Numerische Mathematik (2021) 148:187–222
https://doi.org/10.1007/s00211-021-01202-x

Numerische Mathematik

Stability and convergence of second order backward differentiation schemes for parabolic Hamilton–Jacobi–Bellman equations

Olivier Bokanowski[1,2] · **Athena Picarelli**[3] · **Christoph Reisinger**[4]

Received: 20 May 2020 / Revised: 13 March 2021 / Accepted: 16 April 2021 /
Published online: 20 May 2021

Abstract
We study a second order Backward Differentiation Formula (BDF) scheme for the numerical approximation of linear parabolic equations and nonlinear Hamilton–Jacobi–Bellman (HJB) equations. The lack of monotonicity of the BDF scheme prevents the use of well-known convergence results for solutions in the viscosity sense. We first consider one-dimensional uniformly parabolic equations and prove stability with respect to perturbations, in the L^2 norm for linear and semi-linear equations, and in the H^1 norm for fully nonlinear equations of HJB and Isaacs type. These results are then extended to two-dimensional semi-linear equations and linear equations with possible degeneracy. From these stability results we deduce error estimates in L^2 norm for classical solutions to uniformly parabolic semi-linear HJB equations, with an order that depends on their Hölder regularity, while full second order is recovered in the smooth case. Numerical tests for the Eikonal equation and a controlled diffusion equation illustrate the practical accuracy of the scheme in different norms.

Mathematics Subject Classification 65M12 · 65M06 · 49L12 · 35K45

✉ Athena Picarelli
athena.picarelli@univr.it

Olivier Bokanowski
olivier.bokanowski@math.univ-paris-diderot.fr

Christoph Reisinger
christoph.reisinger@maths.ox.ac.uk

1 Université de Paris, Laboratoire Jacques-Louis Lions (LJLL), F-75013 Paris, France

2 Sorbonne Université, CNRS, LJLL, F-75005 Paris, France

3 Dipartimento di Scienze Economiche, Università di Verona, Via Cantarane 24, 37129 Verona, Italy

4 Mathematical Institute, University of Oxford, Andrew Wiles Building, Woodstock Rd, Oxford OX2 6GG, UK

1 Introduction

This paper provides stability and convergence results for a type of implicit finite difference scheme for the approximation of nonlinear parabolic equations using backward differentiation formulae (BDF).

In particular, we consider Hamilton–Jacobi–Bellman (HJB) equations of the following form:

$$v_t(t,x) + \sup_{a\in\Lambda}\left\{\mathcal{L}^a[v](t,x) + r(t,x,a)v + \ell(t,x,a)\right\} = 0, \tag{1}$$

where $(t,x) \in [0,T]\times\mathbb{R}^d$, $\Lambda \subset \mathbb{R}^m$ is a compact set and

$$\mathcal{L}^a[v](t,x) = -\frac{1}{2}\,\mathrm{tr}[\Sigma(t,x,a)D_x^2 v(t,x)] + b(t,x,a)D_x v(t,x)$$

is a second order differential operator. Here, $(\Sigma)_{ij}$ is symmetric non-negative definite for all arguments.

It is well known that for nonlinear, possibly degenerate equations the appropriate notion of solutions to be considered is that of viscosity solutions [9]. We assume throughout the whole paper the well-posedness of the problem, namely the existence and uniqueness of a solution in the viscosity sense. Under such weak assumptions, convergence of numerical schemes can only be guaranteed if they satisfy certain monotonicity properties, in addition to the more standard consistency and stability conditions for linear equations [2]. This in turn reduces the obtainable consistency order to 1 in the general case [12].

We will therefore not treat (1) in this generality. As we detail further below, the main stability analysis is restricted to the uniformly parabolic case, and full convergence results are given under the additional assumption of semi-linearity, $\Sigma \equiv \Sigma(t,x)$.

It is shown in [16] that a monotone (but inconsistent) P_1–finite element approximation converges in the maximum norm, and in the H^1-norm under a mild non-degeneracy assumption; this assumption is further weakened to possibly degenerate coefficients in [15].

On the other hand, in many cases – especially in non-degenerate ones – solutions exhibit higher regularity and are amenable to higher order approximations. The existence of classical solutions and their regularity properties under a strict ellipticity condition have been investigated, for instance, in [11,17].

The higher order of convergence in both space and time of discontinuous Galerkin approximations is demonstrated theoretically and numerically in [20] for sufficiently regular solutions under a Cordes condition for the diffusion matrix, a measure of the ellipticity.

More recently, it was shown numerically in [6] that some approximation schemes based on a second order backward differentiation formula in both time and space (see, e.g., [22], Section 12.11, for the definition of BDF schemes for ODEs) have good convergence properties. In particular, in an example therein with non-degenerate con-

trolled diffusion where the second order, non-monotone Crank–Nicolson scheme fails to converge, the (also non-monotone) BDF2 scheme shows second order convergence.

The filtered schemes in [6] and ϵ-monotone schemes, e.g. in [7], modify a higher-order scheme to stay ϵ-close to a monotone scheme. This enforces convergence, but in general only at the rate of the monotone scheme, and practically the rate may vary depending on the data and on the strategy to choose the ϵ parameter (see Example 2 in [6, Section 4.2] where a filtered scheme switches back to first order). Here, we directly analyse the stability and the convergence for a non-monotone BDF scheme.

For constant coefficient parabolic PDEs, the L^2-stability and smoothing properties of the BDF scheme are a direct consequence of the strong A-stability of the scheme. Moreover, [3] shows that for the multi-dimensional heat equation the BDF time stepping solution and its first numerical derivative are stable in the maximum norm. The technique, which is strongly based on estimates for the resolvent of the discrete Laplacian, do not easily extend to variable coefficients or the nonlinear case.

A more general linear parabolic setting is considered in [4], where second order convergence is shown for variable demister using energy techniques. This result is extended to a semi-linear example in [10]; the application to incompressible Navier–Stokes equations has been analyzed in [14]. In [5], a closely related BDF scheme is studied for a diffusion problem with an obstacle term (which includes the American option problem in mathematical finance).

The scheme we propose is constructed by using a second order BDF approximation for the first derivatives in both time and space, combined with a standard three-point central finite difference for the second spatial derivative in one dimension. The scheme is therefore second order consistent by construction.

For this scheme, under the assumption of uniform parabolicity, we establish new stability results in the H^1-norm for fully nonlinear HJB and Isaacs equations, and in the L^2-norm for the semilinear case (see Theorems 4 and 5, respectively). These generalize some results of [4,5,10] to more general non-linear situations. From this analysis we deduce error bounds for classical smooth and piecewise smooth solutions in the semilinear uniformly parabolic case (see Theorems 7 and 19).

Our overall approach relies on stability results with respect to perturbations of the right-hand side of the equations. We start by deriving a recursive linear relation satisfied by the approximation error between the original equation and a perturbed one, in the case of HJB and Isaacs equations (Lemma 10); then, we give an inequality between the error norms for three consecutive time steps (Lemma 11) which guarantees an overall stability estimate (Lemma 12). Having proved this generic sufficient condition for stability, we show that this condition is satisfied for different choices of the norm under specific assumptions, which are summarized in Table 1.

The outline of the paper is as follows. In Sect. 2, we define some specific BDF schemes and state the main results concerning well-posedness and stability in discrete H^1- or L^2-norms and our main convergence result for uniformly parabolic semilinear HJB equations. In Sects. 3 and 4 we prove the main stability results and give an extension from HJB to Isaacs equations. In Sect. 5, we give further stability results in the discrete L^2-norm, which are weaker in the sense that they hold only for uncontrolled Lipschitz regular diffusion coefficients, but stronger in the sense that they allow for degenerate diffusion in the linear case and can be extended to two dimensions. In

Table 1 Main stability results

Norm	Dim	Diffusion	Drift	Degenerate	
H^1	1	Controlled	Controlled	No	(Theorem 4)
L^2	1, 2	Lipschitz	Controlled	No	(Theorem 5, Proposition 16)
L^2	1	Semi-convex	Lipschitz	Yes	(Proposition 15)

Sect. 6, we deduce error estimates from the L^2 stability results and from the truncation error of the scheme for sufficiently regular solutions. Section 7 studies carefully two numerical examples, the Eikonal equation and a second order equation with controlled diffusion. Section 8 concludes. An appendix contains a proof of the existence of solutions for our schemes.

2 Definition of the scheme and main results

We focus in the first instance on the one-dimensional equation

$$v_t + \sup_{a\in\Lambda}\Big(-\frac{1}{2}\sigma^2(t,x,a)v_{xx} + b(t,x,a)v_x + r(t,x,a)v + \ell(t,x,a)\Big) = 0, \quad t\in[0,T],\ x\in\mathbb{R}, \tag{2a}$$

$$v(0,x) = v_0(x) \quad x\in\mathbb{R}. \tag{2b}$$

It is known (see Theorem A.1 in [1]) that with the following assumptions:

- Λ is a compact set,
- for some $C_0 > 0$ the functions $\phi \equiv \sigma, b, r, \ell : [0,T]\times\mathbb{R}\times\Lambda \to \mathbb{R}$ and $v_0 : \mathbb{R}\to\mathbb{R}$ satisfy for any $t, s \in [0,T]$, $x, y\in\mathbb{R}$, $a\in\Lambda$

$$|v_0(x)| + |\phi(t,x,a)| \le C_0,$$
$$|v_0(x) - v_0(y)| + |\phi(t,x,a) - \phi(s,y,a)| \le C_0(|x-y| + |t-s|^{1/2}),$$

there exists a unique bounded continuous viscosity solution of (2). We denote by v this solution.

We will make individual assumptions for each result as we go along.

2.1 The BDF2 scheme

For the approximation in the x variable, we will consider the PDE on a truncated domain $\Omega := (x_{\min}, x_{\max})$, where $x_{\min} < x_{\max}$.

Let $N\in\mathbb{N}^* \equiv \mathbb{N}\backslash\{0\}$ be the number of time steps, $\tau := T/N$ the time step size, and $t_n = n\tau$, $n = 1,\dots,N$. Let $I\in\mathbb{N}^*$ the number of interior mesh points in the

spatial direction, and define a uniform mesh $(x_i)_{1\le i\le I}$ with mesh size h by

$$x_i := x_{\min} + ih, \quad i \in \mathbb{I} = \{1, \dots, I\}, \quad \text{where} \quad h := \frac{x_{\max} - x_{\min}}{I+1}.$$

Hereafter, we denote by u a numerical approximation of v, the solution of (1), i.e.

$$u_i^k \sim v(t_k, x_i).$$

For each time step t_k, the unknowns are the values u_i^k for $i = 1, \dots, I$.

Standard Dirichlet boundary conditions use the knowledge of the values at the boundary, $v(t, x_{\min})$ and $v(t, x_{\max})$. Here, as a consequence of the size of the stencil for the spatial BDF2 scheme below, we will assume that values at the two left- and right-most mesh points are given, that is, $v(t, x_j)$ for $j \in \{-1, 0\}$ as well as $j \in \{I+1, I+2\}$ are known (corresponding to the values at the points $(x_{-1}, x_0, x_{I+1}, x_{I+2}) \equiv (x_{\min} - h, x_{\min}, x_{\max}, x_{\max} + h)$).[1]

We then consider the following scheme, for $k \ge 2$, $i \in \mathbb{I}$,

$$\begin{aligned} \mathcal{S}^{(\tau,h)}(t_k, x_i, u_i^k, [u]_i^k) &:= \frac{3u_i^k - 4u_i^{k-1} + u_i^{k-2}}{2\tau} \\ &+ \sup_{a\in\Lambda} \left\{ L^a[u^k](t_k, x_i) + r(t_k, x_i, a)u_i^k + \ell(t_k, x_i, a) \right\} = 0, \end{aligned} \tag{3}$$

where we denote as usual by $[u]_i^k$ the numerical solution excluding at (t_k, x_i), and

$$L^a[u](t_k, x_i) := -\frac{1}{2}\sigma^2(t_k, x_i, a)D^2u_i + b^+(t_k, x_i, a)D^{1,-}u_i - b^-(t_k, x_i, a)D^{1,+}u_i,$$

$$D^2u_i := \frac{u_{i-1} - 2u_i + u_{i+1}}{h^2} \tag{4}$$

(the usual second order approximation of v_{xx}), $b^+ := \max(b, 0)$ and $b^- := \max(-b, 0)$ denote the positive and negative part of b, respectively, and where a second order left- or right-sided BDF approximation is used for the first derivative in space:

$$D^{1,-}u_i := \frac{3u_i - 4u_{i-1} + u_{i-2}}{2h} \quad \text{and} \quad D^{1,+}u_i := -\left(\frac{3u_i - 4u_{i+1} + u_{i+2}}{2h}\right). \tag{5}$$

Note in particular the implicit form of the scheme (3) for the forward Eq. (2). The existence of a unique solution to this nonlinear equation will be addressed later.

[1] In practice, this means that a sufficiently accurate approximation of these "boundary values" has to be available. Boundary approximations with modified schemes are commonly used and are not the focus of this paper; it is seen in [18] that the use of a lower order scheme in the vicinity of the boundary does not affect the global provable convergence order.

We will also define the numerical Hamiltonian associated with the scheme:

$$H[u](t_k, x_i) := \sup_{a\in\Lambda} \Big\{ L^a[u](t_k, x_i) + r(t_k, x_i, a)u_i + \ell(t_k, x_i, a) \Big\}.$$

As discussed above, the scheme is completed by the following boundary conditions:

$$u_i^k := v(t_k, x_i), \quad \text{for } i \in \{-1, 0\} \cup \{I+1, I+2\} \text{ and } 2 \le k \le N.$$

Since (3) is a two-step scheme, for the first time step $k = 1$ (and $i \in \mathbb{I}$), we use a backward Euler approximation scheme :

$$\begin{aligned} &\mathcal{S}^{(\tau,h)}(t_1, x_i, u_i^1, [u]_i^1) \\ &:= \frac{u_i^1 - u_i^0}{\tau} + \sup_{a\in\Lambda} \Big\{ L^a[u^1](t_1, x_i) + r(t_1, x_i, a)u_i^1 + \ell(t_1, x_i, a) \Big\} = 0, \end{aligned} \tag{6}$$

with the initial condition

$$u_i^0 = v_0(x_i), \quad i \in \mathbb{I}. \tag{7}$$

Remark 1 As the backward Euler step is only used once, it does not affect the overall second order of the scheme (see also Sect. 6 below).

Remark 2 Most of our results also apply to the scheme obtained by replacing the BDF approximation (5) of the drift term by a centered finite difference approximation:

$$\widetilde{D}^{1,\pm} u_i := \frac{u_{i+1} - u_{i-1}}{2h}. \tag{8}$$

However, numerical tests (see Sect. 7.1) show that the BDF upwind approximation as in (5) has a better behavior in some extreme cases where the diffusion vanishes. We shall give a rigorous stability estimate for the BDF scheme in the linear case for possibly vanishing diffusion in Sect. 5.2.

2.2 Definitions and main results

In the remainder of this paper, we prove various stability and convergence results for the scheme (3). We state in this section the first main well-posedness and stability results.

Throughout the paper, A will denote the finite difference matrix associated with the second order derivative, i.e.

$$A := \frac{1}{h^2}\begin{pmatrix} 2 & -1 & 0 & & \\ -1 & 2 & -1 & \ddots & \ddots \\ 0 & -1 & \ddots & \ddots & 0 \\ \ddots & \ddots & \ddots & \ddots & -1 \\ & & 0 & -1 & 2 \end{pmatrix}. \tag{9}$$

Let $\langle x, y\rangle_A := \langle x, Ay\rangle$. Then we consider the A-norm defined as follows:

$$|x|_A^2 := \langle x, Ax\rangle = \sum_{1\le i\le I+1} \left(\frac{x_i - x_{i-1}}{h}\right)^2 \tag{10}$$

(with the convention in (10) that $x_0 = x_{I+1} = 0$). Hence, $\sqrt{h}|x|_A$ approximates the H^1 semi-norm in Ω. Similarly, we will consider later the standard Euclidean norm defined by $\|x\|^2 := \langle x, x\rangle$, such that $\sqrt{h}\|x\|$ approximates the L^2-norm. We define therefore the following rescaled norms on $\mathbb{R}^I$:

$$|u|_0 := \left(\sum_{i\in\mathbb{I}} u_i^2\, h\right)^{1/2} = \|u\|\sqrt{h}, \qquad |u|_1 := \left(\sum_{i\in\mathbb{I}} \left(\frac{u_i - u_{i-1}}{h}\right)^2 h\right)^{1/2} = |u|_A\sqrt{h}.$$

Both these norms will be used in the numerical section.

Our first result concerns the solvability of the numerical scheme $\mathcal{S}^{(\tau,h)}$ defined by (3) with respect to its third argument, i.e. seen as an equation for u_i^k, with (t_n, x_i) and $[u]_i^k$ given.

ASSUMPTION (A1). The functions σ, b and r are bounded.

Theorem 3 *Let (A1) and the following CFL condition hold:*

$$\|b\|_\infty \frac{\tau}{h} < C. \tag{11}$$

Then, for τ small enough and $C = 3/2$ (resp. $C = 1$) there exists a unique solution of the scheme (3) for $k \ge 2$ (resp. $k = 1$, for scheme (6)).

The scheme is hence well-defined even if σ vanishes. A uniform ellipticity condition for σ as per Assumption (A2) below will be needed for proving the H^1 stability of the scheme. We provide a relaxation of the ellipticity condition for stability in the Euclidean norm in Sect. 5.2.

ASSUMPTION (A2). There exists $\eta > 0$ such that

$$\inf_{t\in[0,T]} \inf_{x\in\Omega} \inf_{a\in\Lambda} \sigma^2(t, x, a) \ \ge\ \eta.$$

Let u denote the solution of (3), that is,

$$\mathcal{S}^{(\tau,h)}(t_k, x_i, u_i^k, [u]_i^k) = \mathcal{E}_i^k(u), \quad i \in \mathbb{I},\ 1 \le k \le N,$$

with $\mathcal{E}_i^k(u) \equiv 0$, and let w be the solution of a perturbed equation

$$\mathcal{S}^{(\tau,h)}(t_k, x_i, w_i^k, [w]_i^k) = \mathcal{E}_i^k(w), \quad i \in \mathbb{I},\ 1 \le k \le N, \tag{12}$$

with the same boundary values as u:

$$w_i^k = u_i^k, \quad i \in \{-1, 0\} \cup \{I+1, I+2\},\ 2 \le k \le N \tag{13}$$

(but potentially different initial values w^0 and w^1).

Denote further

$$E^k := (E_1^k, \dots, E_I^k)^T = u^k - w^k, \quad 0 \le k \le N.$$

Our main stability result in this setting (which also holds when $\mathcal{E}^k(u) \neq 0$) is the following.

Theorem 4 *Assume (A1), (A2), as well as the CFL condition (11). Then there exists a constant $C \ge 0$ (independent of τ and h) and $\tau_0 > 0$ such that, for any $\tau \le \tau_0$, for any $u = (u_i^k)$ and $w = (w_i^k)$ with same boundary values (13), it holds:*

$$\max_{2\le k\le N} |E^k|_A^2 \le C\Big(|E^0|_A^2 + |E^1|_A^2 + \tau \sum_{2\le k\le N} |\mathcal{E}^k(u) - \mathcal{E}^k(w)|_A^2\Big). \tag{14}$$

The proof of Theorem 4 will be the subject of Sect. 4.

As a corollary we can deduce the $|\cdot|_1$-seminorm boundedness of the scheme. For instance, let us assume that $\ell \equiv 0$, and let u be a solution of the scheme (3) (that is, $\mathcal{E}_i^k(u) \equiv 0$), with 0 boundary conditions ($u_i^k = 0$ for all $k \ge 2$ and $i \in \{-1, 0, I+1, I+2\}$). Then by taking $w = 0$ in (14), we obtain

$$\max_{2\le k\le N} |u^k|_1^2 \le C\Big(|u^0|_1^2 + |u^1|_1^2\Big). \tag{15}$$

A more general bound of $|u^k|_1$ could also be obtained in the case of non-zero boundary values and non-vanishing ℓ, the bound then depending on these data.

In order to obtain stability estimates in other norms, one typically needs some uniform continuity of the coefficients. The analysis of the controlled case, associated with the presence of the supremum operator in (2a), is then made complicated by the fact that even if the solution to (2) is classical and the supremum is attained at some $a^*(t, x)$ for each t and x [and similarly for each k and i in (3)], the optimal control a^* in general does not have any regularity as a function of t and x (or k and i, respectively).

However, in certain circumstances, the previous bound holds with the A-norm replaced by the Euclidean norm. In particular, we consider the following assumption:

ASSUMPTION (A3). The diffusion coefficient is independent of the control and Lipschitz continuous, i.e. $\sigma \equiv \sigma(t,x)$ and there exists $L \geq 0$ such that

$$|\sigma^2(t,x) - \sigma^2(t,y)| \leq L|x-y| \quad \forall x, y \in \Omega, t \in [0,T].$$

Theorem 5 *Assume (A1), (A2), (A3), as well as the CFL condition (11). Then there exists $C \geq 0$ (independent of τ and h) and $\tau_0 > 0$ such that, for any $\tau \leq \tau_0$, for any $u = (u_i^k)$ and $w = (w_i^k)$ with the same boundary values (13), it holds:*

$$\max_{2\leq k\leq N} \|E^k\|^2 \leq C\Big(\|E^0\|^2 + \|E^1\|^2 + \tau \sum_{2\leq k\leq N} \|\mathcal{E}^k(u) - \mathcal{E}^k(w)\|^2\Big). \tag{16}$$

As a consequence, we obtain error estimates under the main assumptions (A1), (A2) and (A3), or under some specific regularity assumptions.

We define the following semi-norm on some interval $\mathcal{I} = (a,b)$, for $\alpha \in (0,1]$:

$$||\phi||_{C^{0,\alpha}(\mathcal{I})} := \sup\left\{\frac{|\phi(x)-\phi(y)|}{|x-y|^\alpha},\ x \neq y,\ x, y \in \mathcal{I}\right\}.$$

For a given open subset Ω_T^* of $(0,T)\times\Omega$, we define $C^{k,\ell}(\Omega_T^*)$ as the set of functions $\phi : \Omega_T^* \to \mathbb{R}$ which admit continuous derivatives $(\frac{\partial^i \phi}{\partial t^i})_{0\leq i\leq k}$ and $(\frac{\partial^j \phi}{\partial x^j})_{0\leq j\leq \ell}$ on Ω_T^*. We also denote by $C_b^{k,\ell}(\Omega_T^*)$ the subset of functions with bounded derivatives on Ω_T^*.

ASSUMPTION (A4). $v \in C^{1,2}((0,T)\times\Omega)$ and for some $C \geq 0, \delta \in (0,1]$, it holds:

$$\sup_{x\in\Omega} \|v_t(.,x)\|_{C^{0,\delta}([0,T])} \leq C, \qquad \sup_{t\in(0,T)} \|v_{xx}(t,.)\|_{C^{0,\delta}(\bar{\Omega})} \leq C. \tag{17}$$

Remark 6 By results in [11] and [17], assumption (A4) is satisfied for sufficiently smooth data and given a uniform ellipticity condition.

We have the following error estimates:

Theorem 7 *We assume (A1), (A2), (A3), and the CFL condition (11).*

(i) *If (A4) holds for some $\delta \in (0,1]$, then the numerical solution u of (3), (6) converges to v in the L^2-norm with*

$$\max_{0\leq k\leq N} |v^k - u^k|_0 \leq Ch^\delta,$$

for some constant C (possibly different from the one in (A4)).

(ii) *If, moreover, $v \in C_b^{3,4}((0,T)\times\Omega)$, then*

$$\max_{0\leq k\leq N} |v^k - u^k|_0 \leq Ch^2,$$

where C is a constant which depends on the derivatives of v of order 3 and 4 in t and x, respectively.

The proof of these and further error estimates will be the subject of Sect. 6.

The extension of the presented results to other types of nonlinear operators (inf, sup inf or inf sup) and corresponding equations will also be discussed.

Hereafter, for simplicity, we will consider $\mathcal{E}^k(u) \equiv 0$ and will denote $\mathcal{E}_i^k := \mathcal{E}_i^k(w)$.

3 Proof of Theorem 3 (well-posedness of the scheme)

The scheme (3) at time t_k (for $k \geq 2$) can be written in the following form:

$$\sup_{a\in\Lambda}(M_a^k X - q_a^k) = 0,$$

where $q_a^k \in \mathbb{R}^I$ and $M_a^k \in \mathbb{R}^{I\times I}$ with the following non-zero entries:

$$(M_a^k)_{i,i} := \frac{3}{2} + \tau\left\{2\frac{\sigma^2}{h^2} + \frac{3b^+}{2h} + \frac{3b^-}{2h} + r\right\} \tag{18}$$

$$(M_a^k)_{i,i+1} := \tau\left\{-\frac{\sigma^2}{h^2} - \frac{4b^-}{2h}\right\}, \quad (M_a^k)_{i,i-1} := \tau\left\{-\frac{\sigma^2}{h^2} - \frac{4b^+}{2h}\right\} \tag{19}$$

$$(M_a^k)_{i,i+2} := \tau\frac{b^-}{2h} \quad (M_a^k)_{i,i-2} := \tau\frac{b^+}{2h} \tag{20}$$

with $\sigma \equiv \sigma(t_k, x_i, a)$, $b^\pm \equiv b^\pm(t_k, x_i, a)$ and $r \equiv r(t_k, x_i, a)$. For $k = 1$, the terms are different but the form (and analysis) is similar. The fact that $(M_a)_{i,i\pm2}$ are nonnegative breaks the monotonicity of the scheme and makes the analysis more difficult compared to the non-degenerate setting and central differences, where M is a diagonally dominant M-matrix for h small enough.

We will use the following lemma, whose proof is given in Appendix A:

Lemma 8 *Assume that Λ is some set, $(q_a)_{a\in\Lambda}$ is a family of vectors in $\mathbb{R}^I$, $(M_a)_{a\in\Lambda}$ is a family of matrices in $\mathbb{R}^{I\times I}$ such that:*

(i) *for all $a \in \Lambda$,*

$$(M_a)_{ii} > 0;$$

(ii) *(a form of diagonal dominance)*

$$\sup_{a\in\Lambda}\max_{i\in\mathbb{I}} \frac{\sum_{j>i}|(M_a)_{ij}|}{|(M_a)_{ii}| - \sum_{j<i}|(M_a)_{ij}|} < 1. \tag{21}$$

Then there exists a unique solution X in $\mathbb{R}^n$ of

$$\sup_{a\in\Lambda}(M_a X - q_a) = 0. \tag{22}$$

Remark 9 For a fixed $a \in \Lambda$, we have

$$\max_{i\in\mathbb{I}} \frac{\sum_{j>i} |(M_a)_{ij}|}{|(M_a)_{ii}| - \sum_{j<i} |(M_a)_{ij}|} < 1 \Leftrightarrow \min_{i\in\mathbb{I}} \left(|(M_a)_{ii}| - \sum_{j\neq i} |(M_a)_{ij}| \right) > 0.$$

Moreover, if Λ is compact and $a \to M_a$ is continuous, then (21) is equivalent to

$$\inf_{a\in\Lambda} \min_{i\in\mathbb{I}} \left(|(M_a)_{ii}| - \sum_{j\neq i} |(M_a)_{ij}| \right) > 0.$$

Proof of Theorem 3 Condition (i) in Lemma 8 is immediately verified, and we turn to proving (ii). We have

$$\mu_1 := \sum_{j>i} |(M_a)_{ij}| \leq \tau \left(\frac{\sigma_i^2}{h^2} + \frac{5b_i^-}{2h} \right)$$

(omitting the dependency on k and a in $\sigma, b^{\pm}, r$) and

$$\mu_2 := |(M_a)_{ii}| - \sum_{j<i} |(M_a)_{ij}| \geq \frac{3}{2} + \tau \left(\frac{\sigma_i^2}{h^2} - \frac{2b_i^+}{2h} + \frac{3b_i^-}{2h} + r \right).$$

By the CFL condition (11), there exists $\epsilon > 0$ such that $\frac{\tau\|b\|_\infty}{h} \leq \frac{3}{2} - \epsilon$. This implies

$$\frac{3}{2} - \frac{\epsilon}{2} + \tau \left(-\frac{2b_i^+}{2h} + \frac{3b_i^-}{2h} \right) \geq \frac{\epsilon}{2} + \tau \frac{5b_i^-}{2h}$$

and therefore

$$\mu_2 \geq \left(\tau \frac{\sigma_i^2}{h^2} + \frac{\epsilon}{2} + \tau r \right) + \left(\tau \frac{5b_i^-}{2h} + \frac{\epsilon}{2} \right).$$

Then by using $\frac{a_1 + a_2}{c_1 + c_2} \leq \max\left(\frac{a_1}{c_1}, \frac{a_2}{c_2} \right)$ for numbers $a_i, c_i \geq 0$, we obtain

$$\frac{\mu_1}{\mu_2} \leq \max \left(\frac{\tau \frac{\sigma_i^2}{h^2}}{\tau \frac{\sigma_i^2}{h^2} + \frac{\epsilon}{2} + \tau r}, \frac{\tau \frac{5b_i^-}{2h}}{\tau \frac{5b_i^-}{2h} + \frac{\epsilon}{2}} \right).$$

Taking τ small enough such that for instance $\frac{\epsilon}{2} + \tau r \geq \frac{\epsilon}{4}$, and since $b(.)$ and $\sigma(.)$ are bounded functions (by (A1)), we obtain the bound

$$\sup_{a\in A} \max_{i\in\mathbb{I}} \frac{\sum_{j>i} |(M_a)_{ij}|}{|(M_a)_{ii}| - \sum_{j<i} |(M_a)_{ij}|} \leq \max \left(\frac{\tau \frac{\|\sigma^2\|_\infty}{h^2}}{\tau \frac{\|\sigma^2\|_\infty}{h^2} + \frac{\epsilon}{4}}, \frac{\tau \frac{5\|b^-\|_\infty}{2h}}{\tau \frac{5\|b^-\|_\infty}{2h} + \frac{\epsilon}{2}} \right).$$

Since the last bound is a constant < 1, we can apply Lemma 8 to obtain the existence and uniqueness of the solution of the BDF2 scheme.

4 Proof of Theorem 4 (stability in the *A*-norm)

The proof consists of three main steps: first, we show a "linear" recursion for the error (Lemma 10); second, we pass from such a recursion for the error in vector form to a scalar recursion (Lemma 11); finally, we show the stability estimate from this scalar recursion (Lemma 12).

4.1 Treatment of the nonlinearity

Given a function $\phi : [0, T] \times \mathbb{R} \times \Lambda \to \mathbb{R}$, for any $(t, x) \in [0, T] \times \mathbb{R}$ we will make use of the notation $co(\phi(t, x, \Lambda))$ to indicate the convex hull of ϕ with respect to its third variable, i.e.

$$co(\phi(t, x, \Lambda)) = \left\{ \sum_{n\in\mathbb{N}} \gamma_n \phi(t, x, a_n) : a_n \in \Lambda, \gamma_n \geq 0, \sum_{n\in\mathbb{N}} \gamma_n = 1 \right\}.$$

First, we have the following:

Lemma 10 *Let u be the solution of (3) and w the solution of (12). There exist coefficients $\tilde{\sigma}_i^k$, $(\tilde{b}^{\pm})_i^k$, $\tilde{r}_i^k$, such that the error $E^k = u^k - w^k$ satisfies*

$$\frac{3E_i^k - 4E_i^{k-1} + E_i^{k-2}}{2\tau} - \frac{1}{2}(\tilde{\sigma}^2)_i^k D^2 E_i^k + (\tilde{b}^+)_i^k D^{1,-} E_i^k - (\tilde{b}^-)_i^k D^{1,+} E_i^k + \tilde{r}_i^k E_i^k = -\mathcal{E}_i^k \tag{23}$$

for any $k \geq 2$ and $i \in \mathbb{I}$, where $(\tilde{\sigma}^2)_i^k$, $(\tilde{b}^{\pm})_i^k$, $\tilde{r}_i^k$ belong, respectively, to the convex hulls $co(\sigma^2(t_k, x_i, \Lambda))$, $co(b^{\pm}(t_k, x_i, \Lambda))$, $co(r(t_k, x_i, \Lambda))$.

Proof By (12) one has (for $k \geq 2$, $1 \leq i \leq I$)

$$\frac{3w_i^k - 4w_i^{k-1} + w_i^{k-2}}{2\tau} + H[w^k](t_k, x_i) = \mathcal{E}_i^k. \tag{24}$$

The scheme simply reads

$$\frac{3u_i^k - 4u_i^{k-1} + u_i^{k-2}}{2\tau} + H[u^k](t_k, x_i) = 0. \tag{25}$$

Subtracting (24) from (25), denoting also $H[u^k] \equiv (H[u^k](t_k, x_i))_{1\leq i\leq I}$, the following recursion is obtained for the error in $\mathbb{R}^I$:

$$\frac{3E^k - 4E^{k-1} + E^{k-2}}{2\tau} + H[u^k] - H[w^k] = -\mathcal{E}^k. \tag{26}$$

For simplicity of presentation, we first consider the case when b and r vanish, i.e. $b(.) \equiv 0$ and $r(.) \equiv 0$, and defer a sketch of the general case to the end of the proof. In this case,

$$H[u^k]_i = \sup_{a\in\Lambda}\Big\{ -\frac{1}{2}\sigma^2(t_k, x_i, a)(D^2u^k)_i + \ell(t_k, x_i, a)\Big\}. \tag{27}$$

Let us assume that σ and ℓ are continuous functions of a so that the supremum is attained.[2] For each given k, i, let then $\bar{a}_i^k \in \Lambda$ denote an optimal control in (27).
In the same way, let $\bar{b}_i^k$ denote an optimal control for $H[v^k]_i$. By using the optimality of $\bar{a}_i^k$, it holds

$$\begin{aligned}
&H[u^k]_i - H[w^k]_i\\
&=-\frac{1}{2}\sigma^2(t_k, x_i, \bar{a}_i^k)(D^2u^k)_i+\ell(t_k, x_i, \bar{a}_i^k)-\sup_{a\in\Lambda}\Big\{-\frac{1}{2}\sigma^2(t_k, x_i, a)(D^2w^k)_i+\ell(t_k, x_i, a)\Big\}\\
&\le-\frac{1}{2}\sigma^2(t_k, x_i, \bar{a}_i^k)(D^2u^k)_i-\Big(-\frac{1}{2}\sigma^2(t_k, x_i, \bar{a}_i^k)(D^2w^k)_i\Big)\\
&=-\frac{1}{2}\sigma^2(t_k, x_i, \bar{a}_i^k)(D^2E^k)_i
\end{aligned} \tag{28}$$

and, in the same way,

$$H[u^k]_i - H[w^k]_i \ge -\frac{1}{2}\sigma^2(t_k, x_i, \bar{b}_i^k)(D^2E^k)_i. \tag{29}$$

Therefore, combining (28) and (29), $H[u^k]_i - H[w^k]_i$ is a convex combination of $-\frac{1}{2}\sigma^2(t_k, x_i, \bar{a}_i^k)(D^2E^k)_i$ and $-\frac{1}{2}\sigma^2(t_k, x_i, \bar{b}_i^k)(D^2E^k)_i$. In particular, we can write

$$H[u^k]_i - H[w^k]_i = -\frac{1}{2}\tilde{\sigma}^2(t_k, x_i)(D^2E^k)_i, \tag{30}$$

where $\tilde{\sigma}^2(t_k, x_i)$ is a convex combination of $\sigma^2(t_k, x_i, \bar{a}_i^k)$ and $\sigma^2(t_k, x_i, \bar{b}_i^k)$. In the general case (i.e. $b, r \not\equiv 0$) one can argue in the exact same way to get

$$H[u^k]_i - H[w^k]_i = -\frac{1}{2}(\tilde{\sigma}^2)_i^k D^2E_i^k + (\tilde{b}^+)_i^k D^{1,-}E_i^k - (\tilde{b}^-)_i^k D^{1,+}E_i^k + \tilde{r}_i^k E_i^k, \tag{31}$$

where, for $\phi \in \{\sigma^2, b, r\}$,

$$\tilde{\phi}_i^k := \gamma_i^k \phi(t_k, x_i, \bar{a}_i^k) + (1-\gamma_i^k)\phi(t_k, x_i, \bar{b}_i^k)$$

for some $\gamma_i^k \in [0, 1]$. □

[2] The general case is obtained easily by considering sequences of ϵ-optimal controls and letting $\epsilon \to 0$, such that (31) below still holds for a suitably defined $\tilde{\sigma}^2, \tilde{b}^+, \tilde{b}^-, \tilde{r}$.

The same technique used above to deal with the nonlinear operator applies also to Isaacs equations, i.e. equations of the following form:

$$v_t + \sup_{a\in\Lambda_1}\inf_{b\in\Lambda_2}\Big\{ -\mathcal{L}^{(a,b)}[v](t,x) + r(t,x,a,b)v + \ell(t,x,a,b)\Big\} = 0, \quad (32)$$

where $(t,x)\in[0,T]\times\mathbb{R}^d$, $\Lambda_1,\Lambda_2\subset\mathbb{R}^m$ are compact sets and

$$\mathcal{L}^{(a,b)}[v](t,x) = \frac{1}{2}\sigma^2(t,x,a,b)v_{xx} + b(t,x,a,b)v_x.$$

To simplify the presentation, let us consider again $b, r\equiv 0$, and now also $\ell\equiv 0$ (indeed, one can easily verify that as in (28) the term ℓ would not appear in (34) and (35), and the case of non-zero b and r is treated similar to the Proof of Lemma 10). By analogous definitions and reasoning to above, we get (26), where, for $\phi\in\{u,w\}$,

$$H[\phi^k]_i = \sup_{a\in\Lambda_1}\inf_{b\in\Lambda_2}\Big\{ -\frac{1}{2}\sigma^2(t,x,a,b)(D_x^2\phi^k)_i\Big\}. \quad (33)$$

Making use of the general inequality (for any real-valued functions $(a,b)\to F_{a,b}$ and $(a,b)\to G_{a,b}$)

$$\sup_{a\in\Lambda_1}\inf_{b\in\Lambda_2}F_{a,b} - \sup_{a\in\Lambda_1}\inf_{b\in\Lambda_2}G_{a,b} \ge \inf_{a\in\Lambda_1}\inf_{b\in\Lambda_2}(F_{a,b}-G_{a,b}),$$

we obtain

$$H[u^k]_i - H[w^k]_i \ge \inf_{a\in\Lambda_1}\inf_{b\in\Lambda_2}\Big\{ -\frac{1}{2}\sigma^2(t,x,a,b)(D_x^2E^k)_i\Big\}. \quad (34)$$

Analogously, one can prove

$$H[u^k]_i - H[w^k]_i \le \sup_{a\in\Lambda_1}\sup_{b\in\Lambda_2}\Big\{ -\frac{1}{2}\sigma^2(t,x,a,b)(D_x^2E^k)_i\Big\}. \quad (35)$$

From these inequalities, an equation exactly as in (23) can be derived, with a suitable convex combination $(\tilde{\sigma^2})_i^k$ of diffusion coefficients, and similar for the drift and other terms.

As a consequence, Lemma 10 – and by extension Theorem 4 – also hold for Isaacs equations of type (32), with the obvious modifications to the definition of the scheme.

4.2 A scalar error recursion

From the recursion (23) on E^k, (or its corresponding formulation for Isaacs equations), we can derive the following:

Lemma 11 *Let assumptions (A1) and (A2) in Theorem 4 be satisfied. Then there exists a constant $C \geq 0$ such that*

$$\frac{1}{2}\Big((3-C\tau)|E^k|_A^2 - 4|E^{k-1}|_A^2 + |E^{k-2}|_A^2\Big) + |E^k - E^{k-1}|_A^2 - |E^{k-1} - E^{k-2}|_A^2 \leq 2\tau |E^k|_A \, |\mathcal{E}^k|_A. \tag{36}$$

Proof For simplicity of presentation we will assume that b has constant positive sign. The case of b with variable sign can be treated in a similar way obtaining estimates analogous to those below separately for the positive and negative part of b and then summing up.
We remark that for $E \in \mathbb{R}^I$, $-D^2E = AE$, where A is the finite difference matrix defined in (9) and D^2 as in (4). By (23), we get the following:

$$\frac{3E^k - 4E^{k-1} + E^{k-2}}{2\tau} + \Delta^k A E^k + F^k B E^k + R^k E^k = -\mathcal{E}^k, \tag{37}$$

where $\Delta^k := \frac{1}{2}\operatorname{diag}((\tilde{\sigma}^2)_i^k)$, $F^k = \operatorname{diag}(\tilde{b}_i^k)$, $R_k = \operatorname{diag}(\tilde{r}_i^k)$ and

$$B = \frac{1}{2h}\begin{pmatrix} 3 & 0 & & & \\ -4 & 3 & 0 & & \\ 1 & -4 & \ddots & \ddots & \\ 0 & \ddots & \ddots & \ddots & 0 \\ \ddots & \ddots & 1 & -4 & 3 \end{pmatrix}.$$

We form the scalar product of (37) with AE^k. By using the identity $2\langle a-b, a\rangle_A = |a|_A^2 + |a-b|_A^2 - |b|_A^2$, one has:

$$\begin{aligned} &\langle 3E^k - 4E^{k-1} + E^{k-2}, E^k\rangle_A \\ &= 4\langle E^k - E^{k-1}, E^k\rangle_A - \langle E^k - E^{k-2}, E^k\rangle_A \\ &= \frac{1}{2}\left(4|E^k|_A^2 + 4|E^k - E^{k-1}|_A^2 - 4|E^{k-1}|_A^2\right) - \frac{1}{2}\left(|E^k|_A^2 + |E^k - E^{k-2}|_A^2 - |E^{k-2}|_A^2\right) \\ &\geq \frac{1}{2}\left(3|E^k|_A^2 - 4|E^{k-1}|_A^2 + |E^{k-2}|_A^2\right) + |E^k - E^{k-1}|_A^2 - |E^{k-1} - E^{k-2}|_A^2, \end{aligned} \tag{38}$$

where we have also used $|a+b|^2 \leq 2|a|^2 + 2|b|^2$. From $(\sigma^2)_i^k \geq \eta > 0$ for all k, i:

$$\langle \Delta^k A E^k, AE^k\rangle \geq \frac{\eta}{2}\|AE^k\|^2, \tag{39}$$

where $\|\cdot\|$ denotes the canonical Euclidean norm in $\mathbb{R}^I$.

In order to estimate the drift component, let us introduce the notation

$$\delta E := (E_i - E_{i-1})_{1\le i\le I}, \quad \delta_2 E := (E_{i-1} - E_{i-2})_{1\le i\le I} \tag{40}$$

with the convention that $E_i = 0$ for all indices i which are not in $\mathbb{I}$. It holds:

$$\begin{aligned}|\langle F^k BE^k, AE^k\rangle| &= \left|\frac{1}{2h}\langle F^k(3E_i^k - 4E_{i-1}^k + E_{i-2}^k)_{i\in\mathbb{I}}, AE^k\rangle\right| \\ &= \left|\frac{1}{2h}\langle F^k\big(3\delta E^k - \delta_2 E^k\big),\ AE^k\rangle\right| \\ &\le \frac{1}{2h}\Big\{3\|F^k\delta E^k\|\,\|AE^k\| + \|F^k\delta_2 E^k\|\,\|AE^k\|\Big\}.\end{aligned}$$

By using the boundedness of the drift term, and $\|\delta E^k\|, \|\delta_2 E^k\| \le h|E^k|_A$,

$$\begin{aligned}|\langle F^k BE^k, AE^k\rangle| &\le \frac{\|b\|_\infty}{2h}\Big\{3\|AE^k\|\|\delta E^k\| + \|AE^k\|\|\delta_2 E^k\|\Big\} \\ &\le 2\|b\|_\infty\,\|AE^k\|\,|E^k|_A. \end{aligned}\tag{41}$$

For the last term, using the boundedness of r and the Cauchy-Schwarz inequality,

$$|\langle R^k E^k, AE^k\rangle| \le \|r\|_\infty \|E^k\|\|AE^k\|. \tag{42}$$

Therefore, putting (39), (41) and (42) together,

$$\begin{aligned}&\langle \Delta^k AE^k + F^k BE^k + R^k E^k, AE^k\rangle \\ &\quad\ge \frac{\eta}{2}\|AE^k\|^2 - 2\|b\|_\infty\|AE^k\|\,|E^k|_A - \|r\|_\infty\|AE^k\|\|E^k\|.\end{aligned}\tag{43}$$

Easy calculus shows that the minimal eigenvalue of A is $\lambda_{\min}(A) = \frac{4}{h^2}\sin^2(\frac{\pi h}{2}) \ge 4$. Hence $\langle X, AX\rangle \ge 4\langle X, X\rangle$ and therefore $\|X\| \le \frac{1}{2}|X|_A$. In the same way, we have also $|X|_A \le \frac{1}{2}\|AX\|$. Hence it holds

$$\langle \Delta^k AE^k + F^k BE^k + R^k E^k, AE^k\rangle \ge \frac{\eta}{2}\|AE^k\|^2 - C_1\|AE^k\||E^k|_A \tag{44}$$

with $C_1 := 2\|b\|_\infty + \frac{1}{2}\|r\|_\infty$. By using $C_1\|AE^k\||E^k|_A \le \frac{\eta}{2}\|AE^k\|^2 + \frac{1}{2\eta}C_1^2|E^k|_A^2$,

$$\langle \Delta^k AE^k + F^k BE^k + R^k E^k, AE^k\rangle \ge -\frac{1}{2\eta}C_1^2|E^k|_A^2. \tag{45}$$

Then, combining (38) and (45), we obtain the desired inequality with $C := \frac{2}{\eta}C_1^2$. □

4.3 A universal stability lemma

In the following, it is assumed that $|\cdot|$ is any vectorial norm. Combined with Lemmas 10 and 11, the following Lemma 12 with the A-norm $|\cdot| \equiv |\cdot|_A$ immediately gives Theorem 14. In Sect. 5, we will use the result for the canonical Euclidean norm $|\cdot| \equiv \|\cdot\|$ to prove Theorem 5.

In order to prove the following Lemma 12, we will exploit properties of the matrix

$$M_\tau := \begin{pmatrix} (3-C\tau) & -4 & 1 & 0 & & \\ 0 & (3-C\tau) & -4 & \ddots & \ddots & \\ & 0 & \ddots & \ddots & 1 & \\ & & \ddots & \ddots & -4 & \\ & & & 0 & (3-C\tau) & \end{pmatrix}, \tag{46}$$

in particular the fact that $(M_\tau)^{-1} \geq 0$ for τ small enough (which we prove).

Lemma 12 *Assume that there exists a constant $C \geq 0$ such that $\forall k = 2, \ldots, N$:*

$$\frac{1}{2}\Big((3-C\tau)|E^k|^2 - 4|E^{k-1}|^2 + |E^{k-2}|^2\Big) + |E^k - E^{k-1}|^2 - |E^{k-1} - E^{k-2}|^2 \\ \leq 2\tau|E^k|\,|\mathcal{E}^k|. \tag{47}$$

Then there exists a constant $C_1 \geq 0$ and $\tau_0 > 0$ such that $\forall 0 < \tau \leq \tau_0$, $\forall n \leq N$:

$$\max_{2\leq k\leq n} |E^k|^2 \leq C_1\Big(|E^0|^2 + |E^1|^2 + \tau \sum_{2\leq j\leq n} |\mathcal{E}^j|^2\Big). \tag{48}$$

Proof Let us denote

$$x_k := |E^k|^2 \quad \text{and} \quad y_k := |E^k - E^{k-1}|^2,$$

so that (47) reads

$$\Big((3-C\tau)x_k - 4x_{k-1} + x_{k-2}\Big) \leq 2(y_{k-1} - y_k) + 4\tau|E^k|\,|\mathcal{E}^k|. \tag{49}$$

For a given $\tau > 0$ and given k, let $M_\tau \in \mathbb{R}^{(k-1)\times(k-1)}$ as defined in (46). Let $z, q \in \mathbb{R}^{k-1}$ be defined by

$$z := (x_k, x_{k-1}, \ldots, x_2)^T \quad \text{and} \quad q := (2(y_{j-1} - y_j) + 4\tau|E^j|\,|\mathcal{E}^j|)_{j=k,\ldots,2}.$$

By (49), we have

$$M_\tau z \leq q. \tag{50}$$

We notice that $M_\tau = (3 - C\tau)I - 4J + J^2$ with

$$J := \text{tridiag}(0, 0, 1).$$

Hence, with

$$\lambda_1 = 2 + \sqrt{1 + C\tau} \quad \text{and} \quad \lambda_2 = 2 - \sqrt{1 + C\tau},$$

the roots of $\lambda^2 - 4\lambda + (3 - C\tau) = 0$ for $3 - C\tau \geq 0$, we can write

$$M_\tau = (\lambda_1 I - J)(\lambda_2 I - J) = \lambda_1\lambda_2 \left(I - \frac{J}{\lambda_1}\right)\left(I - \frac{J}{\lambda_2}\right).$$

Furthermore, since $J^{k-1} = 0$, it holds

$$\begin{aligned} M_\tau^{-1} &= \frac{1}{\lambda_1\lambda_2}\left(I - \frac{J}{\lambda_1}\right)^{-1}\left(I - \frac{J}{\lambda_2}\right)^{-1} \\ &= \frac{1}{\lambda_1\lambda_2}\left(\sum_{0\leq\xi\leq k-2}\left(\frac{J}{\lambda_1}\right)^\xi\right)\left(\sum_{0\leq\xi\leq k-2}\left(\frac{J}{\lambda_2}\right)^\xi\right) = \sum_{p=0}^{k-2} a_p J^p, \end{aligned}$$

where

$$a_p := \sum_{j=0}^{p} \frac{1}{\lambda_1^{j+1}\lambda_2^{p-j+1}} = \frac{1}{\lambda_2^{p+2}} \sum_{j=0}^{p} \left(\frac{\lambda_2}{\lambda_1}\right)^{j+1}.$$

Therefore $M_\tau^{-1} \geq 0$ componentwise (for $\tau < 3/C$), and using (50) it holds $z \leq M_\tau^{-1} q$.

It is possible to prove that there exists $\tau_0 > 0$ and a constant $C_0 \geq 0$ (depending only on T) such that $\forall 0 < \tau \leq \tau_0$ and $\forall p \geq 0$:

$$0 \leq a_p \leq C_0 \quad \text{and} \quad a_p - a_{p-1} \geq 0. \tag{51}$$

We postpone the Proof of (51) to the end. For the first component of z, we deduce

$$\begin{aligned} x_k &\leq \sum_{j=0}^{k-2} a_j q_{j+1} \\ &\leq 2\sum_{j=0}^{k-2} a_j (y_{k-j-1} - y_{k-j}) + 4C_0\tau \sum_{j=2}^{k} |E^j|\,|\mathcal{E}^j| \\ &= -2a_0 y_k + 2\sum_{j=0}^{k-3}(a_j - a_{j+1}) y_{k-j+1} + 2a_{k-2} y_1 + 4C_0\tau \sum_{j=2}^{k} |E^j|\,|\mathcal{E}^j|, \end{aligned} \tag{52}$$

for all $k \geq 2$, where, for (52), we have used the fact that $a_p \leq C_0$. Since $y_j \geq 0, \forall j$, by definition, $a_{k-2} \leq C_0$, $a_0 = \frac{1}{\lambda_1 \lambda_2} \geq 0$ and $a_j - a_{j-1} \geq 0, \forall j$, we obtain

$$x_k \leq 2C_0 y_1 + 4C_0 \tau \sum_{j=2}^{k} |E^j| \, |\mathcal{E}^j|. \tag{53}$$

Recalling the definition of x_k and y_k, for any $2 \leq k \leq n$ one has:

$$\begin{aligned} |E^k|^2 &\leq 2C_0 |E^1 - E^0|^2 + 4C_0 \tau \sum_{j=2}^{k} |E^j| \, |\mathcal{E}^j| \\ &\leq 4C_0 (|E^0|^2 + |E^1|^2) + 4C_0 \tau \Big(\max_{2 \leq k \leq n} |E^k| \Big) \sum_{j=2}^{n} |\mathcal{E}^j| \\ &\leq 4C_0 (|E^0|^2 + |E^1|^2) + \frac{1}{2} \Big(\max_{2 \leq k \leq n} |E^k| \Big)^2 + 8C_0^2 \tau^2 \Big(\sum_{j=2}^{n} |\mathcal{E}^j| \Big)^2 \end{aligned}$$

(where we made use of $2ab \leq \frac{a^2}{K} + Kb^2$ for any $a, b \geq 0$ and $K > 0$). Hence, we obtain

$$\Big(\max_{2 \leq k \leq n} |E^k| \Big)^2 \leq C_1 \Big(|E^0|^2 + |E^1|^2 + \tau \sum_{j=2}^{n} |\mathcal{E}^j|^2 \Big)$$

with $C_1 := \max(8C_0, 16C_0^2 T)$ (we used $\Big(\sum_{j=2}^{n} |\mathcal{E}^j| \Big)^2 \leq n \sum_{j=2}^{n} |\mathcal{E}^j|^2$ and $n\tau \leq T$).

It remains to prove (51). From the definition of a_p one has

$$a_p = \frac{1}{\lambda_2^{p+2}} \sum_{j=0}^{p} \left(\frac{\lambda_2}{\lambda_1} \right)^{j+1} \leq \frac{1}{\lambda_2^{p+2}} \left(1 - \frac{\lambda_2}{\lambda_1} \right)^{-1}$$

for $p = 0, \ldots, k-2$. Observing that $\frac{\lambda_2}{\lambda_1} \leq \frac{1}{3}$, it follows that

$$a_p \leq \frac{3}{2\lambda_2^{p+2}} \leq \frac{3}{2(2 - \sqrt{1 + C\tau})^n}.$$

Notice that $\sqrt{1 + C\tau} \leq 1 + C\tau$, and also that $e^{-x} \leq 1 - x/2, \forall x \in [0, 1]$. Hence $(2 - \sqrt{1 + C\tau})^n \geq (2 - (1 + C\tau))^n = (1 - C\tau)^n \geq (e^{-2C\tau})^n = e^{-2Ct_n}$ for $C\tau \leq \frac{1}{2}$, and therefore $a_p \leq \frac{3}{2} e^{2Ct_n}$. The desired result follows with $C_0 := \frac{3}{2} e^{2CT}$ and $\tau_0 := \frac{1}{2C}$.

Moreover, one has

$$a_p - a_{p-1} = \frac{1}{\lambda_2^{p+1}} \left(\frac{1}{\lambda_2} \sum_{j=0}^{p} \left(\frac{\lambda_2}{\lambda_1} \right)^{j+1} - \sum_{j=0}^{p-1} \left(\frac{\lambda_2}{\lambda_1} \right)^{j+1} \right),$$

which is nonnegative for τ small enough thanks to the fact that $\lambda_1, \lambda_2 \geq 0$ and $\lambda_2 \leq 1$. □

Remark 13 From the previous proof and the Proof of Lemma 11 one can deduce that the restriction

$$\tau \leq \frac{\eta}{C_1^2}$$

(where $C_1 = 2\|b\|_\infty + \frac{1}{2}\|r\|_\infty$) has to be imposed on the time step. From the theoretical point of view this makes the scheme not suitable for nearly-degenerate equations. However, in our numerical tests we did not observe any stability issue even in the case of degenerate problems (see Sect. 7.1).

5 Stability in the Euclidean norm

The fundamental stability result given by Lemma 12 applies to any vectorial norm. In this section, we discuss some special cases where (47) can be obtained for the Euclidean norm $|\cdot| = \|\cdot\|$.

We first prove the stability result for this norm under the extra assumption (A3), i.e., the control may appear everywhere except in the diffusion term, which must also be Lipschitz continuous in the following proof (see Assumption (A3)).

5.1 Proof of Theorem 5 (stability in the Euclidean norm)

We consider the scalar product of (37) directly with E^k (instead of AE^k previously used), again in the situation where $b \geq 0$ to simplify the argument (but the general case follows analogous to the Proof of Lemma 11). We obtain:

$$\langle E^k, 3E^k - 4E^{k-1} + E^{k-2} \rangle + 2\tau \langle E^k, \Delta^k A E^k + F^k B E^k + R^k E^k \rangle = -2\tau \langle E^k, \mathcal{E}^k \rangle. \tag{54}$$

As in Sect. 4.2, we have

$$\begin{aligned} &\langle E^k, 3E^k - 4E^{k-1} + E^{k-2} \rangle \\ &\quad \geq \frac{1}{2} \left(3\|E^k\|^2 - 4\|E^{k-1}\|^2 + \|E^{k-2}\|^2 \right) + \|E^k - E^{k-1}\|^2 - \|E^{k-1} - E^{k-2}\|^2. \end{aligned} \tag{55}$$

We now focus on bounding the other terms on the left-hand side of (54).

By using the Lipschitz continuity of σ^2 one has

$$\begin{aligned}\langle E^k, \Delta^k A E^k\rangle &= \sum_{i\in\mathbb{I}} \frac{(\sigma_i^k)^2}{2h^2}(-E_{i+1}^k + 2E_i^k - E_{i-1}^k)E_i^k \\ &= \sum_{i\in\mathbb{I}} \frac{(\sigma_{i-1}^k)^2}{2h^2}(E_{i-1}^k - E_i^k)^2 + \sum_{i\in\mathbb{I}} \left(\frac{(\sigma_{i-1}^k)^2}{2h^2} - \frac{(\sigma_i^k)^2}{2h^2}\right)(E_{i-1}^k - E_i^k)E_i^k \\ &\geq \frac{\eta}{2h^2}\sum_{i\in\mathbb{I}}(E_{i-1}^k - E_i^k)^2 - \frac{L}{2h}\sum_{i\in\mathbb{I}}|E_{i-1}^k - E_i^k||E_i^k|. \end{aligned} \tag{56}$$

Therefore, by the Cauchy–Schwarz inequality, one obtains

$$\langle E^k, \Delta_k A E^k\rangle \geq \frac{\eta}{2h^2}\|\delta E^k\|^2 - \frac{L}{2h}\|\delta E^k\|\|E^k\|, \tag{57}$$

where δE^k is defined by (40). Moreover, for the first order term one has

$$\begin{aligned}\langle E^k, F^k B E^k\rangle &= \sum_{i\in\mathbb{I}} \frac{b_i}{2h}(3E_i^k - 4E_{i-1}^k + E_{i-2}^k)E_i^k \\ &\geq -\frac{3\|b\|_\infty}{2h}\sum_{i\in\mathbb{I}}|E_i^k - E_{i-1}^k||E_i^k| - \frac{\|b\|_\infty}{2h}\sum_{i\in\mathbb{I}}|E_{i-1}^k - E_{i-2}^k||E_i^k| \\ &\geq -\frac{2\|b\|_\infty}{h}\|\delta E^k\|\|E^k\|, \end{aligned} \tag{58}$$

where for the last equality we have used that $\|\delta_2 E^k\| \leq \|\delta E^k\|$. Putting together estimates (57) and (58), using the fact that $\langle E^k, R^k E^k\rangle \geq -\|r\|_\infty\|E^k\|^2$, we get

$$\begin{aligned}\langle E^k, \Delta_k A E^k + F_k B E^k + R^k E^k\rangle &\geq \frac{\eta}{2h^2}\|\delta E^k\|^2 - \frac{C_1}{2h}\|\delta E^k\|\|E^k\| - \|r\|_\infty\|E^k\|^2 \\ &\geq \frac{\eta}{4h^2}\|\delta E^k\|^2 - \left(\frac{C_1^2}{4\eta} + \|r\|_\infty\right)\|E^k\|^2, \end{aligned}$$

where we have denoted $C_1 := L + 4\|b\|_\infty$ and have used again the Cauchy–Schwarz inequality. Hence, together with (55), this gives (47) with $|\cdot| = \|\cdot\|$ and the constant $C := 4(\frac{C_1^2}{4\eta} + \|r\|_\infty)$. By using Lemma 12, this concludes the proof of Theorem 5. □

Remark 14 The step (56) highlights the need for Assumption (A3), Lipschitz regularity of the diffusion coefficient, in order to obtain the one-step stability inequality (47). This can be avoided in the A-norm stability analysis, Lemma 11, by using a different inner product, which directly gives (39) and only requires uniform ellipticity.

The adaptation of (A3) to the controlled case would impose some Lipschitz continuity of the feedback control with respect to the state variable. Such regularity of the control cannot usually be expected (see for instance the tests in Sect. 7.2).

5.2 Linear equation with degenerate diffusion term

The next result concerns the case of a possibly degenerate diffusion term. It will require more restrictive assumptions on the drift and diffusion terms, and we shall assume that there is no control here. Indeed, in this case, one cannot count on the positive term coming from the non-degenerate diffusion which, in the proof of Theorem 5, is used to compensate the negative correction terms coming from the drift term. This leads us to consider the following assumption:

ASSUMPTION (A5).

(i) The function $r(.)$ is bounded.
(ii) The drift and diffusion coefficients are independent of the control : $b \equiv b(t,x)$ and $\sigma \equiv \sigma(t,x)$.
(iii) there exist $L_1, L_2 \geq 0$ such that, for all t, x, h:

$$|b(t,x) - b(t,y)| \leq L_1 |x-y|, \tag{59}$$

$$\frac{\sigma^2(t,x-h) - 2\sigma^2(t,x) + \sigma^2(t,x+h)}{h^2} \geq -L_2. \tag{60}$$

(The last condition is equivalent to $(\sigma^2)_{xx} \geq -L_2$ in the differentiable case.)

Proposition 15 *Let assumption (A5) be satisfied. Then* (47) *holds for* $|\cdot| = \|\cdot\|$.

Proof We consider again the scalar recursion (54). For any vector $E = (E_i)_{1\leq i\leq I}$ (with $E_j = 0$ for $j \in \{-1, 0, I+1, I+2\}$), it holds:

$$\begin{aligned} E_i(2E_i - E_{i-1} - E_{i+1}) &\geq 2|E_i|^2 - \frac{1}{2}(|E_i|^2 + |E_{i-1}|^2) - \frac{1}{2}(|E_i|^2 + |E_{i+1}|^2) \\ &\geq \frac{1}{2}\,(2|E_i|^2 - |E_{i-1}|^2 - |E_{i+1}|^2). \end{aligned}$$

Hence, by the semi-concavity assumption (60) on σ^2,

$$\begin{aligned} \langle E^k, \Delta^k A E^k \rangle &= \sum_{1\leq i\leq I} \frac{\sigma_i^2}{2h^2} E_i^k (2E_i^k - E_{i-1}^k - E_{i+1}^k) \\ &\geq \sum_{1\leq i\leq I} \frac{\sigma_i^2}{4h^2} (-|E_{i-1}^k|^2 + 2|E_i^k|^2 - |E_{i+1}^k|^2) \\ &\geq \sum_{1\leq i\leq I} \left(\frac{-\sigma_{i-1}^2 + 2\sigma_i^2 - \sigma_{i+1}^2}{4h^2} \right) |E_i^k|^2. \\ &\geq -\frac{L_2}{4} \|E^k\|^2. \end{aligned} \tag{61}$$

Now we focus on a lower bound for $\langle E^k,\, F^k B E^k\rangle$. Let $y_i^k = |E_i^k - E_{i-1}^k|^2$. First,

$$\begin{aligned}(3E_i^k - 4E_{i-1}^k + E_{i-2}^k)E_i^k &= \frac{1}{2}(3|E_i^k|^2 - 4|E_{i-1}^k|^2 + |E_{i-2}^k|^2)\\ &\quad + \frac{1}{2}(4|E_i^k - E_{i-1}^k|^2 - |E_i^k - E_{i-2}^k|^2)\\ &\geq \frac{1}{2}(3|E_i^k|^2 - 4|E_{i-1}^k|^2 + |E_{i-2}^k|^2) + \frac{1}{2}(2y_i^k - 2y_{i-1}^k).\end{aligned}$$

We assume again $b_i \geq 0$ for all i to simplify the presentation. The case where $b_i \leq 0$ for some i is similar. Then, the following bound holds:

$$\begin{aligned}\langle E^k,\, F^k B E^k\rangle &= \sum_{i=1}^{I} \frac{b_i}{2h}(3E_i^k - 4E_{i-1}^k + E_{i-2}^k)E_i^k = \sum_{i=1}^{I+2} \frac{b_i}{2h}(3E_i^k - 4E_{i-1}^k + E_{i-2}^k)E_i^k\\ &\geq \sum_{i=1}^{I+2} \frac{b_i}{4h}(3|E_i^k|^2 - 4|E_{i-1}^k|^2 + |E_{i-2}^k|^2) + \sum_{i=1}^{I+2} \frac{b_i}{h}(y_i^k - y_{i-1}^k)\\ &\geq \sum_{i=1}^{I} \left(\frac{3b_i - 4b_{i+1} + b_{i+2}}{4h}\right)|E_i^k|^2 + \sum_{i=1}^{I+1} \left(\frac{b_i - b_{i+1}}{h}\right)y_i^k\end{aligned}$$

(where we have used $y_0^k = y_{I+2}^k = 0$ and $\sum_{1\leq i\leq I+2} b_i (E_{i-2}^k)^2 = \sum_{1\leq i\leq I} b_{i+2}(E_i^k)^2$ as well as $\sum_{1\leq i\leq I+2} b_i (E_{i-1}^k)^2 = \sum_{0\leq i\leq I+1} b_{i+1}(E_i^k)^2 = \sum_{1\leq i\leq I} b_{i+1}(E_i^k)^2$). Then, by the Lipschitz continuity of $b(.)$ and the bound $y_i^k \leq 2(E_i^k)^2 + 2(E_{i-1}^k)^2$, we have

$$\langle E^k,\, F^k B E^k\rangle \geq -L_1 \sum_{i=1}^{I} |E_i^k|^2 - L_1 \sum_{i=1}^{I+1} y_i^k \geq -3L_1 \|E^k\|^2. \tag{62}$$

By combining the bounds (61) and (62), we obtain

$$\langle E^k,\, \Delta^k A E^k\rangle + \langle E^k,\, F^k B E^k\rangle + \langle E^k,\, R^k E^k\rangle \geq -(\frac{L_2}{4} + 3L_1 + \|r\|_\infty)\|E^k\|^2.$$

Therefore, inequality (47) is obtained with $C := 4(\frac{L_2}{4} + 3L_1 + \|r\|_\infty)$, which leads to the desired stability estimate. □

5.3 Extension to a two-dimensional case

Under suitable assumptions, the result of Theorem 5 can be extended to multi-dimensional equations. We only sketch the main extra features and analysis steps as the notation is significantly lengthier. In the nonlinear case of HJB and Isaacs equations, the derivation of a linear error recursion can be carried out exactly as in

Sect. 4.1 so that we can restrict ourselves to the following linear case with appropriate assumptions on the coefficients specified below,

$$v_t - \frac{1}{2}\,\mathrm{tr}[\Sigma(t,x)D_x^2 v] + b(t,x)D_x v + r(t,x)v + \ell(t,x) = 0$$

for a positive definite matrix Σ and a drift vector b. We consider the two-dimensional case ($d = 2$), as the approximation of the diffusion term with suitable properties is better understood here for diagonally dominant diffusion tensor (see also Remark 17, (iii)). For simplicity, we take $r, \ell \equiv 0$, but this condition can easily be removed as in earlier sections. Lastly, we omit for brevity the dependence of the coefficients on the time variable, which is inconsequential for the stability analysis.

Then with

$$\Sigma(x,y) := \begin{pmatrix} \sigma_1^2(x,y) & \rho\sigma_1\sigma_2(x,y) \\ \rho\sigma_1\sigma_2(x,y) & \sigma_2^2(x,y) \end{pmatrix} \quad \text{and} \quad b(x,y) := \begin{pmatrix} b_1(x,y) \\ b_2(x,y) \end{pmatrix},$$

where $\sigma_1, \sigma_2 \geq 0$ and $\rho \in [-1,1]$ is the correlation parameter, the equation reads (by slight abuse of notation)

$$v_t - \frac{1}{2}\sigma_1^2(x,y)v_{xx} - \rho\sigma_1\sigma_2(x,y)v_{xy} - \frac{1}{2}\sigma_2^2(x,y)v_{yy} + b_1(x,y)v_x + b_2(x,y)v_y = 0.$$

The computational domain is given by $\Omega := (x_{\min}, x_{\max}) \times (y_{\min}, y_{\max})$. We introduce the discretization in space defined by the steps $h_x, h_y > 0$ and we denote by $\mathcal{G}_{(h_x,h_y)}$ the associated mesh. In what follows, given any function ϕ of $(x,y) \in \Omega$, we will denote $\phi_{ij} = \phi(x_i, y_j)$ for $(i,j) \in \mathbb{I} := \mathbb{I}_1 \times \mathbb{I}_2$, where $\mathbb{I}_1 = \{1, \ldots, I_1\}, \mathbb{I}_2 = \{1, \ldots, I_2\}$.

Assuming that $\rho \geq 0$ everywhere (the case when $\rho \leq 0$ is similar), we consider a 7-point stencil for the second order derivatives (see [13, Section 5.1.4]):

$$v_{xx} \sim \frac{v_{i-1,j} - 2v_{ij} + v_{i+1,j}}{h_x^2} =: D_{xx}^2 v_{ij}, \qquad v_{yy} \sim \frac{v_{i,j-1} - 2v_{ij} + v_{i,j+1}}{h_y^2} =: D_{yy}^2 v_{ij}$$

$$v_{xy} \sim \frac{-v_{i,j-1} - v_{i,j+1} - v_{i-1,j} - v_{i+1,j} + v_{i-1,j-1} + v_{i+1,j+1} + 2v_{ij}}{2h_x h_y} =: D_{xy}^2 v_{ij}$$

and the BDF approximation of the first order derivatives

$$D_x^{1,-}u_{ij} := \frac{3u_{ij} - 4u_{i-1,j} + u_{i-2,j}}{2h_x} \quad \text{and} \quad D_x^{1,+}u_{ij} := -\left(\frac{3u_{ij} - 4u_{i+1,j} + u_{i+2,j}}{2h_x}\right),$$

$$D_y^{1,-}u_{ij} := \frac{3u_{ij} - 4u_{i,j-1} + u_{i,j-2}}{2h_y} \quad \text{and} \quad D_y^{1,+}u_{ij} := -\left(\frac{3u_{ij} - 4u_{i,j+1} + u_{i,j+2}}{2h_y}\right).$$

The scheme is therefore defined, for $k \geq 2$, by

$$0 = \frac{u_{ij}^k - 4u_{ij}^{k-1} + u_{ij}^{k-2}}{2\tau}$$

$$-\frac{1}{2}\sigma_1^2(x_i,y_j)D_{xx}^2u_{ij}^k-\rho\sigma_1\sigma_2(x_i,y_j)D_{xy}^2u_{ij}^k-\frac{1}{2}\sigma_2^2(x_i,y_j)D_{yy}^2u_{ij}^k$$
$$+b_1^+(x_i,y_j)D_x^{1,-}u_{ij}^k-b_1^-(x_i,y_j)D_x^{1,+}u_{ij}^k+b_2^+(x_i,y_j)D_y^{1,-}u_{ij}^k-b_2^-(x_i,y_j)D_y^{1,+}u_{ij}^k. \tag{63}$$

A straightforward calculation shows that

$$\begin{aligned}&\sigma_1^2(x_i,y_j)D_{xx}^2u_{ij}+2\rho\sigma_1\sigma_2(x_i,y_j)D_{xy}^2u_{ij}+\sigma_2^2(x_i,y_j)D_{yy}^2u_{ij}\\&\quad=\alpha_{ij}D_{xx}^2u_{ij}+\beta_{ij}D_{yy}^2u_{ij}+\gamma_{ij}\left(u_{i-1,j-1}-2u_{ij}+u_{i+1,j+1}\right),\end{aligned} \tag{64}$$

with

$$\alpha_{ij}:=\frac{\sigma_1(x_i,y_j)}{h_x}\left(\frac{\sigma_1(x_i,y_j)}{h_x}-\frac{\rho\sigma_2(x_i,y_j)}{h_y}\right),$$
$$\beta_{ij}:=\frac{\sigma_2(x_i,y_j)}{h_y}\left(\frac{\sigma_2(x_i,y_j)}{h_y}-\frac{\rho\sigma_1(x_i,y_j)}{h_x}\right),\qquad \gamma_{ij}:=\frac{\rho(x_i,y_j)\sigma_1(x_i,y_j)\sigma_2(x_i,y_j)}{h_yh_x}.$$

The scheme is completed with the following boundary conditions:

$$u_{i,j}^k=v(t_k,x_i,y_j),\quad \forall i\in\{-1,0\}\cup\{I_1+1,I_1+2\},\ j\in\mathbb{I}_2,$$
$$u_{i,j}^k=v(t_k,x_i,y_j),\quad \forall j\in\{-1,0\}\cup\{I_2+1,I_2+2\},\ i\in\mathbb{I}_1.$$

For simplicity, assume $h_x=h_y=:h$. We consider the following assumptions:

ASSUMPTIONS.

(A1'): $\|b_i\|_\infty<\infty$ for $i=1,2$;
(A2'): $\exists\eta>0,\forall(x,y)\in\Omega,\forall i\neq j$: $\sigma_i^2(x,y)-\rho(x,y)\sigma_i(x,y)\sigma_j(x,y)\geq\eta$;
(A3'): $\forall i,j=1,2$, $\sigma_i\sigma_j$ is Lipschitz continuous on Ω.

We then have the following result.

Proposition 16 *Let assumptions (A1'),(A2') and (A3') be satisfied. Then the stability estimate* (16) *holds for* $|\cdot|=\|\cdot\|$.

Proof The proof follows by similar steps to those of Theorem 5, using (64) with $\alpha_{ij},\beta_{ij}\geq\eta/h^2$ and $\gamma_{ij}\geq0$ by assumption (A2'). □

Remark 17 (i) If $h_x\neq h_y$ and for instance $h_y=Ch_x$ for some $C\geq1$, (A2') has to hold with σ_2 replaced by σ_2/C as a result of the scaling properties of the scheme.

(ii) Observe that assumption (A2') is equivalent to requiring strong diagonal dominance of the covariance matrix.

(iii) When the strong diagonal dominance of the matrix Σ is not guaranteed, one can consider the generalized finite difference scheme in [8]. However, determining the precise set of assumptions on the coefficients needed to apply the previous arguments does not seem easy from the construction in [8].

6 Error estimates

In this section, we derive detailed error estimates for the implicit BDF2 scheme (3). For brevity, we restrict ourselves to the one-dimensional case.

In the following, we define specific instances of w_i^k, E_i^k and $\mathcal{E}_i^k$, to which we can apply the results from the preceding sections.

Let u denote the solution of (3) and let w be the solution of (1), i.e. the function v. The error associated with the scheme is then defined by

$$E_i^k := u_i^k - v(t_k, x_i), \quad i \in \mathbb{I}, 0 \le k \le N.$$

For any function ϕ we will also use the notation $\phi_i^k := \phi(t_k, x_i)$ as well as $\phi^k := (\phi_i^k)_{1\le i\le I}$ and $[\phi]_i^k := (\phi_j^m)_{(j,m)\neq(i,k)}$, and the error vector at time t_k is defined by

$$E^k := (E_1^k, \dots, E_I^k)^T = u^k - v^k, \quad 0 \le k \le N.$$

The consistency error will be denoted by $\mathcal{E}^k(\phi) := (\mathcal{E}_i^k(\phi))_{1\le i\le I} \in \mathbb{R}^I$ and for any smooth enough function ϕ is defined, in this section, as follows:

$$\begin{aligned}\mathcal{E}_i^k(\phi) := {} & \mathcal{S}^{(\tau,h)}(t_k, x_i, \phi_i^k, [\phi]_i^k) \\ & -\left(\phi_t + \sup_{a\in\Lambda}\left\{\mathcal{L}^a[\phi](t_k, x_i) + r(t_k, x_i, a)\phi + \ell(t_k, x_i, a)\right\}\right). \end{aligned} \tag{65}$$

By extension, for the exact solution v of (1), we will simply define

$$\mathcal{E}_i^k(v) := \mathcal{S}^{(\tau,h)}(t_k, x_i, v_i^k, [v]_i^k). \tag{66}$$

Note that (66) is well-defined for any continuous function.

In particular, for the scheme (3) it is clear that we have second order consistency in space and time, that is,

$$|\mathcal{E}_i^k(\phi)| \le c_1(\phi)\tau^2 + c_2(\phi)h^2 \tag{67}$$

for any sufficiently regular test function ϕ.

To prove convergence of a certain order, we can now follow the standard approach of considering the exact solution to the PDE as a solution of a perturbed finite difference scheme with the truncation error as the right-hand side. The error therefore satisfies precisely Eqs. (14) and (16) under the pertaining assumptions.

When the Euler timestepping scheme (6) is used at the first time step, by the stability of the scheme we expect to have

$$|E^1|_A \le C\tau|\mathcal{E}^1(v)|_A$$

and (14) simply reads

$$\max_{2\le k\le N} |E^k|_A^2 \le C\Big(|E^0|_A^2 + \tau^2 |\mathcal{E}^1(v)|_A^2 + \tau \sum_{2\le k\le N} |\mathcal{E}^k(v)|_A^2\Big),$$

and similarly for the L^2 error.

6.1 Proof of Theorem 7

We first prove (i). By Taylor expansion, we can write for instance, for some $\theta_1, \theta_2 \in [0, 1]$,

$$\left| v_t(t,x) - \frac{v(t,x) - v(t-\tau,x)}{\tau} \right| = |v_t(t,x) - v_t(t-\theta_1\tau,x)| \le C\tau^\delta$$

and

$$\begin{aligned}
&\left| v_t(t,x) - \frac{3v(t,x) - 4v(t-\tau,x) + v(t-2\tau,x)}{2\tau} \right| \\
&\le \left| v_t(t,x) - \frac{1}{2}\left(3v_t(t-\theta_1\tau,x) - v_t(t-(1+\theta_2)\tau,x)\right) \right| \\
&\le |v_t(t,x) - v_t(t-\theta_1\tau,x)| + \frac{1}{2}|v_t(t-\theta_1\tau,x) - v_t(t-(1+\theta_2)\tau,x)| \\
&\le C\tau^\delta + \frac{1}{2}C(2\tau)^\delta \le 2C\tau^\delta.
\end{aligned}$$

Similarly, using the higher spatial regularity, there exists a constant $C_0 \ge 0$ such that

$$\begin{aligned}
\left| v_x(t,x) - \frac{3v(t,x) - 4v(t,x-h) + v(t,x-2h)}{2h} \right| &\le C_0 C h^{\delta+1}, \\
\left| v_{xx}(t,x) - \frac{v(t,x+h) - 2v(t,x) + v(t,x-h)}{h^2} \right| &\le C_0 C h^\delta.
\end{aligned}$$

The result (i) now follows directly by inserting the obtained truncation error into the stability estimate of Theorem 5.

For the proof of (ii) (smooth case), expansion up to order 3 and 4 gives the truncation error of higher order for $k \ge 2$, and we use the fact that the error from the first backward Euler step is bounded by $\|E^1\| \le C\tau(\tau + h^2)$; in particular, we use that $(E^1 - E^0)/\tau + (\Delta^1 A + F^1 B + R^1)E^1 = -\mathcal{E}^1$, with $\|\mathcal{E}^1\| \le C(\tau + h^2)$, $E^0 = 0$ and the bound is otherwise similar and simpler than that for $k \ge 2$.

6.2 Piecewise smooth solutions

The previous arguments can also be used to derive error estimates for piecewise smooth solutions. In this case, we will need to limit the number of non-regular points that may appear in the exact solution (assumption (A6)(i) is similar to [5]).

ASSUMPTION (A6). There exists an integer $p \geq 1$ and functions $t \to (x_j^*(t))_{1\leq j\leq p}$ for $t \in [0, T]$, such that, with $\Omega_T^* := (\Omega \times (0, T))\backslash \bigcup_{1\leq j\leq p}\{(t, x_j^*(t)),\ t \in (0, T)\}$, the following holds:

(i) $v \in C_b^{3,4}(\Omega_T^*)$;
(ii) $\forall j, t \to x_j^*(t)$ is Lipschitz regular.

We give the following straightforward preliminary result without proof:

Lemma 18 *Assume (A6) and the CFL condition (11). Then for all t*

$$Card\{j,\ x \to v(t, x) \text{ not regular in } [x_{j-2}, x_{j+2}]\} \leq 5p$$

and

$$Card\{j,\ \theta \to v(\theta, x_j) \text{ not regular in } [t - 2\tau, t]\} \leq Cp$$

for some constant $C \geq 0$ independent of τ, h ("not regular" meaning not C^4 in the first case and not C^3 in the second one).

Such a situation will be illustrated in the numerical example of Sect. 7.2.

Theorem 19 *We assume (A1), (A2), (A3) and the CFL condition (11). Let (A4) for some $\delta \in (0, 1]$ and (A6) hold, then the numerical solution u of (3), (6) converges to v in the L^2-norm with*

$$\max_{2\leq k\leq N} |v^k - u^k|_0 \leq Ch^{1/2+\delta},$$

where C is a constant independent of h.

Proof Let $\mathbb{I}^k$ be the (finite) set of indices i such that v is not regular in $\{t_k\} \times (x_i - 2h, x_i + 2h) \cup (t_k - 2\tau, t_k) \times \{x_i\}$. Then

$$\begin{aligned}|\mathcal{E}^k|_0^2 &= \sum_{i\in\mathbb{I}} |\mathcal{E}_i^k|^2 h = \sum_{i\in\mathbb{I}^k} |\mathcal{E}_i^k|^2 h + \sum_{i\in\mathbb{I}\backslash\mathbb{I}^k} |\mathcal{E}_i^k|^2 h \\ &\leq C|\mathbb{I}^k|(\tau^\delta + h^\delta)^2 h + C(\tau^2 + h^2)^2.\end{aligned}$$

We then use the fact that $|\mathbb{I}^k| \leq C$ for some (different) constant C by Lemma 18 and that $(\tau^2 + h^2)^2 = O(h^4) = O(h^{2+\delta})$, $\tau^\delta + h^\delta = O(h^\delta)$ by the CFL condition (11), in order to obtain the desired result. □

Remark 20 (i) Similar results can be derived for errors in the A-norm, however derivatives of one order higher are required due to the derivative in the definition of the norm.

(ii) The estimates in Theorem 7 are not always sharp, as symmetries and the smoothing behaviour of the scheme can result in higher order convergence. We discuss such special cases for Examples 1 and 2 in Sect. 7, Remarks 22 and 23, respectively.

(iii) These error estimates can be compared with [5], where an error bound of order $h^{1/2}$ was obtained for diffusion problems with an obstacle term, under the main assumption that v_{xx} is a.e. bounded with a finite number of singularities (instead of (A4)) . In the present context it seems natural to assume the Hölder regularity of u_t and u_{xx} coming from the ellipticity assumption (see Remark 6).

7 Numerical tests

We now compare the performance of the BDF2 scheme with other second order finite difference schemes on two examples.

7.1 Test 1: Eikonal equation

The first example is based on a deterministic control problem ($\sigma \equiv 0$) and motivates the choice of the BDF2 approximation for the drift term in (5), compared to the more classical centered scheme (8). We consider

$$\begin{cases} v_t + |v_x| = 0, & x \in (-2, 2),\ t \in (0, T), \\ v(0, x) = v_0(x), & x \in (-2, 2), \end{cases}$$

with $v_0(x) = \max(0, 1 - x^2)^4$ and $T = 0.2$. The initial datum is shown in Fig. 1 (dashed line). The exact solution is

$$v(t, x) = \min(v_0(x - t), v_0(x + t)).$$

Remark 21 The Eikonal equation can be written as $v_t + \max_{a\in\{-1,1\}}(a v_x) = 0$ in HJB form. Note that our theoretical analysis does not cover this example, however, since in the degenerate case assumption (A5) is required, which is not satisfied here.

In Fig. 1, we show the results obtained at the terminal time $T = 0.2$ using schemes (3) with (8) (left) and (3) with(5) (right) with $\tau/h = 0.5$. We numerically observe that the centered approximation generates undesirable oscillations, whereas the BDF2 scheme preserves the total variation.

As stated in Theorem 3, in the case of a degenerate diffusion, a CFL condition of the form $\tau \le Ch$ has to be satisfied for well-posedness of the BDF2 scheme. Table 2 shows numerical convergence of order 2 in both time and space, although the solution is globally only Lipschitz.

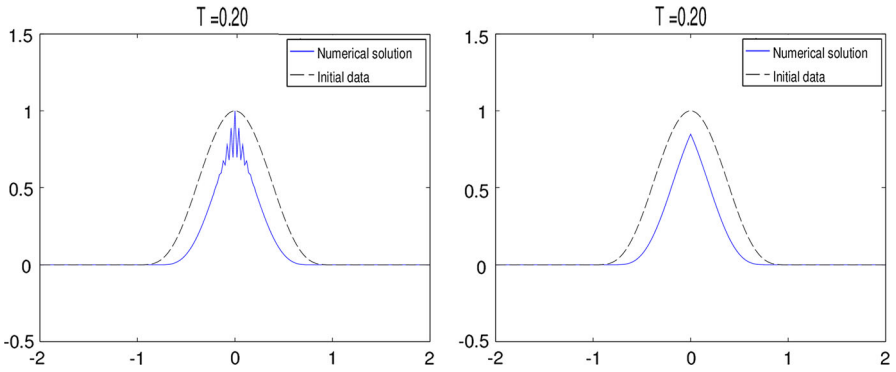

Fig. 1 Test 1: Initial data (dashed line) and numerical solution at time $T = 0.2$ computed for $I + 1 = 200$ and $N = 20$ ($\tau/h = 0.5$) using BDF in time and centered approximation of the drift (left), BDF in time and space (right)

Table 2 Test 1. Error and convergence rate to the exact solution for the BDF2 scheme with $\tau/h = 0.1$ and initial data $v_0(x) = \max(0, 1 - x^2)^4$

N	$I+1$	H^1 norm		L^2-norm		L^∞ norm		CPU (s)
		Error	Order	Error	Order	Error	Order	
5	10	5.35E-01	–	1.25E-01	–	1.36E-01	–	0.094
10	20	2.42E-01	1.14	4.51E-02	1.47	6.83E-02	0.99	0.096
20	40	8.25E-02	1.55	1.55E-02	1.55	2.01E-02	1.77	0.126
40	80	2.38E-02	1.80	4.32E-03	1.84	5.23E-03	1.94	0.147
80	160	6.26E-03	1.92	1.11E-03	1.96	1.31E-03	2.00	0.194
160	320	1.61E-03	1.96	2.79E-04	1.99	3.24E-04	2.01	0.335
320	640	4.09E-04	1.98	7.10E-05	1.99	8.19E-05	2.00	0.759
640	1280	1.03E-04	1.99	1.78E-05	2.00	2.05E-05	2.00	2.306

Remark 22 The full convergence order here is due to the particular symmetry of the solution. To confirm this, we report in Table 3 the results obtained for the same equation with initial data

$$v(0, x) = -\max(0, 1 - x^2)^4$$

(see also Fig. 2). In this case, there is no such symmetry around the two singular points and as a result the full convergence order is lost numerically: the scheme is globally only of order 1 in the H^1 norm and roughly 1.5 in the L^2 and L^∞ norms.

Table 3 Test 1. Error and convergence rate to the exact solution for the BDF2 scheme with $\tau/h = 0.1$ and initial data $v_0(x) = -\max(0, 1-x^2)^4$

N	$I+1$	H^1 norm		L^2 norm		L^∞ norm		CPU (s)
		Error	Order	Error	Order	Error	Order	
5	10	5.84E-01	–	1.62E-01	–	1.51E-01	–	0.006
10	20	2.69E-01	1.12	5.23E-02	1.63	6.20E-02	1.28	0.008
20	40	1.45E-01	0.89	1.86E-02	1.49	2.08E-02	1.58	0.018
40	80	6.74E-02	1.10	5.95E-03	1.64	7.89E-03	1.40	0.039
80	160	3.20E-02	1.08	1.81E-03	1.72	3.57E-03	1.15	0.093
160	320	1.60E-02	1.00	5.44E-04	1.73	1.51E-03	1.24	0.233
320	640	8.16E-03	0.97	1.65E-04	1.72	6.33E-04	1.25	0.695
640	1280	4.20E-03	0.96	5.09E-05	1.70	2.64E-04	1.26	2.163

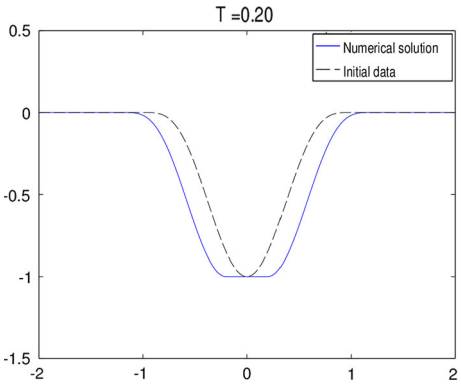

Fig. 2 Test 1: Initial data (dashed line) $v_0(x) = -\max(0, 1-x^2)^4$ and numerical solution at time $T = 0.2$ computed for $I+1 = 200$ and $N = 20$ ($\tau/h = 0.5$) using the BDF2 scheme. The convergence rates for this example are reported in Table 3

7.2 Test 2: A simple controlled diffusion model equation

The second test we propose is a problem with controlled diffusion. We consider

$$\begin{cases} v_t + \sup_{\sigma\in\{\sigma_1,\sigma_2\}} \left(-\frac{1}{2}\sigma^2 v_{xx} \right) = 0, & x \in (-1,1), t \in (0,T), \\ v(0,x) = \sin(\pi x), & x \in (-1,1), \end{cases}$$

with parameters $\sigma_1 = 0.1$, $\sigma_2 = 0.5$, $T = 0.5$.

In spite of the apparent simplicity of the equation under consideration, in [19] an example of non-convergence of the Crank-Nicolson scheme is given for a similar optimal control problem. The BDF2 scheme, in contrast, has shown good performance for that same problem in [6].

Figure 3 (top row) shows the initial data and the value function at terminal time computed using the BDF2 scheme. The error and convergence rate in different norms are reported in Table 4. Here an accurate numerical solution computed by an implicit Euler scheme (which is monotone and hence guaranteed to converge) is used for comparison.

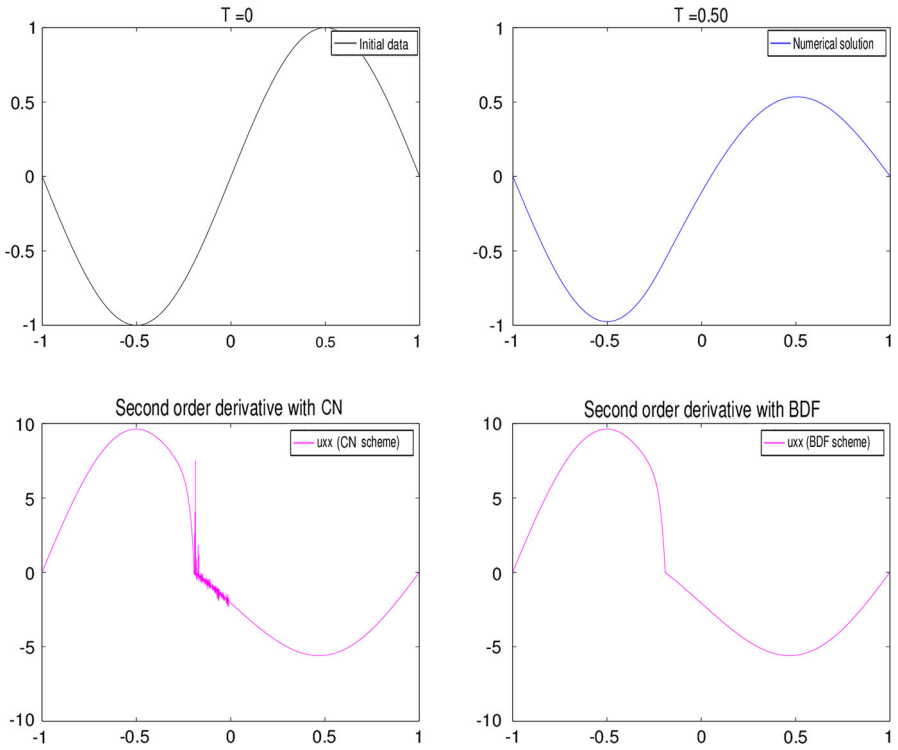

Fig. 3 Test 2: Initial data (top, left), numerical solution at time $T = 0.5$ (top, right) computed by the BDF2 scheme, second order derivative computed with CN scheme (bottom, left) and BDF2 (bottom, right) for $N = 256$ and $I + 1 = 5120$

Table 4 Test 2. Error and convergence rate for the BDF2 scheme with high CFL number $\tau = 5h$. A reference solution computed by the implicit Euler scheme (6) with $I + 1 = 20 \times 2^9$, $N = 2^{22}$ is used

N	$I+1$	H^1 norm		L^2 norm		L^∞ norm		CPU (s)
		Error	Order	Error	Order	Error	Order	
1	20	1.54E-01	–	5.11E-02	–	7.24E-02	–	0.131
2	40	5.53E-02	1.48	1.88E-02	1.45	2.63E-02	1.46	0.112
4	80	1.47E-02	1.91	5.17E-03	1.86	6.99E-03	1.91	0.111
8	160	3.59E-03	2.04	1.27E-03	2.03	1.66E-03	2.08	0.122
16	320	8.98E-04	2.00	3.14E-04	2.02	4.09E-04	2.02	0.146
32	640	2.26E-04	1.99	7.84E-05	2.00	1.02E-04	2.00	0.183
64	1280	5.65E-05	2.00	1.96E-05	2.00	2.56E-05	2.00	0.267
128	2560	1.42E-05	2.00	4.90E-06	2.00	6.42E-06	2.00	0.598
256	5120	1.21E-06	2.01	1.21E-06	2.01	1.59E-06	2.01	1.879

Table 5 Test 2. Error and convergence rate for the CN scheme with high CFL number $\tau = 5h$. A reference solution computed by the implicit Euler scheme (6) with $I + 1 = 20 \times 2^9$, $N = 2^{22}$ is used

N	$I+1$	H^1 norm		L^2 norm		L^∞ norm		CPU (s)
		Error	Order	Error	Order	Error	Order	
1	20	4.11E-02	–	7.01E-03	–	9.44E-03	–	0.149
2	40	7.82E-03	2.39	1.45E-03	2.27	2.29E-03	2.04	0.113
4	80	1.97E-03	1.99	3.87E-04	1.91	5.62E-04	2.03	0.111
8	160	5.16E-04	1.94	1.02E-04	1.92	1.45E-04	1.95	0.128
16	320	1.09E-04	2.24	2.67E-05	1.94	3.77E-05	1.95	0.166
32	640	2.96E-05	1.88	7.15E-06	1.90	9.87E-06	1.93	0.188
64	1280	7.64E-06	1.96	2.03E-06	1.82	2.61E-06	1.92	0.310
128	2560	9.50E-05	−3.64	1.98E-05	−3.29	3.49E-05	−3.74	0.992
256	5120	7.18E-04	−2.92	8.40E-05	−2.08	1.62E-04	−2.22	4.251

Taking $\tau \sim h$ the BDF2 scheme gives clear second order convergence, as seen in Table 4. This is not the case for CN as shown in Table 5. The CN scheme also exhibits some instability in the second order derivative for high CFL number, i.e. τ/h, see Fig. 3 (this is analogous to the finding in [19]). One can verify that for a small CFL number, i.e. $\tau \sim h^2$, the CN scheme shows convergence of second order.

Remark 23 In this example, due to the strict ellipticity, Assumption (A4) is guaranteed for some $\delta > 0$ (see Remark 6). Then Theorem 7 gives convergence with order δ. Furthermore, Fig. 3, bottom row, suggests Hölder continuity of u_{xx} in x, which is expected by virtue of the control being piecewise constant. Therefore, we conjecture that Assumption (A6) is satisfied, such that Theorem 19 would give the higher order $1/2 + \delta$. In the test, in fact the full order 2 is observed (see Table 4).

8 Conclusions

We have proved the well-posedness and stability in L^2 and H^1 norms of a second order BDF scheme for HJB equations with enough regularity of the coefficients. The significance of the results is that this was achieved for a second order (and hence) non-monotone scheme.

One can use the recursion we derived to bound the error of the numerical solution in terms of the truncation error of the scheme. The latter depends on the regularity of the solution and has to be estimated for individual examples. A full analysis was carried out for the semi-linear, uniformly parabolic case.

The numerical tests demonstrate convergence at least as good as predicted by the theoretical results, and often better, due to symmetries of the solution or smoothing properties of the equation and the scheme. This is in contrast to some alternative second order schemes, such as the central spatial difference in the case of a first order

equation, or the Crank-Nicolson time stepping scheme for a second order equation, which can show poor or no convergence.

Funding Open access funding provided by Universitá degli Studi di Verona within the CRUI-CARE Agreement.

Appendix A. Proof of Lemma 8

In order to prove the existence and uniqueness of a solution to (22), we consider a fixed-point approach. The initial problem (22) can be written as follows:

$$\sup_{a\in\Lambda}(L_a X - (q_a - U_a X)) = 0, \tag{68}$$

where L_a and U_a are two matrices such that $M_a \equiv L_a + U_a$. We consider in particular L_a to be the lower triangular part of M_a including the diagonal terms, $(L_a)_{ij} := (M_a)_{ij} 1_{i\geq j}$, and U_a the remaining upper triangular part, $(U_a)_{ij} := (M_a)_{ij} 1_{i<j}$.

For a given vector $c \in \mathbb{R}^I$, let $g(c) := X$ denote the (unique) solution of the following simplified problem:

$$\sup_{a\in\Lambda}(L_a X - (q_a - U_a c)) = 0. \tag{69}$$

Indeed, because $(L_a)_{ii} = (M_a)_{ii} > 0$, denoting $v_a := q_a - U_a c$, it is easy to see by recursion in i that the unique solution of

$$\sup_{a\in\Lambda}(L_a X - v_a) = 0$$

is given by

$$x_i := \inf_{a\in\Lambda}\left(\left((v_a)_i - \sum_{k=1}^{i-1}(L_a)_{ik}x_k\right)/(L_a)_{ii}\right).$$

Therefore, solving (68) amounts to solving $g(X) = X$. By elementary compuations one can show that g is δ-Lipschitz for the $\|.\|_\infty$ norm, with $\delta := \sup_a \|(L_a)^{-1}U_a\|_\infty$.

For a diagonally dominant matrix, the following classical estimate holds

$$\|(L_a)^{-1}U_a\|_\infty \leq \sup_{i\in\mathbb{I}} \frac{\sum_{j>i} |(M_a)_{ij}|}{|(M_a)_{ii}| - \sum_{j<i} |(M_a)_{ij}|}$$

(this is related to the Gauss–Seidel relaxation method; see for instance, Th. 8.2.12 in [21]). By using the assumptions on the matrices M_a, we have $\delta < 1$. Hence, g is a contraction mapping on $\mathbb{R}^I$ and therefore we obtain the existence and uniqueness of a solution of (68) as desired. □

References

1. Barles, G., Jakobsen, E.R.: Error bounds for monotone approximation schemes for parabolic Hamilton–Jacobi–Bellman equations. Math. Comput. **74**(260), 1861–1893 (2007)
2. Barles, G., Souganidis, P.E.: Convergence of approximation schemes for fully nonlinear second order equations. Asymptotic Anal. **4**, 271–283 (1991)
3. Beale, T.: Smoothing properties of implicit finite difference methods for a diffusion equation in maximum norm. SIAM J. Numer. Anal. **47**(4), 2476–2495 (2009)
4. Becker, J.: A second order backward difference method with variable steps for a parabolic problem. BIT Numer. Math. **38**(4), 644–662 (1998)
5. Bokanowski, O., Debrabant, K.: Backward differentiation formula finite difference schemes for diffusion equations with an obstacle term. IMA J Numer. Anal. **41**(2), 900–934 (2021). https://doi.org/10.1093/imanum/draa014
6. Bokanowski, O., Picarelli, A., Reisinger, C.: High-order filtered schemes for time-dependent second order HJB equations. ESAIM Math. Model. Numer. Anal. **54**(1), 69–97 (2018)
7. Bokanowski, Olivier, Falcone, Maurizio, Ferretti, Roberto, Grüne, Lars, Kalise, Dante, Zidani, Hasnaa: Value iteration convergence of ϵ-monotone schemes for stationary Hamilton–Jacobi equations. Discrete Contin. Dynam. Syst. A **35**(9), 4041–4070 (2015)
8. Bonnans, J.F., Zidani, H.: Consistency of generalized finite difference schemes for the stochastic HJB equation. SIAM J. Numer. Anal. **41**(3), 1008–1021 (2003)
9. Crandall, M.G., Ishii, H., Lions, P.L.: User's guide to viscosity solutions of second order partial differential equations. Bull. Amer. Math. Soc. **27**(1), 1–67 (1992)
10. Emmrich, E.: Stability and error of the variable two-step BDF for semilinear parabolic problems. J. Appl. Math. Comput. **19**(1–2), 33–55 (2005)
11. Evans, L.C., Lenhart, S.: The parabolic Bellman equation. Nonlinear Anal. **5**(7), 765–773 (1981)
12. Godunov, S.K.: A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. Matematicheskii Sbornik **89**(3), 271–306 (1959)
13. Hackbusch, W.: Elliptic Differential Equations: Theory and Numerical Treatment, Springer Series in Computational Mathematics, vol. 18. Springer, Berlin (2010)
14. Hill, A., Süli, E.: Approximation of the global attractor for the incompressible Navier–Stokes equations. IMA J. Numer. Anal. **20**(4), 633–667 (2000)
15. Jensen, M.: $L^2(H^1_\gamma)$ finite element convergence for degenerate isotropic Hamilton–Jacobi–Bellman equations. IMA J. Numer. Anal. **37**(3), 1300–1316 (2017)
16. Jensen, M., Smears, I.: On the convergence of finite element methods for Hamilton–Jacobi–Bellman equations. SIAM J. Numer. Anal. **51**(1), 137–162 (2013)
17. Krylov, N.V.: Boundedly nonhomogeneous elliptic and parabolic equations. Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya **46**(3), 487–523 (1982)
18. Picarelli, A., Reisinger, C., Rotaetxe, J.: Some regularity and convergence results for parabolic Hamilton–Jacobi–Bellman equations in bounded domains. J. Differ. Equ. **268**(12), 7843–7876 (2020)
19. Pooley, D.M., Forsyth, P.A., Vetzal, K.R.: Numerical convergence properties of option pricing pdes with uncertain volatility. IMA J. Numer. Anal. **23**(2), 241–267 (2003)
20. Smears, I., Süli, E.: Discontinuous Galerkin finite element methods for time-dependent Hamilton–Jacobi–Bellman equations with Cordes coefficients. Numer. Math. **133**(1), 141–176 (2016)

21. Stoer, J., Bulirsch, R.: Introduction to Numerical Analysis, Texts in Applied Mathematics, vol. 12, 3rd edn, p. xvi+744. Springer-Verlag, New York, (1993). Translated from the German by R. Bartels, W. Gautschi and C. Witzgall. https://doi.org/10.1007/978-0-387-21738-3
22. Süli, E., Mayers, D.F.: An Introduction to Numerical Analysis. Cambridge University Press, Cambridge (2003)