On the Accuracy of the Inner Newton Iteration in Nonlinear Domain Decomposition

Axel Klawonn, Martin Lanser, Oliver Rheinbach, and Matthias Uran

1 Introduction

Nonlinear FETI-DP methods [4, 5, 6, 7] belong to the family of nonoverlapping nonlinear domain decomposition methods and can be used to solve discrete nonlinear problems $A(u) = 0$ arising from the discretization of nonlinear partial differential equations. They can be characterized by decomposition before linearization, and they can be interpreted as nonlinearly right-preconditioned Newton-Krylov methods; see [6]. These methods localize work and have shown to be highly scalable to more than 131 072 cores [6].

We decompose the computational domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, into N nonoverlapping subdomains Ω_i, $i = 1, \ldots, N$, such that $\Omega = \bigcup_i^N \Omega_i$. The associated local finite element spaces are denoted by $W^{(i)}$ and the product space by $W = W^{(1)} \times \cdots \times W^{(N)}$. We introduce $\widetilde{W} \subset W$ as the space of all finite element functions from W which are continuous in certain primal variables, e.g., subdomain vertices.

The fully assembled original finite element problem is equivalent to the nonlinear FETI-DP saddle point system

$$A(\tilde{u}, \lambda) = \begin{bmatrix} \widetilde{K}(\tilde{u}) + B^T \lambda - \tilde{f} \\ B\tilde{u} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \tilde{u},\ \tilde{f},\ \widetilde{K}(\tilde{u}) \ \in \widetilde{W}; \tag{1}$$

see [4]. Nonlinear FETI-DP methods are based on solving (1). Here, Lagrange multipliers $\lambda \in V$ are used to decompose the nonlinear problem into parallel local problems on subdomains, and the linear constraint $B\tilde{u} = 0$ enforces the continuity of

Axel Klawonn, Martin Lanser, Matthias Uran
Mathematisches Institut, Universität zu Köln, Weyertal 86-90, 50931 Köln, Germany, e-mail: {axel.klawonn, martin.lanser, m.uran}@uni-koeln.de

Oliver Rheinbach
Institut für Numerische Mathematik und Optimierung, Fakultät für Mathematik und Informatik, Technische Universität Bergakademie Freiberg, Akademiestr. 6, 09596 Freiberg, e-mail: oliver.rheinbach@math.tu-freiberg.de

1. Mapping: $M : \widetilde{W} \times V \to \widetilde{W} \times V$.
2. M puts the current iterate into the neighborhood of the solution; see also [1].
3. $M(\tilde{u}, \lambda)$ is easily computable compared to the inverse action of $A(\tilde{u}, \lambda)$.

Fig. 1 Properties on the nonlinear preconditioner M for nonlinear FETI-DP methods.

the solution across the interface for nonprimal variables. Here, B is the standard finite element jump operator and the space of Lagrange multipliers is defined as $V := \text{range}(B)$.

Instead of solving $A(\tilde{u}, \lambda) = 0$ directly with Newton's method, which was denoted Nonlinear-FETI-DP-1 in [4, 6], we introduce a nonlinear right-preconditioner $M(\tilde{u}, \lambda)$; see Figure 1 for some desirable properties the preconditioner should fulfill. The resulting nonlinear equation

$$A(M(\tilde{u}, \lambda)) = 0 \tag{2}$$

is solved by a Newton-Krylov method. In each Newton iteration the evaluation of the preconditioner $g^{(k)} = M(\tilde{u}^{(k)}, \lambda^{(k)})$ is computed. The nonlinear right-preconditioner can be used to describe a (partial) nonlinear elimination of variables [3]. We introduce the index sets E and L, where E is the set of variables which will be eliminated nonlinearly by the application of M and L is the set of variables which will be linearized. According to these two index sets, we split the variables $\tilde{u}$, and the jump operator B, $\tilde{u} = (\tilde{u}_E, \tilde{u}_L)$, $B = \begin{bmatrix} B_E & B_L \end{bmatrix}$. Using this splitting, the nonlinear system (1) writes

$$A(\tilde{u}_E, \tilde{u}_L, \lambda) = \begin{bmatrix} A_E(\tilde{u}_E, \tilde{u}_L, \lambda) \\ A_L(\tilde{u}_E, \tilde{u}_L, \lambda) \\ B_E\tilde{u}_E + B_L\tilde{u}_L \end{bmatrix} = \begin{bmatrix} \widetilde{K}_E(\tilde{u}_E, \tilde{u}_L) + B_E^T\lambda - \tilde{f}_E \\ \widetilde{K}_L(\tilde{u}_E, \tilde{u}_L) + B_L^T\lambda - \tilde{f}_L \\ B_E\tilde{u}_E + B_L\tilde{u}_L \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{3}$$

Since the nonlinear elimination process is restricted to the variables $\tilde{u}_E$, the nonlinear preconditioner $M(\tilde{u}, \lambda)$ is linear in $\tilde{u}_L$ and λ. Therefore, we introduce the following notation

$$M(\tilde{u}, \lambda) = M(\tilde{u}_E, \tilde{u}_L, \lambda) := (M_{\tilde{u}_E}(\tilde{u}_E, \tilde{u}_L, \lambda), \tilde{u}_L, \lambda) = (M_{\tilde{u}_E}(\tilde{u}_L, \lambda), \tilde{u}_L, \lambda) \tag{4}$$

and $M_{\tilde{u}_E}(\tilde{u}_E, \tilde{u}_L, \lambda)$ is defined implicitly by

$$\widetilde{K}_E(M_{\tilde{u}_E}(\tilde{u}_E, \tilde{u}_L, \lambda), \tilde{u}_L) + B_E^T\lambda - \tilde{f}_E = 0. \tag{5}$$

Hence, for the evaluation of $g^{(k)} := M(\tilde{u}_E^{(k)}, \tilde{u}_L^{(k)}, \lambda^{(k)})$, the nonlinear system

$$A_E(g^{(k)}) = 0 \tag{6}$$

has to be solved for fixed $\tilde{u}_L^{(k)}$ and $\lambda^{(k)}$ until a sufficient tolerance ε_I is reached, e.g., by Newton's method with the partial update

$g_0^{(k)} = (\tilde{u}^{(k)}, \lambda^{(k)})$ and $l = 0$
while $||A_E(g_l^{(k)})|| > \varepsilon_I$ **do**
  Newton update to $g_{l+1}^{(k)}$
  $l = l + 1$
  $g^{(k)} = g_l^{(k)}$
end while

$g_0^{(k)} = (\tilde{u}^{(k)}, \lambda^{(k)})$, $l = 0$, $J_{\text{old}} = \frac{1}{2}||A(g_0^{(k)})||^2$
while $||A_E(g_l^{(k)})|| > \varepsilon_I$ **do**
  Newton update to $g_{l+1}^{(k)}$
  Compute: $J_{\text{new}} = \frac{1}{2}||A(g_{l+1}^{(k)})||^2$
  if $J_{\text{new}} > \tau J_{\text{old}}$ **then**
    $g^{(k)} = g_l^{(k)}$
    break while
  else
    $J_{\text{old}} = J_{\text{new}}$
  end if
  $l = l + 1$
  $g^{(k)} = g_l^{(k)}$
end while

Fig. 2 Left: Computation of M. **Right:** Computation of $\mathscr{M}$.

$$g_{E,l+1}^{(k)} = g_{E,l}^{(k)} - (D_{\tilde{u}_E} A_E(g_l^{(k)}))^{-1} A_E(g_l^{(k)}); \tag{7}$$

see also Figure 2 on the left. Thus, the application of the nonlinear right-preconditioner is nothing else than minimizing the energy $J_E(\tilde{u}, \lambda) := \frac{1}{2}||A_E(\tilde{u}, \lambda)||^2$.

Replacing $\tilde{u}_E$ in the second and third line of (3) by $M_{\tilde{u}_E}(\tilde{u}_L, \lambda)$ yields the nonlinear Schur complement

$$S_L(\tilde{u}, \lambda) := \begin{bmatrix} \widetilde{K}_L(M_{\tilde{u}_E}(\tilde{u}_L, \lambda), \tilde{u}_L) + B_L^T \lambda - \tilde{f}_L \\ B_E M_{\tilde{u}_E}(\tilde{u}_L, \lambda) + B_L \tilde{u}_L \end{bmatrix}. \tag{8}$$

Finally, we can solve the resulting nonlinear Schur complement system $S_L(\tilde{u}, \lambda) = 0$ with standard Newton-Krylov-FETI-DP (see [4]). For more details, we also refer to [6].

2 Nonlinear FETI-DP Methods Using Energy Reducing Nonlinear Preconditioning

It is possible that the nonlinear elimination presented above leads to an increase in the global energy $J(\tilde{u}, \lambda) = \frac{1}{2}||A(\tilde{u}, \lambda)||^2$, e.g., if the strong nonlinearities are not contained in the index set E. In this case, our nonlinear FETI-DP methods can show a loss of robustness and performance compared to the traditional Newton-Krylov-FETI-DP approach; see Section 3. It can also happen, that our nonlinear FETI-DP methods do not converge to a solution due to an inappropriate coarse space.

To increase the convergence radius for Newton type methods it is standard to enforce a sufficient decrease in the global energy J in each Newton step [9]. This can be achieved by controlling the Newton update. If the Newton update does not result in a sufficient decrease of the energy, the Newton step is rejected and replaced, e.g., by

a steepest descent step. To prove global convergence properties, usually additional assumptions about the step length have to be fulfilled, which can be controlled by a line search approach enforcing certain conditions of, e.g., Armijo or Wolfe type [8]. For the use of line search in nonlinear FETI-DP methods, see [4].

Analogously to classical Newton-Krylov approaches, it is also possible to apply these strategies to nonlinear right-preconditioned Newton-Krylov methods, which is not considered in this paper. Nevertheless, we additionally have to control the application of the nonlinear preconditioner to enforce an energy decrease in each step, or, at least, to avoid an increase with respect to J.

To enlarge the convergence radius of our nonlinear FETI-DP methods we therefore have to compute $g^{(k)}$ not only with respect to $J_E = \frac{1}{2}||A_E||^2$ but also $J = \frac{1}{2}||A||^2$; cf. (3).

As described above, the application of the nonlinear preconditioner M in our nonlinear FETI-DP methods leads to a minimization of $\frac{1}{2}||A_E(\tilde{u},\lambda)||^2$, but we do not control how the global energy J evolves during this update process. To do so, we introduce an approximation $\mathscr{M}(\tilde{u},\lambda)$ of $M(\tilde{u},\lambda)$, which at least does not increase the global energy J. The idea is, to stop the Newton iteration and choose $\mathscr{M}(\tilde{u},\lambda) = g_l$ whenever the updated g_{l+1} does not fulfill the simple decrease property $J(g_{l+1}) \leq \tau J(g_l)$ for the global energy functional. We thus avoid oversolving in the inner Newton iteration, somewhat analogously to inexact Newton methods with carefully chosen forcing terms [2]. To make this property a robust decrease condition, we choose $0 < \tau \leq 1$ and, if not noted otherwise, we use $\tau = 0.8$ in our experiments. For more details see Figure 2 on the right.

It is obvious that this approach never leads to an increased number of inner Newton iterations but it can end up with two extreme cases. First, if the decrease property is fulfilled for all inner Newton steps we have $\mathscr{M}(\tilde{u},\lambda) = M(\tilde{u},\lambda)$. Second, if the decrease condition is not fulfilled for the first inner Newton step, we obtain $\mathscr{M}(\tilde{u},\lambda) = (\tilde{u},\lambda)$ and the application of $\mathscr{M}$ reduces to the identity. The latter case is identical to a single step of Nonlinear-FETI-DP-1, regardless which set of variables E is chosen. Let us briefly recall the definition of Nonlinear-FETI-DP-1 from [6], where the variable set E is chosen to be the empty set. Let us also remark that in the second case all factorizations from the inner Newton iteration can be recycled for the subsequent outer Newton iteration and therefore no additional work compared with a Nonlinear-FETI-DP-1 step is necessary.

Let us remark that we handle the very first computation of $\mathscr{M}$ in a slightly different way, since we do not want to rely on the initial value $\tilde{u}^{(0)}$. We do not stop the Newton iteration if $J(g_1) > \tau J(g_0)$ but we also compute g_l until $J(g_l) \leq \tau J(g_{l-1})$, $l \geq 2$, is not fulfilled. In a similar way we can control the computation of the initial value $\widetilde{K}(\tilde{u}^{(0)}) = \tilde{f} - B^T\lambda^{(0)}$ (see [4]) in the Nonlinear-FETI-DP-1 approach.

In each outer Newton iteration, we now have to solve the linear system

$$DA\left(\mathscr{M}(\tilde{u}^{(k)},\lambda^{(k)})\right)\left(\delta\tilde{u}^{(k)},\lambda^{(k)}\right)^T = A\left(\mathscr{M}(\tilde{u}^{(k)},\lambda^{(k)})\right). \tag{9}$$

Here, the entries in the right hand side belonging to the index set E can not be guaranteed to be zero due to the fact that $\mathscr{M}$ might just be an approximation to M.

3 Numerical Results

In this section, we present numerical results for nonlinear FETI-DP methods using the newly introduced energy reducing and robust preconditioner and compare them to the nonlinear FETI-DP methods introduced in [4, 5, 6, 7] and to the traditional Newton-Krylov-FETI-DP approach. To provide a fair comparison, we choose for all methods the same initial values $u^{(0)}(x_1,x_2) = x_1 \cdot x_2 \cdot (1-x_1) \cdot (1-x_2)$, $\lambda^{(0)} = 0$, and the same tolerances ε_I and ε_O. Inner Newton iterations are stopped if $\frac{1}{2}||A_E||^2 \leq \varepsilon_I = 1e-12$ or the decrease condition is not fulfilled and the global Newton iteration is stopped if $\frac{1}{2}||A||^2 \leq \varepsilon_O = 1e-12$.

We refer to the nonlinear FETI-DP methods als NL-i, $i = 1,\ldots,4$, and to the nonlinear FETI-DP methods using the new nonlinear preconditioner as NL-ane-i, $i = 1,\ldots,4$. The traditional Newton-Krylov-FETI-DP method is denoted NK. Let us briefly recall the different nonlinear variants from [6] by specifying the nonlinear elimination sets. We choose $E = \emptyset$ in NL-1, $E = [I,\Delta,\Pi]$ in NL-2, $E = [I,\Delta]$ in NL-3, and $E = I$ in NL-4, where I denotes the set of variables inside subdomains, Π denotes the set of primal variables, and Δ denotes the set of all remaining interface variables.

As a model problem, we choose a two dimensional problem based on the scaled p-Laplace operator for $p = 4$

$$\alpha\Delta_p u := \operatorname{div}(\alpha|\nabla u|^{p-2}\nabla u).$$

We consider

$$\begin{aligned} -\alpha\Delta_4 u - \beta\Delta_2 u &= 1 \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned}$$

with the computational domain $\Omega = (0,1)^2$ and the coefficients $\alpha = 1e5$ and $\beta = 1$.

The computational domain is decomposed into square subdomains and discretized by piecewise linear finite elements. We choose a problem, where the nonlinearities have a nonlocal character. Here, columns of subdomains are intersected by channels of width $H/2$ from the upper to the lower boundary of Ω, where H is the width of a subdomain; see the left picture in Figure 3. To simulate a less structured domain decomposition, we also consider subdomains with ragged edges; see the right picture in Figure 3 for details. For all our tests we used a sequential MATLAB implementation and we exclusively consider subdomain vertices as primal constraints. Due to our sequential implementation, we choose and evaluate different metrics or indicators to obtain a good estimation of the parallel potential of the different nonlinear FETI-DP preconditioner variants. As a metric for the global communication, we count the number of Krylov iterations (denoted # Krylov It.). For the local work, we count the number of factorizations of $D\widetilde{K}_{BB}$ or $D\widetilde{K}_{II}$ (denoted by "Local Fact."), and we also count the factorizations of the FETI-DP coarse problem (denoted by "Coarse Fact."). Factorizations of the coarse problem are necessary in the computation of the initial value for NL-1 and in the evaluation of the

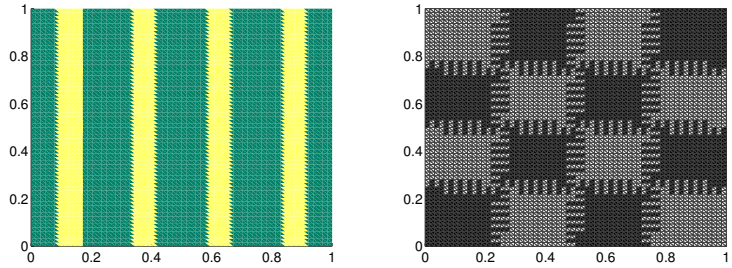

Fig. 3 Left: Channels with a width of $H/3$, where H is the width of a subdomain; $\alpha = 1e5$. **Right:** Domain decomposition with ragged edges, $H/h = 16$.

nonlinear preconditioner for NL-2, while the evaluation of the preconditioner for NL-3 and NL-4 does not include factorizations of the coarse problem. Therefore, we subdivide the section "Coarse Fact." into factorizations of the coarse problem in the first/inner loop (denoted by "in.") and in the main loop (denoted by "out."). For all methods the number of outer coarse factorizations is equal to the number of Newton steps.

For our model problem, the index set E does not contain the nonlinearities for the NL-4 and NL-ane-4 method. As a result the performance of NL-4 is worse than the performance of the traditional NK approach and the number of local factorizations of NL-ane-4 is equal to the number of Newton steps plus one. This shows that the elimination of the interior variables is inappropriate for this problem, but NL-ane-4 detects this and avoids spending time in the evaluation of the inappropriate nonlinear preconditioner. As a consequence, NL-ane-4 is nearly equivalent to NL-1 without the computation of the initial value or to NK and thus superior compared to NL-4. The difference of one factorization results from the additional step in the inner loop in the very first Newton step.

For the structured decomposition into square subdomains NL-2, NL-3, NL-ane-2, and NL-ane-3 perform quite similar. The number of local solves for NL-ane-2 and NL-ane-3 is half as large as for NL-2 and NL-3, but the number of Krylov iterations is slightly higher.

For the less structured decomposition with ragged edges the chosen coarse space (subdomain vertices) is insufficient for NL-2 and NL-3, so these methods do not converge, but using the new approach leads to convergence and saves about 50% of Newton steps and Krylov iterations compared to the traditional NK approach. The new strategy thus increases the convergence radius for NL-2 and NL-3.

4 Conclusion

We have introduced a strategy to automatically decide on the computational effort to be spent in the inner Newton iteration in nonlinear domain decomposition. The

Table 1 Model problem "Nonlocal Nonlinearities"; comparison of standard nonlinear FETI-DP methods and nonlinear FETI-DP methods using the new approach ("NL-ane-*"); channels with a width of $H/2$; $\alpha = 1e5$ inside channels and $\beta = 1$ elsewhere; see also Figure 3; domain $\Omega = (0,1)^2$; decomposed into square subdomains; $H/h = 16$; $\varepsilon_I = 1e{-}12$; $\varepsilon_O = 1e{-}12$; $\tau = 0.8$; computed on Schwarz.

Channels 2D $H/h = 16$**; exact FETI-DP; computed on Schwarz,** $\alpha = 1e5$											
				Normal Edges				**Ragged Edges**			
N	**Problem Size**	$\lvert E\rvert$	**Nonlinear Solver**	**Local Factor. in.**	**Coarse Factor. out.**		**Krylov It.**	**Local Factor. in.**	**Coarse Factor. out.**		**Krylov It.**
16	4225		NK	**13**	-	13	173	**13**	-	13	2108
		0	NL-1 no Init	**13**	-	13	188	**13**	-	13	2227
		0	NL-ane-1	15	6	9	124	14	5	9	1211
		0	NL-1	22	12	10	150	26	17	9	1170
		4225	NL-ane-2	16	11	**5**	**68**	16	10	**6**	**794**
		4225	NL-2	30	24	6	86	div	div	div	div
		4216	NL-ane-3	23	**0**	7	95	25	**0**	9	1134
		4216	NL-3	30	**0**	6	86	div	div	div	div
		3856	NL-ane-4	17	**0**	13	254	14	**0**	13	2227
		3856	NL-4	47	**0**	13	284	43	**0**	13	2277
256	66049		NK	15	-	15	1391	15	-	15	3064
		0	NL-1 no Init	**14**	-	14	1471	**14**	-	14	3139
		0	NL-ane-1	16	6	10	741	**15**	5	10	2149
		0	NL-1	23	13	10	730	36	25	11	2387
		66049	NL-ane-2	16	10	**6**	447	19	11	**8**	**1664**
		66049	NL-2	31	25	**6**	395	div	div	div	div
		65824	NL-ane-3	17	**0**	**6**	429	18	**0**	**8**	1683
		65824	NL-3	35	**0**	**6**	**379**	div	div	div	div
		58624	NL-ane-4	19	**0**	14	1647	**15**	**0**	14	3139
		58624	NL-4	54	**0**	14	1681	50	**0**	14	3156

strategy considers the reduction of the global energy resulting from performing local Newton steps on the subdomains. The Newton iteration performed for the local elimination is stopped (and the step is discarded) when the resulting decrease in the global energy is not satisfactory. This can also be interpreted as an inexact nonlinear elimination. We have shown, that the local work can be significantly reduced compared to standard nonlinear FETI-DP methods while the number of Newton steps and Krylov iterations remains nearly constant. We have also shown, that the dependency on the coarse space is reduced for nonlinear FETI-DP methods and that the robustness of the resulting methods is dramatically increased.

Acknowledgements This work was supported in part by the German Research Foundation (DFG) through the Priority Programme 1648 "Software for Exascale Computing" (**SPPEXA**) under grants KL 2094/4-1, KL 2094/4-2, RH 122/2-1, and RH 122/3-2.

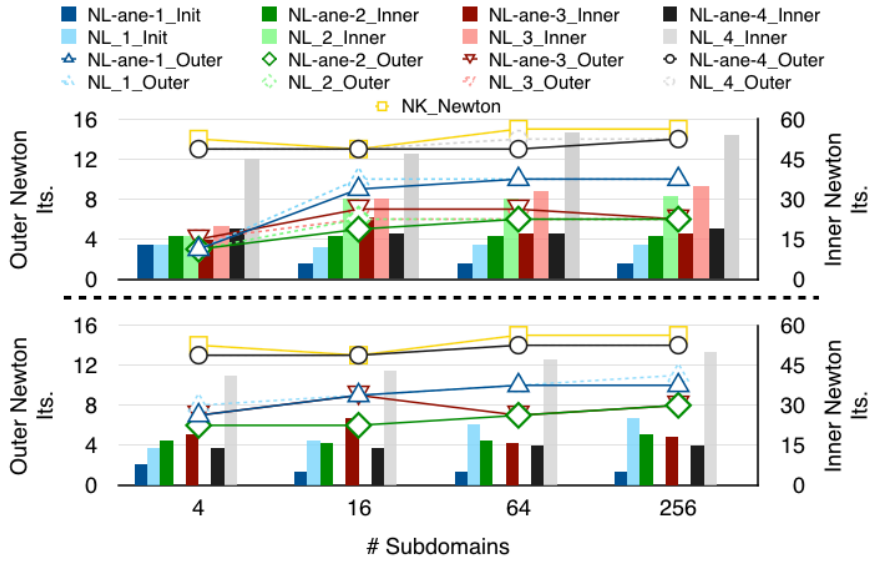

Fig. 4 Model problem "Nonlocal Nonlinearities"; comparison of nonlinear FETI-DP methods and nonlinear FETI-DP methods using energy minimizing preconditioning; channels with a width of $H/3$; $p = 4$ and $\alpha = 1e5$ in the channels and $\beta = 1$ elsewhere; see also Figure 3; domain $\Omega = (0,1)^2$; decomposed into square subdomains; $H/h = 16$; $\varepsilon_I = 1e{-}12$; $\varepsilon_O = 1e{-}12$; $\tau = 0.8$; computed on Schwarz. **Top:** Normal edges; **Bottom:** Ragged edges

References

1. Brune, P. R., Knepley, M. G., Smith, B. F., Tu, X.: Composing scalable nonlinear algebraic solvers. In: SIAM Rev., 57, pp. 535-565. (2015)
2. Eisenstat, S. C., Walker, S. F.: Choosing the forcing terms in an inexact Newton method. In: SIAM J. Sci. Comput., 17(1), pp. 16-32. (1996)
3. Lanzkorn, P. J., Rose, D. J., Wilkes, J. T.: An analysis of approximate nonlinear elimination. In: SIAM J. Sci. Comput., 17, pp. 538-559. (1996)
4. Klawonn, A., Lanser, M., Rheinbach, O.: Nonlinear FETI-DP and BDDC Methods. In: SIAM J. Sci. Comput., 36, pp. A737-A765. (2014)
5. Klawonn, A., Lanser, M., Rheinbach, O., Uran, M.: New nonlinear FETI-DP methods based on partial nonlinear elimination of variables: In: Lect. Notes Comp. Sci. Eng., vol 116, Proceedings of the 23rd International Conference on Domain Decomposition, Springer-Verlag, pp. 207-215. (2017)
6. Klawonn, A., Lanser, M., Rheinbach, O., Uran, M.: Nonlinear FETI-DP and BDDC Methods: A Unified Framework and Parallel Results. In: SIAM J. Sci. Comput., 39, pp. C417-C451. (2017)
7. Klawonn, A., Lanser, M., Radtke, P., Rheinbach, O.: On an adaptive coarse space and on nonlinear domain decomposition. In: Domain Decomposition Methods in Science and Engineering XXI, Erhel, J. and Gander, M. J. and Halpern, L. and Pichot, G. and Sassi, T. and Widlund, O., eds., vol. 98 of Lecture Notes in Computational Science and Engineering, Springer International Publishing, pp. 71-83. (2014)
8. Nocedal, J., Wright, S. J.: Numerical Optimization, 2nd Edition, Springer, Berlin. (2006)
9. Ulbrich, M., Ulbrich, S.: Nichtlineare Optimierung. Springer Basel, (2012).