Numerical schemes for partial differential equations related to optimal stopping time problems

Olivier Bokanowski
Laboratory Jacques Louis Lions
University Paris-Diderot (Paris 7)
and Ensta ParisTech

Joint work with
K. Debrabant (University of Southern Denmark)

Workshop AASS, IMPA, March 30, 2016

Outline

- **I. Diffusion equations with an obstacle term**
- **II. Schemes ... and difficulties**
- **III. A second order scheme**
- **VI. Numerical examples**

I. Motivation

Diffusion equation with an obstacle term:

$$(1)\quad \begin{cases} \min\left(u_t - \frac{\sigma(x)^2}{2}u_{xx} - b(x)u_x + ru,\ u - \varphi(x)\right) = 0, \\ \qquad\qquad\qquad\qquad t > 0,\ x \in \mathbb{R} \\ u(0,x) = \varphi(x), \quad x \in \mathbb{R} \end{cases}$$

Link with stochastic optimal stopping time problem:

$$u(t,x) = \sup_{\tau \in \mathcal{T}_{[0,t]}} \mathbb{E}\left[e^{-r\tau}\varphi(X_\tau^{0,x})\right], \quad t \geq 0, x \in \mathbb{R}$$

where

- $\mathcal{T}_{[0,t]}$ is the set of "stopping times" with values in $[0,t]$
- $X_\theta = X_\theta^{0,x}$ satisfies the following SDE :

$$\begin{cases} dX_\theta = b(X_\theta)d\theta + \sigma(X_\theta)dW_\theta, \quad \theta \geq 0 \\ X_0 = x. \end{cases}$$

I. Motivation

Diffusion equation with an obstacle term:

$$(1) \quad \begin{cases} \min\Big(u_t \overbrace{-\frac{\sigma(x)^2}{2}u_{xx} - b(x)u_x + ru}^{\mathcal{A}u},\ u-\varphi(x)\Big) = 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad t>0,\ x\in\mathbb{R} \\ u(0,x) = \varphi(x), \quad x\in\mathbb{R} \end{cases}$$

Link with stochastic optimal stopping time problem:

$$u(t,x) = \sup_{\tau\in\mathcal{T}_{[0,t]}} \mathbb{E}\Big[e^{-r\tau}\varphi(X_\tau^{0,x})\Big], \quad t\geq 0, x\in\mathbb{R}$$

where

- $\mathcal{T}_{[0,t]}$ is the set of "stopping times" with values in $[0,t]$
- $X_\theta = X_\theta^{0,x}$ satisfies the following SDE :

$$\begin{cases} dX_\theta = b(X_\theta)d\theta + \sigma(X_\theta)dW_\theta, \quad \theta\geq 0 \\ X_0 = x. \end{cases}$$

I. Motivation

Diffusion equation with an obstacle term:

$$(1)\quad \begin{cases} \min\Big(u_t \overbrace{-\frac{\sigma(x)^2}{2}u_{xx} - b(x)u_x + ru}^{\mathcal{A}u},\ u-\varphi(x)\Big) = 0, \\ \qquad\qquad\qquad\qquad t>0,\ x\in\mathbb{R} \\ u(0,x) = \varphi(x), \quad x\in\mathbb{R} \end{cases}$$

Link with stochastic optimal stopping time problem:

$$u(t,x) = \sup_{\tau\in\mathcal{T}_{[0,t]}} \mathbb{E}\Big[e^{-r\tau}\varphi(X_\tau^{0,x})\Big], \quad t\geq 0, x\in\mathbb{R}$$

where

- $\mathcal{T}_{[0,t]}$ is the set of "stopping times" with values in $[0,t]$
- $X_\theta = X_\theta^{0,x}$ satisfies the following SDE :

$$\begin{cases} dX_\theta = b(X_\theta)d\theta + \sigma(X_\theta)dW_\theta, \quad \theta\geq 0 \\ X_0 = x. \end{cases}$$

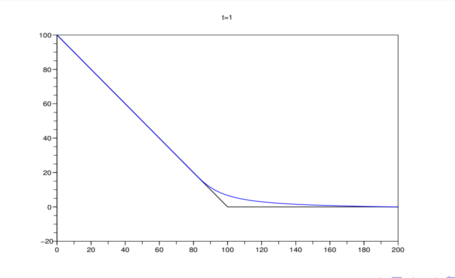
Financial problem: American put option

SDE $X_0 = x \geq 0$ and $\frac{dX_\theta}{X_\theta} = rd\theta + \sigma(X_\theta)dW_\theta$, $(\theta \geq 0)$

COST $\varphi(x) := \max(K - x, 0)$.

$$\Rightarrow \begin{cases} \min(u_t - \frac{x^2\sigma^2(x)}{2}u_{xx} - rxu_x + ru, u - \varphi(x)) = 0, & t \in [0, T],\ x \geq 0 \\ u(0,x) = \varphi(x), & x \geq 0 \end{cases}$$

Figure: american option



Dewynne, Howison, Rupf and Wilmott (1993): "Some mathematical results in the pricing of American options";

- There exists a singular point $x_s(t)$ such that

$$u(t,x) = \varphi(x), \quad x \leq x_s(t), \tag{1}$$

$$u_t(t,x) + \mathcal{A}u(t,x) = 0, \quad x > x_s(t), \tag{2}$$

$$u(t,x) = \varphi(t,x) \text{ and } u_x(t,x) = \varphi_x(t,x), \ \text{ for } x = x_s(t). \tag{3}$$

- $x_s(t) \simeq K - c_0 \log(t)\sqrt{t}$ for small t.
- regularity : u_t, u_{xx} bounded, regular for $x \neq x_s(t)$

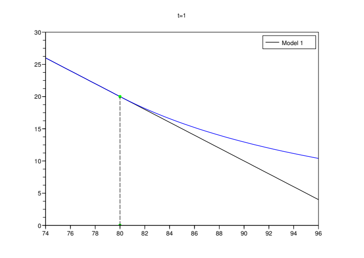


Figure: Zoom around the singular point x_s

Numerical difficulties

- Non linear PDE
- Singular: u_t, u_{xx} bounded but no more at $x = x_s(t)$ (for $t > 0$)
- Singular initial data at $t = 0$: $\max(K - x, 0)$
- Fast variation ($\sqrt{t}$) of the singularity $x_s(t)$ for $t \simeq 0$

A recent work of Reisinger and Forsyth shows that by using the change of variable of the form $v(t, x) = u(t^2, x)$ allows to smoothen the problem at $t = 0$ and partly remove the numerical problems coming from the initial singularity.

Numerical difficulties

- Non linear PDE
- Singular: u_t, u_{xx} bounded but no more at $x = x_s(t)$ (for $t > 0$)
- Singular initial data at $t = 0$: $\max(K - x, 0)$
- Fast variation ($\sqrt{t}$) of the singularity $x_s(t)$ for $t \simeq 0$

A recent work of Reisinger and Forsyth shows that by using the change of variable of the form $v(t, x) = u(t^2, x)$ allows to smoothen the problem at $t = 0$ and partly remove the numerical problems coming from the initial singularity.

An equivalent PDE (Thanks to R. Eymard)

Using the fact that $u_t \geq 0$, we can show that PDE (1) is equivalent to an other PDE:

$$(2) \begin{cases} \min\left(u_t - \frac{\sigma(x)^2}{2} u_{xx} - b(x)u_x + ru,\ u_t\right) = 0, & t > 0,\ x \in \mathbb{R} \\ u(0,x) = \varphi(x), & x \in \mathbb{R} \end{cases}$$

that is, the obstacle term is replaced by u_t.

It is important that the coefficients r, b, σ does not depend of the time for this equivalence to hold (in the financial context).

Equivalently,

$$(2) \Leftrightarrow \begin{cases} u_t + \min(-\frac{\sigma(x)^2}{2} u_{xx} - b(x)u_x + ru,\ 0) = 0, & t > 0,\ x \in \mathbb{R} \\ u(0,x) = g(x), & x \in \mathbb{R} \end{cases}$$

An equivalent PDE (Thanks to R. Eymard)

Using the fact that $u_t \geq 0$, we can show that PDE (1) is equivalent to an other PDE:

$$(2) \begin{cases} \min\left(u_t - \frac{\sigma(x)^2}{2} u_{xx} - b(x)u_x + ru,\ u_t\right) = 0, \quad t > 0,\ x \in \mathbb{R} \\ u(0,x) = \varphi(x), \quad x \in \mathbb{R} \end{cases}$$

that is, the obstacle term is replaced by u_t.

It is important that the coefficients r, b, σ does not depend of the time for this equivalence to hold (in the financial context).

Equivalently,

$$(2) \Leftrightarrow \begin{cases} u_t + \min(-\frac{\sigma(x)^2}{2} u_{xx} - b(x)u_x + ru,\ 0) = 0, \quad t > 0,\ x \in \mathbb{R} \\ u(0,x) = g(x), \quad x \in \mathbb{R} \end{cases}$$

All truth passes through three stages :
First, it is ridiculed
Second, it is violently opposed
Third, it is accepted as being self-evident
(Schopenhauer 1788-1860)

Idea of the proof of (1) $\Leftrightarrow$ (2)

1)

$$u_t + \mathcal{A}u \geq 0.$$

$\Big($ Using the semi-martingale property (or semi-DPP) :

$u(t+h,x) \geq \mathbb{E}(e^{-rh}u(t, X_h^{0,x}))$ $\Big)$

2) Notice that $u_t \geq 0$. Indeed,

$$\begin{aligned} u(t+h,x) &= \sup_{\tau \in \mathcal{T}_{[0,t+h]}} \mathbb{E}[g(X_\tau^{0,x})] \\ &\geq \sup_{t \in \mathcal{T}_{[0,t]}} \mathbb{E}[g(X_\tau^{0,x})] = u(t,x). \end{aligned}$$

Hence

$$\min(u_t + \mathcal{A}u,\ u_t) \geq 0.$$

Idea of the proof of (1) $\Leftrightarrow$ (2)

1)

$$u_t + \mathcal{A}u \geq 0.$$

$$\left(\text{Using the semi-martingale property (or semi-DPP) :} \quad u(t+h,x) \geq \mathbb{E}(e^{-rh}u(t,X_h^{0,x})) \right)$$

2) Notice that $u_t \geq 0$. Indeed,

$$\begin{aligned} u(t+h,x) &= \sup_{\tau\in\mathcal{T}_{[0,t+h]}} \mathbb{E}[g(X_\tau^{0,x})] \\ &\geq \sup_{t\in\mathcal{T}_{[0,t]}} \mathbb{E}[g(X_\tau^{0,x})] = u(t,x). \end{aligned}$$

Hence

$$\min(u_t + \mathcal{A}u,\ u_t) \geq 0.$$

Idea of the proof of (1) $\Leftrightarrow$ (2) (continued)

3) Suppose that $u_t(t,x) > 0$ (with $t > 0$). Then for $h > 0$ small enough,

$$u(t,x) > u(t-h,x) \geq \varphi(x).$$

Hence one can show that the optimal stopping time $\tau^*_{t,x} := \max\{0 \leq \theta \leq t,\ u(\theta, X^{0,x}_\theta) = \varphi(X^{0,x}_\theta)\}$ satisfies $\tau^*_{t,x} < t$ a.s., from which we can conclude using a DPP and Ito's formula that

$$u_t(t,x) + (\mathcal{A}u)(t,x) = 0.$$

4) Hence u (solution of (1)) satisfies also (2). By using a uniqueness argument on the (viscosity) solutions of (2), solutions of both systems coincide.

Idea of the proof of (1) $\Leftrightarrow$ (2) (continued)

3) Suppose that $u_t(t,x) > 0$ (with $t > 0$). Then for $h > 0$ small enough,

$$u(t,x) > u(t-h,x) \geq \varphi(x).$$

Hence one can show that the optimal stopping time $\tau^*_{t,x} := \max\{0 \leq \theta \leq t,\ u(\theta, X_\theta^{0,x}) = \varphi(X_\theta^{0,x})\}$ satisfies $\tau^*_{t,x} < t$ a.s., from which we can conclude using a DPP and Ito's formula that

$$u_t(t,x) + (\mathcal{A}u)(t,x) = 0.$$

4) Hence u (solution of (1)) satisfies also (2). By using a uniqueness argument on the (viscosity) solutions of (2), solutions of both systems coincide.

Application of PDE (2) :

$$\begin{cases} u_t + \min(\mathcal{A}u, 0) = 0, & t > 0,\ x \in \mathbb{R} \\ u(0,x) = \varphi(x), & x \in \mathbb{R} \end{cases}$$

• Related stochastic optimal control problem: u is also obtained as

$$u(t,x) = \sup_{\alpha(.)\in[0,1]} \mathbb{E}\left[e^{-\int_0^t \alpha(s) r\, ds} \varphi(X_t^{0,x,\alpha}) \right]$$

where $X_\theta = X_\theta^{0,x,\alpha}$ is the solution of $X_0 = x$ and

$$dX_\theta = \alpha(\theta)(b(X_\theta)d\theta + \sigma(X_\theta)dW_\theta).$$

• In particular the following DPP holds:

$$u(t_{n+1},x) = \sup_{\alpha(.)\in[0,1]} \mathbb{E}\left[e^{-\int_0^h \alpha(s) r\, ds} u(t_n, X_h^{0,x,\alpha}) \right]$$

• Taking $\alpha \equiv 0$, or $\alpha \equiv 1$ leads to the approximation

$$u(t_{n+1},x) \simeq \max\left(u(t_n,x),\ \mathbb{E}\left[e^{-rh} u(t_n, X_h^{0,x}) \right] \right).$$

II. Schemes and difficulties

- finite difference schemes
- efficient tools for solving non-linear implicit schemes.
- the "1st order barrier"

Simplified obstacle problem: $(\sigma^2/2 = 1,\ b = 0,\ r = 0)$

$$\min(v_t - v_{xx},\ v - \varphi(x)) = 0, \quad t \in (0,T),\ x \in \Omega = (X_{min}, X_{max}),$$
$$v(0,x) = \varphi(x)$$

with dirichlet boundary conditions on $\partial\Omega$.

Explicit finite difference scheme: Mesh $(x_i = ih)$, $(t_n = n\tau)$,

$$\min\left(\frac{u_i^{n+1} - u_i^n}{\tau} - \left(\frac{u_{i-1}^n - 2u_i^n + u_{i-1}^n}{h^2}\right),\ u_i^{n+1} - \varphi(x_i)\right) = 0, \quad 1 \le i \le J$$

with $u_0^{n+1} = u_{J+1}^{n+1} = 0$

... is really explicit ! :

$$\Rightarrow \quad u_i^{n+1} = \max\left(u_i^n - \tau\left(\frac{u_{i-1}^n - 2u_i^n + u_{i-1}^n}{h^2}\right),\ \varphi(x_i)\right)$$

Simplified obstacle problem: $(\sigma^2/2 = 1,\ b = 0,\ r = 0)$

$$\min(v_t - v_{xx},\ v - \varphi(x)) = 0, \quad t \in (0,T),\ x \in \Omega = (X_{min}, X_{max}),$$
$$v(0,x) = \varphi(x)$$

with dirichlet boundary conditions on $\partial\Omega$.

Explicit finite difference scheme: Mesh $(x_i = ih)$, $(t_n = n\tau)$,

$$\min\left(\frac{u_i^{n+1} - u_i^n}{\tau} - \left(\frac{u_{i-1}^n - 2u_i^n + u_{i-1}^n}{h^2}\right),\ u_i^{n+1} - \varphi(x_i)\right) = 0,$$
$$1 \le i \le J$$

with $u_0^{n+1} = u_{J+1}^{n+1} = 0$

... is really explicit ! :

$$\Rightarrow \quad u_i^{n+1} = \max\left(u_i^n - \tau\left(\frac{u_{i-1}^n - 2u_i^n + u_{i-1}^n}{h^2}\right),\ \varphi(x_i)\right)$$

- **Linear case**

$$v_t - v_{xx} = 0$$

- **Explicit scheme:**

$$\frac{u_i^{n+1} - u_i^n}{\tau} - \left(\frac{-u_{i-1}^n + 2u_i^n - u_{i-1}^n}{h^2} \right) = 0 \quad 1 \le i \le J$$

hence with $k := \dfrac{\tau}{h^2}$:

$$u_i^{n+1} = ku_{i-1}^n + (1 - 2k)u_i^n + ku_{i+1}^n \;\equiv (Su^n)_i$$

- CONSISTENCY: $\dfrac{v^{n+1} - Sv^n}{\tau} = (v_t - v_{xx})(t_n, x_i) + O(\tau) + O(h^2)$

- STABILITY : **CFL condition** $\boxed{\frac{2\tau}{h^2} \le 1} \Rightarrow \|U^{n+1}\|_\infty \le \|U^n\|_\infty$

- **Linear case**

$$v_t - v_{xx} = 0$$

- **Explicit scheme:**

$$\frac{u_i^{n+1} - u_i^n}{\tau} - \left(\frac{-u_{i-1}^n + 2u_i^n - u_{i-1}^n}{h^2} \right) = 0 \quad 1 \leq i \leq J$$

hence with $k := \frac{\tau}{h^2}$:

$$u_i^{n+1} = k u_{i-1}^n + (1-2k) u_i^n + k u_{i+1}^n \ \equiv (Su^n)_i$$

- CONSISTENCY: $\frac{v^{n+1} - Sv^n}{\tau} = (v_t - v_{xx})(t_n, x_i) + O(\tau) + O(h^2)$
- STABILITY : **CFL condition** $\boxed{\frac{2\tau}{h^2} \leq 1} \Rightarrow \|U^{n+1}\|_\infty \leq \|U^n\|_\infty$

- **IMPLICIT scheme:** (for linear case $v_t - v_{xx} = 0$)

$$\frac{u_i^{n+1} - u_i^n}{\tau} - \left(\frac{-u_{i-1}^{\mathbf{n+1}} + 2u_i^{\mathbf{n+1}} - u_{i-1}^{\mathbf{n+1}}}{h^2} \right) = 0 \quad 1 \leq i \leq J$$

$\Rightarrow AU^{n+1} = U^n$, with

$$U^n = \begin{pmatrix} u_1^n \\ \vdots \\ u_J^n \end{pmatrix} \quad A = \begin{bmatrix} 1+2k & -k & & \\ -k & \ddots & \ddots & \\ & & \ddots & -k \\ & & -k & 1+2k \end{bmatrix} \quad \text{and } k := \frac{\tau}{h^2} \geq 0.$$

- CONSISTENCY: idem, $O(\tau) + O(h^2)$
- STABILITY : **NO CFL condition !**

$$A \text{ "}\delta\text{-diag. dominant"} \Rightarrow \|A^{-1}\|_\infty \leq \tfrac{1}{\delta} \leq 1 \Rightarrow \|U^{n+1}\|_\infty \leq \|U^n\|_\infty \quad \forall \tau > 0$$

• **implicit for the nonlinear case:** **Can we do the same ?**

Implicit finite difference scheme

$$\min\left(\frac{u_i^{n+1}-u_i^n}{\tau}-\left(\frac{-u_{i-1}^{n+1}+2u_i^{n+1}-u_{i-1}^{n+1}}{h^2}\right),\ u_i^{n+1}-\varphi(x_i)\right)=0,$$

$$1\le i\le J$$

After multiplication of the left part of the min by $\tau>0$, we get:

$$\min\Big(\underbrace{(1+2k)u_i^{n+1}-ku_{i-1}^{n+1}-ku_{i+1}^{n+1}}_{=(Au^{n+1})_i}-\underbrace{u_i^n}_{\equiv b_i},\ u_i^{n+1}-\underbrace{\varphi(x_i)}_{\equiv g_i}\Big)=0$$

$$\Leftrightarrow \quad \text{find } x=U^{n+1},\quad \min((Ax-b)_i, x_i-g_i)=0,\quad 1\le i\le J$$

• EXISTENCE, STABILITY :

- If $A_{ii}\ge 0$ & A: $\delta\ge 1$-diag. domi. $\Rightarrow \exists! x$
- Moreover $\|U^{n+1}\|_\infty \le \max(\|U^n\|_\infty, \|g\|_\infty)$ $\quad\Rightarrow$ **NO CFL condition.**

• **implicit for the nonlinear case:** **Can we do the same ?**

Implicit finite difference scheme

$$\min\left(\frac{u_i^{n+1}-u_i^n}{\tau}-\left(\frac{-u_{i-1}^{n+1}+2u_i^{n+1}-u_{i-1}^{n+1}}{h^2}\right),\ u_i^{n+1}-\varphi(x_i)\right)=0,$$
$$1\le i\le J$$

After multiplication of the left part of the min by $\tau>0$, we get:

$$\min\Big(\underbrace{(1+2k)u_i^{n+1}-ku_{i-1}^{n+1}-ku_{i+1}^{n+1}}_{=(Au^{n+1})_i}-\underbrace{u_i^n}_{\equiv b_i},\ u_i^{n+1}-\underbrace{\varphi(x_i)}_{\equiv g_i}\Big)=0$$

$$\Leftrightarrow\quad \text{find } x=U^{n+1},\quad \min((Ax-b)_i, x_i-g_i)=0,\quad 1\le i\le J$$

• EXISTENCE, STABILITY :

- If $A_{ii}\ge 0$ & A: $\delta\ge$ 1-diag. domi. $\Rightarrow \exists! x$

- Moreover $\|U^{n+1}\|_\infty\le\max(\|U^n\|_\infty,\|g\|_\infty)$ $\Rightarrow$ **NO CFL condition.**

• **implicit for the nonlinear case:** **Can we do the same ?**

Implicit finite difference scheme

$$\min\left(\frac{u_i^{n+1}-u_i^n}{\tau}-\left(\frac{-u_{i-1}^{n+1}+2u_i^{n+1}-u_{i-1}^{n+1}}{h^2}\right),\ u_i^{n+1}-\varphi(x_i)\right)=0,$$
$$1\le i\le J$$

After multiplication of the left part of the min by $\tau>0$, we get:

$$\min\Big(\underbrace{(1+2k)u_i^{n+1}-ku_{i-1}^{n+1}-ku_{i+1}^{n+1}}_{=(Au^{n+1})_i}-\underbrace{u_i^n}_{\equiv b_i},\ u_i^{n+1}-\underbrace{\varphi(x_i)}_{\equiv g_i}\Big)=0$$

$$\Leftrightarrow\quad \text{find } x=U^{n+1},\quad \min((Ax-b)_i,x_i-g_i)=0,\quad 1\le i\le J$$

• EXISTENCE, STABILITY :

- If $A_{ii}\ge 0$ & A: $\delta\ge 1$-diag. domi. $\Rightarrow \exists! x$
- Moreover $\|U^{n+1}\|_\infty\le\max(\|U^n\|_\infty,\|g\|_\infty)$ $\Rightarrow$ **NO CFL condition.**

efficient tools for solving non-linear implicit schemes

PB: solve $F(x) := \min(Ax - b, x - g) = 0$.

- Known also as the "linear complementary problem" (when $g = 0$).
- Exist. and Uniq. result based on
  - positivity of all principal minors of A (A is a "P-matrix"[1])
  - the eigenvalues of A (Iain Smears, Lecture Notes, $\sim$2013)
- From the numerical and financial math. point of view, the following approaches have been considered (non exhaustive list)

(1) PSOR ($A = L + U$)

(2)

(3)

(4) Splitting (not solving $F = 0$ but rather $x = max(A^{-1}b, g)$).

[1] K. Murty, 1972

efficient tools for solving non-linear implicit schemes

PB: solve $F(x) := \min(Ax - b, x - g) = 0$.

• Known also as the "linear complementary problem" (when $g = 0$).
• Exist. and Uniq. result based on

- positivity of all principal minors of A (A is a "P-matrix"[1])
- the eigenvalues of A (Iain Smears, Lecture Notes, $\sim$2013)

• From the numerical and financial math. point of view, the following approaches have been considered (non exhaustive list)

(1) PSOR ($A = L + U$)
(2)
(3)
(4) Splitting (not solving $F = 0$ but rather $x = max(A^{-1}b, g)$).

[1] K. Murty, 1972

efficient tools for solving non-linear implicit schemes

PB: solve $F(x) := \min(Ax - b, x - g) = 0.$

- Known also as the "linear complementary problem" (when $g = 0$).
- Exist. and Uniq. result based on
  - positivity of all principal minors of A (A is a "P-matrix"[1])
  - the eigenvalues of A (Iain Smears, Lecture Notes, $\sim$2013)

- From the numerical and financial math. point of view, the following approaches have been considered (non exhaustive list)

(1) PSOR ($A = L + U$)

(2) Brennan Schwartz ($A = UL$)

(3)

(4) Splitting (not solving $F = 0$ but rather $x = max(A^{-1}b, g)$).

[1] K. Murty, 1972

efficient tools for solving non-linear implicit schemes

PB: solve $F(x) := \min(Ax - b, x - g) = 0$.

- Known also as the "linear complementary problem" (when $g = 0$).
- Exist. and Uniq. result based on
  - positivity of all principal minors of A (A is a "P-matrix"[1])
  - the eigenvalues of A (Iain Smears, Lecture Notes, $\sim$2013)

- From the numerical and financial math. point of view, the following approaches have been considered (non exhaustive list)

(1) PSOR ($A = L + U$)

(2) Brennan Schwartz ($A = UL$)

(3) Newton / Howard

(4) Splitting (not solving $F = 0$ but rather $x = max(A^{-1}b, g)$).

[1] K. Murty, 1972

efficient tools for solving non-linear implicit schemes

PB: solve $F(x) := \min(Ax - b, x - g) = 0$.

• Known also as the "linear complementary problem" (when $g = 0$).
• Exist. and Uniq. result based on

- positivity of all principal minors of A (A is a "P-matrix"[1])
- the eigenvalues of A (Iain Smears, Lecture Notes, ∼2013)

• From the numerical and financial math. point of view, the following approaches have been considered (non exhaustive list)

(1) PSOR ($A = L + U$)
(2) Brennan Schwartz ($A = UL$)
(3) Newton / Howard
(4) Splitting (not solving $F = 0$ but rather $x = max(A^{-1}b, g)$).

[1] K. Murty, 1972

efficient tools for solving non-linear implicit schemes

PB: solve $F(x) := \min(Ax - b, x - g) = 0$.

- Known also as the "linear complementary problem" (when $g = 0$).
- Exist. and Uniq. result based on
  - positivity of all principal minors of A (A is a "P-matrix"[1])
  - the eigenvalues of A (Iain Smears, Lecture Notes, $\sim$2013)

- From the numerical and financial math. point of view, the following approaches have been considered (non exhaustive list)

(1) PSOR ($A = L + U$)

(2) **Brennan Schwartz ($A = UL$)** $\Leftarrow$

(3) **Newton / Howard** $\Leftarrow$

(4) Splitting (not solving $F = 0$ but rather $x = max(A^{-1}b, g)$).

[1] K. Murty, 1972

PB: solve $F(x) := \min(Ax - b, x - g) = 0$.

Approach (2): Brennan and Schwartz : based on $A = UL$ decomposition with

$$U = tridiag(0, 1, u_i), \quad \text{and} \quad L = tridiag(l_i, d_i, 0), \ (d_i > 0)$$

Formally assumes that

$$\min(ULx - b, x - g) = 0 \quad \overset{(*)}{\Longleftrightarrow} \quad \min(Lx - U^{-1}b, x - g) = 0.$$

Then uses the descent algorithm (cost $O(J)$)

Proposition

(i) The equivalence $(*)$ is true for "profiled" solutions $x \in \mathbb{R}^J$ such that

$$\left\{ \exists i_0, \ x_i = g_i \text{ for } i \leq i_0, \text{ and } x_i > g_i \text{ otherwise} \right\}$$

(ii) Otherwise false (or should be adapted...)

PB: solve $F(x) := \min(Ax - b, x - g) = 0.$

Approach (3): semi-smooth Newton's method

- **Definition.** For $F(x)_i = \min((Ax - b)_i, x_i - g_i)$, let

$$F'(x)_{ij} = \begin{cases} A_{ij} \text{ if } (Ax - b)_i \leq (x - g)_i \\ I_{ij} = \delta_{ij} \text{ otherwise.} \end{cases}$$

- **Newton's algorithm:**

$$x^{k+1} = x^k - F(x^k)^{-1} F(x^k).$$

- More generally, consider for instance Merton's portfolio problem:

$$v(T-t,x) := \operatorname*{ess\,sup}_{\alpha:(t,T)\to\mathcal{K}} \mathbb{E}[\varphi(X_T^{t,x,\alpha})|\mathcal{F}_t], \qquad \boxed{\mathcal{K} \text{ compact}}$$

with $dX_\theta/X_\theta = b(X_\theta,\alpha_\theta)d\theta + \sigma(X_\theta,\alpha_\theta)dW_\theta$, leads to

$$\min_{a\in K}\left(v_t - \frac{1}{2}\sigma^2(x,a)x^2\ v_{xx} - b(x,a)x\ v_x\right) = 0.$$

Implicit finite difference scheme : leads to a matrix A_a and vector b_a depending of the parameter a, and the implicit scheme to solve:

$$\boxed{\min_{a\in K}(A_a x - b_a) = 0, \quad x \in \mathbb{R}^J}$$

• More generally, consider for instance Merton's portfolio problem:

$$v(T-t,x) := \operatorname*{ess\,sup}_{\alpha:(t,T)\to\mathcal{K}} \mathbb{E}[\varphi(X_T^{t,x,\alpha})|\mathcal{F}_t], \qquad \boxed{\mathcal{K} \text{ compact}}$$

with $dX_\theta/X_\theta = b(X_\theta,\alpha_\theta)d\theta + \sigma(X_\theta,\alpha_\theta)dW_\theta$, leads to

$$\min_{a\in K}\left(v_t - \frac{1}{2}\sigma^2(x,a)x^2\ v_{xx} - b(x,a)x\ v_x\right) = 0.$$

Implicit finite difference scheme : leads to a matrix A_a and vector b_a depending of the parameter a, and the implicit scheme to solve:

$$\boxed{\min_{a\in K}(A_a x - b_a) = 0, \quad x \in \mathbb{R}^J}$$

- **Definition 1.** B is a monotone matrix if, equivalently

  (i) $B^{-1} \geq 0$ *componentwise*
  (ii) $BX \geq 0 \Rightarrow X \geq 0$ *for all* X.

- **Definition 2.** B is an M matrix if

  B is diagonal dominant,
  $B_{ii} \geq 0$
  $B_{ij} \leq 0.$

- **Definition 3.** (Mixed matrices) For $a = (a_1, \ldots, a_J) \in \{0,1\}^J$, let

$$\begin{cases} B^a_{ij} := A_{ij},\ b^a_i := b_i \text{ if } a_i = 0 \\ B^a_{ij} := I_{ij},\ b^a_i := g_i \text{ if } a_i = 1 \end{cases}$$

Then $F(x) = \min(Ax - b, x - g) = \min_{a \in \{0,1\}^J}(B^a x - b^a)$

- **Monotony assumption (Mono):**

$$\forall a \in \{0,1\}^J,\ B^a \text{ is a monotone matrix}$$

- **Example:** A is an M-matrix $\Rightarrow$ **(Mono)**

Theorem ((Rust & Santos 04') (O.B., Maroso, Zidani 09'))

Assume **(Mono)**,
(i) *there exists a unique* $x \in \mathbb{R}^J$ *s.t.* $F(x) = 0$;
(ii) $\forall x^0,\ \lim_{k\to\infty} x^k = x$. *(Furthermore* $x^k \leq x^{k+1}$*)*
(iii) *The convergence is in at most* 2^J *iterations.*

Proof: equivalence of Howard's algorithm (1958') and Newton's method.

Theorem ((Ito,Hintermuller,Kunisch) (O.B., Maroso ,Zidani))

Obstacle pb: the convergence is in at most J *iterations !*

- **Monotony assumption (Mono):**

$$\forall a \in \{0,1\}^J, \;\; B^a \text{ is a monotone matrix}$$

- **Example:** A is an M-matrix $\Rightarrow$ **(Mono)**

Theorem ((Rust & Santos 04') (O.B., Maroso, Zidani 09'))

Assume **(Mono)**,
(i) *there exists a unique* $x \in \mathbb{R}^J$ *s.t.* $F(x) = 0$;
(ii) $\forall x^0, \; \lim_{k\to\infty} x^k = x$. *(Furthermore* $x^k \leq x^{k+1}$*)*
(iii) *The convergence is in at most* 2^J *iterations.*

Proof: equivalence of Howard's algorithm (1958') and Newton's method.

Theorem ((Ito,Hintermuller,Kunisch) (O.B., Maroso ,Zidani))

Obstacle pb: the convergence is in at most J *iterations !*

- **Monotony assumption (Mono):**

$$\forall a \in \{0,1\}^J,\ B^a \text{ is a monotone matrix}$$

- **Example:** A is an M-matrix $\Rightarrow$ **(Mono)**

Theorem ((Rust & Santos 04') (O.B., Maroso, Zidani 09'))

Assume **(Mono)**,
(*i*) *there exists a unique* $x \in \mathbb{R}^J$ *s.t.* $F(x) = 0$;
(*ii*) $\forall x^0$, $\lim_{k\to\infty} x^k = x$. *(Furthermore* $x^k \leq x^{k+1}$*)*
(*iii*) *The convergence is in at most* 2^J *iterations.*

Proof: equivalence of Howard's algorithm (1958') and Newton's method.

Theorem ((Ito,Hintermuller,Kunisch) (O.B., Maroso ,Zidani))

Obstacle pb: the convergence is in at most J iterations !

- **Monotony assumption (Mono):**

$$\forall a \in \{0,1\}^J,\ B^a \text{ is a monotone matrix}$$

- **Example:** A is an M-matrix $\Rightarrow$ **(Mono)**

Theorem ((Rust & Santos 04') (O.B., Maroso, Zidani 09'))

Assume **(Mono)**,
(*i*) *there exists a unique* $x \in \mathbb{R}^J$ *s.t.* $F(x) = 0$;
(*ii*) $\forall x^0,\ \lim_{k\to\infty} x^k = x$. *(Furthermore* $x^k \le x^{k+1}$*)*
(*iii*) *The convergence is in at most* 2^J *iterations.*

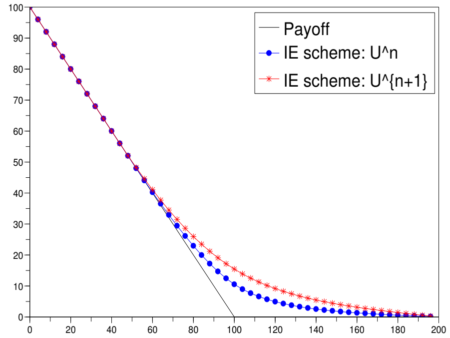
Proof: equivalence of Howard's algorithm (1958') and Newton's method.

Theorem ((Ito,Hintermuller,Kunisch) (O.B., Maroso ,Zidani))

Obstacle pb: the convergence is in at most J iterations !

Application to american options: (OB, Maroso, Zidani 2009)
Limitation of the total number's of Newton's iteration := bounded by the number of mesh points where the value "takes off" the payoff function.

Figure: Two successive iteration for the american option



The first order "barrier"

- Can we improve the error in time from $O(\tau)$ to $O(\tau^2)$ or better ?
- **(Godunov's Th.)** For $u_t + bu_x = 0$, monotone linear schemes $u_i^{n+1} = \sum a_j u_{i+j}^n$ ($a_j \geq 0$) are limited to at most first order.
- For other reasons (see later on..) for non linear diffusion problems some limitations hold, and we are led to try implicit schemes.

- Let us try a Crank - Nicolson scheme (RK2):

$$S_j^{1,n}(u) := \min\left(\frac{u_i^{n+1} - u_i^n}{\tau} + \frac{1}{2}(Au^n + Au^{n+1})_i,\ u_i^{n+1} - g_i\right) = 0,$$

$$1 \leq i \leq J$$

For A: M-matrix, we can solve it by a Newton's Algorithm.

- Numerical experiments show
  - - using $\tau = h$, stable, **second order** (**ObservationA**)
  - - using $\tau = 10h$, stable, but only **first order** (**ObservationB**)

What is wrong ??

The first order "barrier"

• Can we improve the error in time from $O(\tau)$ to $O(\tau^2)$ or better ?

• **(Godunov's Th.)** For $u_t + bu_x = 0$, monotone linear schemes $u_i^{n+1} = \sum a_j u_{i+j}^n$ ($a_j \geq 0$) are limited to at most first order.

• For other reasons (see later on..) for non linear diffusion problems some limitations hold, and we are led to try implicit schemes.

• Let us try a Crank - Nicolson scheme (RK2):

$$S_j^{1,n}(u) := \min\left(\frac{u_i^{n+1} - u_i^n}{\tau} + \frac{1}{2}(Au^n + Au^{n+1})_i,\ u_i^{n+1} - g_i\right) = 0, \quad 1 \leq i \leq J$$

For A: M-matrix, we can solve it by a Newton's Algorithm.

• Numerical experiments show

- - using $\tau = h$, stable, **second order** (**ObservationA**)
- - using $\tau = 10h$, stable, but only **first order** (**ObservationB**)

What is wrong ??

The first order "barrier"

- Can we improve the error in time from $O(\tau)$ to $O(\tau^2)$ or better ?
- **(Godunov's Th.)** For $u_t + bu_x = 0$, monotone linear schemes $u_i^{n+1} = \sum a_j u_{i+j}^n$ ($a_j \geq 0$) are limited to at most first order.
- For other reasons (see later on..) for non linear diffusion problems some limitations hold, and we are led to try implicit schemes.

- Let us try a Crank - Nicolson scheme (RK2):

$$S_j^{1,n}(u) := \min\left(\frac{u_i^{n+1} - u_i^n}{\tau} + \frac{1}{2}(Au^n + Au^{n+1})_i,\ u_i^{n+1} - g_i\right) = 0, \quad 1 \leq i \leq J$$

For A: M-matrix, we can solve it by a Newton's Algorithm.

- Numerical experiments show
  - - using $\tau = h$, stable, **second order** (**ObservationA**)
  - - using $\tau = 10h$, stable, but only **first order** (**ObservationB**)

What is wrong ??

The first order "barrier"

- Can we improve the error in time from $O(\tau)$ to $O(\tau^2)$ or better ?
- **(Godunov's Th.)** For $u_t + bu_x = 0$, monotone linear schemes $u_i^{n+1} = \sum a_j u_{i+j}^n$ ($a_j \geq 0$) are limited to at most first order.
- For other reasons (see later on..) for non linear diffusion problems some limitations hold, and we are led to try implicit schemes.

- Let us try a Crank - Nicolson scheme (RK2):

$$S_j^{1,n}(u) :\equiv \min\left(\frac{u_i^{n+1} - u_i^n}{\tau} + \frac{1}{2}(Au^n + Au^{n+1})_i,\ u_i^{n+1} - g_i\right) = 0,$$

$$1 \leq i \leq J$$

For A: M-matrix, we can solve it by a Newton's Algorithm.

- Numerical experiments show
  - - using $\tau = h$, stable, **second order** (**ObservationA**)
  - - using $\tau = 10h$, stable, but only **first order** (**ObservationB**)

What is wrong ??

The first order "barrier"

• Can we improve the error in time from $O(\tau)$ to $O(\tau^2)$ or better ?
• **(Godunov's Th.)** For $u_t + bu_x = 0$, monotone linear schemes $u_i^{n+1} = \sum a_j u_{i+j}^n$ ($a_j \geq 0$) are limited to at most first order.
• For other reasons (see later on..) for non linear diffusion problems some limitations hold, and we are led to try implicit schemes.

• Let us try a Crank - Nicolson scheme (RK2):

$$S_j^{1,n}(u) :\equiv \min\left(\frac{u_i^{n+1}-u_i^n}{\tau} + \frac{1}{2}(Au^n + Au^{n+1})_i,\ u_i^{n+1} - g_i\right) = 0,$$

$$1 \leq i \leq J$$

For A: M-matrix, we can solve it by a Newton's Algorithm.
• Numerical experiments show

- - using $\tau = h$, stable, **second order** (**ObservationA**)
- - using $\tau = 10h$, stable, but only **first order** (**ObservationB**)

What is wrong ??

The first order "barrier"

• Can we improve the error in time from $O(\tau)$ to $O(\tau^2)$ or better ?
• **(Godunov's Th.)** For $u_t + bu_x = 0$, monotone linear schemes $u_i^{n+1} = \sum a_j u_{i+j}^n$ ($a_j \geq 0$) are limited to at most first order.
• For other reasons (see later on..) for non linear diffusion problems some limitations hold, and we are led to try implicit schemes.

• Let us try a Crank - Nicolson scheme (RK2):

$$S_j^{1,n}(u) :\equiv \min\left(\frac{u_i^{n+1} - u_i^n}{\tau} + \frac{1}{2}(Au^n + Au^{n+1})_i,\ u_i^{n+1} - g_i\right) = 0, \quad 1 \leq i \leq J$$

For A: M-matrix, we can solve it by a Newton's Algorithm.
• Numerical experiments show

- - using $\tau = h$, stable, **second order** (**ObservationA**)
- - using $\tau = 10h$, stable, but only **first order** (**ObservationB**)

What is wrong ??

Consistency analysis:

Notice that for a regular v and $v_i^n = v(t_n, x_i)$:

$$\frac{v_i^{n+1} - v_i^n}{\tau} + \frac{1}{2}(Av^n + Av^{n+1})_i$$
$$= v_t(t_{n+1/2}, x_i) + Av(t_{n+1/2}, x_i) + O(\tau^2)$$
$$= (v_t - v_{xx})(t_{n+1/2}, x_i) + O(\tau^2 + h^2)$$

Hence

$$S_j^{1,n}(v) = \min((v_t - v_{xx})_i^{n+1/2} + O(\tau^2 + h^2),\ v_i^{n+1} - g_i)$$
$$= \min((v_t - v_{xx})_i^{n+1/2},\ v_i^{n+1} - g_i) + O(\tau^2 + h^2)$$

But we cannot assume that $v_t - v_{xx} \equiv 0$!

In general

$$S_j^{1,n}(v) = \min((v_t - v_{xx})_i^{n+1/2},\ v_i^{n+1} - g_i) + O(\tau^2 + h^2)$$
$$= \min((v_t - v_{xx})_i^{n+1} - \frac{\tau}{2}\partial_t(v_t - v_{xx}),,\ v_i^{n+1} - g_i) + O(\tau^2 + h^2)$$

so we have in general **only first order consistency**, in the sense

$$S_j^{1,n}(v) = \min(v_t - v_{xx}, v - \varphi(x_j))_{(t_{n+1},x_i)} + O(\tau + h^2)$$

More generally, consider a high order scheme for the diffusion part, i.e. such that, for instance,

$$\frac{v^{n+1} - S(u^n)}{\tau} = (v_t - v_{xx})(t_n, x_i) + O(\tau^q) + O(h^2)$$

for some $q \geq 2$. (Ex: Weak Taylor schemes, Platen's scheme, Kloeden and Platen 1995,...)
The corresponding scheme for the obstacle equation,

$$u_i^{n+1} = \max(S(u^n)_i, g_i)$$

and equivalent to

$$\min\left(\frac{v^{n+1} - S(u^n)}{\tau},\ u_i^{n+1} - g_i\right) = 0$$

will **not be high order** in time for the obstacle PDE.

Then how the CN scheme could be of second order ?!!

Looking at PDE (2), an other scheme can be (Scheme (2) for PDE (2))

$$\mathcal{S}_j^{2,n} := \min\left(\frac{u_j^{n+1} - u_j^n}{\tau} + \frac{1}{2}(Au^{n+1} + Au^n)_j,\ \frac{u_j^{n+1} - u_j^n}{\tau} \right) = 0 \qquad 1 \le j \le J \tag{4}$$

This is second order in time consistent, at time $t = t_{n+1/2}$!

Numerical experiments show:

- - using $\tau = h$, stable, **second order**
- - using $\tau = 10h$, switch to **first order** (**ObservationC**)

Some explanations...

Related to observations (**A**) and (**B**):

Lemma

(i) If $u^{n+1} \geq u^n$ for the scheme (1) (CN-obstacle scheme for PDE (1), with the constraint $u_j^n \geq g_j$), then u^{n+1} is also solution of the scheme (2) : hence both schemes give identical values.

(ii) This is the case in particular when
- the matrix $I - \frac{\tau}{2}A$ is positive componentwise (which is the case under an appropriate CFL condition of the form $\frac{\tau}{h^2} \leq c_0$)
- the matrix $I + \frac{\tau}{2}A$ is an M-matrix (always the case here).

Related to observation (**C**): the lost of the second order behavior can be explained by the fact that the CN scheme does not have very nice stability behavior for bad CFL numbers.

Some explanations...

Related to observations (**A**) and (**B**):

Lemma

(i) If $u^{n+1} \geq u^n$ for the scheme (1) (CN-obstacle scheme for PDE (1), with the constraint $u_j^n \geq g_j$), then u^{n+1} is also solution of the scheme (2) : hence both schemes give identical values.

(ii) This is the case in particular when
- the matrix $I - \frac{\tau}{2}A$ is positive componentwise (which is the case under an appropriate CFL condition of the form $\frac{\tau}{h^2} \leq c_0$)
- the matrix $I + \frac{\tau}{2}A$ is an M-matrix (always the case here).

Related to observation (**C**): the lost of the second order behavior can be explained by the fact that the CN scheme does not have very nice stability behavior for bad CFL numbers.

III. A BDF Scheme

In order to get the right consistency error, we use the following 3-point formula:

$$\frac{3v_i^{n+1} - 4v_i^n + v_i^{n-1}}{2\tau} \simeq v_t(t_{n+1}, x_i) + O(\tau^2)$$

Known also as a "Backward Difference Formula" of second order

Hence the corresponding implicit BDF2 scheme, for $n \geq 1$:

$$\mathcal{H}_j^{n+1}(u) :\equiv \min\left(\frac{3u_j^{n+1} - 4u_j^n + u_j^{n-1}}{2\tau} + (Au^{n+1})_j,\ u_j^{n+1} - g_j\right) = 0$$

$$1 \leq j \leq J$$

- Multi Step scheme, needs u^0 and u^1 estimate to start
- For the linear part, known also as the "Gear" Scheme
- Scheme already proposed by Oosterlee (2003) together with multigrid idea for 2d problems.

III. A BDF Scheme

In order to get the right consistency error, we use the following 3-point formula:

$$\frac{3v_i^{n+1} - 4v_i^n + v_i^{n-1}}{2\tau} \simeq v_t(t_{n+1}, x_i) + O(\tau^2)$$

Known also as a "Backward Difference Formula" of second order

Hence the corresponding implicit BDF2 scheme, for $n \geq 1$:

$$\mathcal{H}_j^{n+1}(u) :\equiv \min\left(\frac{3u_j^{n+1} - 4u_j^n + u_j^{n-1}}{2\tau} + (Au^{n+1})_j,\ u_j^{n+1} - g_j\right) = 0$$

$$1 \leq j \leq J$$

- Multi Step scheme, needs u^0 and u^1 estimate to start
- For the linear part, known also as the "Gear" Scheme
- Scheme already proposed by Oosterlee (2003) together with multigrid idea for 2d problems.

Proposition

(i) Scheme is second-order consistent in time and space.
(ii) Scheme is fastly implementable for $B = A + \frac{2\tau}{3}$: M-matrix.

Proof of (ii): Equivalent scheme:

$$\min\left((I + \frac{2\tau}{3}A)u^{n+1} - \frac{4}{3}u^n + \frac{1}{3}u^{n-1},\ u^{n+1} - g\right) = 0$$

$$1 \leq j \leq J$$

• Other temptative consistent schemes can be proposed, but the miracle here is that BDF2 for the obstacle problem can be shown to be stable - and convergent ! (although non monotone)

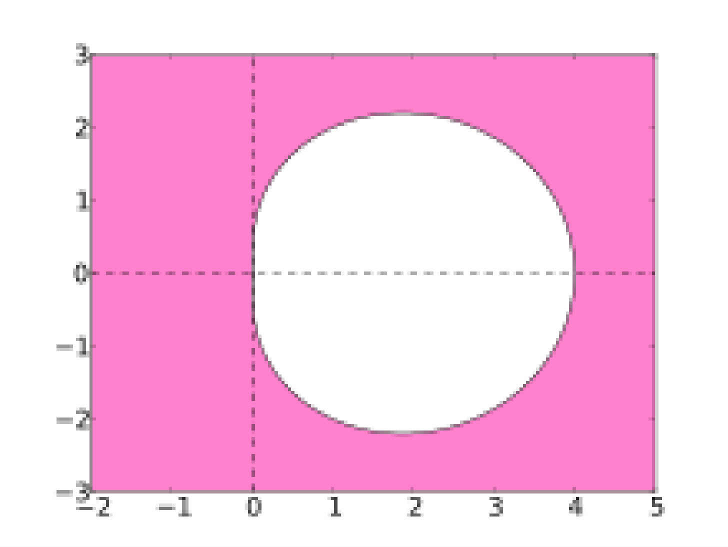
***A*-stability :** for the ODE

$$\dot{y} = \lambda y$$

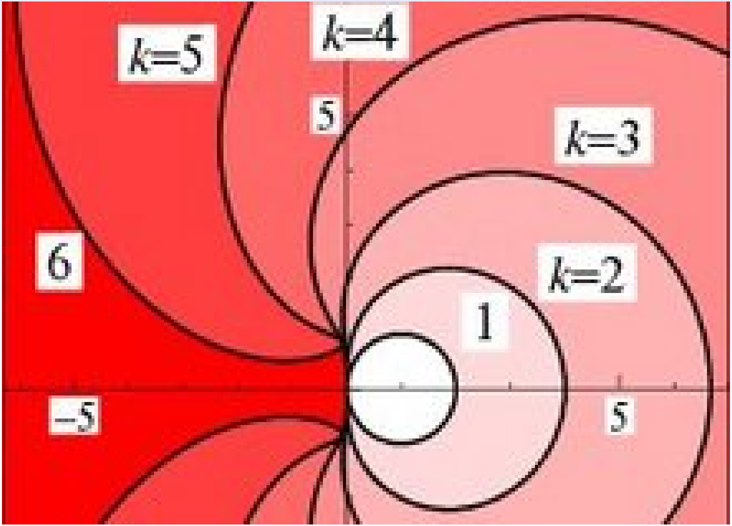
we look for which $\lambda \in \mathbb{C}$ the scheme is stable. If it is stable for all $Re(\lambda) \leq 0$ then we say the scheme is A-stable.

- BDF2 is A-stable
- BDFk , $k \geq 3$ are not A-stable.

Wikipedia - Stability region for BDF2:



Ernst Hairer and Gerhard Wanner (2010), Scholarpedia, 5(4):4591.



Numerical experiments show:

- using $\tau = h$: stable, second order
- using $\tau = 10h$: (incondtionnally) stable, second order

$\Rightarrow$ very efficient !

STABILITY AND CONVERGENCE RESULT

For a given vector $x = (x_j)_{1 \le j \le J}$, let

$$\|x\|_2 = (\sum_{j=1}^{J} |x_j|^2)^{1/2} \tag{5}$$

$$N_h(x) := \left(\sum_{j=1}^{J+1} |\frac{x_j - x_{j-1}}{h}|^2 \right)^{1/2} \tag{6}$$

(with the convention $x_0 := 0$ and $x_{J+1} := 0$).

• Assume the following coercivity on A: there exists constants $\eta > 0$, $\gamma \ge 0$ such that:

$$\langle e, Ae \rangle \ge \eta N_h(e)^2 - \gamma \|e\|_2^2. \tag{7}$$

• Coercivity holds for the matrix A coming from the F.D. approximation of $\mathcal{A}v = -a(x)v_{xx} + b(x)v_x + c(x)v$ with $a(x) \ge \eta > 0$ and $a \in Lip$, $b, c \in L^\infty$.

Scheme

$$\mathcal{H}_j^{n+1}(u) :\equiv \min\left(\frac{3u_j^{n+1} - 4u_j^n + u_j^{n-1}}{2\tau} + (Au^{n+1})_j,\ u_j^{n+1} - g_j\right) = 0$$

Perturbed scheme for the exact solution $v_j^n = v(t_n, x_j)$:

$$\mathcal{H}_j^{n+1}(u) :\equiv \min\left(\frac{3v_j^{n+1} - 4v_j^n + v_j^{n-1}}{2\tau} + (Av^{n+1})_j - \epsilon_j^n,\ v_j^{n+1} - g_j\right) = 0$$

Error: $e^n := v^n - u^n$

Proposition (Stability estimate)

Let $\tau > 0$ be sufficiently small. There exists C_1 independent of n, for all $t_n \leq T$

$$\|e^n\|_2^2 + \tau \sum_{k=1}^{n} N_h(e^k)^2 \leq C_1\left(\|e^0\|_2^2 + \|e^1\|_2^2 + \tau \sum_{k=1,\ldots,n} \|\epsilon_n\|_2^2\right).$$

For solution v with bounded v_{xx} and v_t (typical regularity for $t > 0$), with an "isolated" singularity $x = x_s(t)$ we can show that $\epsilon_j^n = O(1)$ in worst case (otherwise high-order), therefore $\|\epsilon_n\|_2^2 = O(1)$ and in the end, $\|e^n\|_2^2 = O(1)$.
As a consequence:

Theorem

For the BDF2 implicit obstacle scheme, it holds

$$\|e^n\|_{L^2} \simeq (h\|e^n\|_2^2)^{1/2} = O(h^{1/2})$$

STABILITY : ELEMENTS OF THE PROOF

Let $\langle x, y\rangle$ denote the usual scalar product on $\mathbb{R}^J$.

Lemma

For any matrix B, the following equivalence holds:

$$\min(Bx-b, x-g)=0 \quad \Leftrightarrow \quad x \geq g \text{ and } \Big(\langle Bx-b, v-x\rangle \geq 0,\ \forall v \geq g\Big)$$

Proof in the case B is a positive definite symmetric matrix:

$$\begin{aligned}\min(Bx-b, x-g)=0 \quad &\Leftrightarrow \quad x \text{ solves } \min_{x\geq g} \frac{1}{2}\langle x, Bx\rangle - \langle b, x\rangle \\ &\Leftrightarrow \quad x \geq g \text{ and } \Big(\langle Bx-b, v-x\rangle \geq 0,\ \forall v \geq g\Big)\end{aligned}$$

Let

$$B := I + \frac{2\tau}{3}A,$$

and vectors

$$b_u := \frac{4}{3}u^n - \frac{1}{3}u^{n-1}$$
$$b_v := \frac{4}{3}v^n - \frac{1}{3}v^{n-1}$$

Equation in u^{n+1} is equivalent to

$$\min(Bu^{n+1} - b_u, u^{n+1} - g) = 0$$

By the Lemma, this is equivalent to $u^{n+1} \geq g$ and

$$\left\langle Bu^{n+1} - b_u,\ w - u^{n+1} \right\rangle \geq 0, \quad \forall w \geq g. \tag{8}$$

Similarly, equation for v^{n+1} is equivalent to $v^{n+1} \geq g$ and

$$\left\langle Bv^{n+1} - (b_v + \tau\epsilon^n),\ w - v^{n+1} \right\rangle \geq 0, \quad \forall w \geq g. \tag{9}$$

Plugging $w = v^{n+1}$ in (8) and $w = u^{n+1}$ into (9) gives

$$\left\langle Be^{n+1} - \frac{4}{3}e^n + \frac{1}{3}e^{n-1} - \frac{2\tau}{3}\epsilon^n,\ e^{n+1}\right\rangle \leq 0$$

and therefore

$$\left\langle 3e^{n+1} - 4e^n + e^{n-1},\ e^{n+1}\right\rangle + 2\tau\langle e^{n+1}, Ae^{n+1}\rangle \leq 2\tau\langle \epsilon^n,\ e^{n+1}\rangle.$$

Using the coercivity of A, the argument is then technical but "classical" as for the analysis of BDF2 scheme for a parabolic linear problem.

DONE !

Plugging $w = v^{n+1}$ in (8) and $w = u^{n+1}$ into (9) gives

$$\left\langle Be^{n+1} - \frac{4}{3}e^n + \frac{1}{3}e^{n-1} - \frac{2\tau}{3}\epsilon^n,\ e^{n+1}\right\rangle \leq 0$$

and therefore

$$\left\langle 3e^{n+1} - 4e^n + e^{n-1},\ e^{n+1}\right\rangle + 2\tau\langle e^{n+1}, Ae^{n+1}\rangle \leq 2\tau\langle \epsilon^n,\ e^{n+1}\rangle.$$

Using the coercivity of A, the argument is then technical but "classical" as for the analysis of BDF2 scheme for a parabolic linear problem.

DONE !

IV. Numerical examples

MODEL TEST PROBLEMS:

$$\min\left(v_t - \frac{\lambda^2}{2}x^2 v_{xx} - rxv_x + rv,\ v - \varphi(x)\right) = f(t,x)$$
$$v(0,x) = \varphi(x).$$

- Set $x_s(t) := K - c_0\sqrt{t}$
- We construct explicit $v(t,x)$ on $x \in [0, X_{max}]$ s.t.

  (*i*) $v(t,x) = \varphi(x) = K - x$ for $x \leq x_s(t)$,

  (*ii*) $v(t,x) > \varphi(x) = \max(K-x, 0)$ for $x \in]x_s(t), X_{max}]$,

  (*iii*) v is at least C^1 on $[0, X_{max}]$,

  (*iv*) $v(t, X_{max}) = 0$.
- **MODEL 1:** s.t. u_{xx} bounded, regular away from $x = x_s(t)$.
- **MODEL 2:** s.t. u_{3x} bounded, regular away from $x = x_s(t)$.

IV. Numerical examples

MODEL TEST PROBLEMS:

$$\min\left(v_t - \frac{\lambda^2}{2}x^2 v_{xx} - rxv_x + rv,\ v - \varphi(x)\right) = f(t,x)$$
$$v(0,x) = \varphi(x).$$

- Set $x_s(t) := K - c_0\sqrt{t}$
- We construct explicit $v(t,x)$ on $x \in [0, X_{max}]$ s.t.

  (i) $v(t,x) = \varphi(x) = K - x$ for $x \leq x_s(t)$,

  (ii) $v(t,x) > \varphi(x) = \max(K - x, 0)$ for $x \in]x_s(t), X_{max}]$,

  (iii) v is at least C^1 on $[0, X_{max}]$,

  (iv) $v(t, X_{max}) = 0$.
- **MODEL 1:** s.t. u_{xx} bounded, regular away from $x = x_s(t)$.
- **MODEL 2:** s.t. u_{3x} bounded, regular away from $x = x_s(t)$.

IV. Numerical examples

MODEL TEST PROBLEMS:

$$\min\left(v_t - \frac{\lambda^2}{2}x^2 v_{xx} - rxv_x + rv,\ v - \varphi(x)\right) = f(t,x)$$
$$v(0,x) = \varphi(x).$$

- Set $x_s(t) := K - c_0\sqrt{t}$
- We construct explicit $v(t,x)$ on $x \in [0, X_{max}]$ s.t.

(i) $v(t,x) = \varphi(x) = K - x$ for $x \le x_s(t)$,

(ii) $v(t,x) > \varphi(x) = \max(K - x, 0)$ for $x \in]x_s(t), X_{max}]$,

(iii) v is at least C^1 on $[0, X_{max}]$,

(iv) $v(t, X_{max}) = 0$.

- **MODEL 1:** s.t. u_{xx} bounded, regular away from $x = x_s(t)$.
- **MODEL 2:** s.t. u_{3x} bounded, regular away from $x = x_s(t)$.

IV. Numerical examples

MODEL TEST PROBLEMS:

$$\min\left(v_t - \frac{\lambda^2}{2}x^2 v_{xx} - rxv_x + rv,\ v - \varphi(x)\right) = f(t,x)$$

$$v(0,x) = \varphi(x).$$

- Set $x_s(t) := K - c_0\sqrt{t}$
- We construct explicit $v(t,x)$ on $x \in [0, X_{max}]$ s.t.

(*i*) $v(t,x) = \varphi(x) = K - x$ for $x \leq x_s(t)$,

(*ii*) $v(t,x) > \varphi(x) = \max(K - x, 0)$ for $x \in]x_s(t), X_{max}]$,

(*iii*) v is at least C^1 on $[0, X_{max}]$,

(*iv*) $v(t, X_{max}) = 0$.

- **MODEL 1:** s.t. u_{xx} bounded, regular away from $x = x_s(t)$.
- **MODEL 2:** s.t. u_{3x} bounded, regular away from $x = x_s(t)$.

IV. Numerical examples

MODEL TEST PROBLEMS:

$$\min\left(v_t - \frac{\lambda^2}{2}x^2 v_{xx} - rxv_x + rv,\ v - \varphi(x)\right) = f(t,x)$$
$$v(0,x) = \varphi(x).$$

- Set $x_s(t) := K - c_0\sqrt{t}$
- We construct explicit $v(t,x)$ on $x \in [0, X_{max}]$ s.t.

(*i*) $v(t,x) = \varphi(x) = K - x$ for $x \leq x_s(t)$,

(*ii*) $v(t,x) > \varphi(x) = \max(K-x, 0)$ for $x \in]x_s(t), X_{max}]$,

(*iii*) v is at least C^1 on $[0, X_{max}]$,

(*iv*) $v(t, X_{max}) = 0$.

- **MODEL 1:** s.t. u_{xx} bounded, regular away from $x = x_s(t)$.
- **MODEL 2:** s.t. u_{3x} bounded, regular away from $x = x_s(t)$.

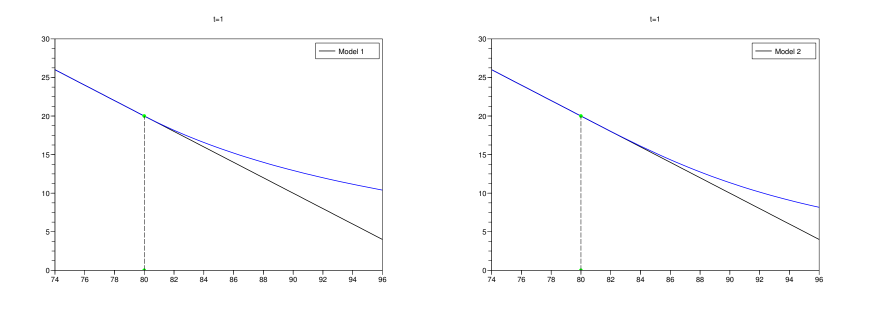


Figure: Zooming around the singular point $(x_s, g(x_s))$ for model 1 (left) and 2 (right).

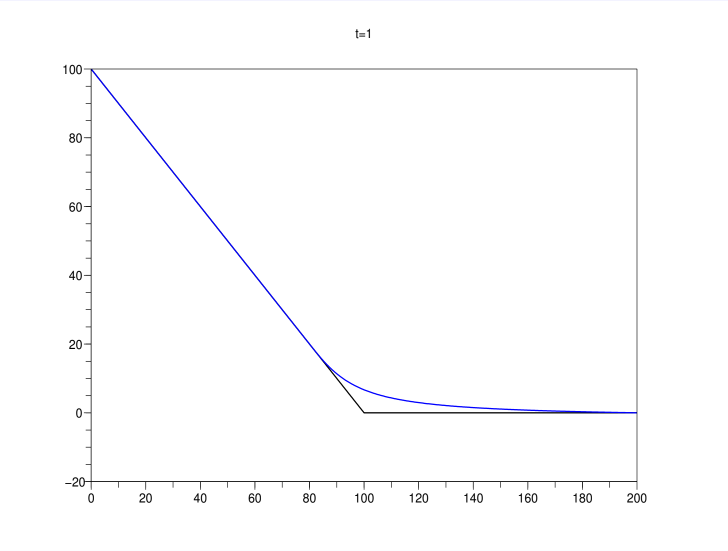


Figure: Model 2.

Numerical examples

Crank Nicolson schemes

Mesh		Error L^1		Error L^2		Error L^∞		CPU time
J	N	error	order	error	order	error	order	
80	80	1.74E-02	-	2.16E-02	-	4.49E-02	-	0.14
160	160	3.15E-03	2.47	3.73E-03	2.54	5.25E-03	3.10	0.25
320	320	8.42E-04	1.91	9.87E-04	1.92	1.39E-03	1.92	0.52
640	640	2.23E-04	1.92	2.58E-04	1.93	3.56E-04	1.97	1.54
1280	1280	6.06E-05	1.88	6.96E-05	1.89	9.15E-05	1.96	6.65
2560	2560	1.66E-05	1.86	1.91E-05	1.86	2.69E-05	1.77	37.66

Table: EDP (1), Crank-Nicolson scheme with $N = J$ ($\tau \equiv h$)

Mesh		Error L^1		Error L^2		Error L^∞		time(s)
J	N	error	order	error	order	error	order	
80	8	1.83E-02	-	2.47E-02	-	4.90E-02	-	0.05
160	16	4.67E-03	1.97	7.23E-03	1.77	2.18E-02	1.17	0.07
320	32	1.68E-03	1.48	2.97E-03	1.28	9.90E-03	1.14	0.13
640	64	6.89E-04	1.28	1.38E-03	1.11	4.75E-03	1.06	0.38
1280	128	2.99E-04	1.20	6.59E-04	1.06	2.31E-03	1.04	1.99
2560	256	1.37E-04	1.13	3.22E-04	1.03	1.14E-03	1.02	16.04

Table: EDP (1), Crank-Nicolson scheme with $N = J/10$ ($\tau \gg h$)

Mesh		Error L^1		Error L^2		Error L^∞		time(s)
J	N	error	order	error	order	error	order	
80	8	2.68E-02	-	3.11E-02	-	4.90E-02	-	0.25
160	16	6.91E-03	1.96	8.28E-03	1.91	1.14E-02	2.11	0.13
320	32	2.53E-03	1.45	2.97E-03	1.48	3.91E-03	1.54	0.18
640	64	9.35E-04	1.43	1.11E-03	1.43	1.50E-03	1.38	0.41
1280	128	3.74E-04	1.32	4.44E-04	1.32	6.07E-04	1.31	1.66
2560	256	1.56E-04	1.26	1.86E-04	1.26	2.56E-04	1.25	11.36

Table: EDP (2) - Crank-Nicolson scheme, $N = J/10$

BDF schemes (American option problem)

BDF2 implicit obstacle scheme:

$$\min\left(\frac{3u_j^{n+1}-4u_j^n+u_j^{n-1}}{2\tau}+(Au^{n+1}+q(t_{n+1}))_j,\ u_j^{n+1}-g_j\right)=0$$

Mesh		Error L^1		Error L^2		Error L^∞		time(s)
J	N	error	order	error	order	error	order	
80	80	1.72E-02	-	2.14E-02	-	4.48E-02	-	0.13
160	160	3.02E-03	2.51	3.57E-03	2.58	5.06E-03	3.15	0.26
320	320	7.93E-04	1.93	9.30E-04	1.94	1.32E-03	1.93	0.49
640	640	2.06E-04	1.94	2.40E-04	1.96	3.39E-04	1.97	1.54
1280	1280	5.46E-05	1.92	6.30E-05	1.93	8.75E-05	1.95	6.43
2560	2560	1.44E-05	1.93	1.67E-05	1.92	2.64E-05	1.73	36.70

Table: BDF2-implicit scheme ($N = J$)

Mesh		Error L^1		Error L^2		Error L^∞		time(s)
J	N	error	order	error	order	error	order	
80	8	8.23E-03	-	1.25E-02	-	3.59E-02	-	0.06
160	16	9.61E-04	3.10	1.28E-03	3.29	2.20E-03	4.02	0.09
320	32	4.18E-04	1.20	5.41E-04	1.24	8.83E-04	1.32	0.13
640	64	1.54E-04	1.44	1.92E-04	1.49	2.99E-04	1.56	0.32
1280	128	4.77E-05	1.69	5.82E-05	1.73	8.71E-05	1.78	1.14
2560	256	1.14E-05	2.07	1.39E-05	2.06	2.01E-05	2.12	6.21

Table: BDF2-implicit scheme, with high CFL numbers.

BDF3 implicit obstacle scheme:

$$\min\left(\frac{\frac{11}{6}u_j^{n+1} - 3u_j^n + \frac{3}{2}u_j^{n-1} - \frac{1}{3}u_j^{n-2}}{\tau} + (Au^{n+1})_j,\right.$$

$$\left. u_j^{n+1} - g_j\right) = 0, \quad \text{for } n \geq 2$$

- No stability estimate !
- test on a more regular data (MODEL 2) to see the order
- changing A matrix to get 4-th order approximation in space
- Use **smoother initial data** $\varphi(x) = v(t_0, x)$, $t_0 > 0$ to avoid initial singularity

Mesh		Error L^1		Error L^2		Error L^∞		time(s)
J	N	error	order	error	order	error	order	
80	8	2.56E+00	-	5.93E-01	-	1.94E-01	-	0.10
160	16	6.48E-01	1.98	1.48E-01	2.00	5.11E-02	1.93	0.11
320	32	1.77E-01	1.87	4.00E-02	1.89	1.33E-02	1.95	0.24
640	64	4.72E-02	1.90	1.06E-02	1.92	3.41E-03	1.96	0.48
1280	128	1.25E-02	1.92	2.77E-03	1.94	8.60E-04	1.99	1.28
2560	256	3.24E-03	1.95	7.14E-04	1.96	2.16E-04	2.00	4.64
5120	512	8.06E-04	2.01	1.78E-04	2.00	5.39E-05	2.00	19.30

Table: (Model 1) BDF2 implicit scheme with 4th order spatial approximation, using high CFL numbers.

Mesh		Error L^1		Error L^2		Error L^∞		time(s)
J	N	error	order	error	order	error	order	
80	8	2.50E+00	-	5.99E-01	-	2.09E-01	-	0.09
160	16	6.42E-01	1.96	1.48E-01	2.01	5.34E-02	1.96	0.13
320	32	1.69E-01	1.93	3.85E-02	1.95	1.35E-02	1.99	0.24
640	64	4.41E-02	1.94	9.96E-03	1.95	3.36E-03	2.00	0.53
1280	128	1.15E-02	1.94	2.57E-03	1.96	8.43E-04	2.00	1.66
2560	256	2.96E-03	1.96	6.55E-04	1.97	2.11E-04	2.00	7.27
5120	512	7.33E-04	2.01	1.63E-04	2.01	5.28E-05	2.00	39.23

Table: (Model 1) BDF3 - third order - implicit scheme with 4th order spatial approximation, using high CFL numbers.

Mesh		Error L^1		Error L^2		Error L^∞		time(s)
J	N	error	order	error	order	error	order	
80	8	2.53E-01	-	4.23E-02	-	9.90E-03	-	0.10
160	16	5.02E-02	2.34	8.54E-03	2.31	2.02E-03	2.29	0.15
320	32	1.09E-02	2.20	1.86E-03	2.20	4.37E-04	2.21	0.26
640	64	2.58E-03	2.08	4.39E-04	2.08	1.04E-04	2.07	0.60
1280	128	6.24E-04	2.05	1.06E-04	2.05	2.50E-05	2.05	1.60
2560	256	1.54E-04	2.02	2.62E-05	2.02	6.16E-06	2.02	5.65
5120	512	3.81E-05	2.01	6.48E-06	2.01	1.53E-06	2.01	24.37

Table: (Model 2) **BDF2** implicit scheme with 4th order spatial approximation, using high CFL numbers, and $t_0 = 0.3$. (SMOOTHER INITIAL DATA)

Mesh		Error L^1		Error L^2		Error L^∞		time(s)
J	N	error	order	error	order	error	order	
80	8	2.51E-02	-	4.42E-03	-	1.21E-03	-	0.11
160	16	3.65E-03	2.78	6.48E-04	2.77	1.76E-04	2.78	0.14
320	32	6.00E-04	2.61	1.04E-04	2.64	2.62E-05	2.75	0.25
640	64	8.56E-05	2.81	1.46E-05	2.83	3.55E-06	2.89	0.65
1280	128	1.09E-05	2.97	1.83E-06	3.00	4.40E-07	3.01	2.11
2560	256	1.42E-06	2.94	2.32E-07	2.98	5.52E-08	2.99	8.96
5120	512	1.92E-07	2.89	3.02E-08	2.94	7.00E-09	2.98	47.05

Table: (Model 2) **BDF3** implicit scheme with 4th order spatial approximation, using high CFL numbers, and $t_0 = 0.3$. (SMOOTHER INITIAL DATA)

CONCLUSION

Errors in terms of number of operations N_{opt}:

BDF2 using $N \equiv J$:

- $N_{opt} \equiv N \cdot J \equiv N^2$,
- observed error : $err = O(\tau^2) = O(\frac{1}{N^2}) = O(\frac{1}{N_{opt}})$.
- theoretical bound : $err_{L^2} = O(h^{1/2}) = O(\frac{1}{N_{opt}^{1/4}})$.

Tree method with N levels (Lamberton 2015), constant coefficient case:

- N levels $\Rightarrow N_{opt} \equiv N^2/2$
- observed error : $err = O(\frac{1}{N}) = O(\frac{1}{N_{opt}^{1/2}})$.
- theoretical bound : $err_{L^\infty} = O(\frac{log(N)^\alpha}{N}) \equiv O(\frac{log(N_{opt})^\alpha}{N_{opt}^{1/2}})$

CONCLUSION

Errors in terms of number of operations N_{opt}:

BDF2 using $N \equiv J$:

- $N_{opt} \equiv N \cdot J \equiv N^2$,
- observed error : $err = O(\tau^2) = O(\frac{1}{N^2}) = O(\frac{1}{N_{opt}})$.
- theoretical bound : $err_{L^2} = O(h^{1/2}) = O(\frac{1}{N_{opt}^{1/4}})$.

Tree method with N levels (Lamberton 2015), constant coefficient case:

- N levels $\Rightarrow N_{opt} \equiv N^2/2$
- observed error : $err = O(\frac{1}{N}) = O(\frac{1}{N_{opt}^{1/2}})$.
- theoretical bound : $err_{L^\infty} = O(\frac{log(N)^\alpha}{N}) \equiv O(\frac{log(N_{opt})^\alpha}{N_{opt}^{1/2}})$

CONCLUSION

Errors in terms of number of operations N_{opt}:

BDF2 using $N \equiv J$:

- $N_{opt} \equiv N \cdot J \equiv N^2$,
- observed error : $err = O(\tau^2) = O(\frac{1}{N^2}) = O(\frac{1}{N_{opt}})$.
- theoretical bound : $err_{L^2} = O(h^{1/2}) = O(\frac{1}{N_{opt}^{1/4}})$.

Tree method with N levels (Lamberton 2015), constant coefficient case:

- N levels $\Rightarrow N_{opt} \equiv N^2/2$
- observed error : $err = O(\frac{1}{N}) = O(\frac{1}{N_{opt}^{1/2}})$.
- theoretical bound : $err_{L^\infty} = O(\frac{log(N)^\alpha}{N}) \equiv O(\frac{log(N_{opt})^\alpha}{N_{opt}^{1/2}})$

CONCLUSION 2

Related ongoing works:

- Discontinuous Galerkin approaches for obstacle problems (with C.-W Shu, Y. Cheng)
- Implicit schemes for nonlinear diffusion problems (with A. Picarelli, C. Reisinger)

Thanks !

CONCLUSION 2

Related ongoing works:

- Discontinuous Galerkin approaches for obstacle problems (with C.-W Shu, Y. Cheng)
- Implicit schemes for nonlinear diffusion problems (with A. Picarelli, C. Reisinger)

Thanks !