)_i\right\}. \tag{8}
$$

On one hand, it follows from (6) and (7) that

$$
\left(A\left(\bar{Q}\right)\left(\bar{x}_\lambda - x_\lambda\right)\right)_i \le 0, \quad \forall i \in I_1.
$$

On the other hand, by virtue of the monotonicity of the operator $[\cdot]_+^{1/k}$, (8) implies

$$
\left(A\left(Q^\star\right)\bar{x}_\lambda\right)_i \le \left(A\left(Q^\star\right)x_\lambda\right)_i, \quad \text{i.e.,} \quad \left(A\left(Q^\star\right)\left(\bar{x}_\lambda - x_\lambda\right)\right)_i \le 0, \quad \forall i \in I_2.
$$

Now, introducing a matrix, denoting $A^* \in \mathcal{M}$ to be the matrix having the ith row as that of of $(A(\bar{Q}))_i$, $i \in I_1$ and of $(A(Q^\star))_i$, $i \in I_2$. Hence, we have

$$
A^*\left(\bar{x}_\lambda - x_\lambda\right) \le 0,
$$

which implies on the whole index set $\mathbb{I}$

$$
\bar{x}_\lambda \le x_\lambda,
$$

since A^* is also a strictly diagonally dominant M-matrix. □

Lemma 4.2. *Let $\lambda_2 > \lambda_1 > 1$, and x_{λ_1} and x_{λ_2} be the solutions of Problem 3.1 with $\lambda = \lambda_1, \lambda_2$, respectively. Then x_{λ_1} is a lower solution of Problem 3.1 with $\lambda = \lambda_2$. Moreover, $x_{\lambda_1} < x_{\lambda_2}$.*

Proof. From the fact x_{λ_1} is the solution of Problem 3.1 with $\lambda = \lambda_1$ and $\lambda_2 > \lambda_1 > 1$, it follows that

$$\begin{aligned}&A\left(\bar{Q}\right)x_{\lambda_1} - b\left(\bar{Q}\right) - \lambda_2 \max_{Q\in\mathcal{Q}^N}\left[b\left(Q\right) - A\left(Q\right)x_{\lambda_1}\right]_+^{1/k}\\ \leq &A\left(\bar{Q}\right)x_{\lambda_1} - b\left(\bar{Q}\right) - \lambda_1 \max_{Q\in\mathcal{Q}^N}\left[b\left(Q\right) - A\left(Q\right)x_{\lambda_1}\right]_+^{1/k} = 0.\end{aligned}$$

This means that x_{λ_1} is a lower solution of Problem 3.1 with $\lambda = \lambda_2$. Furthermore, we have

$$x_{\lambda_1} < x_{\lambda_2},$$

which is a consequence of Lemma 4.1. □

Lemma 4.3. *Let $\lambda > 1$ and $k > 0$. Assume that x_λ and x^* are the solutions of Problem 3.1 and that of the discrete HJB Problem 2.1, respectively. Then x_λ is a lower solution of the discrete HJB Problem 2.1. Moreover, $x_\lambda \leq x^*$.*

Proof. First, define $Q^\lambda \in \mathcal{Q}^N$ to be such that for $i \in \mathbb{I}$

$$\max_{Q\in\mathcal{Q}^N}\left(\left[b\left(Q\right) - A\left(Q\right)x_\lambda\right]_+^{1/k}\right)_i = \left(\left[b\left(Q^\lambda\right) - A\left(Q^\lambda\right)x_\lambda\right]_+^{1/k}\right)_i.$$

Then, (4) becomes

$$A\left(\bar{Q}\right)x_\lambda - b\left(\bar{Q}\right) - \lambda\left[b\left(Q^\lambda\right) - A\left(Q^\lambda\right)x_\lambda\right]_+^{1/k} = 0. \tag{9}$$

Define two disjoint nonempty index subsets I_1 and I_2 of $\mathbb{I}$ as follows

$$I_1 = \left\{i \,\middle|\, \left(b\left(Q^\lambda\right) - A\left(Q^\lambda\right)x_\lambda\right)_i \leq 0\right\}, \tag{10}$$

$$I_2 = \left\{i \,\middle|\, \left(b\left(Q^\lambda\right) - A\left(Q^\lambda\right)x_\lambda\right)_i > 0\right\}. \tag{11}$$

Thus, we distinguish the following two cases.

- For $i \in I_1$, based on (10), we have

  $$\left(b\left(\bar{Q}\right) - A\left(\bar{Q}\right)x_\lambda\right)_i \leq \left(b\left(Q^\lambda\right) - A\left(Q^\lambda\right)x_\lambda\right)_i = \max_{Q\in\mathcal{Q}^N}\left(b\left(Q\right) - A\left(Q\right)x_\lambda\right)_i \leq 0.$$

  Nevertheless, it follows from (9) and (10) that

  $$\left(A\left(\bar{Q}\right)x_\lambda - b\left(\bar{Q}\right)\right)_i = 0, \quad i \in I_1. \tag{12}$$

  Combining the above two equations, we obtain

  $$\max_{Q\in\mathcal{Q}^N}\left(b\left(Q\right) - A\left(Q\right)x_\lambda\right)_i = 0, \quad i \in I_1,$$

  i.e.,

  $$\min_{Q\in\mathcal{Q}^N}\left(A\left(Q\right)x_\lambda - b\left(Q\right)\right)_i = 0, \quad i \in I_1. \tag{13}$$

- For $i \in I_2$, based on (11), we have

$$\left(b\left(Q^{\lambda}\right) - A\left(Q^{\lambda}\right)x_{\lambda}\right)_i = \max_{Q \in \mathcal{Q}^N} \left(b\left(Q\right) - A\left(Q\right)x_{\lambda}\right)_i > 0,$$

which means

$$\min_{Q \in \mathcal{Q}^N} \left(A\left(Q\right)x_{\lambda} - b\left(Q\right)\right)_i < 0, \quad i \in I_2. \tag{14}$$

Summarizing (13) and (14) we deduce that on the whole index set $\mathbb{I}$

$$\min_{Q \in \mathcal{Q}^N} \left\{A\left(Q\right)x_{\lambda} - b\left(Q\right)\right\} \leq 0.$$

This means that x_{λ} is a lower solution of the discrete HJB Problem 2.1.

Now, we will show that $x_{\lambda} \leq x^*$. We still distinguish the following two cases.

- For $i \in I_1$, we already have (see (12))

$$\left(A\left(\bar{Q}\right)x_{\lambda} - b\left(\bar{Q}\right)\right)_i = 0.$$

Moreover, from the fact x^* is the solution of the discrete HJB Problem 2.1 it follows that

$$\left(A\left(\bar{Q}\right)x^* - b\left(\bar{Q}\right)\right)_i \geq \left(A\left(Q^*\right)x^* - b\left(Q^*\right)\right)_i = \min_{Q \in \mathcal{Q}^N} \left(A\left(Q\right)x - b\left(Q\right)\right)_i = 0,$$

with $Q^* = \arg\min_{Q \in \mathcal{Q}^N} \left\{A\left(Q\right)x - b\left(Q\right)\right\}$. Thus, combining the above two equations, we obtain

$$\left(A\left(\bar{Q}\right)\left(x^* - x_{\lambda}\right)\right)_i \geq 0, \quad i \in I_1.$$

- For $i \in I_2$, we have $(b(Q^{\lambda}) - A(Q^{\lambda})x_{\lambda})_i > 0$, which is equivalent to

$$\left(A\left(Q^{\lambda}\right)x_{\lambda} - b\left(Q^{\lambda}\right)\right)_i < 0.$$

As the first case, it also holds that

$$\left(A\left(Q^{\lambda}\right)x^* - b\left(Q^{\lambda}\right)\right)_i \geq \min_{Q \in \mathcal{Q}^N} \left(A\left(Q\right)x - b\left(Q\right)\right)_i = 0.$$

Thus, combining the above two equations, we obtain

$$\left(A\left(Q^{\lambda}\right)\left(x^* - x_{\lambda}\right)\right)_i > 0, \quad i \in I_2.$$

Now, again introducing a matrix, still denoting $A^* \in \mathcal{M}$, to be the matrix having the ith row as that of $(A(\bar{Q}))_i$, $i \in I_1$ and of $(A(Q^{\lambda}))_i$, $i \in I_2$. Therefore, we have

$$A^*\left(x^* - x_{\lambda}\right) \geq 0.$$

Hence, providing that A^* is an M-matrix, we have that on the whole index set $\mathbb{I}$

$$x^* \geq x_{\lambda}.$$

$\square$

With the above lemmas, we now establish the following monotonic convergence result for the $l_{1/k}$ penalty method.

Theorem 4.1. *Let $\{\lambda_m\}, m = 1, 2, \cdots$, be a monotonically increasing sequence tending to positive infinity as $m \to \infty$. Assume that x_{λ_m} is the solution of Problem 3.1 with $\lambda = \lambda_m$. Then the sequence $\{x_{\lambda_m}\}$ is monotonically increasing and convergent to the solution x^* of Problem 2.1.*

Proof. It follows from Lemmas 4.2 and 4.3 that

$$x_{\lambda_1} \leq x_{\lambda_2} \leq \cdots \leq x_{\lambda_i} \leq \cdots \leq x^*.$$

This implies that there exists some $x^\star$ such that

$$\lim_{m\to\infty} x_{\lambda_m} = x^\star.$$

Since x_{λ_m} is the solution of Problem 3.1 with $\lambda = \lambda_m$, there must hold

$$A\left(\bar{Q}\right) x_{\lambda_m} - b\left(\bar{Q}\right) = \lambda_m \max_{Q\in\mathcal{Q}^N} \left[b\left(Q\right) - A\left(Q\right) x_{\lambda_m}\right]_+^{1/k} \geq 0. \tag{15}$$

Letting $m \to \infty$ in (15), we get

$$A\left(\bar{Q}\right) x^\star - b\left(\bar{Q}\right) \geq 0.$$

Furthermore, reforming (15) gives

$$\max_{Q\in\mathcal{Q}^N} \left[b\left(Q\right) - A\left(Q\right) x_{\lambda_m}\right]_+ = \left(\frac{A(\bar{Q})x_{\lambda_m} - b(\bar{Q})}{\lambda_m}\right)^k. \tag{16}$$

Thus, letting $m \to \infty$ in (16), we get

$$\max_{Q\in\mathcal{Q}^N} \left[b\left(Q\right) - A\left(Q\right) x^\star\right]_+ = 0,$$

since $A(\bar{Q})$, $b(\bar{Q})$ and x_{λ_m} are bounded. This implies that

$$b\left(Q\right) - A\left(Q\right) x^\star \leq 0, \quad \forall Q \in \mathcal{Q}^N.$$

Hence,

$$A\left(Q\right) x^\star - b\left(Q\right) \geq 0, \quad \forall Q \in \mathcal{Q}^N.$$

Specifically,

$$\min_{Q\in\mathcal{Q}^N} \left\{A\left(Q\right) x^\star - b\left(Q\right)\right\} \geq 0. \tag{17}$$

Nevertheless, it follows from Lemma 4.3 that $\{x_{\lambda_m}\}$ are lower solutions of the discrete HJB Problem 2.1, and hence

$$\min_{Q\in\mathcal{Q}^N} \left\{A\left(Q\right) x_{\lambda_m} - b\left(Q\right)\right\} \leq 0. \tag{18}$$

Letting $m \to \infty$ in (18), we get

$$\min_{Q \in \mathcal{Q}^N} \{A(Q)x^\star - b(Q)\} \leq 0. \tag{19}$$

In view of (17) and (19), we have

$$\min_{Q \in \mathcal{Q}^N} \{A(Q)x^\star - b(Q)\} = 0.$$

This shows that $x^\star$ solves the discrete HJB Problem 2.1. Since the discrete HJB Problem 2.1 has a unique solution (see Lemma 3.1), we obtain

$$\lim_{m \to \infty} x_{\lambda_m} = x^\star = x^*.$$

□

Remark 4.1. Though not stated explicitly, it follows from the above proof and the fact $\mathcal{Q}$ is compact that for every $i \in \{1, 2, \ldots, N\}$ there exits a $Q_i^* \in \mathcal{Q}$ such that $Q_i^\lambda \to Q_i^*$ as $\lambda \to \infty$ with

$$Q_i^\lambda := \arg\max_{Q \in \mathcal{Q}^N} \left([b(Q) - A(Q)x_\lambda]_+\right)_i, \text{ and } Q_i^* := \arg\min_{Q \in \mathcal{Q}^N} (A(Q)x - b(Q))_i.$$

4.2 Exponential convergence rate

To establish the convergence rate of the $l_{1/k}$ penalty method with respect to the penalty parameter λ, we first give an error estimation of the solution to Problem 3.1.

Theorem 4.2. *Assume that x_λ is the solution of Problem 2.1 for every $\lambda > 1$. There exists a constant $C > 0$, independent of λ, such that*

$$\left\| \min_{Q \in \mathcal{Q}^N} \{A(Q)x_\lambda - b(Q)\} \right\|_\infty \leq \frac{C}{\lambda^k}.$$

Proof. For $Q \in \mathcal{Q}^N$, we have

$$\lambda [b(Q) - A(Q)x_\lambda]_+^{1/k} \leq \lambda \max_{Q \in \mathcal{Q}^N} [b(Q) - A(Q)x_\lambda]_+^{1/k} = A(\bar{Q})x_\lambda - b(\bar{Q}) \leq C,$$

since it follows from Lemma 3.1 and [22, Corollary 2.6] that both x_λ and $A(\bar{Q})$ are bounded. Hence,

$$[b(Q) - A(Q)x_\lambda]_+ \leq \frac{C}{\lambda^k}. \tag{20}$$

Furthermore, for every $i \in \mathbb{I}$, we either have

$$(A(Q)x_\lambda - b(Q))_i \geq 0, \text{ and } \left(A(\bar{Q})x_\lambda - b(\bar{Q})\right)_i = 0 \leq \frac{C}{\lambda^k},$$

or $\exists\, Q^\lambda \in \mathcal{Q}$ such that $b(Q^\lambda) - A(Q^\lambda)x_\lambda \geq 0$, which, based on (20), gives

$$\left(A(Q^\lambda)x_\lambda - b(Q^\lambda)\right)_i = -\left(b(Q^\lambda) - A(Q^\lambda)x_\lambda\right)_i \geq -\frac{C}{\lambda^k}, \text{ and } \left(A(\bar{Q})x_\lambda - b(\bar{Q})\right)_i > 0.$$

Hence, both cases reduce to

$$\left\| \min_{Q\in\mathcal{Q}^N} \{A(Q)\, x_\lambda - b(Q)\} \right\|_\infty \leq \frac{C}{\lambda^k}.$$

□

We are now ready to show that the solution of Problem 3.1 is indeed a good approximation to that of the discrete HJB Problem 2.1, in the sense that the approximation converges exponentially with respect to the penalty parameter.

Theorem 4.3. *Assume that x_λ and x^* are the solution of Problem 3.1 and that of Problem 2.1, respectively. Then for sufficiently large λ, we have*

$$\|x^* - x_\lambda\|_\infty \leq \frac{C}{\lambda^k}, \tag{21}$$

where $C > 0$ is a constant, independent of x^*, x_λ and λ.

Proof. For $\lambda > 0$, since $A(\bar{Q})x_\lambda - b(\bar{Q}) = \lambda \max_{Q\in\mathcal{Q}^N}[b(Q) - A(Q)x_\lambda]_+^{1/k} \geq 0$, we may define $Q_\lambda^* \in \mathcal{Q}^N$ to be such that for every $i \in \mathbb{I}$,

$$(Q_\lambda^*)_i = \begin{cases} \left(\arg\max_{Q\in\mathcal{Q}^N} [b(Q) - A(Q)\, x_\lambda]_+^{1/k}\right)_i, & \text{if } \left(A(\bar{Q})x_\lambda - b(\bar{Q})\right)_i > 0, \\ \bar{q}, & \text{if } \left(A(\bar{Q})x_\lambda - b(\bar{Q})\right)_i = 0, \end{cases} \tag{22}$$

which means, as seen in Theorem 4.2, that

$$\left|\left(A(Q_\lambda^*)\, x_\lambda - b(Q_\lambda^*)\right)_i\right| = \left|\min_{Q\in\mathcal{Q}^N} \left(A(Q)\, x_\lambda - b(Q)\right)_i\right| \leq \frac{C_1}{\lambda^k} \tag{23}$$

for some constant $C_1 > 0$ independent of λ. This implies, based on Theorem 4.1 and Remark 4.1, that

$$\left|\left(A(Q^*)\, x^* - b(Q^*)\right)_i\right| = 0, \tag{24}$$

where $\lim_{\lambda\to\infty} Q_\lambda^* = Q^* = \arg\min_{Q\in\mathcal{Q}^N} \{A(Q)\, x - b(Q)\}$.

It follows from (24) that

$$\left(A(Q_\lambda^*)\, x^* - b(Q_\lambda^*)\right)_i \geq \min_{Q\in\mathcal{Q}^N} \left(A(Q)\, x - b(Q)\right)_i = \left(A(Q^*)\, x^* - b(Q^*)\right)_i = 0.$$

Hence,

$$\left(A(Q_\lambda^*)(x_\lambda - x^*)\right)_i = \left(A(Q_\lambda^*)\, x_\lambda - b(Q_\lambda^*)\right)_i - \left(A(Q_\lambda^*)\, x^* - b(Q_\lambda^*)\right)_i \leq \left(A(Q_\lambda^*)\, x_\lambda - b(Q_\lambda^*)\right)_i.$$

Now, using (23), we get

$$(A(Q_\lambda^*)(x_\lambda - x^*))_i \le \frac{C_1}{\lambda^k}.$$

Meanwhile,

$$\begin{aligned}(A(Q^*)(x^* - x_\lambda))_i &= [(A(Q^*)x^* - b(Q^*)) - (A(Q_\lambda^*)x_\lambda - b(Q_\lambda^*))]_i \\ &\quad + [(A(Q_\lambda^*)x_\lambda - b(Q_\lambda^*)) - (A(Q^*)x_\lambda - b(Q^*))]_i \\ &\le [(A(Q^*)x^* - b(Q^*)) - (A(Q_\lambda^*)x_\lambda - b(Q_\lambda^*))]_i\,,\end{aligned}$$

since the definition Q_λ^* in (22) implies $[(A(Q_\lambda^*)x_\lambda - b(Q_\lambda^*)) - (A(Q^*)x_\lambda - b(Q^*))] \le 0$. Moreover, it follows from (23) and (24) that

$$[(A(Q^*)x^* - b(Q^*)) - (A(Q_\lambda^*)x_\lambda - b(Q_\lambda^*))]_i = -(A(Q_\lambda^*)x_\lambda - b(Q_\lambda^*))_i \le \frac{C_1}{\lambda^k}.$$

Hence,

$$(A(Q^*)(x^* - x_\lambda))_i \le \frac{C_1}{\lambda^k}.$$

Denoting by $A_1^*, A_2^* \in \mathcal{M}$ the matrices having the ith rows, $i \in \mathbb{I}$, as those of of $A(Q_\lambda^*)$ and $A(Q^*)$, respectively, we obtain that

$$x_\lambda - x^* \le \frac{C_1 \|(A_1^*)^{-1}\|_\infty}{\lambda^k}, \quad \text{and} \quad x^* - x_\lambda \le \frac{C_1 \|(A_2^*)^{-1}\|_\infty}{\lambda^k},$$

since it follows that both A_1^* and A_2^* are strictly diagonally dominant M-matrices. Now, following from [22, Corollary 2.6], we know that both A_1^* and A_2^* are bounded. Hence, we infer that

$$\|x^* - x_\lambda\|_\infty \le \frac{C}{\lambda^k},$$

for some constant $C > 0$ independent of λ, x_λ and x^*. □

5 Solution method

In this section, we will develop a solution method to solve the power penalized equation (4). When the linear $(l_{1/k}, k = 1)$ penalty method is applied, the generalized Newton method is well suited to solving the linear penalized equation (4) because of the semi-smoothness of the l_1 penalty function. This method has been well studied in [22], hence we omit its implementation details. But, when the lower order $(l_{1/k}, k > 1)$ penalty method is applied, the generalized Newton method cannot be used to solve the lower order penalized problem 3.1. This is because the lower order penalty function is not semi-smooth. To solve the lower penalized problem, we propose the following damped Newton method.

Algorithm 5.1 (Damped Newton Method)**.**

Step 1. *Choose* $\varepsilon > 0$ *sufficiently small,* $\bar{Q} \in \mathcal{Q}^N$ *and an initial guess* $x^0 \in \mathbb{R}^n$ *such that* $A(\bar{Q})x^0 - b(\bar{Q}) \geq 0$. *Let* $l := 0$.

Step 2. *Compute*

$$Q^l = \arg\max_{Q\in\mathcal{Q}^N} \left(\left[b(Q) - A(Q)x^l \right]_+^{1/k} \right),$$

and solve the following linear system for p^{l+1}*:*

$$\left[A(\bar{Q}) + \lambda D(Q^l)A(Q^l) \right] p^{l+1} = A(\bar{Q})x^l + \lambda \left[b(Q^l) - A(Q^l)x^l \right]_+^{1/k} + b(\bar{Q}), \qquad (25)$$

where $D(\cdot)$ *is defined by*

$$\begin{aligned} D(Q^l) &= \frac{1}{k}\mathrm{diag}\left(\left[b(Q^l) - A(Q^l)x \right]_+^{1/k-1} \right) \\ &= \frac{1}{k}\mathrm{diag}\left(([b(q_1^l) - A(q_1^l)x]_+)_1^{1/k-1}, ..., ([b(q_N^l) - A(q_N^l)x]_+)_N^{1/k-1} \right). \end{aligned} \qquad (26)$$

Step 3. *Set*

$$x^{l+1} = x^l + \nu p^{l+1},$$

where $0 < \nu < 1$ *is a damping parameter determined by the Armijo linear search method [6].*

Step 4. *If* $\max_{i\in\mathbb{I}} \frac{|x_i^{l+1} - x_i^l|}{\max(1, |x_i^{l+1}|)} < \varepsilon$, *then stop. Otherwise, set* $l := l+1$ *and go to Step 2.*

In view of numerical solution of (25), from the last part of the proof of Lemma 3.2 we see that the system matrix $A(\bar{Q}) + \lambda D(Q^l)A(Q^l)$ of (25) is an M-matrix, and thus efficient iterative methods, such as the preconditioned CGS or BiCGSTAB ([10]), can be used for solving (25). Note that for many practical problems, $A(Q)$ is an upper/lower triangular matrix as will be seen in the next section. Hence, the system matrix $A(\bar{Q}) + \lambda D(Q^l)A(Q^l)$ of (25) is also an upper/lower triangular matrix. In practice, we can use the LU decomposition to solve the linear system (25) very efficiently.

Remark 5.1. It is noted that for some $i \in \mathbb{I}$ and $y \in \mathbb{R}^N$, $([y]_+)_i^{1/k}$ may not exist at 0 when $k > 1$. In this case, we set it to be $\lim_{y\to 0^-} ([y]_+)_i^{1/k} = 0$. We notice that in the numerical experiments this treatment rarely affects the empirical performance of the algorithm.

6 Numerical experiment

In this section, we design two numerical examples to demonstrate the effectiveness, rates of convergence and efficiency of the power penalty approach to the discrete HJB equations. The first example is a Markovian dynamic programming problem taken from [18], which is used to show the advantages of the proposed penalty method over the classical policy iteration method. The second example, arising from the optimal investment problem under an incomplete market (see, [22]), is used to verify the rates of convergence and efficiency of the power penalty method.

6.1 A Markovian dynamic programming model

Consider the Markovian dynamic programming (MDP) problem in [18] which can be written as

$$V_i = \max\left\{V_{i-1} + f_i^1, V_{i+1} + f_i^2\right\}, \quad i = 0, \dots, M,$$

where $f_0^1 = f_0^2 = f_M^1 = f_M^2 = 0$, $f_i^1 = -1, f_i^2 = -2$ for all $i = 1, \dots, M-1$, and $f_{M-1}^1 = -1$, $f_{M-1}^2 = 2M$. We apply the l_1 penalty method to solve this problem, which results in

$$A^1 V_\lambda - \lambda[b^2 - A^2 V_\lambda]_+ = b^1,$$

with

$$A^1 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad A^2 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

are two $(M+1) \times (M+1)$ matrices and

$$b^1 = \begin{bmatrix} 0 & -2 & \cdots & -2 & 2M & 0 \end{bmatrix}^\top, \quad b^2 = \begin{bmatrix} 0 & -1 & \cdots & -1 & -1 & 0 \end{bmatrix}^\top$$

are two $M+1$ vectors.

In our numerical tests we increase M from 100 to 2000. It is worth noting that though both A^1 and A^2 are not strictly diagonally dominant M-matrices, the penalty method works very well. All the numerical results show that the number of iterations of the l_1 penalty method stays between 1 and 2 when the initial guess is set to be $V_0 = 0$. However, as stated in [11, 18], with the same initial guess ($V_0 = 0$), the number of iteration of the policy iteration is $M-1$ since it will correct the optimal control one by one, from grid $M-1$ to grid 1. This example shows that though the policy iteration is convergent for HJB equations, it can take up to the number of grid points to converge. On the contrary, the proposed penalty method can still be very efficient.

6.2 Investment model

In the test problem, a bond price, a stock price and a volatility are assumed to follow the processes

$$\mathrm{d}B_t = rB_t\mathrm{d}t, \quad \mathrm{d}S_t = \mu S_t\mathrm{d}t + \sigma(Y_t)\, S_t\mathrm{d}W_t^1, \quad \mathrm{d}Y_t = b(Y_t)\,\mathrm{d}t + a(Y_t)\,\mathrm{d}W_t^2,$$

respectively. Here, $\mathrm{d}W_t^1$ and $\mathrm{d}W_t^2$ are two Brownian motions with a correlation coefficient $-1 \le \rho \le 1$. Let $X_t = \pi_t^0 + \pi_t$ denote an investor's portfolio or wealth at $t \in [0, T]$, where π_t^0 and π_t are the amounts invested, respectively in the bond and in the stock. Then,

$$\mathrm{d}X_t = rX_t\mathrm{d}t + (\mu - r)\,\pi_t\mathrm{d}t + \sigma\,(Y_t)\,\pi_t\mathrm{d}W_t^1.$$

Suppose the investor's utility function to be of CRRA-type and given by

$$U\,(x) = \frac{1}{\gamma}x^\gamma, \quad x \in \mathbb{R}, \gamma \in (0, 1).$$

It is well-known that the investor's value function is given by

$$\phi\,(x, y, t) = \sup_{\pi \in \mathcal{A}} \mathbb{E}\left[U\,(X_T)\,|X_t = x, Y_t = Y\right], \quad (x, y, t) \in [0, \infty) \times \mathbb{R} \times [0, T]\,.$$

In [24], the value function is reformed as

$$\phi\,(x, y, t) = \frac{x^\gamma}{\gamma}\varphi\,(y, t)\,,$$

which is the viscosity solution of the following continuous HJB problem.

Problem 6.1 (Continuous HJB)**.**

$$\begin{aligned}\frac{1}{\gamma}&\left[\varphi_t + \frac{1}{2}a^2\,(y)\,\varphi_{yy} + b\,(y)\,\varphi_y\right] + r\varphi \\ &+ \max_{u \in \mathcal{Q}}\left[\frac{1}{2}\,(\gamma - 1)\,\sigma^2\,(y)\,u^2\varphi + \rho\sigma\,(y)\,a\,(y)\,u\varphi_y + (\mu - r)\,u\varphi\right] = 0, \quad (27)\end{aligned}$$

with an appropriately chosen compact set $\mathcal{Q} \subset \mathbb{R}$ and $\varphi\,(y, T) = 1$.

We set the parameters in Problem 6.1 to be $r = 0.3$, $\mu = 0.7$, $\rho = 0.2$, $\gamma = 0.5$, and $T = 1$. We also set $y_{\min} = 0.1$ and $y_{\max} = 1$ and for $y \in [y_{\min}, y_{\max}]$ we use

$$\begin{aligned}a\,(y) &= -2.5\,(y - 0.5 - 0.5y_{\min})^2 + 2.5\,(-0.5 + 0.5y_{\min})^2\,,\\ b\,(y) &= -y + 0.55,\\ \sigma\,(y) &= y.\end{aligned}$$

6.2.1 Discretization of the continuous HJB equation

To numerically solve the continuous Problem 6.1, we first localize $y \in \mathbb{R}$ in (27) to $y \in [y_{\min}, y_{\max}]$ and $\mathcal{Q} \subset \mathbb{R}$ to $\mathcal{Q} = [q_{\min}, q_{\max}]$, and then present the mesh grids of the space variable y, the time variable t and the control variable q as follows:

- Space grid: $y_i = y_{\min} + i\mathit{\Delta} y$, $0 \le i \le N$ with $\mathit{\Delta} y = (y_{\max} - y_{\min})/N$.

- Time grid: $t_j = T - j\Delta t$, $0 \le j \le N$, with $\Delta t = T/M$.
- Control grid: $\widetilde{\mathcal{Q}} = \{q_s | q_s = q_{\min} + s\Delta q, \quad s = 0, \cdots, (q_{\max} - q_{\min})/\Delta q\}$.

By using these mesh grids we perform a fully implicit finite difference discretization, using one-sided differences for all first derivatives (including the time derivative) and central differences for all second derivatives. To guarantee the M-matrix property, we use the upwind scheme to handle the first order spatial derivative terms. All other terms are approximated by their respective nodal values at (y_i, t_j). We approximate $A(Q)$ and $b(Q)$, $Q \in \mathcal{Q}^N$ by piecewise constant functions. This discretization method results in a sequence of discrete HJB problems of the form (3) in $\mathbb{R}^{N-1}$ with a discrete HJB equation at each time step j.

6.2.2 Numerical implementation

To compare the convergence properties of different power penalty methods, we choose the parameters in Step 1 of Algorithm 5.1 to be $\varepsilon = 10^{-8}$, and use a unified grid mesh with $M = 800$ and $N = 800$. At the same time, we choose $q_{\min} = -150$, $q_{\max} = 150$ and $\Delta q = 0.005$. Hence, the continuous control set $\mathcal{Q}$ is approximated by 60000 piecewise constant functions. All experiments were performed in double precision under Matlab environment.

We first investigate the convergence property of the l_1 penalty method. To compute the rates of convergence of the linear penalty method, we solve the discrete HJB equation using the semi-smooth Newton method with $\lambda = 10^i$ for $i = 4, 5, ..., 11$. The l_∞-norms of the errors in the last time point $t = 0$ between the numerical solutions with two consecutive λ values are calculated. Then, the ratios of errors from two consecutive values of λ are presented. All these results are listed in Table 1. From the table we see that the computed rates of convergence are close to $\mathcal{O}(\lambda^{\log_{10} 10}) = \mathcal{O}(\lambda)$, consistence with the theoretical result in (21). The average numbers of Newton iterations for all λ and all time steps range from 1.16 to 1.18, indicating the numerical method is very robust with respect to the penalty parameter λ. We also note that when $\lambda \ge 10^{11}$, the l_1 penalty method failed due to the ill-condition caused by too large penalty parameters.

We then investigate the convergence property of the lower order $l_{1/2}$ penalty method. As we did for the l_1 penalty method, the lower order penalized problem is first solved by the damped Newton method (Algorithm 5.1) with $\lambda = 10^i$ for $i = 2, 3, ..., 6$, then l_∞-norms and ratios of the errors are computed. It is known that the damping parameter plays a key role in the numerical convergence speed of the Newton method. To optimize the convergence speed, as stated in Algorithm, the damping parameter is determined via the classic Armijo linear search method. Our numerical tests show that in most

$\lambda = 10^i$	Error ($\|\cdot\|_\infty$)	Ratio ($\log_{10}$)	Total Iter.	Aver. Iter.
$i = 4$	0.28910484		944	1.18
$i = 5$	0.02924957	0.995	928	1.16
$i = 6$	0.00292809	0.999	933	1.17
$i = 7$	0.00029263	1.000	936	1.17
$i = 8$	0.00002901	1.003	937	1.17
$i = 9$	0.00000275	1.031	937	1.17
$i = 10$	0.00000026	1.033	937	1.17
$i = 11$	*Failed*	*Failed*	*Failed*	*Failed*

Table 1: Computed rates of convergence in λ and iteration numbers for the l_1 penalty method on the mesh grids 800×800. 'Total Iter.' stands for the total number of iterations at all time steps. 'Aver. Iter.' stands for the average number of iterations at each time step.

$\lambda = 10^i$	Error ($\|\cdot\|_\infty$)	Ratio ($\log_{10}$)	Total Iter.	Aver. Iter.
$i = 2$	2.986963996		4360	5.45
$i = 3$	0.036247934	1.92	3593	4.50
$i = 4$	0.000374588	1.99	2735	3.42
$i = 5$	0.000000475	1.90	1994	2.53
$i = 6$	0.000000005	1.98	1960	2.45

Table 2: Computed rates of convergence in λ and iteration numbers for the $l_{1/2}$ penalty method on the mesh grids 800×800. 'Total Iter.' stands for the total number of iterations at all time steps. 'Aver. Iter.' stands for the average number of iterations at each time step. The damping parameter is determined via Armijo linear search method.

discretization scheme, the optimal damping parameter is found to be about 0.5. We then list all the results in Table 2. From the table we see that the computed rates of convergence are close to $\mathcal{O}(\lambda^{\log_{10} 100}) = \mathcal{O}(\lambda^2)$, again consistent with the theoretical result in (21). Furthermore, the average numbers of Newton iterations for all λ and all time steps now range from 2.45 to 5.45, indicating the numerical method is also robust with respect to the penalty parameter λ. Moreover, comparing Tables 1 with 2, we can see that to achieve the same level of accuracy, the $l_{1/2}$ penalty method requires much less penalty parameter than l_1 penalty method needs. It is also clear that we can achieve a much more accurate result by the $l_{1/2}$ penalty method. This verifies the advantage of lower order penalty method.

Finally, to further demonstrate the computational efficiency, we carry out some numerical comparisons between l_1 and $l_{1/2}$ penalty methods and the popular policy iteration method. With the same level of accuracy ($\varepsilon = 10^{-3}$), we list all the the average numbers of iterations for all time steps and computation times on different mesh grids in Table 3. The table clearly shows that the computational efficiency of the l_1 and $l_{1/2}$ penalty methods are comparable to that of the policy iteration method. Since the policy iteration

$M \times N$	Policy iteration method		l_1 penalty method		$l_{1/2}$ penalty method	
	Aver. Iter.	CPU	Aver. Iter.	CPU	Aver. Iter.	CPU
25×25	2.00	0.142	2.00	0.145	3.61	0.154
50×50	1.96	0.370	1.96	0.406	3.13	0.417
100×100	1.75	3.097	1.75	3.554	2.81	3.713
200×200	1.89	11.330	1.88	12.895	2.62	13.062
400×400	1.70	24.610	1.70	28.086	2.44	30.165
800×800	1.17	74.538	1.18	81.017	2.13	84.029

Table 3: Comparison of the l_1, $l_{1/2}$ penalty methods and the policy iteration method. $\lambda = 10^{10}$ is used for the l_1 penalty method. $\lambda = 10^6$ is used for the $l_{1/2}$ penalty method. 'Aver. Iter.' stands for the average number of iterations at each time step. 'CPU' means the computation time. The damping parameter is determined via Armijo linear search method.

method is commonly regarded as the standard method to solving discrete HJB equations, we can conclude that the power penalty method is an efficient and effective method.

7 Conclusions

Motivated by the power penalty approach to complementarity problems, in this work we have proposed a power penalty approach to the discrete HJB equations arising from finance. The approach is to approximate the discrete HJB by a nonlinear algebraic equation containing a penalty term. We have shown that under some mild conditions the penalty equation is uniquely solvable. An exponential rate of convergence with respect to the penalty parameters, for the solution to the penalty equation, has been established. A damped Newton method has been proposed for solving the lower order penalized equation. Numerical experiments using an optimal investment problem under an incomplete market have been carried out. Numerical results have demonstrated that the computed rates of convergence are consistent with theoretical one and that the method is efficient and effective for solving practical problems. Moreover, a numerical test has been carried out to verify the advantage of the penalty method over the policy iteration method in some circumstances.

Acknowledgement

The authors would like to thank the anonymous referees and the editor for their helpful comments and suggestions toward the improvement of this paper. Kai Zhang's research was supported by National Natural Science Foundation of China (Project 11871347). Yang's research was supported by the Research Grant Council of Hong Kong (PolyU

152342/16E).

References

[1] AVRAMESCU C. (2002). A Generalization of Miranda's Theorem, *Seminar on Fixed Point Theory Cluj-Napoca*, 3, 121-128.

[2] Barles, G. (1997). Convergence of Numerical Schemes for Degenerate Parabolic Equations Arising in Finance Theory. *Numerical Methods in Finance*, Cambridge Univ. Press, Cambridge.

[3] Bokanowski, O., Maroso, S., & Zidani, H. (2009). Some convergence results for Howard's algorithm. *SIAM Journal on Numerical Analysis*, 47(4), 3001-3026.

[4] Crandall, M. G., Ishii, H., & Lions, P. (1992). User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Mathematical Society*, 27(1), 1-67.

[5] Dai, M., & Zhong, Y. (2008). Penalty methods for continuous-time portfolio selection with proportional transaction costs. *Journal of Computational Finance*, 13(3), 1-31.

[6] Dennis, J.E., & Schnabel, B. (1983). Numerical methods for unconstrained optimization and nonlinear equations. *Prentice Hall series in computational mathematics.*

[7] Fleming W. H. & Soner, H. M. (1993). *Controlled Markov Processes and Viscosity Solutions.* Springer-Verlag, New York.

[8] Forsyth, P. A., & Labahn, G. (2007). Numerical methods for controlled Hamilton-Jacobi-Bellman PDEs in finance. *Journal of Computational Finance*, 11(2), 1-43.

[9] Frey, R., & Polte, U. (2011). Nonlinear Black-Scholes equations in finance: associated control problems and properties of solutions. *SIAM Journal on Control and Optimization*, 49(1), 185-204.

[10] Faccinei, F., & Pang, J. S. (2003). *Finite-Dimensional Variational Inequalities and Complementarity Problems*, New York, Springer.

[11] Han, D., & Wan, J. W. (2013). Multigrid Methods for Second Order Hamilton–Jacobi–Bellman and Hamilton–Jacobi–Bellman–Isaacs Equations. *SIAM Journal on Scientific Computing*, 35(5).

[12] Ortega J. M. & Rheinboldt W. C. (1970). *Iterative Solution of Nonlinear Equations in Several Variables*. New York/London. Academic Press.

[13] Pham, H. (2009). Continuous-Time Stochastic Control and Optimization with Financial Applications, *Stochastic Modelling and Applied Probability, 61*, Springer-Verlag, Berlin.

[14] Ralph, D., & Scholtes, S. (1997). Sensitivity analysis of composite piecewise smooth equations. *Mathematical Programming*, 76(3), 593-612.

[15] Rubinov, A., & Yang, X. Q. (2003). *Lagrange-type Functions in Constrained Non-convex Optimization*, Kluwer.

[16] Shreve, S. E. (2004). *Stochastic Calculus for Finance II: Continuous-Time Models*. New York, Springer-Verlag.

[17] Wang, S., Yang, X. Q., & Teo, K. L. (2006). Power penalty method for a linear complementarity problem arising from American option valuation. *Journal of Optimization Theory and Applications*, 129, 227-254.

[18] Santos, M. S., & Rust, J. (2003). Convergence Properties of Policy Iteration. *Siam Journal on Control and Optimization*, 42(6), 2094-2115.

[19] Sun, Z., Liu, Z. M., & Yang, X. Q. (2015). On power penalty methods for linear complementarity problems arising from American option pricing. *Journal of Global Optimization*, 63(1), 165-180.

[20] Varga, R. S. (1962). *Matrix Iterative Analysis*. Prentice-Hall, Engelwood Cliffs, NJ.

[21] Wahib, M. I. M., Lin, Z., & Edirisingh, C. P. (2010). Pricing swing options in the electricity market under regime switching uncertainty. *Quantitative Finance*, 10, 975–994.

[22] Witte, J. H., & Reisinger, C. (2011). A penalty method for the numerical solution of Hamilton-Jacobi-Bellman (HJB) equations in finance. *SIAM Journal on Numerical Analysis*, 49(1), 213-231.

[23] Yiu, K. F., Liu, J., Siu, T. K., & Ching, W. (2010). Optimal portfolios with regime switching and value-at-risk constraint. *Automatica*, 46(6), 979-989.

[24] Zariphopoulou T. (2001), A solution approach to valuation with unhedgeable risks. *Finance and Stochastics*, 5(1), 61-82.

[25] Zhang, K., & Wang, S. (2009). A computational scheme for uncertain volatility model in option pricing. *Applied Numerical Mathematics*, 59(8), 1754-1767.

[26] Zhang, K., Yang, X. Q., & Teo, K. L. (2010). A power penalty approach to American option pricing with jump diffusion processes. *Journal of Industrial and Management Optimization*, 4(4) : 783-799.

[27] Zhang, K., Yang, X. Q., Wang, S., & Teo, K. L. (2010). Numerical performance of penalty method for American option pricing. *Optimization Methods & Software*, 25(5), 737-752.