Editorial Manager(tm) for Journal of Computational Physics
Manuscript Draft

Manuscript Number:

Title: An Implementation of an Exact Finite Reduction Scheme for Semilinear Dirichlet Problems

Article Type: Regular Article

Section/Category:

Keywords: Nonlinear PDEs,Nonlinear equations, Finite Reduction, Newton Method, Fixed Point iteration

Corresponding Author: Dr Alberto Lovison, PhD

Corresponding Author's Institution: University of Padua, Italy

First Author: Franco Cardin

Order of Authors: Franco Cardin; Alberto Lovison; Mario Putti

Manuscript Region of Origin:

Abstract:

An Implementation of an Exact Finite Reduction Scheme for Semilinear Dirichlet Problems

Franco Cardin, [a] Alberto Lovison [a] and Mario Putti [b]

[a] *Dipartimento di Matematica Pura ed Applicata*
Università degli Studi di Padova
via Belzoni 7, I – 35131 Padova, Italy

[b] *Dipartimento di Metodi e Modelli Matematici per le Scienze Applicate*
Università degli Studi di Padova
via Belzoni 7, I – 35131 Padova, Italy

Abstract

In this paper we study a semilinear Dirichlet problem applying a non–local Lyapunov-Schmidt type reduction originally devised for field theory. A numerical algorithm is developed on the basis of the discretization of the differential operator by means of simple finite differences. The eigendecomposition of the resulting matrix is used to implement a discrete version of the reduction process. By the new algorithm the problem is decomposed into two coupled subproblems of different dimensions. A large subproblem is solved by means of a fixed point iteration completely controlled by the features of the original equation. The other problem has dimensions that can be made much smaller than the former, and inherits most of the nonlinear difficulties of the original system. The advantage of this approach is that sofisticated linearization strategies can be used to solve this small nonlinear system, at the expense of a partial eigendecomposition of the discretized linear differential operator. The proposed scheme is used for the solution of a simple nonlinear one dimensional problem. The applicability of the procedure is tested and experimental convergence estimates are consolidated. Numerical results are used to show the performance of the new algorithm.

Key words: Nonlinear PDEs, Nonlinear equations, Finite Reduction, Newton Method, Fixed point iteration

Introduction

The numerical solution of nonlinear Partial Differential Equations (PDEs) of diffusion type relies almost exclusively on the Newton linearization algorithm, possibly complemented by globalization techniques and implementation approaches that try to decrease the computational burden of the scheme (e.g. inexact Newton, quasi-Newton Jacobian updates, etc.) [8; 3; 9; 11]. The Jacobian matrix can be difficult to evaluate and can often be nearly singular because of the presence of non smooth derivatives in the nonlinear terms. In these cases derivative terms of the Jacobian matrix can be neglected but convergence of the resulting iteration is not guaranteed anymore. Techniques based on secant matrices or on finite difference evaluation of the Jacobian have been devised [7; 5; 11]. Alternatively, the nonsmooth nonlinear functions are replaced by smooth, often spline based, interpolations, where the Newton approach can be safely applied [12].

For the cases of nonlinearities confined in the forcing term F, as in the semilinear PDE $-Lu = F(u)$ (L elliptic), a reformulation can be effectively used to alleviate these problems. To this aim we propose a numerical implementation of a technique based on a global finite parameters reduction proposed in field theory by [2; 4] for a steady state nonlinear diffusion problem. This reduction approach originates from the ideas developed by [1; 6] for Hamiltonian systems. Exploiting the eigendecomposition of L, the solution is first expressed by means of the Green operator. The solution space is then splitted into the sum of a finite (head) and an infinite (tail) dimensional subspaces, leading to a reformulation of the original PDE into two functional fixed point problems. Because of the contractivity feature arising from the natural hierarchy among the eigenspaces of the elliptic operator, the infinite dimensional fixed point problem (the tail) always possesses a unique solution and its contraction factor can be completely controlled. As a matter of fact, the fixed point of the tail is uniquely determined for any assigned candidate solution of the head. As opposed to the infinite problem, nothing can be said about existence and multiplicity for the head, as it inherits most of the potential nonlinear difficulties of the original PDE. However, after the splitting is performed, the fixed point of the tail can be substituted into the head, obtaining a formally finite dimensional system. The solution sets of the original PDE and of the new finite dimensional system correspond each other isomorphically, and thus the two problems are fully equivalent [1]. Note that the dimension of the reduced system can be made as small as allowed by the intrisic features of the PDE, such as for example the topological complexity of the solution set. This is

[1] In other words, a sort of "holography principle" takes place, as for example, by Gabor procedure, the information contained in a real life three-dimensional scene can be completely recorded on a two-dimensional picture.

essentially what we have called exact finite reduction.

Translation of the reduction approach in a numerical framework requires the discretization of the differential operator. To this aim we project the PDE onto a finite dimensional space by means of a finite difference scheme. The discrete operator corresponding to L is a matrix whose eigenvalues and eigenvectors are employed to define the numerical Green function and to determine the splitting of the space $\mathbb{R}^n$, where n is the size of the finite difference grid. For any fixed n, the reduction technique applies directly to the discretized problem without loss of accuracy or information. More precisely, the splitting produces two finite fixed point problems, one defined in $\mathbb{R}^m$ and one in its complement $\mathbb{R}^{n-m}$. By construction, the latter problem always admits a unique solution, and the contraction factor is determined as a function of m. Thus, convergence of this problem by means of Picard iteration can be attained in a pre-assigned number of iterations, while most of the nonlinear difficulties are concentrated in the fixed point problem defined in $\mathbb{R}^m$. The main appeal of this procedure is that, whenever we are allowed to choose $m \ll n$, sofisticated linearization techniques can be effectively used on a much smaller system, maintaining a fixed number of simple iterations in the solution of the larger dimensional problem. Furthermore, because of the elliptic character of L, the Green operator can be defined by a partial eigenspectrum, drastically alleviating the overall computational burden.

The applicability of the developed algorithm is tested for the solution of a simple one dimensional model problem admitting an analytical solution. The numerical results are used to verify the theoretical convergence estimates and show that the proposed scheme is competitive with the more standard Newton Raphson method.

1 Analytical setting

Our investigation takes place in $H := H_0^1(\Omega, \mathbb{R}^k)$, $\Omega \subset \mathbb{R}^d$, where we consider a non linear perturbation F of an elliptic operator L. We are trying to solve the following simple Dirichlet boundary value problem:

$$\begin{cases} -Lu = F(u), & \text{in } \Omega, \\ u = 0, & \text{on } \partial\Omega. \end{cases} \tag{1}$$

Assume the nonlinear operator $F : H \to H$ is a Nemitski operator, i.e., $F(u) := f \circ u$, where $f : \mathbb{R}^k \to \mathbb{R}^k$ is Lipschitz,

$$|f(s_1) - f(s_2)| \leq C\,|s_1 - s_2|\,. \tag{2}$$

The core of the method consists in the spectral decomposition of H w.r.t. the eigenspace of $-L$, and in the exploitation of the Green operator $g = (-L)^{-1}$, i.e., $g : H \to H$, $g \circ (-L) = -L \circ g = id_H$. The problem is translated through g and successively decomposed into a finite and an infinite part by means of a suitable cut-off.

Here is an outline of these steps. The spectral decomposition and the Green operator of (1) are given by:

$$-Lw_j = \lambda_j w_j, \qquad \langle w_i, w_j \rangle = \delta_{ij}, \qquad 0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots. \tag{3}$$

$$g(v) = g\left(\sum_{j=1}^{+\infty} v_j w_j\right) = \sum_{j=1}^{+\infty} v_j \frac{1}{\lambda_j} w_j, \tag{4}$$

thus the cut-off of the space H is written as:

$$v = \sum_{j=1}^{+\infty} v_j w_j = \sum_{j=1}^{m} v_j w_j + \sum_{j=m+1}^{+\infty} v_j w_j \in H, \tag{5}$$

$$v = \mathbb{P}_m v + \mathbb{Q}_m v = \mu + \eta, \qquad H = \mathbb{P}_m H \oplus \mathbb{Q}_m H. \tag{6}$$

Here the crucial starting point: we are going to search solutions of (1) represented by the form: $u = g(v)$, for suitable $v \in H$,

$$\begin{aligned} -Lu &= F(u), \\ -L\left(g\left(v\right)\right) &= F\left(g\left(v\right)\right), \quad v = \mu + \eta, \\ \mu + \eta &= F(g(\mu + \eta)), \end{aligned} \tag{7}$$

so the problem is splitted into

$$\begin{aligned} &\eta = \mathbb{Q}_m F(g(\mu + \eta)) \quad \text{infinite part} \quad \text{(tail)} \\ &\\ &\mu = \mathbb{P}_m F(g(\mu + \eta)) \quad \text{finite part} \quad \text{(head)} \end{aligned} \tag{8}$$

The infinite part of the equation, for suitable fixed cut-off m, is uniquely solved, for every fixed finite part $\mu \in \mathbb{P}_m H$. Indeed the map

$$\begin{aligned} \mathbb{Q}_m H &\longrightarrow \mathbb{Q}_m H \\ \eta &\longmapsto \mathbb{Q}_m F(g(\mu + \eta)), \end{aligned} \tag{9}$$

is contractive provided m is suitably large. Using the Lipschitz constant C of F and recalling the monotone character of the spectral sequence $\{\lambda_j\}$, we

obtain:

$$\|\mathbb{Q}_m F(g(\mu+\eta_1)) - \mathbb{Q}_m F(g(\mu+\eta_2))\| \leq$$
$$\leq C\,\|g(\mu+\eta_1) - g(\mu+\eta_2)\| \leq C\frac{1}{\lambda_{m+1}}\,\|\eta_1-\eta_2\|\,.$$

Thus, we can choose $m \in \mathbb{N}$ large enough to achieve $\frac{C}{\lambda_{m+1}} < 1$, so that the *unique* fixed point $\tilde{\eta}(\mu)$ of this contraction solves the tail of equation (8). It can be easily proved that the fixed point $\tilde{\eta}(\mu)$ inherits the regularity of F, being expressible as the implicit function of an equation involving F:

$$\mathcal{F}(\mu,\eta) = 0, \qquad \mathcal{F}(\mu,\eta) := \mu - \mathbb{Q}_m F(g(\mu+\eta)).$$

By substituting $\tilde{\eta}(\mu)$ into the head, we get a *finite dimensional problem*:

$$\mu = \mathbb{P}_m F\left(g\left(\mu+\tilde{\eta}\left(\mu\right)\right)\right), \qquad \mu \in \mathbb{R}^m. \tag{10}$$

In spite of the finiteness of equation (10), in general we have not an *a priori* control about existence and uniqueness of the solution; more precisely, we could find no solutions, or many solutions, and possible bifurcation phenomena could happen for increasing (Lipschitz constant C of) F. Every solution μ^* of (10) gives rise to a solution of the original nonlinear Dirichlet problem:

$$u = g\left(\mu^* + \tilde{\eta}\left(\mu^*\right)\right).$$

Conversely, in corrispondence to each solution u of (1) there exists exactly one solution μ of (10).

Remark 1 Clearly the proposed procedure recalls the Lyapunov–Schmidt reduction technique. In particular equation (10) recalls the 'bifurcation' equation. Hypothesis (2) overcomes the locality feature of the classical Lyapunov–Schmidt technique.

2 Numerical discretization

The previously outlined procedure can be implemented in a numerical framework by substituting the differential operator of the PDE with its discretized version. Using a finite difference approach, denoting by T_h a generic discretization of Ω, formed by n nodes and N subdivisions with characteristic mesh size h, the discrete elliptic operator reduces to a symmetric positive-definite matrix $-L_h$ (we assume at least one Dirichlet boundary condition is imposed). The numerical solution vector $u_h \in H_h := \mathbb{R}^n$, is given by the solution of the system of the nonlinear algebraic equations:

$$-L_h u_h = F_h(u_h), \tag{11}$$

where F_h is the discretization of the nonlinear function operator F.

The symmetric eigenproblem of the corresponding linear system can be written as:

$$-L_h w = \lambda w, \tag{12}$$

where

$$\begin{gathered} 0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n, \\ w_1, w_2, \ldots, w_n, \qquad \langle w_i, w_j \rangle = \delta_{ij}, \end{gathered}$$

are the real positive eigenvalues and the corresponding eigenvectors. Note that the eigenvalues and eigenvectors thus defined converge to the eigenvalues and eigenfunctions of the continuous problem (1) in the limit when $h \to 0$ and $n \to +\infty$ [10].

Remark 2 The smallest eigenvalues of the finite difference problem (12) converge to fixed values when the nodal spacing of the discretization is made arbitrarily small, i.e., *when the dimension of the linear system (12) goes to infinity. Hence the leftmost eigenpairs of the eigenproblems converge to the same leftmost eigenspectrum.*

Remark 3 The largest eigenvalues cannot converge and grow as n^2 *when* $h \to 0$, *consistently with theory. This suggests that the highest frequencies are in essence inversely related to the magnitude of the discretization error, while the lowest frequencies represent the fundamental natural modes of the physical system described by the PDE [10].*

In analogy to the continuous case (3), the discrete Green operator g_h of $-L_h$ can be written with respect to the basis of H_h given by the eigenvectors of $-L_h$:

$$g_h(w_k) = \left(-L_h^{-1}\right)(w_k) := \frac{1}{\lambda_k} w_k \qquad k = 1, \ldots, n.$$

Since $-L_h$ is s.p.d., any vector $v \in H_h$ can be expressed as:

$$v = a_1 w_1 + \cdots + a_n w_n = Wa,$$

where W denotes the matrix whose columns are the eigenvectors w_k of $-L_h$:

$$W := [w_1, \ldots, w_n].$$

The Green operator applied to v gives:

$$g_h(v) = \frac{a_1}{\lambda_1} w_1 + \cdots + \frac{a_n}{\lambda_n} w_n = W \Lambda^{-1} a,$$

where Λ is the diagonal matrix of the eigenvalues (ordered accordingly to the corresponding eigenvectors).

The algorithm described in the previous section applies directly to the discretized problem, and proceeds as follows. For a given m, the vector space H_h is split into two subspaces $P_m H_h$ and $Q_m H_h$, where $P_m H_h \subseteq H_h$ is generated by the first m eigenvectors $w_1, \ldots, w_m$, while $Q_m H_h \subseteq H_h$ is generated by $w_{m+1}, \ldots, w_n$. Consequently, the projectors P_m and Q_m, which are the discrete counterparts of $\mathbb{P}_m$ and $\mathbb{Q}_m$ in (6), can be explicitly written by means of the two matrices V_1 and V_2:

$$V_1 := [w_1, \ldots, w_m], \qquad V_2 := [w_{m+1}, \ldots, w_n], \qquad [V_1, V_2] = W.$$

For every $v = \hat{\mu} + \hat{\eta} \in H_h$, we have:

$$\hat{\mu} := P_m v = V_1 V_1^T v = V_1 a', \qquad a' \in \mathbb{R}^m, \tag{13}$$
$$\hat{\eta} := Q_m v = V_2 V_2^T v = V_2 a'', \qquad a'' \in \mathbb{R}^{n-m}. \tag{14}$$

The discrete version of (7) becomes then:

$$-L_h u = F_h(u), \tag{15}$$
$$-L_h\left(g_h(v)\right) = F_h\left(g_h(v)\right), \tag{16}$$
$$\hat{\mu} + \hat{\eta} = F_h\left(g_h(\hat{\mu} + \hat{\eta})\right). \tag{17}$$

The numerical algorithm is thus formed by two finite dimensional fixed point iterations:

$$\hat{\eta} = Q_m F_h\left(g_h(\hat{\mu} + \hat{\eta})\right), \qquad \in Q_m H_h \cong \mathbb{R}^{n-m}, \tag{18}$$
$$\hat{\mu} = P_m F_h\left(g_h(\hat{\mu} + \hat{\eta})\right), \qquad \in P_m H_h \cong \mathbb{R}^m, \tag{19}$$

with (18) satisfying:

$$\left\| Q_m F_h\left(g_h(\hat{\mu} + \hat{\eta}_1)\right) - Q_m F_h\left(g_h(\hat{\mu} + \hat{\eta}_2)\right) \right\| \leq \frac{C}{\lambda_{m+1}} \left\| \hat{\eta}_1 - \hat{\eta}_2 \right\|.$$

To prove the last assertion, we first note that F_h is Lipschitz whenever f is, i.e., for every $u = \{u_i\}, \bar{u} = \{\bar{u}_i\} \in H_h$,

$$\left\| F_h(u) - F_h(\bar{u}) \right\| = \left\| \begin{pmatrix} f(u_1) - f(\bar{u}_1) \\ \vdots \\ f(u_n) - f(\bar{u}_n) \end{pmatrix} \right\| =$$
$$= \left\| \begin{pmatrix} c_1(u_1 - \bar{u}_1) \\ \vdots \\ c_n(u_n - \bar{u}_n) \end{pmatrix} \right\| \leq C \left\| u - \bar{u} \right\|,$$

where C is the Lipschitz constant of f, while $c_1, \ldots, c_n$ are suitable positive numbers such that $c_j \leq C, \forall j$. The proof is completed by considering that:

$$\|g_h(\hat{\mu} + \hat{\eta}_1) - g_h(\hat{\mu} + \hat{\eta}_2)\| = \|g_h(\hat{\eta}_1 - \hat{\eta}_2)\| \leq \frac{1}{\lambda_{m+1}} \|\hat{\eta}_1 - \hat{\eta}_2\| .$$

The value of m is chosen so that the Lipschitz constant $\frac{C}{\lambda_{m+1}}$ becomes sufficiently smaller than 1 and guarantees the convergence towards the fixed point $\tilde{\eta}(\mu)$ within a predefined number of iterations of (18). By means of the solution of (18) the following fixed point problem of size m can be defined:

$$\hat{\mu} = P_m F_h \left(g_h \left(\hat{\mu} + \tilde{\eta}(\hat{\mu}) \right) \right) . \tag{20}$$

This system can be solved by any iterative strategy (including Peano-Picard or Newton–Raphson), but we do not have estimates on contractivity.

Numerical performance does not allow the full solution of the eigenproblem (12), which has been assumed so far. However, the ellipticity of $-L_h$ suggests that the discrete Green function tends to zero rapidly, according to the continuous case. In the following section we will apply the proposed scheme to a sample problem and will evaluate experimentally the minimum size of the eigensolution of (12) that gives a predefined accuracy.

Remark 4 In analogy with the continuous case, the reduced problem (20) defined in the small space $\mathbb{R}^m$ is equivalent to the original problem (11) in $\mathbb{R}^n$.

2.1 Implementation and sample test

Consider the one-dimensional Laplacian operator on the unit interval;

$$L := \Delta = \frac{\partial^2}{\partial x^2}, \qquad \Omega := [0,1] \subseteq \mathbb{R}, \qquad \mathcal{H} := H_0^1 \left([0,1], \mathbb{R} \right) , \tag{21}$$

which, after the discretization with characteristic length $h = \frac{1}{n+1}$, and elimination from $-L_h$ of the two Dirichlet boundary conditions, is represented by the well known tridiagonal matrix:

$$-L_h = \frac{1}{h^2} \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & \ddots & \ddots & \\ & & \ddots & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} . \tag{22}$$

operating on the finite dimensional space $H_h := \mathbb{R}^n$. By very classical computations, the eigenpairs of (12) are found to be:

$$\lambda_k = 4(n+1)^2 \sin^2\left(\frac{k\pi}{2(n+1)}\right), \qquad k = 1, \ldots, n, \tag{23}$$

$$w_k = \{w_{k,i}\}_{i=1,\ldots,n} = \left\{\sqrt{\frac{2}{n+1}} \sin\left(\frac{k\pi}{n+1} i\right)\right\}, \qquad k = 1, \ldots, n. \tag{24}$$

For $n \to \infty$, the leftmost eigenvalues converge to the quantities

$$\lambda_k(-L_h) \longrightarrow (k\pi)^2, \qquad k = 1, \ldots, n, \tag{25}$$

while the eigenvectors behave as:

$$w_k \longrightarrow \sin(k\pi x), \qquad k = 1, \ldots, n, \quad 0 \le x \le 1. \tag{26}$$

The largest eigenvalue is provided by (23) with $k = n$:

$$\lambda_n(-L_h) = \frac{4}{h^2} \sin^2\left(\frac{\pi}{2} \frac{n}{n+1}\right) \approx \frac{4}{h^2}. \tag{27}$$

Note that the eigenpairs satisfy Remarks 2 and 3.

Our sample test considers the Nemitski operator F associated to the function

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = \alpha\left(1 - e^{-\frac{x^2}{2}}\right).$$

Note that F is Lipschitz, with constant

$$C = \sup |f'| = \alpha/\sqrt{e} \approx 18.1959 \qquad (\alpha = 30).$$

Discretizing with $n = 80$ nodes, we can look for a suitable eigenvalue to perform the cutoff. The first candidates for contractive factors M are found to be

$$M^{(1)} = \frac{|f'|}{\lambda_1} \approx 1.84364, \qquad M^{(2)} = \frac{|f'|}{\lambda_2} \approx 0.460912, \qquad M^{(3)} = \frac{|f'|}{\lambda_3} \approx 0.204852, \qquad \ldots$$

We set $m = 2$, so that the contractive factor in (18) is $M^{(3)}$, thus at most $k := \text{int}\left(-i/\log(M^{(3)})\right) + 1$ iterations are needed to reduce the initial error by a factor 10^{-i}. The splitting is organized as follows:

$$H = \mathbb{R}^n = P_2 H \oplus Q_2 H,$$
$$v = \hat{\mu} + \hat{\eta} = a'_1 w_1 + a'_2 w_2 + a''_1 w_3 + \cdots + a''_{n-2} w_n,$$

Table 1
Convergence of the Peano-Picard iterations applied to (18) with $m = 2$, $\mu_0 = 500w_1 + 500w_2$ and using the complete eigenspectrum of (22), $(n = 640, l = 640)$.

j	$\varepsilon_\eta^{(j)}$	M	j	$\varepsilon_\eta^{(j)}$	M
1	8.21310×10^{1}	$\star$	11	6.33852×10^{-7}	0.169058
2	5.56597×10^{0}	0.067770	12	1.07158×10^{-7}	0.169058
3	9.43390×10^{-1}	0.169492	13	1.81159×10^{-8}	0.169059
4	1.60178×10^{-1}	0.169790	14	3.06254×10^{-9}	0.169053
5	2.71269×10^{-2}	0.169355	15	5.17750×10^{-10}	0.169059
6	4.58876×10^{-3}	0.169159	16	8.75862×10^{-11}	0.169167
7	7.75914×10^{-4}	0.169090	17	1.47106×10^{-11}	0.167956
8	1.31182×10^{-4}	0.169067	18	2.57688×10^{-12}	0.175171
9	2.21776×10^{-5}	0.169060	19	4.18135×10^{-13}	0.162264
10	3.74931×10^{-6}	0.169058	20	1.28513×10^{-13}	0.307348

To show the convergence characteristics of map (18), we perform several iterations by applying the Peano-Picard procedure. All the eigenpairs of (22) are employed in the calculations. The effects of using a partial eigenspectrum ($l \ll n$) will be addressed in the next section.

We fix as a first guess $\hat{\mu}^{(0)} = 500w_1 + 500w_2$ (w_1 and w_2 are the first two eigenvectors), and randomly generate $\hat{\eta}^{(0)}$. Denoting by $\hat{\eta}^{(j)}$ the j-th iterate of the contraction map, we expect the following estimate to be fulfilled:

$$\left\| \hat{\eta}^{(j+1)} - \hat{\eta}^{(j)} \right\| \leq 0.21 \left\| \hat{\eta}^{(j)} - \hat{\eta}^{(j-1)} \right\|,$$

The results are reported in Table 1. After 18 iterations, the L^2 norm of the difference between two successive iterations, $\varepsilon_\eta^{(j)} = \left\| \hat{\eta}^{(j)} - \hat{\eta}^{(j-1)} \right\|_{L^2}$, becomes smaller than 10^{-11} with a contractive factor of approximately 0.17. Note that the actual value of M is always smaller than the theoretical predictions and seems to stabilize after the 3^{rd} iteration. After 16 iterations, small oscillations appear due to round-off errors. Similar behavior is found when changing the initial guess $\mu^{(0)}$ or increasing the number of nodes $n = 160, 320, 640$. The number of iterations changes slightly, while M always converges to about 0.17.

The solution of the full problem is obtained by solving (20). Here we iterate by means again of the Peano-Picard procedure, though we do not possess any contraction result. At each of the iterations of the $\hat{\mu}$–map we have to solve the $\hat{\eta}$–map. To ensure convergence of the latter we perform a fixed number of iterations equal to 20. This allows $\varepsilon_\eta^{(20)}$ to become always smaller than 10^{-12}.

Table 2
Convergence behavior of the complete map (20) starting with $m = 2$, $\mu^{(0)} = 100w_1 - 100w_2$ and using the complete eigenspectrum of (22), $(n = 640, l = 640)$.

j	$\mu^{(j)}$	$\varepsilon_\mu^{(j)}$	M
1	$(230.589, -40.3486)$	6.7037×10^1	$\star$
2	$(283.236, -16.8716)$	5.7644×10^1	0.859881
3	$(369.349, -6.96820)$	8.6680×10^1	1.503710
4	$(482.325, -2.46623)$	1.1307×10^2	1.304410
5	$(573.516, -0.605729)$	9.1209×10^1	0.806688
6	$(612.605, -0.103476)$	3.9092×10^1	0.428595
7	$(623.499, -0.0150424)$	1.0895×10^1	0.278693
$\vdots$	$\vdots$	$\vdots$	$\vdots$
16	$(626.853, -2.61551 \times 10^{-10})$	1.7451×10^{-5}	0.225253
17	$(626.853, -3.56424 \times 10^{-11})$	3.9308×10^{-6}	0.225253
18	$(626.853, -4.93335 \times 10^{-12})$	8.8543×10^{-7}	0.225253
19	$(626.853, -5.30742 \times 10^{-13})$	1.9945×10^{-7}	0.225253
20	$(626.853, -7.67118 \times 10^{-14})$	4.4925×10^{-8}	0.225252

Table 2 reports the fixed point μ for the complete equation. Note that the numerically calculated contractive factor of this small scale ($m = 2$) problem is rather small, achieving a value of about 0.22 (see table 2, 4^{th} column).

As apparent from table 2, the map converges to

$$\hat{\mu} = 626.853w_1 - (7.67118 \times 10^{-14})w_2,$$

in 20 iterations. By means of the contraction map we can build the approximate solution of the discretized problem,

$$\bar{u}_h = g_h(\hat{\mu} + \tilde{\eta}(\hat{\mu})).$$

The accuracy of the numerical solution is verified by looking at the experimental convergence rate of the residual function $E(x) := -\frac{\partial^2}{\partial x^2}\tilde{u}(x) - F(\tilde{u}(x))$. The computation of an analytical solution, that would be needed to evaluate the error function, actually showed a slower convergence and a worse accuracy if compared with the numerical solution, because of the difficulties in the evaluation of the improper integral appearing in the exact formula. To evaluate the residual $E(x)$ we construct a candidate solution for the analytical problem (1) by means of a cubic spline interpolation of the point values. Thus we verify that the accuracy of the finite difference method does not degenerate as the

Table 3
Behavior of the residual of the approximate solution when the number of subdivisions n increases ($m = 2, l = n$).

	L^1–norm of the residual			L^2–norm of the residual	
n	$\|E_n(x)\|$	$\frac{\|E_n(x)\|}{\|E_{2n}(x)\|}$	n	$\|E_n(x)\|$	$\frac{\|E_n(x)\|}{\|E_{2n}(x)\|}$
10	9.07257×10^{-1}	$\star$	10	1.53448	$\star$
20	2.81267×10^{-1}	3.22561	20	7.778759×10^{-1}	1.97042
40	7.59899×10^{-2}	3.701369	40	2.47547×10^{-1}	3.14590
80	1.55879×10^{-2}	4.874926	80	5.12011×10^{-2}	4.83480
160	3.30348×10^{-3}	4.718633	160	9.88305×10^{-3}	5.18070
320	7.46062×10^{-4}	4.427889	320	1.94486×10^{-3}	5.08160
640	1.76359×10^{-4}	4.230356	640	4.02237×10^{-4}	4.83512

discretization scale n =10, 20, ..., 320, 640 varies. Theoretically, the norm of the residual function should decrease proportionally to the square of the number of subdivisions. As apparent from Table 3, doubling the subdivisions, the residual asymptotically decreases by a factor of approximately 0.25, as expected.

2.1.1 Using a partial eigenspectrum

As claimed in the previous section, in general the complete eigensolution of the elliptic operator $-L_h$ is computationally too demanding and is seldom calculated. Nevertheless, theoretical and numerical considerations suggest that a not so large number of eigenvectors could suffice to evaluate a good approximate solution of the PDE. We consider the solution $\bar{u}_h$ so far determined employing $l = n = 640$ eigenvectors as the "exact" solution, and we try to approximate it by progressively reducing the number $l \ll n$ of eigenvectors involved to generate the solution.

First we test the contractiveness of the generator of the tail $\hat{\eta}$. Figure 1 reports the behavior of the Picard-Peano iteration in the cases $l = 4, 10, 40, 160$ by plotting $\varepsilon_\eta^{(j)} := \left\|\hat{\eta}^{(j+1)} - \hat{\eta}^{(j)}\right\|$ against the iteration index j. Linear convergence is clearly visible and the fact that the lines are parallel indicates that the contractive factor is similar for all runs. A simple calculation allows the estimation of the actual contractive factor $\widetilde{M}$ by evaluating the slopes of the four lines in their asymptotic regime. In the case $l = 4$, we obtain a value of $\widetilde{M} \approx 0.10$, smaller than the theoretical bound given by $M \leq M^{(3)}$. Note that the convergence curves almost coincide for $l = 10 \div 160$. Differences are visible only for $l = 4$, suggesting that in this case the number of the employed eigenpairs is too small to obtain sufficient accuracy in the tail. This fact can

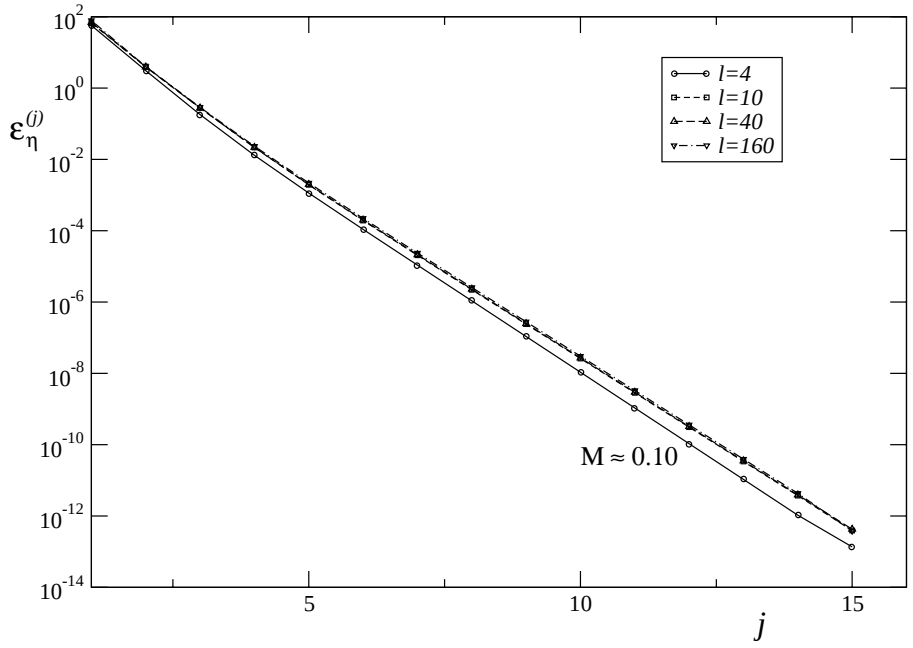

Fig. 1. Convergence of Peano-Picard iteration applied to the tail starting with $m = 2$, $\mu^{(0)} = 500w_1 + 500w_2$, and $n = 640, l = 160, 40, 10, 4$.

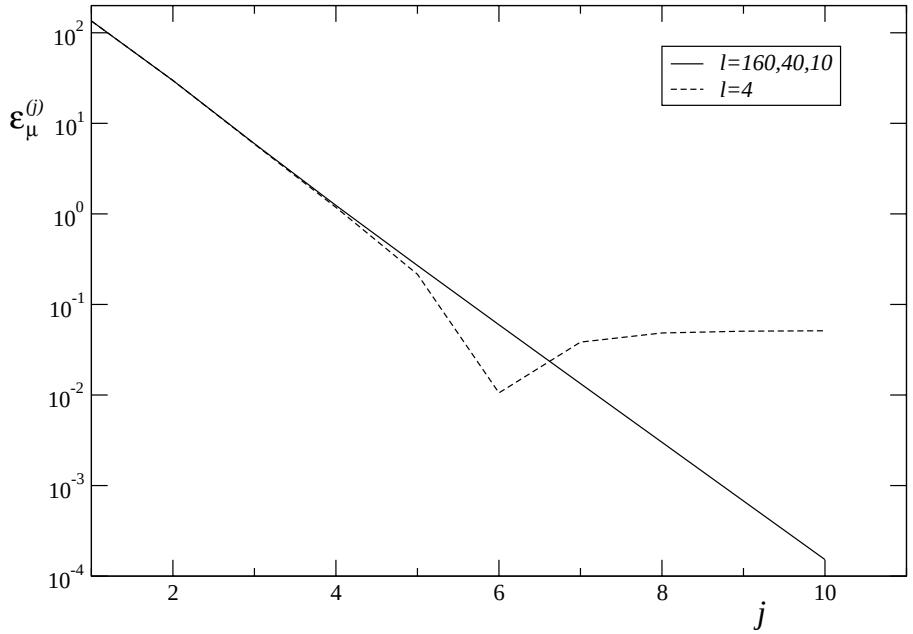

Fig. 2. Convergence of the Peano-Picard iteration applied to the head with $m = 2$, $\mu^{(0)} = 500w_1 + 500w_2$, and $n = 640, l = 160, 40, 10, 4$.

be better appreciated by looking at the convergence behavior of the Picard-Peano procedure applied to the solution of the fixed point iteration defined for the "head" space. To this aim we look at successive approximate solutions $\bar{u}^{(j)} = g_h(\hat{\mu}^{(j)} + \tilde{\eta}(\hat{\mu}^{(j)}))$ of the PDE obtained from the solution of the complete map calculated with an incresing number l (< 640) of eigenpairs. The rate of

convergence to the exact solution is verified by evaluating the errors against a "pseudo analytical" solution $\tilde{u} = g_h(\tilde{\mu} + \tilde{\eta}(\tilde{\mu}))$, which is obtained by means of the same calculations but based on the entire eigenspace (i.e. $l = n = 640$). Figure 2 shows the behavior of the error norm $\epsilon_\mu^{(j)} = \|\mu^{(j)} - \tilde{\mu}\|$ as a function of the iteration counter j. The results show that a clear loss of accuracy occurs for $l = 4$, while $l \geq 10$ is sufficient to obtain accurate solutions. This observations reflect the elliptic property of the differential system at hand.

2.2 Newton-Raphson procedure

The exact finite reduction scheme can be better exploited by means of faster converging linearization strategies, such as the Newton-Raphson approach. To this aim, the solutions μ of the reduced equation (20) can be considered as the fixed points of the map

$$\mu \mapsto PP(\mu) := P_m F_h(g_h(\mu + \tilde{\eta}(\mu))).$$

The Peano-Picard procedure consists in the iterated application of $PP(\cdot)$ from a tentative starting point μ_0:

$$\mu_1 = PP(\mu_0), \ldots, \mu_i = PP^i(\mu_0), \ldots$$

If a limit is reached then a solution of the original problem is found. Alternatively, the solutions of (20) can also be considered as the zeros of the map:

$$\begin{aligned} NR : \mathbb{R}^m &\to \mathbb{R}^m, \\ \mu &\mapsto NR(\mu) := \mu - P_m F_h(g_h(\hat{\mu} + \tilde{\eta}(\hat{\mu}))), \end{aligned}$$

whose solution can be sought by means of the Newton-Raphson procedure. Namely, we search for a limit in the sequence defined as follows:

$$\begin{cases} J_{NR}(\mu)s &= -NR(\mu), \\ \mu &\mapsto \mu + s. \end{cases}$$

The exact finitely reduced equation (20) allows the determination of the Jacobian of the non linear system:

$$J_{NR}(\mu) = \left(\frac{\partial NR_i}{\partial \mu_j}(\mu) \right)_{i,j=1,\ldots,m},$$

The Jacobian J_{NR} can be calculated with respect to the eigenvector coordinates, i.e., by expressing the solution vector $u = c_1 w_1 + \cdots + c_r w_r + \cdots + c_l w_l$:

$$\frac{\partial NR_i}{\partial \mu_j}(\mu) = \delta_{ij} - \sum_{r=1}^{l} \frac{\partial (F_h)_i}{\partial c_r} \cdot \left(\frac{1}{\lambda_r} \frac{\partial}{\partial \mu_j}(\mu + \tilde{\eta}(\mu)) \right) =$$
$$= \delta_{ij} - \frac{\partial (F_h)_i}{\partial c_j} \frac{1}{\lambda_j} - \sum_{r=m+1}^{l} \frac{\partial (F_h)_i}{\partial c_r} \cdot \frac{1}{\lambda_r} \frac{\partial \tilde{\eta}_r(\mu)}{\partial \mu_j}. \quad (28)$$

The elements of the gradient of F_h are calculated as:

$$\frac{\partial (F_h)_i}{\partial c_r}(u) = \frac{\partial}{\partial c_r} \langle F_h(c_1 w_1 + \cdots + c_l w_l), w_i \rangle =$$
$$= \frac{\partial}{\partial c_r} \sum_{k=1}^{n} f(c_1 w_{1,k} + \cdots + c_l w_{l,k}) w_{i,k} = \sum_{k=1}^{n} f'(u_k) w_{r,k} w_{i,k}. \quad (29)$$

The r-th component of the derivative of $\tilde{\eta}(\mu)$ is calculated from equation (18) as:

$$\frac{\partial \tilde{\eta}_r}{\partial \mu_j}(\mu) = \frac{\partial (F_h)_r}{\partial c_j} \frac{1}{\lambda_j} + \sum_{k=m+1}^{l} \frac{\partial (F_h)_r}{\partial c_k} \frac{1}{\lambda_k} \frac{\partial \tilde{\eta}_k}{\partial \mu_j}(\mu),$$

which defines a linear system whose r-th equation is given by:

$$\sum_{k=m+1}^{l} \left(\delta_{rk} - \frac{\partial (F_h)_r}{\partial c_k} \frac{1}{\lambda_k} \right) \frac{\partial \tilde{\eta}_k}{\partial \mu_j} = \frac{1}{\lambda_j} \frac{\partial (F_h)_r}{\partial c_j}, \qquad r = m+1, \ldots, l.$$

From (29) it is easy to see that the system matrix is symmetric. Furthermore, since the derivatives $\partial (F_h)_r / \partial c_k$ are uniformly bounded by the Lipschitz constant C of F, and we have chosen m such that $C/\lambda_k < 1$ for every $k > m$, the system matrix is also positive definite. Note that this also shows the differentiability of the fixed point map $\tilde{\eta}(\mu)$.

Implementation of the previous algorithm to the sample test described in Section 2.1 using the cut-off $m = 2$ produces two different solutions when starting from different initial guesses. This is evidenced by the convergence behavior of the Newton-Raphson scheme shown in Table 2.2. The second solution, not found by the Peano-Picard map, corresponds to a fixed point where the map (20) is not contractive. It can be shown that these two fixed points are the only solutions of the original problem. In other words, by means of Newton-Raphson we have verified Remark 4. Note that convergence of Newton-Raphson is optimal as can be seen in Table 2.2 where a quadratic error reduction is clearly visible.

Table 4
Solutions found starting from $\mu_0 = (100, 0)$ and $\mu_0 = (600, 0)$, with $n = 640$ subdivisions, employing $k = 32$ eigenvectors

$\boldsymbol{\mu_0 = (100, 0)}$			
j	μ_j	$\varepsilon_\mu^{(j)}$	$\frac{\varepsilon_\mu^{(j-1)}}{[\varepsilon_\mu^{(j)}]^2}$
1	$(205.949, 1.66223 \times 10^{-15})$	105.95	105.95
2	$(160.633, -1.08118 \times 10^{-15})$	45.316	0.00403693
3	$(159.585, -5.48825 \times 10^{-16})$	1.04871	0.000510684
4	$(159.582, 6.32719 \times 10^{-16})$	0.00283031	0.0025735
5	$(159.582, 2.17691 \times 10^{-15})$	2.10223×10^{-8}	0.00262429
6	$(159.582, 3.53473 \times 10^{-15})$	1.70535×10^{-13}	385.881
L^2-residual $= 0.00103257$			

$\boldsymbol{\mu_0 = (600, 0)}$			
j	μ_j	$\varepsilon_\mu^{(j)}$	$\frac{\varepsilon_\mu^{(j-1)}}{[\varepsilon_\mu^{(j)}]^2}$
1	$(627.622, 1.59563 \times 10^{-15})$	27.6225	27.6225
2	$(626.853, -4.35704 \times 10^{-17})$	0.769375	0.00100835
3	$(626.852, 1.88864 \times 10^{-16})$	0.000522536	0.000882756
4	$(626.852, 1.73901 \times 10^{-15})$	2.41811×10^{-10}	0.000885613
5	$(626.852, 2.70311 \times 10^{-15})$	9.64099×10^{-16}	16487.9
L^2-residual $= 0.0205632$			

3 Conclusions

A numerical algorithm that translates the global finite parameter reduction techinque arising from the Amann–Conley–Zenhder idea has been presented for the solution of semilinear Dirichlet problems. The infinite dimensional fixed point problem is translated by means of a finite difference approximation of the linear differential operator into a finite size fixed point problem. The resulting discrete algorithm inherits all the properties of the continuous counterpart. Thus the fixed point problem defined on the tail space is a contractive map controlled by the eigenvalues of the linear discrete operator, while the fixed point defined for the head space, that in general can be made much smaller than the other one, concentrates most of the nonlinear difficulties of the original PDE. The advantage of the proposed reduction approach is that it allows the use of sophisticated nonlinear solvers, which, if employed on the original

system defined in $\mathbb{R}^n$, would result much more computationally expensive.

As a perspective, the reduction technique could be promisingly applied to more interesting situations, e.g., greater dimensions or time evolution problems.

References

[1] A. Amann and E. Zehnder. Periodic solutions of asymptotically linear Hamiltonian systems. *Manuscripta Math.*, 32(1-2):149–189, 1980.

[2] H. Amann and E. Zehnder. Nontrivial solutions for a class of nonresonance problems and applications to nonlinear differential equations. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 7(4):539–603, 1980.

[3] P. N. Brown and Y. Saad. Convergence theory of nonlinear Newton-Krylov algorithms. *SIAM J. Sci. Opt.*, 4(2):297–330, 1994.

[4] F. Cardin. Global finite generating functions for field theory. In *Classical and quantum integrability (Warsaw, 2001)*, volume 59 of *Banach Center Publ.*, pages 133–142. Polish Acad. Sci., Warsaw, 2003.

[5] S. S. Clift and P. A. Forsyth. Linear and non-linear iterative methods for the incompressible Navier-Stokes equations. *Int. J. Numer. Methods Fluids*, 18:229–256, 1994.

[6] C. Conley and E. Zehnder. A global fixed point theorem for symplectic maps and subarmonic solutions of Hamiltonian equations in tori. In *Nonlinear functional analysis and its applications, Part 1*, volume 45 of *Ptoc. Sympos. Pure Math.*, pages 283–299, Providence, RI, 1986. Amer. Math. Cos.

[7] J. E. Dennis and J. J. Moré. Quasi-Newton methods, motivation and theory. *SIAM Review*, 19(1):46–89, 1977.

[8] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization.* Prentice Hall, Englewood Cliffs, NJ, 1983.

[9] S. C. Eisenstat and H. F. Walker. Globally convergent Inexact Newton methods. *SIAM J. Sci. Opt.*, 4(2):393–422, 1994.

[10] G. Gambolati, G. Pini, and M. Putti. Nested iterations for symmetric eigenproblems. *SIAM J. Sci. Comput.*, 16(1):173–192, 1995.

[11] C. T. Kelley. *Iterative Methods for Linear and Nonlinear Equations.* SIAM, Philadelphia, 1995.

[12] C. T. Miller, G. A. Williams, C. T. Kelley, and M. D. Tocci. Robust solution of Richards' equation for nonuniform porous media. *Water Resour. Res.*, 34(9):2599–2610, 1998.