Numerical Stability of Tangents and Adjoints of Implicit Functions

Uwe Naumann[1][0000−0002−7518−5922]

Department of Computer Science, RWTH Aachen University, 52056 Aachen, Germany naumann@stce.rwth-aachen.de
www.stce.rwth-aachen.de

Abstract. We investigate errors in tangents and adjoints of implicit functions resulting from errors in the primal solution due to approximations computed by a numerical solver.
Adjoints of systems of linear equations turn out to be unconditionally numerically stable. Tangents of systems of linear equations can become instable as well as both tangents and adjoints of systems of nonlinear equations, which extends to optima of convex unconstrained objectives. Sufficient conditions for numerical stability are derived.

Keywords: algorithmic differentiation · implicit function.

1 Introduction

We consider twice differentiable implicit functions

$$F : \mathbb{R}^m \to \mathbb{R}^n : \mathbf{p} \mapsto \mathbf{x} = F(\mathbf{p}) \tag{1}$$

defined by the roots of residuals

$$R : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n : (\mathbf{x}, \mathbf{p}) \mapsto R(\mathbf{x}, \mathbf{p}) \; . \tag{2}$$

R is referred to as the primal residual as opposed to tangent and adjoint residuals to be considered later. Primal roots of the residual satisfying

$$R(\mathbf{x}, \mathbf{p}) = 0 \tag{3}$$

are assumed to be approximated by numerical solvers

$$S : \mathbb{R}^m \to \mathbb{R}^n : \mathbf{p} \mapsto \mathbf{x} + \Delta\mathbf{x} = S(\mathbf{p})$$

with an absolute error $\Delta\mathbf{x}$ yielding a relative error $\delta\mathbf{x}$ of norm

$$\|\delta\mathbf{x}\| = \frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{\|S(\mathbf{p}) - F(\mathbf{p})\|}{\|F(\mathbf{p})\|} \; .$$

We investigate (relative) errors in corresponding tangents

$$\dot{\mathbf{x}} = \dot{F}(\mathbf{x}, \dot{\mathbf{p}}) \equiv \frac{dF}{d\mathbf{p}} \cdot \dot{\mathbf{p}} \tag{4}$$

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_17

and adjoints

$$\bar{\mathbf{p}} = \bar{F}(\mathbf{x}, \bar{\mathbf{x}}) \equiv \frac{dF}{d\mathbf{p}}^T \cdot \bar{\mathbf{x}} \tag{5}$$

due to $\Delta\mathbf{x}$. *Algorithmic* tangents and adjoints result from the application of algorithmic differentiation (AD) [3, 4] to the solver S. *Symbolic* tangents and adjoints can be derived at the solution of Equation (3) in terms of tangents and adjoints of the residual [2, 5]. AD of the solver can thus be avoided which typically results in a considerably lower computational complexity.

2 Prerequisites

We perform standard first-order error analysis. For a given absolute error $\Delta\mathbf{p}$ in the input of a function F the absolute error in the result is estimated as

$$\Delta\mathbf{x} \approx \frac{dF}{d\mathbf{p}} \cdot \Delta\mathbf{p} \; . \tag{6}$$

Equation (1) is differentiated with respect to $\mathbf{p}$ in the direction of the absolute error $\Delta\mathbf{p}$. From the Taylor series expansion of

$$\mathbf{x} + \Delta\mathbf{x} = \mathbf{x} + \frac{dF}{d\mathbf{p}} \cdot \Delta\mathbf{p} + O(\|\Delta\mathbf{p}\|^2)$$

it follows that negligence of the remainder within a neighborhood of $\mathbf{x}$ containing $\Delta\mathbf{x}$ is reasonable for $\|\Delta\mathbf{p}\| \to 0$ and assuming convergence of the Taylor series to the correct function value. For linear F we get $\Delta\mathbf{x} = \frac{dF}{d\mathbf{p}} \cdot \Delta\mathbf{p}$ due to the vanishing remainder.

Tangents and adjoints of Equation (1) can be expressed as matrix equations over derivatives of the residual. The fundamental operations involved are scalar multiplications and additions, outer vector products, matrix-vector products and solutions of systems of linear equations.

It is well-known that scalar multiplication $y = x_1 \cdot x_2$ is numerically stable with relative error $|\delta y| = |\delta x_1| + |\delta x_2|$. A similar result holds for scalar division. It generalizes naturally to element-wise multiplication and division of vectors, matrices, and higher-order tensors as well as to the outer product of two vectors.

Scalar addition $y = x_1 + x_2$ on the other hand is known to be numerically unstable $|\delta y| \frac{|\Delta x_1 + \Delta x_2|}{|x_1 + x_2|} \to \infty$ for $\Delta x_1 \neq -\Delta x_2$ and $x_1 \to -x_2$. A similar result holds for scalar subtraction.

Numerical instability of scalar addition prevents unconditional numerical stability of inner vector products as well as matrix-vector/matrix products and solutions of systems of linear equations. Sufficient conditions for numerical stability need to be formulated.

The relative error of a matrix-vector product $\mathbf{x} = A \cdot \mathbf{b}$ for $A \in \mathbb{R}^{n \times n}$ and $\mathbf{x}, \mathbf{b} \in \mathbb{R}^n$ is easily shown to be equal to

$$\|\delta\mathbf{x}\| \approx \kappa(A) \cdot (\|\delta A\| + \|\delta\mathbf{b}\|) \; . \tag{7}$$

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_17

Depending on the magnitude of the condition number $\kappa(A) \equiv \|A^{-1}\| \cdot \|A\|$ of A the relative error of the matrix-vector product can suffer from a potentially dramatic amplification of the relative errors in the arguments.

We take a closer look at the derivation of a similar result for systems of linear equations $A \cdot \mathbf{x} = \mathbf{b}$ for $A \in \mathbb{R}^{n \times n}$ and $\mathbf{x}, \mathbf{b} \in \mathbb{R}^n$. Differentiation in the direction of non-vanishing absolute errors $\Delta A \in \mathbb{R}^{n \times n}$ and $\Delta \mathbf{x}, \Delta \mathbf{b} \in \mathbb{R}^n$ yields

$$\Delta \mathbf{x} = A^{-1} \cdot (\Delta \mathbf{b} - \Delta A \cdot \mathbf{x})$$

and hence the first-order error estimate

$$\begin{aligned}
\frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|} &= \frac{\|A^{-1} \cdot (\Delta \mathbf{b} - \Delta A \cdot \mathbf{x})\|}{\|\mathbf{x}\|} \\
&\leq \frac{\|A^{-1} \cdot \Delta \mathbf{b}\|}{\|\mathbf{x}\|} + \frac{\|A^{-1} \cdot \Delta A \cdot \mathbf{x}\|}{\|\mathbf{x}\|} \\
&\leq \frac{\|A^{-1}\| \cdot \|\Delta \mathbf{b}\|}{\|\mathbf{x}\|} + \frac{\|A^{-1} \cdot \Delta A \cdot \mathbf{x}\|}{\|\mathbf{x}\|} = \frac{\|A\| \cdot \|A^{-1}\| \cdot \|\Delta \mathbf{b}\|}{\|A\| \cdot \|\mathbf{x}\|} + \frac{\|A^{-1} \cdot \Delta A \cdot \mathbf{x}\|}{\|\mathbf{x}\|} \\
&\leq \kappa(A) \cdot \frac{\|\Delta \mathbf{b}\|}{\|A \cdot \mathbf{x}\|} + \frac{\|A^{-1} \cdot \Delta A \cdot \mathbf{x}\|}{\|\mathbf{x}\|} = \kappa(A) \cdot \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|} + \frac{\|A^{-1} \cdot \Delta A \cdot \mathbf{x}\|}{\|\mathbf{x}\|} \\
&\leq \kappa(A) \cdot \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|} + \frac{\|A^{-1}\| \cdot \|\Delta A\| \cdot \|\mathbf{x}\|}{\|\mathbf{x}\|} = \kappa(A) \cdot \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|} + \|A^{-1}\| \cdot \|\Delta A\| \\
&= \kappa(A) \cdot \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|} + \frac{\|A\| \cdot \|A^{-1}\| \cdot \|\Delta A\|}{\|A\|} = \kappa(A) \cdot \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|} + \kappa(A) \cdot \frac{\|\Delta A\|}{\|A\|} \\
&= \kappa(A) \cdot \left(\frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|} + \frac{\|\Delta A\|}{\|A\|} \right) .
\end{aligned}$$

As for matrix-vector products we get

$$\|\delta \mathbf{x}\| \approx \kappa(A) \cdot (\|\delta A\| + \|\delta \mathbf{b}\|) . \tag{8}$$

Again, a low condition number of A is sufficient for numerical stability.

3 Errors in Tangents and Adjoints of Implicit Functions

Differentiation of Equation (3) with respect to $\mathbf{p}$ yields

$$\frac{\partial R}{\partial \mathbf{x}} \cdot \frac{d\mathbf{x}}{d\mathbf{p}} + \frac{\partial R}{\partial \mathbf{p}} = R_{\mathbf{x}} \cdot \frac{d\mathbf{x}}{d\mathbf{p}} + R_{\mathbf{p}} = 0 , \tag{9}$$

where ∂ denotes partial differentiation. Multiplication with $\dot{\mathbf{p}}$ from the right yields the tangent residual

$$R_{\mathbf{x}} \cdot \frac{d\mathbf{x}}{d\mathbf{p}} \cdot \dot{\mathbf{p}} + R_{\mathbf{p}} \cdot \dot{\mathbf{p}} = R_{\mathbf{x}} \cdot \dot{\mathbf{x}} + R_{\mathbf{p}} \cdot \dot{\mathbf{p}} = 0 . \tag{10}$$

The tangent $\dot{\mathbf{x}}$ can be computed as the solution of the system of linear equations

$$R_{\mathbf{x}} \cdot \dot{\mathbf{x}} = -R_{\mathbf{p}} \cdot \dot{\mathbf{p}} .$$

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_17

The right-hand side is obtained by a single evaluation of the tangent residual. Tangents in the directions of the Cartesian basis of $\mathbb{R}^n$ yields $R_{\mathbf{x}}$. Potential sparsity can and should be exploited [1]. An error $\Delta\mathbf{x}$ in the primal solution yields a corresponding error in the tangent for $R_{\mathbf{x}} = R_{\mathbf{x}}(\mathbf{x})$ and/or $R_{\mathbf{p}} = R_{\mathbf{p}}(\mathbf{x})$.

From Equation (9) it follows that for regular $R_{\mathbf{x}}$

$$\frac{d\mathbf{x}}{d\mathbf{p}} = -R_{\mathbf{x}}^{-1} \cdot R_{\mathbf{p}} \; .$$

Transposition of the latter followed by multiplication with $\bar{\mathbf{x}}$ from the right yields

$$\bar{\mathbf{p}} = \frac{d\mathbf{x}}{d\mathbf{p}}^T \cdot \bar{\mathbf{x}} = -R_{\mathbf{p}}^T \cdot R_{\mathbf{x}}^{-T} \cdot \bar{\mathbf{x}} \; . \tag{11}$$

The adjoint $\bar{\mathbf{p}}$ can be computed as the solution of the system of linear equations

$$R_{\mathbf{x}}^T \cdot \mathbf{z} = -\bar{\mathbf{x}}$$

followed by the evaluation of the adjoint residual yielding

$$\bar{\mathbf{p}} = R_{\mathbf{p}}^T \cdot \mathbf{z}$$

. Again, an error $\Delta\mathbf{x}$ in the primal solution yields a corresponding error in the adjoint.

3.1 Systems of Linear Equations

The tangent of the solution of the primal system of linear equations

$$A \cdot \mathbf{x} = \mathbf{b} \tag{12}$$

is defined as $\dot{\mathbf{x}} = \dot{\mathbf{x}}_A + \dot{\mathbf{x}}_{\mathbf{b}}$, where

$$A \cdot \dot{\mathbf{x}}_{\mathbf{b}} = \dot{\mathbf{b}} \tag{13}$$

and

$$A \cdot \dot{\mathbf{x}}_A = -\dot{A} \cdot \mathbf{x} \tag{14}$$

[2]. An error $\Delta\mathbf{x}$ in the primal solution which, for example, might result from the use of an indirect solver yields an erroneous tangent

$$\dot{\mathbf{x}} + \Delta\dot{\mathbf{x}} = (\dot{\mathbf{x}}_A + \Delta\dot{\mathbf{x}}_A) + (\dot{\mathbf{x}}_{\mathbf{b}} + \Delta\dot{\mathbf{x}}_{\mathbf{b}}) \; .$$

Application of Equation (8) to Equation (13) yields

$$\|\delta\dot{\mathbf{x}}_{\mathbf{b}}\| \approx \kappa(A) \cdot (\|\delta A\| + \|\delta\dot{\mathbf{b}}\|) \; .$$

Independence of $\dot{\mathbf{x}}_{\mathbf{b}}$ from $\mathbf{x}$ (and hence from $\Delta\mathbf{x}$) implies $\delta\dot{\mathbf{x}}_{\mathbf{b}} = 0$ for error-free A and $\dot{\mathbf{b}}$, that is $\Delta\dot{\mathbf{x}} = \Delta\dot{\mathbf{x}}_A$, respectively $\delta\dot{\mathbf{x}} = \delta\dot{\mathbf{x}}_A$. Let $\mathbf{c} = -\dot{A} \cdot \mathbf{x}$. With Equation (7) it follows that

$$\|\delta\mathbf{c}\| \approx \kappa(\dot{A}) \cdot \|\delta\mathbf{x}\|$$

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_17

as $\delta\dot{A}=0$ Moreover, application of Equation (8) to $A\cdot\dot{\mathbf{x}}_A=\mathbf{c}$ yields

$$\|\delta\dot{\mathbf{x}}_A\|\approx\kappa(A)\cdot\|\delta\mathbf{c}\|\ .$$

Consequently,

$$\|\delta\dot{\mathbf{x}}_A\|\approx\kappa(A)\cdot\kappa(\dot{A})\cdot\|\delta\mathbf{x}\|\ . \tag{15}$$

Low condition numbers of both A and $\dot{A}$ ensure numerical stability of tangent systems of linear equations.

The adjoint of the primal linear system in Equation (12) is defined as

$$A^T\cdot\bar{\mathbf{b}}=\bar{\mathbf{x}} \tag{16}$$

and

$$\bar{A}=-\bar{\mathbf{b}}\cdot\mathbf{x}^T \tag{17}$$

[2]. Application of Equation (8) to Equation (16) yields

$$\delta\bar{\mathbf{b}}\approx\kappa(A)\cdot(\delta A+\delta\bar{\mathbf{x}})\ .$$

Independence of $\bar{\mathbf{b}}$ from $\mathbf{x}$ (and hence from $\Delta\mathbf{x}$) implies $\delta\bar{\mathbf{b}}=0$ for error-free A and $\bar{\mathbf{x}}$. The outer product $\bar{A}=-\bar{\mathbf{b}}\cdot\mathbf{x}^T$ is numerically stable as scalar multiplication is. Consequently, adjoint systems of linear equations are numerically stable.

3.2 Systems of Nonlinear Equations

Differentiation of Equation (10) in the direction of absolute errors $\Delta R_{\mathbf{x}}\in\mathbb{R}^{n\times n}$, $\Delta\dot{\mathbf{x}}\in\mathbb{R}^n$, $\Delta R_{\mathbf{p}}\in\mathbb{R}^{n\times m}$ and $\Delta\dot{\mathbf{p}}\in\mathbb{R}^m$ yields

$$\Delta R_{\mathbf{x}}\cdot\dot{\mathbf{x}}+R_{\mathbf{x}}\cdot\Delta\dot{\mathbf{x}}+\Delta R_{\mathbf{p}}\cdot\dot{\mathbf{p}}\ [+\underbrace{R_{\mathbf{p}}\cdot\Delta\dot{\mathbf{p}}}_{=0}]=0$$

as $\Delta\dot{\mathbf{p}}=0$ and hence

$$\Delta\dot{\mathbf{x}}=R_{\mathbf{x}}^{-1}\cdot(\Delta R_{\mathbf{x}}\cdot\dot{\mathbf{x}}+\Delta R_{\mathbf{p}}\cdot\dot{\mathbf{p}})\ .$$

First-order estimates for

$$\Delta R_{\mathbf{x}}\cdot\dot{\mathbf{x}}=[\Delta R_{\mathbf{x}}\cdot\dot{\mathbf{x}}]_i\approx[R_{\mathbf{x},\mathbf{x}}]_{i,j,k}\cdot[\dot{\mathbf{x}}]_j\cdot[\Delta\mathbf{x}]_k\equiv\Delta\dot{R}_{\mathbf{x}}\cdot\Delta\mathbf{x}$$

and

$$\Delta R_{\mathbf{p}}\cdot\dot{\mathbf{p}}=[\Delta R_{\mathbf{p}}\cdot\dot{\mathbf{p}}]_i\approx[R_{\mathbf{p},\mathbf{x}}]_{i,j,k}\cdot[\dot{\mathbf{p}}]_j\cdot[\Delta\mathbf{x}]_k\equiv\Delta\dot{R}_{\mathbf{p}}\cdot\Delta\mathbf{x}$$

in index notation (summation over the shared index) yield

$$\Delta\dot{\mathbf{x}}\approx R_{\mathbf{x}}^{-1}\cdot(\Delta\dot{R}_{\mathbf{x}}+\Delta\dot{R}_{\mathbf{p}})\cdot\Delta\mathbf{x}$$

and hence, with Equation (7),

$$\|\delta\dot{\mathbf{x}}\|\approx\kappa(R_{\mathbf{x}})\cdot\kappa(\Delta\dot{R}_{\mathbf{x}}+\Delta\dot{R}_{\mathbf{p}})\cdot\|\delta\mathbf{x}\|\ . \tag{18}$$

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_17

Low condition numbers of the respective first and second derivatives of the residual ensure numerical stability of tangent systems of nonlinear equations. Both $\Delta\dot{R}_{\mathbf{x}}$ and $\Delta\dot{R}_{\mathbf{p}}$ can be computed by algorithmic differentiation (AD) [3, 4].

Application of Equation (8) to the system of linear equations

$$R_{\mathbf{x}}^T \cdot \mathbf{z} = -\bar{\mathbf{x}}$$

for $\Delta\bar{\mathbf{x}} = 0$ yields

$$\Delta\mathbf{z} = R_{\mathbf{x}}^{-T} \cdot \Delta R_{\mathbf{x}}^T \cdot \mathbf{z}$$

and hence

$$\|\delta\mathbf{z}\| \approx \kappa(R_{\mathbf{x}}) \cdot \kappa(\Delta\bar{R}_{\mathbf{x}}) \cdot \|\delta\mathbf{x}\| \;,$$

where

$$[\Delta R_{\mathbf{x}}^T \cdot \mathbf{z}]_j \approx [R_{\mathbf{x},\mathbf{x}}]_{i,j,k} \cdot [\mathbf{z}]_i \cdot [\Delta\mathbf{x}]_k \equiv \Delta\bar{R}_{\mathbf{x}} \cdot \Delta\mathbf{x} \;.$$

Differentiation of $\bar{\mathbf{p}} = R_{\mathbf{p}}^T \cdot \mathbf{z}$ in the direction of the non-vanishing absolute errors $\Delta R_{\mathbf{p}}^T \in \mathbb{R}^{m \times n}$ and $\Delta\mathbf{z} \in \mathbb{R}^n$ yields

$$\Delta\bar{\mathbf{p}} = \Delta R_{\mathbf{p}}^T \cdot \mathbf{z} + R_{\mathbf{p}}^T \cdot \Delta\mathbf{z} = \Delta R_{\mathbf{p}}^T \cdot \mathbf{z} + R_{\mathbf{p}}^T \cdot R_{\mathbf{x}}^{-T} \cdot \Delta R_{\mathbf{x}}^T \cdot \mathbf{z}$$

and hence

$$\|\delta\bar{\mathbf{p}}\| \approx \big(\kappa(\Delta\bar{R}_{\mathbf{p}}) + \kappa(R_{\mathbf{p}}) \cdot \kappa(R_{\mathbf{x}}) \cdot \kappa(\Delta\bar{R}_{\mathbf{x}})\big) \cdot \|\delta\mathbf{x}\| \;, \tag{19}$$

where

$$[\Delta R_{\mathbf{p}}^T \cdot \mathbf{z}]_j \approx [R_{\mathbf{p},\mathbf{x}}]_{i,j,k} \cdot [\mathbf{z}]_i \cdot [\Delta\mathbf{x}]_k \equiv \Delta\bar{R}_{\mathbf{p}} \cdot \Delta\mathbf{x} \;.$$

Low condition numbers of the respective first and second derivatives of the residual ensure numerical stability of adjoint systems of nonlinear equations. Both $\Delta\bar{R}_{\mathbf{x}}$ and $\Delta\bar{R}_{\mathbf{p}}$ can be computed by AD.

3.3 Convex Unconstrained Objectives

The first-order optimality condition for a parameterized convex unconstrained objective

$$f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} : (\mathbf{x}, \mathbf{p}) \mapsto y = f(\mathbf{x}, \mathbf{p})$$

yields the residual $f_{\mathbf{x}}(\mathbf{x}, \mathbf{p}) = 0$. Consequently, assuming f to be three times differentiable,

$$\|\delta\dot{\mathbf{x}}\| \approx \kappa(f_{\mathbf{x},\mathbf{x}}) \cdot \kappa(\Delta\dot{f}_{\mathbf{x},\mathbf{x}} + \Delta\dot{f}_{\mathbf{x},\mathbf{p}}) \cdot \|\delta\mathbf{x}\| \;, \tag{20}$$

where

$$\Delta f_{\mathbf{x},\mathbf{x}} \cdot \dot{\mathbf{x}} = [\Delta f_{\mathbf{x},\mathbf{x}} \cdot \dot{\mathbf{x}}]_i \approx [f_{\mathbf{x},\mathbf{x},\mathbf{x}}]_{i,j,k} \cdot [\dot{\mathbf{x}}]_j \cdot [\Delta\mathbf{x}]_k \equiv \Delta\dot{f}_{\mathbf{x},\mathbf{x}} \cdot \Delta\mathbf{x}$$

and

$$\Delta f_{\mathbf{x},\mathbf{p}} \cdot \dot{\mathbf{p}} = [\Delta f_{\mathbf{x},\mathbf{p}} \cdot \dot{\mathbf{p}}]_i \approx [f_{\mathbf{x},\mathbf{p},\mathbf{x}}]_{i,j,k} \cdot [\dot{\mathbf{p}}]_j \cdot [\Delta\mathbf{x}]_k \equiv \Delta\dot{f}_{\mathbf{x},\mathbf{p}} \cdot \Delta\mathbf{x} \;.$$

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_17

Similarly,

$$\|\delta \bar{\mathbf{p}}\| \approx \left(\kappa(\Delta \bar{f}_{\mathbf{x},\mathbf{p}}) + \kappa(f_{\mathbf{x},\mathbf{p}}) \cdot \kappa(f_{\mathbf{x},\mathbf{x}}) \cdot \kappa(\Delta \bar{f}_{\mathbf{x},\mathbf{x}})\right) \cdot \|\delta \mathbf{x}\| \, , \tag{21}$$

where

$$[\Delta f^T_{\mathbf{x},\mathbf{x}} \cdot \mathbf{z}]_j = [\Delta f_{\mathbf{x},\mathbf{x}} \cdot \mathbf{z}]_j \approx [f_{\mathbf{x},\mathbf{x},\mathbf{x}}]_{i,j,k} \cdot [\mathbf{z}]_i \cdot [\Delta \mathbf{x}]_k \equiv \Delta \bar{f}_{\mathbf{x},\mathbf{x}} \cdot \Delta \mathbf{x}$$

and

$$[\Delta f^T_{\mathbf{x},\mathbf{p}} \cdot \mathbf{z}]_j \approx [f_{\mathbf{x},\mathbf{p},\mathbf{x}}]_{i,j,k} \cdot [\mathbf{z}]_i \cdot [\Delta \mathbf{x}]_k \equiv \Delta \bar{f}_{\mathbf{x},\mathbf{p}} \cdot \Delta \mathbf{x} \, .$$

Low condition numbers of the respective second and third derivatives of the objective ensure numerical stability of tangent and adjoint optima of convex unconstrained objectives. Both $\Delta \dot{f}_{\mathbf{x},\mathbf{x}}$ and $\Delta \dot{f}_{\mathbf{x},\mathbf{p}}$ as well as $\Delta \bar{f}_{\mathbf{x},\mathbf{x}}$ and $\Delta \bar{f}_{\mathbf{x},\mathbf{p}}$ can be computed by AD.

4 Conclusion

Adjoint systems of linear equations are numerically stable with respect to errors in the primal solution. However, numerical stability of tangents and adjoints of implicit functions cannot be guaranteed in general. Sufficient conditions in terms of derivatives of the residual are given by Equations (15), (18), (19), (20) and (21). AD can be used to compute these derivatives. Corresponding symbolic tangents and adjoints should be augmented with optional estimation of conditions of the relevant derivatives.

References

1. A. Gebremedhin, F. Manne, and A. Pothen. What color is your jacobian? graph coloring for computing derivatives. *SIAM Review*, 47(4):629–705, 2005.
2. M. Giles. Collected matrix derivative results for forward and reverse mode algorithmic differentiation. In C. Bischof, M. Bücker, P. Hovland, U. Naumann, and J. Utke, editors, *Advances in Automatic Differentiation*, pages 35–44. Springer, 2008.
3. A. Griewank and A. Walther. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation.* Number 105 in Other Titles in Applied Mathematics. SIAM, Philadelphia, PA, 2nd edition, 2008.
4. U. Naumann. *The Art of Differentiating Computer Programs. An Introduction to Algorithmic Differentiation.* Number SE24 in Software, Environments, and Tools. SIAM, 2012.
5. U. Naumann, J. Lotz, K. Leppkes, and M. Towara. Algorithmic differentiation of numerical methods: Tangent and adjoint solvers for parameterized systems of nonlinear equations. *ACM Transactions on Mathematical Software*, 41:26, 2015.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_17