Advances in Applied Mathematics and Mechanics
Adv. Appl. Math. Mech., Vol. **xx**, No. x, pp. 1-30

DOI: 10.4208/aamm.OA-2020-0270
xx 2021

Anderson Acceleration of Nonlinear Solvers for the Stationary Gross-Pitaevskii Equation

Dominique Forbes, Leo G. Rebholz[1,*] and Fei Xue

School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC 29634, USA

Received 26 August 2020; Accepted (in revised version) 2 February 2021

Abstract. We consider Anderson acceleration (AA) applied to two nonlinear solvers for the stationary Gross-Pitaevskii equation: a Picard type nonlinear iterative solver and a normalized gradient flow method. We formulate the solvers as fixed point problems and show that they both fit into the recently developed AA analysis framework. This allows us to prove that both methods' linear convergence rates are improved by a factor (less than one) from the gain of the AA optimization problem at each step. Numerical tests for finding ground state solutions in 1D and 2D show that AA significantly improves convergence behavior in both solvers, and additionally some comparisons between the solvers are drawn. A local convergence analysis for both methods are also provided.

AMS subject classifications: 65J15, 65B05, 65N30

Key words: Gross-Pitaevskii, Anderson acceleration, convergence analysis.

1 Introduction

We consider numerical solvers for the following nonlinear eigenvalue problem,

$$\mu\phi(x) = -\frac{1}{2}\Delta\phi(x) + V(x)\phi(x) + \beta|\phi(x)|^2\phi(x), \qquad x \in \Omega, \tag{1.1a}$$

$$\phi(x) = 0, \qquad x \in \partial\Omega, \tag{1.1b}$$

$$\int_\Omega |\phi(x)|^2 \, dx = 1, \tag{1.1c}$$

where V is a given trapping potential of the form

$$V(x) = \frac{1}{2}(\gamma_1^2 x_1^2 + \cdots + \gamma_d^2 x_d^2)$$

*Corresponding author.
Emails: dforbes@clemson.edu (D. Forbes), rebholz@clemson.edu (L. G. Rebholz), fxue@clemson.edu (F. Xue)

with $\gamma_i > 0$, $\forall i$, real parameter β, with ϕ being the unknown and the eigenvalue μ can be calculated as

$$\mu = \int_\Omega \left(\frac{1}{2}|\nabla\phi|^2 + V|\phi|^2 + \beta|\phi|^4 \right) dx. \tag{1.2}$$

This system describes stationary solutions of the nonlinear Schrödinger (NLS) equation, which is also known as the the non-rotational Gross-Pitaevskii equation (GPE) in the context of Bose-Einstein condensates (BEC) [7, 25, 27]; in the GPE setting ϕ represents the macroscopic wave function of the condensate. The parameter β being positive/negative represents defocusing/focusing in NLS, and attraction/repulsion of the condensate atoms in GPE. Following [7, 9], we will assume $\beta \geq 0$ for simplicity, although a more technical analysis would allow for our results to hold with β taking small negative values. Herein we assume a solution $\phi \in H_0^1(\Omega)$ exists to (1.1a)-(1.1c) that is locally unique; see [9] for more on this assumption and conditions for its validity.

Solutions of (1.1a)-(1.1c) represent local minima of the Gross-Pitaevskii energy

$$E_\beta(\phi) = \int_\Omega \left(\frac{1}{2}|\nabla\phi|^2 + V|\phi|^2 + \frac{\beta}{2}|\phi|^4 \right) dx,$$

and are used to numerically create initial conditions for real time dynamics of BEC, and to consider experimentally observed physical features through direct investigation [36]. Solutions to (1.1a)-(1.1c) that globally minimize energy are called ground state solutions, and it is stated in 2004 in [7] that "One of the fundamental problems in numerical simulation of BEC lies in computing the ground state solution". Existing methods include the normalized gradient flow (or gradient flow with discrete normalization) [3, 6, 7], which is most well-known, optimization methods such as steepest descent (without preconditioning) [36], preconditioned steepest descent (PSD, which is also effective in related problems [11, 14–16, 23]) and conjugate gradient (PCG) [4], and regularized Newton's method [33], as well as specialized non-optimization methods such as nonlinear inverse iterations [20], self-consistent field iteration [5], and implicit algorithms based on inexact Newton's method [21]. While normalized gradient flow methods have been widely used in the literature to find solutions of (1.1a)-(1.1c), improving the convergence and robustness of these methods remains an important problem and is the purpose of this paper.

We consider herein Anderson acceleration (AA) applied to two nonlinear solvers for (1.1a)-(1.1c). AA was originally developed in 1965 by D.G Anderson [2] as an extrapolation technique which forms the next iterate in a nonlinear fixed point iteration from an (in a sense) optimal linear combination of previous iterates. AA has increased in popularity since the work of [37], which showed how to implement it efficiently as a post-processor for a fixed point iteration and that it could be used successfully on a variety of application problems. Theoretical justification for how AA works was proved in [12], building on the theory from [13, 24, 34], and then sharpened and generalized to the case of noncontractive operators in [30]. AA has been used in a wide variety of applications including e.g., computing nearest correlation matrices in [19], geometry optimization [26], electronic

structure computations [13], radiation diffusion and nuclear physics [1, 35], molecular interaction [32], and fluid mechanics [28, 29]. This paper extends the AA methodology to improve solvers for the ground state solution of GPE.

The first solver we consider, which we call the Picard-projection iteration, takes the form Picard-projection

Step 1 Find $\hat{\phi}_{k+1}$ satisfying

$$\begin{aligned} &-\frac{1}{2}\Delta\hat{\phi}_{k+1}+V\hat{\phi}_{k+1}+\beta|\phi_k|^2\hat{\phi}_{k+1}=\phi_k,\\ &\hat{\phi}_{k+1}|_{\partial\Omega}=0. \end{aligned}$$

Step 2 Calculate ϕ_{k+1} by

$$\phi_{k+1}=\frac{\hat{\phi}_{k+1}}{\|\hat{\phi}_{k+1}\|_{L^2}}.$$

We note that μ is scaled out of the iteration, but can be recovered after convergence via (1.2). This iteration is considered a Picard-projection scheme since Step 1 consists of a linearization created by lagging part of the nonlinear term in the iteration, followed by Step 2 which projects the solution from Step 1 back to the unit sphere.

The second solver is a type of normalized gradient flow (NGF) studied in [7], where pseudo-timestepping is employed to find the steady solution of the time dependent system

$$\begin{aligned} &\phi_t=\frac{1}{2}\Delta\phi-V\phi-\beta|\phi|^2\phi,\\ &\phi_{\partial\Omega}=0,\\ &\int_\Omega|\phi|^2\,dx=1. \end{aligned}$$

We use the two-step backward Euler (BE) based method proposed in [7], which leads to the "time-stepping" iteration below.

BENGF Step 1 Find $\hat{\phi}_{k+1}$ satisfying

$$\frac{\hat{\phi}_{k+1}-\phi_k}{\Delta t}=\frac{1}{2}\Delta\hat{\phi}_{k+1}-V\hat{\phi}_{k+1}-\beta|\phi_k|^2\hat{\phi}_{k+1}, \tag{1.3a}$$

$$\hat{\phi}_{k+1}|_{\partial\Omega}=0. \tag{1.3b}$$

BENGF Step 2 Calculate ϕ_{k+1} by

$$\phi_{k+1}=\frac{\hat{\phi}_{k+1}}{\|\hat{\phi}_{k+1}\|_{L^2}}.$$

In Step 1, the time step size Δt is a user defined parameter, often taken large e.g., $\mathcal{O}(0.1)$ or larger [7]. Note that as $\Delta t \to \infty$, BENGF becomes formally equivalent to Picard-projection, making the methods strongly related. Note that BENGF Step 1 can be rearranged as

$$-\frac{1}{2}\Delta\left(\Delta t\hat{\phi}_{k+1}\right)+\left(V+\frac{1}{\Delta t}\right)\left(\Delta t\hat{\phi}_{k+1}\right)+\beta|\phi_k|^2\left(\Delta t\hat{\phi}_{k+1}\right)=\phi_k,$$

which reveals that as $\Delta t \to \infty$, $V+\frac{1}{\Delta t} \to V$, and therefore $\Delta t\hat{\phi}_{k+1}$ obtained from BENGF Step 1 would finally be identical in direction to $\hat{\phi}_{k+1}$ computed from Picard-projection Step 1.

We analyze and test AA applied to both the Picard-projection and BENGF solvers. For each, we formulate them as fixed point iterations together with a finite element spatial discretization. We then prove that both methods fit in the AA analysis framework recently developed in [30], and thus that their linear convergence rate will be improved at each iteration by a scaling factor (less than 1) associated with the gain of the AA optimization problem. Numerical experiments are then presented that show a significant improvement in the convergence of both methods, often cutting the number of iterations required for convergence by more than half. Comparisons between the methods are also given, and in fact we show that given a small $\beta>0$, for smaller time step sizes, the BENGF method tends to be slower than Picard-projection method but for larger time step sizes has similar convergence behavior. While we consider just these two particular solvers herein, based on our results we expect that AA will also provide improvement to other solvers for (1.1a)-(1.1c).

We remark that this PDE takes a form similar to the Allen-Cahn and Cahn-Hilliard equations, and energy stable numerical schemes have been developed for these systems, e.g., [10, 17, 18, 22, 31, 38, 39] and references therein. Extending those methods to work for this system and using AA, or applying AA to those systems are interesting ideas the authors will explore in future work.

This paper is arranged as follows. In Section 2, mathematical preliminaries are given as well as some notation and definitions. Section 3 analyzes AA applied to the two solvers, and Section 4 gives results of the associated numerical tests. Section 5 provides local convergence analysis of the methods, and finally conclusions are drawn in Section 6.

2 Background and preliminaries

We consider a domain $\Omega \subset \mathbb{R}^d$, $d=1,2,3$. The $L^2(\Omega)$ norm and inner product will be denoted by $\|\cdot\|$ and $(\cdot,\cdot)$ respectively. All other norms will be labeled with subscripts.

Define the space $X=H_0^1(\Omega)$, and recall the Poincare inequality, which guarantees that for all $\phi \in X$ there exists a constant C_P depending only on Ω such that

$$\|\phi\| \le C_P\|\nabla\phi\|.$$

Hence for functions in X, the H^1 and $\|\nabla\cdot\|$ norms are equivalent.

2.1 Discretization details

Let $\tau_h(\Omega)$ be a regular conforming mesh of Ω, with maximum element width h. Define $X_h = X \cap P_k(\tau_h)$, where $P_k(\tau_h)$ denotes degree k polynomials over each element of the mesh τ_h. We assume the mesh is sufficiently regular for the following inverse inequalities to hold [8]: there exists a constant C_I that is independent of h such that for every $v \in X_h$,

$$\|\nabla v\|_{L^p} \leq C_I h^{\frac{d}{2}\frac{2-p}{p}} \|\nabla v\|, \quad \|\nabla v\|_{L^p} \leq C_I h^{-1} \|v\|_{L^p}.$$

Lastly, we define $Y_h = \{v \in X_h,\ \|v\| = 1\}$ to be the subset of X_h on the unit sphere.

2.2 Anderson acceleration

Anderson acceleration is an extrapolation technique which is used to improve the convergence of fixed-point iterations. It may be stated as follows [34,37], for a given fixed-point operator $g : Y \to Y$, where Y is any normed vector space:

Algorithm 2.1. Anderson acceleration with depth $m \geq 0$ and damping factors $0 < \eta_k \leq 1$.

Step 0 Choose $x_0 \in Y$.

Step 1 Calculate $w_1 = g(x_0) - x_0$. Set $x_1 = x_0 + w_1$.

Step k For $k = 2,3,\cdots$. Set $m_k = \min\{k-1, m\}$.

a. Calculate $w_k = g(x_{k-1}) - x_{k-1}$.

b. Solve the minimization problem for parameters $\{\alpha_j^k\}_{k-m_k}^{k-1}$

$$\min \left\| \left(1 - \sum_{j=k-m_k}^{k-1} \alpha_j^k\right) w_k + \sum_{j=k-m_k}^{k-1} w_j \right\|_Y . \tag{2.1}$$

c. For damping factor $0 < \eta_k \leq 1$, set

$$\begin{aligned} x_k = &\Big(1 - \sum_{j=k-m_k}^{k-1} \alpha_j^k\Big) x_{k-1} + \sum_{j=k-m_k}^{k-1} \alpha_j^k x_{j-1} \\ &+ \eta_k \left(\Big(1 - \sum_{j=k-m_k}^{k-1} \alpha_j^k\Big) w_k + \sum_{j=k-m_k}^{k-1} \alpha_j^k w_j \right), \end{aligned} \tag{2.2}$$

In the AA algorithm, $w_j = g(x_{j-1}) - x_{j-1}$ is referred to as the nonlinear residual. The goal of AA is to accelerate convergence compared to the original fixed point iteration $x_{k+1} = g(x_k)$, and we note that $m = 0$ returns exactly this original fixed point iteration.

For implementation with depth $m>0$, it is helpful to write the algorithm in terms of an unconstrained optimization problem [13,30,37]. Define the matrices E_k and F_k, whose columns are the consecutive differences between iterates and residuals, respectively:

$$E_{k-1} := \begin{pmatrix} e_{k-1} & e_{k-2} & \cdots & e_{k-m_k} \end{pmatrix}, \quad e_j = x_j - x_{j-1}, \tag{2.3a}$$

$$F_k := \begin{pmatrix} (w_k - w_{k-1}) & (w_{k-1} - w_{k-2}) & \cdots & (w_{k-m_k+1} - w_{k-m_k}) \end{pmatrix}. \tag{2.3b}$$

Then defining

$$\gamma^k = \text{argmin}_{\gamma \in \mathbb{R}^m} \|w_k - F_k \gamma\|_Y,$$

we can write the update step (2.2) as

$$x_k = x_{k-1} + \eta_k w_k - (E_{k-1} + \eta_k F_k)\gamma^k = x_{k-1}^\alpha + \eta_k w_k^\alpha, \tag{2.4}$$

where $w_k^\alpha = w_k - F_k\gamma^k$ and $x_{k-1}^\alpha = x_{k-1} - E_{k-1}\gamma^k$. The optimization gain factor ξ_k is defined by

$$\|w_k^\alpha\|_Y = \xi_k \|w_k\|_Y, \tag{2.5}$$

and this gain factor ξ_k plays a critical role in the convergence theory to follow. Specifically, AA reduces the contribution from the first-order residual term by a factor of ξ_k, but introduces higher-order terms into the residual expansion.

The next two assumptions are sufficient conditions on the fixed point operator g for the AA convergence results developed in [12,30] to hold.

Assumption 2.1. Assume $g \in C^1(Y)$ has a fixed point x^* in Y, and there are positive constants c_0 and c_1 with

1. $\|g'(x)\| \le c_0$ for all $x \in Y$, and
2. $\|g'(x) - g'(y)\| \le c_1 \|x - y\|$ for all $x, y \in Y$.

Note that for the AA theory below, it is not required that $c_0 < 1$, or even that $\|g'(x^*)\| < 1$.

Assumption 2.2. Assume there is a constant $\sigma > 0$ for which the differences between consecutive residuals and iterates satisfy

$$\|w_{k+1} - w_k\|_Y \ge \sigma \|x_k - x_{k-1}\|_Y, \quad k \ge 1.$$

For contractive g, such a σ can be automatically found [30]; if g is non-contractive, existence of such σ's is related to the Jacobian of the function $f(x) = x - g(x)$ being non-singular [30].

Under Assumptions 2.1 and 2.2, the following result summarized from [30] produces a one-step bound on the residual $\|w_{k+1}\|$ in terms of the previous residual $\|w_k\|$.

Theorem 2.1 (Pollock, Rebholz, 2020)**.** *Consider Algorithm* 2.1 *with depth* m_k*. Suppose Assumptions* 2.1 *and* 2.2 *hold, and the direction sines between columns of* F_j *defined by* (2.3b) *are bounded below by a constant* $c_s>0$*, for* $j=k-m_k,\cdots,k-1$*. Then the residual* $w_{k+1}=g(x_k)-x_k$ *satisfies the following bound:*

$$\|w_{k+1}\| \le \|w_k\| \left(\xi_k((1-\eta_k)+c_0\eta_k) + \frac{Cc_1\sqrt{1-\theta_k^2}}{2} \left(\|w_k\| h(\xi_k) + 2\sum_{n=k-m_k+1}^{k-1} (k-n)\|w_n\| h(\xi_n) + m_k \left\|w_{k-m_k}\right\| h(\xi_{k-m_k}) \right) \right), \tag{2.6}$$

where $h(\xi_j)\le C\sqrt{1-\xi_j^2}+\eta_j\xi_j$*, and* C *depends on* c_s *and the implied upper bound on the direction cosines.*

Remark 2.1. As discussed in [30], the assumption that $c_s>0$ enforces that the columns of F_j are linearly independent. This is necessary for solving the minimization problem, and moreover the constant C in (2.6) blows up as $c_s\to 0$. This criteria can be easily enforced in practice and on the fly, by calculating c_s and reducing m_k accordingly so that c_s is always bounded above a certain tolerance (in tests in [30], a tolerance of 0.1 worked well).

The one-step estimate (2.6) shows how the gain factor ξ_k from the optimization problem scales the linear convergence rate, and also the summation shows how increasing the depth m_k adds higher order terms to the residual expansion. Hence the improvement in the linear convergence rate comes at a cost if recent residuals are not small. If there is no gain from the optimization problem, then since the higher order terms are scaled by $\sqrt{1-\xi_k^2}$ the usual fixed point residual estimate is recovered. The theorem suggests that greater depths in the early iterations may slow or prevent convergence in many cases, while alternating between small m_k for early iterations and increasing it in later iterations when the residual is small may be more effective; indeed, in [29] this is shown to be an effective strategy for choosing the depth.

3 Anderson accelerated iterations for solving GPE

In this section we prove that the Anderson accelerated Picard-projection and BENGF iterations, once discretized with the finite element method (FEM), fit the AA analysis framework developed in [30]. This allows us to prove that AA improves the linear convergence rate for both iterations.

3.1 Anderson acceleration Picard-projection iteration

We consider here AA applied to a FEM discretization of the Picard-projection iteration, which we call PP_h. It is defined as follows:

PP$_h$ Step 1 Given $\phi_k \in Y_h$, find $\hat{\phi}_{k+1} \in X_h$ satisfying

$$\frac{1}{2}(\nabla\hat{\phi}_{k+1},\nabla\chi)+(V\hat{\phi}_{k+1},\chi)+\beta(|\phi_k|^2\hat{\phi}_{k+1},\chi)=(\phi_k,\chi).$$

PP$_h$ Step 2 Calculate

$$\phi_{k+1}=\frac{\hat{\phi}_{k+1}}{\|\hat{\phi}_{k+1}\|}.$$

We establish next the well-posedness of PP$_h$, which allows us to define the solution operator $g: Y_h \to Y_h$ for each PP$_h$ iteration so that it can be written as

$$\phi_{k+1}=g(\phi_k)=(p\circ f)(\phi_k),$$

where f represents the Step 1 solution operator, and p represents the Step 2 projection.

Lemma 3.1. *For any given mesh sufficiently regular for the inverse inequality to hold, the PP$_h$ iteration is well-posed.*

Proof. Given $u \in V_h$, define the bilinear form $a_u(\cdot,\cdot): X_h \times X_h \to \mathbb{R}$ by

$$a_u(v,\chi)=\frac{1}{2}(\nabla v,\nabla\chi)+(Vv,\chi)+\beta(|u|^2v,\chi).$$

Since $\beta \geq 0$ and $V(x) \geq 0$ for all x, a_u is observed to be continuous and coercive: choosing $\chi = v$ gives

$$a_u(v,v)=\frac{1}{2}\|\nabla v\|^2+\|V^{1/2}v\|^2+\beta\int_\Omega |u|^2|v|^2\,dx \geq \frac{1}{2}\|\nabla v\|^2,$$

and Cauchy-Schwarz, Hölder, and Sobolev inequalities provide

$$\begin{aligned}
|a_u(v,\chi)| &\leq \frac{1}{2}\|\nabla v\|\|\nabla\chi\|+\|V\|_{L^\infty}\|v\|\|\chi\|+\beta\||u|^2\|_{L^{3/2}}\|v\|_{L^6}\|\chi\|_{L^6}\\
&\leq \frac{1}{2}\|\nabla v\|\|\nabla\chi\|+\|V\|_{L^\infty}C_P^2\|\nabla v\|\|\nabla\chi\|+C\beta\|u\|_{L^3}^2\|\nabla v\|\|\nabla\chi\|\\
&\leq \left(\frac{1}{2}+\|V\|_{L^\infty}C_P^2+C\beta\|u\|_{L^3}^2\right)\|\nabla v\|\|\nabla\chi\|\\
&\leq \left(\frac{1}{2}+\|V\|_{L^\infty}C_P^2+C\beta C_I h^{-d/3}\right)\|\nabla v\|\|\nabla\chi\|,
\end{aligned}$$

with the last step thanks to the inverse inequality and $\|u\|=1$. We note that in 1D, an improved Agmon inequality will allow for the dependence on h to be removed using

$$\beta(|u|^2v,\chi)\leq\beta\|u\|_{L^2}^2\|v\|_{L^\infty}\|\chi\|_{L^\infty}\leq C\beta\|\nabla v\|\|\nabla\chi\|.$$

Define $F_u : X_h \to \mathbb{R}$ by

$$F_u(\chi) = (u,\chi),$$

and F is easily seen to be continuous using Cauchy-Schwarz and Poincare via

$$|F_u(\chi)| \le \|u\|\|\chi\| \le C_P\|\nabla\chi\|.$$

Now by Lax-Milgram, the problem:

$$\text{Given } u \in V_h, \text{ find } v \in X_h \text{ satisfying } a_u(v,\chi) = F_u(\chi) \text{ for all } \chi \in X_h$$

is well-posed. Since $\phi_k \in V_h \subset X_h$, this well-posedness implies that PP_h Step 1 is well posed, and that the continuity and coercivity constants are independent of k.

Next, since $\phi_k \in Y_h$ in PP_h Step 1, we have that $\phi_k \neq 0$ and thus from the coercivity of a_{ϕ_k}, we obtain $\hat{\phi}_{k+1} \neq 0$. This makes the Step 2 projection well-defined, and thus we have established that the PP_h iteration is well-posed. □

The well-posedness result above (in particular the use of the Lax-Milgram theorem) is sufficient to also establish Lipschitz continuity of g: for all $u,v \in Y_h$, there exists $C_g < \infty$ satisfying

$$\|\nabla(g(u) - g(v))\| \le C_g\|\nabla(u-v)\|.$$

Despite the potential inverse dependence of C_g on the mesh width in the above theory, we argue that near the solution C_g may not, at least for a small $\beta > 0$. Notice that each iteration of PP_h can be considered as one step of an inverse power method using the finite element coefficient matrix A_k arising from the PP_h Step 1. This is an important point when discussing convergence of PP_h and comparing it to BENGF below. Denoting this matrix by

$$A_k = \left(-\frac{1}{2}\Delta + V + \beta|\phi_k|^2\right)_h,$$

and $A = (-\frac{1}{2}\Delta + V + \beta|\phi|^2)_h$ (where ϕ solves (1.1a)-(1.1c) as the unique minimum eigenvalue) then for a good enough initial guess it is reasonable to expect that PP_h defines a contractive iteration that will converge to the minimum eigenvalue of A and its associated eigenvector, for a sufficiently small $\beta > 0$. A more detailed discussion of the local convergence is given in section 5, but for now this suggests that C_g should be at most $O(1)$ if near a root with a small β, since in any vector norm the associated Lipschitz constant would be bounded above by 1.

While the PP_h scheme is simple, converges locally with a small β (see Section 5), and has no parameters to choose, it can suffer from slow convergence if the smallest and second smallest eigenvalues of A are close. We consider now PP_h enhanced with AA, with

depth m_k and relaxation parameters η_k: this is defined simply by the AA algorithm of Section 2 with the fixed point function g above. We will refer to it as Anderson accelerated PP$_h$ (AAPP$_h$). As discussed in Section 2, a theoretical framework was recently developed in [12, 30] that proves an increase in the linear convergence rate of a fixed point iteration by AA in the vicinity of a root, provided several assumptions on the underlying fixed point function. This subsection is dedicated to showing that the PP$_h$ fixed point operator g satisfies the AA theory assumptions from [30], which yields the following result.

Theorem 3.1. *For AAPP$_h$ under the assumption that the direction sines between columns of F_j defined by* (2.3b) *are bounded below by a constant $c_s > 0$, for $j = k-m_k, \cdots, k-1$, and Assumption* 2.3 *holds for g and the previous m_k iterates, then $w_{k+1} = g(\phi_k) - \phi_k$ satisfies*

$$\|w_{k+1}\| \leq \|w_k\| \left(\xi_k((1-\eta_k) + C_g\eta_k) + C\sqrt{1-\theta_k^2} \left(\sum_{n=k-m_k}^{k-1} \|w_n\| h(\xi_n) \right) \right). \tag{3.1}$$

Remark 3.1. The theorem reveals an improvement in the linear convergence rate from the use of AA since the gain factor $\xi_k \in [0,1]$ of the AA optimization problem is reduced as m_k is increased. However, the theorem also shows that as m_k is increased, more higher order terms are introduced into the upper bound. The additional higher order terms that are added can be enough to prevent convergence especially early in an iteration when the residual is large. This phenomena of m_k being too large is illustrated for AAPP$_h$ in the first numerical test in Section 4.1. As discussed in [29, 30], a good strategy for choosing m_k is to choose it small when the residual is large, and choose it large when the residual is small.

Remark 3.2. Regarding the assumptions of the theorem: Remark 2.1 above discusses how $c_s > tol > 0$ can be satisfied in practice by reducing m_k if necessary. Assumption 2.3 holds for g and the previous m_k iterates, for example, when g is contractive in a neighborhood of the root (which can be checked by running PP$_h$ and monitoring convergence, given an initial guess in the neighborhood) and the previous m_k iterates all lie in the contractive region. It is discussed in [30] how Assumption 2.3 can be satisfied if g is not contractive at the root. While the analysis below guarantees an upper bound on C_g, it is not sufficiently refined to produce usable criteria for where g is contractive near a root.

This theorem will be proved in the following series of lemmas, which show that g satisfies Assumptions 2.1-2.2. Once this is established, we are able to invoke Theorem 2.1, which provides the above result for 1 iteration of AAPP$_h$.

To begin our analysis, consider the FEM problem representing

Step 1 Given $u \in X_h$ and $\beta > 0$, find $v \in X_h$ satisfying

$$\frac{1}{2}(\nabla v, \nabla \chi) + (Vv, \chi) + \beta(|u|^2 v, \chi) = (u, \chi), \quad \forall \chi \in X_h. \tag{3.2}$$

It is shown above that the solution operator $f : X_h \to X_h$ given by $v = f(u)$ associated with (3.2) is well-defined.

Next, define the set $S=\{u\in X_h,\ \|u\|\leq K_0\}$, where K_0 can be any constant larger than 1, and note that $Y_h\subset S$. We now prove an a priori solution bound for $v=f(u)$ with $u\in S$.

Lemma 3.2. *For any $\beta>0$ and $u\in S$, the solution $v=f(u)\in X_h$ satisfies*

$$\|\nabla v\|\leq C\min\{2C_PK_0,\beta^{-1/2}(vol(\Omega))^{1/2}\}=:C_1,$$

where C is a constant dependent only on Ω.

Proof. Taking $\chi=v$ in (3.2) yields

$$\frac{1}{2}\|\nabla v\|^2+\|V^{1/2}v\|^2+\beta\||u|v\|^2=(u,v).$$

Thanks to Cauchy-Schwarz and Poincare inequalities and $u\in S$, we have that the right hand side term is bounded by

$$(u,v)\leq\|u\|\|v\|\leq C_P\|u\|\|\nabla v\|\leq C_PK_0\|\nabla v\|,$$

and so

$$\frac{1}{2}\|\nabla v\|^2\leq C_PK_0\|\nabla v\|\Longrightarrow\|\nabla v\|\leq 2C_PK_0.$$

One could alternatively majorize the right hand side using Cauchy-Schwarz and Young's inequalities via

$$(u,v)=\int_\Omega(uv)\cdot 1\,dx\leq\||u|v\|\|1\|\leq\frac{\beta}{2}\||u|v\|^2+\frac{1}{2\beta}\|1\|^2,$$

which gives

$$\|\nabla v\|^2+\beta\||u|v\|^2\leq\beta^{-1}\mathrm{vol}(\Omega),$$

and finishes the proof. □

Lemma 3.3. *Suppose $u_1,u_2\in S$. Then there exists a constant $C_f=\left(8(C_P^4+8C\beta^2K_0^2C_1^2)\right)^{1/2}$ such that*

$$\|\nabla(f(u_1)-f(u_2))\|=\|\nabla(v_1-v_2)\|\leq C_f\|\nabla(u_1-u_2)\|.$$

Proof. Let $v_1=f(u_1)$ and $v_2=f(u_2)$. Then from the definition of f, we have that

$$\frac{1}{2}(\nabla v_1,\nabla\chi)+(Vv_1,\chi)+\beta(|u_1|^2v_1,\chi)=(u_1,\chi),\qquad\forall\chi\in X_h,$$
$$\frac{1}{2}(\nabla v_2,\nabla\chi)+(Vv_2,\chi)+\beta(|u_2|^2v_2,\chi)=(u_2,\chi),\qquad\forall\chi\in X_h.$$

Subtracting then yields

$$\frac{1}{2}(\nabla(v_1-v_2),\nabla\chi)+(V(v_1-v_2),\chi)+\beta(|u_1|^2v_1-|u_2|^2u_2,\chi)=(u_1-u_2,\chi),\quad \forall\chi\in X_h,$$

and choosing $\chi=v_1-v_2$ implies that

$$\begin{aligned}&\frac{1}{2}\|\nabla(v_1-v_2)\|^2+\|V^{1/2}(v_1-v_2)\|^2\\&\quad+\beta\int_\Omega(|u_1|^2v_1-|u_2|^2v_2)(v_1-v_2)dx\\=&(u_1-u_2,v_1-v_2).\end{aligned}\tag{3.3}$$

The right hand side term can be majorized with Cauchy-Schwarz and Young inequalities as

$$(u_1-u_2,v_1-v_2)\le\|u_1-u_2\|_{H^{-1}}\|\nabla(v_1-v_2)\|\le\|u_1-u_2\|_{H^{-1}}^2+\frac{1}{4}\|\nabla(v_1-v_2)\|^2,$$

and using the expansion

$$\begin{aligned}|u_1|^2v_1-|u_2|^2v_2=&(|u_1|^2-|u_2|^2)v_2+|u_1|^2(v_1-v_2)\\=&(|u_1|+|u_2|)(|u_1|-|u_2|)v_2+|u_1|^2(v_1-v_2),\end{aligned}$$

we have from (3.3) that

$$\begin{aligned}&\frac{1}{4}\|\nabla(v_1-v_2)\|^2+\|V^{1/2}(v_1-v_2)\|^2+\beta\int_\Omega|u_1|^2|v_1-v_2|^2\,dx\\\le&\|u_1-u_2\|_{H^{-1}}^2+\beta\Big|\int_\Omega(|u_1|+|u_2|)(|u_1|-|u_2|)v_2(v_1-v_2)\,dx\Big|.\end{aligned}\tag{3.4}$$

Using Hölder's inequality ($L^2-L^6-L^6-L^6$) on the right hand side, the embedding of L^6 into H^1, the triangle inequality, and Lemma 3.2, we obtain the bound

$$\begin{aligned}&\Big|\int_\Omega(|u_1|+|u_2|)(|u_1|-|u_2|)v_2(v_1-v_2)\,dx\Big|\\\le&\|u_1+u_2\|\|\nabla(u_1-u_2)\|\|\nabla v_2\|\|\nabla(v_1-v_2)\|\\\le&2CK_0C_1\|\nabla(u_1-u_2)\|\|\nabla(v_1-v_2)\|.\end{aligned}$$

Combining with the bound (3.4), we now have

$$\frac{1}{4}\|\nabla(v_1-v_2)\|^2\le\|u_1-u_2\|_{H^{-1}}^2+2C\beta K_0C_1\|\nabla(u_1-u_2)\|\|\nabla(v_1-v_2)\|.\tag{3.5}$$

Young's inequality and Poincare now provide

$$\|\nabla(v_1-v_2)\|^2\le8(C_P^4+8C\beta^2K_0^2C_1^2)\|\nabla(u_1-u_2)\|^2,\tag{3.6}$$

which finishes the proof. □

Lemma 3.4. *The solution operator f is Frechet differentiable on S, and its Frechet derivative is Lipschitz continuous.*

Proof. From the definition of f, we have for $u\in S$, $\delta\in S$, and $u+\delta\in S$,

$$\frac{1}{2}(\nabla(f(u+\delta)),\nabla\chi)+(Vf(u+\delta),\chi)+\beta(|u+\delta|^2f(u+\delta),\chi)=(u+\delta,\chi),$$
$$\frac{1}{2}(\nabla(f(u)),\nabla\chi)+(Vf(u),\chi)+\beta(|u|^2f(u),\chi)=(u,\chi).$$

Subtracting gives

$$\begin{aligned}&\frac{1}{2}(\nabla(f(u+\delta)-f(u)),\nabla\chi)+(V(f(u+\delta)-f(u)),\chi)\\&\quad+\beta((|u+\delta|^2-|u|^2)f(u)),\chi)+\beta(|u+\delta|^2(f(u+\delta)-f(u)),\chi)\\&=(\delta,\chi).\end{aligned}\tag{3.7}$$

Define the operator $A_u=A_u(\delta)$ by: find $A_u\in X_h$ satisfying

$$\frac{1}{2}(\nabla A_u,\nabla\chi)+(VA_u,\chi)+\beta(|u|^2A_u,\chi)+2\beta(u\delta f(u),\chi)=(\delta,\chi),\quad\forall\chi\in X_h.\tag{3.8}$$

Since $u,\delta\in S$, and with the a priori bounds established in Lemma 3.2, A_u is well-defined thanks to the Lax-Milgram theorem. Further, choosing $\chi=A_u$, dropping positive left hand side terms, and using Sobolev inequalities yields the bound

$$\begin{aligned}\frac{1}{2}\|\nabla A_u\|^2&\le 2\beta\|f(u)\|_{L^6}\|A_u\|_{L^6}\|u\|_{L^2}\|\delta\|_{L^6}+\|\delta\|\|A_u\|\\&\le C\beta C_1K_0\|\nabla A_u\|\|\nabla\delta\|+C_P^2\|\nabla\delta\|\|\nabla A_u\|,\end{aligned}$$

and thus

$$\|\nabla A_u\|\le(C\beta C_1K_0+C_P^2)\|\nabla\delta\|.\tag{3.9}$$

Writing $e=f(u+\delta)-f(u)-A_u$ after subtracting (3.8) from (3.7) provides

$$\begin{aligned}&\frac{1}{2}(\nabla e,\nabla\chi)+(Ve,\chi)+\beta(|u|^2e,\chi)\\&=-\beta(|\delta|^2f(u),\chi)-\beta((2u\delta+\delta^2)(f(u+\delta)-f(u)),\chi).\end{aligned}\tag{3.10}$$

Choosing $\chi=e$, we obtain using Hölder ($L^2-L^6-L^6-L^6$), Sobolev and Poincare inequalities along with Lemma 3.2 that

$$\begin{aligned}&\frac{1}{2}\|\nabla e\|^2+\|V^{1/2}e\|^2+\beta\||u|e\|^2\\&=-\beta(|\delta|^2f(u),e)-\beta((2u\delta+\delta^2)(f(u+\delta)-f(u)),e)\\&\le CC_P\beta\|\nabla\delta\|^2\|\nabla f(u)\|\|\nabla e\|+C\beta\|\nabla e\|\|\nabla(f(u+\delta)-f(u))\|\|\nabla\delta\|(\|\delta\|+\|u\|)\\&\le CC_PC_1\beta\|\nabla\delta\|^2\|\|\nabla e\|+CC_PK_0\beta\|\nabla e\|\|\nabla(f(u+\delta)-f(u))\|\|\nabla\delta\|.\end{aligned}\tag{3.11}$$

Using Lemma 3.3 and reducing gives

$$\|\nabla e\| \leq CC_P\beta\left(C_1+K_0C_f\right)\|\nabla\delta\|^2. \tag{3.12}$$

This proves that A_u is the Frechet derivative of f at u.

The bound (3.9) with the assumption that $u\in S$ implies continuity of A_u, and Lipschitz continuity of the Frechet derivative of f on S since the bound depends only on global constants. □

We have now established that f is Lipshitz continuously Frechet differentiable on S, and that

$$\|\nabla(f(u)-f(v))\| \leq C_f\|\nabla(u-v)\| \quad \text{for} \quad u,v\in S.$$

Next, we show that the function $p(\psi)=\psi/\|\psi\|$ is Lipschitz continuously Frechet differentiable on bounded sets that are bounded away from 0.

Define the space $Y_h=\{v\in X_h,\ \|v\|=1\}$, and consider now the function $p: f(X_h\backslash\{0\})\to Y_h$ defined by

$$p(\psi)=\frac{\psi}{\|\psi\|}.$$

In fact, for any given $\psi\in X_h\backslash\{0\}$ and a $\delta\psi\in X_h\backslash\{0\}$ sufficiently small in norm,

$$\begin{aligned}
&p(\psi+\delta\psi)-p(\psi)=\frac{\psi+\delta\psi}{\|\psi+\delta\psi\|}-\frac{\psi}{\|\psi\|}\\
=&\frac{\psi+\delta\psi}{\left((\psi,\psi)+(\psi,\delta\psi)+(\delta\psi,\psi)+(\delta\psi,\delta\psi)\right)^{\frac{1}{2}}}-\frac{\psi}{\|\psi\|}\\
=&\frac{\psi+\delta\psi}{\|\psi\|\left(1+(\psi,\psi)^{-1}((\psi,\delta\psi)+(\delta\psi,\psi)+(\delta\psi,\delta\psi))\right)^{\frac{1}{2}}}-\frac{\psi}{\|\psi\|}\\
=&\frac{\psi+\delta\psi}{\|\psi\|}\left(1-\frac{1}{2}(\psi,\psi)^{-1}((\psi,\delta\psi)+(\delta\psi,\psi)+(\delta\psi,\delta\psi))\right)-\frac{\psi}{\|\psi\|}+\mathcal{O}(\delta\psi^2)\\
=&\frac{1}{\|\psi\|}\left(\delta\psi-\frac{1}{2}(\psi,\psi)^{-1}((\psi,\delta\psi)+(\delta\psi,\psi))\psi\right)+\mathcal{O}(\delta\psi^2)\\
=&\frac{1}{\|\psi\|}\left(\delta\psi-\frac{(p(\psi),\delta\psi)+(\delta\psi,p(\psi))}{2}p(\psi)\right)+\mathcal{O}(\delta\psi^2).
\end{aligned} \tag{3.13}$$

In other words, the Frechet derivative of $p:\psi\to\frac{\psi}{\|\psi\|}$ is given by the linear mapping

$$Dp(\psi):\delta\psi\to\frac{1}{\|\psi\|}\left(\delta\psi-\frac{(p(\psi),\delta\psi)+(\delta\psi,p(\psi))}{2}p(\psi)\right), \tag{3.14}$$

which is obviously continuous with respect to ψ bounded away from 0. Also, it follows that

$$\begin{aligned}\|Dp(\psi)\| &= \sup_{\|\delta\psi\|\neq 0} \frac{1}{\|\psi\|} \left\| \delta\psi - \frac{(p(\psi),\delta\psi)+(\delta\psi,p(\psi))}{2} p(\psi) \right\| / \|\delta\psi\| \\ &\leq \frac{1}{\|\psi\|} \sup_{\delta\psi\neq 0} \left(1+\|p(\psi)\|^2\right) \|\delta\psi\| / \|\delta\psi\| = \frac{2}{\|\psi\|}. \end{aligned} \tag{3.15}$$

The analysis above can be summarized into the following conclusion:

Lemma 3.5. *The function $g: Y_h \to Y_h$ defined by $g = p \circ f$ is Lipschitz continuously Frechet differentiable, which completes the proof of Theorem* 3.1.

Proof. Taking $K_0 = 2$, we have that $Y_h \subset S$. This is sufficient for Lemmas 3.2-3.4 to hold, making f Lipschitz continuously Frechet differentiable on Y_h. The function $f: u \to v$ maps $u \in Y_h$ to $v \in X_h \setminus \{0\}$ bounded away from 0. The coercivity of the symmetric bilinear form representing the left side of (3.2), which is established in Lemma 3.1, together with the domain of f being Y_h (all elements lie on unit sphere in L^2 norm) is sufficient to guarantee a positive lower bound on $\|\nabla f(u)\|$. Hence p operates only on $f(Y_h)$, which is bounded away from 0, and so p is Lipschitz continuously Frechet differentiable on $f(Y_h)$ and maps into Y_h. Since $g = p \circ f$, g must then also be Lipschitz continuously Frechet differentiable on Y_h. □

3.2 Anderson accelerated BENGF iteration

We consider now AA applied to a FEM discretization of BENGF iteration (1.3a)-(1.3b). While [7] considers BENGF as a pseudo time-stepping method looking for a steady state, we consider it as a fixed point iteration to which we can apply AA.

We now define BENGF$_h$, which is the BENGF iteration equipped with a finite element discretization.

BENGF$_h$ Step 1 Given $\phi_k \in Y_h$, find $\hat{\phi}_{k+1} \in X_h$ satisfying

$$\frac{\Delta t}{2}(\nabla\hat{\phi}_{k+1},\nabla\chi)+((1+V\Delta t)\hat{\phi}_{k+1},\chi)+\beta\Delta t(|\phi_k|^2\hat{\phi}_{k+1},\chi)=(\phi_k,\chi).$$

BENGF$_h$ Step 2 Calculate

$$\phi_{k+1} = \frac{\hat{\phi}_{k+1}}{\|\hat{\phi}_{k+1}\|}.$$

Just as for PP$_h$, we will consider improvements offered by AA to BENGF$_h$.

Define the solution operator for the BENGF$_h$ iteration to be $\tilde{g}: Y_h \to Y_h$, so that BENGF$_h$ can be written as: $\phi_{k+1} = \tilde{g}(\phi_k)$. Since BENGF$_h$ takes the exact same form as PP$_h$, except has different left hand side coefficients in Step 1 that are still the same sign as in PP$_h$, the theory that proves PP$_h$ is well-posed and that its solution operator g is well defined also applies here, and thus BENGF$_h$ is well posed and $\tilde{g}$ is well defined. The Anderson accelerated BENGF$_h$ is then defined by using the AA algorithm with the fixed point function $\tilde{g}$, and we call the resulting method AABENGF$_h$. By the same reasoning as above for BENGF$_h$ well-posedness, we can also invoke the PP$_h$ theory to prove that $\tilde{g}$ satisfies Assumptions 2.1-2.2. Thus, calling $C_{\tilde{g}}$ the BENGF$_h$ Lipschitz constant, repeating the PP$_h$ theory establishes the following theorem for AABENGF$_h$. Remark 3.2 applies in the same way to Theorem 3.2 as it does to Theorem 3.1 with regard to the theorem assumptions. Also similar to Theorem 3.1, the theorem below reveals an improvement in the linear convergence rate from AA through reducing ξ_k by increasing m_k, but with the tradeoff of more higher order terms being introduced into the upper bound.

Theorem 3.2. *For AABENGF$_h$ under the assumption that the direction sines between columns of F_j defined by* (2.3b) *are bounded below by a constant $c_s > 0$, for $j = k-m_k, \cdots, k-1$, and Assumption* 2.3 *holds for g and the previous m_k iterates, the residual $w_{k+1} = \tilde{g}(\phi_k) - \phi_k$ satisfies the following bound.*

$$\|w_{k+1}\| \leq \|w_k\| \left(\xi_k((1-\eta_k) + C_{\tilde{g}}\eta_k) + C\sqrt{1-\theta_k^2} \left(\sum_{n=k-m_k}^{k-1} \|w_n\| h(\xi_n) \right) \right). \tag{3.16}$$

4 Numerical tests

We now test the impact of AA on the PP$_h$ and BENGF$_h$ iterations for computing the ground state solution of several GPE problems in 1D and 2D. Overall we find for both iterations that AA significantly reduces the number of iterations needed for convergence. No relaxation is used, since the methods appeared to be contractive (or at least not growing) in all tests.

4.1 1D tests

We first compare PP$_h$, BENGF$_h$, and their AA accelerated variants on an example problem from [7] in 1D with a harmonic oscillator potential. Specifically, this problem has

$$V(x) = \frac{x^2}{2}, \quad \phi_0(x) = \frac{1}{\pi^{1/4}} e^{-\frac{x^2}{2}}, \quad \Omega = [-20, 20].$$

We discretize using $N = 2^{14}$ equispaced elements (subintervals), and piecewise linear elements. Mass lumping is used, and so this FEM discretization is equivalent to a second order finite difference discretization in space. The stopping criterion of each method is that $\|\phi_{k+1} - \phi_k\| < 10^{-8}$.

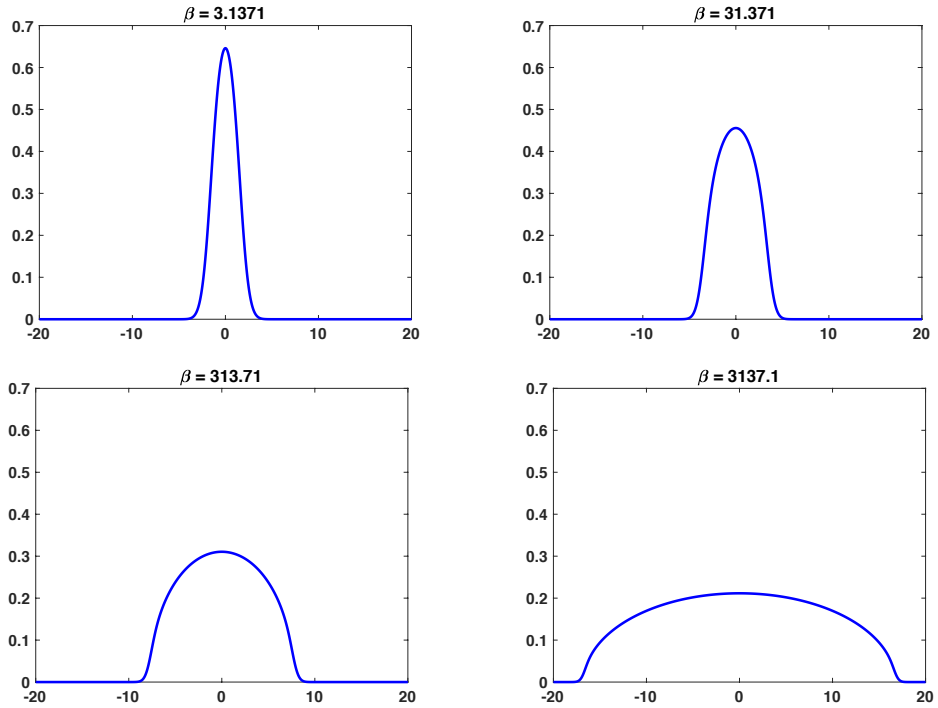


Figure 1: Solution to the 1D GPE with different values of β.

Solutions are computed by PP_h, $BENGF_h$, with and without AA, and with varying β. Table 1 shows the energy and chemical potential of the solutions corresponding to different values of β, and Fig. 1 plots the ground state solutions ϕ.

The performance of PP_h and $AAPP_h$ is summarized and illustrated in Fig. 2 and Table 2. For example, Table 2 shows that PP_h ($m=0$) takes 16 iterations to converge for the GPE with $\beta=3.1371$, and $AAPP_h$ with depth $m=1$ needs 37 iterations to converge for $\beta=31.371$. The best performance (lowest iteration counts) is shown in bold numbers in the table. Note that $m=3$ is the best depth of AA for the two smaller values of β, but for the two larger β, $m=2$ is the best whereas $m=3$ leads to failure of convergence. The fail-

Table 1: The energy and chemical potential (lowest eigenvalue) found for the 1D GPE test problem. The same values were found with PP_h and $BENGF_h$, with and without AA. These values obtained with $N=2^{13}$ and $N=2^{14}$ subintervals agreed to at least 5 digits.

β	$E_\beta(\phi)$	$\mu_g=\mu_\beta(\phi)$
3.1371	1.0441	1.5266
31.371	3.9810	6.5527
313.71	18.171	30.259
3137.1	84.249	140.41

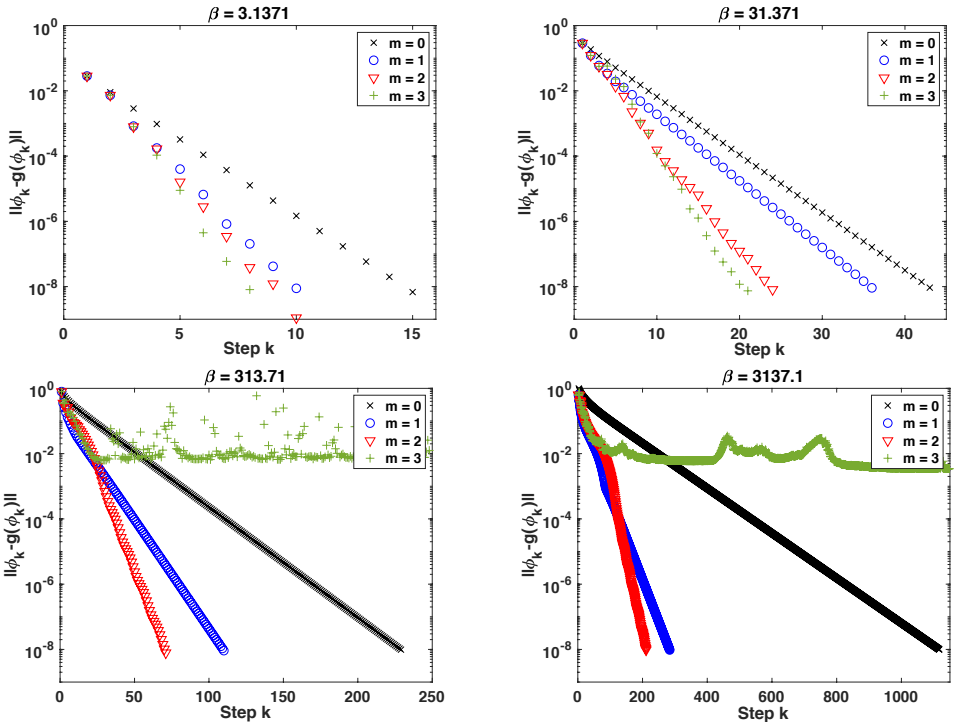


Figure 2: Convergence of Picard-projection and Anderson accelerated Picard-projection for the 1D GPE with different values of β.

ure for larger m is consistent with the convergence analysis, as increasing m introduces extra higher order terms into the right hand side of the residual bound; see Theorem 3.1 and Remark 3.1. We shall see a similar pattern for the 2D test later.

The performance of BENGF$_h$ with different time step size Δt is shown in Table 3. For instance, BENGF$_h$ with Δt=0.01 takes 588 iterations to converge for β=3.1371. We note in particular that for β=3.1371, the performance of BENGF$_h$ improves monotonically with increasing Δt, and PP$_h$ outperforms BENGF$_h$ with any $\Delta t \leq 10$. This is consistent with the local convergence analysis of the methods from section 5 on the potential advantage of Picard-projection for small β. For larger β, it is interesting to note that the (near-)optimal

Table 2: Iteration counts of Picard-projection (m=0) and AA accelerated Picard-projection with different depth $m \geq 1$ for the GPE 1D test problem.

β	m=0	m=1	m=2	m=3
3.1371	16	11	11	**9**
31.371	44	37	25	**22**
313.71	230	111	**72**	∞
3137.1	1118	285	**212**	∞

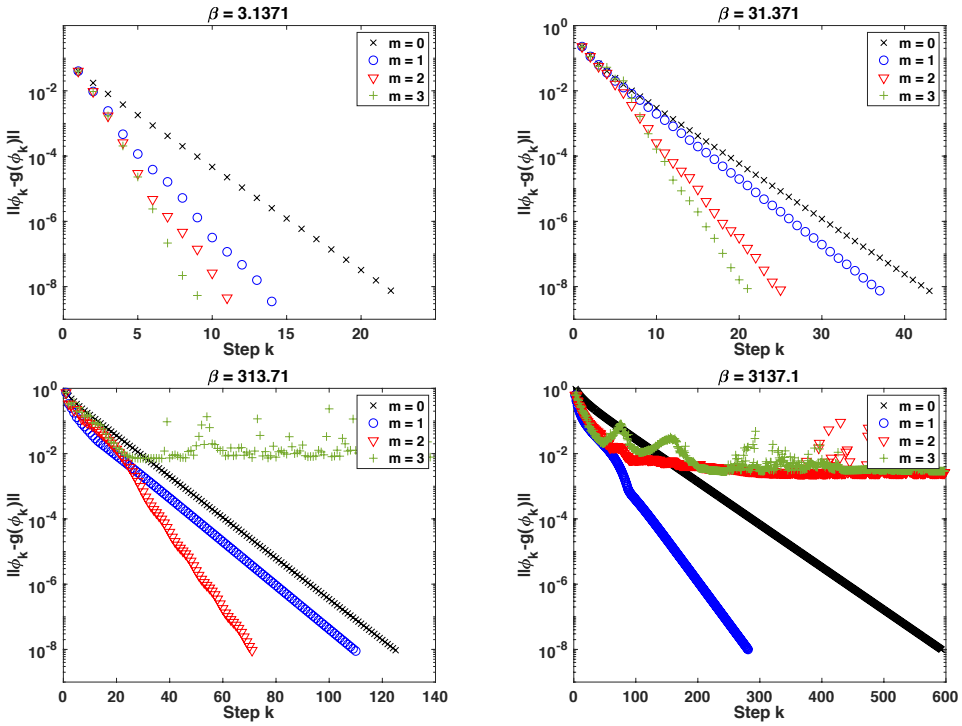


Figure 3: Convergence of BENGF$_h$ and AABENGF$_h$ with time step $\Delta t = 1$ for the 1D GPE with different values of β.

Δt (with which fewest iterations are needed) seems to be inversely proportional to β; with the optimal Δt, BENGF$_h$ converges more rapidly than PP$_h$. However, it is not clear how to find the optimal Δt without repeated experiments.

Finally, the performance of BENGF$_h$ and AABENGF$_h$ is shown in Fig. 3 and Table 4. For example, Table 4 shows that BENGF$_h$ ($m = 0$) and AABENGF$_h$ with depth $m = 1$ need 16 and 11 iterations, respectively, to find the solution for $\beta = 3.1371$ and $\Delta t = 100$. Again, the lowest iteration counts are shown in bold numbers. As in the AAPP$_h$ test above, for larger β the larger choices of m do not allow convergence, which is consistent with the convergence analysis; see Theorem 3.2 and Remark 3.1.

Table 3: Iteration counts of BENGF$_h$ with different time step size Δt for the GPE 1D test problem. The (near-)optimal time step Δt for BENGF$_h$ seems to be inversely proportional to β.

β	$\Delta t = 0.01$	$\Delta t = 0.1$	$\Delta t = 1$	$\Delta t = 10$	$\Delta t = 100$	$\Delta t = \infty$ (PP$_h$)
3.1371	588	82	23	17	**16**	16
31.371	492	88	44	**42**	44	44
313.71	434	142	**126**	213	228	230
3137.1	550	**338**	594	1027	1109	1118

Table 4: Iteration counts of BENGF_h ($m=0$) and AA accelerated BENGF_h with depth $m\geq 1$ for the GPE 1D test problem. The (near-)optimal time step Δt is used with each value of β.

$(\beta,\Delta t)$	$m=0$	$m=1$	$m=2$	$m=3$
(3.1371,100)	16	11	11	**9**
(31.371,10)	42	38	25	**22**
(313.71,1)	126	111	**72**	–
(3137.1,0.1)	338	**263**	–	–

In comparison with Table 2, it is worth noting that the best performance of AAPP_h and AABENGF_h are exactly the same for $\beta=3.1371,31.371$ and 313.71, whereas the former needs fewer iterations for $\beta=3137.1$. Since the AAPP_h does not need Δt, it is a robust algorithm that depends less on the choice of parameters and reliably outperforms PP_h and BENGF_h (with optimal Δt) for this test problem.

4.2 2D tests

We consider now a test from [7] in 2D for a harmonic oscillator potential together with a potential from a stirrer corresponding to a far-blue detuned Gaussian laser beam, i.e.,

$$V(x,y)=\frac{1}{2}\left(x^2+y^2\right)+4e^{-\left((x-1)^2+y^2\right)}.$$

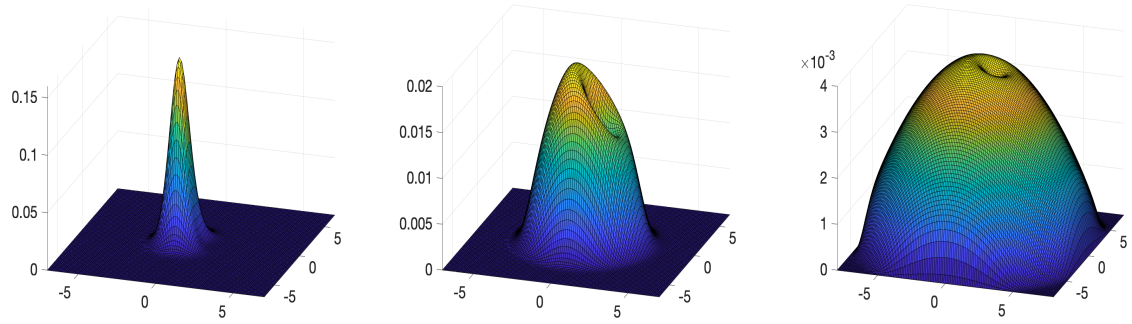
We compute solutions with the Picard-projection iteration and BENGF_h, with and without AA, and for several choices of β. For the initial guess for PP_h and initial condition for BENGF_h, we use

$$\phi_0(x,y)=\frac{1}{\pi^{1/2}}e^{-(x^2+y^2)/2}.$$

For spatial discretizations of the domain $\Omega=[-8,8]^2$, we use uniform meshes (with varying h) and continuous piecewise quadratic elements. Three different meshes were used, $h=\frac{1}{32}$, $\frac{1}{128}$, $\frac{1}{256}$, and all convergence behavior was found to be mesh independent (up to ± 1 iteration for each instance). Plots of the solution to the steady GPE system are shown in Fig. 4 as surface plots, and Table 5 shows the energy and eigenvalues associated with the converged solution to the discrete system.

Table 5: Shown above are the energy and eigenvalues found for the varying β in the 2D GPE test problem. The values for $h=\frac{1}{128}$ and $h=\frac{1}{256}$ agreed to at least 5 digits for each value in the table.

β	$E_\beta(\phi)$	$\mu_g=\mu_\beta(\phi)$
1	1.684	1.7245
400	5.851	8.3150
10000	26.84	40.587



$\beta=1$ $\beta=400$ $\beta=10{,}000$

Figure 4: Shown above is are surface plots of the converged ground state solution on the finest mesh, for varying β.

For AAPP$_h$, we computed solutions with constant $m=0,1,2,3$ and also an adaptive depth chosen so that

$$m_k=\begin{cases} 1, & \|\phi_{k-1}-\phi_{k-2}\|\geq 10^{-2}, \\ 2, & 10^{-3}\leq\|\phi_{k-1}-\phi_{k-2}\|<10^{-2}, \\ 10, & \|\phi_{k-1}-\phi_{k-2}\|<10^{-3}. \end{cases}$$

This adaptive depth is made based on the convergence results in Theorems 3.1 and 3.2, which show that smaller m early in the iteration will decrease the linear convergence rate and introduce less error from higher order terms, whereas once the residual is smaller the higher order terms will be negligible and it is advantageous for the depth to be taken larger.

Convergence results for AAPP$_h$ and PP$_h$ ($m=0$) are displayed in Fig. 5, and we observe that AAPP$_h$ offers considerable improvement over PP$_h$ provided m is not taken too large, reducing the number of iterations substantially, especially for larger β. The best results for AAPP$_h$ came from using the adaptive choice of m, but using constant m still offered a significant improvement over PP$_h$ ($m=0$). Using AAPP$_h$ with constant $m>3$ did not offer any improvement over the $m=3$ results for this problem.

For the AABENGF$_h$ method, we used the same parameters as above for AAPP$_h$. Convergence behavior is shown in Fig. 6, and we observe that AABENGF$_h$ makes a big improvement in BENGF$_h$, and is best when the adaptive m is chosen. Results for AABENGF$_h$ with adaptive m are very similar to those above for AAPP$_h$ with adaptive m. As with the 1D tests, we observed sensitivity in (AA)BENGF$_h$ to the time step size Δt. Shown in Table 6 are iteration counts for BENGF$_h$ (i.e., $m=0$) and AABENGF$_h$ with adaptive m, for varying Δt, and we observe somewhat less sensitivity in choosing the time step size when AA with adaptive m is used.

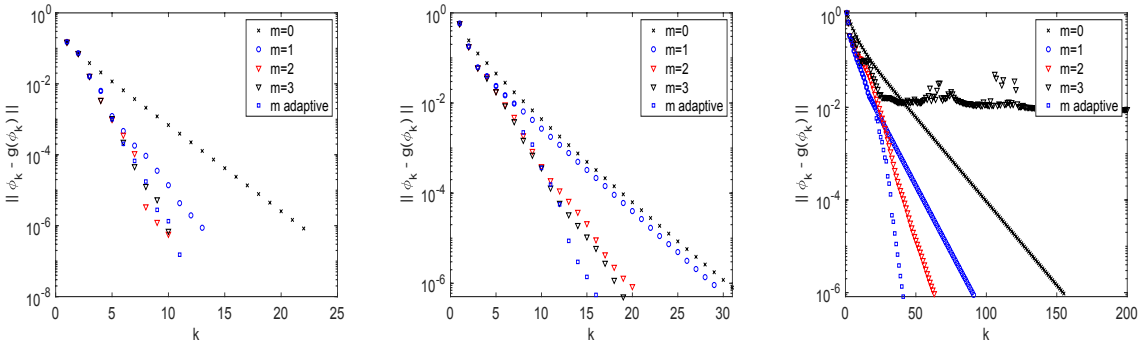


Figure 5: Shown above is convergence behavior for AAPP$_h$ with varying m.

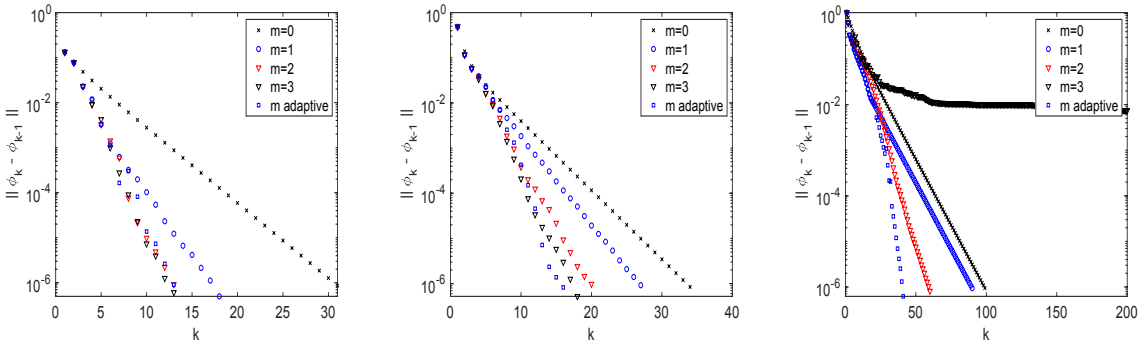


Figure 6: Shown above is convergence behavior for BENGF$_h$ with $\Delta t=1$ and Anderson acceleration with varying m.

Table 6: Shown above are the number of iterations to reach convergence for BENGF$_h$ using $\beta=400$, $m=0$ and adaptive m, for varying Δt.

Δt	AABENGF$_h$ $w/$ adaptive m	BENGF$_h$ $(m=0)$
0.01	45	314
0.05	23	92
0.1	18	62
1	16	34
10	16	31
100	16	31

5 Local convergence of Picard-projection and BENGF

Lastly, we provide a local convergence analysis of the Picard-projection and BENGF iterations, which appears to be absent in the literature. In the previous sections, the Picard-

projection and BENGF iterations are studied and tested under a finite element discretization (and then called PP_h and BENGF_h, respectively) since this readily allowed for analysis. In this section, for simplicity, we will assume a finite difference discretization. Modifications can be made to the proof to make it work also for FEM, but require additional notation and technical details. While we do not consider AA in this section, the results here are relevant to the theory above since they give bounds on the Lipschitz constants C_g and $C_{\tilde{g}}$ (at least locally) which appear in the convergence theorems. We will slightly abuse notation in this section, with vectors ϕ_{k+1} and $\hat{\phi}_{k+1}$ representing finite difference solutions at iteration $k+1$. We will use $\|\cdot\|_{l^p}$ to denote l^p vector norms in this section.

For a given mesh/discretization, in matrix form the scheme reads

$$\text{Step 1:} \quad \text{Solve } A_k\hat{\phi}_{k+1}=\phi_k \text{ for } \hat{\phi}_{k+1}, \text{ i.e., } \hat{\phi}_{k+1}=A_k^{-1}\phi_k, \tag{5.1a}$$

$$\text{Step 2:} \quad \phi_{k+1}=\hat{\phi}_{k+1}/\|\hat{\phi}_{k+1}\|_{l^2}, \tag{5.1b}$$

with the change of norm from L^2 to l^2 expected to be negligible on a sufficiently fine mesh. Here,

$$A_k=-\frac{1}{2}L+\text{diag}(V)+\beta\text{diag}(|\phi_k|^2)$$

is symmetric and positive definite with $\beta\geq 0$, where L is the discrete Laplacian matrix, and $\text{diag}(V)$ and $\text{diag}(|\phi_k|^2)$ denote the diagonal matrices with diagonal entries being the values of $V(x)$ and the squares of the values of $\phi_k(x)$ on mesh nodes.

Assume that ϕ is the (finite difference representation of) the eigenvector of the algebraic nonlinear eigenvalue problem

$$A\phi:=\left(-\frac{1}{2}L+\text{diag}(V)+\beta\text{diag}(|\phi|^2)\right)\phi=\mu\phi \tag{5.2}$$

normalized in ℓ^2 norm corresponding to the ground state solution of the GPE. For our subsequent analysis based on linear algebra, however, it is more convenient to rescale ϕ so that is instead normalized in the vector 2-norm. Let the original $\phi=c\tilde{\phi}$, where $\|\phi\|_{\ell^2}=1$ and $\|\tilde{\phi}\|_2=1$. We have $c=\|\phi\|_{\ell^2}/\|\tilde{\phi}\|_{\ell^2}=\|\tilde{\phi}\|_2/\|\tilde{\phi}\|_{\ell^2}$, and (5.2) can be written as

$$A\tilde{\phi}:=\left(-\frac{1}{2}L+\text{diag}(V)+\beta\frac{\|\tilde{\phi}\|_2^2}{\|\tilde{\phi}\|_{\ell^2}^2}\text{diag}(|\tilde{\phi}|^2)\right)\tilde{\phi}=\mu\tilde{\phi}.$$

For a given mesh/discretization and any vector $\phi\neq 0$, $\|\tilde{\phi}\|_2^2/\|\tilde{\phi}\|_{\ell^2}^2$ is a fixed constant. Therefore, by letting $\tilde{\beta}=\beta\|\tilde{\phi}\|_2^2/\|\tilde{\phi}\|_{\ell^2}^2$, we have a new equation in exactly the same form as (5.2), with ϕ and β replaced with $\tilde{\phi}$ (normalized in vector 2-norm) and $\tilde{\beta}$ (proportional to β but independent of ϕ), respectively. To simplify the notation for our analysis, we still work with the original equation (5.2), with ϕ normalized in vector 2-norm.

Note that A is entirely determined by the desired unknown ϕ. As A is symmetric and positive definite with $\beta\geq 0$, any eigenvalue of A is positive. It is clear that (5.1) is simply a variant of the well-known inverse power method for computing the eigenvalue

of smallest modulus. Define the matrix $D_k=\beta\text{diag}(|\phi_k|^2-|\phi|^2)$ (a diagonal matrix whose diagonal entries are the entries of the vector $\beta(|\phi_k|^2-|\phi|^2)$) such that

$$\begin{aligned}\|D_k\|&=\beta\big\||\phi_k|^2-|\phi|^2\big\|_\infty\le\beta\||\phi_k|+|\phi|\|_\infty\||\phi_k|-|\phi|\|_\infty\\&\le 2\beta\||\phi_k|-|\phi|\|_\infty\le 2\beta\|\phi_k-\phi\|=4\beta\sin\frac{\angle(\phi_k,\phi)}{2}.\end{aligned}\tag{5.3}$$

Suppose that $\sin\angle(\phi_k,\phi)$ is sufficiently small, so that $\|A^{-1}D_k\|\le\|A^{-1}\|\|D_k\|<1$. It follows that

$$\begin{aligned}(A+D_k)^{-1}&=\Big((I+D_kA^{-1})A\Big)^{-1}=A^{-1}(I+D_kA^{-1})^{-1}\\&=A^{-1}\Big(I-D_kA^{-1}(I+D_kA^{-1})^{-1}\Big),\end{aligned}\tag{5.4}$$

where

$$\begin{aligned}&\|D_kA^{-1}(I+D_kA^{-1})^{-1}\|\le\|D_k\|\|A^{-1}\|\|(I+D_kA^{-1})^{-1}\|\\\le&\frac{\|A^{-1}\|\|D_k\|}{1-\|A^{-1}\|\|D_k\|}\lesssim\frac{2\beta}{\lambda_1}\sin\angle(\phi_k,\phi).\end{aligned}$$

From (5.4), Step 1 of Picard-projection (5.1) can be written as

$$\begin{aligned}\hat{\phi}_{k+1}&=A_k^{-1}\phi_k=(A+D_k)^{-1}\phi_k\\&=A^{-1}\Big(\phi_k-D_kA^{-1}(I+D_kA^{-1})^{-1}\phi_k\Big).\end{aligned}\tag{5.5}$$

This shows that (5.1) is a perturbed inverse power method based on the limit matrix A, with a small perturbation vector $D_kA^{-1}(I+D_kA^{-1})^{-1}\phi_k$ subtracted from ϕ_k before the action of A^{-1}. This observation helps us establish a one-step local convergence of Picard-projection closely related to that of the standard inverse power method involving A.

Theorem 5.1. *Consider the algebraic GPE*

$$A\phi=\left(-\frac{1}{2}L+\text{diag}(V)+\beta\text{diag}(|\phi|^2)\right)\phi=\mu\phi,$$

where $V\ge 0$ and $\beta\ge 0$. Let $(\mu,\phi)=(\lambda_1,v_1)$ be the eigenpair representing the GPE ground state solution. Let λ_2 be the second smallest eigenvalue of A, and assume that $\lambda_1<\lambda_2$. Let ϕ_k be an approximation to ϕ, with $\|\phi_k\|=1$ and $\sin\angle(\phi_k,\phi)$ sufficiently small, and define $D_k=\beta\text{diag}(|\phi_k|^2-|\phi|^2)$. Then the new iterate ϕ_{k+1} computed by Picard-projection (5.1) *satisfies*

$$\tan\angle(\phi_{k+1},\phi)\le\frac{\lambda_1+2\beta\sin\angle(v_1,D_kv_1)}{\lambda_2}\tan\angle(\phi_k,\phi)+\mathcal{O}\left(\sin^2\angle(\phi_k,\phi)\right).\tag{5.6}$$

Proof. Let (λ_i,v_i) be the eigenpairs of A, with $0<\lambda_1<\lambda_2\leq\cdots\leq\lambda_n$, and $\|v_i\|=1$. Recall that $\lambda_1=\mu$ and $v_1=\phi$ are the eigenvalue and eigenvector of the algebraic nonlinear eigenvalue problem (GPE) corresponding to the ground state, and $(v_i,v_j)=0$ for $i\neq j$ since A is symmetric. Let

$$\phi_k=\sum_{i=1}^{n}c_iv_i,$$

with

$$\|\phi_k\|=\left(\sum_{i=1}^{n}c_i^2\right)^{\frac{1}{2}}=1,\quad\text{where}\quad\left(\sum_{i=2}^{n}c_i^2\right)^{\frac{1}{2}}=\sin\angle(\phi_k,v_1)$$

is assumed to be sufficiently small, and $c_1=(\phi_k,v_1)=\cos\angle(\phi_k,v_1)\approx1$. Let us define

$$q_k:=D_kA^{-1}(I+D_kA^{-1})^{-1}\phi_k=D_kA^{-1}\phi_k-\left(D_kA^{-1}\right)^2\phi_k+\left(D_kA^{-1}\right)^3\phi_k-\cdots,$$

where

$$\left\|\left(D_kA^{-1}\right)^{\ell}\right\|\leq\|D_k\|^{\ell}\|A^{-1}\|^{\ell}\leq\left(\frac{4\beta\sin\frac{\angle(\phi_k,\phi)}{2}}{\lambda_1}\right)^{\ell}\to0$$

exponentially with ℓ because $\sin\angle(\phi_k,v_1)$ is sufficiently small. Therefore q_k can be simply written as

$$q_k=D_kA^{-1}\phi_k+\mathcal{O}\left(\sin^2\angle(\phi_k,\phi)\right).\tag{5.7}$$

Substitute $\phi_k=\sum_{i=1}^{n}c_iv_i$ into (5.7). Since $\|D_k\|,|c_i|\leq\mathcal{O}\left(\sin\angle(\phi_k,\phi)\right)$ $(2\leq i\leq n)$, we have

$$q_k=\frac{c_1}{\lambda_1}D_kv_1+D_k\sum_{i=2}^{n}\frac{c_i}{\lambda_i}v_i+\mathcal{O}\left(\sin^2\angle(\phi_k,\phi)\right)=\frac{c_1}{\lambda_1}D_kv_1+\mathcal{O}\left(\sin^2\angle(\phi_k,\phi)\right).\tag{5.8}$$

Now let $q_k=\sum_{i=1}^{n}d_iv_i$, such that for each i,

$$\begin{aligned}d_i=(v_i,q_k)&=\frac{c_1}{\lambda_1}(v_i,D_kv_1)+\mathcal{O}\left(\sin^2\angle(\phi_k,\phi)\right)\\&\leq\frac{c_1}{\lambda_1}\|D_k\|\cos\angle(v_i,D_kv_1)+\mathcal{O}\left(\sin^2\angle(\phi_k,\phi)\right).\end{aligned}$$

It follows that

$$\begin{aligned}\left(\sum_{i=2}^{n}d_i^2\right)^{\frac{1}{2}}&\leq\frac{c_1}{\lambda_1}\|D_k\|\left(\sum_{i=2}^{n}\cos^2\angle(v_i,D_kv_1)\right)^{\frac{1}{2}}+\mathcal{O}\left(\sin^2\angle(\phi_k,\phi)\right)\\&=\frac{c_1}{\lambda_1}\|D_k\|\sin\angle(v_1,D_kv_1)+\mathcal{O}\left(\sin^2\angle(\phi_k,\phi)\right).\end{aligned}\tag{5.9}$$

With the work above, (5.5) leads to

$$\hat{\phi}_{k+1}=A^{-1}(\phi_k-q_k)=\frac{c_1-d_1}{\lambda_1}v_1+\sum_{i=2}^{n}\frac{c_i-d_i}{\lambda_i}v_i.$$

Recall that $c_1=\cos\angle(\phi_k,\phi)$, and note that

$$2c_1\sin\frac{\angle(\phi_k,\phi)}{2}\leq 2\cos\frac{\angle(\phi_k,\phi)}{2}\sin\frac{\angle(\phi_k,\phi)}{2}=\sin\angle(\phi_k,\phi).$$

Therefore,

$$\begin{aligned}
\tan\angle(\hat{\phi}_{k+1},\phi)&=\frac{\left(\sum_{i=2}^n\left(\frac{c_i-d_i}{\lambda_i}\right)^2\right)^{\frac{1}{2}}}{\frac{c_1-d_1}{\lambda_1}}\leq\frac{\left(\sum_{i=2}^n\left(\frac{c_i}{\lambda_i}\right)^2\right)^{\frac{1}{2}}+\left(\sum_{i=2}^n\left(\frac{d_i}{\lambda_i}\right)^2\right)^{\frac{1}{2}}}{\frac{c_1-d_1}{\lambda_1}}\\
&\leq\frac{\frac{1}{\lambda_2}\left(\sum_{i=2}^n c_i^2\right)^{\frac{1}{2}}+\frac{1}{\lambda_2}\left(\sum_{i=2}^n d_i^2\right)^{\frac{1}{2}}}{\frac{c_1-d_1}{\lambda_1}}\leq\frac{\lambda_1}{\lambda_2}\frac{\sin\angle(\phi_k,\phi)+2\beta\lambda_1^{-1}\sin\angle(v_1,D_kv_1)\left(2c_1\sin\frac{\angle(\phi_k,\phi)}{2}\right)}{c_1-d_1}\\
&\leq\frac{\lambda_1+2\beta\sin\angle(v_1,D_kv_1)}{\lambda_2}\tan\angle(\phi_k,\phi)+\mathcal{O}\left(\sin^2\angle(\phi_k,\phi)\right).
\end{aligned}\tag{5.10}$$

Finally, since $\phi_{k+1}=\hat{\phi}_{k+1}/\|\hat{\phi}_{k+1}\|$, the conclusion is established as

$$\angle(\hat{\phi}_{k+1},\phi)=\angle(\phi_{k+1},\phi).$$

Thus, we complete the proof. □

We consider now the local convergence of BENGF. Similar to Picard-projection, the local convergence of BENGF can also be analyzed from the perspective of a shift-invert power method. In matrix form, this method reads

$$\text{Step 1:}\quad \text{Solve}(I+\Delta tA_k)\hat{\phi}_{k+1}=\phi_k\text{for}\hat{\phi}_{k+1},\quad(\text{i.e.,}\quad\hat{\phi}_{k+1}=(I+\Delta tA_k)^{-1}\phi_k),\tag{5.11a}$$

$$\text{Step 2:}\quad\phi_{k+1}=\hat{\phi}_{k+1}/\|\hat{\phi}_{k+1}\|.\tag{5.11b}$$

with A_k defined the same as above for Picard-projection.

The local convergence analysis closely follows that of Picard-projection. Specifically, we can write Step 1 of BENGF as

$$\hat{\phi}_{k+1}=(I+\Delta tA+\Delta tD_k)^{-1}\phi_k,$$

where $D_k=\beta\text{diag}(|\phi_k|^2-|\phi|^2)$. Using the inverse of matrix addition (5.4), we have

$$\hat{\phi}_{k+1}=(I+\Delta tA)^{-1}\left(\phi_k-\Delta tD_k(I+\Delta tA)^{-1}\left(I+\Delta tD_k(I+\Delta tA)^{-1}\right)^{-1}\phi_k\right),\tag{5.12}$$

which is a perturbed shift-invert power method based on the limit matrix A, with a small perturbation vector subtracted from the standard result. In fact, (5.12) is obtained by simply replacing A and D_k with $I+\Delta tA$ and ΔtD_k, respectively, in (5.5). Thus, to derive the one-step local convergence, we need only replace D_k with ΔtD_k in (5.9), and replace λ_1 and λ_2 with $1+\Delta t\lambda_1$ and $1+\Delta t\lambda_2$, respectively, in (5.10). Then we have the following convergence result for BENGF.

Theorem 5.2. *Consider the algebraic GPE*

$$A\phi = \left(-\frac{1}{2}L + \text{diag}(V) + \beta\text{diag}(|\phi|^2)\right)\phi = \mu\phi,$$

where $V \geq 0$ and $\beta \geq 0$. Let $(\mu,\phi) = (\lambda_1, v_1)$ be the eigenpair representing the GPE ground state solution. Let λ_2 be the second smallest eigenvalue of A, and assume that $\lambda_1 < \lambda_2$. Let ϕ_k be an approximation to ϕ, with $\|\phi_k\| = 1$ and $\sin\angle(\phi_k,\phi)$ sufficiently small, and define $D_k = \beta\text{diag}(|\phi_k|^2 - |\phi|^2)$. Then the new iterate ϕ_{k+1} computed by BENGF$_h$ (5.11) satisfies

$$\tan\angle(\phi_{k+1},\phi) \leq \frac{1+\Delta t(\lambda_1 + 2\beta\sin\angle(v_1, D_k v_1))}{1+\Delta t\lambda_2}\tan\angle(\phi_k,\phi) + \mathcal{O}\left(\sin^2\angle(\phi_k,\phi)\right).$$

We comment on the connection between Picard-projection and BENGF. First, we pointed earlier that Picard-projection is equivalent to BENGF as $\Delta t \to \infty$. In terms of the local convergence rate, if $\beta\sin\angle(v_1, D_k v_1)$ is sufficiently small such that

$$\lambda_1 + 2\beta\sin\angle(v_1, D_k v_1) < \lambda_2,$$

then the bound on BENGF convergence factor

$$\frac{1+\Delta t(\lambda_1 + 2\beta\sin\angle(v_1, D_k v_1))}{1+\Delta t\lambda_2}$$

achieves its minimum

$$\frac{\lambda_1 + 2\beta\sin\angle(v_1, D_k v_1)}{\lambda_2} \quad \text{as} \quad \Delta t \to \infty,$$

which is exactly the result (5.6) for Picard-projection. This suggests that for small β, BENGF may converge more rapidly with larger step size Δt, and Picard-projection is likely to outperform BENGF asymptotically. This has been verified numerically in our 1D test for $\beta = 3.1371$; see Table 3. On the other hand, if $\beta\sin\angle(v_1, D_k v_1)$ is large such that $\lambda_1 + 2\beta\sin\angle(v_1, D_k v_1) > \lambda_2$, then we would not get a bound on the convergence factor that is less than 1 for either method. The relative merit of the two methods for GPE with large β deserves further exploration, though numerically, Table 3 suggests that BENGF with optimal step size Δt outperforms Picard-projection.

6 Conclusions

We have considered analytically and numerically the effect of AA on Picard-projection and backward Euler normalized gradient flow iterations for solving the stationary GPE. Under a finite element discretization, we proved that both of the iterations fit into the AA analysis framework from [30], which allowed us to prove that AA improves the linear convergence rates of the methods by a multiplicative factor representing the gain of the

AA optimization problem, but adds additional higher order terms to the residual expansion. Numerical tests show substantial improvement provided by AA for both methods, especially for larger β. Finally, a local convergence analysis of both methods is given. There are two main conclusions from this paper: First, AA provides significant improvement in both methods. Second, although there are differences between Picard-projection and BENGF iterations, when AA is used with a larger time step size they behave very similar. This second result is important since BENGF has a parameter Δt, and it seems that by using AA Picard-projection one does not need this parameter. Future directions include extending this work to the important case of rotational GPE.

Acknowledgements

Author LR acknowledges support from National Science Foundation grant DMS 2011490, and author FX acknowledges support from National Science Foundation grant DMS 1819097.

References

[1] H. AN, X. JIA, AND H. F. WALKER, *Anderson acceleration and application to the three-temperature energy equations,* J. Comput. Phys., 347 (2017), pp. 1–19.
[2] D. G. ANDERSON, *Iterative procedures for nonlinear integral equations,* J. Assoc. Comput. Mach., 12(4) (1965), pp. 547–560.
[3] X. ANTIONE AND R. DUBOSCQ, *GPELab, a Matlab toolbox to solve Gross-Pitaevskii equations I: Computation of stationary solutions,* Comput. Phys. Commun., 185 (2014), pp. 2969–2991.
[4] X. ANTOINE, A. LEVITT, AND Q. TANG, *Efficient spectral computation of the stationary states of rotating Bose-Einstein condensates by preconditioned nonlinear conjugate gradient methods,* J. Comput. Phys., 343 (2017), pp. 92–109.
[5] Z. BAI, R.C. LI, AND D. LU, *Optimal convergence rate of self-consistent field iteration for solving eigenvector-dependent nonlinear eigenvalue problems,* arXiv preprint arXiv:2009.09022, 2020.
[6] W. BAO AND Y. CAI, *Mathematical theory and numerical methods for Bose-Einstein condensation,* Kinet. Relat. Models, 6 (2013), pp. 1–135.
[7] W. BAO AND Q. DU, *Computing the ground state solution of Bose-Einstein condensates by a normalized gradient flow,* SIAM J. Sci. Comput., 25(5) (2004), pp. 1674–1697.
[8] S. BRENNER AND L. R. SCOTT, The Mathematical Theory of Finite Element Methods, 3rd edition, Springer-Verlag, 2008.
[9] H. CHEN, L. HE, AND A. ZHOU, *Finite element approximations of nonlinear eigenvalue problems in quantum physics,* Comput. Methods Appl. Mech. Eng., 200 (2011), pp. 1846–1865.
[10] A. DIEGEL, C. WANG, AND S. M. WISE, *Stability and convergence of a second order mixed finite element method for the Cahn-Hilliard equation,* IMA J. Numer. Anal., 36 (2016), pp. 1867–1897.
[11] Q. DU, L. JU, X. LI, AND Z. QIAO, *Maximum principle preserving exponential time differencing schemes for the nonlocal Allen-Cahn equation,* SIAM Review, in press, 2021.
[12] C. EVANS, S. POLLOCK, L. REBHOLZ, AND M. XIAO, *A proof that Anderson acceleration increases the convergence rate in linearly converging fixed point methods (but not in quadratically converging ones),* SIAM J. Numer. Anal., 58 (2020), pp. 788–810.

[13] H. FANG AND Y. SAAD, *Two classes of multisecant methods for nonlinear acceleration,* Numer. Linear Algebra Appl., 16(3) (2009), pp. 197–221.
[14] W. FENG, Z. GUAN, J. LOWENGRUB, C. WANG, S. WISE, AND Y. CHEN, *A uniquely solvable, energy stable numerical scheme for the functionalized Cahn-Hilliard equation and its convergence analysis,* J. Sci. Comput., 76 (2018), pp. 1938–1967.
[15] W. FENG, A. SALGADO, C. WANG, AND S. WISE, *Preconditioned steepest descent methods for some nonlinear elliptic equations involving p-Laplacian terms,* J. Comput. Phys., 334 (2017), pp. 45–67.
[16] W. FENG, C. WANG, S. WISE, AND Z. ZHANG, *A second-order energy stable backward differentiation formula method for the epitaxial thin film equation with slope selection,* Numer. Methods Partial Differential Equations, 34(6) (2018), pp. 1975–2007.
[17] H. GOMEZ AND T. HUGHES, *Provably unconditionally stable, second-order time-accurate, mixed variational methods for phase-field models,* J. Comput. Phys., 230 (2011), pp. 5310–5327.
[18] J. GUO, C. WANG, S. WISE, AND X. YUE, *An h^2 convergence of a second-order convex-splitting, finite difference scheme for the three-dimensional Cahn-Hilliard equation,* Commun. Math. Sci., 14(2) (2016), pp. 489–515.
[19] N. HIGHAM AND N. STRABIC, *Anderson acceleration of the alternating projections method for computing the nearest correlation matrix,* Numer. Algorithms, 72 (2016), pp. 1021–1042.
[20] E. JARLEBRING, S KVAAL, AND W. MICHIELS, *An inverse iteration method for eigenvalue problems with eigenvector nonlinearities,* SIAM J. Sci. Comput., 36 (2014), pp. A1978–A2001.
[21] E. JARLEBRING AND P. UPADHYAYA, *Implicit algorithms for eigenvector nonlinearities,* arXiv preprint arXiv:2002.12805, 2020.
[22] L. JU, J. ZHANG, AND Q. DU, *Fast and accurate algorithms for simulating coarsening dynamics of Cahn-Hilliard equations,* Comput. Mater. Sci., 108 (2015), pp. 272–282.
[23] S. WISE K. CHENG, AND C. WANG, *An energy stable bdf2 fourier pseudo-spectral numerical scheme for the square phase field crystal equation,* submitted, 2021.
[24] C. T. KELLEY, *Numerical methods for nonlinear equations,* Acta Numer., 27 (2018), pp. 207–287.
[25] L. LANDAU AND E. LIFSCHITZ, Quantum Mechanics: Non-Relativistic Theory, Pergamon Press, New York, 1977.
[26] Y. PENG, B. DENG, J. ZHANG, F. GENG, W. QIN, AND L. LIU, *Anderson acceleration for geometry optimization and physics simulation,* ACM Trans. Graph., 37(4) (2018).
[27] L. P. PITAEVSKII, *Vortex lines in an imperfect Bose gas,* Soviet Phys. JETP, 13 (1961), pp. 451–454.
[28] S. POLLOCK, L. REBHOLZ, AND M. XIAO, *Anderson-accelerated convergence of Picard iterations for incompressible Navier-Stokes equations,* SIAM J. Numer. Anal., 57(2) (2019), pp. 615–637.
[29] S. POLLOCK, L. REBHOLZ, AND M. XIAO, *Acceleration of nonlinear solvers for natural convection problems,* J. Numer. Math., in press, 2020.
[30] S. POLLOCK AND L. G. REBHOLZ, *Anderson acceleration for contractive and noncontractive iterations,* IMA J. Numer. Anal., in press, 2020.
[31] J. SHEN AND X. YANG, *Numerical approximations of Allen-Cahn and Cahn-Hilliard equations,* Discret. Contin. Dyn. Syst., 28 (2010), pp. 1669–1691.
[32] P. STASIAK AND M. W. MATSEN, *Efficiency of pseudo-spectral algorithms with Anderson mixing for the SCFT of periodic block-copolymer phases,* Euro. Phys. J. E, 34110 (2011), pp. 1–9.
[33] T. TIAN, Y. CAI, X. WU, AND Z. WEN, *Ground states and their characterization of spin-F Bose-Einstein condensates,* arXiv preprint arXiv:1907.01194, 2019.
[34] A. TOTH AND C. T. KELLEY, *Convergence analysis for Anderson acceleration,* SIAM J. Numer. Anal., 53(2) (2015), pp. 805–819.

[35] A. TOTH, C. T. KELLEY, S. SLATTERY, S. HAMILTON, K. CLARNO, AND R. PAWLOWSKI, *Analysis of Anderson acceleration on a simplified neutronics/thermal hydraulics system,* Proceedings of the ANS MC2015 Joint International Conference on Mathematics and Computation (M&C), Supercomputing in Nuclear Applications (SNA) and the Monte Carlo (MC) Method, ANS MC2015 CD:1–12, 2015.

[36] G. VERGEZ, I. DANAILA, S. AULIAC, AND F. HECHT, *A finite element toolbox for the stationary Gross-Pitaevskii equation with rotation,* Comput. Phys. Commun., 209 (2016), pp. 144–162.

[37] H. F. WALKER AND P. NI, *Anderson acceleration for fixed-point iterations,* SIAM J. Numer. Anal., 49(4) (2011), pp. 1715–1735.

[38] Y. YAN, W. CHEN, C. WANG, AND S. M. WISE, *A second-order energy stable BDF numerical scheme for the Cahn-Hilliard equation,* Commun. Comput. Phys., 23 (2018), pp. 572–602.

[39] C. ZHANG, J. OUYANG, C. WANG, AND S. M. WISE, *Numerical comparison of modified-energy stable SAV-type schemes and classical BDF methods on benchmark problems for the functionalized Cahn-Hilliard equation,* J. Comput. Phys., 423 (2020), pp. 109772.