Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

MARTIN WEISER PETER DEUFLHARD

BODO ERDMANN

Affine conjugate adaptive Newton methods for nonlinear elastomechanics [1]

[1]Supported by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin.

ZIB-Report 04-01 (December 2004)

Affine conjugate adaptive Newton methods for nonlinear elastomechanics †

Martin Weiser, Peter Deuflhard, Bodo Erdmann

December 2004

Abstract

The paper extends affine conjugate Newton methods from convex to nonconvex minimization, with particular emphasis on PDE problems originating from compressible hyperelasticity. Based on well-known schemes from finite dimensional nonlinear optimization, three different algorithmic variants are worked out in a function space setting, which permits an adaptive multilevel finite element implementation. These algorithms are tested on two well-known 3D test problems and a real-life example from surgical operation planning.

MSC 2000: 49M15, 65K10, 74B20, 74G65
Keywords: affine conjugate Newton methods, nonconvex minimization, nonlinear elastomechanics, cranio-maxillofacial surgery

1 Introduction

It is well known, that the equations of compressible linear elastomechanics may lead to unphysical results in the case of large deformations. In particular, regions of local self-penetration are quite common in soft tissue predictions relying on linear elastomechanics. Therefore, realistic models of large stress deformations will require both geometric and material nonlinearity. However, both sources of nonlinearity necessarily lead to nonconvex stored energy functions (see, e.g., Ciarlet [2]). Hence, solving the equations of nonlinear elasticity means minimizing a nonconvex variational functional. In the recently published monograph [5], adaptive affine conjugate Newton methods for convex optimization problems have been elaborated. In the present paper, that paradigm is extended to the nonconvex case.

In Section 2, we introduce three approaches: a Newton-like method (N-lin), a Newton-Truncated-CG method (N-TCG), and two versions of a Newton-Lanczos type method (N-Lanczos A or B). For these approaches, Section 3 presents a common theoretical framework based on some affine conjugate theory, which, in Section 4, is specified to the three approaches. Finally, in Section 5, numerical results for two notorious test problems and a complex real-life problem arising from cranio-maxillofacial surgery are included.

†Supported by the DFG Research Center Matheon "Mathematics for key technologies" in Berlin.

2 Three approaches to non-convex optimization

We consider the minimization problem

$$x = \arg\min f(\xi)$$

with sufficiently smooth but in general nonconvex functional f. A necessary condition for x to be an isolated local minimizer is

$$F(x) := f'(x) = 0 \quad \text{and} \quad F'(x) := f''(x) \text{ positive definite.}$$

An example of such a problem, to be used in Section 5, is the minimization of stored energy in finite strain hyperelasticity, where f may be given, e.g., by an Ogden-type material as

$$f(u) = a\,\mathrm{tr}E + b(\mathrm{tr}E)^2 + c\,\mathrm{tr}E^2 + d\Gamma(\det(I + \nabla u)). \tag{1}$$

The classical quadratic stored energy function of the St. Venant-Kirchhoff material law obtained by setting $a = d = 0$ is convex. With large strains, however, it often leads to unphysical solutions with local interpenetration. As worked out in [2], any hyperelastic material law that prevents local interpenetration is necessarily associated with a nonconvex stored energy function.

In finite-dimensional nonlinear programming, nonconvex optimization has a long tradition that has led to a vast amount of available methods [3, 4, 13]. However, the infinite dimensional PDE setting in the focus here immediately rules out several approaches which are limited to problems of low to medium size dimension.

In *convex* unconstrained optimization, Newton's method

$$F'(x_k)\delta x_k = -F(x_k), \quad x_{k+1} = x_k + \lambda\delta x_k, \quad \lambda > 0 \tag{2}$$

applied to the gradient $F(x) = f'(x)$ of a strictly convex functional $f : X \to \mathbb{R}$ can be used to obtain the solution. Due to $F'(x_k)$ being positive definite, δx_k is always a descent direction. Within an adaptive affine conjugate algorithm [5, 6, 7] the step size λ for the Newton step is chosen such as to minimize an upper bound of f on the one-dimensional subspace

$$\delta X_k = \mathrm{span}\{\delta x_k\}.$$

This basic idea can be extended to the *nonconvex* case as well, when $F'(x_k)$ may be indefinite or even singular. In this case, the Newton direction, even if it exists, need no longer be a descent direction. In view of (1), however, we have a natural decomposition

$$F'(x_k) = M + N(x_k),$$

where $M = F'(0)$ represents the *linear* elastomechanics part and

$$N(x_k) = \mathcal{O}(\|x_k\|) \quad \text{for } \|x_k\| \to 0$$

comprises the nonlinearity. Hence, M induces a convenient energy norm.

For actual computation, the infinite dimensional problem has to be discretized. In order to preserve as much of the original problem's structure as possible, we aim at adaptive multilevel discretizations on a sequence of grids. The Newton-type step is then necessarily chosen from the finite dimensional discretization. In this context, the question of accuracy matching has to be addressed, that is, how to integrate convergence of the Newton-type method and mesh refinement consistently.

The above decomposition inspires three Newton-type algorithms, arranged in increasing order of computational complexity per step. Of course, their relative overall efficiency will additionally depend on their comparative number of required iterations.

Newton-like method (N-lin). In this approach we drop the nonlinear contribution $N(x_k)$ and simply select the positive definite operator M of *linear* elastomechanics to produce a descent direction

$$M\overline{\delta x_k} = -F(x_k), \quad x_{k+1} = x_k + \lambda\overline{\delta x_k}. \tag{3}$$

Again, the step length λ can be chosen adaptively by minimizing an upper bound of f on the one-dimensional subspace

$$\delta X_k = \mathrm{span}\{\overline{\delta x_k}\}.$$

Only the constant stiffness matrix M has to be assembled, which is significantly sparser than $N(x_k)$. A similar approach has been suggested by Glowinski and Le Tallec [10, 12] in the context of augmented Lagrangian methods for incompressible nonlinear elastomechanics. The nonlinearity $N(x_k)$ is only required for few directional derivatives that arise in the adaptive selection of step sizes. The expected convergence rate will typically be linear.

We explicitly want to point out that $M = F'(0)$ has the same affine conjugate transformation behavior as $F'(x)$.

Newton-Truncated-CG (N-TCG). In this approach we use the exact second order information $M+N(x_k)$ and try to cope with its possible indefiniteness directly. Recall that in convex optimization any PCG method will be a candidate of choice within an inexact Newton method; in this case, even arbitrarily poor Galerkin approximations give rise to functional descent. In nonconvex optimization, however, naive application of such an approach will fail as soon as it encounters a direction p^i of negative curvature at the iterate δx_k^i. The idea of using a PCG method nevertheless for nonconvex minimization problems and truncating the iteration in a suitable way dates back to Toint [16] and Steihaug [15]. Several PCG variants in the setting of trust-region methods with empirical adaptation of the trust-region radius are given in Conn, Gould, and Toint [3, Section 7.5]. Inspired by these suggestions, we will define suitable iterates by minimizing an upper bound of f on the two-dimensional subspace

$$\delta X_k = \mathrm{span}\{\delta x_k^i, p^i\}.$$

A particularly appealing feature of this approach is that in regions where the functional is convex, the method reduces to the damped Newton method mentioned in (2) and analyzed in [7].

Newton-Lanczos (N-Lanczos). When directions of negative curvature are encountered early on in the PCG iteration, the Newton-Truncated-CG method essentially reduces to a kind of steepest descent method. Since evaluating the nonlinear stiffness matrix $F'(x)$ is quite expensive, it might be advantageous to exploit r-dimensional Krylov subspaces

$$\delta X_k = \operatorname{span}\{F'(x^k)^j F(x^k) : j = 0, \dots, r-1\} \tag{4}$$

for minimization. GOULD et al. [11] study the Lanczos process for solving nonconvex trust-region subproblems. Their numerical experiments, however, do not indicate a clear performance gain compared to a truncated CG algorithm on a subset of the CUTE test set [1].
In the following Section 3 we will present a unified theoretical framework, which will be specified to the above three approaches in Section 4.

3 Affine conjugate adaptive Newton-type algorithms: a unified derivation

In finite dimensional unconstrained nonconvex optimization, *adaptive* trust region techniques based on constrained *quadratic* models of the functional f are well established — see [3, 4, 13]. In the present paper, we will construct adaptive Newton-type algorithms based on a *cubic* upper bound of the functional f, which arises naturally from affine conjugate theory. In order to follow this line, we perform the four steps given in the recent monograph [5]:

1. derivation of a cubic upper bound for the functional f,
2. construction of a theoretically optimal minimizer within the search subspace δX,
3. identification of computational estimates $[\omega]$ for affine conjugate Lipschitz constants ω, and
4. proof of a bit counting lemma.

For good reasons, the notation to be chosen here will differ slightly from the one in [5]. Throughout the paper, we set the following basic assumptions.

Assumptions 3.1. Let D be an open subset of some Hilbert space X and $M \in \mathcal{L}(X, X^*)$ a symmetric positive definite operator that induces the energy norms $\|x\|_M^2 := \langle x, Mx\rangle$ on X and $\|y\|_{M^{-1}}^2 := \langle y, M^{-1}y\rangle$ on X^*. Here, $\langle\cdot,\cdot\rangle$ denotes the dual pairing between X and its dual X^*. Assume that $f : D \to \mathbb{R}$ is a twice continuously Fréchet-differentiable functional with gradient $F(x) = f'(x)$ and bounded second derivative $F'(x) = f''(x)$ that satisfies the Lipschitz condition

$$\|(F'(x+\delta x) - F'(x))\delta x\|_{M^{-1}} \le \omega \|\delta x\|_M^2 \tag{5}$$

for some constant $\omega < \infty$ and for all $x, \delta x$ such that the convex hull $\operatorname{co}\{x, x+\delta x\}$ of x and $x + \delta x$ is contained in D. Moreover, let $x_0 \in D$ be given and suppose that the level set $\mathcal{L}(x_0) := \{x \in D : f(x) \le f(x_0)\}$ is closed and bounded.

Lemma 3.2. *At each $x \in D$, an upper bound on f is given by*

$$f(x+\delta x) \le f(x) + \langle F(x), \delta x\rangle + \frac{1}{2}\langle \delta x, F'(x)\delta x\rangle + \frac{\omega}{6}\|\delta x\|_M^3 . \tag{6}$$

Proof. In view of the Lipschitz condition (5) we derive an upper bound for functional values by Taylor expansion around x as

$$\begin{aligned} f(x+\delta x) = f(x) + \langle F(x), \delta x\rangle + \frac{1}{2}\langle \delta x, F'(x)\delta x\rangle \\ + \int_0^1 \int_0^1 \langle t\delta x, (F'(x+st\delta x) - F'(x))\delta x\rangle \, ds \, dt \end{aligned}$$

The last term can then be estimated as

$$\begin{aligned} \int_0^1 \int_0^1 \langle t\delta x, &(F'(x+st\delta x) - F'(x))\delta x\rangle \, ds \, dt \\ &= \int_0^1 \int_0^1 t\langle M^{1/2}\delta x, M^{-1/2}(F'(x+st\delta x) - F'(x))\delta x\rangle \, ds \, dt \\ &\le \int_0^1 \int_0^1 t \, \|\delta x\|_M \, \|(F'(x+st\delta x) - F'(x))\delta x\|_{M^{-1}} \, ds \, dt \\ &\le \frac{\omega}{6} \, \|\delta x\|_M^3 \,, \end{aligned}$$

which confirms (6). ☐

On the basis of this lemma, we now study Newton-type algorithms in a unified framework. Starting at some initial iterate $x_0 \in D$, we successively minimize the upper bound (6) of f around the current iterate x_k over some low dimensional subspace δX_k, thus defining

$$\delta x_k := \arg \min_{\delta x \in \delta X_k} \left(\langle F(x_k), \delta x\rangle + \frac{1}{2} \langle F'(x_k)\delta x, \delta x\rangle + \frac{\omega}{6} \|\delta x\|_M^3 \right) \tag{7}$$

Due to the symmetry of the second and third right hand terms w.r.t. $\delta x_k \leftrightarrow -\delta x_k$, we conclude that δx_k is a descent direction, which means that

$$\langle F(x_k), \delta x_k\rangle \le 0. \tag{8}$$

The three Newton-type methods considered in Section 2 above differ only in the choice of the search subspace δX_k.

Functional descent. Given a space δX of search directions δx, we first derive general local descent results by means of the above upper bound for f.

Lemma 3.3. *For simplicity, we drop the iteration index k. Let $\delta X \subset X$ be the low dimensional search subspace and let $\delta x \in \delta X$ be the minimizer as defined by condition* (7)*. Assume that* $\mathrm{co}\{x, x+\delta x\} \subset D$*. Then the functional value reduces as*

$$f(x+\delta x) \le f(x) + \frac{2}{3} \langle F(x), \delta x\rangle + \frac{1}{6} \langle F'(x)\delta x, \delta x\rangle \le f(x),$$

and its derivative satisfies

$$\langle F(x), \delta x\rangle + \langle F'(x)\delta x, \delta x\rangle \le \frac{1}{2}\langle F(x+\delta x), \delta x\rangle \le 0. \tag{9}$$

Proof. First we notice that for $\delta x = 0$ the claims are trivially satisfied. Now assume $\delta x \neq 0$. From (7) we obtain the directional derivative

$$\begin{aligned} 0 &= \left\langle \frac{\partial}{\partial \delta x} \left(\langle F(x), \delta x \rangle + \frac{1}{2} \langle F'(x)\delta x, \delta x \rangle + \frac{\omega}{6} \|\delta x\|_M^3 \right), \delta x \right\rangle \\ &= \left\langle F(x) + F'(x)\delta x + \frac{\omega}{2} \|\delta x\|_M M \delta x, \delta x \right\rangle \\ &= \langle F(x) + F'(x)\delta x, \delta x \rangle + \frac{\omega}{2} \|\delta x\|_M^3 \end{aligned}$$

and hence,

$$\frac{\omega}{2} \|\delta x\|_M^3 = -\big(\langle F(x), \delta x \rangle + \langle F'(x)\delta x, \delta x \rangle \big). \tag{10}$$

Note that (10) is an implicit equation for ω, since δx depends on ω via (7). Inserting (10) into the upper bound (6) yields

$$\begin{aligned} f(x+\delta x) &\le f(x) + \langle F(x), \delta x \rangle + \frac{1}{2} \langle F'(x)\delta x, \delta x \rangle \\ &\qquad - \frac{1}{3} \big(\langle F(x), \delta x \rangle + \langle F'(x)\delta x, \delta x \rangle \big) \\ &= f(x) + \frac{2}{3} \langle F(x), \delta x \rangle + \frac{1}{6} \langle F'(x)\delta x, \delta x \rangle. \end{aligned}$$

With both (8) and $\langle F(x), \delta x \rangle + \langle F'(x)\delta x, \delta x \rangle \le 0$ from (10), we obtain the reduction property

$$\begin{aligned} &f(x) + \frac{2}{3} \langle F(x), \delta x \rangle + \frac{1}{6} \langle F'(x)\delta x, \delta x \rangle \\ &\qquad = f(x) + \frac{3}{6} \langle F(x), \delta x \rangle + \frac{1}{6} \left(\langle F(x), \delta x \rangle + \langle F'(x)\delta x, \delta x \rangle \right) \le f(x). \end{aligned}$$

As for the derivative estimate, we start from

$$F(x+\delta x) = F(x) + F'(x)\delta x + \int_0^1 (F'(x+t\delta x) - F'(x))\delta x \, dx$$

and therefore obtain by (10)

$$|\langle F(x+\delta x) - F(x) - F'(x)\delta x, \delta x \rangle| \le \frac{\omega}{2} \|\delta x\|_M^3 = -\big(\langle F(x), \delta x \rangle + \langle F'(x)\delta x, \delta x \rangle \big),$$

which is (9). □

Computational estimates. We now substitute the unavailable Lipschitz constant ω in actual computation by two different easily computable estimates $[\omega]$, a third order value $[\omega_3] \le \omega$, and a second order value $[\omega_2] \le \omega$, defined as

$$[\omega_3] := \frac{6}{\|\delta x\|_M^3} \left| f(x+\delta x) - f(x) - \langle F(x), \delta x \rangle - \frac{1}{2} \langle F'(x)\delta x, \delta x \rangle \right| \tag{11}$$

$$[\omega_2] := \frac{2}{\|\delta x\|_M^3} |\langle F(x+\delta x) - F(x) - F'(x)\delta x, \delta x \rangle| \tag{12}$$

These estimates $[\omega]$ may possibly be too small. In order to control their relative accuracy, the following bit counting lemma is helpful.

Lemma 3.4. *Let $\delta X \subset X$ be the low dimensional search subspace. Assume that for some descent direction $\delta x \in \delta X$ the following condition holds:*

$$\langle F(x), \delta x\rangle + \frac{1}{2}\langle F'(x)\delta x, \delta x\rangle + \frac{[\omega]}{6}\,\|\delta x\|_M^3 = \min. \tag{13}$$

For $\sigma \le \frac{1}{2}$ and $0 \le \omega - [\omega] \le \sigma[\omega]$, the following reductions of the functional and its derivative are obtained:

$$f(x+\delta x) \le f(x) + \frac{1}{2}\langle F(x), \delta x\rangle - \frac{1-2\sigma}{12}[\omega]\,\|\delta x\|_M^3 \le f(x) \tag{14}$$

and

$$-\frac{2+\sigma}{2}[\omega]\,\|\delta x\|_M^3 \le \langle F(x+\delta x), \delta x\rangle \le \frac{\sigma}{2}[\omega]\,\|\delta x\|_M^3 .$$

Proof. Inserting $\omega \le (1+\sigma)[\omega]$ into (6), we arrive at

$$f(x+\delta x) \le f(x) + \langle F(x), \delta x\rangle + \frac{1}{2}\langle F'(x)\delta x, \delta x\rangle + \frac{1+\sigma}{6}[\omega]\,\|\delta x\|_M^3 . \tag{15}$$

Upon differentiating (13) in the direction δx, we have

$$0 = \langle F(x), \delta x\rangle + \langle F'(x)\delta x, \delta x\rangle + \frac{[\omega]}{2}\,\|\delta x\|_M^3 . \tag{16}$$

Insertion into (15) above yields

$$f(x+\delta x) \le f(x) + \frac{1}{2}\langle F(x), \delta x\rangle - \frac{[\omega]}{4}\,\|\delta x\|_M^3 + \frac{1+\sigma}{6}[\omega]\,\|\delta x\|_M^3 ,$$

which is just the left hand inequality of statement (14). The right hand inequality is evident from (8). Again with (16) the derivative reduction then reads as

$$\begin{aligned}|\langle F(x+\delta x) - F(x) - F'(x)\delta x, \delta x\rangle| &\le \frac{\omega}{2}\|\delta x\|_M^3 \\ &\le \frac{(1+\sigma)[\omega]}{2}\|\delta x\|_M^3 = -(1+\sigma)\big(\langle F(x), \delta x\rangle + \langle F'(x)\delta x, \delta x\rangle\big),\end{aligned}$$

which completes the proof. ☐

Recall from [5] that with such a bit counting lemma, the precise form of a monotonicity test is now prescribed.

Adaptive trust region strategy. In view of (14) we require a sufficient accuracy of the estimate $[\omega]$ and request a functional decrease corresponding to $\sigma \le \frac{1}{3}$. A step δx is accepted, if

$$f(x+\delta x) \le f(x) + \frac{1}{2}\langle F(x), \delta x\rangle - \frac{[\omega_3]}{36}\,\|\delta x\|_M^3 \tag{17}$$

holds. For small δx, e.g. close to the solution, both this acceptance test and the computation of $[\omega_3]$ according to (11) become numerically unstable, which has to be monitored carefully. In this case, the following substitute is to be used together with $[\omega_2]$ as defined by (12):

$$\langle F(x+\delta x), \delta x\rangle \le \frac{[\omega_2]}{6}\,\|\delta x\|_M^3 . \tag{18}$$

Lemma 3.5. *Whenever the monotonicity tests* (17) *or* (18)*, respectively, fail, the new Lipschitz estimate* $[\omega]_{\mathrm{new}}$ *computed by* (11) *or* (12)*, respectively, satisfies*

$$[\omega]_{new} > \frac{4}{3}[\omega]_{\mathrm{old}}.$$

Proof. We start with the third order estimate. From (16) we obtain

$$\begin{aligned}
[\omega_3]_{\mathrm{new}} &= \frac{6}{\|\delta x\|_M^3} \left| f(x+\delta x) - f(x) - \langle F(x), \delta x\rangle - \frac{1}{2}\langle F'(x)\delta x, \delta x\rangle \right| \\
&> \frac{6}{\|\delta x\|_M^3} \left(\frac{1}{2}\langle F(x), \delta x\rangle - \frac{[\omega_3]}{36} \|\delta x\|_M^3 - \langle F(x), \delta x\rangle - \frac{1}{2}\langle F'(x)\delta x, \delta x\rangle \right) \\
&= \frac{6}{\|\delta x\|_M^3} \left(-\frac{[\omega_3]}{36} \|\delta x\|_M^3 + \frac{[\omega]}{4} \|\delta x\|_M^3 \right) \\
&= \frac{4}{3}[\omega_3].
\end{aligned}$$

As for the second order estimate, again (16) yields

$$\begin{aligned}
[\omega_2]_{\mathrm{new}} &= \frac{2}{\|\delta x\|_M^3} |\langle F(x+\delta x), \delta x\rangle - \langle F(x), \delta x\rangle - \langle F'(x)\delta x, \delta x\rangle| \\
&= \frac{2}{\|\delta x\|_M^3} \left| \langle F(x+\delta x), \delta x\rangle + \frac{[\omega]}{2} \|\delta x\|_M^3 \right| \\
&> \frac{2}{\|\delta x\|_M^3} \left(\frac{[\omega]}{6} \|\delta x\|_M^3 + \frac{[\omega_2]}{2} \|\delta x\|_M^3 \right) \\
&= \frac{4}{3}[\omega_2].
\end{aligned}$$

□

Whenever the monotonicity tests (17) or (18), respectively, fail, the updated estimate $[\omega]$ is increased according to Lemma 3.5, until at last $\omega \leq (1+\sigma)[\omega]$ holds. At this point, Lemma 3.4 guarantees that the coresponding step δx_k passes the monotonicity test. The consequence is that after finitely many iterations of this scheme a functional reduction is obtained. Following the notation used for one-dimensional search subspaces δX_k, we refer to this iteration as stepsize reduction loop. In practice, we have rarely rarely observed more than one reduction. In passing we note that we also deliberately use a heuristic increase of the computational estimate $[\omega]$, whenever the computed corrections leave the definition domain D.

Termination criterion. In contrast to convex optimization, where the energy norm of the Newton correction can be used to formulate the termination criterion

$$\langle \delta x_k, M\delta x_k\rangle < \mathrm{ETOL}^2, \tag{19}$$

the iterate x_k may be far away from the solution even if the energy norm of the Newton correction is small. For x_* to be a local minimum it is necessary that $F'(x_*)$ be positive semidefinite. Therefore we will additionally require that

$F'(x_k)$ is positive semidefinite on δX_k for any iterate x_k to be accepted as approximate solution.

A stricter alternative is to require that when $x_k \to x_*$, then $F'(x_*)$ has to be positive definite. Under the assumption that the ordinary Newton method converges towards x_*, the distance $\|x_* - x_k\|_M$ is bounded by

$$\begin{aligned}\omega \|x_* - x_k\|_M &\le \sum_{j=k}^{\infty} \omega \left\|F'(x_j)^{-1} F(x_j)\right\|_M \\ &= \sum_{j=k}^{\infty} h_j \le \sum_{j=k}^{\infty} \left(\frac{h_k}{2}\right)^{j-k} h_k \le \frac{2h_k}{2-h_k}\end{aligned}$$

where $h_k = \omega \left\|F'(x_k)^{-1} F(x_k)\right\|_M$. With a somewhat more general Lipschitz condition than (5) we may then arrive at

$$\begin{aligned}\langle \xi, F'(x_*)\xi\rangle &= \langle \xi, F'(x_k)\xi\rangle + \langle \xi, (F'(x_*) - F'(x_k))\xi\rangle \\ &\ge \langle \xi, F'(x_k)\xi\rangle - \|\xi\|_M \|(F'(x_*) - F'(x_k))\xi\|_{M^{-1}} \\ &\ge \langle \xi, F'(x_k)\xi\rangle - \|\xi\|_M^2 \,\omega \|x_* - x_k\|_M .\end{aligned}$$

Hence, $F'(x_*)$ is positive definite if

$$\inf_{\xi \ne 0} \frac{\langle \xi, F'(x_k)\xi\rangle}{\|\xi\|_M^2} > \frac{2h_k}{2-h_k}.$$

As a computationally available approximation of this criterion we may impose the lower dimensional requirement

$$\min_{\xi \in \delta X_k, \xi \ne 0} \frac{\langle \xi, F'(x_k)\xi\rangle}{\|\xi\|_M} > \frac{2[h_k]}{2-[h_k]} \quad \text{with } [h_k] = [\omega] \left\|\delta x^k\right\|_M$$

in addition to (19) for any iterate x_k to be accepted as local minimizer.

4 Specification to the three approaches

We are now ready to apply the above general framework to each of the three Newton-type methods introduced in Section 2.

4.1 Newton-like algorithm

With the Newton-like correction given by $M\overline{\delta x_k} = -F(x_k)$, the one-dimensional search subspace $\delta X_k = \operatorname{span}(\overline{\delta x_k})$ permits a representation of the above setting in terms of a stepsize λ, which can be computed explicitly by minimizing (13) as

$$\lambda = -\frac{2\langle F(x_k), \overline{\delta x}_k\rangle}{\epsilon + \sqrt{\epsilon^2 - 2[\omega] \left\|\overline{\delta x}_k\right\|_M^3 \langle F(x_k), \overline{\delta x}_k\rangle}} \tag{20}$$

with $\epsilon = \langle \overline{\delta x}, F'(x)\overline{\delta x}_k\rangle$. The resulting algorithm reads as follows.

Algorithm 4.1.

1: *solve* $M\overline{\delta x}_k = -F(x_k)$
2: *compute* λ_k *from* (20)
set $\delta x_k = \lambda_k \overline{\delta x}_k$
check monotonicity test (17)
update $[\omega_k]$ *according to* (11)
if the monotonicity test (17) *has been violated: goto 2*
set $x_{k+1} = x_k + \lambda_k \overline{\delta x_k}$, $[\omega_{k+1}] = [\omega_k]$
if (19) *is satisfied and* $\langle \overline{\delta x_k}, F'(x_k)\overline{\delta x_k}\rangle \geq 0$*: stop*
increase k *and goto 1*

Accuracy matching. In PDE applications, the exact computation of the N-lin correction $\overline{\delta x_k}$ by solving (3) is in general infeasible, since a discretization error remains. Even in finite dimensional problems, truncation errors of iterative solvers are unavoidable, if only the dimension is sufficiently large. In both cases, we can only compute an *inexact* correction $\widehat{\delta x_k}$ by solving

$$M\widehat{\delta x_k} = -F(x_k) + r_k$$

up to some inner residual r_k. Now the question arises, how large an inner residual, may it stem from discretization errors or iteration errors, we may accept without spoiling the convergence of the N-lin algorithm. At least, $\widehat{\delta x_k}$ has to be a descent direction. We suggest to require

$$\langle F(x_k), \widehat{\delta x_k}\rangle \leq (1-\delta_k)\langle F(x_k), \overline{\delta x_k}\rangle \leq 0$$

for some $\delta_k \in]0,1[$, which is equivalent to

$$\|r_k\|_{M^{-1}} \leq \sqrt{\delta_k}\ \|F(x_k)\|_{M^{-1}}\,.$$

If M is a good approximation to $F'(x_*)$ at the solution point, choosing δ_k close to 0 leads to faster convergence. If, on the other hand, M differs significantly from $F'(x_*)$, a more accurate approximation $\widehat{\delta x_k}$ of $\overline{\delta x_k}$ cannot be expected to improve the linear convergence of the method in the same way. Since we aim at nonlinear elasticity including large deformations, we expect the latter case to be quite common, and recommend to just choose $\delta_k = \frac{1}{2}$.

4.2 Newton-Truncated-CG algorithm

The second approach constructs the search subspace δX_k from a PCG algorithm applied to the linear system $\delta F'(x_k)\delta x_k = -F(x_k)$. If at inner iterate δx_k^i in the PCG method a search direction p_k^i with nonpositive curvature is encountered, i.e. if $\langle p_k^i, F'(x_k)p_k^i\rangle \leq 0$, then the PCG iteration is terminated and the search subspace is defined as $\delta X_k = \mathrm{span}\{\delta x_k^i, p_k^i\}$. Any numerical method for solving the low dimensional cubic minimization problem (13) can be used to compute $\delta x_k \in \delta X_k$. The computational complexity of this subproblem is actually negligible compared to the PCG iteration.

If no such direction is encountered until the PCG iteration achieves the required accuracy at some iterate δx_k^j, the search subspace is defined by the inexact Newton direction as $\delta X_k = \mathrm{span}\{\delta x_k^j\}$.

The resulting algorithm then reads as follows.

Algorithm 4.2.

1: *compute* $\delta X_k = PCG(F'(x_k), -F(x_k))$
2: *solve* (13) *for* δx_k
check the monotonicity test (17)
update $[\omega]_k$ *according to* (11)
if the monotonicity test was violated: goto 2
set $x_{k+1} = x_k + \delta x_k$, $[\omega]_{k+1} = [\omega]_k$
if (19) *is satisfied and* $\langle \xi, F'(x_k)\xi \rangle \geq 0$ *for all* $\xi \in \delta X_k$: *stop*
increase k *and goto 1*

In case no nonconvex search direction is encountered, which naturally occurs in the neighborhood of a local minimum with $F'(x_*)$ positive definite, the method reduces to the affine conjugate Newton method for convex minimization [5, 7] and inherits its local convergence properties. As for the Newton-like method above, the step $\delta x_k = \lambda_k \delta x_k^j$ can be expressed in terms of a step size λ given by (20).

Accuracy matching. If the PCG method does not produce a search direction p_k with negative curvature of the functional, we are left with the decision when to terminate the iteration. Since in this case the functional is convex at x_k to our current best knowledge, a situation that is to be expected close to the solution, we rely on the accuracy matching for convex problems worked out in [6, 7]. Minimizing the information gain per unit work, we arrive at different choices for δ_k for finite dimensional problems and infinite dimensional problems, where a mesh refinement loop takes the place of the PCG iteration.

For *finite* dimensional problems we therefore choose a relative tolerance

$$\delta_k = \min\left([\omega] \left\|\delta x_k^j\right\|_M, 10^{-2}\right)$$

for the PCG Algorithm and recover the locally quadratic convergence of Newton's method.

In the *infinite* dimensional setting, we accept a mesh refinement level as soon as (i) the PCG method on this refinement level generates directions of negative curvature of f, or (ii) the error estimator indicates a relative error of $\delta < 1$. Ultimately, we end up with the linear convergence of inexact Newton methods.

4.3 Newton-Lanczos algorithm

The third approach aims at minimizing the cubic model (6) over the complete Krylov space (4) of sufficiently large dimension r. An M-orthonormal basis $v_0, \ldots, v_{r-1}$ of δX_k can be obtained by preconditioning the Lanczos process with M, which transforms (13) into the simpler low-dimensional and sparse problem

$$\begin{aligned} h &= \arg\min_{h \in \mathbb{R}^k} \gamma_0 \langle h, e_1 \rangle + \frac{1}{2}\langle h, Th \rangle + \frac{[\omega]}{6}\|h\|_2^2, \\ \delta x_k &= Vh, \end{aligned} \tag{21}$$

where T is tridiagonal and $V = [v_0, \ldots, v_{r-1}]$. The Newton-type correction defined by (21) may be assumed to provide a better functional decrease than the truncated CG solution for two reasons: (i) a larger search subspace is considered in case nonconvex directions are encountered, and (ii) the cubic term affects the choice of δx_k from a much higher dimensional space than it does in the truncated CG case.

Numerical results for a similar Lanczos approach have been reported in [11] in the context of trust-region methods. Therein the number of iterations tended to be smaller compared to the truncated CG approach. Due to its higher computational effort, however, it provided no clear advantage over the truncated CG version in overall computing time.

Since in nonlinear elastomechanics the assembly of $F'(x)$ tends to dominate the computational cost, a moderate reduction of iteration numbers could be sufficient to decrease the overall computational cost in this application. However, using the stiffness matrix M of linear elastomechanics as a preconditioner, it is necessary to solve a static linear problem in each step of the Lanczos method, which is prohibitively expensive. That is why, in practice, a suitable approximation $\hat{M} \approx M$ will be used, and one of the auxiliary functionals

$$J_A = \langle F(x), \delta x\rangle + \frac{1}{2}\langle \delta x, F'(x)\delta x\rangle + \frac{\omega}{6}\,\|\delta x\|_{\hat{M}}^3$$

or

$$J_B = \langle F(x), \delta x\rangle + \frac{1}{2}\langle \delta x, F'(x)\delta x\rangle + \frac{\omega}{6}\,\|\delta x\|_{M}^3$$

will be minimized.

Variant A preserves the simple tridiagonal structure of (21), but generates only a suboptimal search direction w.r.t. (7). As a consequence, a stepsize has to be chosen as in the N-lin and N-TCG settings. Together with the search direction, the number of Newton-Lanczos steps can be expected to depend directly on the quality of the preconditioner $\hat{M}$.

Variant B preserves the original minimization property (7), but produces dense matrices, the dimension of which is the number of Lanczos iterations. Thus, solving the low dimensional minimization problem (7) is significantly more expensive, but the number of Newton-Lanczos steps may be expected to be essentially independent of the choice of the preconditioner $\hat{M}$.

Accuracy matching. There is no natural exit point in case directions of negative curvature are encountered. In fact, the procedure is designed to proceed past such iterates. However, if the functional is nonconvex at the current iterate x_k, the solution cannot be assumed to be close, even in case $\|F(x_k)\|_{M^{-1}}$ is small. Thus, a highly accurate determination of δx_k is unlikely to improve the overall performance significantly. As a heuristic, we suggest to monitor the eigenvalues σ_i of the projected matrix T. If all of them are positive, we impose the same termination criterion as in the N-TCG. Otherwise we proceed with the Lanczos iteration until an allocated budget of computing time t_k is exhausted. In order to balance possible progress and computing time, we suggest to set t_k to about 20% of the time needed for assembling $F'(x_k)$ and $F(x_k)$. As for mesh refinement, we employ the same strategy as outlined for the N-TCG method.

5 Numerical comparisons

This section is devoted to a numerical evaluation of the algorithms developed up to now. In view of a multilevel approximation of PDE problems on a sequence of grids, we do not employ the CUTE testset of problems with fixed finite dimension. We choose examples from elastomechanics with both geometric and material nonlinearity. Hence we switch to the usual notation in elastomechanics and substitute the variable x by the displacement u.

The algorithms are tested on three different geometries. In all cases we consider a hyperelastic Ogden material [14] as discussed in [2, §4.10], with stored energy function

$$f(u) = a\,\mathrm{tr}E + b(\mathrm{tr}E)^2 + c\,\mathrm{tr}E^2 + d\Gamma(\det(I+\nabla u)), \quad F(u) = f'(u), \tag{22}$$

where

$$\begin{aligned} a &= -d\Gamma'(1) & b &= \frac{1}{2}(\lambda - d(\Gamma'(1)+\Gamma''(1))) \\ c &= \mu + d\Gamma'(1) & d &> 0, \qquad \Gamma(s) = s^2 - \ln s. \end{aligned}$$

Here, $E = \frac{1}{2}(\nabla u^T + \nabla u + \nabla u^T \nabla u)$ is the Green-St. Venant strain tensor and $\lambda = 7.76 \cdot 10^5$ and $\mu = 8.62 \cdot 10^4$ are the Lamé constants corresponding to Young's modulus $2.5 \cdot 10^5$ and Poisson ratio 0.45. The most appealing feature of this material law is that near the undeformed reference state it is a second order approximation to the linear St. Venant-Kirchhoff material to λ and μ [2, Thm. 4.10-2], which is recovered asymptotically for $d \to 0$. We do refer to the linear St. Venant-Kirchhoff material as the case $d = 0$, even though the two material laws coincide only for orientation preserving deformations.

The natural choice for the energy metric is given by the always positive definite stiffness matrix $M = F'(0)$ of linear elastomechanics. It exhibits the same transformation properties as $F'(u)$ and therefore conserves the affine conjugacy of Newton's method. A particularly convenient feature of this choice is that M is usually sparser than $F'(u)$ and has to be assembled only once on each grid refinement level.

In comparing the algorithms we give iteration numbers and general complexity considerations rather than actual CPU times, which are highly subject to implementation details.

Example 1: Ubiquitous cube. The ubiquitous test example is defined on the cube $\Omega = [-1,1]^3$. We impose Dirichlet boundary conditions of $u = 0$ on the bottom face $[-1,1]^2 \times \{-1\}$ and $u = (0,0,-0.8)^T$ on the top face $[-1,1]^2 \times \{1\}$ and no boundary conditions on the remaining four sides.

As initial iterate for the methods we choose the deformation given by linear elastomechanics with the same material constants λ and μ. The computation is performed on fixed uniform grids with Cartesian structure. The obtained solutions are shown in Figure 1. As it is well known, large deformations can lead to local self-penetration of the material ($\det(I+\nabla u) < 0$) unless the stored energy tends to infinity for $\det(I + \nabla u) \to 0$. Thus the case $d = 0$ leads to unphysical solutions with inverted elements. In contrast, the logarithmic term prevents interpenetration and leads to a solution that is quite close to the one given by linear elastomechanics. Iteration numbers for the linear material law

Figure 1: **Example 1.** *Left*: linear elastomechanics. *Center*: geometric nonlinearity with linear St. Venant-Kirchhoff material ($d = 0$). *Right*: geometric nonlinearity with nonlinear OGDEN material ($d = 10^5$).

with only geometric nonlinearity and for the completely nonlinear model are given in Table 1. As expected, the former case is more difficult due to the larger distance between initial iterate and solution, as well as the existence of a large number of local minima.

material	nodes	N-lin	N-TCG	N-Lanczos (A/B)
$d = 0$	729	265	27	24/fail
	4913	401	91	54/fail
$d = 10^5$	729	21	7	7/7
	4913	34	7	11/8

Table 1: **Example 1.** Number of iterations for different uniform mesh sizes. Termination criterion (19) with relative tolerance $\mathrm{ETOL} = 10^{-3} \left\| u^k \right\|_M$.

Example 2: Hexagonal 3D beam. A more flexible structure with still trivial reference geometry is given by a hexagonal prism with aspect ratio 1:10. Again the top face is moved downwards about 40% of the length of the beam. The solution of linear elasticity shown in Figure 2 (left) is chosen as initial iter-

Figure 2: **Example 2.** *Left*: linear elastomechanics. *Center*: geometric nonlinearity with linear St. Venant-Kirchhoff material. *Right*: geometric nonlinearity with OGDEN material ($d = 10^5$).

material	nodes	N-lin	N-TCG	N-Lanczos (A/B)
$d = 0$	779	(3000)	69	81/fail
	4142		59	242/fail
$d = 10^5$	779	(2500)	44	93/23
	4142		56	1556/29

Table 2: **Example 2.** Number of iterations for different uniform mesh sizes. Termination criterion (19) with relative tolerance $\mathrm{ETOL} = 10^{-3} \left\| u^k \right\|_M$. Values in parentheses estimated.

ate. A similar example with quadratic base has been considered by Glowinski and LeTallec [10, 12]. Here, the stable non-symmetric solutions of nonlinear elastomechanics (with or without material nonlinearity) deviate significantly from the unstable symmetric solution, which, in turn, is close to the solution of linear elastomechanics.

Numerical solutions for different settings are shown in Figure 2. Iteration numbers are given in Table 2. The result of an adaptive computation starting at a regular mesh with 147 nodes and resulting in 4236 nodes is shown in Figure 3.

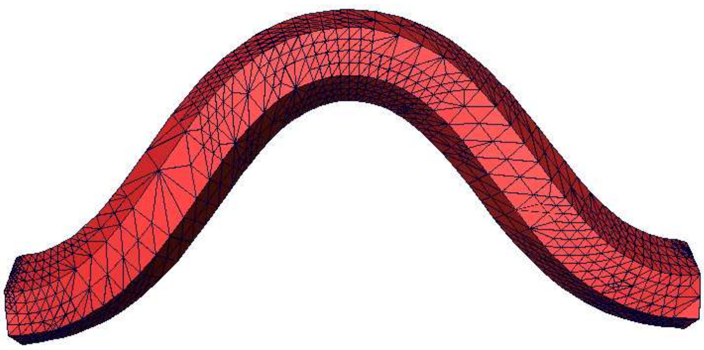

Figure 3: **Example 2.** Result of mesh adaptation.

Example 3: Cranio-maxillofacial surgery. In cranio-maxillofacial surgery, the post-operative appearance of the patient is of vital interest. Therefore, a prediction of the facial soft tissue deformations induced by surgical bone displacement is a decisive tool for therapy planning. In cooperation with surgeons, a therapy planning tool that allows to cut and move bone parts of virtual patient models has been developed by the ZIB working group *Computer Assisted Surgery* inside the visualization package Amira (Zachow et al. [17, 18]). First predictions of the post-operative apprearence of patients have been computed by Gladilin [8] predominantly using linear elastomechanics. However, the bone displacements usually lead to large deformations of the soft tissue in certain areas, so that linear elastomechanics cannot be expected to yield reliable results. In view of this property, first steps towards nonlinear models have already been undertaken in [9]. With the algorithms presented here, a robust and reliable

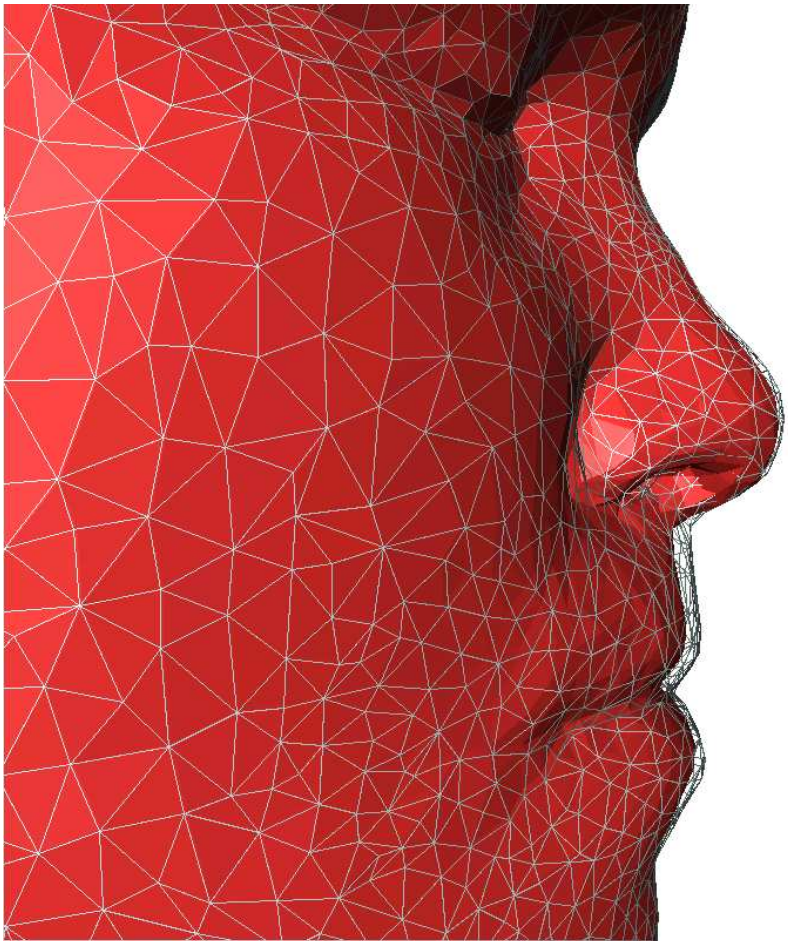

Figure 4: **Example 3.** Displacement of soft tissue due to bone movement. *Opaque surface*: with geometric nonlinearity. *Mesh lines*: linear elastomechanics.

computation of stable configurations is now possible.

Out of a large set of problems (`http://www.zib.de/visual/projects/cas/cas-gallery.en.html`), we pick one test case where the maxilla is advanced by 5mm. Due to geometrical constraints we chose a St. Venant-Kirchhoff material law together with geometric nonlinearity. The result obtained by the N-TCG algorithm within 17 steps is compared in Figure 4 with the solution of linear elastomechanics.

Remark 5.1. When using constitutive laws that are defined only on the orientation-preserving subset of all possible configurations, e.g. Ogden materials, it is necessary to start the Newton-type algorithms from an admissible initial deformation u_0. Depending on the complexity of the geometry and the boundary conditions, such a starting point may be difficult to obtain. In simple situations, explicit interpolation of boundary conditions (Example 1) or the solution of linear elastomechanics (Example 2) can provide an orientation-preserving initial iterate. More complex cases require some kind of homotopy, e.g. incremental load or a shift of the logarithmic barrier in (22).

In biomedical applications with their complex geometries, it may even happen that the prescribed boundary conditions *enforce* a self-penetrating deformation. This is indeed the case for all examples from cranio-maxillofacial surgery tested so far, which is the reason why we restricted Example 3 to geometric nonlinearity and used a St. Venant-Kirchhoff material. In order to overcome this kind of difficulty, a closer cooperation between geometric modeling, grid generation, and PDE solution will be necessary. This topic is certainly beyond the scope of the present paper and on the agenda for further investigation.

Comparative performance. It is evident from Tables 1 and 2, that the N-lin method needs much more iterations especially in more complex situations. However, since only $F(u^k)$ but not $F'(u^k)$ has to be assembled in every step, each step can be significantly cheaper to compute than in the N-TCG or the N-Lanczos methods. In our code, the CPU time ratio for computing $F(u^k)$ and $F'(u^k)$, respectively, is about 1:5. Neglecting any overhead such as solving linear systems, error estimation and evaluation of $f(u)$ and directional derivatives $\langle F'(u^k)\delta u^k, \delta u^k\rangle$, we can roughly estimate that N-lin needs more CPU time than N-TCG, if it requires more than six times as many iterations. Note that this estimate is quite conservative in favor of N-lin. Taking this correction factor into account, we see that N-lin is on par with N-TCG in Example 1 and much slower in the more challenging Example 2.

The comparison between N-TCG and N-Lanczos/A comes out less clear. The computational complexity per iteration is about the same. However, the iteration count of N-Lanczos/A depends highly on the effectivity of the preconditioner, which may lead to extremely high iteration numbers as encountered in Table 2. Here the mesh dependence of the SSOR preconditioner is inherited by the N-Lanczos/A method. Additionally, the implementation complexity of the N-Lanczos methods is significantly higher than that of N-TCG.

In contrast to that, the N-Lanczos/B version shows the lowest number of iterations. However, its computational complexity for solving the system is significantly higher than that of N-TCG or N-Lanczos/A, which easily outweighs the lower number of iterations reported in Table 2.

Conclusion

In this paper we have developed three different approaches to nonconvex problems in finite strain elasticity: Newton-like, Newton-Truncated-CG, and Newton-Lanczos.

Summarizing, for both theoretical and computational complexity reasons, the Newton-Truncated-CG method comes out as our preferred candidate.

Acknowledgment The authors are extremely grateful to S. Zachow (ZIB), who provided the head model. They also wish to thank J. Liesen (TU Berlin and DFG Research Center MATHEON) for helpful discussions on Krylov subspace methods.

References

[1] I. Bongartz, A.R. Conn, N. Gould, and Ph.L. Toint. Cute: Constrained and unconstrained testing environment. *ACM Trans. Math. Softw.*, 21(1):123–160, 1995.

[2] P.G. Ciarlet. *Mathematical elasticity. Volume I: Three-dimensional elasticity*, volume 20 of *Studies in Mathematics and its Applications*. North-Holland, 1988.

[3] A.R. Conn, N.I.M. Gould, and P.L. Toint. *Trust-Region Methods*. SIAM, 2000.

[4] J.E. Dennis and R.B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. Prentice-Hall, 1983.

[5] P. Deuflhard. *Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms*, volume 35 of *Computational Mathematics*. Springer, 2004.

[6] P. Deuflhard and M. Weiser. Local inexact Newton multilevel FEM for nonlinear elliptic problems. In M.-O. Bristeau, G. Etgen, W. Fitzgibbon, J.-L. Lions, J. Periaux, and M. Wheeler, editors, *Computational science for the 21st century*, pages 129–138. Wiley, 1997.

[7] P. Deuflhard and M. Weiser. Global inexact Newton multilevel FEM for nonlinear elliptic problems. In W. Hackbusch and G. Wittum, editors, *Multigrid Methods V*, Lecture Notes in Computational Science and Engineering, pages 71–89. Springer, 1998.

[8] E. Gladilin. *Biomechanical Modeling of Soft Tissue and Facial Expressions for Craniofacial Surgery Planning*. PhD thesis, Free University Berlin, 2003.

[9] E. Gladilin, S. Zachow, P. Deuflhard, and H.-C. Hege. Adaptive nonlinear elastic fem for realistic soft tissue prediction in craniofacial surgery simulations. In S. K. Mun, editor, *Proc. SPIE Medical Imaging 2002: Visualization, Image-Guided Procedures*, volume 4681, pages 1–8, 2002.

[10] R. Glowinski and P. Le Tallec. *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, volume 9 of *Studies in Applied Mathematics*. SIAM, 1989.

[11] N.I.M. Gould, S. Lucidi, M. Roma, and Ph.L. Toint. Solving the trust-region subproblem using the Lanczos method. *SIAM J. Optim.*, 9(2):504–525, 1999.

[12] P. Le Tallec. Numerical methods for nonlinear three-dimensional elasticity. In P.G. Ciarlet and J.L. Lions, editors, *Handbook of Numerical Analysis*, volume III, pages 465–622. North-Holland, Elsevier, 1994.

[13] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer, 1999.

[14] R.W. Ogden. Large deformation isotropic elasticity: on the correlation of theory and experiment for incompressible rubber-like solids. *Proc. Roy. Soc. London*, A328:567–583, 1972.

[15] T. Steihaug. The conjugate gradient method and trust regions in large scale optimization. *SIAM J. Numer. Anal.*, 20(3):626–637, 1983.

[16] P.L. Toint. Towards an efficient sparsity exploiting Newton method for minimization. In I.S. Duff, editor, *Sparse Matrices and Their Use*, pages 57–88. Academic Press, 1981.

[17] S. Zachow, E. Gladilin, H.-F. Zeilhofer, and R. Sader. Improved 3D osteotomy planning in cranio-maxillofacial surgery. In *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI 2001)*, Lecture Notes in Computer Science, pages 473–481, Utrecht, 2001. Springer.

[18] S. Zachow, T. Hierl, and B. Erdmann. A quantitative evaluation of 3D soft tissue prediction in maxillofacial surgery planning. In *Proc. 3. Jahrestagung der Deutschen Gesellschaft für Computer- und Roboter-assistierte Chirurgie e.V.*, München, 2004.