

# SEMI-LAGRANGIAN SCHEMES FOR LINEAR AND FULLY NON-LINEAR DIFFUSION EQUATIONS

KRISTIAN DEBRABANT AND ESPEN R. JAKOBSEN

ABSTRACT. For linear and fully non-linear diffusion equations of Bellman-Isaacs type, we introduce a class of approximation schemes based on differencing and interpolation. As opposed to classical numerical methods, these schemes work for general diffusions with coefficient matrices that may be non-diagonal dominant and arbitrarily degenerate. In general such schemes have to have a wide stencil. Besides providing a unifying framework for several known first order accurate schemes, our class of schemes includes new first and higher order versions. The methods are easy to implement and more efficient than some other known schemes. We prove consistency and stability of the methods, and for the monotone first order methods, we prove convergence in the general case and robust error estimates in the convex case. The methods are extensively tested.

## 1. INTRODUCTION

In this paper we introduce and analyze a class of approximation schemes for fully non-linear diffusion equations of Bellman-Isaacs type,

$$(1.1) \quad u_t - \inf_{\alpha \in \mathcal{A}} \sup_{\beta \in \mathcal{B}} \left\{ L^{\alpha, \beta}[u](t, x) + c^{\alpha, \beta}(t, x)u + f^{\alpha, \beta}(t, x) \right\} = 0 \quad \text{in } Q_T,$$

$$(1.2) \quad u(0, x) = g(x) \quad \text{in } \mathbb{R}^N,$$

where  $Q_T := (0, T] \times \mathbb{R}^N$ ,  $\mathcal{A}$  and  $\mathcal{B}$  are complete metric spaces, and

$$L^{\alpha, \beta}[u](t, x) = \text{tr}[a^{\alpha, \beta}(t, x)D^2u(t, x)] + b^{\alpha, \beta}(t, x)Du(t, x).$$

The coefficients  $a^{\alpha, \beta} = \frac{1}{2}\sigma^{\alpha, \beta}\sigma^{\alpha, \beta \top}$ ,  $b^{\alpha, \beta}$ ,  $c^{\alpha, \beta}$ ,  $f^{\alpha, \beta}$  and the initial data  $g$  take values respectively in  $\mathbb{S}^N$ , the space of  $N \times N$  symmetric matrices,  $\mathbb{R}^N$ ,  $\mathbb{R}$ ,  $\mathbb{R}$ , and  $\mathbb{R}$ . We will only assume that  $a^{\alpha, \beta}$  is positive semi-definite, thus the equation is allowed to degenerate and hence not have smooth solutions in general. Under suitable assumptions (see Section 2), the initial value problem (1.1)-(1.2) has a unique, bounded, Hölder continuous, viscosity solution  $u$ . This function is the upper or lower value of a stochastic differential game, or, if  $\mathcal{A}$  or  $\mathcal{B}$  is a singleton, the value function of a finite horizon, optimal stochastic control problem [27].

---

2010 *Mathematics Subject Classification.* 65M12, 65M15, 65M06, 35K10, 35K55, 35K65, 49L25, 49L20.

*Key words and phrases.* Monotone approximation schemes, difference-interpolation methods, stability, convergence, error bound, degenerate parabolic equations, Hamilton-Jacobi-Bellman equations, viscosity solution.

Jakobsen was supported by the Research Council of Norway through the project “Integro-PDEs: Numerical methods, Analysis, and Applications to Finance”.

We introduce a family of schemes that we call Semi-Lagrangian (SL) schemes. They are a type difference-interpolation schemes and arise as time-discretizations of the following semi-discrete equation

$$u_t - \inf_{\alpha \in \mathcal{A}} \sup_{\beta \in \mathcal{B}} \left\{ L_k^{\alpha, \beta} [\mathcal{I}_{\Delta x} u](t, x) + c^{\alpha, \beta}(t, x)u + f^{\alpha, \beta}(t, x) \right\} = 0 \quad \text{in } X_{\Delta x} \times (0, T),$$

where  $L_k^{\alpha, \beta}$  is a monotone difference approximation of  $L^{\alpha, \beta}$  and  $\mathcal{I}_{\Delta x}$  is an interpolation operator on the spatial grid  $X_{\Delta x}$ . For more details see Section 3. Typically these schemes are first order accurate wide-stencil schemes, and if the matrix  $a^{\alpha, \beta}$  is bad enough, the stencil has to keep increasing as the grid is refined to have convergence. They include as special cases monotone schemes from [16, 20, 7, 11], new versions of these schemes, and a new non-monotone spatially second order accurate compact stencil scheme. There are three main advantages of these schemes: (i) they are easy to understand and implement, (ii) they are faster than some alternative methods, and (iii) they are consistent and, if  $\mathcal{I}_{\Delta x}$  is monotone, monotone for every positive semi-definite diffusion matrix  $a^{\alpha, \beta} = \frac{1}{2} \sigma^{\alpha, \beta} \sigma^{\alpha, \beta \top}$ . The last point is important because monotone methods are known to converge to the correct solution [3], while non-monotone methods need not converge [23] or can even converge to false solutions [25].

Classical finite difference approximations (FDMs) of (1.1) are not monotone (of positive type) unless the matrix  $a^{\alpha, \beta}$  satisfies additional assumptions like e. g. being diagonally dominant [19]. More general assumptions are given in e. g. [6, 14] but at the cost of increased stencil length. In fact, Dong and Krylov [14] proved that *no fixed stencil FDM* can approximate equations with a second derivative term involving a general positive semi-definite matrix function  $a^{\alpha, \beta}$ . Note that this type of result has been known for a long time, see e. g. [21, 11]. Some very simple examples of such “bad” matrices are given by

$$\left( \begin{array}{cc} x_1^2 & \frac{1}{2}x_1x_2 \\ \frac{1}{2}x_1x_2 & x_2^2 \end{array} \right) \text{ in } [0, 1]^2, \quad \left( \begin{array}{cc} \alpha^2 & \alpha\beta \\ \alpha\beta & \beta^2 \end{array} \right) \text{ for } \mathcal{A} = \mathcal{B} = [0, 1], \quad I - \frac{(Du)(Du)^\top}{|Du|^2},$$

and these types of matrices appear in Finance, Stochastic Control Theory, and Mean Curvature Motion. The third example leads to quasi-linear equations and will not be considered here, we refer instead to [11].

To obtain convergent or monotone methods for problems involving non-diagonally dominant matrices, we know of two strategies: (i) The classical method of rotating the coordinate system locally to obtain diagonally dominant matrices  $a^{\alpha, \beta}$ , see e. g. Section 5.4 in [19], and (ii) the use of wide stencil methods. The two solutions seem to be somewhat related, but at least the defining ideas and implementation are different. Both ways lead to methods that have reduced order compared to standard schemes for diagonally dominant problems, but the first strategy seems much more difficult to implement.

We mention that the schemes of [22] for stationary Bellman equations have much in common with the schemes in this paper. However the two types of schemes and their derivation and error analysis are different in general. Other related wide-stencil schemes are the method of [6], which is not of SL type, and various SL like schemes for other types of equations, e. g. the Mean Curvature Motion equation [11], Monge-Ampère equations [24], and non-local Bellman equations [8]. The terminology SL schemes is already used for schemes for transport equations, conservation laws,

and first order Hamilton-Jacobi equations. In the Hamilton-Jacobi setting, these schemes go back to the 1983 paper [9] of Capuzzo-Dolcetta.

The rest of this paper is organized as follows. In the next section we explain the notation and state a well-posedness and regularity result for (1.1)–(1.2). The SL schemes are motivated and defined in an abstract setting in Section 3, and in Section 4 we prove that they are consistent,  $L^\infty$ -stable, and, if  $\mathcal{I}_{\Delta x}$  is monotone, also monotone and convergent. We provide several examples of SL schemes in Section 5, including the linear interpolation SL (LISL) scheme. This is the basic example of this paper, a first order accurate wide stencil scheme that can be defined on unstructured grids. Higher order interpolation is not monotone. But for essentially monotone solutions, we use third order monotonicity preserving cubic Hermite interpolation [17, 15] to obtain new schemes called monotonicity preserving cubic interpolation SL (MPCSL) schemes in Section 5.3. These compact stencil schemes are second order accurate in space and first or second order accurate in time.

In Section 6 we discuss various issues concerning the SL schemes. We compare the LISL scheme to the scheme of Bonnans-Zidani [6] and find that the LISL scheme is easier to understand and implement and is faster in general. We also explain that the SL schemes can be interpreted as collocation methods for derivative free equations, and as dynamic programming equations of discrete stochastic differential games or optimal control problems.

In Sections 7, 8 and Appendix B, we derive robust error estimates for monotone schemes for convex equations, i. e. when  $\mathcal{I}_{\Delta x}$  is monotone and  $\mathcal{B}$  is a singleton in (1.1). They are obtained through the regularization method of Krylov [18] and apply to degenerate equations and non-smooth solutions. Finally, in Section 9 our methods are extensively tested. In particular we find a first indication that the LISL and MPCSL schemes yield much faster methods to solve the finance problem of [4].

## 2. NOTATION AND WELL-POSEDNESS

In this section we introduce notation and assumptions, and give a well-posedness and comparison result for the initial value problem (1.1) – (1.2).

We denote by  $\leq$  the component by component ordering in  $\mathbb{R}^M$  and the ordering in the sense of positive semi-definite matrices in  $\mathbb{S}^N$ . The symbols  $\wedge$  and  $\vee$  denote the minimum respectively the maximum. By  $|\cdot|$  we mean the Euclidean vector norm in any  $\mathbb{R}^p$  type space, e. g. if  $X \in \mathbb{R}^{N \times P}$  (an  $N \times P$  matrix), then  $|X|^2 = \sum_{i,j} |X_{ij}|^2 = \text{tr}(XX^\top)$  where  $X^\top$  is the transpose of  $X$ . If  $w$  is a bounded function from some set  $Q' \subset \overline{Q}_\infty$  into  $\mathbb{R}$ ,  $\mathbb{R}^M$ , or  $\mathbb{R}^{N \times P}$ , we set

$$|w|_0 = \sup_{(t,y) \in Q'} |w(t,y)|, \quad [w]_\delta = \sup_{(t,x) \neq (s,y)} \frac{|w(t,x) - w(s,y)|}{(|x-y| + |t-s|^{1/2})^\delta},$$

and  $|w|_\delta = |w|_0 + [w]_\delta$  for any  $\delta \in (0, 1]$ . Let  $C_b(Q')$  and  $\mathcal{C}^{0,\delta}(Q')$ ,  $\delta \in (0, 1]$ , denote respectively the space of bounded continuous functions on  $Q'$  and the subset of  $C_b(Q')$  in which the norm  $|\cdot|_\delta$  is finite. Typically  $Q' = Q_T$  or  $Q' = \mathbb{R}^N$ , and we will always suppress the domain  $Q'$  when writing norms.

For simplicity, we will use the following assumptions on the data of (1.1)–(1.2):

- (A1) For any  $\alpha \in \mathcal{A}$  and  $\beta \in \mathcal{B}$ ,  $a^{\alpha,\beta} = \frac{1}{2} \sigma^{\alpha,\beta} \sigma^{\alpha,\beta \top}$  for some  $N \times P$  matrix  $\sigma^{\alpha,\beta}$ . Moreover, there is a constant  $K$  independent of  $\alpha, \beta$  such that

$$|g|_1 + |\sigma^{\alpha,\beta}|_1 + |b^{\alpha,\beta}|_1 + |c^{\alpha,\beta}|_1 + |f^{\alpha,\beta}|_1 \leq K.$$

These assumptions are standard and ensure comparison and well-posedness of (1.1)–(1.2) in the class of bounded  $x$ -Lipschitz functions.

**Proposition 2.1.** *Assume that (A1) holds. Then there exist a unique solution  $u$  of (1.1)–(1.2) and a constant  $C$  only depending on  $T$  and  $K$  from (A1) such that*

$$|u|_1 \leq C.$$

*Furthermore, if  $u_1$  and  $u_2$  are sub- and supersolutions of (1.1) satisfying  $u_1(0, \cdot) \leq u_2(0, \cdot)$ , then  $u_1 \leq u_2$ .*

The proof is standard. Assumption (A1) can be relaxed in many ways, e. g. using weighted norms, Hölder or uniform continuity, etc. But in doing so, solutions can become unbounded and less smooth, and the analysis becomes harder and less transparent. Therefore we will not consider such extensions in this paper.

By solutions in this paper we always mean viscosity solutions, see e. g. [10, 27].

### 3. DEFINITION OF SL SCHEMES

In this section we propose a class of approximation schemes for (1.1)–(1.2) which we call Semi-Lagrangian or SL schemes. This class includes (parabolic versions of) the “control schemes” of Menaldi [20] and Camilli and Falcone [7] and the monotone schemes of Crandall and Lions [11]. It also includes SL schemes for first order Bellman equations [9, 16], and some new versions as discussed in Section 5. For a motivation for the name, see Remark 6.2. For the time discretization we use a new generalized mid-point rule that includes explicit, implicit, and a second order Crank-Nicolson like approximations. Since the equation is non-smooth in general, the usual way of defining a Crank-Nicolson scheme [12] only gives a first order accurate scheme in time.

The schemes are defined on a possibly unstructured family of grids  $\{G_{\Delta t, \Delta x}\}$ ,

$$G = G_{\Delta t, \Delta x} = \{(t_n, x_i)\}_{n \in \mathbb{N}_0, i \in \mathbb{N}} = \{t_n\}_{n \in \mathbb{N}_0} \times X_{\Delta x},$$

for  $\Delta t, \Delta x > 0$ . Here  $0 = t_0 < t_1 < \dots < t_n < t_{n+1}$  satisfy

$$\max_n \Delta t_n \leq \Delta t \quad \text{where} \quad \Delta t_n = t_n - t_{n-1},$$

and  $X_{\Delta x} = \{x_i\}_{i \in \mathbb{N}}$  is the set of vertices or nodes for a non-degenerate polyhedral subdivision  $\mathcal{T}^{\Delta x} = \{T_j^{\Delta x}\}_{j \in \mathbb{N}}$  of  $\mathbb{R}^N$ , i. e., the polyhedrons  $T_j^{\Delta x}$  satisfy

$$\text{int}(T_j^{\Delta x} \cap T_i^{\Delta x}) \stackrel{i \neq j}{=} \emptyset, \quad \bigcup_{j \in \mathbb{N}} T_j^{\Delta x} = \mathbb{R}^N,$$

$$\rho \Delta x \leq \sup_{j \in \mathbb{N}} \{\text{diam } B_{T_j^{\Delta x}}\} \leq \sup_{j \in \mathbb{N}} \{\text{diam } T_j^{\Delta x}\} \leq \Delta x$$

for some  $\rho \in (0, 1)$ , where  $\text{diam}$  is the diameter of the set and  $B_{T_j^{\Delta x}}$  is the greatest ball contained in  $T_j^{\Delta x}$ .

To motivate the numerical schemes, we write  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_m, \dots, \sigma_P)$  where  $\sigma_m$  is the  $m$ -th column of  $\sigma$  and observe that for  $k > 0$  and smooth functions  $\phi$ ,

$$\begin{aligned} \frac{1}{2} \text{tr}[\sigma \sigma^\top D^2 \phi(x)] &= \frac{1}{2} \sum_{m=1}^P \text{tr}[\sigma_m \sigma_m^\top D^2 \phi(x)] \\ &= \sum_{m=1}^P \frac{1}{2} \frac{\phi(x + k\sigma_m) - 2\phi(x) + \phi(x - k\sigma_m)}{k^2} + \mathcal{O}(k^2), \end{aligned}$$

$$\begin{aligned} bD\phi(x) &= \frac{\phi(x + k^2b) - \phi(x)}{k^2} + \mathcal{O}(k^2) \\ &= \frac{1}{2} \frac{\phi(x + k^2b) - 2\phi(x) + \phi(x + k^2b)}{k^2} + \mathcal{O}(k^2). \end{aligned}$$

These approximations are monotone (of positive type) and the errors are bounded by  $\frac{1}{48}P|\sigma|_0^4|D^4\phi|_0k^2$  and  $\frac{1}{2}|b|_0^2|D^2\phi|_0k^2$  respectively. To relate these approximations to a grid  $G$ , we replace  $\phi$  by its interpolant  $\mathcal{I}_{\Delta x}\phi$  on that grid and obtain

$$\begin{aligned} \frac{1}{2}\text{tr}[\sigma\sigma^\top D^2\phi(x)] &\approx \sum_{m=1}^P \frac{1}{2} \frac{(\mathcal{I}_{\Delta x}\phi)(x + k\sigma_m) - 2(\mathcal{I}_{\Delta x}\phi)(x) + (\mathcal{I}_{\Delta x}\phi)(x - k\sigma_m)}{k^2}, \\ bD\phi(x) &\approx \frac{1}{2} \frac{(\mathcal{I}_{\Delta x}\phi)(x + k^2b) - 2(\mathcal{I}_{\Delta x}\phi)(x) + (\mathcal{I}_{\Delta x}\phi)(x + k^2b)}{k^2}. \end{aligned}$$

If the interpolation is monotone (positive) then the full discretization is still monotone and represents a typical example of the discretizations we consider below.

To construct the general scheme, we generalize the above construction. Consider general finite difference approximations of the differential operator  $L^{\alpha,\beta}[\phi]$  in (1.1) defined as

$$(3.1) \quad L_k^{\alpha,\beta}[\phi](t, x) := \sum_{i=1}^M \frac{\phi(t, x + y_{k,i}^{\alpha,\beta,+}(t, x)) - 2\phi(t, x) + \phi(t, x + y_{k,i}^{\alpha,\beta,-}(t, x))}{2k^2},$$

for  $k > 0$  and some  $M \geq 1$ , where for all smooth functions  $\phi$ ,

$$(3.2) \quad |L_k^{\alpha,\beta}[\phi] - L^{\alpha,\beta}[\phi]| \leq C(|D\phi|_0 + \dots + |D^4\phi|_0)k^2.$$

This approximation and interpolation yield a semi-discrete approximation of (1.1),

$$U_t - \inf_{\alpha \in \mathcal{A}} \sup_{\beta \in \mathcal{B}} \left\{ L_k^{\alpha,\beta}[\mathcal{I}_{\Delta x}U](t, x) + c^{\alpha,\beta}(t, x)U + f^{\alpha,\beta}(t, x) \right\} = 0 \quad \text{in } (0, T) \times X_{\Delta x},$$

and the final scheme can then be found after discretizing in time using a parameter  $\theta \in [0, 1]$ ,

$$(3.3) \quad \delta_{\Delta t_n} U_i^n = \inf_{\alpha \in \mathcal{A}} \sup_{\beta \in \mathcal{B}} \left\{ L_k^{\alpha,\beta}[\overline{\mathcal{I}_{\Delta x}U}^{\theta,n}]_i^{n-1+\theta} + c_i^{\alpha,\beta,n-1+\theta} \overline{U}_i^{\theta,n} + f_i^{\alpha,\beta,n-1+\theta} \right\}$$

in  $G$ , where  $U_i^n = U(t_n, x_i)$ ,  $f_i^{\alpha,\beta,n-1+\theta} = f^{\alpha,\beta}(t_{n-1} + \theta\Delta t_n, x_i)$ ,  $\dots$  for  $(t_n, x_i) \in G$ ,

$$\begin{aligned} \delta_{\Delta t}\phi(t, x) &= \frac{\phi(t, x) - \phi(t - \Delta t, x)}{\Delta t}, \quad \text{and} \quad \overline{\phi}^{\theta,n} = (1 - \theta)\phi^{n-1} + \theta\phi^n, \\ \overline{\mathcal{I}_{\Delta x}\phi}^{\theta,n} &= (1 - \theta)\mathcal{I}_{\Delta x}\phi^{n-1} + \theta\mathcal{I}_{\Delta x}\phi^n. \end{aligned}$$

As initial conditions we take

$$(3.4) \quad U_i^0 = g(x_i) \quad \text{in } X_{\Delta x}.$$

*Remark 3.1.* For the choices  $\theta = 0, 1$ , and  $1/2$  the time discretization corresponds to respectively explicit Euler, implicit Euler and midpoint rule. For  $\theta = 1/2$ , the full scheme can be seen as generalized Crank-Nicolson type discretization.

## 4. ANALYSIS OF SL SCHEMES

In this section we prove that the SL scheme (3.3) is consistent and  $L^\infty$ -stable, and in the case when the interpolation (and hence the scheme) is monotone, we present existence, uniqueness, and convergence results for the schemes. Error estimates are given in Section 7 for the monotone convex case.

For the approximation  $L_k^{\alpha,\beta}$  and interpolation  $\mathcal{I}_{\Delta x}$  we assume

$$(Y1) \quad \left\{ \begin{array}{l} \sum_{i=1}^M [y_{k,i}^{\alpha,\beta,+} + y_{k,i}^{\alpha,\beta,-}] = 2k^2 b^{\alpha,\beta} + \mathcal{O}(k^4), \\ \sum_{i=1}^M [y_{k,i}^{\alpha,\beta,+} y_{k,i}^{\alpha,\beta,+ \top} + y_{k,i}^{\alpha,\beta,-} y_{k,i}^{\alpha,\beta,- \top}] = 2k^2 \sigma^{\alpha,\beta} \sigma^{\alpha,\beta \top} + \mathcal{O}(k^4), \\ \sum_{i=1}^M [y_{k,i,j_1}^{\alpha,\beta,+} y_{k,i,j_2}^{\alpha,\beta,+} y_{k,i,j_3}^{\alpha,\beta,+} + y_{k,i,j_1}^{\alpha,\beta,-} y_{k,i,j_2}^{\alpha,\beta,-} y_{k,i,j_3}^{\alpha,\beta,-}] = \mathcal{O}(k^4), \\ \sum_{i=1}^M [y_{k,i,j_1}^{\alpha,\beta,+} y_{k,i,j_2}^{\alpha,\beta,+} y_{k,i,j_3}^{\alpha,\beta,+} y_{k,i,j_4}^{\alpha,\beta,+} + y_{k,i,j_1}^{\alpha,\beta,-} y_{k,i,j_2}^{\alpha,\beta,-} y_{k,i,j_3}^{\alpha,\beta,-} y_{k,i,j_4}^{\alpha,\beta,-}] = \mathcal{O}(k^4), \end{array} \right.$$

for all  $j_1, j_2, j_3, j_4 = 1, 2, \dots, N$  indicating components of the  $y$ -vectors.

(I1) There are  $K \geq 0, r \in \mathbb{N}$  such that for all smooth functions  $\phi$ ,

$$|(\mathcal{I}_{\Delta x} \phi) - \phi|_0 \leq K |D^r \phi|_0 \Delta x^r.$$

(I2) There is a set of non-negative functions  $\{w_j(x)\}_j$  such that

$$(\mathcal{I}_{\Delta x} \phi)(x) = \sum_j \phi(x_j) w_j(x),$$

and for all  $i, j \in \mathbb{N}$ ,

$$w_j(x) \geq 0, \quad w_i(x_j) = \delta_{ij}, \quad \text{and} \quad \sum_i w_i(x) \equiv 1.$$

(I2') Assumption (I2) holds, but  $w_j = w_{\phi,j}$  is allowed to depend on  $\phi$ .

Under assumption (Y1), a Taylor expansion shows that  $L_k^{\alpha,\beta}$  is a second order consistent approximation satisfying (3.2). If we assume also (I1), it then follows that  $L_k^{\alpha,\beta}[\mathcal{I}_{\Delta x} \phi]$  is a consistent approximation of  $L^{\alpha,\beta}[\phi]$  if  $\frac{\Delta x^r}{k^2} \rightarrow 0$ . Indeed

$$|L_k^{\alpha,\beta}[\mathcal{I}_{\Delta x} \phi] - L^{\alpha,\beta}[\phi]| \leq |L_k^{\alpha,\beta}[\mathcal{I}_{\Delta x} \phi] - L_k^{\alpha,\beta}[\phi]| + |L_k^{\alpha,\beta}[\phi] - L^{\alpha,\beta}[\phi]|,$$

where  $|L_k^{\alpha,\beta}[\phi] - L^{\alpha,\beta}[\phi]|$  is estimated in (3.2), and by (I1),

$$|L_k^{\alpha,\beta}[\mathcal{I}_{\Delta x} \phi] - L_k^{\alpha,\beta}[\phi]| \leq C |D^r \phi|_0 \frac{\Delta x^r}{k^2}.$$

*Remark 4.1.* Assumption (Y1) is similar to the local consistency conditions used in [19]. The  $O(k^4)$  terms insure that the method is second order accurate as  $k \rightarrow 0$ . Convergence will still be achieved if we relax  $O(k^4)$  to  $o(k^2)$  as  $k \rightarrow 0$ .

*Remark 4.2.* An interpolation satisfying (I2') is said to be *positive* and preserves positivity of the data. Such an interpolation does not use (exact) derivatives to reconstruct the function  $\phi$ , and it may be a non-monotone and non-linear operator, as in the case of monotonicity preserving cubic interpolation (see Section 5.3).

When (I2) holds and  $w_{\phi,i} = w_i$  is independent of  $\phi$ , the interpolation is a linear operator and *monotone* in the sense that  $U \leq V$  implies that  $\mathcal{I}_{\Delta x} U \leq \mathcal{I}_{\Delta x} V$ .

The  $w_j$ 's form a basis and the relation  $\sum_i w_{\phi,i}(x) \equiv 1$  follows readily from the other assumptions in (I1) and (I2). Examples are constant, linear, or multi-linear interpolation (i. e.  $r \leq 2$  in (I1)) since higher order interpolation is not monotone.

The scheme is said to be of class  $\tilde{L}$  if it can be written as

$$(4.1) \quad \sup_{\alpha} \inf_{\beta} \left\{ B_{U^n,j,j}^{\alpha,\beta,n,n} U_j^n - \sum_{i \neq j} B_{U^n,j,i}^{\alpha,\beta,n,n} U_i^n - \sum_i B_{U^{n-1},j,i}^{\alpha,\beta,n,n-1} U_i^{n-1} - F_j^{\alpha,\beta,n} \right\} = 0$$

in  $G$ , where  $B_{U^m,i,j}^{\alpha,\beta,n,m} \geq 0$  might depend on  $U^m$ . In the case of monotone interpolation,  $B_{U^m,i,j}^{\alpha,\beta,n,m}$  are independent of  $U^m$ , and the  $\tilde{L}$ -property implies monotonicity of the approximation scheme in the sense of Barles-Souganidis [3].

In the following, we denote by  $c^{\alpha,\beta,+}$  the positive part of  $c^{\alpha,\beta}$ . We now show consistency and stability of the scheme.

**Lemma 4.1** (All SL schemes). *Assume (I1), (I2'), and (Y1) hold.*

(a) *The scheme (3.3) is consistent with (1.1) with truncation error bounded by*

$$\begin{aligned} & \frac{|1-2\theta|}{2} |\phi_{tt}|_0 \Delta t + C \left( \Delta t^2 (|\phi_{tt}|_0 + |\phi_{ttt}|_0 + |D\phi_{tt}|_0 + |D^2\phi_{tt}|_0) \right. \\ & \quad \left. + |D^r \phi|_0 \frac{\Delta x^r}{k^2} + (|D\phi|_0 + \dots + |D^4\phi|_0) k^2 \right). \end{aligned}$$

(b) *The scheme (3.3) is of class  $\tilde{L}$  (see (4.1) for the definition) if the following CFL condition holds,*

$$(4.2) \quad (1-\theta)\Delta t \left[ \frac{M}{k^2} - c_i^{\alpha,\beta,n-1+\theta} \right] \leq 1 \quad \text{and} \quad \theta \Delta t c_i^{\alpha,\beta,n-1+\theta} \leq 1 \quad \text{for all } \alpha, \beta, n, i.$$

(c) *If in addition (A1) and (4.2) hold and  $2\theta\Delta t \sup_{\alpha,\beta} |c^{\alpha,\beta,+}|_0 \leq 1$ , then any solution  $U$  of (3.3)–(3.4) is  $L^\infty$ -stable satisfying*

$$|U^n|_0 \leq e^{2 \sup_{\alpha,\beta} |c^{\alpha,\beta,+}|_0 t_n} \left[ |g|_0 + t_n \sup_{\alpha,\beta} |f^{\alpha,\beta}|_0 \right].$$

*Remark 4.3.* By parabolic regularity  $D^2 \sim \partial_t$ , so  $|D^2\phi_{tt}|_0 \sim |\phi_{ttt}|_0$ . When  $\theta = 1/2$ , the scheme (3.3) is second order accurate in time.

*Proof.* (a) The scheme (3.3) is consistent with (1.1) with a truncation error bound

$$\begin{aligned} & \frac{|1-2\theta|}{2} |\phi_{tt}|_0 \Delta t + \frac{1}{3} |\phi_{ttt}|_0 \Delta t^2 + \sup_{\alpha,\beta,n} \left\{ |L^{\alpha,\beta}[\bar{\phi}^{\theta,n}] - L_k^{\alpha,\beta}[\overline{\mathcal{T}_{\Delta x} \phi}^{\theta,n}]|_0 \right\} \\ & \quad + \sup_{\alpha,\beta,n} \left\{ |L^{\alpha,\beta}[\phi^{n-1+\theta} - \bar{\phi}^{\theta,n}]|_0 + |c^{\alpha,\beta,n-1+\theta}(\phi^{n-1+\theta} - \bar{\phi}^{\theta,n})|_0 \right\} \end{aligned}$$

for smooth  $\phi$ . Part (a) now follows since by (I1), (3.2), and simple computations,

$$\begin{aligned} |L^{\alpha,\beta}[\bar{\phi}^{\theta,n}] - L_k^{\alpha,\beta}[\overline{\mathcal{T}_{\Delta x} \phi}^{\theta,n}]| & \leq C |D^r \phi|_0 \frac{\Delta x^r}{k^2} + C (|D\phi|_0 + \dots + |D^4\phi|_0) k^2, \\ |L^{\alpha,\beta}[\phi^{n-1+\theta} - \bar{\phi}^{\theta,n}]| & \leq \Delta t^2 \theta (1-\theta) \sup_{\alpha,\beta} |L^{\alpha,\beta}[\phi_{tt}]|_0 \\ & \leq C \Delta t^2 (|D\phi_{tt}|_0 + |D^2\phi_{tt}|_0), \\ |c^{\alpha,\beta,n-1+\theta}(\phi^{n-1+\theta} - \bar{\phi}^{\theta,n})| & \leq C \theta (1-\theta) \Delta t^2 |\phi_{tt}|_0. \end{aligned}$$

(b) Note that since  $\sum_i w_{\phi,i} \equiv 1$ ,

$$L_k^{\alpha,\beta}[\mathcal{I}_{\Delta x}\phi(t, \cdot)](t_{n-1+\theta}, x_j) = \sum_{i \in \mathbb{N}} l_{\phi,j,i}^{\alpha,\beta,n-1+\theta} [\phi(t, x_i) - \phi(t, x_j)],$$

where

$$l_{\phi,j,i}^{\alpha,\beta,n-1+\theta} = \sum_{l=1}^M \frac{w_{\phi,i}(x_j + y_{k,l}^{\alpha,\beta,+}(t_{n-1+\theta}, x_j)) + w_{\phi,i}(x_j + y_{k,l}^{\alpha,\beta,-}(t_{n-1+\theta}, x_j))}{2k^2}$$

with  $\sum_i l_{\phi,j,i}^{\alpha,\beta,n-1+\theta} = \frac{M}{k^2}$ . The  $l_{\phi,j,i}^{\alpha,\beta,n-1+\theta}$ 's are non-negative by (I2). The coefficients in (4.1) can now be written as

$$\begin{aligned} B_{U^n,j,j}^{\alpha,\beta,n,n} &= 1 + \theta \Delta t_n \left( \frac{M}{k^2} - l_{U^n,j,j}^{\alpha,\beta,n-1+\theta} - c_j^{\alpha,\beta,n-1+\theta} \right), \\ B_{U^{n-1},j,j}^{\alpha,\beta,n,n-1} &= 1 - (1 - \theta) \Delta t_n \left( \frac{M}{k^2} - l_{U^{n-1},j,j}^{\alpha,\beta,n-1+\theta} - c_j^{\alpha,\beta,n-1+\theta} \right), \\ B_{U^n,j,i}^{\alpha,\beta,n,n} &= \theta \Delta t_n l_{U^n,j,i}^{\alpha,\beta,n-1+\theta}, & B_{U^{n-1},j,i}^{\alpha,\beta,n,n-1} &= (1 - \theta) \Delta t_n l_{U^{n-1},j,i}^{\alpha,\beta,n-1+\theta}, \end{aligned}$$

where  $j \neq i$ . These coefficients are positive if (4.2) holds.

(c) Fix any  $\varepsilon > 0$  and let  $j$  be such that  $|U_j^n| \geq |U|_0 - \varepsilon$ . Assume first that  $U_j^n \geq 0$ . By the definition and sign of the  $B$ -coefficients (see part (b)),

$$\begin{aligned} B_{U^n,j,j}^{\alpha,\beta,n,n} U_j^n &\geq \left( 1 - \theta \Delta t_n \sup_{\alpha,\beta} |c^{\alpha,\beta,+}|_0 \right) U_j^n + \theta \Delta t_n \left( \frac{M}{k^2} - l_{U^n,j,j}^{\alpha,\beta,n-1+\theta} \right) (|U^n|_0 - \varepsilon), \\ - \sum_{i \neq j} B_{U^n,j,i}^{\alpha,\beta,n,n} U_i^n &\geq -\theta \Delta t_n \left( \frac{M}{k^2} - l_{U^n,j,j}^{\alpha,\beta,n-1+\theta} \right) |U^n|_0, \\ - \sum_i B_{U^{n-1},j,i}^{\alpha,\beta,n,n-1} U_i^{n-1} &\geq - \left( 1 + (1 - \theta) \Delta t_n \sup_{\alpha,\beta} |c^{\alpha,\beta,+}|_0 \right) |U^{n-1}|_0. \end{aligned}$$

By (4.1) we then find that

$$\begin{aligned} |U_j^n| &= U_j^n \\ &\leq \left( \frac{1 + (1 - \theta) \Delta t \sup_{\alpha,\beta} |c^{\alpha,\beta,+}|_0}{1 - \theta \Delta t \sup_{\alpha,\beta} |c^{\alpha,\beta,+}|_0} \right) \left[ |U^{n-1}|_0 + \Delta t_n \sup_{\alpha,\beta} |f^{\alpha,\beta}|_0 + \theta \Delta t_n \frac{M}{k^2} \varepsilon \right] \\ &\leq e^{2 \sup_{\alpha,\beta} |c^{\alpha,\beta,+}|_0 t_n} \left[ |g|_0 + t_n \sup_{\alpha,\beta} |f^{\alpha,\beta}|_0 \right] + O_{k,n}(\varepsilon). \end{aligned}$$

If  $U_j^n < 0$ , a similar argument shows that  $\tilde{U} = -U$  satisfies the same inequality. Since  $|U^n|_0 \leq |U_j^n| + \varepsilon$  and  $\varepsilon$  is arbitrary, the result then follows.  $\square$

If monotone interpolation is used, we also prove existence, uniqueness, and convergence of the schemes.

**Theorem 4.2** (Monotone SL schemes). *Assume (A1), (I1), (I2), (Y1), and (4.2).*

(a) *There exists a unique bounded solution  $U$  of (3.3)–(3.4).*

(b)  *$U$  converges uniformly to the solution  $u$  of (1.1)–(1.2) as  $\Delta t, k, \frac{\Delta x^r}{k^2} \rightarrow 0$ .*

*Proof.* (a) *Existence and uniqueness* follow by induction. Let  $t = t_n$  and assume  $U^{n-1}$  is a known bounded function. For  $\varepsilon > 0$  we define the operator  $T$  by

$$TU_j^n = U_j^n - \varepsilon \cdot (\text{left hand side of (4.1)}) \quad \text{for all } j \in \mathbb{Z}^M.$$

Note that the fixed point equation  $U^n = TU^n$  is equivalent to equation (3.3). By the definition and sign of the  $B$ -coefficients (which do not depend on  $U^n$  in this proof!) we see that

$$\begin{aligned} & TU_j^n - T\tilde{U}_j^n \\ & \leq \sup_{\alpha,\beta} \left\{ \left[ 1 - \varepsilon \left( 1 + \Delta t_n \theta \left( \frac{M}{k^2} - l_{j,j}^{\alpha,\beta,n-1+\theta} - c_j^{\alpha,\beta,n-1+\theta} \right) \right) \right] (U_j^n - \tilde{U}_j^n) \right. \\ & \quad \left. + \varepsilon \Delta t_n \theta \left( \frac{M}{k^2} - l_{j,j}^{\alpha,\beta,n-1+\theta} \right) |U^n - \tilde{U}^n|_0 \right\} \\ & \leq \left( 1 - \varepsilon \left[ 1 - \Delta t_n \theta \sup_{\alpha,\beta} |c^{\alpha,\beta,+}|_0 \right] \right) |U^n - \tilde{U}^n|_0 \end{aligned}$$

for  $\varepsilon$  such that  $1 - \varepsilon(1 + \Delta t_n \theta (\frac{M}{k^2} - c_j^{\alpha,\beta,n-1+\theta})) \geq 0$  and  $\varepsilon(1 - \Delta t_n \theta \sup_{\alpha,\beta} |c^{\alpha,\beta,+}|_0) < 1$  for all  $j, n, \alpha, \beta$ . Taking the supremum over all  $j$  and interchanging the role of  $U$  and  $\tilde{U}$  proves that  $T$  is a contraction on the Banach space of bounded functions on  $X_{\Delta x}$  under the sup-norm. Existence and uniqueness then follows from the fixed point theorem (for  $U^n$ ) and for all of  $U$  by induction since  $U^0 = g$  is bounded.

(b) In view of the  $L^\infty$ -stability of the scheme (Lemma 4.1 (c)), *convergence* of  $U$  to the solution  $u$  of (1.1)–(1.2) follows from the Barles-Souganidis result in [3].  $\square$

## 5. EXAMPLES OF SL SCHEMES

**5.1. Examples of approximations  $L_k^{\alpha,\beta}$ .** We present several examples of approximations of the term  $L^{\alpha,\beta}[\phi]$  of the form  $L_k^{\alpha,\beta}[\phi]$ , including previous approximations that have appeared in [16, 20, 7, 11] plus more computational efficient variants.

1. The approximation of Falcone [16] (see also [9]),

$$b^{\alpha,\beta} D\phi \approx \frac{\mathcal{I}_{\Delta x} \phi(x + hb^{\alpha,\beta}) - \mathcal{I}_{\Delta x} \phi(x)}{h},$$

corresponds to our  $L_k^{\alpha,\beta}$  if  $k = \sqrt{h}$ ,  $y_k^{\alpha,\beta,\pm} = k^2 b^{\alpha,\beta}$ .

2. The approximation of Crandall-Lions [11],

$$\frac{1}{2} \text{tr}[\sigma^{\alpha,\beta} \sigma^{\alpha,\beta \top} D^2 \phi] \approx \sum_{j=1}^P \frac{\mathcal{I}_{\Delta x} \phi(x + k\sigma_j^{\alpha,\beta}) - 2\mathcal{I}_{\Delta x} \phi(x) + \mathcal{I}_{\Delta x} \phi(x - k\sigma_j^{\alpha,\beta})}{2k^2},$$

corresponds to our  $L_k^{\alpha,\beta}$  if  $y_{k,j}^{\alpha,\beta,\pm} = \pm k\sigma_j^{\alpha,\beta}$  and  $M = P$ .

3. The corrected version of the approximation of Camilli-Falcone [7] (see also [20]),

$$\begin{aligned} & \frac{1}{2} \text{tr}[\sigma^{\alpha,\beta} \sigma^{\alpha,\beta \top} D^2 \phi] + b^{\alpha,\beta} D\phi \\ & \approx \sum_{j=1}^P \frac{\mathcal{I}_{\Delta x} \phi(x + \sqrt{h}\sigma_j^{\alpha,\beta} + \frac{h}{P} b^{\alpha,\beta}) - 2\mathcal{I}_{\Delta x} \phi(x) + \mathcal{I}_{\Delta x} \phi(x - \sqrt{h}\sigma_j^{\alpha,\beta} + \frac{h}{P} b^{\alpha,\beta})}{2h}, \end{aligned}$$

corresponds to our  $L_k^{\alpha,\beta}$  if  $k = \sqrt{h}$ ,  $y_{k,j}^{\alpha,\beta,\pm} = \pm k\sigma_j^{\alpha,\beta} + \frac{k^2}{P} b^{\alpha,\beta}$  and  $M = P$ .

4. The new approximation obtained by combining approximations 1 and 2,

$$\begin{aligned} & \frac{1}{2} \text{tr}[\sigma^{\alpha,\beta} \sigma^{\alpha,\beta \top} D^2 \phi] + b^{\alpha,\beta} D\phi \\ & \approx \frac{\mathcal{I}_{\Delta x} \phi(x + k^2 b^{\alpha,\beta}) - \mathcal{I}_{\Delta x} \phi(x)}{k^2} + \sum_{j=1}^P \frac{\mathcal{I}_{\Delta x} \phi(x + k\sigma_j^{\alpha,\beta}) - 2\mathcal{I}_{\Delta x} \phi(x) + \mathcal{I}_{\Delta x} \phi(x - k\sigma_j^{\alpha,\beta})}{2k^2}, \end{aligned}$$

corresponds to our  $L_k^{\alpha,\beta}$  if  $y_{k,j}^{\alpha,\beta,\pm} = \pm k\sigma_j^{\alpha,\beta}$  for  $j \leq P$ ,  $y_{k,P+1}^{\alpha,\beta,\pm} = k^2b^{\alpha,\beta}$  and  $M = P + 1$ .

5. The new, more efficient version of approximation 3,

$$\frac{1}{2}\text{tr}[\sigma^{\alpha,\beta}\sigma^{\alpha,\beta\top}D^2\phi] + b^{\alpha,\beta}D\phi \approx \sum_{j=1}^{P-1} \frac{\mathcal{I}_{\Delta x}\phi(x + k\sigma_j^{\alpha,\beta}) - 2\mathcal{I}_{\Delta x}\phi(x) + \mathcal{I}_{\Delta x}\phi(x - k\sigma_j^{\alpha,\beta})}{2k^2} \\ + \frac{\mathcal{I}_{\Delta x}\phi(x + k\sigma_P^{\alpha,\beta} + k^2b^{\alpha,\beta}) - 2\mathcal{I}_{\Delta x}\phi(x) + \mathcal{I}_{\Delta x}\phi(x - k\sigma_P^{\alpha,\beta} + k^2b^{\alpha,\beta})}{2k^2},$$

corresponds to our  $L_k^{\alpha,\beta}$  if  $y_{k,j}^{\alpha,\beta,\pm} = \pm k\sigma_j^{\alpha,\beta}$  for  $j < P$ ,  $y_{k,P}^{\alpha,\beta,\pm} = \pm k\sigma_P^{\alpha,\beta} + k^2b^{\alpha,\beta}$  and  $M = P$ .

Approximation 5 is always more efficient than 3 in the sense that it requires fewer arithmetic operations. The most efficient of approximations 3, 4, and 5, is 4 when  $\sigma^{\alpha,\beta}$  does not depend on  $\alpha, \beta$  but  $b^{\alpha,\beta}$  does, and 5 in the other cases.

**5.2. Linear interpolation SL scheme (LISL).** To keep the scheme (3.3) monotone, linear or multi-linear interpolation is the most accurate interpolation one can use in general. In this typical case we call the full scheme (3.3)–(3.4) the LISL scheme, and we will now summarize the results of Section 4 for this special case.

**Corollary 5.1.** *Assume that (A1) and (Y1) hold.*

(a) *The LISL scheme is monotone if the CFL conditions (4.2) hold.*

(b) *The truncation error of the LISL scheme is  $O(|1 - 2\theta|\Delta t + \Delta t^2 + k^2 + \frac{\Delta x^2}{k^2})$ , so it is first order accurate when  $k = O(\Delta x^{1/2})$  and  $\Delta t = O(\Delta x)$ , or if  $\theta = \frac{1}{2}$ ,  $\Delta t = O(\Delta x^{1/2})$ .*

(c) *If  $2\theta\Delta t \sup_{\alpha,\beta} |c^{\alpha,\beta,+}|_0 \leq 1$  and (4.2) hold, then there exists a unique bounded and  $L^\infty$ -stable solution  $U$  of the LISL scheme converging uniformly to the solution  $u$  of (1.1)–(1.2) as  $\Delta t, k, \frac{\Delta x}{k} \rightarrow 0$ .*

From this result it follows that the scheme is at most *first order accurate*, has *wide and increasing stencil* and a *good CFL condition*. From the truncation error and the definition of  $L_k^{\alpha,\beta}$  the stencil is wide since the scheme is consistent only if  $\Delta x/k \rightarrow 0$  as  $\Delta x \rightarrow 0$  and has stencil length proportional to

$$l := \frac{\max_{t,x,\alpha,\beta,i} |y_{k,i}^{\alpha,\beta,-}| \vee |y_{k,i}^{\alpha,\beta,+}|}{\Delta x} \sim \frac{k}{\Delta x} \rightarrow \infty \quad \text{as } \Delta x \rightarrow 0.$$

Here we have used that if (A1) holds and  $\sigma \neq 0$ , then typically  $y_{k,i}^{\alpha,\beta,\pm} \sim k$ . Note that if  $k = \Delta x^{1/2}$ , then  $l \sim \Delta x^{-1/2}$ . Finally, in the case  $\theta \neq 1$  the CFL condition for (3.3) is  $\Delta t \leq Ck^2 \sim \Delta x$  when  $k = O(\Delta x^{1/2})$ , and it is much less restrictive than the usual parabolic CFL condition,  $\Delta t = O(\Delta x^2)$ .

**5.3. A high order SL scheme for monotone solutions.** In this section we introduce spatially second order accurate SL schemes (3.3)–(3.4) for non-degenerate tensor product grids. These schemes are based on monotonicity preserving cubic (MPC) Hermite interpolation [17, 15] and will be denoted MPCSL schemes in short. They are consistent for monotone (in coordinate directions) solutions of the scheme, but they are not monotone.

The MPC interpolation is obtained by a careful modification of cubic Hermite interpolation [15], and for a function of one variable on the interval  $[x_i, x_{i+1}]$  it takes the form

$$(5.1) \quad (\mathcal{I}_{\Delta x} \phi)(x) = \phi_i + (\phi_{i+1} - \phi_i)P_i(x)$$

where

$$(5.2) \quad P_i(x) = \alpha_i \frac{x - x_i}{\Delta x} + (3 - \beta_i - 2\alpha_i) \left( \frac{x - x_i}{\Delta x} \right)^2 - (2 - \alpha_i - \beta_i) \left( \frac{x - x_i}{\Delta x} \right)^3,$$

where  $\alpha_i, \beta_i$  are bounded coefficients depending on  $\phi_{i-2}, \phi_{i-1}, \dots, \phi_{i+3}$ . The algorithm is described in Appendix A. Multidimensional interpolation operators are obtained as tensor products of one-dimensional interpolation operators, i. e. by interpolating dimension by dimension.

*Remark 5.1.* Rewriting  $\mathcal{I}_{\Delta x} \phi$ , we find that  $(\mathcal{I}_{\Delta x} \phi)(x) = \sum_i \phi_i w_{\phi,i}(x)$  for

$$w_{\phi,i}(x) = (1 - P_i(x))1_{[x_i, x_{i+1})}(x) + P_{i-1}(x)1_{[x_{i-1}, x_i)}(x)$$

and  $1_{[x_i, x_{i+1})}(x)$  is the indicator function that is 1 in  $[x_i, x_{i+1})$  and 0 otherwise. It is immediate that  $\sum_i w_{\phi,i}(x) \equiv 1$ , and  $w_{\phi,i} \geq 0$  since  $P_i(x_i) = 0$ ,  $P_i(x_{i+1}) = 1$ , and  $P_i$  is monotone in between.

**Lemma 5.2.** *The above monotonicity preserving cubic interpolation satisfies (I2'). If the interpolated function is strictly monotone between grid points, then (I1) holds with  $r = 4$  and the method is fourth order accurate.*

*Proof.* Assumption (I2) holds by construction, see remark 5.1. The error estimate follows from [15], since the above algorithm coincides with the two sweep algorithm given there when  $n = 1$  interval is considered. In [15] it is proved that this algorithm gives third order accurate approximations to the exact derivatives and hence the cubic Hermite polynomial constructed using this approximation is fourth order accurate.  $\square$

By Lemma 5.2 and the results in Section 4 we have the following result:

**Corollary 5.3.** *Assume (A1), (Y1) hold, and that for all  $\Delta x \in (0, 1)$ , solutions  $U$  of the MPCSL scheme are such that  $\mathcal{I}_{\Delta x} U$  is strictly  $x$ -monotone between points in the  $x$ -grid  $X_{\Delta x}$ .*

(a) *The truncation error of the MPCSL scheme is*

$$O\left(1 - 2\theta|\Delta t + \Delta t^2 + k^2 + \frac{\Delta x^4}{k^2}\right),$$

*and hence the scheme is second order accurate in space when  $k = O(\Delta x)$  and first or second order accurate in time when  $\theta \neq \frac{1}{2}$  or  $\theta = \frac{1}{2}$  respectively.*

(b) *If  $2\theta\Delta t \sup_{\alpha, \beta} |c^{\alpha, \beta}|_0 \leq 1$ , then the solution  $U$  is  $L^\infty$ -stable.*

## 6. DISCUSSION

**6.1. Comparison with the scheme of Bonnans-Zidani (BZ).** In [6] (see also [5, 4]) Bonnans and Zidani suggest an alternative approach to discretize degenerate diffusion equations. Their idea is to approximate the diffusion matrix  $a^{\alpha, \beta}$  by a nicer matrix  $a_k^{\alpha, \beta}$  which admits monotone finite difference approximations. For every  $k \in \mathbb{N}$  they find a stencil

$$\mathcal{S}_k \subset \{\xi = (\xi_1, \dots, \xi_N) \in \mathbb{Z}^N : 0 < \max_{i=1}^N |\xi_i| \leq k, i = 1, \dots, N\}$$

and positive numbers  $a_{k,\xi}^{\alpha,\beta}$  such that

$$a^{\alpha,\beta} \approx a_k^{\alpha,\beta} := \sum_{\xi \in \mathcal{S}_k} a_{k,\xi}^{\alpha,\beta} \xi \xi^\top.$$

This leads to a diffusion term that is a linear combination of directional derivatives which are again approximated by central difference approximations,

$$\text{tr}[a^{\alpha,\beta} D^2 \phi] \approx \text{tr}[a_k^{\alpha,\beta} D^2 \phi] = \sum_{\xi \in \mathcal{S}_k} a_{k,\xi}^{\alpha,\beta} D_\xi^2 \phi \approx \sum_{\xi \in \mathcal{S}_k} a_{k,\xi}^{\alpha,\beta} \Delta_\xi \phi,$$

where  $D_\xi^2 = \text{tr}[\xi \xi^\top D^2] = (\xi \cdot D)^2$  and

$$\Delta_\xi w(x) = \frac{1}{|\xi|^2 \Delta x^2} \{w(x + \xi \Delta x) - 2w(x) + w(x - \xi \Delta x)\}.$$

This approximation is monotone by construction and respects the grid. In two space dimensions,  $a_k^{\alpha,\beta}$  can be chosen such that  $|a^{\alpha,\beta} - a_k^{\alpha,\beta}| = O(k^{-2})$  (cf. [5]), and then it is easy to see that the truncation error is

$$O(k^{-2} + k^2 \Delta x^2).$$

When  $b^{\alpha,\beta} \equiv 0$ , the BZ scheme can be obtained from (3.3) by replacing our  $L_k^{\alpha,\beta}$  by the above Bonnans-Zidani diffusion approximation. This scheme shares many properties with the LISL scheme, it is at most *first order accurate* (take  $k \sim \Delta x^{-1/2}$ ), it has a similar *wide and increasing stencil*, and it has a similar *good CFL condition*  $\Delta t \leq Ck^2 \Delta x^2$  ( $\sim \Delta x$  when  $k \sim \Delta x^{-1/2}$ ). To understand why the stencil is wide, simply note that  $k$  by definition is the stencil length and that the scheme is consistent only if  $k \rightarrow \infty$  and  $k\Delta x \rightarrow 0$ . The typical stencil length is  $k \sim \Delta x^{-1/2}$ , just as it was for the LISL scheme.

The main drawback of this method is that it is costly since we must compute the matrix  $a_k^{\alpha,\beta}$  for every  $x, t, \alpha, \beta$  in the grid. In the fast two dimensional implementation in [5], the number of operations for computing the coefficients for a fixed  $x, t, \alpha, \beta$  is  $\mathcal{O}(k)$  and thus goes to infinity as  $k \rightarrow \infty$  in bad cases. The LISL scheme is easier to understand and implement and is faster in the sense that the computational cost for approximating the diffusion matrix for fixed  $x, t, \alpha, \beta$  is independent of the stencil size. Later we will see some numerical indication that the LISL scheme could be faster than the BZ scheme in some test problems.

The MPCSL scheme in the typical case when  $k = \Delta x$ , is a *second order accurate in space* and *compact stencil* scheme having the usual (not so good) *CFL conditions* for parabolic problems  $\Delta t \sim k^2 = \Delta x^2$ . When it can be used, it is far more efficient than the other two schemes, see Section 9. However there is no proof that the method will converge to the correct solution, and it is formally convergent only when the exact solutions are essentially monotone, meaning monotone at least between grid points. Both the BZ and LISL schemes “always” converge.

**6.2. Boundary conditions.** When solving PDEs on bounded domains, the SL (and BZ) schemes may exceed the domain if they are not modified near the boundary. The reason is of course the wide stencil. This may or may not be a problem depending on the equation and the type of boundary condition: (i) For Dirichlet conditions the scheme needs to be modified near the boundary or boundary conditions must

be extrapolated. This may result in a loss of accuracy or monotonicity near the boundary. (ii) Homogeneous Neumann conditions can be implemented exactly by extending in the normal direction the values of the solution on the boundary to the exterior. (iii) If the boundary has no regular points, no boundary conditions can be imposed. In this case the SL schemes will not leave the domain if the normal diffusion tends to zero fast enough when the boundary is approached. Typical examples are equations of Black-Scholes type.

**6.3. Interpretation as a collocation method.** In the case the functions  $w_{\phi,i}$  in (I2) do not depend on  $\phi$  (and form a basis), the scheme (3.3)–(3.4) can then be interpreted as a collocation method for a derivative free equation, this is essentially the approach of Falcone et al. [16, 7]. The idea is that if

$$W^{\Delta x}(Q_T) = \{u : u \text{ is a function on } Q_T \text{ satisfying } u \equiv \mathcal{I}_{\Delta x} u \text{ in } Q_T\}$$

denotes the interpolant space associated to the interpolation  $\mathcal{I}_{\Delta x}$ , equation (3.3) can be stated in the following equivalent way: Find  $U \in W^{\Delta x}(Q_T)$  solving

$$(6.1) \quad \delta_{\Delta t_n} U_i^n = \inf_{\alpha \in \mathcal{A}} \sup_{\beta \in \mathcal{B}} \left\{ L_k^{\alpha, \beta} [\bar{U}^{\theta, n}]_i^{n-1+\theta} + c_i^{\alpha, \beta, n-1+\theta} \bar{U}_i^{\theta, n} + f_i^{\alpha, \beta, n-1+\theta} \right\} \text{ in } G.$$

In general  $W^{\Delta x}$  can be any space of approximations which is interpolating on the grid  $X_{\Delta x}$ , e. g. a space of splines, but we do not consider this generality here.

**6.4. Stochastic game/control interpretation.** The scheme (3.3)–(3.4) can be interpreted as the dynamical programming equation of a discrete stochastic differential game. We will explain this in the less technical case when  $\mathcal{B}$  is a singleton and the game simplifies to an optimal stochastic control problem.

Assume that (A1) holds, and for simplicity, that  $c^\alpha(t, x) \equiv 0$  and the other coefficients are independent of  $t$ . Then it is well-known (cf. [27]) that the (viscosity) solution  $u$  of (1.1)–(1.2) is the value function of the stochastic control problem:

$$(6.2) \quad u(T-t, x) = \min_{\alpha(\cdot) \in \mathcal{A}} E \left[ \int_t^T f^{\alpha(s)}(X_s) ds + g(X_T) \right],$$

where  $\mathcal{A}$  is a set of admissible  $\mathcal{A}$ -valued controls and the diffusion process  $X_s = X_s^{t, x, \alpha(\cdot)}$  satisfies the SDE

$$(6.3) \quad X_t = x \quad \text{and} \quad dX_s = \sigma^{\alpha(s)}(X_s) dW_s + b^{\alpha(s)} ds \quad \text{for } s > t.$$

This follows from dynamical programming (DP), and (1.1) is called the DP equation for the control problem (6.2)–(6.3). Similarly, the schemes (3.3)–(3.4) are DP equations (at least in the explicit case) of suitably chosen discrete time and space control problems approximating (6.2)–(6.3). We refer to [19] for more details.

We take the slightly different approach explored in [9, 16, 20, 7] to show the relation to control theory. The idea is to write the SL scheme in collocation form (6.1) and show that (6.1) is the DP equation of a discrete time continuous space optimal control problem. We illustrate this approach by deriving an explicit scheme involving  $L_k^\alpha$  as defined in part 4 Section 5.1. Let  $\{t_0 = 0, t_1, \dots, t_M = T\}$  be discrete times and consider the discrete time approximation of (6.2)–(6.3) given by

$$(6.4) \quad \tilde{u}(T-t_m, x) = \min_{\alpha \in \mathcal{A}_M} E \left[ \sum_{k=m}^{M-1} f^{\alpha_k}(\tilde{X}_k) \Delta t_{k+1} + g(\tilde{X}_M) \right],$$

$$(6.5) \quad \tilde{X}_m = x, \quad \tilde{X}_n = \tilde{X}_{n-1} + \sigma^{\alpha_n}(\tilde{X}_{n-1}) k_n \xi_n + b^{\alpha_n}(\tilde{X}_{n-1}) k_n^2 \eta_n, \quad n > m,$$

where  $k_n = \sqrt{(P+1)\Delta t_n}$ ,  $A_M \subset A$  is an appropriate subset of piecewise constant controls, and  $\xi_n = (\xi_{n,1}, \dots, \xi_{n,P})^\top$  and  $\eta_n$  are mutually independent sequences of i. i. d. random variables satisfying

$$P\left((\xi_{n,1}, \dots, \xi_{n,P}, \eta_n) = \pm e_j\right) = \frac{1}{2(P+1)} \quad \text{if } j \in \{1, \dots, P\},$$

$$P\left((\xi_{n,1}, \dots, \xi_{n,P}, \eta_n) = e_{P+1}\right) = \frac{1}{P+1},$$

( $e_j$  denotes the  $j$ -th unit vector) and all other values of  $(\xi_{n,1}, \dots, \xi_{n,P}, \eta_n)$  have probability zero. Here we have used a weak Euler approximation of the SDE coupled with a quadrature approximation of the integral. By DP

$$\tilde{u}(T - t_m, x) = \min_{\alpha \in A_M} E\left[\sum_{k=m}^{n-1} f^{\alpha_k}(\tilde{X}_k) \Delta t_{k+1} + \tilde{u}(T - t_n, \tilde{X}_n)\right] \quad \text{for all } n > m,$$

and taking  $n = m + 1$ ,  $s_{M-m} = T - t_m$ ,  $\Delta s_m = s_m - s_{m-1}$ ,  $\bar{k}_m = k_{M-m}$ , and evaluating the expectation using (6.5), we see that

$$\tilde{u}(s_{M-m}, x) = \min_{\alpha \in A} \left\{ f^\alpha(x) \Delta s_{M-m} + \frac{\bar{k}_{M-m-1}^2}{P+1} L_{\bar{k}_{M-m-1}}^\alpha [\tilde{u}(s_{M-m-1}, x) + \tilde{u}(s_{M-m-1}, x)] \right\},$$

where  $L_k^\alpha$  is as in Section 5.1 part 4. If we subtract  $\tilde{u}(s_{M-m-1}, x)$  from both sides and divide by  $\Delta s_{M-m} = \frac{\bar{k}_{M-m-1}^2}{P+1}$ , we find (6.1) with  $\theta = 0$ .

In [7], a similar argument is given in the stationary case for schemes involving the  $L_k^\alpha$  of part 3 Section 5.1. In fact it is possible to identify all  $L_k^\alpha$ 's appearing in Section 5.1 with DP equations of suitably chosen discrete time continuous space control problems. However assumption (Y1) is not strong enough for this approach to work for the general  $L_k^\alpha$  defined in Sections 3 and 4.

*Remark 6.1.* A DP approach naturally leads to explicit methods for time dependent PDEs. But implicit methods can be derived from a trick: Discretize the PDE in time by backward Euler to find a (sequence of) stationary PDEs and use the DP approach on each stationary PDE. This leads to an implicit iteration scheme since the DP equations of stationary problems are always implicit.

*Remark 6.2.* By the definition of  $L_k^\alpha$  and (Y1),  $x + y_{i,k}^{\alpha,\pm}$  can be seen as a short time approximation of (6.3). Hence the scheme (3.3) tracks particle paths approximately. In view of the discussion above we might say that the scheme follows particles in the mean because of the expectation. For first order PDEs, schemes defined in this way are called SL schemes by e. g. Falcone. Moreover, in this case our schemes will coincide with the SL schemes of Falcone [16] in the explicit case. This explains why we choose to call these schemes SL schemes also in the general case.

## 7. ERROR ESTIMATES IN THE MONOTONE CONVEX CASE

We derive error bounds when  $\mathcal{I}_{\Delta x}$  is monotone and  $\mathcal{B}$  is a singleton and hence (1.1) is convex. It is not known how to prove such results in the general case. In the following we do not indicate the trivial  $\beta$  dependence any more and we take

a uniform time-grid,  $G = \Delta t \{0, 1, \dots, N_T\} \times X_{\Delta x}$ , for simplicity. Let  $Q_{\Delta t, T} := \Delta t \{0, 1, \dots, N_T\} \times \mathbb{R}^N$  and consider the intermediate equation

$$(7.1) \quad \delta_{\Delta t} V^n(x) = \inf_{\alpha \in \mathcal{A}} \left\{ L_k^\alpha[\bar{V}^{\theta, n}](t, x) + c^\alpha(t, x) \bar{V}^{\theta, n}(x) + f^\alpha(t, x) \right\}_{t=t_{n-1+\theta}} \quad \text{in } \mathbb{R}^N$$

for  $n = 1, 2, 3, \dots$ , with initial condition

$$(7.2) \quad V(0, x) = g(x) \quad \text{in } \mathbb{R}^N.$$

**Lemma 7.1.** *Assume that (I2) and the CFL condition (4.2) hold and that  $\sup_n |V^n|_1 \leq C_V$ . If  $V$  solves (7.1)–(7.2) and  $U$  solves (3.3)–(3.4), then*

$$|U - V| \leq C \frac{\Delta x}{k^2} \quad \text{in } G.$$

*Proof.* Let  $W = U - V$  and subtract the equation for  $V$  from the one for  $U$  to find

$$\begin{aligned} W_i^n &\leq W_i^{n-1} + \Delta t \sup_{\alpha \in \mathcal{A}} \left\{ L_k^\alpha[\mathcal{I}_{\Delta x} \bar{W}^{\theta, n}]_i^{n-1+\theta} + c_i^{\alpha, n-1+\theta} \bar{W}_i^{\theta, n} \right. \\ &\quad \left. + L_k^\alpha[\mathcal{I}_{\Delta x} \bar{V}^{\theta, n} - \bar{V}^{\theta, n}]_i^{n-1+\theta} \right\} \quad \text{in } G. \end{aligned}$$

Let  $C_c = \max_\alpha |c^{\alpha, +}|_0$ . If  $W_i^n \geq 0$ , we rearrange using

$$\begin{aligned} |\mathcal{I}_{\Delta x} V^n - V^n|_0 &\stackrel{(I2)}{=} \left| \sum_j w_j(\cdot) (V_j^n - V^n(\cdot)) \right|_0 \leq \left| \sum_j w_j(\cdot) |V_j^n - V^n(\cdot)| \right|_0 \\ &\leq \Delta x |V^n|_1 \left| \sum_j w_j(\cdot) \right|_0 \stackrel{(I2)}{=} \Delta x |V^n|_1 \end{aligned}$$

to see that

$$\begin{aligned} &\left( 1 + \theta \Delta t \left( \frac{M}{k^2} - C_c \right) \right) W_i^n \\ &\leq W_i^{n-1} + \Delta t \sup_{\alpha \in \mathcal{A}} \left\{ \theta \left( L_k^\alpha[\mathcal{I}_{\Delta x} W^n]_i^{n-1+\theta} + \frac{M}{k^2} W_i^n \right) \right. \\ &\quad \left. + (1 - \theta) \left( L_k^\alpha[\mathcal{I}_{\Delta x} W^{n-1}]_i^{n-1+\theta} + c_i^{\alpha, n-1+\theta} W_i^{n-1} \right) \right\} \\ &\quad + 2 \Delta t \sup_n |V^n|_1 \frac{\Delta x}{k^2} \quad \text{in } G. \end{aligned}$$

By the CFL condition (4.2), the coefficients of the above inequality are all non-negative. Hence since  $W^n \leq |W^n|_0 := \sup_i |W_i^n|$ , we may replace  $W^n$  by  $|W^n|_0$  on the right hand side. Moreover, since  $\mathcal{I}_{\Delta x} |W^n|_0 = |W^n|_0$  and  $L_k^\alpha[|W^n|_0] = 0$ , the upper bound on the right hand side then reduces to

$$(1 + \Delta t(1 - \theta)C_c) |W^{n-1}|_0 + \theta \Delta t \frac{M}{k^2} |W^n|_0 + 2C_V \Delta t \frac{\Delta x}{k^2}.$$

If  $W_i^n < 0$ , then the same bound also holds for  $-W_i^n$ , and hence

$$(1 + \Delta t \theta \left( \frac{M}{k^2} - C_c \right)) |W^n|_0 \leq (1 + \Delta t(1 - \theta)C_c) |W^{n-1}|_0 + \theta \Delta t \frac{M}{k^2} |W^n|_0 + 2C_V \Delta t \frac{\Delta x}{k^2} \Delta t.$$

Since  $W^0 \equiv 0$  in  $X_{\Delta x}$ , an iteration then reveals that

$$|W^n|_0 \leq 2C_V \Delta t \frac{\Delta x}{k^2} \sum_{m=0}^n \left( \frac{1 + \Delta t(1 - \theta)C_c}{1 - \Delta t \theta C_c} \right)^m \leq t_n \frac{\Delta x}{k^2} 4C_V e^{C_c t_n}$$

when  $\Delta t$  is small enough, which implies the lemma.  $\square$

Next we estimate  $|V - u|$ , where  $u$  solves (1.1)–(1.2), by the regularization method of Krylov [18]. To do that we need a continuity and continuous dependence result for the scheme that relies on the following additional (covariance-type) assumptions: Whenever two sets of data  $\sigma, b$  and  $\tilde{\sigma}, \tilde{b}$  are given, the corresponding approximations  $L_k^\alpha, y_{k,i}^{\alpha,\pm}$  and  $\tilde{L}_k^\alpha, \tilde{y}_{k,i}^{\alpha,\pm}$  in (3.1) satisfy

$$(Y2) \quad \left\{ \begin{array}{l} \sum_{i=1}^M [y_{k,i}^{\alpha,+} + y_{k,i}^{\alpha,-}] - [\tilde{y}_{k,i}^{\alpha,+} + \tilde{y}_{k,i}^{\alpha,-}] \leq 2k^2(b^\alpha - \tilde{b}^\alpha), \\ \sum_{i=1}^M [y_{k,i}^{\alpha,+} y_{k,i}^{\alpha,+ \top} + y_{k,i}^{\alpha,-} y_{k,i}^{\alpha,- \top}] + [\tilde{y}_{k,i}^{\alpha,+} \tilde{y}_{k,i}^{\alpha,+ \top} + \tilde{y}_{k,i}^{\alpha,-} \tilde{y}_{k,i}^{\alpha,- \top}] \\ \quad - [y_{k,i}^{\alpha,+} \tilde{y}_{k,i}^{\alpha,+ \top} + \tilde{y}_{k,i}^{\alpha,+} y_{k,i}^{\alpha,+ \top} + y_{k,i}^{\alpha,-} \tilde{y}_{k,i}^{\alpha,- \top} + \tilde{y}_{k,i}^{\alpha,-} y_{k,i}^{\alpha,- \top}] \\ \leq 2k^2(\sigma^\alpha - \tilde{\sigma}^\alpha)(\sigma^\alpha - \tilde{\sigma}^\alpha)^\top + 2k^4(b^\alpha - \tilde{b}^\alpha)(b^\alpha - \tilde{b}^\alpha)^\top, \end{array} \right.$$

when  $\sigma, b, y_k^\pm$  are evaluated at  $(t, x)$  and  $\tilde{\sigma}, \tilde{b}, \tilde{y}_k^\pm$  are evaluated at  $(t, y)$  for all  $t, x, y$ .

In Section 8 we will prove the following error estimate.

**Theorem 7.2.** *Assume that  $\mathcal{B}$  is a singleton, that (A1), (Y1), (Y2), and the CFL conditions (4.2) hold, and that  $k \in (0, 1)$  and  $\Delta t \leq (2k_0 \wedge 2k_1)^{-1}$ . If  $u$  and  $V$  are bounded solutions of (1.1)–(1.2) and (7.1)–(7.2), then*

$$|V - u| \leq C(|1 - 2\theta|\Delta t^{1/4} + \Delta t^{1/3} + k^{1/2}) \quad \text{in } Q_{\Delta t, T}.$$

It also follows from the regularity results in Section 8 (see Proposition 8.4) that  $|V^n|_1 \leq 2C_T$ , so by Lemma 7.1 and Theorem 7.2 we have the following result.

**Corollary 7.3** (Error Bound). *Under (I1), (I2), and the assumptions of Theorem 7.2, if  $u$  solves (1.1)–(1.2) and  $U$  solves (3.3)–(3.4), then*

$$|u - U| \leq |u - V| + |V - U| \leq C(|1 - 2\theta|\Delta t^{1/4} + \Delta t^{1/3} + k^{1/2} + \frac{\Delta x}{k^2}) \quad \text{in } G.$$

This error bound applies to the LISL schemes, and it also holds for unstructured grids. For more regular solutions it is possible to obtain better error estimates, but general and optimal results are not available. The best estimate in our case is  $O(\Delta x^{1/5})$  which is achieved when  $k = O(\Delta x^{2/5})$  and  $\Delta t = O(k^2)$ . Note that the CFL conditions (4.2) already imply that  $\Delta t = O(k^2)$  if  $\theta < 1$ . Also note that the above bound does not show convergence when  $k$  is optimal for the LISL scheme ( $k = O(\Delta x^{1/2})$ ).

*Remark 7.1.* These results are consistent with results for special LISL type schemes for stationary Bellman equations. In fact if all coefficients are independent of time and  $c^\alpha(x) < -c < 0$ , then by combining the results of [7] and [1], exactly the same error estimate is obtained for the solution of a particular stationary LISL scheme and the unique *stationary* Lipschitz solution of (1.1).

## 8. PROOF OF THEOREM 7.2

We start by an existence and uniqueness result.

**Lemma 8.1.** *Assume that (A1), (Y1), and the CFL conditions (4.2) hold. Then there exists a unique solution  $U \in C_b(Q_{T, \Delta t})$  of (7.1)–(7.2).*

The proof is similar to (but simpler than) the proof of Theorem 4.2 with the modification that the fixed point is achieved in the Banach space  $C_b(\mathbb{R}^N)$  instead of the space of bounded functions on  $X_{\Delta x}$ .

We now give a result comparing subsolutions of (7.1) to supersolutions of

$$(8.1) \quad \begin{aligned} \delta_{\Delta t} U^n(x) &= \inf_{\alpha \in \mathcal{A}} \left\{ \tilde{L}_k^\alpha[\bar{U}^{\theta, n}](t, x) + \tilde{c}^\alpha(t, x) \bar{U}^{\theta, n} + \tilde{f}^\alpha(t, x) \right\}_{t=t_{n-1+\theta}} \quad \text{in } \mathbb{R}^N, n \geq 1, \\ U(0, x) &= \tilde{g}(x) \quad \text{in } \mathbb{R}^N, \end{aligned}$$

where  $\tilde{L}_k^\alpha$  is the operator defined in (3.1), (Y1), (Y2) when  $\sigma^\alpha, b^\alpha$  are replaced by  $\tilde{\sigma}^\alpha, \tilde{b}^\alpha$ .

**Theorem 8.2.** *Assume that (A1), (Y2), (4.2) hold for both (7.1) and (8.1). If  $U \in C(Q_{T, \Delta t})$  is a bounded above subsolution of (7.1) and  $\tilde{U} \in C(Q_{T, \Delta t})$  a bounded below supersolution of (8.1), then for all  $k \in (0, 1)$ ,  $\Delta t \leq (k_0 \wedge k_1)^{-1} \wedge \frac{L_0}{2L}$  (see below),  $x, y \in \mathbb{R}^N$ ,  $n \in \{0, 1, \dots, N_T\}$ ,*

$$\begin{aligned} U(t_n, x) - \tilde{U}(t_n, y) &\leq R_{k_0}(t_n) |U(0, \cdot) - \tilde{U}(0, \cdot)|_0 \\ &\quad + R_{k_0}(t_n) R_{k_1}(t_n) (L_0 + t_n L) |x - y| \\ &\quad + t_n R_{k_0}(t_n) \sup_{\alpha \in \mathcal{A}} [|(f - \tilde{f})^+|_0 + (|U|_0 \wedge |\tilde{U}|_0) |c - \tilde{c}|_0] \\ &\quad + t_n^{1/2} 2K_T \sup_{\alpha \in \mathcal{A}} [|b - \tilde{b}|_0 + |\sigma - \tilde{\sigma}|_0] \end{aligned}$$

where  $R_k(t) = 1/(1 - k\Delta t)^{t/\Delta t}$ ,  $K_T \leq R_{k_0}(T) R_{k_1}(T) (L_0 + TL)$ ,

$$\begin{aligned} L_0 &= [g]_1 \vee [\tilde{g}]_1 + 1, \quad L = ([c^\alpha]_1 \vee [\tilde{c}^\alpha]_1) (|U|_0 \wedge |\tilde{U}|_0) + [f^\alpha]_1 \vee [\tilde{f}^\alpha]_1, \\ k_0 &= \sup_{\alpha} |c^{\alpha,+}|_0, \quad k_1 = 8 \sup_{\alpha} \{[\sigma^\alpha]_1^2 + [b^\alpha]_1^2 + 1\}. \end{aligned}$$

*Remark 8.1.* The function  $R_k(n\Delta t) = 1/(1 - k\Delta t)^n$  satisfies  $\delta_{\Delta t} R_k(t_n) = kR_k(t_n)$ ,  $R_k(0) = 1$ , and  $R_k(t_n) \leq e^{2kt_n}$  when  $\Delta t \leq \frac{1}{2k}$ .

This is a key result in this paper, and the proof is given in Appendix B. In the stationary case, results of this type have been obtained in [1, 8] for simpler schemes. The result is a joint uniqueness (take  $(\tilde{\sigma}, \tilde{b}, \tilde{c}, \tilde{f}, \tilde{g}) = (\sigma, b, c, f, g)$ ), continuous dependence (take  $x = y$ ), boundedness, and  $x$ -Lipschitz continuity result:

**Corollary 8.3.** *Under the assumptions of Theorem 8.2, if  $k \in (0, 1)$  and  $\Delta t \leq (2k_0 \wedge 2k_1)^{-1}$ , then any bounded solution  $U \in C_b(Q_{T, \Delta t})$  of (7.1) satisfies*

- (i)  $|U(t_n, \cdot)|_0 \leq e^{2k_0 t_n} (|g|_0 + t_n \sup_{\alpha} |f^\alpha|_0)$ ,
- (ii)  $|U(t_n, x) - U(t_n, y)| \leq e^{2(k_0 + k_1)t_n} (L_0 + t_n L) |x - y|$ ,

where the constants, which are defined in Theorem 8.2, are independent of  $k, \Delta t, \Delta x$ .

*Proof.* Part (i) follows from Theorem 8.2 (with  $x = y$ ) and Remark 8.1 since  $\tilde{U} \equiv 0$  satisfies (8.1) with  $(\tilde{\sigma}^\alpha, \tilde{b}^\alpha, \tilde{c}^\alpha, \tilde{f}^\alpha, \tilde{g}^\alpha) = (\sigma^\alpha, b^\alpha, c^\alpha, 0, 0)$ . Part (ii) follows by taking  $U = \tilde{U}$  and  $x \neq y$ .  $\square$

Now we extend the scheme (7.1) to the whole space  $Q_T$ . One way to do this and to obtain continuous in time solutions is to pose initial conditions on  $[0, \Delta t)$  by interpolating between  $g(x)$  and  $U(\Delta t, x)$  where  $U$  is the solution of (7.1)–(7.2).

$$(8.2) \quad \delta_{\Delta t} V(t, x) = \inf_{\alpha \in \mathcal{A}} \left\{ L_k^\alpha[\bar{V}^\theta(t, \cdot)](t^\theta, x) + c^\alpha(t^\theta, x) \bar{V}^\theta(t, x) + f^\alpha(t^\theta, x) \right\} \\ \text{in } (\Delta t, T] \times \mathbb{R}^N,$$

$$(8.3) \quad V(t, x) = \left(1 - \frac{t}{\Delta t}\right) g(x) + \frac{t}{\Delta t} U(\Delta t, x) \quad \text{in } [0, \Delta t] \times \mathbb{R}^N.$$

where  $\bar{V}^\theta(t, x) = (1 - \theta)V(t - \Delta t, x) + \theta V(t, x)$  and  $t^\theta = t - (1 - \theta)\Delta t$ . From the previous results for  $U$  the existence, uniqueness, and properties of  $V$  easily follow.

**Proposition 8.4.** *Assume that (A1), (Y1), (Y2), and the CFL conditions (4.2) hold, and that  $k \in (0, 1)$  and  $\Delta t \leq (2k_0 \wedge 2k_1)^{-1}$ .*

(a) *There exists a unique solution  $V \in C_b(Q_T)$  of (8.2)–(8.3).*

(b) *There is a constant  $C_T \geq 0$  independent of  $k, \Delta t, \Delta x$  such that*

- (i)  $|V|_0 \leq C_T$ ,
- (ii)  $|V(t, x) - V(t, y)| \leq C_T |x - y|$  for all  $t \in [0, T], x, y \in \mathbb{R}^N$ ,
- (iii)  $|V(s_1, x) - V(s_2, x)| \leq C_T |s_1 - s_2|^{1/2}$  for all  $s_1, s_2 \in [0, T], x \in \mathbb{R}^N$ .

(c) *Let  $V \in C_b(Q_T)$  and  $\tilde{V} \in C_b(Q_T)$  be sub- and supersolutions of (8.2)–(8.3) corresponding to coefficients  $(\sigma^\alpha, b^\alpha, c^\alpha, f^\alpha, g)$  and  $(\tilde{\sigma}^\alpha, \tilde{b}^\alpha, \tilde{c}^\alpha, \tilde{f}^\alpha, \tilde{g})$  respectively. Then there is a constant  $C_T \geq 0$  independent of  $k, \Delta t, \Delta x$  such that for all  $t \in [0, T]$ ,*

$$|V(t, \cdot) - \tilde{V}(t, \cdot)|_0 \leq C_T \left( |g - \tilde{g}|_0 + t \sup_\alpha [ (|U|_0 \wedge |\tilde{U}|_0) |c^\alpha - \tilde{c}^\alpha|_0 + |f^\alpha - \tilde{f}^\alpha|_0 ] \right. \\ \left. + t^{1/2} \sup_\alpha [ |\sigma^\alpha - \tilde{\sigma}^\alpha|_0 + |b^\alpha - \tilde{b}^\alpha|_0 ] \right).$$

*Proof.* First note that the initial data on  $[0, \Delta t]$  is uniformly bounded and Lipschitz continuous in  $x$  and  $t$  by construction and Corollary 8.3.

(a) Existence of a bounded  $x$ -continuous solution follows from repeated use of Lemma 8.1 since we have initial conditions on  $[0, \Delta t]$ . Continuity in time follows from Theorem 8.2 (with  $x = y$ ) since the data is  $t$ -continuous.

(b) Part (i) and (ii) follow from Corollary 8.3 since the initial data is uniformly bounded and  $x$ -Lipschitz in  $[0, \Delta t]$ . To prove part (iii) we assume  $s_1 < s_2$  and let  $U(t, x)$  and  $\tilde{U}(t, x)$  solve (8.2) with data  $(\sigma^\alpha(t+s_1, x), b^\alpha(t+s_1, x), c^\alpha(t+s_1, x), f^\alpha(t+s_1, x), V(s_1, x))$  and  $(0, 0, 0, 0, V(s_1, x))$  respectively. Note that for  $t \in [0, T - s_1]$ ,  $\tilde{U}(t, x) \equiv V(s_1, x)$  and  $U(t, x) \equiv V(t + s_1, x)$  where  $V$  is the unique solution of (8.2)–(8.3). By part (c) we then get

$$|V(t + s_1, \cdot) - V(s_1, \cdot)|_0 = |U(t, \cdot) - \tilde{U}(t, \cdot)|_0 \\ \leq C_T \left( 0 + t \sup_\alpha [ |f^\alpha|_0 + |V|_0 |c^\alpha|_0 ] + t^{1/2} \sup_\alpha [ |\sigma^\alpha|_0 + |b^\alpha|_0 ] \right) \quad \text{for } t > 0,$$

and hence part (iii) follows.

(c) Note that by construction of the initial data and Theorem 8.2 with  $x = y$ , the result holds for  $t \in [0, \Delta t]$ , and then the result holds for any  $t > \Delta t$  by another application of Theorem 8.2 with  $x = y$ .  $\square$

Using Krylov's method of shaking the coefficients [18], we will now find smooth subsolutions of (8.2). First we introduce the auxiliary equation

$$(8.4) \quad \delta_{\Delta t} V^\varepsilon(t, x) = \inf_{\substack{0 \leq s \leq \varepsilon^2 \\ |e| \leq \varepsilon \\ \alpha \in \mathcal{A}}} \left\{ L_k^\alpha[\tau_{-e} \bar{V}^{\varepsilon, \theta}(t, \cdot)](r + s, x + e) \right. \\ \left. + c^\alpha(r + s, x + e) \bar{V}^{\varepsilon, \theta}(t, x) + f^\alpha(r + s, x + e) \right\}_{r=t^\theta - \Delta t - \varepsilon^2} \text{ in } (\Delta t, T] \times \mathbb{R}^N,$$

$$(8.5) \quad V^\varepsilon(t, x) = \left(1 - \frac{t}{\Delta t}\right) g(x) + \frac{t}{\Delta t} V^\varepsilon(\Delta t, x) \text{ in } [0, \Delta t] \times \mathbb{R}^N,$$

where  $\tau_e \phi(t, x) = \phi(t, x + e)$  and  $V^\varepsilon(\Delta t, x)$  is obtained by first solving (8.4) for discrete times  $t_n = n\Delta t$ . For this equation to be well-defined for  $t \in (\Delta t, T]$ , the data and  $y_{k,i}^{\alpha, \pm}$  must be defined for  $t \in (-\Delta t - \varepsilon^2, T + \varepsilon^2]$ . But this is ok since one can easily extend these functions to  $t \in [-r, T + r]$  for any  $r > 0$  in such a way that (A1), (Y1), (Y2) still hold. Also note that

$$(8.6) \quad L_k^\alpha[\tau_{-e} \bar{V}^{\varepsilon, \theta}(t, \cdot)](r + s, x + e) = \frac{1}{2k^2} \sum_{i=1}^M \left\{ \bar{V}^{\varepsilon, \theta}(t, x + y_{k,i}^{\alpha,+}(r + s, x + e)) \right. \\ \left. - 2\bar{V}^{\varepsilon, \theta}(t, x) + \bar{V}^{\varepsilon, \theta}(t, x + y_{k,i}^{\alpha,-}(r + s, x + e)) \right\},$$

and hence (8.4) is an equation of the same type as (8.2) (with different  $\mathcal{A}$  and shifted coefficients) satisfying (A1), (Y1), (Y2) whenever (8.2) does.

By Proposition 8.4 there is a unique solution  $V^\varepsilon$  of (8.4)–(8.5) in  $[0, T + \Delta t + \varepsilon^2] \times \mathbb{R}^N$ . Let  $U^\varepsilon(t, x) := V^\varepsilon(t + \Delta t + \varepsilon^2, x)$  and define by convolution,

$$(8.7) \quad U_\varepsilon(t, x) = \int_{\mathbb{R}^N} \int_0^\infty U^\varepsilon(t - s, x - e) \rho_\varepsilon(s, e) ds de,$$

where  $\varepsilon > 0$ ,  $\rho_\varepsilon(t, x) = \frac{1}{\varepsilon^{N+2}} \rho\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right)$ , and

$$\rho \in C^\infty(\mathbb{R}^{N+1}), \quad \rho \geq 0, \quad \text{supp } \rho \subset [0, 1] \times \{|x| \leq 1\}, \quad \int_{\mathbb{R}^N} \rho(e) de = 1.$$

Note that  $U_\varepsilon$  is well defined on the time interval  $[-\Delta t, T]$ . By the next result it is the sought after smooth subsolution of (8.2).

**Proposition 8.5.** *Under the assumptions of Proposition 8.4, the function  $U_\varepsilon$  defined in (8.7) satisfies*

(i)  $U_\varepsilon \in C^\infty((-\Delta t, T) \times \mathbb{R}^N)$ ,  $|U_\varepsilon|_1 \leq C$ ,  $|D^m \partial_t^n U_\varepsilon|_0 \leq C\varepsilon^{1-m-2n}$  for  $n, m \in \mathbb{N}$ .

(ii) If  $V$  is the solution of (8.2)–(8.3), then  $|U_\varepsilon - V| \leq C(\varepsilon + \Delta t^{1/2})$  in  $Q_T$ .

(iii)  $U_\varepsilon$  is a subsolution of (8.2) in  $Q_T$ .

*Proof.* The regularity estimates in (i) are immediate from properties of convolutions and the regularity of  $V^\varepsilon$ . The bound on  $U_\varepsilon - V$  (in  $[0, T]$ ) in (ii) follows from Proposition 8.4 (c) and (A1) which imply

$$|V^\varepsilon - V|_0 \leq C(\varepsilon + \Delta t^{1/2}),$$

and regularity of  $V^\varepsilon$  along with properties of convolutions,

$$|U_\varepsilon - V^\varepsilon|_0 \leq |U_\varepsilon - U^\varepsilon|_0 + |V^\varepsilon(\cdot + \Delta t + \varepsilon^2, \cdot) - V^\varepsilon|_0 \leq |V_\varepsilon|_1(\varepsilon + \Delta t^{1/2}).$$

To see that  $U_\varepsilon$  is a subsolution of (8.2), first note that from the definition of  $U^\varepsilon$  and (8.4) it follows that

$$\delta_{\Delta t} U^\varepsilon(t, x) \leq L_k^\alpha[\tau_{-e} \bar{U}^{\varepsilon, \theta}(t, \cdot)](t^\theta + s, x + e)$$

$$+ c^\alpha(t^\theta + s, x + e)\overline{U}^{\varepsilon, \theta}(t, x) + f^\alpha(t^\theta + s, x + e)$$

for all  $(t, x) \in [-\varepsilon^2, T] \times \mathbb{R}^N$ ,  $|e|, s^2 \leq \varepsilon$ , and  $\alpha \in \mathcal{A}$ . Now we change variables from  $(t + s, x + e)$  to  $(t, x)$  to find that

$$\begin{aligned} \delta_{\Delta t} U^\varepsilon(t - s, x - e) &\leq L_k^\alpha[\tau_{-e}\overline{U}^{\varepsilon, \theta}(t - s, \cdot)](t^\theta, x) \\ &\quad + c^\alpha(t^\theta, x)\overline{U}^{\varepsilon, \theta}(t - s, x - e) + f^\alpha(t^\theta, x) \end{aligned}$$

for all  $(t, x) \in [0, T] \times \mathbb{R}^N$ ,  $|e|, s^2 \leq \varepsilon$ , and  $\alpha \in \mathcal{A}$ . Then we multiply by  $\rho_\varepsilon(s, e)$  and integrate w. r. t.  $(s, e)$ . To see what the result is, note that

$$\begin{aligned} L_k^\alpha[\tau_{-e}U^\varepsilon(t - s, \cdot)](r, x) &= \frac{1}{2k^2} \sum_{i=1}^M \left\{ U^\varepsilon(t - s, x + y_{k,i}^{\alpha,+}(r, x) - e) \right. \\ &\quad \left. - 2U^\varepsilon(t - s, x - e) + U^\varepsilon(t - s, x + y_{k,i}^{\alpha,-}(r, x) - e) \right\}, \end{aligned}$$

and hence

$$\int \int L_k^\alpha[\tau_{-e}U^\varepsilon(t - s, \cdot)](r, x)\rho_\varepsilon(s, e) ds de = L_k^\alpha[U_\varepsilon(t, \cdot)](r, x).$$

For the whole equation we then have

$$\delta_{\Delta t} U_\varepsilon(t, x) \leq L_k^\alpha[\overline{U}_\varepsilon^\theta(t, \cdot)](t^\theta, x) + c^\alpha(t^\theta, x)\overline{U}_\varepsilon^\theta(t, x) + f^\alpha(t^\theta, x)$$

for all  $(t, x) \in Q_T$  and  $\alpha \in \mathcal{A}$ . Since this inequality holds for all  $\alpha$ , it follows that  $U_\varepsilon$  is a subsolution of (8.2) in all of  $Q_T$ .  $\square$

We are now in a position to prove the error estimate given in Theorem 7.2.

*Proof of Theorem 7.2.* Let  $U_\varepsilon$  be defined in (8.7). By Proposition 8.5 (i) and Lemma 4.1 (a),

$$\begin{aligned} \partial_t U_\varepsilon - \inf_{\alpha \in \mathcal{A}} \left\{ L^\alpha[\overline{U}_\varepsilon^\theta(t, \cdot)](t^\theta, x) + c^\alpha(t^\theta, x)\overline{U}_\varepsilon^\theta(t, x) + f^\alpha(t^\theta, x) \right\} \\ \leq \frac{|1 - 2\theta|}{2} |\partial_t^2 U_\varepsilon|_0 \Delta t + C \left\{ (|\partial_t^2 U_\varepsilon|_0 + |\partial_t^3 U_\varepsilon|_0 + |\partial_t^2 D U_\varepsilon|_0 + |\partial_t^2 D^2 U_\varepsilon|_0) \Delta t^2 \right. \\ \left. + (|D U_\varepsilon|_0 + \dots + |D^4 U_\varepsilon|_0) k^2 \right\} \\ \leq C \left\{ |1 - 2\theta| \varepsilon^{-3} \Delta t + \varepsilon^{-5} \Delta t^2 + \varepsilon^{-3} k^2 \right\} \end{aligned}$$

in  $Q_T$ . Moreover, by Proposition 8.5 (ii),

$$g(x) = U(0, x) \geq U_\varepsilon(0, x) - C(\varepsilon + \Delta t^{1/2}).$$

It follows that there is a constant  $C \geq 0$  such that

$$U_\varepsilon - C e^{\sup_\alpha |c^\alpha|_0 t} \left\{ \varepsilon + \Delta t^{1/2} + t \left( |1 - 2\theta| \varepsilon^{-3} \Delta t + \varepsilon^{-5} \Delta t^2 + \varepsilon^{-3} k^2 \right) \right\}$$

is a classical subsolution of (1.1)–(1.2) with time shifted coefficients. By continuous dependence and the comparison principle

$$U_\varepsilon - C e^{\sup_\alpha |c^\alpha|_0 t} \left\{ \varepsilon + \Delta t^{1/2} + t \left( |1 - 2\theta| \varepsilon^{-3} \Delta t + \varepsilon^{-5} \Delta t^2 + \varepsilon^{-3} k^2 \right) \right\} \leq u \text{ in } Q_T,$$

and hence by Proposition 8.5 (ii),

$$U - u = (U - U_\varepsilon) + (U_\varepsilon - u) \leq C \left\{ \varepsilon + \Delta t^{1/2} + |1 - 2\theta| \varepsilon^{-3} \Delta t + \varepsilon^{-5} \Delta t^2 + \varepsilon^{-3} k^2 \right\}.$$

We minimize w. r. t.  $\varepsilon$  and find that

$$u - U \leq \begin{cases} C(\Delta t^{1/4} + k^{1/2}) & \text{if } \theta \neq \frac{1}{2} \\ C(\Delta t^{1/3} + k^{1/2}) & \text{if } \theta = \frac{1}{2} \end{cases} \quad \text{in } Q_T.$$

The lower bound on  $u - U$  follows with symmetric – but much easier – arguments where a smooth supersolution of the equation (1.1) is constructed. Consistency and comparison for the scheme (8.2) is then used to conclude. In view of Lemma 4.1, the lower bound is a direct consequence of Theorem 3.1 (a) in [2].  $\square$

## 9. NUMERICAL RESULTS

In the following, we apply the LISL and MPCSL schemes to linear and convex test problems in two space-dimensions, and hence have no dependence of  $\beta$ . For the LISL scheme, we choose  $k = \sqrt{\Delta x}$  and a regular triangular grid, whereas for the MPCSL scheme we choose  $k = \Delta x$  and a regular rectangular grid. If not stated otherwise, we use  $\theta = 0$  (explicit methods), CFL condition  $\Delta t = k^2$ , and approximation 5.1.5 for  $L^{\alpha,\beta}$ . As error measure we will always use the  $L^\infty$ -norm, and the error rates are calculated as  $r_i = \frac{\ln \|e_i\| - \ln \|e_{i-1}\|}{\ln \|\Delta x_i\| - \ln \|\Delta x_{i-1}\|}$ . All calculations are done in Matlab, on an INTEL(R) Core(TM)2 Duo P8700, 2.54Ghz Laptop.

**9.1. Linear problem with smooth solution.** Our first problem is taken from [5] and has exact solution  $u(t, x) = (2 - t) \sin x_1 \sin x_2$ , its coefficients in (1.1) are

$$\begin{aligned} f^\alpha(t, x) &= \sin x_1 \sin x_2 [(1 + 2\beta^2)(2 - t) - 1] \\ &\quad - 2(2 - t) \cos x_1 \cos x_2 \sin(x_1 + x_2) \cos(x_1 + x_2), \\ c^\alpha(t, x) &= 0, \quad b^\alpha(t, x) = 0, \quad \sigma^\alpha(t, x) = \sqrt{2} \begin{pmatrix} \sin(x_1 + x_2) & \beta & 0 \\ \cos(x_1 + x_2) & 0 & \beta \end{pmatrix}. \end{aligned}$$

We consider  $\beta^2 = 0.1$  and  $\beta = 0$ . Note that in the second case, the scheme considered in [5] is not consistent. Table 1 gives the (spatial) errors and rates obtained at  $t = 1$  applying the LISL and the MPCSL scheme, as well as the CPU time needed. As the solution is linear in  $t$ , one time step suffices.

As expected for smooth solutions, in both cases we obtain order one for the LISL scheme and order two for the MPCSL scheme, and the CPU time needed is proportional to the number of grid points  $\frac{1}{\Delta x^2}$ . Here, we have chosen the grid points such that the solution is monotone in between. If not, we would get order one for the LISL scheme but no convergence for the MPCSL scheme (see Section 5.3).

**9.2. Linear problem with non-smooth solution.** The second problem we test has a non-smooth exact solution

$$u(t, x) = (1 + t) \sin \frac{x_2}{2} \begin{cases} \sin \frac{x_1}{2} & \text{for } -\pi < x_1 < 0, \\ \sin \frac{x_1}{4} & \text{for } 0 < x_1 < \pi \end{cases}$$

in  $[-\pi, \pi]^2$  and coefficients in (1.1) given by

$$\begin{aligned} f^\alpha(t, x) &= \sin \frac{x_2}{2} \begin{cases} \sin \frac{x_1}{2} (1 + \frac{1+t}{4} (\sin^2 x_1 + \sin^2 x_2)) & \text{for } -\pi < x_1 < 0 \\ \sin \frac{x_1}{4} (1 + \frac{1+t}{16} (\sin^2 x_1 + 4 \sin^2 x_2)) & \text{for } 0 < x_1 < \pi \end{cases} \\ &\quad - \sin x_1 \sin x_2 \cos \frac{x_2}{2} \begin{cases} \frac{1+t}{2} \cos \frac{x_1}{2} & \text{for } -\pi < x_1 < 0 \\ \frac{1+t}{4} \cos \frac{x_1}{4} & \text{for } 0 < x_1 < \pi \end{cases}, \end{aligned}$$

(A)  $\beta^2 = 0.1$

$\Delta x$	LISL			MPCSL		
	error	rate	time in s	error	rate	time in s
3.93e-2	3.79e-2		0.07	1.03e-3		0.11
1.96e-2	1.93e-2	0.97	0.30	2.57e-4	2.00	0.54
9.82e-3	9.45e-3	1.03	1.53	6.42e-5	2.00	3.75
4.91e-3	4.50e-3	1.07	6.25	1.61e-5	2.00	15.16
2.45e-3	2.43e-3	0.89	24.77	4.01e-6	2.00	62.02

(B)  $\beta = 0$

$\Delta x$	LISL			MPCSL		
	error	rate	time in s	error	rate	time in s
3.93e-2	3.94e-2		0.04	1.03e-3		0.06
1.96e-2	1.98e-2	0.99	0.14	2.57e-4	2.00	0.24
9.82e-3	9.94e-3	0.99	0.68	6.43e-5	2.00	1.47
4.91e-3	4.70e-3	1.08	2.61	1.61e-5	2.00	5.94
2.45e-3	2.45e-3	0.94	10.64	4.02e-6	2.00	25.29

TABLE 1. Results for the smooth linear problem at  $t = 1$ , grid adapted to monotonicity

$$c^\alpha(t, x) = 0, \quad b^\alpha(t, x) = 0, \quad \sigma^\alpha(t, x) = \sqrt{2} \begin{pmatrix} \sin x_1 \\ \sin x_2 \end{pmatrix},$$

and we pose Dirichlet boundary conditions. This is a monotone non-smooth problem, and we obtain order one half applying the LISL scheme and order one applying the MPCSL scheme, i. e. reduced rates, see Table 2. Again, one time step suffices, and the CPU time needed is thus proportional to  $\frac{1}{\Delta x^2}$ .

$\Delta x$	LISL			MPCSL		
	error	rate	time in s	error	rate	time in s
7.76e-2	1.24e-2		0.02	7.56e-3		0.03
3.90e-2	8.75e-3	0.51	0.04	4.19e-3	0.86	0.06
1.96e-2	6.19e-3	0.50	0.14	2.20e-3	0.93	0.28
9.80e-3	4.38e-3	0.50	0.76	1.12e-3	0.97	1.79
4.90e-3	3.10e-3	0.50	2.99	5.69e-4	0.98	7.16
2.45e-3	2.19e-3	0.50	11.48	2.86e-4	0.99	28.66

TABLE 2. Results for the non-smooth linear problem at  $t = 1$

### 9.3. Optimal control problems with smooth solutions.

- (A) We test an example from [5] with exact solution  $u(t, x_1, x_2) = \left(\frac{3}{2} - t\right) \sin x_1 \sin x_2$ . The corresponding coefficients and control set in (1.1) are

$$f^\alpha = \left(\frac{1}{2} - t\right) \sin x_1 \sin x_2 + \left(\frac{3}{2} - t\right) \left[ \sqrt{\cos^2 x_1 \sin^2 x_2 + \sin^2 x_1 \cos^2 x_2} \right]$$

$$- 2 \sin(x_1 + x_2) \cos(x_1 + x_2) \cos x_1 \cos x_2 \Big],$$

$$c^\alpha = 0, \quad b^\alpha = \alpha, \quad \sigma^\alpha = \sqrt{2} \begin{pmatrix} \sin(x_1 + x_2) \\ \cos(x_1 + x_2) \end{pmatrix}, \quad \mathcal{A} = \{\alpha \in \mathbb{R}^2 : \alpha_1^2 + \alpha_2^2 = 1\}.$$

As  $\sigma^\alpha$  does not depend on  $\alpha$  but  $b^\alpha$  does, we choose approximation 5.1.4 for  $L^{\alpha, \beta}$  and thus need only about half of the number of interpolations we would need if we had chosen approximation 5.1.5.

(B) The next test problem has exact solution  $u(t, x_1, x_2) = (2 - t) \sin x_1 \sin x_2$  and coefficients and control set given by

$$f^\alpha(t, x) = (1 - t) \sin x_1 \sin x_2 - 2\alpha_1 \alpha_2 (2 - t) \cos x_1 \cos x_2,$$

$$c^\alpha(t, x) = 0, \quad b^\alpha(t, x, \alpha) = 0, \quad \sigma^\alpha = \sqrt{2} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \quad \mathcal{A} = \{\alpha \in \mathbb{R}^2 : \alpha_1^2 + \alpha_2^2 = 1\}.$$

In both examples, due to the solution being linear in  $t$ , one time step suffices. The results at  $t = 0.5$  are given in Table 3, where again the grid is adapted to monotonicity. As expected for smooth solutions, the LISL scheme yields a numerical order of convergence of one, whereas the MPCSL scheme yields order two. The CPU time is now proportional to  $\frac{1}{\Delta x^3}$ , reflecting that we use  $\frac{4\pi}{\Delta x}$  grid points to discretize the control.

(A)						
$\Delta x$	LISL			MPCSL		
	error	rate	time in s	error	rate	time in s
3.93e-2	3.01e-2		2.00	8.40e-4		4.74
1.96e-2	1.61e-2	0.91	23.22	2.12e-4	1.98	53.06
9.82e-3	8.03e-3	1.00	268.64	5.30e-5	2.00	743.35
4.91e-3	3.94e-3	1.03	2161.63	1.33e-5	2.00	5995.62
2.45e-3	2.03e-3	0.96	17366.16	3.32e-6	2.00	48150.42

  

(B)						
$\Delta x$	LISL			MPCSL		
	error	rate	time in s	error	rate	time in s
3.93e-2	2.18e-2		4.31	5.14e-4		9.51
1.96e-2	1.07e-2	1.03	49.49	1.29e-4	2.00	109.50
9.82e-3	5.45e-3	0.97	571.99	3.21e-5	2.00	1515.24
4.91e-3	2.55e-3	1.10	4608.88	8.03e-6	2.00	12241.84
2.45e-3	1.34e-3	0.92	36995.81	2.01e-6	2.00	98378.95

TABLE 3. Results for optimal control problems at  $t = 0.5$ , grid adapted to monotonicity

**9.4. Convergence test for a super-replication problem.** We consider a test problem from [4] which was used to test convergence rates for numerical approximations of a super-replication problem from finance. The corresponding PDE is

(9.1)

$$\inf_{\alpha_1^2 + \alpha_2^2 = 1} \left\{ \alpha_1^2 u_t(t, x) - \frac{1}{2} \text{tr} \left( \sigma^\alpha(t, x) \sigma^{\alpha \top}(t, x) D^2 u(t, x) \right) \right\} = f(t, x), \quad 0 \leq x_1, x_2 \leq 3$$

with  $\sigma^\alpha(t, x) = \begin{pmatrix} \alpha_1 x_1 \sqrt{x_2} \\ \alpha_2 \eta(x_2) \end{pmatrix}$  and  $\eta(x) = x(3-x)$ . We take  $u(t, x) = 1 + t^2 - e^{-x_1^2 - x_2^2}$  as exact solution as in [4], and then  $f$  is forced to be

$$f(t, x) = \frac{1}{2} \left( u_t - \frac{1}{2} x_1^2 x_2 u_{x_1 x_1} - \frac{1}{2} x_2^2 (3 - x_2)^2 u_{x_2 x_2} \right. \\ \left. - \sqrt{\left( -u_t + \frac{1}{2} x_1^2 x_2 u_{x_1 x_1} - \frac{1}{2} x_2^2 (3 - x_2)^2 u_{x_2 x_2} \right)^2 + \left( x_1 \sqrt{x_2}^3 (3 - x_2) u_{x_1 x_2} \right)^2} \right).$$

In [4]  $\eta(x) = x$ , while we take  $\eta(x) = x(3-x)$  to prevent the LISL scheme from overstepping the boundaries. Note that changing  $\eta$  does *not* change the solutions as long as  $\eta > 0$  in the interior of the domain, see [4], and hence the above equation is equivalent to the equation used in [4]. The initial values and Dirichlet boundary values at  $x_1 = 0$  and  $x_2 = 0$  are taken from the exact solution. As in [4], at  $x = 3$  and  $y = 3$  homogeneous Neumann boundary conditions are implemented. To approximate the values of  $\alpha_1, \alpha_2$ , the Howard algorithm is used (see [4]), which requires an implicit time discretization, so we choose  $\theta = 1$ . As stop criterion of the iterations we require that the change of the maximal component and the sum over all components of the residual in Howard's algorithm are both smaller than 0.01. The minimization is done over  $\alpha_{1,k} + i\alpha_{2,k} = e^{2\pi i k / 2N_{\Delta x}}$ ,  $k = 1, \dots, N_{\Delta x}$ , where  $N_{\Delta x} = 3/\Delta x$  is the number of space grid points in one dimension. The linear systems involved are solved by the standard MATLAB back slash operator, using internally UMFPACK [13]. The numbers of time steps are chosen as  $\frac{1}{\Delta x}$  for the LISL scheme and  $\frac{1}{\Delta x^2}$  for the MPCSL scheme, respectively.

The results at  $t = 1$  are given in Table 4. Again, the numerical order of convergence is approximately one when the LISL scheme is used and approximately two for the MPCSL scheme. The CPU times are better than expected for both the LISL and MPCSL schemes: They get multiplied roughly by 10 when  $\Delta x$  is divided by 2, a property which can also be observed in [4]. The reason is that the Howard algorithm needs fewer iterations when the time step becomes smaller.

(A) LISL				(B) MPCSL			
$\Delta x$	error	rate	time in s	$\Delta x$	error	rate	time in s
1.50e-1	2.01e-1		0.71	3.00e-1	8.21e-2		1.40
7.50e-2	9.49e-2	1.08	5.52	1.50e-1	1.83e-2	2.17	11.38
3.75e-2	4.29e-2	1.15	59.32	7.50e-2	5.03e-3	1.86	124.25
1.87e-2	1.94e-2	1.15	803.26				

TABLE 4. Results for the convergence test for the super-replication problem at  $t = 1$

*Remark 9.1.* Equation (9.1) can not be written in a form (1.1) satisfying the assumptions of this paper, so the results of this paper do not apply to this problem.

However, it seems possible to extend the analysis of Section 4 to cover this problem using comparison results from [4] along with  $L^\infty$ -bounds on the numerical solution that follow from the maximum principle. Because of the unusual structure of (9.1), this analysis is not standard and outside the scope of this paper.

*Remark 9.2.* If we compare naively these results to the results of [4], we find that the LISL and MPCSL schemes are about 10 and up to a 1000 times faster than the method of [4]. Of course, this comparison is not fair, e. g. it could be that a less efficient linear solver is used in [4].

#### APPENDIX A. MONOTONICITY PRESERVING CUBIC INTERPOLATION

To define this type of interpolation, we start by a 1D function  $\phi$ . For each sub-interval  $[x_i, x_{i+1}]$ ,  $i \in \mathbb{Z}$ , we construct a cubic Hermite interpolant

$$\mathcal{I}_{\Delta x}\phi(x) = c_0 + c_1(x - x_i) + c_2(x - x_i)^2 + c_3(x - x_i)^3$$

fulfilling

$$\mathcal{I}_{\Delta x}\phi(x_i) = \phi_i, \quad (\mathcal{I}_{\Delta x}\phi)'(x_i) = d_i, \quad \mathcal{I}_{\Delta x}\phi(x_{i+1}) = \phi_{i+1}, \quad (\mathcal{I}_{\Delta x}\phi)'(x_{i+1}) = d_{i+1},$$

where  $\phi_i = \phi(x_i)$  and  $d_i$  is an estimate of the derivative of  $\phi$  at  $x_i$ . It follows that

$$(A.1) \quad c_0 = \phi_i, \quad c_1 = d_i, \quad c_2 = \frac{3\Delta_i - d_{i+1} - 2d_i}{\Delta x}, \quad c_3 = -\frac{2\Delta_i - d_{i+1} - d_i}{\Delta x^2},$$

where  $\Delta_i = \frac{\phi_{i+1} - \phi_i}{\Delta x}$ . To get a fourth order accurate interpolant,  $\phi'_i$  must be at least third order accurate, and we take the symmetric fourth order approximation

$$(A.2) \quad d_i = \frac{\phi_{i-2} - 8\phi_{i-1} + 8\phi_{i+1} - \phi_{i+2}}{12\Delta x}, \quad i \in \mathbb{Z}.$$

The resulting interpolation is not monotonicity preserving. Necessary and sufficient conditions for preserving monotonicity were found by Fritsch and Carlson [17] (see also [26]): If  $\Delta_i = 0$ , then monotonicity follows if and only if  $d_i = d_{i+1} = 0$ , and if

$$\alpha_i = \frac{d_i}{\Delta_i} \quad \text{and} \quad \beta_i = \frac{d_{i+1}}{\Delta_i},$$

then monotonicity for  $\Delta_i \neq 0$  follows if and only if  $(\alpha_i, \beta_i) \in \mathcal{M} = \mathcal{M}_e \cup \mathcal{M}_b$  where

$$\mathcal{M}_e = \{(\alpha, \beta) : (\alpha - 1)^2 + (\alpha - 1)(\beta - 1) + (\beta - 1)^2 - 3(\alpha + \beta - 2) \leq 0\},$$

$$\mathcal{M}_b = \{(\alpha, \beta) : 0 \leq \alpha \leq 3, 0 \leq \beta \leq 3\}.$$

Eisenstat, Jackson and Lewis [15] give an algorithm that modifies the derivative approximation  $d_i$  such that the above conditions are fulfilled, and for monotone data the resulting interpolant is a  $C^1$  fourth order approximation. We will only consider  $C^0$  interpolants, and in that case their algorithm simplifies to the following steps to compute  $(\mathcal{I}_{\Delta x}\phi)(x)$  on the interval  $[x_i, x_{i+1}]$ :

Step 1 Compute the initial  $d_i$  using (A.2).

Step 2 Compute  $\Delta_i$ . If  $\Delta_i \neq 0$  compute  $\alpha_i$  and  $\beta_i$ , else set  $\alpha_i = \beta_i = 1$ .

Step 3 Set  $\alpha_i := \max\{\alpha_i, 0\}$  and  $\beta_i := \max\{\beta_i, 0\}$ .

Step 4 If  $(\alpha_i, \beta_i) \notin \mathcal{M}$ , modify  $(\alpha_i, \beta_i)$  as follows:

- If  $\alpha_i \geq 3$  and  $\beta_i \geq 3$ , set  $\alpha_i = \beta_i = 3$ ,
- else if  $\beta_i > 3$  and  $\alpha_i + \beta_i \geq 4$ , decrease  $\beta_i$  such that  $(\alpha_i, \beta_i) \in \partial\mathcal{M}$ ,

- else if  $\beta_i > 3$  and  $\alpha_i + \beta_i < 4$ , increase  $\alpha_i$  such that  $(\alpha_i, \beta_i) \in \partial\mathcal{M}$  or  $\alpha_i = 4 - \beta_i$ , in the last case subsequently decrease  $\beta_i$  until  $(\alpha_i, \beta_i) \in \partial\mathcal{M}$ ,
- else if  $\alpha_i > 3$  and  $\alpha_i + \beta_i \geq 4$ , decrease  $\alpha_i$  such that  $(\alpha_i, \beta_i) \in \partial\mathcal{M}$ ,
- else if  $\alpha_i > 3$  and  $\alpha_i + \beta_i < 4$ , increase  $\beta_i$  such that  $(\alpha_i, \beta_i) \in \partial\mathcal{M}$  or  $\beta_i = 4 - \alpha_i$ , in the last case subsequently decrease  $\alpha_i$  until  $(\alpha_i, \beta_i) \in \partial\mathcal{M}$ .

Step 5 Finally, replace  $d_i$  by  $\alpha_i \Delta_i$  and  $d_{i+1}$  by  $\beta_i \Delta_i$  in (A.1) and compute  $\mathcal{I}_{\Delta x} \phi$  from the resulting formula which then equals (5.1).

## APPENDIX B. THE PROOF OF THEOREM 8.2

We will prove the result when  $k_0 = 0$ . The general case can be reduced to this case in a standard way by considering  $U/R_{k_0}$  and  $\tilde{U}/R_{k_0}$  instead of  $U$  and  $\tilde{U}$ . We use doubling of variables techniques similar to those used to prove this type of results for equation (1.1). We take

$$m_0 = |(U(0, \cdot) - \tilde{U}(0, \cdot))^+|_0, \quad m = \sup_{\alpha} \left[ |(f^\alpha - \tilde{f}^\alpha)^+|_0 + (|U|_0 \wedge |\tilde{U}|_0) |c^\alpha - \tilde{c}^\alpha|_0 \right],$$

$$M^2 = 4 \sup_{\alpha} \left[ |\sigma^\alpha - \tilde{\sigma}^\alpha|_0^2 + |b^\alpha - \tilde{b}^\alpha|_0^2 \right],$$

where  $\phi^+$  denotes the positive part of  $\phi$ , and define  $W(t, x, y) = U(t, x) - \tilde{U}(t, y)$ ,

$$\phi(t, x, y) = m_0 + tm + \frac{1}{2\varepsilon} K_T t M^2 + \frac{1}{2} R_{k_1}(t)(L_0 + tL)(\varepsilon + \frac{1}{\varepsilon}|x - y|^2) + \delta(|x|^2 + |y|^2),$$

$$\psi(t, x, y) = W(t, x, y) - \phi(t, x, y) - \eta(1 + t), \quad \tilde{m} = \sup_{\substack{t \in \Delta t \mathbb{N}_0 \\ x, y \in \mathbb{R}^N}} \psi(t, x, y) = \psi(\tilde{t}, \tilde{x}, \tilde{y}),$$

for  $\varepsilon, \delta, \eta > 0$  and a maximum point  $(\tilde{t}, \tilde{x}, \tilde{y})$ . A maximum point exists because of the  $\delta$ -terms in  $\phi$ . We will prove that for any sequence  $\eta_l \rightarrow 0$ , there is another sequence  $\delta_l \rightarrow 0$  such that  $\psi(\tilde{t}_l, \tilde{x}_l, \tilde{y}_l) \leq o(1)$  as  $l \rightarrow \infty$ . This implies Theorem 8.2 when  $k_0 = 0$ . To see this, fix  $t > 0, x, y$  and note that for any  $\varepsilon > 0$ ,

$$U(t, x) - \tilde{U}(t, y) - m_0 - tm - \frac{1}{2\varepsilon} K_T t M^2 - \frac{1}{2} R_{k_1}(t)(L_0 + tL)(\varepsilon + \frac{1}{\varepsilon}|x - y|^2) \\ \leq \psi(\tilde{t}_l, \tilde{x}_l, \tilde{y}_l) + \delta_l(|x|^2 + |y|^2) + \eta_l(1 + t) \leq o(1) \quad \text{as } l \rightarrow \infty.$$

In this inequality we send  $l \rightarrow \infty$  and choose  $\varepsilon = |x - y| \vee t^{1/2} M$  to find that

$$U(t, x) - \tilde{U}(t, y) \leq m_0 + tm + t^{1/2} K_T M + R_{k_1}(t)(L_0 + tL)|x - y|,$$

and hence Theorem 8.2 follows since  $t > 0, x, y$  were arbitrary. We will not be explicit about the form of the  $\delta$ -terms below. Their role is only to guarantee that the maximum is attained at a (finite) point  $(\tilde{t}, \tilde{x}, \tilde{y})$ , and their contribution will always be  $o(1)$  as  $\delta \rightarrow 0$  (see also Section 3 in [1]).

It is enough to prove that for every  $\eta > 0$ ,  $\psi(\tilde{t}, \tilde{x}, \tilde{y}) \leq o(1)$  as  $\delta \rightarrow 0$ . We proceed by contradiction assuming there is an  $\eta > 0$  such that  $\lim_{\delta \rightarrow 0} \psi(\tilde{t}, \tilde{x}, \tilde{y}) > 0$ . By the definition of  $\psi$  we now have  $W(\tilde{t}, \tilde{x}, \tilde{y}) > 0$  and  $\tilde{t} > 0$  for all  $\delta > 0$  small enough. The last statement is true since

$$\psi(0, \tilde{x}, \tilde{y}) \leq m_0 + L_0 |\tilde{x} - \tilde{y}| - m_0 - \frac{L_0}{2} (\varepsilon + \frac{1}{\varepsilon} |\tilde{x} - \tilde{y}|^2) - \eta < 0.$$

The rest of the proof will aim at getting a contradiction for the case  $\tilde{t} > 0$ . Even if we do not write it like that, what we show below is that  $\frac{\psi(\tilde{t}, \tilde{x}, \tilde{y}) - \psi(\tilde{t} - \Delta t, \tilde{x}, \tilde{y})}{\Delta t} \leq o(1) - \eta$  as  $\delta \rightarrow 0$ , and this is impossible since  $(\tilde{t}, \tilde{x}, \tilde{y})$  is a maximum point of  $\psi$ .

We proceed by defining the operator  $\Pi^\alpha$ ,

$$\begin{aligned} \Pi^\alpha[\phi(t, \cdot, \cdot)](r, x, y) &= \sum_{i=1}^M \left\{ \phi(t, x + y_{k,i}^{\alpha,+}(r, x), y + \tilde{y}_{k,i}^{\alpha,+}(r, y)) \right. \\ &\quad \left. - 2\phi(t, x, y) + \phi(t, x + y_{k,i}^{\alpha,-}(r, x), y + \tilde{y}_{k,i}^{\alpha,-}(r, y)) \right\}. \end{aligned}$$

By the definition of  $L_k^\alpha$  and  $\tilde{L}_k^\alpha$ , it follows that

$$\Pi^\alpha[W(t, \cdot, \cdot)](r, x, y) = 2k^2 \left\{ L_k^\alpha[U(t, \cdot)](r, x) - \tilde{L}_k^\alpha[\tilde{U}(t, \cdot)](r, y) \right\}.$$

We set  $\lambda := \frac{\Delta t}{k^2}$  and subtract the inequalities defining  $U$  and  $\tilde{U}$  (see (7.1) and (8.1)) to find that for  $(t, x), (t, y) \in Q_T$

$$\begin{aligned} W(t, x, y) &\leq W(t - \Delta t, x, y) + \sup_\alpha \left\{ \frac{\lambda}{2} \Pi^\alpha[\overline{W}^\theta(t, \cdot, \cdot)](t^\theta, x, y) \right. \\ &\quad \left. + \Delta t c^\alpha(t^\theta, x) \overline{W}^\theta(t, x, y) \right\} + \Delta t L|x - y| + \Delta t m, \end{aligned}$$

where  $\overline{W}^\theta(t, x, y) = (1 - \theta)W(t - \Delta t, x, y) + \theta W(t, x, y)$  and  $t^\theta = t - (1 - \theta)\Delta t$ . Note that this new ‘‘scheme’’ is still monotone by the definition of  $\Pi^\alpha$  and the CFL condition. Hence we may replace  $W$  in the above inequality by any bigger function coinciding with  $W$  at  $(t, x, y)$ . By the definition of  $\tilde{m}$ ,

$$W \leq \phi + \eta(1 + t) + \tilde{m} \quad \text{in} \quad \Delta t \mathbb{N}_0 \times \mathbb{R}^N \times \mathbb{R}^N,$$

and equality holds at  $(\tilde{t}, \tilde{x}, \tilde{y})$ . Therefore we find that

$$\begin{aligned} (\star) \quad \phi(\tilde{t}, \tilde{x}, \tilde{y}) + \eta(1 + \tilde{t}) &\leq \phi(\tilde{t} - \Delta t, \tilde{x}, \tilde{y}) + \eta(1 + \tilde{t} - \Delta t) \\ &\quad + \sup_\alpha \frac{\lambda}{2} \Pi^\alpha[\overline{\phi}^\theta(\tilde{t}, \cdot, \cdot)](\tilde{t}^\theta, \tilde{x}, \tilde{y}) + \Delta t L|\tilde{x} - \tilde{y}| + \Delta t m. \end{aligned}$$

Here we also used the fact that  $\Pi^\alpha[\eta(1 + t) + \tilde{m}] = 0$  and  $c^\alpha \leq 0$ . Moreover we can Taylor expand to see that

$$\begin{aligned} \Pi^\alpha[\phi(t, \cdot, \cdot)](r, x, y) &= \sum_{i=1}^M \left\{ (Y_i^+ + Y_i^-) \cdot D_x \phi + (\tilde{Y}_i^+ + \tilde{Y}_i^-) \cdot D_y \phi \right. \\ &\quad + \frac{1}{2} \text{tr}[D_{xx}^2 \phi \cdot (Y_i^+ Y_i^{+\top} + Y_i^- Y_i^{-\top})] + \frac{1}{2} \text{tr}[D_{yy}^2 \phi \cdot (\tilde{Y}_i^+ \tilde{Y}_i^{+\top} + \tilde{Y}_i^- \tilde{Y}_i^{-\top})] \\ &\quad \left. + \frac{1}{2} \text{tr}[D_{xy}^2 \phi \cdot (Y_i^+ \tilde{Y}_i^{+\top} + \tilde{Y}_i^+ Y_i^{+\top} + Y_i^- \tilde{Y}_i^{-\top} + \tilde{Y}_i^- Y_i^{-\top})] \right\}, \end{aligned}$$

where  $Y_i^\pm = y_{k,i}^{\alpha,\pm}(r, x)$  and  $\tilde{Y}_i^\pm = \tilde{y}_{k,i}^{\alpha,\pm}(r, y)$ . Now we use (Y2) along with the definition of  $\phi$ , to see that

$$\begin{aligned} \Pi^\alpha[\phi(t, \cdot, \cdot)](r, x, y) &\leq \frac{1}{\varepsilon} R_{k_1}(t)(L_0 + tL) \left\{ 2k^2 (b^\alpha(r, x) - \tilde{b}^\alpha(r, y))(x - y) \right. \\ &\quad + k^2 \text{tr} \left[ (\sigma^\alpha(r, x) - \tilde{\sigma}^\alpha(r, y)) (\sigma^\alpha(r, x) - \tilde{\sigma}^\alpha(r, y))^\top \right] \\ &\quad \left. + k^4 \text{tr} \left[ (b^\alpha(r, x) - \tilde{b}^\alpha(r, y)) (b^\alpha(r, x) - \tilde{b}^\alpha(r, y))^\top \right] \right\} + o(1), \end{aligned}$$

as  $\delta \rightarrow 0$ . These considerations lead to the following simplification of  $(\star)$ ,

$$\begin{aligned}
& \eta + \frac{\phi(\tilde{t}, \tilde{x}, \tilde{y}) - \phi(\tilde{t} - \Delta t, \tilde{x}, \tilde{y})}{\Delta t} \\
& \leq \theta \frac{1}{\varepsilon} R_{k_1}(\tilde{t})(L_0 + \tilde{t}L) \left( \frac{1}{2} M^2 + \frac{1}{4} k_1 |\tilde{x} - \tilde{y}|^2 \right) \\
& \quad + (1 - \theta) \frac{1}{\varepsilon} R_{k_1}(\tilde{t} - \Delta t)(L_0 + (\tilde{t} - \Delta t)L) \left( \frac{1}{2} M^2 + \frac{1}{4} k_1 |\tilde{x} - \tilde{y}|^2 \right) \\
& \quad + L |\tilde{x} - \tilde{y}| + m + o(1) \\
& \leq \frac{1}{\varepsilon} R_{k_1}(\tilde{t})(L_0 + \tilde{t}L) \left( \frac{1}{2} M^2 + \frac{1}{4} k_1 |\tilde{x} - \tilde{y}|^2 \right) + L |\tilde{x} - \tilde{y}| + m + o(1) := RHS,
\end{aligned}$$

as  $\delta \rightarrow 0$ . Now we proceed to calculate  $\delta_{\Delta t} \phi(t, x, y) = \frac{\phi(t, x, y) - \phi(t - \Delta t, x, y)}{\Delta t}$ . To do that we note that  $\delta_{\Delta t}(uv) = (\delta_{\Delta t} u)v + u\delta_{\Delta t} v - \Delta t(\delta_{\Delta t} u)(\delta_{\Delta t} v)$ . Since  $\delta_{\Delta t} R_{k_1}(t) = k_1 R_{k_1}(t)$  we then see that

$$\delta_{\Delta t}[R_{k_1}(t)(L_0 + tL)] = k_1 R_{k_1}(t)(L_0 + tL) + R_{k_1}(t)L - \Delta t L k_1 R_{k_1}(t),$$

and hence

$$\delta_{\Delta t} \phi(\tilde{t}, \tilde{x}, \tilde{y}) = m + \frac{1}{2} K_T \frac{1}{\varepsilon} M^2 + \frac{1}{2} R_{k_1}(\tilde{t}) [k_1(L_0 + (\tilde{t} - \Delta t)L) + L] (\varepsilon + \frac{1}{\varepsilon} |\tilde{x} - \tilde{y}|^2).$$

All of this leads to  $\eta \leq RHS - \delta_{\Delta t} \phi(\tilde{t}, \tilde{x}, \tilde{y}) \leq o(1)$  as  $\delta \rightarrow 0$  and  $2\Delta t L \leq L_0$ . The last inequality follows from the bound on  $K_T$ . We have our contradiction and the proof is complete.

#### REFERENCES

1. Guy Barles and Espen Robstad Jakobsen, *On the convergence rate of approximation schemes for Hamilton-Jacobi-Bellman equations*, M2AN Math. Model. Numer. Anal. **36** (2002), no. 1, 33–54. MR 1916291 (2003h:65142)
2. ———, *Error bounds for monotone approximation schemes for parabolic Hamilton-Jacobi-Bellman equations*, Math. Comp. **76** (2007), no. 260, 1861–1893 (electronic). MR 2336272 (2008i:65161)
3. Guy Barles and Panagiotis E. Souganidis, *Convergence of approximation schemes for fully nonlinear second order equations*, Asymptotic Anal. **4** (1991), no. 3, 271–283. MR 1115933 (92d:35137)
4. Olivier Bokanowski, Benjamin Bruder, Stefania Maroso, and Hasnaa Zidani, *Numerical approximation for a superreplication problem under gamma constraints*, SIAM J. Numer. Anal. **47** (2009), no. 3, 2289–2320. MR 2519604 (2010j:91213)
5. Joseph Frédéric Bonnans, Élisabeth Ottenwaelter, and Hasnaa Zidani, *A fast algorithm for the two dimensional HJB equation of stochastic control*, M2AN Math. Model. Numer. Anal. **38** (2004), no. 4, 723–735. MR 2087732 (2005e:93165)
6. Joseph Frédéric Bonnans and Hasnaa Zidani, *Consistency of generalized finite difference schemes for the stochastic HJB equation*, SIAM J. Numer. Anal. **41** (2003), no. 3, 1008–1021. MR 2005192 (2004i:49061)
7. Fabio Camilli and Maurizio Falcone, *An approximation scheme for the optimal control of diffusion processes*, RAIRO Modél. Math. Anal. Numér. **29** (1995), no. 1, 97–122. MR 1326802 (96a:49033)
8. Fabio Camilli and Espen Robstad Jakobsen, *A finite element like scheme for integro-partial differential Hamilton-Jacobi-Bellman equations*, SIAM J. Numer. Anal. **47** (2009), no. 4, 2407–2431. MR 2525605 (2010j:65227)
9. Italo Capuzzo Dolcetta, *On a discrete approximation of the Hamilton-Jacobi equation of dynamic programming*, Appl. Math. Optim. **10** (1983), no. 4, 367–377. MR 713483 (84j:49024)
10. Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. (N.S.) **27** (1992), no. 1, 1–67. MR 1118699 (92j:35050)

11. Michael G. Crandall and Pierre-Louis Lions, *Convergent difference schemes for nonlinear parabolic equations and mean curvature motion*, Numer. Math. **75** (1996), no. 1, 17–41. MR 1417861 (97j:65134)
12. John Crank and Phyllis Nicolson, *A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type*, Proc. Cambridge Philos. Soc. **43** (1947), 50–67. MR 0019410 (8,409b)
13. Timothy A. Davis, *Algorithm 832: UMFPACK V4.3—an unsymmetric-pattern multifrontal method*, ACM Trans. Math. Software **30** (2004), no. 2, 196–199. MR 2075981
14. Hongjie Dong and Nicolai V. Krylov, *On the rate of convergence of finite-difference approximations for Bellman equations with constant coefficients*, Algebra i Analiz **17** (2005), no. 2, 108–132. MR 2159586 (2006f:49050)
15. Stanley C. Eisenstat, Kenneth R. Jackson, and John W. Lewis, *The order of monotone piecewise cubic interpolation*, SIAM J. Numer. Anal. **22** (1985), no. 6, 1220–1237. MR 811195 (87d:65014)
16. Maurizio Falcone, *A numerical approach to the infinite horizon problem of deterministic control theory*, Appl. Math. Optim. **15** (1987), no. 1, 1–13. MR 866164 (88c:49025)
17. Frederic N. Fritsch and Ralph E. Carlson, *Monotone piecewise cubic interpolation*, SIAM J. Numer. Anal. **17** (1980), no. 2, 238–246. MR 567271 (81g:65012)
18. Nicolai V. Krylov, *On the rate of convergence of finite-difference approximations for Bellman's equations with variable coefficients*, Probab. Theory Related Fields **117** (2000), no. 1, 1–16. MR 1759507 (2001j:65134)
19. Harold J. Kushner and Paul Dupuis, *Numerical methods for stochastic control problems in continuous time*, second ed., Applications of Mathematics (New York), vol. 24, Springer-Verlag, New York, 2001, Stochastic Modelling and Applied Probability. MR 1800098 (2001g:93002)
20. José-Luis Menaldi, *Some estimates for finite difference approximations*, SIAM J. Control Optim. **27** (1989), no. 3, 579–607. MR 993288 (90m:65137)
21. Theodore Samuel Motzkin and Wolfgang Wasow, *On the approximation of linear elliptic differential equations by difference equations with positive coefficients*, J. Math. Physics **31** (1953), 253–259. MR 0052895 (14,693i)
22. Rémi Munos and Hasnaa Zidani, *Consistency of a simple multidimensional scheme for Hamilton-Jacobi-Bellman equations*, C. R. Math. Acad. Sci. Paris **340** (2005), no. 7, 499–502. MR 2135230 (2005k:49088)
23. Adam M. Oberman, *Convergent difference schemes for degenerate elliptic and parabolic equations: Hamilton-Jacobi equations and free boundary problems*, SIAM J. Numer. Anal. **44** (2006), no. 2, 879–895 (electronic). MR 2218974 (2007a:65173)
24. ———, *Wide stencil finite difference schemes for the elliptic Monge-Ampère equation and functions of the eigenvalues of the Hessian*, Discrete Contin. Dyn. Syst. Ser. B **10** (2008), no. 1, 221–238. MR 2399429 (2009f:35101)
25. David M. Pooley, Peter A. Forsyth, and Ken R. Vetzal, *Numerical convergence properties of option pricing PDEs with uncertain volatility*, IMA J. Numer. Anal. **23** (2003), no. 2, 241–267. MR 1974225 (2004b:65132)
26. Philip J. Rasch and David L. Williamson, *On shape-preserving interpolation and semi-Lagrangian transport*, SIAM J. Sci. Statist. Comput. **11** (1990), no. 4, 656–687. MR 1054632 (91f:65020)
27. Jiongmin Yong and Xun Yu Zhou, *Stochastic controls*, Applications of Mathematics (New York), vol. 43, Springer-Verlag, New York, 1999, Hamiltonian systems and HJB equations. MR 1696772 (2001d:93004)

(Kristian Debrabant) UNIVERSITY OF SOUTHERN DENMARK, DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, CAMPUSVEJ 55, 5230 ODENSE M, DENMARK

*E-mail address:* `debrabant@imada.sdu.dk`

(Espen R. Jakobsen) DEPARTMENT OF MATHEMATICAL SCIENCES, NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY, 7491 TRONDHEIM, NORWAY

*E-mail address:* `erj@math.ntnu.no`