SIAM Conference on Computational Science and Engineering

Improved Newton Linearization for L^1-Norm-Type Minimization with Application to Viscoplastic Fluid Solvers

Johann Rudi — Argonne National Laboratory
Georg Stadler — New York University
Omar Ghattas — The University of Texas at Austin

Argonne National Laboratory

Summary

- Our challenge: nonlinear applications modelled by optimization problems with singularities in the Hessian due to a L^1-norm type term in the objective.
- Hessian exhibits a null space upon linearization with Newton's method that is problematic for convergence.
- Applications include inverse problems with total variation regularization as well as viscoplastic flows like Earth's mantle convection.
- We analyze issues with the standard Newton linearization through numerical experiments and theoretically in an abstract setting.
- We propose an improved linearization based on a perturbation of an otherwise implicitly assumed equality constraint.
- We achieve robust and fast Newton convergence independent of the discretization. We further show advantageous theoretical properties of the new linearization.

Introduction in 1 Dimension

Minimization problem Given constants $a > 0$, $b \in \mathbb{R}$, and $0 \le \theta \le 1$,

$$\text{find minimizer} \quad x^* \in \mathbb{R} \quad \text{of} \quad \min_{x\in\mathbb{R}} J(x) := \frac{a}{1+\theta}|x|_\epsilon^{1+\theta} - bx$$

Denote $|x|_\epsilon := (|x|^2 + \epsilon^2)^{1/2}$. Note that the parameter θ controls the nonlinearity such that Newton's method *converges well* for $\theta = 1$; and Newton's method *converges poorly* for $\theta \to 0$. The gradient and Hessian of $J(x)$ are

$$g(x) := J'(x) = \frac{ax}{|x|_\epsilon^{1-\theta}} - b, \quad H(x) := J''(x) = \frac{a}{|x|_\epsilon^{1-\theta}}\left(1 - (1-\theta)\frac{x^2}{|x|_\epsilon^2}\right).$$

Standard Newton linearization One Newton step (simplified, e.g., without backtracking line search) is

$$\text{solve for } \hat{x} \text{ in} \quad H(x)\,\hat{x} = -g(x), \quad \text{update} \quad x \leftarrow x + \hat{x},$$

therefore, we need to solve the linearized equation

$$\frac{a}{|x|_\epsilon^{1-\theta}}\left(1 - (1-\theta)\frac{x^2}{|x|_\epsilon^2}\right)\hat{x} = -\left(\frac{ax}{|x|_\epsilon^{1-\theta}} - b\right) \qquad \text{(standard Newton step)}$$

Perturbed Newton linearization Define for $x, y \in \mathbb{R}$

$$\text{model error:} \quad E(x,y) := \frac{x}{|x|_\epsilon^{1-\theta}} - y,$$
$$\text{model perturbation:} \quad D(x,y) := x - |x|_\epsilon^{1-\theta} y.$$

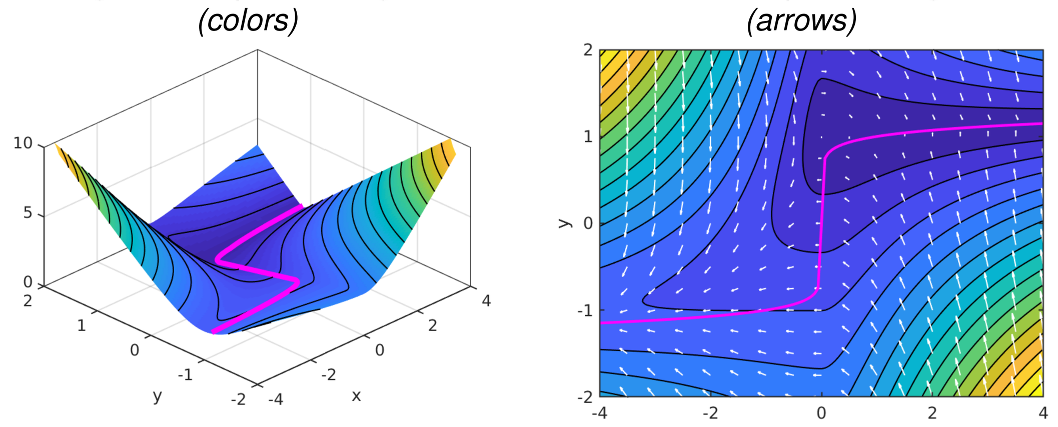
The perturbed Newton linearization for $\hat{x}$ and $\hat{y}$ assumes $D(x,y) \neq 0$, then

$$a\hat{y} = -(ay - b),$$
$$\left(1 - (1-\theta)\frac{xy}{|x|_\epsilon^{1+\theta}}\right)\hat{x} - |x|_\epsilon^{1-\theta}\hat{y} = -\left(x - |x|_\epsilon^{1-\theta} y\right).$$

Substituting for $\hat{y}$ gives the perturbed and reduced Newton linearization for $\hat{x}$

$$\frac{a}{|x|_\epsilon^{1-\theta}}\left(1 - (1-\theta)\frac{xy}{|x|_\epsilon^{1+\theta}}\right)\hat{x} = -\left(\frac{ax}{|x|_\epsilon^{1-\theta}} - b\right) \qquad \text{(improved Newton step)}$$

Magnitude of perturbed gradient *(colors)* — Directions of perturbed gradient *(arrows)*



The *magenta curve* in both plots represents the (absolute value of) the unperturbed gradient $g(x)$ in (x,y)-space, meaning $E(x,y) = 0$. Introducing the perturbation, $D(x,y) \neq 0$, relaxes the rapid changes of the unperturbed $g(x)$ around $x = 0$.

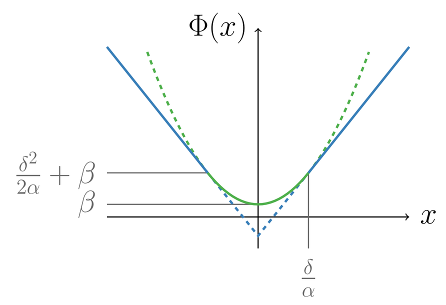
Abstract Derivation of Perturbed Newton Linearizations

Minimization problem Given dimensions $d \in \mathbb{N}$ and $n \in \{1, d, d\times d\}$, domain $\Omega \subseteq \mathbb{R}^d$, and a linear and bounded functional $F : L^2(\Omega)^n \to \mathbb{R}$,

$$\text{find minimizer} \quad U^* : \Omega \to \mathbb{R}^n \quad \text{of} \quad \min_U J(U) := \int_\Omega \Phi(U) - F(U)$$

using a generalization of the *Huber loss:*

$$\Phi(U) := \begin{cases} \frac{1}{2}\alpha|U|^2 + \beta, & \alpha|U| \le \delta, \\ \delta|U| - \frac{1}{2}\frac{\delta^2}{\alpha} + \beta, & \text{otherwise,} \end{cases}$$



with parameters $\alpha, \delta > 0$ and $\beta \in \mathbb{R}$.

Standard Newton linearization The gradient and Hessian are the 1st- and 2nd-order variations of J:

$$g(U)\tilde{U} := \int_\Omega \left(\chi\alpha + (1-\chi)\frac{\delta}{|U|}\right)\langle U, \tilde{U}\rangle - F(\tilde{U}), \quad \chi := \begin{cases} 1, & \alpha|U| \le \delta, \\ 0, & \text{o.w.,} \end{cases}$$
$$\left(H(U)\hat{U}, \tilde{U}\right) := \int_\Omega \left(\chi\alpha + (1-\chi)\frac{\delta}{|U|}\left(\mathbf{I} - \frac{U\otimes U}{|U|^2}\right)\right)\left\langle \hat{U}, \tilde{U}\right\rangle.$$

The standard Newton linearization requires to solve for Newton step $\hat{U}$ in

$$\int_\Omega \left(\chi\alpha + (1-\chi)\frac{\delta}{|U|}\left(\mathbf{I} - \frac{U\otimes U}{|U|^2}\right)\right)\left\langle \hat{U}, \tilde{U}\right\rangle = -g(U)\tilde{U}$$

The outer product term is computationally challenging because:

- Coefficient $\left(\mathbf{I} - \frac{U\otimes U}{|U|^2}\right)$ in Hessian represents an orthogonal projector.
- Hessian has a zero eigenvalue associated to eigenvector $(1-\chi)U$.

Perturbed Newton linearization Define for $U, S \in L^2(\Omega)^n$

$$\text{model error:} \quad E(U,S) := \frac{U}{|U|} - S,$$
$$\text{model perturbation:} \quad D(U,S) := U - |U|\,S.$$

We augment the previous gradient by a model perturbation, $D(U,S) \neq 0$,

$$g(U,S)\tilde{U} = \int_\Omega \langle \chi\alpha U + (1-\chi)\delta S, \tilde{U}\rangle - F(\tilde{U}),$$
$$\left(D(U,S), \tilde{S}\right) = \int_\Omega \langle U - |U|\,S, \tilde{S}\rangle,$$

to get the perturbed Newton linearization for step $(\hat{U}, \hat{S})$

$$\int_\Omega \left\langle \chi\alpha\hat{U} + (1-\chi)\delta\hat{S}, \tilde{U}\right\rangle = -g(U,S)\tilde{U},$$
$$\int_\Omega \left\langle \left(\mathbf{I} - \frac{U\otimes S}{|U|}\right)\hat{U} - |U|\,\hat{S}, \tilde{S}\right\rangle = -\left(D(U,S), \tilde{S}\right).$$

Here, the dual step has the explicit expression

$$\hat{S} = \frac{U}{|U|} - S + \frac{1}{|U|}\left(\mathbf{I} - \frac{U\otimes S}{|U|}\right)\hat{U}.$$

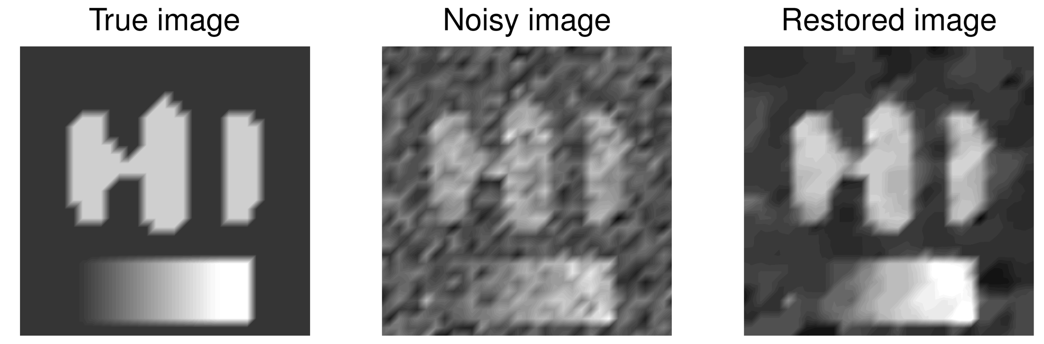
Therefore, substitution leads to the perturbed and reduced Newton linearization for $\hat{U}$, where only the outer product term of the Hessian has changed compared to the standard linearization:

$$\int_\Omega \left(\chi\alpha + (1-\chi)\frac{\delta}{|U|}\left(\mathbf{I} - \frac{U\otimes S}{|U|}\right)\right)\left\langle \hat{U}, \tilde{U}\right\rangle = -g(U)\tilde{U}$$

Improvements gained from perturbation

- Perturbation results in model error dependent regularization:
$$\mathbf{I} - \frac{U\otimes S}{|U|} = \mathbf{I} - \frac{U\otimes U}{|U|^2} + \frac{U\otimes E(U,S)}{|U|}.$$
- Perturbed linearization acts as a nonlinear preconditioner far from the solution while enabling fast Newton convergence close to solution, since
$$\mathbf{I} - \frac{U\otimes S}{|U|} \to \mathbf{I} - \frac{U\otimes U}{|U|^2} \quad \text{as} \quad E(U,S) \to 0.$$
- The explicit expression of $\hat{S}$ enables a simple and computationally cheap update of the dual variable S after computing step $\hat{U}$:
$$S \leftarrow S + \hat{S} = \frac{U}{|U|} + \frac{1}{|U|}\left(\mathbf{I} - \frac{U\otimes S}{|U|}\right)\hat{U}.$$

Image Restoration with Total Variation (TV)



A *true* image is generated *(left)*, to which Gaussian noise with standard deviation $\sigma = 0.2$ is added *(middle)*. The noisy image serves as input data to the image restoration problem with the result shown *right*.

Minimization problem Given blurry and noisy image data, $d : (0,1)^2 \to \mathbb{R}$, the blurring operator, $B : L^2(\Omega) \to L^2(\Omega)$, $\Omega = (0,1)^2$, and the total variation operator

$$\Phi(\nabla u) = \begin{cases} \frac{1}{2\epsilon}|\nabla u|^2, & |\nabla u| \le \epsilon, \\ |\nabla u| - \frac{\epsilon}{2}, & \text{otherwise,} \end{cases}$$

with Huber parameters $\alpha = 1/\epsilon$, $\delta = 1$, $\beta = 0$, where $0 < \epsilon \ll 1$; consider

$$\text{find minimizer} \quad u^* : (0,1)^2 \to \mathbb{R} \quad \text{of} \quad \min_u J(u) := \frac{1}{2}\|Bu - d\|_{L^2}^2 + \gamma\,\Phi(\nabla u)$$

Note: The regularization with $\Phi(\nabla u)$ has the advantage of preserving edges in the restored image u^*. The parameters $\epsilon > 0$ and $\gamma > 0$ regularize the TV operator and the image restoration problem, respectively.

Standard Newton linearization The 1st- and 2nd-order variations of $\Phi(\nabla u)$ are

$$\delta_u[\Phi(\nabla u)](\tilde{u}) = \int_\Omega \left(\frac{\chi}{\epsilon} + \frac{(1-\chi)}{|\nabla u|}\right)\nabla u \cdot \nabla\tilde{u},$$
$$\delta_u\delta_u[\Phi(\nabla u)](\tilde{u})(\hat{u}) = \int_\Omega \left(\frac{\chi}{\epsilon} + \frac{(1-\chi)}{|\nabla u|}\left(\mathbf{I} - \frac{\nabla u\otimes\nabla u}{|\nabla u|^2}\right)\right)\nabla\hat{u}\cdot\nabla\tilde{u}.$$

The standard Newton linearization requires to solve for Newton step $\hat{u}$ in

$$-\nabla\cdot\left[\gamma\left(\frac{\chi}{\epsilon} + \frac{(1-\chi)}{|\nabla u|}\left(\mathbf{I} - \frac{\nabla u\otimes\nabla u}{|\nabla u|^2}\right)\right)\nabla\hat{u}\right] + B^*B\hat{u} = -\gamma\delta_u[\Phi(\nabla u)] - B^*(Bu - d)$$

The left-hand side constitutes a Poisson operator with an anisotropic 2nd-order tensor coefficient.

Perturbed Newton linearization Define for $u \in L^2(\Omega)$, $S \in L^\infty(\Omega)$, $\|S\|_{L^\infty} \le 1$

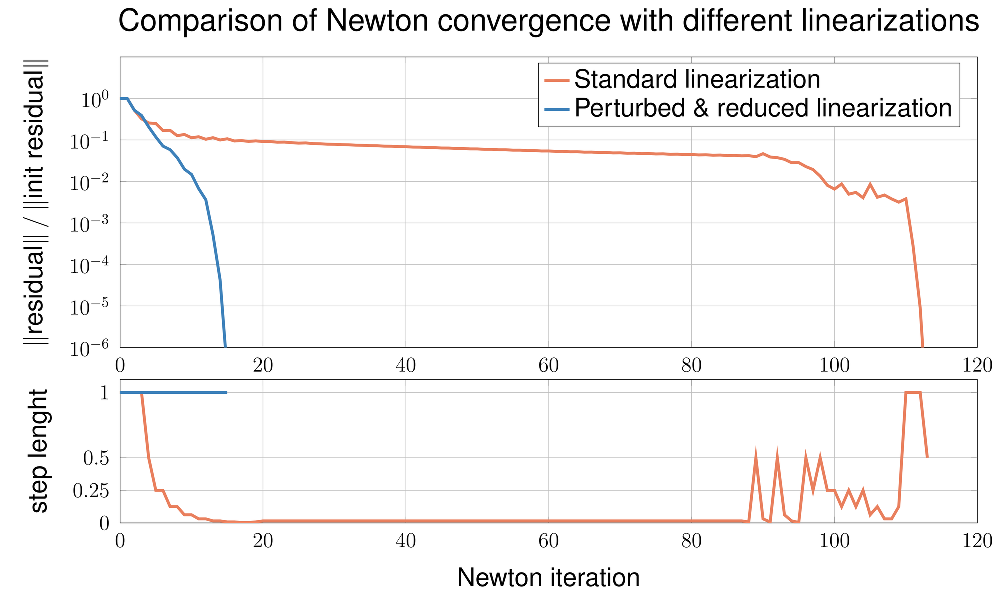
$$\text{model error:} \quad E(u,S) := \frac{\nabla u}{|\nabla u|} - S,$$
$$\text{model perturbation:} \quad D(u,S) := \nabla u - |\nabla u|\,S.$$

The perturbed and reduced Newton linearization for $\hat{u}$ is the system

$$-\nabla\cdot\left[\gamma\left(\frac{\chi}{\epsilon} + \frac{(1-\chi)}{|\nabla u|}\left(\mathbf{I} - \frac{\nabla u\otimes S}{|\nabla u|}\right)\right)\nabla\hat{u}\right] + B^*B\hat{u} = -\gamma\delta_u[\Phi(\nabla u)] - B^*(Bu - d)$$

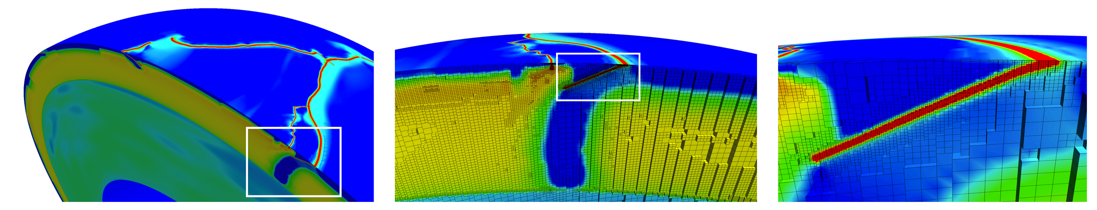
Numerical experiments Comparison of Newton convergence with standard and perturbed linearizations using parameters $\epsilon = 0.01$, $\gamma = 0.002$, and added Gaussian noise to the *true* image with standard deviation $\sigma = 0.2$.

Comparison of Newton convergence with different linearizations



Viscoplastic Stokes Flow

Motivated by rheology of Earth's mantle convection with plate tectonics:



Effective viscosity field *(colors)* and viscoplastic dynamic weakening of lithosphere in hinge zone *(right zoom)*. (Visualization by L. Alisic)

Minimization problem Given viscosity $\mu > 0$, yield strength $\tau_{\text{yield}} > 0$, the 2nd invariant of the strain rate tensor, $\dot{\varepsilon}_{\text{II}} := \frac{1}{\sqrt{2}}|\nabla_s \boldsymbol{u}|$ with $\nabla_s\boldsymbol{u} := \frac{1}{2}(\nabla\boldsymbol{u} + \nabla\boldsymbol{u}^\top)$, and

$$\Phi(2\dot{\varepsilon}_{\text{II}}) = \begin{cases} \frac{\mu}{2}(2\dot{\varepsilon}_{\text{II}})^2 + \frac{1}{2}\frac{\tau_{\text{yield}}^2}{\mu}, & \mu(2\dot{\varepsilon}_{\text{II}}) \le \tau_{\text{yield}}, \\ \tau_{\text{yield}}(2\dot{\varepsilon}_{\text{II}}), & \text{otherwise,} \end{cases}$$

with Huber parameters $\alpha = \mu$, $\delta = \tau_{\text{yield}}$, $\beta = \frac{1}{2}\frac{\delta^2}{\alpha} = \frac{1}{2}\frac{\tau_{\text{yield}}^2}{\mu}$; consider

$$\text{find minimizer} \quad \boldsymbol{u}^* : (0,1)^3 \to \mathbb{R}^3 \quad \text{of} \quad \min_{\boldsymbol{u}} J(\boldsymbol{u}) := \int_\Omega \Phi(\boldsymbol{u}) - \boldsymbol{f}\cdot\boldsymbol{u} \quad \text{s.t.} \quad \nabla\cdot\boldsymbol{u} = 0$$

Standard Newton linearization Solve for Newton step $(\hat{\boldsymbol{u}}, \hat{p})$ in

$$-\nabla\cdot\left[2\left(\chi\mu + (1-\chi)\frac{\tau_{\text{yield}}}{\sqrt{2}\,|\nabla_s\boldsymbol{u}|}\left(\mathbf{I} - \frac{\nabla_s\boldsymbol{u}\otimes\nabla_s\boldsymbol{u}}{|\nabla_s\boldsymbol{u}|^2}\right)\right)\nabla_s\hat{\boldsymbol{u}}\right] + \nabla\hat{p} = -\boldsymbol{r}_{\text{mom}}$$
$$-\nabla\cdot\hat{\boldsymbol{u}} = -r_{\text{mass}}$$

The residuals of the nonlinear Stokes momentum and mass equations appear on the right-hand side. What plays the role of viscosity in the Newton step is an anisotropic 4th-order tensor.

Perturbed Newton linearization Define for $\boldsymbol{u} \in H^1(\Omega)^d$ and $\boldsymbol{S} \in L^\infty(\Omega)^{d\times d}$, $\|\boldsymbol{S}\|_{L^\infty} \le 1$

$$\text{model error:} \quad E(\boldsymbol{u},\boldsymbol{S}) := \frac{\nabla_s\boldsymbol{u}}{|\nabla_s\boldsymbol{u}|} - \boldsymbol{S},$$
$$\text{model perturbation:} \quad D(\boldsymbol{u},\boldsymbol{S}) := \nabla_s\boldsymbol{u} - |\nabla_s\boldsymbol{u}|\,\boldsymbol{S}.$$

The perturbed and reduced Newton linearization of the momentum equation is

$$-\nabla\cdot\left[2\left(\chi\mu + (1-\chi)\frac{\tau_{\text{yield}}}{\sqrt{2}\,|\nabla_s\boldsymbol{u}|}\left(\mathbf{I} - \frac{\nabla_s\boldsymbol{u}\otimes\boldsymbol{S}}{|\nabla_s\boldsymbol{u}|}\right)\right)\nabla_s\hat{\boldsymbol{u}}\right] + \nabla\hat{p} = -\boldsymbol{r}_{\text{mom}}$$

Numerical experiments Comparison of Newton convergence with standard and perturbed linearizations using a model problem with a viscosity that incorporates low-viscosity plumes in a high-viscosity background medium,

$$\mu(\boldsymbol{x}) := (\mu_{\max} - \mu_{\min})(\chi_n(\boldsymbol{x}) - 1) + \mu_{\min}, \quad \boldsymbol{x} \in (0,1)^3,$$

where $\chi_n(\boldsymbol{x}) \in [1,2]$ are C^∞ indicator functions, which accumulate n plumes via products of Gaussians.

Yielding volume	**Mesh level** ℓ	**Standard Newton** It. Newton	#backtr.	It. GMRES	**Perturbed Newton** It. Newton	#backtr.	It. GMRES
∼45%	4	33	20	1469	10	0	379
∼45%	5	36	25	2255	12	0	664
∼45%	6	57	49	4255	13	0	876
∼65%	4	29	21	1559	18	10	965
∼65%	5	37	26	2464	17	9	1245
∼65%	6	48	39	3892	20	9	1707
∼90%	4	35	25	1505	19	11	872
∼90%	5	40	32	2147	21	11	1267
∼90%	6	32	21	2312	23	11	1811

References

[1] T. F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM Journal on Scientific Computing*, 20(6):1964–1977, 1999.

[2] M. Hintermüller and G. Stadler. An infeasible primal-dual algorithm for total variation-based inf-convolution-type image restoration. *SIAM Journal on Scientific Computing*, 28(1):1–23, 2006.

[3] M. Spiegelman, D. A. May, and C. R. Wilson. On the solvability of incompressible Stokes with viscoplastic rheologies in geodynamics. *Geochemistry, Geophysics, Geosystems*, 17(6):2213–2238, 2016.