On the Equivalence of Left and Right Nonlinear Preconditioning for a Class of Algebraic Systems

Konstantin Brenner[0000−0002−9372−0431],
Alessandra Marelli[0009−0001−3544−0335],
Quang Huy Tran[0000−0001−7771−3154]

1 Introduction

Nonlinear preconditioning is a numerical technique that often helps to significantly improve robustness and accelerate the convergence of linearization schemes such as Newton's method. In the context of domain decomposition, the approach was first introduced by Cai and Keyes [4] within the framework of the nonlinear Additive Schwarz method, under the name ASPIN. Since then, numerous variants of additive or multiplicative nonlinear preconditioning methods have been proposed, including Schwarz-inspired methods such as MSPIN [13] or RASPEN [6], as well as the nonlinear extensions of non-overlapping domain decomposition techniques such as FETI-DP [9] or BDDC [10].

In many situations the efficiency of nonlinear preconditioning stems from its ability to deal with *unbalanced nonlinearities*[1]. These typically arise when a small subset of the system's unknowns is responsible for the poor convergence of the linearization scheme. This is notably the case for nonlinear systems resulting from the discretization of PDEs with stiff nonlinear closure laws, equation degeneracies, or material interfaces.

One possible treatment of unbalanced nonlinearities was advocated in [11] (see also [14]), where the authors proposed to reduce the system by performing a nonlinear

Konstantin Brenner
Université Côte d'Azur, Inria, CNRS, LJAD, France, e-mail: konstantin.brenner@univ-cotedazur.fr

Alessandra Marelli
IFP Energies nouvelles & Université Côte d'Azur, Inria, CNRS, LJAD, France, e-mail: alessandra.marelli@univ-cotedazur.fr

Quang Huy Tran
IFP Energies nouvelles, France, e-mail: quang-huy.tran@ifpen.fr

[1] This expression is used by several authors [4, 5, 12, 13] to intuitively describe a situation in which the different terms of an algebraic systems have very contrasting orders of magnitude.

elimination of the problematic unknowns. This leads to a right preconditionning method called NlEm for *Nonlinear Elimination method*.

Let us illustrate the idea of the method. Let $F : D \subset \mathbb{R}^N \to \mathbb{R}^N$ be a function for which we seek a root $\mathbf{u}$ satisfying $F(\mathbf{u}) = 0$. Let the components of $\mathbf{u}$ be partitioned into the subsets of *bad* and *good* unknowns according to some criterion. Up to a reordering, the vector of unknowns can be expressed as $\mathbf{u} = (\mathbf{u}_g, \mathbf{u}_b)$. Furthermore, we consider the splitting of the original system into

$$F_g(\mathbf{u}_g, \mathbf{u}_b) = 0, \tag{1a}$$

$$F_b(\mathbf{u}_g, \mathbf{u}_b) = 0, \tag{1b}$$

where, for any given $\mathbf{u}_g$, subsystem (1b) is assumed to admit a unique solution in $\mathbf{u}_b$. Although we do not assume any explicit relation between the splitting of the variables and that of the equations, our motivation comes from PDE-based applications, where a natural correspondence exists between the node/cell unknowns and the discretized equations via the mesh.

We then introduce the nonlinear map G_b, defined for all $\mathbf{u}_g$, by

$$F_b(u_g,\ G_b(\mathbf{u}_g)) = 0. \tag{2}$$

With this notation, subsystem (1b) is equivalent to

$$\mathbf{u}_b = G_b(\mathbf{u}_g). \tag{3}$$

Substituting (3) into (1a) yields a reduced system depending exclusively on the good unknowns, namely,

$$F_g(\mathbf{u}_g,\ G_b(\mathbf{u}_g)) = 0. \tag{4}$$

Let

$$\mathcal{F}(\mathbf{u}_g) \coloneqq F_g(\mathbf{u}_g,\ G_b(\mathbf{u}_g)). \tag{5}$$

Then, Newton's method applied to (4) takes the form

$$\mathbf{u}_g^{n+1} = \mathbf{u}_g^n - \mathcal{F}'(\mathbf{u}_g^n)^{-1}\,\mathcal{F}(\mathbf{u}_g^n). \tag{6}$$

This method based on elimination of a portion of the unknowns has been further developed and extended by Hwang et al. [7, 8]. In particular, the authors suggested an elegant implementation of (6) as the following two-step method:

- Update the linearization point by solving the sub-problem

$$\widetilde{\mathbf{u}}_b^n = G_b(\mathbf{u}_g^n), \quad \widetilde{\mathbf{u}}^n = (\mathbf{u}_g^n, \widetilde{\mathbf{u}}_b^n). \tag{7}$$

- Perform a single global Newton step on the original system

$$\mathbf{u}^{n+1} = \widetilde{\mathbf{u}}^n - F'(\widetilde{\mathbf{u}}^n)^{-1}F(\widetilde{\mathbf{u}}^n). \tag{8}$$

It is straightforward to verify that, given the same initial point, the iterates produced by (6) and (7)–(8) are identical. Nevertheless, the latter turns out to be more practical, insofar as it avoids the computation of a Schur complement matrix.

Furthermore, the formulation (7)–(8) of the NlEm makes it possible to introduce substantial modifications to the algorithm. As suggested in [8], it is judicious to adaptively update the bad/good splitting over the iterations so as to enhance the efficiency of the algorithm, especially for PDEs involving moving fronts and interfaces. In the following, we will use the term NlEm in a broader sense, allowing the composition of the bad and good subsets to vary between iterations.

An alternative left nonlinear preconditioner based on nonlinear elimination was proposed by Liu et al. [12] under the name *Nonlinear Elimination Preconditioned Inexact Newton*'s method (NEPIN). Instead of modifying the linearization point, NEPIN employs nonlinear elimination to adjust the Newton direction. The update can subsequently be damped using a backtracking line search. In this article, we consider a more academic variant of the method called NEPEN, in which no damping of the Newton step is applied and the linear systems are assumed to be solved exactly.

In order to introduce NEPEN, let us again consider system (1a)–(1b). Instead of substituting (3) into (1a) we keep the former equation, which leads to the system

$$F_g(\mathbf{u}_g, \mathbf{u}_b) = 0 \tag{9a}$$

$$\mathbf{u}_b - G_b(\mathbf{u}_g) = 0. \tag{9b}$$

Applying Newton's method to (9) yields the NEPEN method. Here again, formulation (9) offers us the opportunity to dynamically change the definition of bad and good subsets of unknowns from one iteration to another.

NlEm belongs to the class of right preconditioners, as it modifies the unknowns through the nonlinear substitution $(\mathbf{u}_g, \mathbf{u}_b) = (\mathbf{u}_g, G_b(\mathbf{u}_g))$ while leaving the residual function unchanged (note that (1b) is satisfied trivially). NEPEN, on the other hand, provides a form of left preconditioning, as it modifies the residual function while leaving the system's unknowns unchanged. We refer the readers to [12] for further details.

While in general NlEm and NEPEN produce different sequences of approximate solutions, it can be shown that under certain conditions the two methods coincide. The objective of the present work is to identify those conditions and to provide examples of physically relevant PDEs that give rise to systems for which NlEm and NEPEN are equivalent. Specifically, we show that the updates produced by NlEm and NEPEN coincide if the good subsystems can be expressed in the form

$$F_g(\mathbf{u}_g, \mathbf{u}_b) = f_g(\mathbf{u}_g) + A_{gb}(\mathbf{u}_g)\mathbf{u}_b, \tag{10}$$

with some vector-valued and matrix-valued functions f_g and A_{gb}.

We note that a similar equivalence result regarding left block-Jacobi preconditioning and the corresponding two-step method, introduced in [5] under the name NKS-RAS, has been pointed out in [2] for the case of diagonally nonlinear systems.

The remainder of the article is organized as follows. In Section 2, we prove the equivalence between the two methods in the case of block-diagonal nonlinear systems; in Section 3, we present two examples of such systems, based on the porous medium equation and the heterogeneous Poisson equation with a nonlinear transmission, for which we provide numerical illustrations of the performance of the NEPEN–NlEm method.

2 Comparison of NlEm and NEPEN

Let $\mathbf{u} = (\mathbf{u}_g, \mathbf{u}_b)$ denote the current approximate solution, and let $\widetilde{\mathbf{u}}_b = G_b(\mathbf{u}_g)$. Furthermore, for $\alpha, \beta \in \{b, g\}$, let $J_{\alpha\beta} = \partial_{\mathbf{u}_\beta} F_\alpha(\mathbf{u}_g, \mathbf{u}_b)$ and $\widetilde{J}_{\alpha\beta} = \partial_{\mathbf{u}_\beta} F_\alpha(\mathbf{u}_g, \widetilde{\mathbf{u}}_b)$ denote the partial Jacobian matrices. By formally applying the Implicit Function Theorem [15, §5.2.4], we obtain $G'_b(\mathbf{u}_g) = -\widetilde{J}_{bb}^{-1}\widetilde{J}_{bg}$.

Let us first consider the NEPEN method. By linearizing (9) around $(\mathbf{u}_g, \mathbf{u}_b)$, we obtain

$$\begin{bmatrix} J_{gg} & J_{gb} \\ \widetilde{J}_{bb}^{-1}\widetilde{J}_{bg} & I \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_g \\ \boldsymbol{\delta}_b \end{bmatrix} = - \begin{bmatrix} F_g(\mathbf{u}_g, \mathbf{u}_b) \\ \mathbf{u}_b - \widetilde{\mathbf{u}}_b \end{bmatrix}, \tag{11}$$

where $(\boldsymbol{\delta}_g, \boldsymbol{\delta}_b)$ is the update vector. Eliminating $\boldsymbol{\delta}_b$ from the bad sub-system, we get

$$\boldsymbol{\delta}_b = \widetilde{\mathbf{u}}_b - \mathbf{u}_b - \widetilde{J}_{bb}^{-1}\widetilde{J}_{bg}\boldsymbol{\delta}_g, \tag{12}$$

which yields

$$\big(J_{gg} - J_{gb}\widetilde{J}_{bb}^{-1}\widetilde{J}_{bg}\big)\,\boldsymbol{\delta}_g = -\big(F_g(\mathbf{u}_g, \mathbf{u}_b) + J_{gb}(\widetilde{\mathbf{u}}_b - \mathbf{u}_b)\big). \tag{13}$$

Now, let us consider the two-step formulation (7)–(8) of NlEm. System (8) can be expressed as

$$\begin{bmatrix} \widetilde{J}_{gg} & \widetilde{J}_{gb} \\ \widetilde{J}_{bg} & \widetilde{J}_{bb} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_g \\ \boldsymbol{\delta}_b + \mathbf{u}_b - \widetilde{\mathbf{u}}_b \end{bmatrix} = - \begin{bmatrix} F_g(\mathbf{u}_g, \widetilde{\mathbf{u}}_b) \\ 0 \end{bmatrix}. \tag{14}$$

From the bad linear subsystem we deduce that $\boldsymbol{\delta}_b + \mathbf{u}_b - \widetilde{\mathbf{u}}_b = -\widetilde{J}_{bb}^{-1}\widetilde{J}_{bg}\boldsymbol{\delta}_g$, which coincides with (12). Substituting the latter into the good subsystem leads to

$$\big(\widetilde{J}_{gg} - \widetilde{J}_{gb}\widetilde{J}_{bb}^{-1}\widetilde{J}_{bg}\big)\,\boldsymbol{\delta}_g = -F_g(\mathbf{u}_g, \widetilde{\mathbf{u}}_b). \tag{15}$$

By comparing (13) and (15), we establish that the iterations produced by the two methods are identical as soon as

$$J_{gg} = \widetilde{J}_{gg}, \qquad J_{gb} = \widetilde{J}_{gb}, \tag{16}$$

and

$$F_g(\mathbf{u}_g, \widetilde{\mathbf{u}}_b) = F_g(\mathbf{u}_g, \mathbf{u}_b) + J_{gb}(\widetilde{\mathbf{u}}_b - \mathbf{u}_b). \tag{17}$$

As a side note, we remark that (16) is not required for a *simplified* version of the left preconditioner that is suggested for practical use in [12]. On the other hand, any function F_g satisfying (10) also satisfies (16) and (17). The converse can be established if (16) holds for all $\mathbf{u}_g$ and $\mathbf{u}_b$ and not merely for a given iterate.

Theorem 1 *Assume that the domain of F is convex and that the Gâteaux-derivative of F_g is hemicontinuous. Then, both* (16) *and* (17) *are satisfied for all $\mathbf{u}_g$ and $\mathbf{u}_b$ if and only if* (10) *holds.*

Proof. As discussed above, (10) implies (16)–(17). Assume that (16) holds for all $\mathbf{u}_g$ and $\mathbf{u}_b$. Then, J_{gg} and J_{gb} only depend on $\mathbf{u}_g$. Applying component-wise the Mean Value Theorem [15, §3.2.2], one shows that the second statement of (16) implies (17). Furthermore, it follows from the Gradient Theorem [15, §3.2.7] that, for any given $\widehat{\mathbf{u}}_b$, $F_g(\mathbf{u}_g, \mathbf{u}_b) = F_g(\mathbf{u}_g, \widehat{\mathbf{u}}_b) + J_{gb}(\mathbf{u}_b - \widehat{\mathbf{u}}_b)$ which implies (10). □

3 Numerical experiment

In this section, we analyze two examples of nonlinear systems for which NEPEN and NlEm are identical. The first example, based on the discretization of the porous medium equation, features a diagonal nonlinearity, for which NEPEN and NlEm coincide regardless of the good/bad partitioning. The second example, motivated by flows in heterogeneous porous media, involves a system that is linear except for two equations. In this case, equivalence between NEPEN and NlEm is achieved for those partitions of the unknowns that group the nonlinear equations to the bad subsystem.

Porous medium equation. The test case considered here is similar to the one presented in [3] and [2] to which we refer for more detailed discussion. In brief, we are interested in the algebraic system resulting from the implicit in time discretization of the porous medium equation [16]. Specifically, we consider the nonlinear system resulting from the discretization of the boundary value problem consisting of the differential equation

$$\beta(u) - \beta(u_{ini}) = \partial^2_{xx} u, \qquad x \in (0, 1), \tag{18a}$$

and the boundary conditions

$$\partial_x u(0) = -q, \qquad \partial_x u(1) = 0, \tag{18b}$$

where $\beta(u) = u^{1/10}$, $q = 0.5$ and $\beta(u_{ini}) = 10^{-6}$. Problem (18) is discretized using a mass-lumped $\mathbf{P}_1$ Finite Element method with $N = 100$, yielding the algebraic system

$$\beta(\mathbf{u}) - \beta(\mathbf{u}_{ini}) + A\mathbf{u} = \mathbf{b}, \tag{19}$$

where $\beta(\mathbf{u})$ is understood component-wise, $\mathbf{u}_{ini}$ is a vector with all components equal to u_{ini}, and the matrix A is the product of the inverse mass matrix and the stiffness matrix. Clearly, (19) satisfies (10) regardless the choice of bad/good splitting.

In the left column of Figure 1, we report the approximate solution of (18). The solution profile exhibits a relatively sharp front separating the *dry region* on the right, characterized by small values of $\beta(u)$, from the *wet region* on the left.

In view of the Monotone Newton Theorem (see e.g. [15]), it can be shown that Newton's method converges monotonically from the initial guess. Nevertheless, the observed convergence is very slow due to the singularity of β' at $u = 0$. In fact, when applied without preconditioning, Newton's method requires more than 700 iterations to achieve an accuracy below 10^{-12}. This convergence difficulty can be efficiently addressed by eliminating a subset of degrees of freedom associated with the transition from the dry to the wet state.

To identify this subset, we proceed as follows. First, we partition the domain into two disjoint subsets of wet and dry degrees of freedom based on the value of u, using a threshold of 10^{-12}. The bad subset is then defined as the set of dry degrees of freedom that have at least one wet neighbor. Experimentally, we observe that the approximate solutions are monotone along the domain; consequently, at every iteration of the method such a bad degree of freedom turns out to be unique. Optionally, the eliminated subset may be extended by some safety width to include a few additional layers of neighboring degrees of freedom.

For both plain and preconditioned Newton's methods, the initial guess is chosen as the vector with all components equal to u_{ini}, except for the leftmost degree of freedom, which is set to 0.1. This choice is motivated by our specific method of selecting bad degrees of freedom.

The right column of Figure 1 reports convergence history of the preconditioned Newton's method for 4 safety extension widths. The error is measured in l^∞ and is computed based on a reference solution. Elimination of a single degree of freedom downstream to the front leads to significant speedup of factor about 9. A further reduction in the number of outer iterations is achieved when an extended elimination domain is considered.

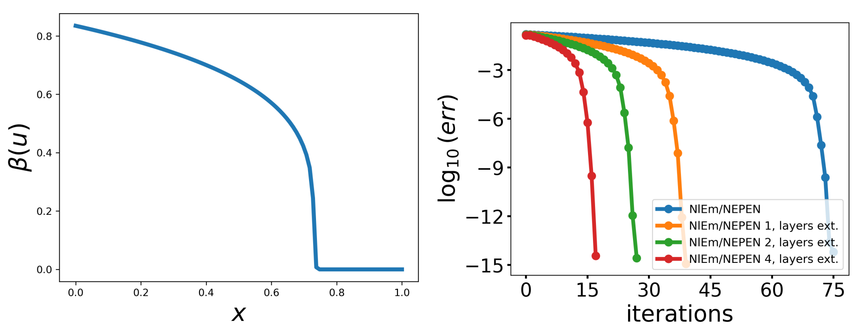

Fig. 1 Approximate solution of (18) (left) and convergence history for preconditioned Newton's method (right).

Poisson equation with nonlinear transmission condition. Let $\Omega_1 = (-1, 0)$ and $\Omega_2 = (0, 1)$. We look for a pair of functions $u_1 \in H^1(\Omega_1)$ and $u_2 \in H^1(\Omega_2)$ satisfying, in a weak sense, the differential equation

$$-k_i \partial^2_{xx} u_i = 1, \qquad x \in \Omega_i, \tag{20a}$$

equipped with the transmission and boundary conditions

$$\partial_x u_1(0) = \partial_x u_2(0), \qquad u_2(0) = \phi(u_1(0)), \tag{20b}$$

$$u_1(-1) = 0, \qquad u_2(1) = 0, \tag{20c}$$

with $k_1 = 0.5$, $k_2 = 1$ and $\phi(u) = u^{10}$. This problem is inspired by models of unsaturated water flow in heterogeneous porous media [1]. The left panel of Figure 2 shows the approximate solution of (20), which displays a discontinuity at $x = 0$.

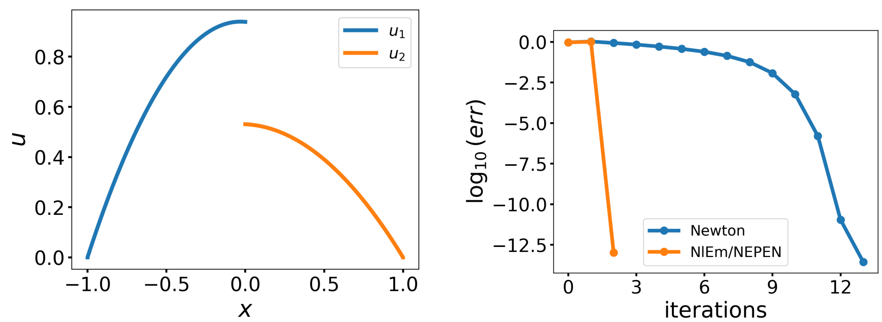

Fig. 2 Approximate solution of (20) (left) and convergence history for plane and preconditioned Newton's methods (right).

We discretize the problem using $\mathbf{P}_1$ Finite Element method based on a uniform partitioning of the interval $(-1, 1)$ resolving the interface point $x = 0$. Specifically, we use 100 elements, leading to $N = 99$ after elimination the degrees of freedom associated to Dirichlet boundary conditions. Denoting by $\mathbf{u}_1$ and $\mathbf{u}_2$ the nodal values associated to the interior of Ω_1 and Ω_2 and by u_Γ the approximate value of $u_1(0)$, the Finite Element discretization of (20) yields the following system of algebraic equations

$$\begin{bmatrix} A_{11} & A_{1\Gamma} & 0 \\ A_{\Gamma 1} & A^1_{\Gamma\Gamma} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ u_\Gamma \\ \mathbf{u}_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & A^2_{\Gamma\Gamma} & A_{\Gamma 2} \\ 0 & A_{2\Gamma} & A_{22} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \phi(u_\Gamma) \\ \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ b_\Gamma \\ \mathbf{b}_2 \end{bmatrix} \tag{21}$$

where the nonzero blocks within the two matrices in the left-hand side are obtained by discretizing the sub-domain problems with homogeneous Neumann boundary condition at $x = 0$ and zero Dirichlet conditions at $x = -1$ and $x = 1$ respectively.

We consider the nonlinear preconditioning based on elimination of unknowns consisting of the interface unknown u_Γ and its direct neighbor from Ω_1 and Ω_2. With such a bad subdomain, it is easy to verify that (21) can be written in the form (10) with affine f_g and constant A_{gb}. Therefore, NlEm and NEPEN coincide.

In the right panel of Figure 2, we report convergence history of Newton's method with and without nonlinear preconditioning. The error is evaluated in l^∞ norm with respect to a converged reference solution. The domain has been discretized with 100 elements and we have used zero initial guess. The NlEm/NEPEN method converges within only 2 iterations, while the standard Newton method requires 13 iterations to achieve the same precision. Similar improvement due to nonlinear preconditioning is observed in 2-D configurations (that we do not report here for brevity), with a slight increase in the NlEm/NEPEN iteration count, typically reaching 4 iterations.

4 Conclusion

In this paper, we highlighted the equivalence between two nonlinear elimination methods—NlEm and NEPEN—for a class of systems. Both can be viewed as preconditioners, one acting on the right and the other on the left. To our knowledge, this result seems to be new. The power of the method is illustrated on two examples arising from steady-state problems. The next step is to apply it to evolutionary problems, where the time-step has a critical role on the stiffness of the system to be solved.

References

1. Berninger, H., Kornhuber, R., Sander, O.: A multidomain discretization of the Richards equation in layered soil. Comput. Geosci. **19**(1), 213–232 (2015)
2. Brenner, K.: On global and monotone convergence of the preconditioned Newton's method for some mildly nonlinear systems. In: International Conference on Domain Decomposition Methods, pp. 85–92. Springer, Switzerland (2022)
3. Brenner, K.: On the monotone convergence of Jacobi–Newton method for mildly nonlinear systems. J. Comput. Appl. Math. **419**, 114719 (2023)
4. Cai, X.C., Keyes, D.E.: Nonlinearly preconditioned inexact Newton algorithms. SIAM J. Sci. Comp. **24**(1), 183–200 (2002)
5. Cai, X.C., Li, X.: Inexact Newton methods with restricted additive Schwarz based nonlinear elimination for problems with high local nonlinearity. SIAM J. Sci. Comp. **33**(2), 746–762 (2011)
6. Dolean, V., Gander, M.J., Kheriji, W., Kwok, F., Masson, R.: Nonlinear preconditioning: How to use a nonlinear Schwarz method to precondition Newton's method. SIAM J. Sci. Comp. **38**(6), A3357–A3380 (2016)
7. Hwang, F.N., Lin, H.L., Cai, X.C.: Two-level nonlinear elimination based preconditioners for inexact Newton methods with application in shocked duct flow calculation. Electron. Trans. Numer. Anal. **37**, 239–251 (2010)
8. Hwang, F.N., Su, Y.C., Cai, X.C.: A parallel adaptive nonlinear elimination preconditioned inexact Newton method for transonic full potential equation. Comput. Fluids **110**, 96–107 (2015)
9. Klawonn, A., Lanser, M., Rheinbach, O.: Nonlinear FETI-DP and BDDC methods. SIAM J. Sci. Comp. **36**(2), A737–A765 (2014)
10. Klawonn, A., Lanser, M., Rheinbach, O.: Nonlinear BDDC methods with approximate solvers. Electron. Trans. Numer. Anal. **49**, 244–273 (2018)
11. Lanzkron, P.J., Rose, D.J., Wilkes, J.T.: An analysis of approximate nonlinear elimination. SIAM J. Sci. Comp. **17**(2), 538–559 (1996)
12. Liu, L., Hwang, F.N., Luo, L., Cai, X.C., Keyes, D.E.: A nonlinear elimination preconditioned inexact Newton algorithm. SIAM J. Sci. Comp. **44**(3), A1579–A1605 (2022)
13. Liu, L., Keyes, D.E.: Field-split preconditioned inexact Newton algorithms. SIAM J. Sci. Comp. **37**, A1388–A1409 (2015)
14. Milaszewicz, J.P., Abdel Masih, S.: On elimination and fixed point iterations. Comput. Math. Appl. **25**, 43–53 (1993)
15. Ortega, J.M., Rheinboldt, W.C.: Iterative solution of nonlinear equations in several variables. Society for Industrial and Applied Mathematics (2000)
16. Vázquez, J.L.: The Porous Medium Equation: Mathematical Theory. Oxford Mathematical Monographs. Oxford University Press, Oxford (2007)