BIT

Check for
updates

A note on optimal H^1-error estimates for Crank-Nicolson approximations to the nonlinear Schrödinger equation

Patrick Henning[1] · Johan Wärnegård[1]

© The Author(s) 2020

Abstract
In this paper we consider a mass- and energy–conserving Crank-Nicolson time discretization for a general class of nonlinear Schrödinger equations. This scheme, which enjoys popularity in the physics community due to its conservation properties, was already subject to several analytical and numerical studies. However, a proof of optimal $L^\infty(H^1)$-error estimates is still open, both in the semi-discrete Hilbert space setting, as well as in fully-discrete finite element settings. This paper aims at closing this gap in the literature. We also suggest a fixed point iteration to solve the arising nonlinear system of equations that makes the method easy to implement and efficient. This is illustrated by numerical experiments.

1 Introduction

In this paper we consider nonlinear Schrödinger equations (NLS) seeking a complex function $u(t, x)$ such that

✉ Patrick Henning
  pathe@kth.se

  Johan Wärnegård
  jwar@kth.se

[1] Department of Mathematics, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden

⚛ Springer

$$i\partial_t u = -\triangle u + V u + \gamma(|u|^2)u$$

in a bounded domain $\mathscr{D} \subset \mathbb{R}^d$, with a homogenous Dirichlet boundary condition on $\partial\mathscr{D}$ and a given initial value. Here, $V(x)$ is a known real-valued potential and $\gamma : [0, \infty) \to \mathbb{R}$ is a smooth (and possibly nonlinear) function that depends on the unknown density $|u|^2$. Of particular interest are cubic nonlinearities of the form $\gamma(|u|^2)u = \kappa|u|^2u$, for some $\kappa \in \mathbb{R}$. In this case, the equation is called Gross–Pitaevskii equation. It has applications in optics [1,2], fluid dynamics [3,4] and, most importantly, in quantum physics, where it models for example the dynamics of Bose-Einstein condensates in a magnetic trapping potential [5–7]. Another relevant class are saturated nonlinearities, such as $\gamma(|u|^2) = \kappa|u|^2(1 + \alpha|u|^2)^{-1}$ for some $\alpha \geq 0$, which appear in the context of nonlinear optical wave propagation in layered metallic structures [8,9] or the propagation of light beams in plasmas [10]. In order to discretize nonlinear Schrödinger equations in time, splitting methods and exponential integrators yield typically highly efficient solution schemes that can be easily combined with a spectral discretization in space (cf. [11–18] and the references therein). If the exact solution to the NLS admits high regularity, such discretization schemes typically show a remarkably good performance. However, if the regularity of the solution is strongly reduced, either by rough potentials V (e.g. disorder potentials or optical lattices) or by rough initial values $u(0)$ (e.g. when effects close to phase transitions are studied), then the performance of these methods can drop dramatically. Here we refer exemplarily to the recent numerical experiments reported in [19–21]. To overcome this issue, Ostermann and Schratz proposed new low-regularity time-integrators [19,20] which improve the convergence in low regularity regimes significantly. However, the approach still relies on a Fourier discretization in space, which is not an optimal choice due to the loss of spectral convergence for non-smooth solutions. Practically, the usage of a (low order) finite element space discretization is often desirable in order to account for spatial low regularity. In the following we will only discuss approaches that can be easily combined with finite elements in space, meaning that we put ourselves into the situation that we assume that the solution to the NLS does not admit much smoothness.

Nonlinear Schrödinger equations come with important physical invariants, where the mass and the energy are considered as two of the most crucial ones. When solving a NLS numerically it is therefore of great importance to also reproduce this conservation on the discrete level. This aspect was emphasized by various numerical studies [21,22], where it was also found that the complexity of the physical setup (or low-regularity) can stress this issue even further .

For the subclass of power law nonlinearities of the form $\gamma(|u|^2) = \sum_{k=1}^{K} \kappa_k|u|^{2\sigma_k}$ for $\sigma_k \geq 0$ and $\alpha_k \in \mathbb{R}$, a mass and energy conserving relaxation scheme was proposed and analyzed by Besse [23,24]. Thanks to its properties, the scheme shows a very good performance in realistic physical setups [21]. Despite the large variety of different numerical approaches for solving the time-dependent NLS (cf. [16–18,25–34] and the references therein) the literature knows however only one time discretization that conserves both mass and energy simultaneously for arbitrary (smooth) nonlinearities. This discretization, which was first mathematically studied by Sanz-Serna [31] and which is long-known in the physics community, is a Crank-Nicolson-type (CN) approach where the nonlinearity is approximated by a suitable difference quotient

involving the primitive integral of γ. This is also the time discretization that we shall consider in this paper. Here we note it was analytically and numerically demonstrated that this method is applicable and reliable in low-regularity regimes [21,35].

A combination of the method with a finite difference space discretization was proposed and analyzed by Bao and Cai [36,37]. Combining the Crank-Nicolson time discretization with a $P1$ finite element discretization in space, the first a priori error estimates for the arising method were obtained by Sanz-Serna 1984 [31] for cubic nonlinearities. He considers the case $d = 1$ and derives optimal $L^\infty(L^2)$-error estimates under the coupling constraint $\tau \lesssim h$, where τ denotes the time step size of the Crank-Nicolson method and h the mesh size of the finite element discretization in space. In 1991, Akrivis et al. [25] improved this result by showing optimal convergence rates in $L^\infty(L^2)$ in dimension $d = 1, 2, 3$ and under the relaxed coupling constraint $\tau \lesssim h^{d/4}$. Finally, in 2017 [35], the $L^\infty(L^2)$-error estimates could be improved yet another time by showing that the coupling constraint can be fully removed. Furthermore, general nonlinearities could be considered, the influence of potentials could be taken into account and even convergence under weak regularity assumptions could be proved (with reduced convergence rates). However, so far, optimal error estimates in $L^\infty(H^1)$ for this particular CN-discretization are still open in the literature.

One reason for this absence of H^1-results could be related to the techniques used for the error analysis in previous works (cf. [25,28–31,34]) which is based on the following steps: *1.* Appropriate truncation of the nonlinearity to obtain a problem with bounded growth. *2.* Analyzing the scheme with truncation in the FE space and deriving corresponding $L^\infty(L^2)$- and/or $L^\infty(H^1)$-error estimates. *3.* Using inverse estimates in the finite element space to show that the truncated approximations are uniformly bounded in $L^\infty(L^\infty)$ by a term of the form $C(1 + h^{-s}(\tau^2 + h^p))$, with appropriate powers $p > s > 0$ that depend on the considered space discretization, regularity and space dimension. *4.* Concluding that if τ and h are coupled in an appropriate way, then the truncated approximations are all uniformly bounded by a constant C and hence coincide with a solution to the scheme without truncation.

This strategy does not only have the disadvantage that it produces unnecessary coupling conditions, but also that it becomes impractically technical when considering $L^\infty(H^1)$-error estimates for the Crank-Nicolson FEM. This is because it requires a suitable truncation of the primitive integral of γ that is on the one hand consistent with the energy conservation and on the other hand allows for uniform bounds of the approximations in $L^\infty(W^{1,\infty})$. However, thanks to the new techniques developed in [33] and the CN error analysis suggested in [35] in the context of $L^\infty(L^2)$-error estimates, the truncation step is no longer necessary and the desired $L^\infty(L^\infty)$-bounds can be derived with elliptic regularity theory. With this, it is now possible to obtain $L^\infty(H^1)$ estimates in a direct way, not only in the finite element setting, but also in the semi-discrete Hilbert space setting.

In this paper we will therefore build upon the results from [33,35] to fill the gap in the literature and prove optimal $L^\infty(H^1)$-error estimates for the energy-conservative Crank-Nicolson approach without coupling constraints and for a general class of nonlinearities. The paper is structured as follows. In Sect. 2 we present the notation and the analytical assumptions on the problem. In Sect. 3 we present the time–discrete Crank-Nicolson method, we recall its well-posedness and optimal error estimates in

$L^\infty(L^2)$. Furthermore, we present and prove the new error estimate in $L^\infty(H^1)$. The paper continues with the fully-discrete setting presented in Sect. 4, where the time discretization is combined with a finite element discretization in space. We recall what is known about this discretization and finally prove corresponding $L^\infty(H^1)$-error estimates, which is the main result of this paper. The paper concludes with a note on how to efficiently implement the method and two numerical experiments to confirm the convergence rates and to illustrate a setting in which it makes computational sense to use the CN-FEM instead of for example a spectral method.

2 Notation and assumptions

We start with introducing the analytical setting of this work. Throughout the paper we assume that $\mathscr{D} \subset \mathbb{R}^d$ (for $d = 2, 3$) is a convex bounded domain with polyhedral boundary. On $\mathscr{D}$, the Sobolev space of complex-valued, weakly differentiable functions with a zero trace on $\partial\mathscr{D}$ and L^2-integratable partial derivatives is as usual denoted by $H_0^1(\mathscr{D}) := H_0^1(\mathscr{D}, \mathbb{C})$. The potential $V \in L^\infty(\mathscr{D}; \mathbb{R})$ is assumed to be real and nonnegative. Indirectly, we also assume that V is sufficiently smooth so that it is compatible with the regularity assumptions for u listed below (see [35] for a discussion on this aspect). The (possibly nonlinear) function

$$\gamma : [0, \infty) \to [0, \infty)$$

is assumed to be two times differentiable, fulfills $\gamma(0) = 0$ and its growth can be characterized with

$$|\gamma(|v|^2)v - \gamma(|w|^2)w| \le L(K)|v - w| \quad \text{for all } v, w \in \mathbb{C} \text{ with } |v|, |w| \le K$$

where L is a function with the following growth properties

$$0 \le L(s) \le Cs^q \quad \text{for } s \ge 0 \quad \text{and} \quad \begin{cases} q \in [0, \infty) & \text{for } d = 2, \\ q \in [0, 4) & \text{for } d = 3. \end{cases}$$

Note that in [35] the admissible growth condition in $3d$ requires $q \in [0, 2)$, which is however a typo and should be, as above, $q \in [0, 4)$ (cf. [38, Proposition 3.2.5 and Remark 3.2.7] for the original result). Examples for nonlinearities that fulfill these assumptions are mentioned in the introduction. The most common and physically relevant choices covered by our setting are power law nonlinearities $\gamma(\rho) = \kappa\rho^q$ for $\kappa \ge 0$ and $0 \le q < \infty$ in $2d$ and $0 \le q < 4$ in $3d$. Other physically relevant nonlinearities that fulfill the conditions are saturated nonlinearities appearing in the modeling of optical wave propagation such as $\gamma(\rho) = \kappa\rho(1 + \alpha\rho)^{-1}$ for $\alpha, \kappa \ge 0$.

The above assumptions cover the regime of so-called defocussing (positive) nonlinearities and guarantees that the NLS and its Crank-Nicolson discretization are well-posed. For focussing (negative) nonlinearities, i.e. $\gamma : [0, \infty) \mapsto (-\infty, 0]$, the well-posedness (of both the continuous and discrete models) can no longer be

guaranteed without making additional technical assumptions. Typically, effects such as finite time blow ups can occur in this regime. To avoid constantly having to invoke a saving clause we restrict our attention to the defocussing case. We do however point out that under the assumptions that the NLS and the Crank-Nicolson discretizations are well-posed (without blow-up in the time interval $[0, T]$) then all our error estimates hold without changes.

For the initial value we assume that $u^0 \in H_0^1(\mathcal{D}) \cap H^2(\mathcal{D})$ and, without loss of generality, that it has a normalized mass, i.e. $\int_{\mathcal{D}} |u^0(x)|^2 \, dx = 1$. With this, the considered nonlinear Schrödinger equation (NLS) reads as follows. For a maximum time $T > 0$ and an initial value u^0, we seek

$$u \in L^\infty([0, T], H_0^1(\mathcal{D})) \quad \text{and} \quad \partial_t u \in L^\infty([0, T], H^{-1}(\mathcal{D}))$$

such that $u(\cdot, 0) = u^0$ and

$$i\partial_t u = -\Delta u + V u + \gamma(|u|^2) u \tag{2.1}$$

in the sense of distributions. Problem (2.1) admits at least one solution, that is even unique for repulsive cubic nonlinearities in $1d$ and $2d$ (cf. [38] in general and [35, Remark 2.1] for precise references). We assume that the solution admits the following additional regularity, which is

$$u_{ttt} \in L^2(0, T; H^1(\mathcal{D}))$$
$$\text{and} \quad u \in L^\infty(0, T; W^{1,\infty}(\mathcal{D})), \quad \partial_t^{(k)} u \in L^2(0, T; H^2(\mathcal{D})) \quad \text{for } 0 \le k \le 2, \tag{2.2}$$

where we note that any solution with such increased regularity must be unique (cf. [35, Lemma 3.1]). In the rest of the paper u hence always refers to this uniquely characterized solution.

It is well known that solutions to the NLS (2.1) preserve the mass, i.e.

$$\int_{\mathcal{D}} |u(t, x)|^2 \, dx = \int_{\mathcal{D}} |u^0(x)|^2 \, dx = 1$$

and the energy, i.e.

$$E[u(t)] = E[u^0], \quad \text{where} \quad E[u] := \frac{1}{2} \int_{\mathcal{D}} |\nabla u(x)|^2 + V(x) |u(x)|^2 + \Gamma(|u(x)|^2) \, dx,$$

with $\Gamma(\rho) := \int_0^\rho \gamma(r) \, dr$.

For brevity, we shall denote the L^2-norm of a function $v \in L^2(\mathcal{D}) := L^2(\mathcal{D}, \mathbb{C})$ by $\|v\|$. The L^2-inner product is denoted by $\langle v, w \rangle = \int_{\mathcal{D}} v(x) \overline{w(x)} \, dx$. Here, $\overline{w}$ denotes the complex conjugate of w.

Throughout the paper we will use the notation $A \lesssim B$, to abbreviate $A \le CB$, where C is a constant that only depends on u, T, d, $\mathcal{D}$, V and γ, but not on the discretization.

Remark 2.1 In the analysis we restrict our attention to homogeneous Dirichlet boundary conditions. Typically these boundary conditions can be motivated by physical reasoning. For example in the context of Bose Einstein condensates, the magnetic potential V is a trapping potential that becomes very quickly very large and hence traps the condensate in a bounded region. Mathematically this leads to an exponential decay of the solution u to zero (in moderate distances from the origin of the coordinate system) and hence justifies to truncate the computational domain to a simple geometric object on which the problem is solved with zero boundary conditions. A typical alternative found in the literature are periodic boundary conditions which are e.g. favorable for spectral methods. Both the formulation of the Crank-Nicolson method and its error analysis can be easily generalized to that case.

3 Time-discrete Crank-Nicolson scheme

In this section we will state the semi-discrete Crank-Nicolson scheme, recall its well-posedness and available stability bounds, and then use these results to prove optimal $L^\infty(H^1)$-error estimates in the Hilbert space setting. For that, let T denote the final time of computation, N the number of time-steps, and $\tau = T/N$ the time step size. By t_n we shall mean $t_n = n\tau$. The exact solution at time t_n shall be denoted by $u^n := u(t_n, \cdot)$. We also introduce a short hand notation for discrete time derivatives which is $D_\tau u^n := (u^{n+1} - u^n)/\tau$ and analogously $D_\tau u_\tau^n := (u_\tau^{n+1} - u_\tau^n)/\tau$.

3.1 Method formulation and main result

With the notation above, the semi-discrete Crank-Nicolson approximation $u_\tau^{n+1} \in H_0^1(\mathscr{D})$ to u^{n+1} is given recursively as the solution (in the sense of distributions) to the equation

$$
i D_\tau u_\tau^n = -\Delta u_\tau^{n+\frac{1}{2}} + V u_\tau^{n+\frac{1}{2}} + \frac{\Gamma(|u_\tau^{n+1}|^2) - \Gamma(|u_\tau^n|^2)}{|u_\tau^{n+1}|^2 - |u_\tau^n|^2} u_\tau^{n+\frac{1}{2}}, \tag{3.1}
$$

where $u_\tau^{n+\frac{1}{2}} := (u_\tau^n + u_\tau^{n+1})/2$. The initial value is selected as $u_\tau^0 = u^0$. It is easily seen that the discretization conserves both mass and energy, i.e.

$$
\int_{\mathscr{D}} |u_\tau^n|^2 \, dx = \int_{\mathscr{D}} |u^0|^2 \, dx \quad \text{and} \quad E[u_\tau^n] = E[u^0] \quad \text{for all } n \geq 0.
$$

The scheme (3.1) is well-posed and admits a set of a priori error estimates. The properties are summarized in the following theorem that is proved in [35, Theorem 4.1].

Theorem 3.1 *Under the general assumptions of this paper, there exists a constant $C(u) > 0$ and a solution $u_\tau^n \in H_0^1(\mathscr{D})$ to the semi-discrete Crank-Nicolson scheme (3.1) that is uniquely characterized by the property that*

$$\sup_{0 \le n \le N} \left(\|u_\tau^n\|_{L^\infty(\mathscr{D})} + \|u_\tau^n\|_{H^2(\mathscr{D})} \right) \le C(u), \tag{3.2}$$

and the a priori estimate for the L^2-error

$$\sup_{0 \le n \le N} \|u_\tau^n - u^n\| \lesssim \tau^2,$$

where u is the (unique) exact solution with the regularity property (2.2).

Our main result on optimal error estimates in the $L^\infty(H^1)$ reads as follows.

Theorem 3.2 (Optimal H^1-error estimates for the semi-discrete method) *Consider the setting of Theorem 3.1, then the $L^\infty(H^1)$-error converges with optimal order in τ, i.e.*

$$\sup_{0 \le n \le N} \|u_\tau^n - u^n\|_{H^1(\mathscr{D})} \lesssim \tau^2.$$

The theorem is proved in Sect. 3.2 below.

3.2 Proof of Theorem 3.2

In this section we will prove Theorem 3.2. Let us introduce some notation that is used throughout the proofs. We recall $D_\tau e^n = (e^{n+1} - e^n)/\tau$. Furthermore, we let $e^{n+1/2} := (e^{n+1} + e^n)/2$ and $u^{n+1/2} := (u^{n+1} + u^n)/2$. For time derivatives at fixed time t^n, we also write $\partial_t u^n := \partial_t u(t^n, \cdot)$.

We begin by establishing a differential equation for the time discrete error $e^n = u^n - u_\tau^n$. This is stated in the following lemma.

Lemma 3.1 (Consistency error) *The error $e^n = u^n - u_\tau^n$ fulfills the identity*

$$i D_\tau e^n + \Delta e^{n+1/2} - V e^{n+1/2} - e_\gamma^n = T^n, \tag{3.3}$$

where the the consistency error T^n is given by

$$T^n := i \left(D_\tau u^n - \partial_t u(t_{n+1/2}) \right) + \Delta(u^{n+1/2} - u(t_{n+1/2})) - V(u^{n+1/2} - u(t_{n+1/2}))$$
$$- \left(\gamma(\xi^n) - \gamma(|u(t_{n+1/2})|^2) \right). \tag{3.4}$$

Here, $e_\gamma^n := \gamma(\xi^n)u^{n+1/2} - \gamma(\xi_\tau^n)u_\tau^{n+1/2}$ for some bounded functions $\xi^n, \xi_\tau^n \in L^\infty(\mathscr{D})$ with the properties that

$$\xi^n(x) \in [\min(|u^n|^2, |u^{n+1}|^2), \max(|u^n|^2, |u^{n+1}|^2)] \quad and$$
$$\xi_\tau^n(x) \in [\min(|u_\tau^n|^2, |u_\tau^{n+1}|^2), \max(|u_\tau^n|^2, |u_\tau^{n+1}|^2)]$$

for almost all $x \in \mathscr{D}$.

Proof It is easily verified that exact solution fulfills

$$iD_\tau u^n + \Delta u^{n+1/2} - V u^{n+1/2} - \frac{\Gamma(|u^{n+1}|^2) - \Gamma(|u^n|^2)}{|u^{n+1}|^2 - |u^n|^2} = T^n. \qquad (3.5)$$

By the regularity assumptions we can apply Taylor expansion arguments to T^n to see:

$$\sum_{k=0}^{N} \|T^k\|^2 \leq C\tau^3 \qquad (3.6)$$

The argument that proves (3.6) is elaborated in "Appendix A", where it becomes also visible how the regularity assumptions enter explicitly in the estimate. Next, subtracting (3.5) from (3.1) we find that $e^n = u^n - u^n_\tau$ satisfies:

$$iD_\tau e^n + \Delta e^{n+1/2} - V e^{n+1/2} - e^n_\gamma = T^n$$

where e^n_γ denotes the error coming from the nonlinear term, defined by

$$e^n_\gamma = \frac{\Gamma(|u^{n+1}|^2) - \Gamma(|u^n|^2)}{|u^{n+1}|^2 - |u^n|^2} u^{n+1/2} - \frac{\Gamma(|u^{n+1}_\tau|^2) - \Gamma(|u^n_\tau|^2)}{|u^{n+1}_\tau|^2 - |u^n_\tau|^2} u^{n+1/2}_\tau.$$

Recalling the definition of Γ we have:

$$\frac{\Gamma(|u^{n+1}|^2) - \Gamma(|u^n|^2)}{|u^{n+1}|^2 - |u^n|^2} = \frac{1}{|u^{n+1}|^2 - |u^n|^2} \int_{|u^n|^2}^{|u^{n+1}|^2} \gamma(r)\, dr =: \gamma(\xi^n),$$

likewise

$$\frac{\Gamma(|u^{n+1}_\tau|^2) - \Gamma(|u^n_\tau|^2)}{|u^{n+1}_\tau|^2 - |u^n_\tau|^2} = \frac{1}{|u^{n+1}_\tau|^2 - |u^n_\tau|^2} \int_{|u^n_\tau|^2}^{|u^{n+1}_\tau|^2} \gamma(r)\, dr =: \gamma(\xi^n_\tau).$$

The expression for e^n_γ is thus simplified to

$$e^n_\gamma := \gamma(\xi^n) u^{n+1/2} - \gamma(\xi^n_\tau) u^{n+1/2}_\tau,$$

where ξ^n is a function taking values between $|u^n|^2$ and $|u^{n+1}|^2$ and ξ^n_τ takes values between $|u^n_\tau|^2$ and $|u^{n+1}_\tau|^2$. $\qquad \square$

The differential equation in Lemma 3.1 is now used to derive a recurrence formula for the H^1-norm of the error. Multiplying (3.3) by $D_\tau e^n$, integrating and taking the real part yields:

$$\frac{\|\nabla e^{n+1}\|^2 - \|\nabla e^n\|^2}{2\tau} = \underbrace{\mathrm{Re}(\langle e^n_\gamma, D_\tau e^n \rangle)}_{\mathrm{I}} - \underbrace{\mathrm{Re}(\langle T^n, D_\tau e^n \rangle)}_{\mathrm{II}}. \qquad (3.7)$$

The idea is to bound the terms I and II in such a way that Grönwall's inequality can be used. We proceed to bound term I. Multiplying the error PDE (3.3) by e_γ^n results in:

$$i\langle D_\tau e^n, e_\gamma^n\rangle = \langle \nabla e^{n+1/2}, \nabla e_\gamma^n\rangle + \langle V e^{n+1/2}, e_\gamma^n\rangle + \|e_\gamma^n\|^2 + \langle T^n, e_\gamma^n\rangle$$

and consequently

$$
\begin{aligned}
|I| &= |\mathrm{Re}(\langle D_\tau e^n, e_\gamma^n\rangle)| \\
&\leq |\mathrm{Im}(\langle \nabla e^{n+1/2}, \nabla e_\gamma^n\rangle)| + |\mathrm{Im}\langle V e^{n+1/2}, e_\gamma^n\rangle| + |\mathrm{Im}(\langle T^n, e_\gamma^n\rangle)| \\
&\lesssim \|\nabla e^{n+1/2}\|^2 + \|\nabla e_\gamma^n\|^2 + \|V\|_\infty(\|e^{n+1/2}\|^2 + \|e_\gamma^n\|^2) + \tau^4 + \|e_\gamma^n\|^2 + \|T^n\|^2 \\
&\lesssim \|\nabla e^{n+1}\|^2 + \|\nabla e^n\|^2 + \|\nabla e_\gamma^n\|^2 + \|e_\gamma^n\|^2 + \|T^n\|^2 + \tau^4.
\end{aligned}
\tag{3.8}
$$

In order to use Grönwall's inequality we need to bound $\|e_\gamma^n\|$ and $\|\nabla e_\gamma^n\|$ in terms of $\|e^n\|, \|\nabla e^n\|$ and terms of $\mathcal{O}(\tau^2)$. These bounds are formulated in the two following lemmas.

Lemma 3.2 *Given the optimal L^2-convergence of Theorem 3.1 and the uniform bounds (3.2), the error coming from the nonlinear term behaves as τ^2, i.e. $\|e_\gamma^n\| \lesssim \tau^2$.*

Proof We introduce the function f by

$$f(a, b) = \frac{1}{b-a}\int_a^b \gamma(r)\, dr.\tag{3.9}$$

The partial derivatives of f with respect to the i'th variable are denoted by $\partial_i f$. A standard application of the mean value theorem yields that

$$
\begin{aligned}
e_\gamma^n &= (\gamma(\xi^n) - \gamma(\xi_\tau^n))u^{n+1/2} + \gamma(\xi_\tau^n)(u^{n+1/2} - u_\tau^{n+1/2}) \\
&= (f(|u^n|^2, |u^{n+1}|^2) - f(|u_\tau^n|^2, |u_\tau^{n+1}|^2))u^{n+1/2} + \gamma(\xi_\tau^n)e^{n+1/2} \\
&= \big(\partial_1 f(\eta^n, \eta^{n+1})(|u^n|^2 - |u_\tau^n|^2) + \partial_2 f(\eta^n, \eta^{n+1})(|u^{n+1}|^2 - |u_\tau^{n+1}|^2)\big)u^{n+1/2} \\
&\quad + \gamma(\xi_\tau^n)e^{n+1/2}
\end{aligned}
$$

for some functions η^n and $\eta^{n=1}$ with

$$\min\{|u^n(x)|^2, |u_\tau^n(x)|^2\} \leq \eta^n(x) \leq \max\{|u^n(x)|^2, |u_\tau^n(x)|^2\}.\tag{3.10}$$

A quick sanity check shows that the partial derivatives of f are bounded by the derivative of γ:

$$
\begin{aligned}
|\partial_1 f(a, b)| &= |\frac{1}{b-a}(\gamma(c) - \gamma(a))| = |\gamma'(\theta^-)\frac{c-a}{b-a}| \\
|\partial_2 f(a, b)| &= |\frac{1}{b-a}(\gamma(b) - \gamma(c))| = |\gamma'(\theta^+)\frac{b-c}{b-a}|,
\end{aligned}
\tag{3.11}
$$

where c and θ lie somehwere between a and b. Hence

$$
\begin{aligned}
\|e_\gamma^n\| \leq{} & \|\partial_1 f(\eta^n, \eta^{n+1})\|_{L^\infty}\|u^{n+1/2}\|_{L^\infty}\||u^n|^2 - |u_\tau^n|^2\| \\
& + \|\partial_2 f(\eta^n, \eta^{n+1})\|_{L^\infty}\|u^{n+1/2}\|_{L^\infty}\||u^{n+1}|^2 - |u_\tau^{n+1}|^2\| \\
& + \|\gamma(\xi_\tau^n)\|_{L^\infty}\|e^{n+1/2}\|.
\end{aligned}
$$

Since $\||u^n|^2 - |u_\tau^n|^2\| \leq \||u^n| + |u^{n+1}|\|_{L^\infty}\||e^n|\|$ it now follows with the max-norm bounds for the exact solution and the semi-discrete solution that $\|e_\gamma^n\| \lesssim \|e^{n+1}\| + \|e^n\| \lesssim \tau^2$. $\qquad\square$

Lemma 3.3 *Given Theorem* 3.1, *the gradient of the error coming from the nonlinear term is bounded as* $\|\nabla e_\gamma^n\| \lesssim \|\nabla e^{n+1}\| + \|\nabla e^n\| + \tau^2$.

Proof The steps are much the same as in the previous lemma, with the exception that we need to use max-norm estimates for the gradient of the exact solution, which are however not available for the numerical approximation. We begin with a suitable error splitting of the form

$$
\begin{aligned}
\nabla e_\gamma^n ={} & \nabla[\gamma(\xi^n)u^{n+1/2} - \gamma(\xi_\tau^n)u_\tau^{n+1/2}] \\
={} & \nabla\gamma(\xi^n)u^{n+1/2} + \gamma(\xi^n)\nabla u^{n+1/2} - \nabla\gamma(\xi_\tau^n)u_\tau^{n+1/2} - \gamma(\xi_\tau^n)\nabla u_\tau^{n+1/2} \\
={} & u_\tau^{n+1/2}\nabla(\gamma(\xi^n) - \gamma(\xi_\tau^n)) + \nabla\gamma(\xi^n)(u^{n+1/2} - u_\tau^{n+1/2}) \\
& + (\gamma(\xi^n) - \gamma(\xi_\tau^n))\nabla u^{n+1/2} + \gamma(\xi_\tau^n)\nabla(u^{n+1/2} - u_\tau^{n+1/2}).
\end{aligned}
$$

Using the previous lemma and the max-norm bounds for the exact solution we may conclude,

$$
\begin{aligned}
\|\nabla e_\gamma^n\| \leq{} & \|u_\tau^{n+1/2}\|_{L^\infty}\|\nabla(\gamma(\xi^n) - \gamma(\xi^n))\| + \|\nabla\gamma(\xi^n)\|_{L^\infty}\|e^{n+1/2}\| \\
& + \|\nabla u^{n+1/2}\|_{L^\infty}\tau^2 + \|\gamma(\xi_\tau^n)\|_{L^\infty}\|\nabla e^{n+1/2}\|.
\end{aligned} \tag{3.12}
$$

What is left to bound is the term $\|\nabla(\gamma(\xi^n) - \gamma(\xi^n))\|$. We consider its dependence on $|u^n|^2, |u^{n+1}|^2, |u_\tau^n|^2$ and $|u_\tau^{n+1}|^2$ and find:

$$
\begin{aligned}
\nabla(\gamma(\xi^n) - \gamma(\xi_\tau^n)) ={} & \nabla(f(|u^n|^2, |u^{n+1}|^2) - f(|u_\tau^n|^2, |u_\tau^{n+1}|^2)) \\
={} & \partial_1 f^n \nabla|u^n|^2 + \partial_2 f^n \nabla|u^{n+1}|^2 - \partial_1 f_\tau^n \nabla|u_\tau^n|^2 - \partial_2 f_\tau^n\nabla|u_\tau^{n+1}|^2 \\
={} & (\partial_1 f^n - \partial_1 f_\tau^n)\nabla|u^n|^2 + \partial_1 f_\tau^n \nabla(|u^n|^2 - |u_\tau^n|^2) \\
& + (\partial_2 f^n - \partial_2 f_\tau^n)\nabla|u^{n+1}|^2 + \partial_2 f_\tau^n \nabla(|u^{n+1}|^2 - |u_\tau^{n+1}|^2),
\end{aligned} \tag{3.13}
$$

where $\partial_i f^n := \partial_i f(|u^n|^2, |u^{n+1}|^2)$ and $\partial_i f_\tau^n := f_i(|u_\tau^n|^2, |u_\tau^{n+1}|^2)$. Another application of the mean value theorem yields:

$$
\begin{aligned}
\partial_1(f^n - f_\tau^n) = \partial_{1,1} f(\theta^n, \theta^{n+1})\,(|u^n|^2 - |u_\tau^n|^2) + \partial_{1,2} f(\theta^n, \theta^{n+1})\,(|u^{n+1}|^2 - |u_\tau^{n+1}|^2), \\
\partial_2(f^n - f_\tau^n) = \partial_{2,1} f(\vartheta^n, \vartheta^{n+1})(|u^n|^2 - |u_\tau^n|^2) + \partial_{2,2} f(\vartheta^n, \vartheta^{n+1})(|u^{n+1}|^2 - |u_\tau^{n+1}|^2),
\end{aligned} \tag{3.14}
$$

for suitable mean value functions that can be bounded pointwise by the maximum of the exact solution and the semi-discrete CN-approximation. The following quick calculations show that the second partial derivatives of f are bounded by the second derivative of γ:

$$\partial_{1,1} f(a, b) = \frac{1}{(b-a)^2} 2(\gamma(c) - \gamma(a)) - \frac{\gamma'(a)}{b-a} = \frac{\gamma'(\theta^-) - \gamma'(a)}{b-a} + C_{\gamma''},$$

$$\partial_{1,2} f(a, b) = \frac{1}{(b-a)^2} (\gamma(b) - 2\gamma(c) + \gamma(a)) = \gamma''(c) + C_{\gamma''},$$

$$\partial_{2,2} f(a, b) = \frac{\gamma'(b)}{b-a} - \frac{1}{(b-a)^2} 2(\gamma(b) - \gamma(c)) = \frac{\gamma'(b) - \gamma'(\theta^+)}{b-a} + C_{\gamma''},$$

$$\tag{3.15}$$

where it was used that $c = (b + a)/2 + C_{\gamma''}(a - b)^2$. It thus becomes clear that $\|\partial_{1,1} f + \partial_{1,2} f + \partial_{2,2} f\|_{L^\infty} \leq C_{\gamma''}$. This gives us the following L^2-bound on (3.13)

$$\|\nabla(\gamma(\xi^n) - \gamma(\xi_\tau^n))\| \leq C_{\gamma''}(\|e^n\| + \|e^{n+1}\|) + C_{\gamma'}\|\nabla(|u^n|^2 - |u_\tau^n|^2 + |u^{n+1}|^2 - |u_\tau^{n+1}|^2)\|.$$

We conclude that

$$\|\nabla e_\gamma^n\| \lesssim \|\nabla(|u^n|^2 - |u_\tau^n|^2)\| + \|\nabla(|u^{n+1}|^2 - |u_\tau^{n+1}|^2)\| + \tau^2.$$

Here it is noted that $\nabla(|u^n|^2 - |u_\tau^n|^2)$ may be written as

$$\nabla(|u^n|^2 - |u_\tau^n|^2) = 2\text{Re}((u^n - u_\tau^n)\nabla\overline{u}^n + u_\tau^n\nabla(\overline{u}^n - \overline{u}_\tau^n)).$$

Hence, using the max-norm bounds for the gradient of the exact solution we have

$$||\nabla e_\gamma^n|| \lesssim ||\nabla e^{n+1}|| + ||\nabla e^n|| + \tau^2.$$

$$\square$$

With Lemmas 3.2 and 3.3 we now have the following bound on term I.

Lemma 3.4 *For term* I *which is given by* (3.8), *we have the estimate*

$$|I| \lesssim \|\nabla e^{n+1}\|^2 + \|\nabla e^n\|^2 + \|T^n\|^2 + \tau^4. \tag{3.16}$$

We can now proceed to bound term II. Here we explicate the Taylor term using (3.4) to see

$$\begin{aligned}
II &= -\text{Re}(\langle T^n, D_\tau e^n\rangle) \\
&\leq \underbrace{|\langle (D_\tau u^n - \partial_t u(t_{n+1/2})), D_\tau e^n\rangle|}_{\text{IIa}} + \underbrace{\text{Re}\langle -\Delta(u^{n+1/2} - u(t_{n+1/2})), D_\tau e^n\rangle}_{\text{IIb}} \\
&\quad + \underbrace{|\langle V(u^{n+1/2} - u(t_{n+1/2})), D_\tau e^n\rangle|}_{\text{IIc}}
\end{aligned}$$

$$+ \underbrace{\left| \left\langle \left(\frac{\Gamma(|u^{n+1}|^2) - \Gamma(|u^n|^2)}{|u^{n+1}|^2 - |u^n|^2} - \gamma(|u(t_{n+1/2})|^2), D_\tau e^n \right\rangle \right) \right|}_{\text{IId}}.$$

We start with estimating IIa, IIc and IId, which can be bounded in a similar way.

Step 1, bounding IIa:

By replacing $D_\tau e^n$ using (3.3) (i.e. time derivative is replaced by regularity in space) we have

$$|\langle D_\tau u^n - \partial_t u(t_{n+1/2}), D_\tau e^n \rangle|$$
$$\leq |\langle D_\tau u^n - \partial_t u(t_{n+1/2}), \Delta e^{n+1/2} \rangle| + |\langle D_\tau u^n - \partial_t u(t_{n+1/2}), V e^{n+1/2} \rangle|$$
$$+ |\langle D_\tau u^n - \partial_t u(t_{n+1/2}), e_\gamma^n \rangle| + |\langle D_\tau u^n - \partial_t u(t_{n+1/2}), T^n \rangle|$$
$$\overset{(3.6)}{\lesssim} \|\nabla e^{n+1}\|^2 + \|\nabla e^n\|^2 + \|D_\tau u^n - \partial_t u(t_{n+1/2})\|_{H^1}^2 + \|T^n\|^2 + \tau^4.$$

Here we see how the assumption that $\partial_{ttt} u \in L^2(0, T; H^1(\mathscr{D}))$ will enter, as it allows us to conclude that $\sum \|D_\tau u^n - \partial_t u(t_{n+1/2})\|_{H^1}^2 \lesssim \tau^3$.

Step 2, bounding IIc:

We use the same idea as for IIa to get

$$|\langle V(u^{n+1/2} - u(t_{n+1/2})), D_\tau e^n \rangle| \leq \|V\|_{L^\infty} |\langle u^{n+1/2} - u(t_{n+1/2}), D_\tau e^n \rangle|$$
$$\leq \|V\|_{L^\infty} \left(|\langle u^{n+1/2} - u(t_{n+1/2}), \Delta e^{n+1/2} \rangle| + \right.$$
$$+ |\langle u^{n+1/2} - u(t_{n+1/2}), V e^{n+1/2} \rangle| + |\langle u^{n+1/2} - u(t_{n+1/2}), e_\gamma^n \rangle|$$
$$\left. + |\langle u^{n+1/2} - u(t_{n+1/2}), T^n \rangle| \right)$$
$$\lesssim \|\nabla e^{n+1}\|^2 + \|\nabla e^n\|^2 + \|u^{n+1/2} - u(t_{n+1/2})\|_{H^1}^2 + \|T^n\|^2 + \tau^4.$$

Step 3, bounding IId:

We start from

$$\left| \left\langle \frac{\Gamma(|u^{n+1}|^2) - \Gamma(|u^n|^2)}{|u^{n+1}|^2 - |u^n|^2} - \gamma(|u(t_{n+1/2})|^2), D_\tau e^n \right\rangle \right|$$
$$= |\langle \gamma(\xi^n) - \gamma(|u(t_{n+1/2})|^2), D_\tau e^n \rangle|$$

replace $D_\tau e^n$ again using (3.3). Furthermore, in virtue of the assumptions it holds that $\|\nabla(\gamma(\xi^n) - \gamma(|u(t_{n+1/2})|^2))\| \lesssim \|u^{n+1/2} - u(t_{n+1/2})\|_{H^1}$. This is made explicit in the appendix. We thus obtain

$$|\langle \gamma(\xi^n) - \gamma(|u(t_{n+1/2})|^2), D_\tau e^n \rangle| = |\langle \gamma(\xi^n) - \gamma(|u(t_{n+1/2})|^2), -\Delta e^{n+1/2} + V e^{n+1/2} + e_\gamma^n + T^n \rangle|$$
$$\leq |\langle \nabla(\gamma(\xi^n) - \gamma(|u(t_{n+1/2})|^2)), \nabla e^{n+1/2} \rangle| + \|T^n\|^2 + \mathscr{O}(\tau^4)$$
$$\lesssim \|\nabla e^{n+1}\|^2 + \|\nabla e^n\|^2 + \|T^n\|^2 + \|u^{n+1/2} - u(t_{n+1/2})\|_{H^1}^2 + \tau^4.$$

Step 4, bounding IIb:

The previous technique does not work on this term since replacing the discrete time derivative with regularity in space would give rise to the term $\nabla \Delta (u^{n+1/2} - u(t_{n+1/2}))$, which we can not afford. Instead we use summation by parts in time to get the factor $D_\tau \Delta (u^{n+1/2} - u(t_{n+1/2}))$, which when integrated against $e^{n+1/2}$ can be handled. First we recall:

$$D_\tau[a^k b^k] = a^k D_\tau b^k + b^{k+1} D_\tau a^k \quad \Longleftrightarrow \quad -a^k D_\tau b^k = D_\tau[a^k]b^{k+1} - D_\tau[a^k b^k].$$

Using this on term IIb yields:

$$\begin{aligned}
\langle -\Delta(u^{n+1/2} - u), D_\tau e^n\rangle &= \langle \nabla(u^{n+1/2} - u), D_\tau \nabla e^n\rangle \\
&= D_\tau[\langle \nabla(u^{n+1/2} - u(t_{n+1/2})), \nabla e^n\rangle] - \langle D_\tau \nabla(u^{n+1/2} - u(t_{n+1/2})), \nabla e^{n+1}\rangle \\
&\le D_\tau[\langle \nabla(u^{n+1/2} - u(t_{n+1/2})), \nabla e^n\rangle] + |\langle D_\tau \nabla(u^{n+1/2} - u(t_{n+1/2})), \nabla e^{n+1}\rangle| \\
&\le D_\tau[\langle \nabla(u^{n+1/2} - u(t_{n+1/2})), \nabla e^n\rangle] + \|D_\tau(u^{n+1/2} - u(t_{n+1/2}))\|_{H^1}^2 + \|\nabla e^{n+1}\|^2.
\end{aligned}$$

Collecting the estimates we have the following estimate for term II.

Lemma 3.5 *For term* $\text{II} = -Re(\langle T^n, D_\tau e^n\rangle)$ *it holds the estimate*

$$\begin{aligned}
\text{II} \le{}& D_\tau[\langle \nabla(u^{n+1/2} - u(t_{n+1/2})), \nabla e^n\rangle] + C\big(\|\nabla e^{n+1}\|^2 + \|\nabla e^n\|^2 + \tau^4 + \|T^n\|^2 \\
&+ \|D_\tau u^n - \partial_t u(t_{n+1/2})\|_{H^1}^2 + \|u^{n+1/2} - u(t_{n+1/2})\|_{H^1}^2\big).
\end{aligned}$$

(3.17)

Here we note the importance of not estimating the absolute value of the first term since it is necessary to use the fact that n of these terms cancel when summed up, i.e. $\sum_k D_\tau a^k = \frac{1}{\tau}(a^{n+1} - a^0)$. We are now ready to finish the proof of the first main result.

Proof of Theorem 3.2 We pick off where we left (3.7) and find by using Lemmas 3.4 and 3.5:

$$\begin{aligned}
&\frac{\|\nabla e^{n+1}\|^2 - \|\nabla e^n\|^2}{2\tau} \\
&\le D_\tau[\langle \nabla(u^{n+1/2} - u(t_{n+1/2})), \nabla e^n\rangle] + C\big(\|\nabla e^{n+1}\|^2 + \|\nabla e^n\|^2 + \tau^4 \\
&\quad + \|D_\tau u^n - \partial_t u(t_{n+1/2})\|_{H^1}^2 + \|T^n\|^2 + \|u^{n+1/2} - u(t_{n+1/2})\|_{H^1}^2\big).
\end{aligned}$$

Summing this up and using $e^0 = 0$ gives

$$\begin{aligned}
\frac{\|\nabla e^{n+1}\|^2}{2\tau} \le{}& C\left(\sum_{k=0}^n \|\nabla e^k\|^2\right) + \frac{1}{\tau}\langle \nabla(u^{n+3/2} - u(t_{n+3/2})), \nabla e^{n+1}\rangle \\
&+ C\tau^3 + \sum_{k=0}^n \|T^k\|^2 + \|u^{k+1/2} - u(t_{k+1/2})\|_{H^1}^2 + \|D_\tau u^k \\
&- \partial_t u(t_{k+1/2})\|_{H^1}^2
\end{aligned}$$

and therefore, recalling (3.6),

$$\|\nabla e^{n+1}\|^2 \leq C \left(\sum_{k=0}^{n} \tau \|\nabla e^k\|^2 \right) + C\tau^4 + |\langle \nabla(u^{n+3/2} - u(t_{n+3/2})), \nabla e^{n+1}\rangle|.$$

Young's inequality with $\epsilon > 0$ is used on the last term:

$$|\langle \nabla(u^{n+3/2} - u(t_{n+3/2})), \nabla e^{n+1}\rangle| \leq C \left(\frac{\tau^4}{\epsilon} + \epsilon\|\nabla e^{n+1}\|^2 \right). \tag{3.18}$$

Which holds since,

$$\|\nabla(u^{n+3/2} - u(t_{n+3/2}))\| \lesssim \tau^2 \|\partial_{tt} u\|_{L^\infty(H^1)},$$

where we have $\|\partial_{tt} u\|_{L^\infty(H^1)} \lesssim \|\partial_{tt} u\|_{L^2(H^1)} + \|\partial_{ttt} u\|_{L^2(H^1)}$ by Sobolev embeddings. Finally we arrive at

$$\|\nabla e^{n+1}\|^2 \leq C \left(\sum_{k=0}^{n} \tau \|\nabla e^k\|^2 \right) + C\tau^4 + \frac{\tau^4}{\epsilon} + \epsilon\|\nabla e^{n+1}\|^2$$

and for e.g. $\epsilon = 1/2$ we can absorb $\epsilon\|\nabla e^{n+1}\|^2$ in the left hand side and conclude

$$\|\nabla e^{n+1}\|^2 \leq C \left(\tau^4 + \sum_{k=0}^{n} \tau \|\nabla e^k\|^2 \right).$$

Grönwall's inequality now yields:

$$\|\nabla e^{n+1}\| \lesssim \tau^2. \tag{3.19}$$

$\square$

4 Fully-discrete Crank-Nicolson scheme

We shall now consider the fully-discrete setting that is based on a finite element discretization in space. For that, we let $S_h \subset H_0^1(\mathscr{D})$ denote the space of P1 Lagrange finite elements on a quasi-uniform simplicial mesh on $\mathscr{D}$ with mesh size h. In this setting we have by standard finite element theory (cf. [39]) the following inequality for any $u \in H_0^1(\mathscr{D}) \cap H^2(\mathscr{D})$:

$$\|u - P_h u\| + h\|\nabla(u - P_h u)\| \lesssim h^2 \|u\|_{H^2}. \tag{4.1}$$

Here $P_h : H_0^1(\mathscr{D}) \to S_h$ denotes (for example) the Ritz-projection into the finite element space. For a given discrete initial value $u_h^0 \in S_h$ the CN-FEM approximation

$u^n_{\tau,h} \in S_h$ to u^n is given by the fully discrete equation

$$i\left\langle \frac{u^{n+1}_{\tau,h} - u^n_{\tau,h}}{\tau}, v \right\rangle = \left\langle \nabla u^{n+1/2}_{\tau,h}, \nabla v \right\rangle + \left\langle V u^{n+1/2}_{\tau,h}, v \right\rangle$$

$$+ \left\langle \frac{\Gamma(|u^{n+1}_{\tau,h}|^2) - \Gamma(|u^n_{\tau,h}|^2)}{|u^{n+1}_{\tau,h}|^2 - |u^n_{\tau,h}|^2} u^{n+1/2}_{\tau,h}, v \right\rangle \quad (4.2)$$

for all $v \in S_h$. The initial value is selected as $u^0_{\tau,h} = P_h u^0$. As in the semi-discrete case, the discretization conserves both mass and energy, i.e.

$$\int_{\mathscr{D}} |u^n_{\tau,h}|^2 \, dx = \int_{\mathscr{D}} |u^0_{\tau,h}|^2 \, dx \quad \text{and} \quad E[u^n_{\tau,h}] = E[u^0_{\tau,h}] \quad \text{for all } n \geq 0.$$

The scheme is well-posed and the corresponding approximations converge in the $L^\infty(L^2)$-norm with optimal order in space and time to the exact solution. A proof of this statement can be easily extracted from [35, Theorem 3.1 and Lemma 5.3]. In particular, we have the following result.

Theorem 4.1 *Under the general assumptions of this paper, there exists a solution $u^n_{\tau,h} \in H^1_0(\mathscr{D})$ to the fully discrete Crank-Nicolson scheme* (4.2) *such that the following a priori error estimates hold*

$$\sup_{0 \leq n \leq N} \|u^n_{\tau,h} - u^n_\tau\|_{L^2(\mathscr{D})} \lesssim h^2 \quad \text{and} \quad \sup_{0 \leq n \leq N} \|u^n_{\tau,h} - u^n\|_{L^2(\mathscr{D})} \lesssim h^2 + \tau^2.$$

With this we are ready to state our final theorem.

Theorem 4.2 (Optimal H^1-error estimates for the fully discrete method) *Let $u^n_\tau \in H^1_0(\mathscr{D})$ denote the fully-discrete Crank-Nicolson approximations from Theorem 4.1, then it holds*

$$\sup_{0 \leq n \leq N} \|u^n_{\tau,h} - u^n\|_{H^1(\mathscr{D})} \lesssim \tau^2 + h.$$

Proof First, we recall the inverse estimate on quasi-uniform meshes (cf. [39]), i.e. $\|\nabla v_h\| \leq Ch^{-1}\|v_h\|$ for all $v_h \in S_h$, which implies

$$\|\nabla(P_h(u^n_\tau) - u^n_{\tau,h})\| \leq Ch^{-1}\|P_h(u^n_\tau) - u^n_{\tau,h}\|. \quad (4.3)$$

With this, the H^1 convergence result (3.19) together with Theorem 4.1 suffice to show optimal H^1-convergence rates for the fully discrete method. This is made clear by the following splitting.

$$\|\nabla(u^n - u^n_{\tau,h})\| \leq \|\nabla(u^n - u^n_\tau)\| + \|\nabla(u^n_\tau - P_h(u^n_\tau))\| + \|\nabla(P_h(u^n_\tau) - u^n_{\tau,h})\|$$

$$\overset{(4.3)}{\leq} \|\nabla(u^n - u^n_\tau)\| + Ch + Ch.$$

Here we have made use of the inequality (4.1), the uniform H^2-regularity of u^n_τ, i.e. $\|u^n_\tau\|_{H^2} \le C(u)$ (cf. (3.2)) and the optimal L^2-estimates. In virtue of Theorem 3.2 we may thus conclude:

$$\|\nabla(u - u^n_{\tau,h})\| \le C(\tau^2 + h). \tag{4.4}$$

$\square$

Detailed numerical studies that confirm the optimal convergence rates stated in Theorem 4.1 and Theorem 4.2 are presented in [21,35].

5 Implementation and numerical examples

In this section we will discuss how the Crank-Nicolson FEM discretization can be efficiently implemented and practically used. Afterwards, we present two numerical experiments. The first one is to confirm the theoretically predicted convergence rates in Theorem 4.1 and the second experiment demonstrates that our approach is fully competitive in low regularity regimes, where we compare it with a time-splitting spectral method.

5.1 Efficient implementation

The Crank Nicolson method (4.2), albeit popular, suffers from the drawback that it requires solving a fully nonlinear system of equations in each time step. Furthermore, this system of equations is often solved through a Newton step, the implementation of which can become complicated and expensive for general nonlinearities. We present here a competitive fixed point solver which makes the method perform on par in terms of computational time with linearized time-stepping methods such as the RE-FEM proposed by C. Besse [23] which was found to be best performing in [21].

To detail the proposed fixed-point iteration, let $U^n \in \mathbb{R}^N$ denote the vector of nodal values that belongs to the function $u^n_{\tau,h} \in S_h$. Introducing the following matrix notation:

$$(M)_{ij} = \langle v_j, v_i \rangle \qquad A_{ij} = \langle \nabla v_j, \nabla v_i \rangle$$
$$(M_V)_{ij} = \langle V v_j, v_i \rangle \quad (M_\Gamma)_{ij}(U^{n+1}, U^n) = \langle \gamma(\xi^{n+1/2}_{\tau,h}) v_j, v_i \rangle$$

the equation (4.2) in matrix form becomes :

$$\mathrm{i}M \frac{U^{n+1} - U^n}{\tau} = \left(A + M_V + M_\Gamma(U^{n+1}, U^n)\right) \frac{U^{n+1} + U^n}{2}.$$

Let $L_1 = M + \mathrm{i}\tau/2(A + M_V)$ and $L_2 = M - \mathrm{i}\tau/2(A + M_V)$. Our fixed point iteration takes the form:

$$U^{n+1}_{i+1} = L_1^{-1} L_2 U^n + L_1^{-1} M_\Gamma(U^{n+1}_i, U^n)(U^{n+1}_i + U^n)/2. \tag{5.1}$$

Here we note that matrix L_1 does not change with time. Hence, the above iteration can be done efficiently by precomputing the LU-factorization of L_1. After it is precomputed, each time step only involves matrix-vector multiplications, but no longer the solving of a linear system of equations. In fact, the main cost in each time step account for the assembly of the updated mass matrix $M_\Gamma(U_i^{n+1}, U^n)$ with the densities from the previous time step and the previous iteration. Compared to this, all other costs are essentially negligible. We find that typically, but dependent on τ, 4–8 iterations are required to reach a tolerance of machine epsilon. To illustrate the efficiency we conclude with two numerical test problems.

5.2 Harmonic potential

First we consider a smooth potential to confirm the expected convergences rates for a type of nonlinearity that complements previous test cases [21,35]. Here we seek $u(x, t)$ with

$$\begin{cases} i\partial_t u = -\Delta u + Vu + \gamma(|u|^2)u & \text{in } \mathscr{D} \times (0, T], \\ u(\cdot, t) = 0 & \text{on } \partial\mathscr{D} \times (0, T], \\ u(\cdot, 0) = u_0 & \text{in } \mathscr{D}, \end{cases} \tag{5.2}$$

where we consider the saturated nonlinearity $\gamma(r) := r/(1+r)$ (cf. [8–10]) . Furthermore, $\mathscr{D} = [-5, 5]^2$ is the computational domain, the maximum time is selected as $T = 1$ and the trapping potential $V(x, y) = (\nu_x x)^2 + (\nu_y y)^2$. For the time-dependent problem we set the trapping frequencies to $\nu_x = 2$ and $\nu_y = 3$. The initial value u_0 is the unique positive ground state with $\int_\mathscr{D} |u_0|^2 = 1$ to the problem with $\nu_x = \nu_y = 1$, i.e. it solves the eigenvalue problem

$$\lambda_0 u_0 = -\Delta u_0 + Vu_0 + \gamma(|u_0|^2)u_0,$$

with ground state eigenvalue (chemical potential) $\lambda_0 > 0$. The H^1-errors are presented in Table 1. The $\mathscr{O}(h)$-convergence is best seen in column $\tau = 2^{-9}$, where initially the convergence is $\mathscr{O}(h^{1.5})$ but flattens out to $\mathscr{O}(h^{1.2})$ for the last data point. Since the reference solution is also computed with $h = 0.0125$, it is expected that this last order of convergence is an overestimate. Using the values in row $h = 0.0125$ we estimate the order of convergence with respect to τ to be 1.9, hence, confirming the theoretically predicted rates from Theorem 4.1.

5.3 Discontinuous potential

This example illustrates a moderate setting where, due to reduced regularity of the exact solution, finite element based methods are preferable over spectral methods. In the following we compare the Crank-Nicolson approach with a Strang splitting spectral method of order 2 (SP2) [40] which is known to show a very good performance in smooth settings.

Table 1 H^1-errors for harmonic potential test case (5.2)

	$\|\nabla(u_{\tau,h} - u_{\text{ref}})\|_{L^2}$						
	$\tau = 2^{-5}$	$\tau = 2^{-6}$	$\tau = 2^{-7}$	$\tau = 2^{-8}$	$\tau = 2^{-9}$	$\tau = 2^{-10}$	$\tau = 2^{-11}$
$h = 0.2$	0.829	0.400	0.485	0.526	0.538	0.542	0.543
$h = 0.1$	1.048	0.396	0.144	0.179	0.191	0.194	0.194
$h = 0.05$	1.109	0.498	0.129	0.054	0.062	0.066	0.066
$h = 0.025$	1.124	0.526	0.155	0.037	0.022	0.023	0.023
$h = 0.0125$	1.128	0.533	0.163	0.039	0.009	0.002	0.0005

The reference solution u_{ref} is calculated with $h = 0.0125$ and $\tau = 2^{-13}$. The energy is $E[u^n_{h,\tau}] = 3.86874$
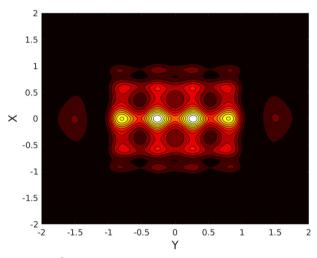

Fig. 1 Plot of the density ($|u|^2$) of the reference solution to (5.3) at $T = 1$

In this test problem we seek $u(x, t)$ with

$$\begin{cases} i\partial_t u = -\Delta u + Vu + \gamma(|u|^2)u & \text{in } \mathscr{D} \times (0, T], \\ u(\cdot, 0) = u_0 & \text{in } \mathscr{D}, \end{cases} \tag{5.3}$$

where we consider the saturated nonlinearity $\gamma(r) := 10r/(1 + r)$. For a fair comparison with the SP2, we consider our problem with periodic boundary conditions which are easier to handle by the spectral method. The generalization of the Crank-Nicolson method to periodic boundary conditions is straightforward. Furthermore, $\mathscr{D} = [-5, 5]^2$ is again the computational domain and the maximum time is selected as $T = 1$. The trapping potential $V(x, y) = (v_x x)^2 + (v_y y)^2 + 100(\mathbb{1}_{|x|\geq 1}(x) + \mathbb{1}_{|y|\geq 1}(y))$, with trapping frequencies $v_x = 1$ and $v_y = 3$, is discontinuous and causes a slight loss of regularity. We stress that this is a moderate test case, as illustrated in Fig. 1 most of the dynamics take place within the unit cube where the potential is smooth. The initial value u_0 is the unique positive ground state with $\int_{\mathscr{D}} |u_0|^2 = 1$ and

Table 2 Errors and computational times for the CN-FEM

	$\|\,|u_{\tau,h}|^2 - |u_{\text{ref}}|^2\,\|_{L^1}$	$\|\nabla(u_{h,\tau} - u_{\text{ref}})\|_{L^2}$	CPU [h]
CN-FEM, $h = 0.025$, $N_{DoF} = 400^2$			
$\tau = 2^{-8}$	0.40	4.91	0.4
$\tau = 2^{-9}$	0.18	3.20	0.8
$\tau = 2^{-10}$	0.12	1.43	1.4
CN-FEM, $h = 0.0125$, $N_{DoF} = 800^2$			
$\tau = 2^{-8}$	0.42	5.38	1.8
$\tau = 2^{-9}$	0.22	3.61	3.3
$\tau = 2^{-10}$	0.13	1.77	6.0
$\tau = 2^{-11}$	0.04	0.80	11.5

For a relative comparison of the errors we recall $\|\,|u_{\tau,h}|^2\,\|_{L^1} = 1$ and note $E[u_{\tau,h}] = 25.539397$

Table 3 Errors and computational times for the SP2

	$\|\,|u_{\tau,h}|^2 - |u_{\text{ref}}|^2\,\|_{L^1}$	$\|\nabla(u_{h,\tau} - u_{\text{ref}})\|_{L^2}$	CPU [h]
SP2, $N_{DoF} = 800^2$			
$\tau = 2^{-12}$	0.28	4.29	0.24
$\tau = 2^{-13}$	0.28	2.42	0.49
$\tau = 2^{-14}$	0.28	2.35	0.96
SP2, $N_{DoF} = 1600^2$			
$\tau = 2^{-12}$	0.12	2.62	1.15
$\tau = 2^{-13}$	0.12	1.28	2.26
$\tau = 2^{-14}$	0.12	1.18	5.2

For a relative comparison of the errors we recall $\|\,|u_{\tau,h}|^2\,\|_{L^1} = 1$ and note $E[u_{\tau,h}] = 25.539397$

$V_0(x, y) = x^2 + y^2$, i.e. it solves the eigenvalue problem

$$\lambda_0 u_0 = -\Delta u_0 + V_0 u_0 + \gamma(|u_0|^2)u_0.$$

The errors and the computational times of the CN-FEM are presented in Table 2 and the errors and computational times of the SP2 in Table 3. The reference solution, u_{ref}, is computed using the CN-FEM with $h = 0.0125$ and $\tau = 2^{-13}$. The implantation was done in `Julia`. It is important to keep in mind that the SP2 uses mostly inbuilt functions such as the fast Fourier transform from the `C` subroutine library (`FFTW`). These functions are heavily optimized and show an extremely good performance. In spite of this we see, comparing the errors of the CN-FEM for $h = 0.025$ and $\tau = 2^{-10}$ to those of the SP2 for $N_{DoF} = 3.2 \cdot 10^6$ and $\tau = 2^{-13}$, that they are on par with respect to CPU time relative to accuracy, with a slight computational advantage for the CN-FEM. This advantage becomes clearer, the larger the region of reduced regularity (e.g. in the context of optical lattices or disorder potentials). This justifies the usage

of CN-FEM in low regularity regimes. Furthermore, it is clearly seen that the space discretization dominates the error of the SP2, in order for the spectral method to catch up in terms of accuracy with the CN-FEM with $h = 0.0125$, an estimated 10 to 40 million degrees of freedom would be needed and thus the memory cost would become an issue.

A Explicit decomposition of the consistency error

Here we make the consistency error (3.4) and its estimate (3.6) explicit and highlight where the regularity assumptions come in. For this we use the following standard integral remainder of Taylor expansion (for Sobolev functions $v \in H^{n+1}(a, b)$):

$$v(b) - T_n^{a,b}(v) = \int_a^b \frac{(b-t)^n}{n!} v^{(n+1)}(t) dt,$$

where a is the point of expansion and $T_n^{a,b}(v)$ the Taylor polynomial of degree n of v evaluated in a and b. Recalling that $u^{n+1/2} = (u^{n+1} + u^n)/2$, the Taylor expansion implies in our case:

$$
\begin{aligned}
\|V(u^{n+1/2} - u(t_{n+1/2}))\|^2 &= \frac{1}{2} \| V \left(\int_{t_{n+1/2}}^{t_{n+1}} (t_{n+1} - s) \partial_{tt} u(s) ds \right. \\
&\qquad \left. - \int_{t_{n+1/2}}^{t_n} (t_n - s) \partial_{tt} u(s) ds \right) \|^2 \\
&\lesssim \left(\int_{t_n}^{t_{n+1}} (t_{n+1} - s) \|\partial_{tt} u(s)\| ds \right)^2 \\
&\lesssim \left(\tau^{3/2} \|\partial_{tt} u(s)\|_{L^2((t_n, t_{n+1}); L^2(\mathscr{D}))} \right)^2 \\
&\lesssim \tau^3 \int_{t_n}^{t_{n+1}} \|\partial_{tt} u\|^2 ds
\end{aligned}
$$

Thus

$$\sum_{k=1}^{n} \|u^{n+1/2} - u(t_{n+1/2})\|^2 \leq \tau^3 \int_0^T \|\partial_{tt} u(s)\|^2 ds \leq \tau^3 \|\partial_{tt} u\|^2_{L^2((0,T);L^2(\mathscr{D}))}.$$

Likewise we have

$$\|i(D_\tau u^n - \partial_t u(t_{n+1/2}))\|^2 = \frac{1}{\tau^2} \| \int_{t_{n+1/2}}^{t_{n+1}} \frac{(t_{n+1} - s)^2}{2} \partial_{ttt} u(s) ds$$
$$- \int_{t_{n+1/2}}^{t_n} \frac{(t_n - s)^2}{2} \partial_{ttt} u(s) ds \|^2$$
$$\lesssim \frac{1}{\tau^2} \tau^5 \int_{t_n}^{t_{n+1}} \|\partial_{ttt} u(s)\|^2 ds$$

and

$$\|\Delta(u^{n+1/2} - u(t_{n+1/2}))\|^2 = \| \int_{t_{n+1/2}}^{t_{n+1}} \frac{(t_{n+1} - s)^2}{2} \Delta \partial_{tt} u(s) ds$$
$$- \int_{t_{n+1/2}}^{t_n} \frac{(t_n - s)^2}{2} \Delta \partial_{tt} u(s) ds \|^2$$
$$\lesssim \tau^3 \int_{t_n}^{t_{n+1}} \|\Delta \partial_{tt} u(s)\|^2 ds.$$

For the estimate of the term that comes from the nonlinearity, we set for the sake of brevity $a := |u^n|^2$, $b := |u^{n+1}|^2$ and $c := |u(t_{n+1/2})|^2$ to obtain

$$\left\| \frac{1}{|u^{n+1}|^2 - |u^n|^2} \int_{|u^n|^2}^{|u^{n+1}|^2} \gamma(r) dr - \gamma |u(t_{n+1/2})|^2 \right\|^2 = \left\| \frac{1}{b-a} \int_a^b \gamma(r) dr - \gamma(c) \right\|^2$$
$$= \left\| \frac{1}{b-a} \int_a^b \int_c^r \gamma'(s) ds dr \right\|^2 = \left\| \gamma'(c) \left(\frac{b+a}{2} - c \right) \right\|$$
$$\lesssim \tau^3 \|\partial_{tt} |u|^2\|^2_{L^2((t_n, t_{n+1}); L^2)} \lesssim \tau^3 \|\partial_{tt} u\|^2_{L^2((t_n, t_{n+1}); L^2)}.$$

where c' lies between a and b. Thus

$$\sum_{k=0}^{n} \|T^k\|^2 \lesssim \tau^3 (\|\partial_{ttt} u\|^2_{L^2((0,T);L^2)} + \|\partial_{tt} u\|^2_{L^2((0,T);H^2)}$$
$$+ \|\partial_t u\|^2_{L^2((0,T);H^2)} + \|u\|^2_{L^2((0,T);H^2)}),$$

which proves (3.6).

References

1. Agrawal, G.P.: Nonlinear fiber optics. In: Christiansen, P.L., Sørensen, M.P., Scott, A.C. (eds.) Nonlinear Science at the Dawn of the 21st Century, pp. 195–211. Springer, Berlin (2000)
2. Hasimoto, H., Ono, H.: Nonlinear modulation of gravity waves. J. Phys. Soc. Jpn. **33**(3), 805–811 (1972)
3. Yuen, H.C., Lake, B.M.: Instabilities of waves on deep water. Ann. Rev. Fluid Mech. **12**(1), 303–334 (1980)
4. Zakharov, V.: Stability of periodic waves of finite amplitude on a surface of deep fluid. J. Appl. Mech. Tech. Phys. **9**(2), 190–194 (1968)
5. Gross, E.P.: Structure of a quantized vortex in boson systems. Nuovo Cimento **10**(20), 454–477 (1961)
6. Lieb, E.H., Seiringer, R., Yngvason, J.: A rigorous derivation of the Gross–Pitaevskii energy functional for a two-dimensional Bose gas. Comm. Math. Phys. **224**(1), 17–31 (2001). Dedicated to Joel L. Lebowitz
7. Pitaevskii, L.P.: *Vortex lines in an imperfect Bose gas*. Number 13. Soviet Physics JETP-USSR (1961)
8. Mihalache, V.F.D., Nazmitdinov, R.G.: Nonlinear optical waves in layered structures. Sov. J. Part. Nucl. **20**, 86 (1989)
9. Jordan, R., Turkington, B., Zirbel, C.L.: A mean-field statistical theory for the nonlinear Schrödinger equation. Phys. D **137**(3–4), 353–378 (2000)
10. Max, C.E.: Strong self-focusing due to ponderomotive forces in plasmas. Phys. Fluids **19**, 74 (1976)
11. Besse, C., Bidégaray, B., Descombes, S.: Order estimates in time of splitting methods for the nonlinear Schrödinger equation. SIAM J. Numer. Anal. **40**(1), 26–40 (2002)
12. Cano, B., González-Pachón, A.: Exponential time integration of solitary waves of cubic Schrödinger equation. Appl. Numer. Math. **91**, 26–45 (2015)
13. Cohen, D., Gauckler, L.: One-stage exponential integrators for nonlinear Schrödinger equations over long times. BIT **52**(4), 877–903 (2012)
14. Faou, E.: Geometric numerical integration and Schrödinger equations. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich, (2012)
15. Hairer, E., Lubich, C., Wanner, G.: Geometric Numerical Integration. Springer Series in Computational Mathematics, vol. 31, 2nd edn. Springer, BerlinBerlin (2006). (**Structure-preserving algorithms for ordinary differential equations**)
16. Lubich, C.: On splitting methods for Schrödinger–Poisson and cubic nonlinear Schrödinger equations. Math. Comput. **77**(264), 2141–2153 (2008)
17. Thalhammer, M., Abhau, J.: A numerical study of adaptive space and time discretisations for Gross–Pitaevskii equations. J. Comput. Phys. **231**(20), 6665–6681 (2012)
18. Thalhammer, M.: Convergence analysis of high-order time-splitting pseudospectral methods for nonlinear Schrödinger equations. SIAM J. Numer. Anal. **50**(6), 3231–3258 (2012)
19. Ostermann, A., Schratz, K.: Low regularity exponential-type integrators for semilinear Schrödinger equations. Found. Comput. Math. **18**(3), 731–755 (2018)
20. Knöller, M., Ostermann, A., Schratz, K.: A fourier integrator for the cubic nonlinear Schrödinger equation with rough initial data. SIAM J. Numer. Anal. **57**(4), 1967–1986 (2019)
21. Henning, P., Wärnegård, J.: Numerical comparison of mass-conservative schemes for the Gross–Pitaevskii equation. AIMS Kinet. Relat. Models **12**(6), 1247–1271 (2019)
22. Sanz-Serna, J.M., Verwer, J.G.: Conservative and nonconservative schemes for the solution of the nonlinear Schrödinger equation. IMA J. Numer. Anal. **6**(1), 25–42 (1986)
23. Besse, C.: A relaxation scheme for the nonlinear Schrödinger equation. SIAM J. Numer. Anal. **42**(3), 934–952 (2004)
24. Besse, C., Descombes, S., Dujardin, G., Lacroix-Violet, I.: Energy preserving methods for nonlinear Schrödinger equations. arXiv:1812.04890, (2018)
25. Akrivis, G.D., Dougalis, V.A., Karakashian, O.A.: On fully discrete Galerkin methods of second-order temporal accuracy for the nonlinear Schrödinger equation. Numer. Math. **59**(1), 31–53 (1991)
26. Antoine, X., Bao, W., Besse, C.: Computational methods for the dynamics of the nonlinear Schrödinger/Gross–Pitaevskii equations. Comput. Phys. Commun. **184**(12), 2621–2633 (2013)
27. Bao, W., Cai, Y.: Mathematical theory and numerical methods for Bose–Einstein condensation. Kinet. Relat. Models **6**(1), 1–135 (2013)
28. Henning, P., Målqvist, A.: The finite element method for the time-dependent Gross–Pitaevskii equation with angular momentum rotation. SIAM J. Numer. Anal. **55**(2), 923–952 (2017)

29. Karakashian, O., Makridakis, C.: A space-time finite element method for the nonlinear Schrödinger equation: the discontinuous Galerkin method. Math. Comput. **67**(222), 479–499 (1998)
30. Karakashian, O., Makridakis, C.: A space-time finite element method for the nonlinear Schrödinger equation: the continuous Galerkin method. SIAM J. Numer. Anal. **36**(6), 1779–1807 (1999)
31. Sanz-Serna, J.M.: Methods for the numerical solution of the nonlinear Schrödinger equation. Math. Comput. **43**(167), 21–27 (1984)
32. Tourigny, Y.: Optimal H^1 estimates for two time-discrete Galerkin approximations of a nonlinear Schrödinger equation. IMA J. Numer. Anal. **11**(4), 509–523 (1991)
33. Wang, J.: A new error analysis of Crank-Nicolson Galerkin FEMs for a generalized nonlinear Schrödinger equation. J. Sci. Comput. **60**(2), 390–407 (2014)
34. Zouraris, G.E.: On the convergence of a linear two-step finite element method for the nonlinear Schrödinger equation. M2AN Math. Model. Numer. Anal. **35**(3), 389–405 (2001)
35. Henning, P., Peterseim, D.: Crank-Nicolson Galerkin approximations to nonlinear Schrödinger equations with rough potentials. Math. Models Methods Appl. Sci. **27**(11), 2147–2184 (2017)
36. Bao, W., Cai, Y.: Uniform error estimates of finite difference methods for the nonlinear Schrödinger equation with wave operator. SIAM J. Numer. Anal. **50**(2), 492–521 (2012)
37. Bao, W., Cai, Y.: Optimal error estimates of finite difference methods for the Gross–Pitaevskii equation with angular momentum rotation. Math. Comput. **82**(281), 99–128 (2013)
38. Cazenave, T.: Semilinear Schrödinger equations, volume 10 of Courant Lecture Notes in Mathematics. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, (2003)
39. Brenner, S.C., Scott, L.R.: The Mathematical Theory of Finite Element Methods. Texts in Applied Mathematics, vol. 15, 3rd edn. Springer, New York (2008)
40. Bao, W., Jin, S., Markowich, P.A.: Numerical study of time-splitting spectral discretizations of nonlinear Schrödinger equations in the semiclassical regimes. SIAM J. Sci. Comput. **25**(1), 27–64 (2003)