

Nonlinear Schwarz Preconditioning for Nonlinear Optimization Problems with Bound Constraints

Hardik Kothari, Alena Kopaničáková, and Rolf Krause

1 Introduction

We consider a Lipschitz domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, and a triangulation \mathcal{T} on Ω . Now, we define $\mathcal{V} = \text{span}\{\phi_p\}_{p \in \mathcal{N}}$ as a Finite Element (FE) space, where \mathcal{N} denotes a set of nodes of the mesh \mathcal{T} . Furthermore, we introduce the feasible set $\mathcal{F} = \{v \in \mathcal{V} \mid \underline{\psi} \leq v \leq \bar{\psi}\}$, where $\underline{\psi}, \bar{\psi}$ denote the component-wise lower bound and upper bound, respectively.

We consider the following abstract nonlinear minimization problem:

$$\text{Find } v^* = \arg \min_{v \in \mathcal{F}} f(v), \quad (1)$$

where $f: \mathcal{V} \rightarrow \mathbb{R}$ denotes a bounded, twice-Lipschitz-continuously-differentiable objective function. Problems of this type arise in numerous applications, such as contact mechanics [15], or fracture mechanics [12, 13].

Under certain assumptions on the function f , the minimization problem (1) can be equivalently rewritten as a nonlinear complementarity problem (NCP). As the first-order optimality conditions for (1) are given as: Find $v \in \mathcal{V}$ such that

$$\begin{aligned} \nabla f(v) \leq 0, \quad v - \bar{\psi} \leq 0, \quad \langle \nabla f(v), v - \bar{\psi} \rangle = 0, \quad \forall v \in \mathcal{F} \setminus \underline{\mathcal{F}}, \\ \nabla f(v) \geq 0, \quad \underline{\psi} - v \leq 0, \quad \langle \nabla f(v), \underline{\psi} - v \rangle = 0, \quad \forall v \in \mathcal{F} \setminus \bar{\mathcal{F}}, \end{aligned} \quad (2)$$

where $\underline{\mathcal{F}} = \{v \in \mathcal{F} \mid v = \underline{\psi}\}$ and $\bar{\mathcal{F}} = \{v \in \mathcal{F} \mid v = \bar{\psi}\}$ denote boundaries of the feasible set \mathcal{F} . Standard approaches for solving such minimization problems include penalty/augmented Lagrangian methods, interior-point methods, or active-set methods; see [16] for a detailed overview. In this work, we focus our attention on Newton-based active-set methods, namely the semismooth Newton method, and the

Hardik Kothari, Rolf Krause
Università della Svizzera italiana, Switzerland, e-mail: hardik.kothari@usi.ch, rolf.krause@usi.ch

Alena Kopaničáková
Brown University, USA, e-mail: alena.kopanicakova@brown.edu
Università della Svizzera italiana, Switzerland, e-mail: alena.kopanicakova@usi.ch

sequential quadratic programming (SQP) Newton method. Although the active-set methods are fairly efficient, their convergence tends to deteriorate due to three main factors: inability to detect an active-set sufficiently fast; strong and highly unbalanced nonlinearities; and ill-conditioning of the problem.

In the context of unconstrained nonlinear problems, nonlinear additive Schwarz preconditioners have been demonstrated to accelerate the convergence of the Newton methods; see for example Additive Schwarz Preconditioned Inexact Newton (ASPIN) [6], Restricted Additive Schwarz Preconditioned Exact Newton (RASPEN) [8], or Nonlinear Elimination Preconditioned Inexact Newton (NEPIN) [7] methods. In this work, we aim to extend a class of Schwarz preconditioned Newton methods to solve constrained nonlinear optimization problems. To the best of our knowledge, there have been only a few attempts to employ Schwarz methods to solve variational inequalities, for instance [1, 2, 3, 11]. In this work, we introduce a two-level nonlinear additive Schwarz preconditioner for the Newton-SQP method, which ensures that the subdomain and coarse-level corrections remain in the feasible set.

2 Nonlinear preconditioning

We define a residual function $F: \mathcal{V} \rightarrow \mathcal{V}'$ as the gradient of the original objective function, i.e., $F(\cdot) \equiv \nabla f(\cdot)$. Let G be a nonlinear preconditioner of the residual function F , such that in some sense G denotes an approximate inverse of the nonlinear function F , i.e., $G \approx F^{-1}$. Now, we can define nonlinearly-preconditioned residual function as $\mathcal{F}(v) := F(G(v))$. This preconditioner is used to define a nonlinearly-preconditioned variational inequality problem as follows: Find $v \in \mathcal{V}$ such that

$$\begin{aligned} \mathcal{F}(v) \leq 0, \quad v - \bar{\psi} \leq 0, \quad \langle \mathcal{F}(v), v - \bar{\psi} \rangle = 0, \quad \forall v \in \mathcal{F} \setminus \underline{\mathcal{F}}, \\ \mathcal{F}(v) \geq 0, \quad \underline{\psi} - v \leq 0, \quad \langle \mathcal{F}(v), \underline{\psi} - v \rangle = 0, \quad \forall v \in \mathcal{F} \setminus \bar{\mathcal{F}}, \end{aligned} \quad (3)$$

where the solution of (3) is the same as the solution of (1) and (2). Please note that the operator G is used as a “right-preconditioner”, since this type of preconditioning does not change the original nonlinear system, and it also avoids the transformation of bound constraints into general inequality constraints. Generally, the preconditioner G can be constructed implicitly as a fixed-point iteration, i.e., $v^{(k+1)} = G(v^{(k)})$. In this work, we construct G using a variant of the nonlinear Restricted Additive Schwarz (NRAS) method, termed as NRAS-B method. Contrary to the standard NRAS method, the NRAS-B method ensures that the bound constraints are not violated by taking the preconditioning step. Thus, the preconditioner G produces an iterate that remains in the feasible set, i.e., $v^{(k)} \in \mathcal{F}$, for all $k = 1, 2, \dots$

NRAS-B method: We consider a decomposition of the FE space \mathcal{V} into n overlapping and non-overlapping subspaces, denoted by $\{\mathcal{V}_i\}_{i=1}^n$ and $\{\tilde{\mathcal{V}}_i\}_{i=1}^n$, respectively. The overlap between the subspaces is controlled by the variable δ , defined as a multiple of the mesh-width h of the underlying mesh \mathcal{T} . On these subspaces, we

define the standard restriction operator $R_i: \mathcal{V} \rightarrow \mathcal{V}_i$, and the prolongation operator $P_i: \mathcal{V}_i \rightarrow \mathcal{V}$, where $R_i^\top = P_i$. Similarly, we define a restricted prolongation operator $\tilde{P}_i: \tilde{\mathcal{V}}_i \rightarrow \mathcal{V}$ such that $\sum_{i=1}^n \tilde{P}_i R_i = I$.

Utilizing the aforementioned decomposition and the transfer operators, we now define the constrained nonlinear optimization problem on each subspace as follows. For a given initial guess $v_i^{(0)} \in \mathcal{F}_i$, where $v_i^{(0)} \leftarrow R_i v^{(k)}$:

$$\text{Find } v_i^* = \arg \min_{v_i \in \mathcal{F}_i} f_i(v_i), \quad (4)$$

where $f_i: \mathcal{V}_i \rightarrow \mathbb{R}$ denotes a restriction of the function f to the subspace \mathcal{V}_i . The feasible set associated with the subspace \mathcal{V}_i is given by $\mathcal{F}_i = \{v_i \in \mathcal{V}_i \mid \underline{\psi}_i \leq v_i \leq \bar{\psi}_i\}$. Here, we point out that the local minimization problems (4) are solved on overlapping subspaces. However, the global iterate $v^{(k)}$ is updated using the corrections associated with the non-overlapping subspaces, as in

$$v^{(k+1)} = v^{(k)} + \alpha \sum_{i=1}^n \tilde{P}_i (v_i^* - R_i v^{(k)}), \quad (5)$$

where v_i^* denotes a solution of (4) and α is computed using a line-search strategy. Solving (4) and update rule (5) comprise an iteration of the NRAS-B method.

Two-level NRAS-B method: The convergence of additive Schwarz methods is known to deteriorate with an increasing number of subdomains. In order to achieve algorithmic scalability, it is essential to ensure global information transfer through a coarse space. In the context of constrained minimization problems, constructing a coarse space is not a trivial task, as one has to ensure that the prolonged corrections from the coarse level provide a sufficient decrease in the objective function f , and the updated iterate remains in the feasible set. We construct a coarse-level objective function $f_0: \mathcal{V}_0 \rightarrow \mathbb{R}$, where \mathcal{V}_0 denotes a coarse space $\mathcal{V}_0 \subset \mathcal{V}$, and \mathcal{T}_0 denotes a mesh associated with the FE space \mathcal{V}_0 . The transfer of information between the coarse level and the original problem is ensured by the prolongation operator $P_0: \mathcal{V}_0 \rightarrow \mathcal{V}$ and the restriction operator $R_0: \mathcal{V} \rightarrow \mathcal{V}_0$, where $R_0 = P_0^\top$. Moreover, we also employ the projection operator $\Pi_0: \mathcal{V} \rightarrow \mathcal{V}_0$ in order to transfer primal quantities to the coarse level.

In the context of nonlinear multilevel methods, several approaches for constructing the coarse-level feasible set $\mathcal{F}_0 = \{v_0 \in \mathcal{V}_0 \mid \underline{\psi}_0 \leq v_0 \leq \bar{\psi}_0\}$ are considered in the literature [9, 10, 14]. Here, we utilize constraint-projection rules from [9] and construct $\underline{\psi}_0, \bar{\psi}_0$ in a component-wise manner as

$$\begin{aligned} (\underline{\psi}_0)_t &= (v_0)_t + \max_{j \in \mathcal{N} \cap (\omega_0)_t} [(\underline{\psi} - v^{(k)})_j], \\ (\bar{\psi}_0)_t &= (v_0)_t + \min_{j \in \mathcal{N} \cap (\omega_0)_t} [(\bar{\psi} - v^{(k)})_j], \end{aligned} \quad (6)$$

where the symbol $(\cdot)_t$ denotes the value of a function associated with the t -th node of the mesh. The support of the basis function $(\phi_0)_t$ is denoted by $(\omega_0)_t$. Now, we

Algorithm 1: RASPN-B method**Data:** $f : \mathcal{V} \rightarrow \mathbb{R}$, $\underline{\psi} \in \mathcal{V}$, $\bar{\psi} \in \mathcal{V}$, $v^{(0)} \in \mathcal{F}$, $k \leftarrow 0$ **Result:** $v^{(k)}$

-
- 1 **while** $\|[\nabla f(v^{(k)})]_{\mathcal{F}}\| \geq \epsilon_{atol}$ **do**
 - 2 For given $v^{(k)}$, find $v^{(+)}$ by using a step of NRAS-B or TL-NRAS-B method
 - 3 Assemble gradient and Hessian: $g \leftarrow \nabla f(v^{(+)})$, $H \leftarrow \nabla^2 f(v^{(+)})$
 - 4 Find $s^{(k)}$ by solving the following constrained quadratic optimization problem

$$\min_{s^{(k)}} Q(s^{(k)}) := 1/2 \langle H s^{(k)}, s^{(k)} \rangle + \langle g, s^{(k)} \rangle, \text{ s. t. } \underline{\psi} - v^{(+)} \leq s^{(k)} \leq \bar{\psi} - v^{(+)}$$
 - 5 Find $\alpha^{(k)}$ using a line-search algorithm
 - 6 Update the iterate: $v^{(k+1)} \leftarrow v^{(+)} + \alpha^{(k)} s^{(k)}$, $k \leftarrow k + 1$
-

can define the optimization problem on the coarse level as follows. For a given initial guess $v_0^{(0)} \in \mathcal{F}_0$, where $v_0^{(0)} \leftarrow \Pi_0 v^{(k)}$:

$$\text{Find } v_0^* = \arg \min_{v_0 \in \mathcal{F}_0} \hat{f}_0(v_0). \quad (7)$$

Please note that minimization problem (7) is defined using an augmented coarse-level objective function \hat{f}_0 , defined as

$$\hat{f}_0(v_0) = f_0(v_0) + \langle R_0 \nabla f(v^{(k)}) - \nabla f_0(\Pi_0 v^{(k)}), v_0 \rangle, \quad (8)$$

where $v^{(k)}$ denotes the current iterate on the fine level. By adding the first-order consistency term to the objective function f_0 , we ensure that the gradient of the augmented objective function \hat{f}_0 at the first iterate is the restricted fine-level gradient.

We follow an inverted V-cycle approach, where a coarse-level update step is followed by a single step of NRAS-B iteration, i.e., iterate $v^{(k)}$ is updated as follows:

$$\begin{aligned} v^{(k+1/2)} &= v^{(k)} + \hat{\alpha} P_0(v_0^* - \Pi_0 v^{(k)}), \\ v^{(k+1)} &= v^{(k+1/2)} + \alpha \sum_{i=1}^n \tilde{P}_i(v_i^* - R_i v^{(k+1/2)}). \end{aligned} \quad (9)$$

The symbol v_0^* in (9) denotes the solution of the coarse-level minimization problem (7), while v_i^* is the solution of the subproblem (4) associated with the i -th subspace. The step sizes $\hat{\alpha}$ and α are again obtained using a line-search algorithm. Combining solutions of (7) and (4) with update rule (9), we can define an iteration of the TL-NRAS-B method.

Nonlinearly-preconditioned Newton SQP method: Finally, we provide a brief description of the nonlinearly-preconditioned Newton-SQP method for bound-constrained optimization problems. As summarized in Alg. 1, the method consists of two main phases. First, we invoke a step of the NRAS-B/TL-NRAS-B method in order to obtain an updated iterate $v^{(k)}$. Later, we construct a quadratic model Q , which is minimized subject to the bound constraints with the aim of obtaining a new

search direction $s^{(k)}$. In contrast to standard preconditioned Newton methods [5, 7], the minimization of the quadratic model is subjected to pointwise constraints, which ensures that the updated iterates remain in the feasible set \mathcal{F} . We note that the right-preconditioning can also be interpreted as a multiplicative or a composite solver [4].

3 Numerical experiments

In this section, we investigate the performance of nonlinear Schwarz preconditioners using two constrained minimization problems, namely the ignition and the minimal surface problems. Both numerical examples are defined on a domain $\Omega := (0, 1)^2$ with boundary $\Gamma = \partial\Omega$, which is decomposed into four parts: $\Gamma_l = \{0\} \times [0, 1]$, $\Gamma_r = \{1\} \times [0, 1]$, $\Gamma_b = [0, 1] \times \{0\}$ and $\Gamma_t = [0, 1] \times \{1\}$. The discretization is performed using a mesh consisting of 120×120 uniform quadrilaterals which are further decomposed into triangular elements. In the case of two-level methods, we also employ a coarser mesh with 30×30 elements in each direction.

Ignition: We minimize a variant of the Bratu problem, given as:

$$\begin{aligned} \min_{u \in H^1(\Omega)} f_I(u) &:= \frac{1}{2} \int_{\Omega} \|\nabla u\|^2 - (ue^u - e^u) dx - \int_{\Omega} f(x)u dx, \\ \text{subject to } \underline{\psi}(x) &\leq u \leq \overline{\psi}(x), \quad \text{a.e. in } \Omega, \quad u = 0, \quad \text{on } \Gamma, \end{aligned} \tag{10}$$

where $f(x) = (9\pi^2 + e^{(x_1^2 - x_1^3)} \sin(3\pi x_2))(x_1^2 - x_1^3) + 6x_1 - 2) \sin(3\pi x_1)$. The bounds are given as $\underline{\psi}(x) = 0.2 - 8(x_1 - 7/16)^2 - 8(x_2 - 7/16)^2$ and $\overline{\psi}(x) = 0.5$.

Minimal Surface: This experiment aims to find the minimal surface described by a function u by solving the following minimization problem:

$$\begin{aligned} \min_{u \in H^1(\Omega)} f_M(u) &= \int_{\Omega} \sqrt{(1 + \|\nabla u\|^2)} dx, \\ \text{subject to } \begin{cases} \underline{\psi}(x) \leq u \leq \overline{\psi}(x) & \text{a.e. in } \Omega. \\ u = -0.3 \sin(2\pi x_2) & \text{on } \Gamma_l, \quad u = 0.3 \sin(2\pi x_2) & \text{on } \Gamma_r, \\ u = -0.3 \sin(2\pi x_1) & \text{on } \Gamma_b, \quad u = 0.3 \sin(2\pi x_1) & \text{on } \Gamma_t, \end{cases} \end{aligned} \tag{11}$$

where the lower bound is prescribed as $\underline{\psi}(x) = 0.25 - 8(x_1 - 0.7)^2 - 8(x_2 - 0.7)^2$ and the upper bound is $\overline{\psi}(x) = 8(x_1 - 0.3)^2 - 8(x_2 - 0.3)^2 - 0.4$.

Setup of the solution strategies: For all numerical experiments, we prescribe the overlap $\delta = 3$ and obtain the decomposition into the subdomains using the library METIS. All considered solution methods terminate if $\|[\nabla f]_{\mathcal{F}}\| \leq 10^{-8}$, where $[\nabla f]_{\mathcal{F}} = \mathcal{P}_{\mathcal{F}}(x - \nabla f(x)) - x$ denotes the projected gradient. Here, the symbol $\mathcal{P}_{\mathcal{F}}$ denotes the projection onto the feasible set \mathcal{F} . Contrary to the traditional nonlinear RAS methods, the subdomain solvers, coarse-level solvers, and constrained quadratic minimization solvers are terminated using a fairly strict termination crite-

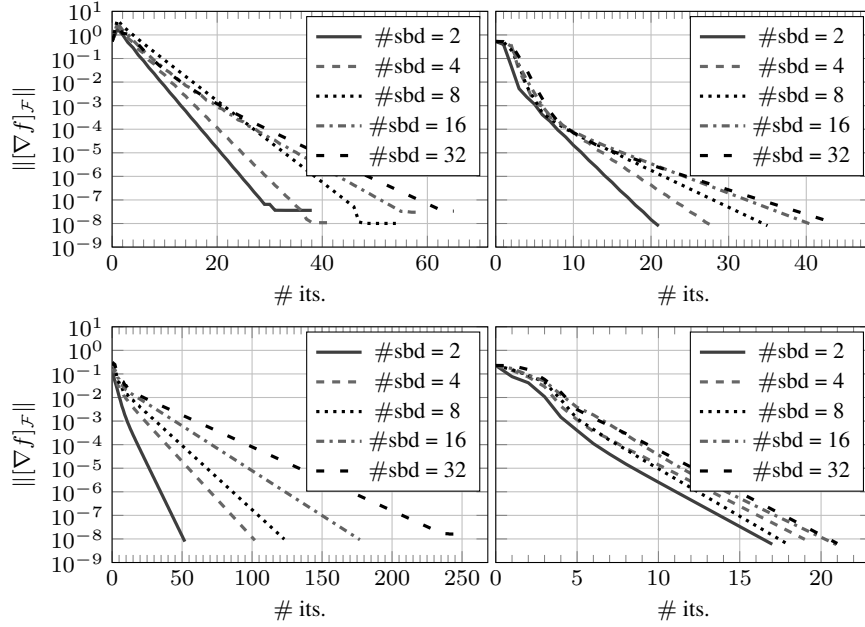


Fig. 1 Convergence history of NRAS-B (Left) and TL-NRAS-B (Right) methods for the ignition problem (Top) and the minimal surface (Bottom) problem. The experiments are performed with an increasing number of subdomains (sbd).

tion, i.e., they terminate if $\|[\nabla f]_{\mathcal{F}}\| \leq 10^{-11}$. Moreover, we employ a line-search method with the Armijo condition for computing the step size in all inner and outer solvers. The local and the coarse-level solvers for the NRAS-B/TL-NRAS-B methods employ the Newton-SQP method. On coarse levels, simply restriction of the objective function f_I and f_M is used to construct f_0 .

Comparison between NRAS-B and TL-NRAS-B methods: The comparison is performed with respect to an increasing number of subdomains. As we can observe from Fig. 1, the standard NRAS-B method requires more iterations to satisfy the termination criterion than the TL-NRAS-B method. Due to the strict termination criterion, we notice that the NRAS-B method stagnates before reaching the termination criterion, while the TL-NRAS-B method manage to converge to the desired tolerance irrespective of number of subdomains. We also notice that the NRAS-B method requires more iterations with an increasing number of subdomains for both problems. For the TL-NRAS-B, we observe scalable convergence with respect to the number of subdomains for the minimal surface problem. However, for the ignition problem, the number of iterations grows with the number of subdomains. This can be attributed to the fact that the coarse grid is not able to represent the constraint function associated with this particular problem sufficiently well. Hence, the coarse-level nonlinear problems become over-constrained, which amounts to small coarse-grid corrections and insufficient global information transfer.

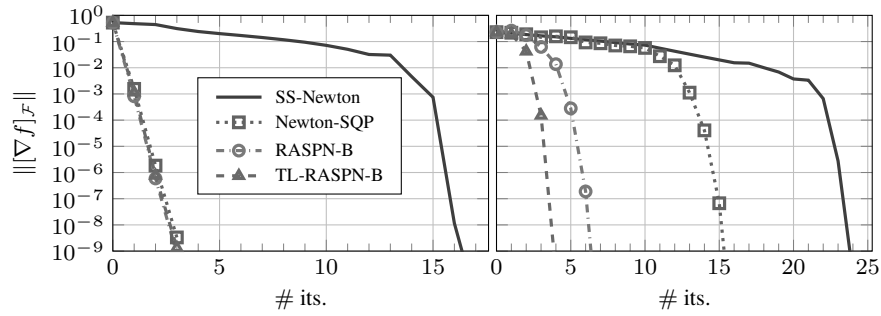


Fig. 2 Convergence history of semismooth Newton (SS-Newton), Newton-SQP, RASP-N-B, and TL-RASP-N-B methods for the ignition problem (Left) and minimal surface problem (Right).

Comparing RASP-N-B method with other methods: In this section, we compare the performance of the NRAS-B and TL-NRAS-B preconditioned Newton methods with the semismooth Newton and Newton-SQP methods. In this study, the NRAS-B and TL-NRAS-B methods employ 16 subdomains. We note that the semismooth Newton method linearizes the nonlinearity of the problem and constraints simultaneously, while the Newton-SQP method first linearizes the nonlinearity of the problem and at each Newton iteration a QP problem is solved with constraints.

From Fig. 2, we can see that the Newton-SQP method preconditioned with NRAS-B and TL-NRAS-B method outperforms the semismooth Newton method for both examples. As the ignition problem is only mildly nonlinear, preconditioned Newton methods as well as the Newton-SQP method manage to satisfy the termination criterion in only 4 iterations. In the case of a minimal surface problem, which is more nonlinear, the benefit of preconditioning the Newton method is more evident. The RASP-N-B and TL-RASP-N-B methods converge in 7 and 4 iterations, respectively. In comparison, the Newton-SQP and semismooth-Newton methods require 16 and 24 iterations to converge, respectively.

4 Conclusion

In this work, we presented a nonlinear additive Schwarz preconditioning method for bound-constrained nonlinear optimization problems. The scalability of the method is enhanced by introducing a coarse level with the first-order consistent objective function and the constraints restricted from the fine level. The developed Schwarz methods are then employed as a right preconditioner for the Newton-SQP method. Our numerical results demonstrate that the proposed preconditioners enhance the convergence of the Newton-SQP method and outperform standard active-set Newton methods. We also show that the two-level preconditioner is algorithmically scalable if a coarse space captures the constraints from the fine level sufficiently well.

Acknowledgements H.K. and R.K. thank the Swiss National Science Foundation (SNSF) and the Deutsche Forschungsgemeinschaft for their support through the project SPP 1962 “Stress-Based Methods for Variational Inequalities in Solid Mechanics: Finite Element Discretization and Solution by Hierarchical Optimization” [186407]. A.K. gratefully acknowledges the support of the SNSF under the project “Multilevel training of DeepONets - multiscale and multiphysics applications” [206745]. Additionally, authors acknowledge the support of Platform for Advanced Scientific Computing through the project “FraNetG: Fracture Network Growth” and SNSF through the project ML2 - Multilevel and Domain Decomposition Methods for Machine Learning” [197041].

References

1. Badea, L. and Krause, R. One-and two-level Schwarz methods for variational inequalities of the second kind and their application to frictional contact. *Numerische Mathematik* **120**(4), 573–599 (2012).
2. Badea, L., Tai, X.-C., and Wang, J. Convergence rate analysis of a multiplicative Schwarz method for variational inequalities. *SIAM Journal on Numerical Analysis* **41**(3), 1052–1073 (2003).
3. Badea, L. and Wang, J. An additive Schwarz method for variational inequalities. *Mathematics of Computation* **69**(232), 1341–1354 (2000).
4. Brune, P. R., Knepley, M. G., Smith, B. F., and Tu, X. Composing scalable nonlinear algebraic solvers. *SIAM Review* **57**(4), 535–565 (2015).
5. Cai, X.-C. Nonlinear overlapping domain decomposition methods. In: *Domain Decomposition Methods in Science and Engineering XVIII*, 217–224. Springer (2009).
6. Cai, X.-C. and Keyes, D. E. Nonlinearly Preconditioned Inexact Newton Algorithms. *SIAM Journal on Scientific Computing* **24**(1), 183–200 (2002).
7. Cai, X.-C. and Li, X. Inexact Newton methods with restricted additive Schwarz based nonlinear elimination for problems with high local nonlinearity. *SIAM Journal on Scientific Computing* **33**(2), 746–762 (2011).
8. Dolean, V., Gander, M. J., Kheriji, W., Kwok, F., and Masson, R. Nonlinear Preconditioning: How to Use a Nonlinear Schwarz Method to Precondition Newton’s Method. *SIAM Journal on Scientific Computing* **38**(6), A3357–A3380 (2016).
9. Gelman, E. and Mandel, J. On multilevel iterative methods for optimization problems. *Mathematical Programming* **48**(1), 1–17 (1990).
10. Gratton, S., Mouffe, M., Toint, P. L., and Weber-Mendonça, M. A recursive ℓ_∞ -trust-region method for bound-constrained nonlinear optimization. *IMA Journal of Numerical Analysis* **28**(4), 827–861 (2008).
11. Gross, C. and Krause, R. A new class of non-linear additively preconditioned trust-region strategies: Convergence results and applications to non-linear mechanics (2009). Available as INS Preprint No. 904.
12. Kopaničáková, A., Kothari, H., and Krause, R. Nonlinear field-split preconditioners for solving monolithic phase-field models of brittle fracture. *Computer Methods in Applied Mechanics and Engineering* **403**, 115733 (2023).
13. Kopaničáková, A. and Krause, R. A recursive multilevel trust region method with application to fully monolithic phase-field models of brittle fracture. *Computer Methods in Applied Mechanics and Engineering* **360**, 112720 (2020).
14. Kopaničáková, A. and Krause, R. A Multilevel Active-Set Trust-Region (MASTR) Method for Bound Constrained Minimization. In: *Domain Decomposition Methods in Science and Engineering XXVI*, 355–363. Springer (2022).
15. Kothari, H. and Krause, R. A generalized multigrid method for solving contact problems in lagrange multiplier based unfitted finite element method. *Computer Methods in Applied Mechanics and Engineering* **392**, 114630 (2022).
16. Nocedal, J. and Wright, S. *Numerical Optimization*. Springer (2000).