Globalization of Nonlinear FETI–DP Methods

S. Köhler, and O. Rheinbach

1 Introduction

Nonlinear FETI-DP (Finite Element Tearing and Interconnection - Dual Primal) methods [10] are nonlinear generalizations of linear FETI-DP domain decomposition methods [16, 5]. Nonlinear FETI-DP domain decomposition methods have shown their robustness and scalability, e.g., for linear and nonlinear structural mechanics problems [11], where results for up to 786 432 cores were presented. Related nonlinear domain decomposition methods (DDMs) are nonlinear BDDC methods [10] (derived from linear Balancing Domain Decomposition by Constraints [4]), nonlinear FETI-1 methods [15] and the ASPIN approach (Additive Schwarz Preconditioned Inexact Newton) method [2, 9, 8].

The idea of nonlinear FETI-DP methods is to decompose the global problem $\widehat{K}(\hat{u}) = \hat{f}$ into local nonlinear problems $K_i(u_i) = f_i$, $i = 1, \ldots, N$, defined on nonoverlapping subdomains $\Omega_i = 1, \ldots, N$, and to enforce continuity on the interface as $\Gamma := \cup_i^N \partial\Omega_i \cap \partial\Omega$ using subassembly of primal variables and Lagrange multipliers λ.

Nonlinear FETI-DP methods make use of nonlinear elimination, where different methods result from different elimination sets. In [12], four different types of static elimination sets were introduced, referred to as Nonlinear-FETI-DP-x (*NL-x*), where $x = 1$ (no elimination is applied), $x = 2$ (primal, dual and inner variables are eliminated), $x = 3$ (dual and inner variables are eliminated) and $x = 4$ (only the inner variables are eliminated). Other choices of elimination sets include automatic strategies to determine the elimination set [7, 18].

If a tangent is available, nonlinear problems are typically solved by Newton's method or related methods such as quasi-Newton, inexact Newton or Newton-like

Stephan Köhler · 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, stephan.koehler@math.tu-freiberg.de, url: http://www.mathe.tu-freiberg.de/nmo/mitarbeiter/oliver-rheinbach

methods [14, 17]. However, without globalization Newton's method may fail to converge.

Common globalization methods are trust-region methods or line search methods. In this paper, we study line search methods for the globalization of nonlinear FETI-DP methods for nonlinear structural mechanics problems. We use an exact differentiable penalty function [1] related to the augmented Lagrange approach, but we can use the Hessian of the standard Lagrange function for a Newton-like descent method.

2 Nonlinear FETI-DP

Nonlinear FETI-DP methods are methods to solve the nonlinear saddle point problem

$$\begin{aligned} \widetilde{K}(\tilde{u}) + B^T\lambda &= \tilde{f}, \\ B\tilde{u} &= 0, \end{aligned} \tag{1}$$

which directly corresponds to the linear FETI-DP saddle point problem [16]. Here, B is the FETI-DP jump operator (as in linear FETI-DP methods), and λ is the vector of corresponding Lagrange multipliers. The nonlinear operator $\widetilde{K}(\tilde{u}) := R_\Pi^T K(R_\Pi \tilde{u})$ is obtained from finite element subassembly of the block operator $K(u) = [K_1(u_1), \ldots, K_N(u_N)]^T$ in the primal variables using the operator R_Π^T as in linear FETI-DP methods [16]. Here, this coupling provides a nonlinear coarse problem for the method. Thus, $\widetilde{K}$ represents a nonlinear coarse approximation of the original problem.

Next, we perform the nonlinear elimination: we split the first row in (1) according to disjoint index sets E, L (eliminate or linearize) and solve in a first step

$$\widetilde{K}_E(\tilde{u}_E, \tilde{u}_L) - \tilde{f}_E + B_E^T\lambda = 0, \tag{2}$$

for $\tilde{u}_E$, given $\tilde{u}_L$ and λ. Then, we can insert $\tilde{u}_E$ into the remaining equations and solve by linearization in $\tilde{u}_L$ and λ, and using the implicit function theorem. Let us recall, that for *NL*-1 we have $E = \emptyset$. For *NL*-2 the elimination set E contains all variables and $L = \emptyset$, for *NL*-3 we eliminate the inner and dual variables [16], and for *NL*-4 we eliminate only the inner variables. Automatic strategies to determine the elimination set E can also be considered but are not discussed here. Note that the local nonlinear elimination uses an exact Newton method in the sense that we perform a Newton iteration using a direct sparse solver for the Newton equation. This can be afforded since this is an operation local to a subdomain. For *NL*-2 the elimination involves also the (small) coarse space.

3 Exact Differentiable Penalty Method with Nonlinear Elimination

For $\nabla \tilde{J}(\tilde{u}) = \widetilde{K}(\tilde{u}) - \tilde{f}$ the equations in (1) are the first order optimality conditions for the minimization of the energy $\min \tilde{J}(\tilde{u})$ subject to the continuity constraint $B\tilde{u} = 0$, where $\tilde{J}(\tilde{u}) := J(R_{\Pi}\tilde{u})$ is obtained from the global energy $J(u) = \sum_{i=1}^{N} J^{(i)}(u_i)$.

Exact penalty method in nonlinear FETI-DP To be consistent with the vast literature in optimization, we will now use the notation

$$\min_{x \in \mathbb{R}^n} J(x) \qquad \text{subject to (s.t.)} \qquad c_i(x) = 0, \quad i = 1, \ldots, p, \tag{3}$$

where $J, c_i \in C^3(\mathbb{R}^n), i = 1, \ldots, p$. In the FETI-DP context, x is $\tilde{u}$, and $c(x) = 0$ are the continuity constraints $B\tilde{u} = 0$.

Penalty methods replace the original constrained minimization problem by a sequence of unconstrained minimization problems, where a penalty term, which measures the constraint violation, is added to objective function. In [3] the exact differentiable penalty function

$$P(x, \lambda; \mu, M) = \mathcal{L}(x, \lambda) + \frac{\mu}{2}\|c(x)\|^2 + \frac{1}{2}\|M(x)\nabla_x \mathcal{L}(x, \lambda)\|^2, \tag{4}$$

was introduced, where $\mathcal{L}$ is the Lagrange function, $\mu > 0$ and $M : \mathbb{R}^n \to \mathbb{R}^{m \times n}$, $p \leq m \leq n$. This penalty function is exact in the sense that for each local solution $\hat{x}$ of the original constrained minimization problem and the related Lagrange multipliers $\hat{\lambda}$, a finite penalty parameter $\overline{\mu}$ exists such that for $\mu > \overline{\mu}$ the point $\hat{x}$ is the first component of a local minimum $(\hat{x}, \hat{\lambda})$ of the penalty function $P(x, \lambda)$. In this sense, $\mu \to \infty$ is not needed. The function $P(\cdot, \cdot; \mu, M)$ is closely related to augmented Lagrange methods, but there are some differences. The most import advantage, compared to standard augmented Lagrange, especially in the nonlinear FETI-DP context, is the fact that we can use the standard Lagrange-Newton equation

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L}(x, \lambda) & \nabla^2_{x\lambda}\mathcal{L}(x, \lambda) \\ \nabla^2_{\lambda x}\mathcal{L}(x, \lambda) & O \end{bmatrix} \begin{bmatrix} \delta x \\ \delta \lambda \end{bmatrix} = - \begin{bmatrix} \nabla_x \mathcal{L}(x, \lambda) \\ \nabla_\lambda \mathcal{L}(x, \lambda) \end{bmatrix} \tag{5}$$

see, e.g. [1], to compute a Newton-like search direction. Therefore, we do not need to modify the Hessian of $\mathcal{L}$, as in the standard augmented Lagrange method.

A detailed analysis of P can be found in [1, Chapter 4.3], including a proof of the exactness of P on X^*, where $X^* := \{x \in \mathbb{R}^n \mid \nabla c(x) \text{ has rank } p\}$, under the assumptions that $M \in C^1(X^*)$ and $M\nabla c$ is invertible on X^*.

We see that (5) is not affected by the penalty parameter μ. Indeed, μ only affects the acceptance criterion for this direction. Let us remark that in our context a good choice for M is $M(x) = \nabla c(x)^T$. Note that we assume that ∇c has full rank.

Let us remark that the standard method for the update of the penalty parameter in [1] needs to compute $(\nabla c(x)^T M(x)^T)^{-1} c(x)$. In our context, this is computationally expensive. Instead, we consider an update strategy inspired by augmented

Lagrange methods [6]: Set $\mu_{k+1} = \varepsilon_{\text{update}}\,\mu_k$ whenever $\|c(x^{(k)})\| \geq \rho\,\|c(x^{(k+1)})\|$ for $\varepsilon_{\text{update}} > 1$ and $\rho \in (0,1)$. The drawback is that we cannot guarantee any more that μ is increased only a finite number of times, which holds for the method suggested in [1].

A standard convergence result (every limit point of a Newton-like algorithm is a stationary point for P) can be obtained under standard assumptions, see e.g. [1], quite similar to Assumption 3.1, which we use later on.

We recall that by nonlinear elimination of x_E, we refer to solving

$$\nabla_{x_E}\mathcal{L}(x_E, x_L, \lambda) = \nabla_{x_E} J(x_E, x_L) + \nabla_{x_E} c(x_E, x_L)\lambda = 0 \tag{6}$$

for x_E, given x_L and λ, which defines the implicit function $g_E(x_L, \lambda)$. For simplicity, we now write ∇_E instead of ∇_{x_E} and ∇_L instead of ∇_{x_L}. We allow $E = \emptyset$ or $L = \emptyset$, then the related matrices or vectors are empty.

Combination with Nonlinear Elimination For the combination of $P(\cdot\,,\cdot\,;\mu, M)$ with nonlinear elimination, we replace x_E by the elimination g_E and define the functions $\mathfrak{L}(x_L, \lambda) := \mathcal{L}(g_E(x_L, \lambda), x_L, \lambda)$, $C(x_L, \lambda) := c(g_E(x_L, \lambda), x_L)$, and the penalty function

$$\begin{aligned} &\mathcal{P}(x_L, \lambda; \mu, \mathcal{M}) \\ &= \mathfrak{L}(x_L, \lambda) + \frac{\mu}{2}\|C(x_L, \lambda)\|^2 + \frac{1}{2}\|\mathcal{M}(x_L, \lambda)\nabla_L \mathfrak{L}(x_L, \lambda)\|^2, \end{aligned} \tag{7}$$

where $\mu > 0$ and $\mathcal{M} : \mathbb{R}^{n_L} \times \mathbb{R}^p \to \mathbb{R}^{p \times n_L}$. According to the considerations above, we define $\mathcal{M}$ as $\mathcal{M}(x_L, \lambda) := \nabla_L c\big|^T_{(g_E(x_L,\lambda), x_L)}$. By $\nabla_L c\big|_{(g_E(x_L,\lambda),x_L)}$ we mean the evaluation of $\nabla_L c$ at the point $(g_E(x_L, \lambda), x_L)$. By our assumptions on c it follows that $\mathcal{M} \in C^1(X_L^* \times \Lambda^*)$, where $X_L^* \times \Lambda^* := \{(x_L, \lambda) \mid \nabla c\big|_{(g_E(x_L,\lambda),x_L)}$ has rank $p\}$.

The special choice of $\mathcal{M}$ has the advantage of being consistent with the case $E = \emptyset$, $L = \mathbb{R}^n$. In this situation, we have $\mathcal{P}(\cdot\,,\cdot\,;\mu, \mathcal{M}) = P(\cdot\,,\cdot\,;\mu, M)$. The drawback is that for general selections of E, L we cannot guarantee that $\mathcal{M}$ has full rank. In the context of four nonlinear FETI-DP *NL*-1, 2, 3, 4 methods this means that only for *NL*-4 (*NL*-1) the matrix $\mathcal{M}$ has full rank. In *NL*-3 the matrix $\mathcal{M}$ has only zero entries and is empty in *NL*-2.

We cannot expect that all theoretical properties of P are transferred to $\mathcal{P}$. However, the exactness remains valid as well as some other properties.

Theorem 1 ([13])

If $(x_E^*, x_L^*, \lambda^*)$ *is a KKT point of* (3) *and* $(x_L^*, \lambda^*) \in X_L^* \times \Lambda^*$*, then* (x_L^*, λ^*) *is a stationary point of* $\mathcal{P}(\cdot\,,\cdot\,;\mu, \mathcal{M})$ *and*

$$\mathcal{P}(x_L^*, \lambda^*; \mu, \mathcal{M}) = \mathcal{J}(x_L^*, \lambda^*) = J(g_E(x_E^*, \lambda^*), x_L^*) = J(x_E^*, x_L^*).$$

Furthermore, if $\nabla^2_{xx}\mathcal{L}\big|_{(x_E^*, x_L^*, \lambda^*)}$ *is positive definite on* $\ker(\nabla c\big|^T_{(x_E^*, x_L^*)})$*, then there exists a* $\overline{\mu} > 0$ *such that* (x_L^*, λ^*) *is a local minimum of* $\mathcal{P}(\cdot, \cdot\,; \mu, \mathcal{M})$ *for all* $\mu > \overline{\mu}$*.*

Init: $(x_L^{(0)}, \lambda^{(0)}) \in \mathbb{R}^{n_L} \times \in \mathbb{R}^p, \beta, \eta_1, \rho \in (0,1),\ \varepsilon_{\text{update}} > 1,\ \varepsilon_{\text{tol}}, \mu_0, \eta_2, \eta_3, p > 0.$

for $k = 0, 1, \ldots$ until convergence **do**

1. If $\|\nabla \mathcal{P}^{(k)}\|_\infty \le \varepsilon_{\text{tol}}$, *STOP*.
2. (a) Compute $\nabla \mathfrak{L}^{(k)}$ and $\nabla^2 \mathfrak{L}^{(k)}$.

   (b) Solve $\begin{bmatrix} \nabla^2_{LL}\mathfrak{L}^{(k)} & \nabla^2_{L\lambda}\mathfrak{L}^{(k)} \\ \nabla^2_{\lambda L}\mathfrak{L}^{(k)} & \nabla^2_{\lambda\lambda}\mathfrak{L}^{(k)} \end{bmatrix} \begin{bmatrix} \delta x_L^{(k)} \\ \delta \lambda^{(k)} \end{bmatrix} = - \begin{bmatrix} \nabla_L \mathfrak{L}^{(k)} \\ \nabla_\lambda \mathfrak{L}^{(k)} \end{bmatrix}$.

   (c) Set $d^{(k)} := \begin{bmatrix} \delta x_L^{(k)} \\ \delta \lambda^{(k)} \end{bmatrix}$.

   if $\nabla \mathcal{P}^{(k)\,T} d^{(k)} \le -\min\{\eta_1, \eta_2 \|d^{(k)}\|^p\} \|d^{(k)}\|^2$. **then**

   Set $\begin{bmatrix} \delta x_L^{(k)} \\ \delta \lambda^{(k)} \end{bmatrix} := - \begin{bmatrix} \nabla_L \mathcal{P}^{(k)} \\ \nabla_\lambda \mathcal{P}^{(k)} \end{bmatrix}$.

   end
3. Compute the largest number $\alpha_k \in \{\beta^l \mid l = 0, 1, 2, \ldots\}$ such that the Armijo rule

   $$\mathcal{P}(x_L^{(k)} + \alpha_k \delta x_L^{(k)}, \lambda^{(k)} + \alpha_k \delta \lambda^{(k)}; \mu_k, \mathcal{M}) - \mathcal{P}(x_L^{(k)}, \lambda^{(k)}; \mu_k, \mathcal{M})$$
   $$\le \eta_3\, \alpha_k \left(\nabla_L \mathcal{P}^{(k)\,T} \delta x_L^{(k)} + \nabla_\lambda \mathcal{P}^{(k)\,T} \delta \lambda^{(k)} \right)$$

   holds.
4. Set $x_L^{(k+1)} = x_L^{(k)} + \alpha_k\, \delta x_L^{(k)}$ and $\lambda^{(k+1)} = \lambda^{(k)} + \alpha_k\, \delta \lambda^{(k)}$.
5. **if** $\|C(x_L^{(k+1)}, \lambda^{(k+1)})\| \ge \rho\, \|C(x_L^{(k)}, \lambda^{(k)})\|$ **then**

   Set $\mu_{k+1} = \varepsilon_{\text{update}}\, \mu_k$.

   else

   Set $\mu_{k+1} = \mu_k$.

   end

end

Fig. 1: Newton-like algorithm for the computation of stationary points of $\mathcal{P}$.

Since $\mathcal{P}$ is an exact penalty function, we consider the unconstrained minimization problem $\min_{x_L, \lambda} \mathcal{P}(x_L, \lambda; \mu, \mathcal{M})$ to solve (3).

The same arguments, which show that (5) is a Newton-like direction for P, imply that

$$\begin{bmatrix} \delta x_L \\ \delta \lambda \end{bmatrix} = - \begin{bmatrix} \nabla^2_{LL}\mathfrak{L}(x_L,\lambda) & \nabla^2_{L\lambda}\mathfrak{L}(x_L,\lambda) \\ \nabla^2_{\lambda L}\mathfrak{L}(x_L,\lambda) & \nabla^2_{\lambda\lambda}\mathfrak{L}(x_L,\lambda) \end{bmatrix}^{-1} \begin{bmatrix} \nabla_L \mathfrak{L}(x_L,\lambda) \\ \nabla_\lambda \mathfrak{L}(x_L,\lambda) \end{bmatrix} \tag{8}$$

is a Newton-like direction for $\mathcal{P}(\cdot, \cdot; \mu, \mathcal{M})$ at (x_L, λ). Let us remark that the solution of (8) is equivalent to the solution of the standard Lagrange-Newton equation at the point $(g_E(x_L, \lambda)), x_L, \lambda)$.

We outline a Newton-like minimization algorithm for $\mathcal{P}$ in Figure 1, where we define $\nabla \mathfrak{L}^{(k)} := \nabla \mathfrak{L}(x_L^{(k)}, \lambda^{(k)}), \nabla^2 \mathfrak{L}^{(k)} := \nabla^2 \mathfrak{L}(x_L^{(k)}, \lambda^{(k)}), \nabla \mathcal{P}^{(k)} := \nabla \mathcal{P}(x_L^{(k)}, \lambda^{(k)}; \mu_k, \mathcal{M})$ and the blocks $\nabla^2_{LL} \mathfrak{L}^{(k)}$, etc.

For the main convergence result of the algorithm presented in Figure 1 we need the following assumptions:

Assumption 3.1 *The sequence* $\left((x_L^{(k)}, \lambda^{(k)}) \right)_k$ *generated by the Algorithm in Figure 1 is contained in a convex set* $\Omega_L \times \Lambda$ *and the following properties hold:*

(a) The nonlinear elimination $g_E(x_L, \lambda)$ *exists for all* $(x_L, \lambda) \in \Omega_L \times \Lambda$.
(b) The functions J *and* c_i, $i = 1, \ldots, p$ *and their first, second and third derivatives are bounded on* $g_E(\Omega_L \times \Lambda) \times \Omega_L$.

(c) The sequence $(\mu_k)_k$ is bounded.

The boundedness assumption 3.1(b) is needed to ensure that 2.(c) in algorithm of Figure 1 is a generalized angle condition. Furthermore, we need 3.1(c) to prove the main convergence result.

Theorem 2 ([13])

Let Assumption 3.1 be fulfilled. Then every limit point of the sequence $((x_L^{(k)}, \lambda^{(k)}))_k$ generated by the algorithm presented in Figure 1 is a stationary point of $\mathcal{P}$.

4 Numerical Results

We consider a Neo-Hookean benchmark problem using stiff or almost incompressible inclusions embedded in each subdomain. The strain energy density function for the compressible part is given by $J(x) = \frac{\mu}{2}(\mathrm{tr}(F(x)^T F(x)) - 2) - \mu \log(\psi(x)) + \frac{\lambda}{2}(\log(\psi(x)))^2$, where $\psi(x) = \det(F(x))$, $F(x) = \nabla\varphi(x)$, $\varphi(x) = x + u(x)$, $u(x)$ denotes the displacement and μ and λ are the Lamé constants. The nearly incompressible part is given by $J(x) = \frac{\mu}{2}(\mathrm{tr}(\frac{1}{\psi(x)} F(x)^T F(x)) - 2) + \frac{\kappa}{2}(\psi(x) - 1)^2$, where $\kappa = \frac{\lambda(1+\mu)}{3\mu}$, see, e.g. [18]. As material parameters, we use $E = 210$ and $\nu = 0.3$ for the matrix material, $E = 210\,000$ and $\nu = 0.3$ for the stiff inclusions, and, finally, $E = 210$ and $\nu = 0.499$ for the (mildly) almost incompressible inclusions. For the discretization, we use $P2$ elements, which are not stable for the incompressible case.

As Krylov methods, we use GMRES or CG: During the factorizations, it is detected whether $D\widetilde{K}$ is positive definite; in this case, we use CG, otherwise GMRES is used. In Table 1 we see that Newton's method, without globalization, will not converge in the case without inclusions for the body force $(0, -20)^T$, and in the cases with inclusions even for the smaller body force $(0, -10)^T$. In Table 2 we see that, using the algorithm in Figure 1 using the four different nonlinear FETI-DP methods *NL*-1, *NL*-2, *NL*-3, and *NL*-4, we have convergence even for the higher body force $(0, -60)^T$. The cases $(0, -10)^T$ and $(0, -20)^T$ converge as well, but are not presented here. The failure of *NL*-1, 2 to converge despite globalization is due to the fact that we reached the stopping criterion, $\frac{\max\{\|x^{(k+1)} - x^{(k)}\|_\infty, \|\lambda^{(k+1)} - \lambda^{(k)}\|_\infty\}}{\max\{\|x^{(k)}\|_\infty, \|\lambda^{(k)}\|_\infty\}} < 10^{-8}$. This indicates that no sufficient progress is reached, and we abort the simulation since we are limited to machine precision. This example also illustrates that nonlinear elimination can help to achieve convergence.

References

1. Bertsekas, D.P.: Constrained optimization and Lagrange multiplier methods. Computer Science and Applied Mathematics. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London (1982)

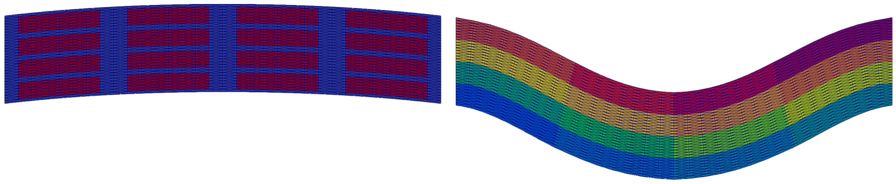

Fig. 2: Model problem with 4×4 subdomains.
Left: Start configuration; *blue*: matrix material, *red*: inclusions.
Right: Deformed state

Table 1: Nonlinear FETI-DP-1 or NL–1; $H/h \approx 8$, see Fig. 2; Newton's method without globalization; the number of Newton iterations is shown; no conv.: $\|\nabla \mathcal{L}^{(k)}\|_\infty \geq 1e5\, \|\nabla \mathcal{L}^{(0)}\|_\infty$

No Globalization					
		body force			
		$f=(0,-10)^T$	$f=(0,-20)^T$	$f=(0,-10)^T$	$f=(0,-10)^T$
d.o.f.	#Sub.	no incl.		comp. incl. ($E = 210\,000$)	incomp. incl. ($\nu = 0.499$)
16 642	16	5	**no conv.**	**no conv.**	**no conv.**
25 922	25	5	**no conv.**	**no conv.**	**no conv.**
37 250	36	5	**no conv.**	**no conv.**	**no conv.**
50 626	49	5	**no conv.**	**no conv.**	**no conv.**
66 050	64	5	**no conv.**	**no conv.**	**no conv.**

Table 2: Nonlinear-FETI-DP-1,2,3,4 or NL–1, 2, 3, 4; body force $f = (0, -60)^T$; $H/h \approx 8$; globalized Newton-like method; number of Newton iterations is shown; stopping criterion: $\|\nabla \mathcal{L}^{(k)}\|_\infty \leq 1e-6\, \|\nabla \mathcal{L}^{(0)}\|_\infty$. Globalization using the algorithm in Figure 1 is used

Using Globalization; see Figure 1													
		body force $f = (0, -60)^T$											
		no incl.				comp. incl. ($E = 210\,000$)				incomp. incl. ($\nu = 0.499$)			
d.o.f.	#Sub.	NL-1	NL-2	NL-3	NL-4	NL-1	NL-2	NL-3	NL-4	NL-1	NL-2	NL-3	NL-4
16 642	16	8	5	5	7	18	6	5	13	-	5	8	19
25 922	25	8	5	6	8	16	6	6	9	-	5	8	21
37 250	36	8	7	5	8	16	7	5	11	-	6	8	21
50 626	49	8	6	6	8	16	7	5	15	-	6	9	22
66 050	64	8	-	7	8	16	7	5	23	-	9	9	23

2. Cai, X.C., Keyes, D.E.: Nonlinearly Preconditioned Inexact Newton Algorithms. SIAM J. Sci. Comput. **24**(1), 183–200 (2002). DOI 10.1137/S106482750037620X
3. Di Pillo, G., Grippo, L.: A new class of augmented lagrangians in nonlinear programming. SIAM J. Control Optim. **17**(5), 618–628 (1979). DOI 10.1137/0317044
4. Dohrmann, C.R.: A preconditioner for substructuring based on constrained energy minimization. SIAM J. Sci. Comput. **25**(1), 246–258 (2003)
5. Farhat, C., Lesoinne, M., Pierson, K.: A Scalable Dual-Primal Domain Decomposition Method. Numer. Linear Algebra Appl. **7**(7–8), 687–714 (2000). DOI 10.1002/1099-1506(200010/12)7:7/8\%3C687::AID-NLA219\%3E3.0.CO;2-S
6. Geiger, C., Kanzow, C.: Theorie und Numerik restringierter Optimierungsaufgaben. Springer (2002). DOI 10.1007/1978-3-642-56004-0

7. Gong, S., Cai, X.C.: A Nonlinear Elimination Preconditioned Inexact Newton Method for Heterogeneous Hyperelasticity. SIAM J. Sci. Comput. **41**(5), S390–S408 (2019). DOI 10.1137/18M1194936
8. Groß, C.: A Unifying Theory for Nonlinear Additively and Multiplicatively Preconditioned Globalization Strategies (2009)
9. Groß, C., Krause, R.: On the globalization of aspin employing trust-region control strategies - convergence analysis and numerical examples. Tech. Rep. 2011-03, Institute of Computational Science, Universita della Svizzera italiana (2011)
10. Klawonn, A., Lanser, M., Rheinbach, O.: Nonlinear FETI–DP and BDDC methods. SIAM J. Sci. Comput. **36**(2), A737–A765 (2014). DOI 10.1137/130920563
11. Klawonn, A., Lanser, M., Rheinbach, O.: FE^2TI: Computational Scale Bridging for Dual-Phase Steels (2015). In: PARCO, pp. 797–806 (2015). DOI 10.3233/978-1-61499-621-7-797
12. Klawonn, A., Lanser, M., Rheinbach, O., Uran, M.: Nonlinear FETI–DP and BDDC methods: a unified framework and parallel results. SIAM J. Sci. Comput. **39**(6), C417–C451 (2017). DOI 10.1137/16M1102495
13. Köhler, S., Rheinbach, O.: Globalization of Nonlinear FETI–DP Methods. In Prep.
14. Nocedal, J., Wright, S.: Numerical Optimization. Springer Science & Business Media (2006). DOI 10.1007/978-0-387-40065-5
15. Pebrel, J., Rey, C., Gosselet, P.: A Nonlinear Dual-Domain Decomposition Method: Application to Structural Problems with Damage. Int. J. Multiscale Comput. Eng. **6**(3), 251–262 (2008). DOI 10.1615/IntJMultCompEng.v6.i3.50
16. Toselli, A., Widlund, O.: Domain decomposition methods—algorithms and theory, *Springer Series in Computational Mathematics*, vol. 34. Springer, Berlin (2005)
17. Ulbrich, M., Ulbrich, S.: Nichtlineare Optimierung. Birkhäuser Basel (2012). DOI 10.1007/978-3-0346-0654-7
18. Uran, M.: High-Performance Computing Two-Scale Finite Element Simulations of a Contact Problem Using Computational Homogenization. Ph.D. thesis, Universität zu Köln (2020)