Peer two-step methods with embedded sensitivity approximation for parameter-dependent ODEs

Bernhard A. Schmitt* Ekaterina Kostina†

Abstract

Peer two-step methods have been successfully applied to initial value problems for stiff and non-stiff ordinary differential equations both on parallel and sequential computers. Their essential property is the use of several stages per time step with the same accuracy. As a new application area these methods are now used for parameter-dependent ODEs where the peer stages approximate the solution also at different places in the parameter space. The main interest here are sensitivity data through an approximation of solution derivatives in different parameter directions. Basic stability and convergence properties are discussed and peer methods of order two and three in the time stepsize are constructed. The computed sensitivity matrix is used in approximate Newton and Gauss-Newton methods for shooting in boundary value problems, where initial values and/or ODE parameters are searched for, and in parameter identification from partial information on trajectories.

Key words. Peer two-step methods, sensitivity analysis for ordinary differential equations, shooting methods, parameter identification.
AMS subject classifications. 65L05, 34A34, 65L07

1 Introduction

Peer methods have been introduced in [10] as parallel methods for the solution of initial value problems of ordinary differential equations $y'(t) = f\big(t, y(t)\big)$. The essential difference to classical time stepping methods like Runge-Kutta and multistep methods is the use of several approximations Y_{mi}, $i = 1, \ldots, s$, of equal quality for the solution $y(t)$ at off-step points t_{mi} in each time step through $[t_m, t_{m+1}]$. The two-step structure of peer methods leads to full stage order of all stages Y_{mi}. Implicit peer methods have been discussed for parallel implementation in [11], [12] and with sequential stages in [2]. Explicit peer methods were investigated in [17] as non-parallel methods and in [13] as parallel methods. Since the stages of peer methods provide several approximations of a single solution trajectory $y(t)$ in the classical case it is quite natural to extend

*Fachbereich Mathematik und Informatik, Philipps-Universität , D-35041 Marburg, Germany (schmitt@mathematik.uni-marburg.de).

†Fachbereich Mathematik und Informatik, Philipps-Universität , D-35041 Marburg, Germany (kostina@mathematik.uni-marburg.de).

this concept to parameter-dependent problems with a solution manifold $y(t,p)$, $(t,p) \in \mathbb{R}^{q+1}$. We discuss an extension of peer methods to the parameter-dependent initial value problem

$$y'(t,p) = f\big(y(t,p),p\big),\ t \in [t_0,t_{end}], \quad y(t_0,p) = u(p), \tag{1}$$

where the prime denotes the time derivative. Here only the autonomous case is considered for simplicity and f is defined on some subset of $\mathbb{R}^n \times \mathbb{R}^q$. The parameters $p \in \mathbb{R}^q$ might be some natural parameters of the ODE model like physical constants and the sensitivity of the solution $y(t,p)$ on changes of the parameters p is of interest. A rather standard application is shooting methods where $f = f(y)$ is independent of p and only the initial values $u(p)$ are variables in some large subspace of $\mathbb{R}^n$ in order to approximate the fundamental solution of the ODE. A special case here are time-periodic solutions of dynamical systems. A further application is the parameter identification in ODEs. In these areas derivatives of the solution $y(t,p)$ with respect to the parameters p_i, $i = 1,\ldots,q$, are needed for the application of Newton methods. The cheap approximation of these derivatives with peer methods is the main topic of this paper.

For this purpose the basic structure of two-step peer methods from [11] needs no modification, only the interpretation will change and is adapted to the more general situation in (1). Each stage Y_{mi}, $1 \le i \le s$, will be associated with some off-step node $t_{mi} = t_m + h_m c_i$ in time again with general time stepsizes h_m. Now, however, there will also be off-steps $r_j = (r_{\nu j})_{\nu=1}^q$, $j = 1,\ldots,s$, to some basis value p_0 also in parameter space, the latter will be assumed to be $p_0 = 0$ without restriction. These fixed parameter off-steps are scaled by a common parameter stepsize $\rho > 0$. In this setting the two-step peer-method computes approximations $Y_{mi} \cong y(t_{mi}, \rho r_i)$ by

$$Y_{mi} - h_m \sum_{j=1}^{i} \gamma_{ij} f(Y_{mj}, \rho r_j) = \sum_{j=1}^{s} b_{ij} Y_{m-1,j} + h_m \sum_{j=1}^{s} a_{ij} f(Y_{m-1,j}, \rho r_j), \tag{2}$$

$i = 1,\ldots,s$. The coefficient matrix $\Gamma = (\gamma_{ij})$ may be lower triangular, in general. However, we concentrate here on parallel implicit methods with a diagonal matrix $\Gamma = \mathrm{diag}(\gamma_i) \ge 0$ and explicit methods with $\Gamma = 0$. We note that an implicit peer method may be implemented in linearly implicit form, [11], [2]. Some of the coefficient matrices $A = (a_{ij})$, $B = (b_{ij})$, Γ of the scheme may depend on the stepsize sequence (h_m) used. In this case they will be distinguished by an additional index, e.g. A_m. There is also a compact representation of the method (2) in Kronecker-product form or by introducing matrices $Y_m^{\mathsf{T}} = (Y_{m1},\ldots,Y_{ms}) \in \mathbb{R}^{s\times n}$, $F_m^{\mathsf{T}} = \big(f(Y_{mi},\rho r_i)\big)_{i=1}^s$ of stage vectors and function evaluations which reads

$$Y_m - h_m \Gamma_m F_m = B_m Y_{m-1} + h_m A_m F_{m-1}. \tag{3}$$

In the original ODE setting the number of stages is fairly low, $2 \le s \le 8$, since it corresponds essentially to the order of the method. But for parameter-dependent ODEs much larger stage numbers $s \ge q$ may be used since one

stage is needed for each parameter direction, at least. For methods with many stages sparse coefficients A, B are attractive, of course. Using peer methods (2) we expect remarkable savings compared to the usual approach of computing parameter derivatives by using variational equations or computing neighboring solutions, [16], [1], [14]. In fact, the parameter derivative $\phi_i(t) = \partial y(t,p)/\partial p_i$ of the solution satisfies the linear variational equation

$$\phi_i'(t) = f_y(y,p)\phi_i + f_{p_i}(y,p),\ t \in [t_0, t_{end}], \quad \phi_i(t_0) = \frac{\partial u(p)}{\partial p_i}, \tag{4}$$

which may be integrated numerically by any appropriate method. In practice, however, great care is taken that the numerical solutions of (1) and of (4) or the neighboring problems have the same characteristics. Hence, all solutions are computed by the same numerical method and with the same order and stepsize sequences, e.g. [3], [14], often by *internal numerical differentiation* using the derivative of the numerical scheme itself. So, if the numerical integration method is of order k this approach uses kqn additional degrees of freedom per time step for the approximation of all parameter derivatives. This may be interpreted as a method using $k(q+1)$ stages or other solution data for the full approximation. Instead, our new approach with the peer method (2) allows to approximate all parameter derivatives by differences of only q additional stages independent of the order of the method. In a simple configuration with the first q off-step directions $r_i - r_s = e_i$ where e_i is the i-th unit vector and $c_i = c_s$, $1 \le i \le q$, the approximation is obtained in the form

$$\frac{\partial y(t_{m+1},0)}{\partial p_i} \cong \frac{Y_{m,j} - Y_{m,s}}{\rho}. \tag{5}$$

A low order approximation (5) can be obtained in this way with a peer method having only $s = k + q$ stages for order k. So the number of stages is roughly a fraction $1/k$ of the standard approach. This paper is mainly intended as a 'proof of concept' for the proposed approach. Hence rather simple situations will be discussed.

In Section 2 we start with accuracy and stability issues for the methods, especially the difficulties with zero-stability since the stability matrix has a multiple eigenvalue one. This requires some low-rank structure of the coefficient matrices and we introduce a subclass called satellite configuration for obvious reasons. It is distinguished by a simple structure and allows a flexible and efficient implementation. A careful analysis is needed in Section 3 to show that time-step independent errors $\mathcal{O}(\varrho^2)$ do not accumulate in the global error. The influence of discretization errors on the convergence of Newton and Gauss-Newton methods is analyzed in Section 4. In order to show that the computed parameter derivatives are accurate enough in practice the peer methods are applied in Sections 5 and 6 to shooting with initial values and parameters for small ODEs and the Brusselator with diffusion in 1D, and to parameter estimation for a given trajectory. By H we will denote the maximal stepsize $H = \max\{h_m : t_0 \le t_m < t_e\}$.

2 Stability and order

Before discussing order conditions some remarks on zero stability are necessary. Application of the scheme (3) to the trivial test equation $y' = 0$ leads to the simple recursion $Y_m = B_m Y_{m-1} = B_m B_{m-1} \cdots B_1 Y_0$. Obviously, a necessary stability condition is that all these products are uniformly bounded.

Definition 1 *The peer method is* zero-stable *if all products of the coefficients B_m are uniformly bounded by some constant K,*

$$\|B_m B_{m-1} \cdots B_{m-i}\| \le K,\ 0 \le i \le m. \tag{6}$$

In case of a constant coefficient $B_m = B$ in all steps the powers B^m have to be uniformly bounded for zero-stability. This is equivalent to the well-known condition that all eigenvalues of B lie inside the unit disk and those with absolute value one be non-defective. Indeed it will be seen that the matrix B will have a multiple eigenvalue one due to accuracy reasons. For the more general test equation $y' = \lambda y$, $\lambda \in \mathbb{C}$, of Dahlquist the peer method acts as a multiplication $Y_m = M_m(h_m\lambda)Y_{m-1}$ with the stability matrix

$$M_m(z) = (I - z\Gamma_m)^{-1}(B_m + zA_m). \tag{7}$$

Special values of M_m are $M_m(0) = B_m$ and $M(-\infty) = -\Gamma_m^{-1}A_m$ if Γ_m is nonsingular. Thus, for stiff problems the limit $\Gamma_m^{-1}A_m$ should be a contraction. Optimal stiff damping is achieved in [11] with the choice $A_m = 0$.

Order conditions for peer methods are derived by Taylor expansion of the residual $h_m\Delta_{mi}$ of the solution y in the scheme (2) given by

$$\begin{aligned} h_m\Delta_{mi} =& y(t_{mi}, \rho r_i) - h_m \sum_{j=1}^{i} \gamma_{ij} y'(t_{mj}, \rho r_j) \\ & - \sum_{j=1}^{s} b_{ij} y(t_{m-1,j}, \rho r_j) - h_m \sum_{j=1}^{s} a_{ij} y'(t_{m-1,j}, \rho r_j). \end{aligned} \tag{8}$$

Now, derivatives with respect to both stepsizes h and ρ have to be considered. The coefficient of the order zero term $h^0\rho^0$ in the expansion of $h\Delta$ leads to the preconsistency condition

$$\sum_{j=1}^{s} b_{ij} = 1,\ i = 1, \ldots, s \iff B\mathbb{1} = \mathbb{1} = (1, \ldots, 1)^{\mathsf{T}} \tag{9}$$

and introduces the eigenvalue 1 of B for the first time. The meaning of the other conditions will also be indicated by displaying the appropriate powers $h^\ell\rho^\nu$ in the expansion of $h\Delta$. Conditions with $\ell > 0$ will depend on the stepsize ratio

$$\sigma_m := \frac{h_m}{h_{m-1}}, \quad m \ge 1, \tag{10}$$

which is assumed to be bounded $0 < \sigma_m \le \bar{\sigma}$. Taylor expansion is applied around the point $(t_{m-1}, 0)$ and yields the conditions

$$h_{m-1}^{\ell}: \quad \frac{1}{\ell}(1+\sigma_m c_i)^{\ell} = \sigma_m \sum_{j=1}^{i} \gamma_{ij}(1+\sigma_m c_j)^{\ell-1} + \frac{1}{\ell}\sum_{j=1}^{s} b_{ij} c_j^{\ell} + \sigma_m \sum_{j=1}^{s} a_{ij} c_j^{\ell-1}, \ i = 1,\dots,s, \ \ell \ge 1, \tag{11}$$

$$\rho: \quad r_{\nu i} = \sum_{j=1}^{s} b_{ij} r_{\nu j}, \ i = 1,\dots,s, \ \nu = 1,\dots,q, \tag{12}$$

$$h_{m-1}\rho: \quad (1+\sigma_m c_i) r_{\nu i} = \sigma_m \sum_{j=1}^{i} \gamma_{ij} r_{\nu j} + \sum_{j=1}^{s} (b_{ij} c_j + \sigma_m a_{ij}) r_{\nu j}, \tag{13}$$
$$i = 1,\dots,s, \ \nu = 1,\dots,q.$$

With these conditions the structure of the local error follows easily.

Lemma 2 *Let (9) and the order condition (12) for ρ, (13) for $h\rho$ and (11) for h^{ℓ}, $1 \le \ell \le k$, be satisfied in the time step at t_m. Then, the local error in this step satisfies*

$$h_m \Delta_m = \mathcal{O}(\rho^2 + h_{m-1}^2 \rho + h_{m-1}^{k+1}).$$

The conditions (11) for h_{m-1}^{ℓ} correspond to those from [12] but the others have important consequences. Evidently, the ρ-condition (12) is equivalent to the identity

$$B_m R^{\mathsf{T}} = R^{\mathsf{T}} \tag{14}$$

where $R = (r_{\nu j}) \in \mathbb{R}^{q \times s}$ is the off-step matrix in the parameter space. By (12) each of the rows $r^{(\nu)} = R^{\mathsf{T}} e_{\nu}$, $\nu = 1,\dots,q$, of R introduces an additional eigenvalue one in B. The multiple eigenvalue one has to be analyzed carefully in view of zero stability (6). In order to discuss this problem in detail a basis transformation is used with the matrix

$$X = (r^{(1)}, \dots, r^{(q)}, \mathbb{1}, \dots) = (R^{\mathsf{T}}, \mathbb{1}, \dots) \in \mathbb{R}^{s \times s}. \tag{15}$$

The last $s-q-1$ columns of X will be specified later on but nonsingularity of X is assumed. Now, the conditions (9) and (12) or (14) are equivalent with the transformed matrix

$$\tilde{B}_m := X^{-1} B_m X = \begin{pmatrix} I_{q+1} & \tilde{B}_{m,2} \\ 0 & \tilde{B}_{m,4} \end{pmatrix} \tag{16}$$

having as leading block the identity matrix of dimension $q+1$.

Lemma 3 *Let the off-diagonal blocks in (16) be uniformly bounded, $\|\tilde{B}_{m,2}\| \le \zeta$ and assume that*

$$\|\tilde{B}_{m,4} \tilde{B}_{m-1,4} \cdots \tilde{B}_{m-i,4}\| \le b\beta^{i+1}, \ \beta < 1,$$

for all $0 \le i \le m$ such that $t_m \le t_{end}$. Then with some $K > 0$ there is a uniform bound for all products

$$\|B_m B_{m-1} \cdots B_{m-i}\| \le \frac{K}{1-\beta},\ 0 \le i \le m.$$

Proof Transformed products have the form

$$\tilde{B}_m \tilde{B}_{m-1} \cdots \tilde{B}_{m-i} = \begin{pmatrix} I & N_{mi} \\ 0 & \tilde{B}_{m,4} \cdots \tilde{B}_{m-i,4} \end{pmatrix}$$

with

$$\begin{aligned} \|N_{mi}\| =& \|\tilde{B}_{m,2}\tilde{B}_{m-1,4} \cdots \tilde{B}_{m-i,4} + \ldots + \tilde{B}_{m-i+1,2}\tilde{B}_{m-i,4} + \tilde{B}_{m-i,2}\| \\ \le& \zeta(1 + b\beta + \ldots + b\beta^{i-1}) \le \zeta\big(1 + \frac{b\beta}{1-\beta}\big). \end{aligned}$$

With some constant K_1 depending on the actual matrix norm it follows that $\|\tilde{B}_m \tilde{B}_{m-1} \cdots \tilde{B}_{m-i}\| \le K_1 \max\{\beta^{i+1}, 1+\zeta+\zeta b/(1-\beta)\} \le K_1(1+\zeta+\zeta b\beta)/(1-\beta)$ and the assertion holds with $K = (1 + \zeta + \zeta b\beta)K_1 cond(X)$. $\square$

It may be seen from (12), (13) that for each $\ell \ge 1$ the condition for $h^\ell\rho$ introduces qs equations and requires q additional stages of the peer method. Thus, the simplest useful case with the fewest stages uses only the $h\rho$-condition (13). Introducing the node matrix $C = \mathrm{diag}(c_i)$ this condition can be written in matrix form as

$$\big(I + \sigma_m C - BC - \sigma_m(A + \Gamma)\big)R^{\mathsf{T}} = 0. \tag{17}$$

Accordingly, each of the h^k-conditions (11), $k \ge 1$, reads as

$$\Big(\frac{1}{k}(I + \sigma_m C)^k - \sigma_m \Gamma(I + \sigma_m C)^{k-1} - \frac{1}{k}BC^k - \sigma_m AC^{k-1}\Big)\mathbb{1} = 0.$$

The combined conditions may be solved either for A or B. Hence, at least one of these two coefficient matrices will depend on σ_m. Step-dependent matrices B_m may be difficult with respect to zero-stability (6) while σ-dependence of A may be critical in the stiff limit $M(\infty)$ of (7). In this paper we concentrate on non-stiff problems and will use a constant matrix $B_m = B$ as in [17], [13].

2.1 Low-rank representations

Considering condition (17) as an equation for σA the similarity with the conditions (12) on B is striking. Hence it is natural to specify now the remaining columns of the basis matrix in (15) and define X by

$$X = (r^{(1)}, \ldots, r^{(q)}, \mathbb{1}, C\mathbb{1}, \ldots, C^{s-q-1}\mathbb{1}) = (R^{\mathsf{T}}, V_{s-q}) \in \mathbb{R}^{s\times s}, \tag{18}$$

where the last $s - q$ columns represent a tall Vandermonde matrix $V_{s-q} = (c_i^{j-1}) \in \mathbb{R}^{s\times(s-q)}$. Combining the $h\rho$ conditions with those for $h^1, \ldots, h^{s-q}$ is equivalent with the identity

$$\begin{aligned} \sigma AX =& (I + \sigma C)(R^{\mathsf{T}}, \mathbb{1}, \frac{1}{2}(I + \sigma C)\mathbb{1}, \ldots) \\ & - BC(R^{\mathsf{T}}, \mathbb{1}, \frac{1}{2}C\mathbb{1}, \ldots) - \sigma\Gamma(R^{\mathsf{T}}, \mathbb{1}, (I + \sigma C)\mathbb{1}, \ldots). \end{aligned} \tag{19}$$

In this equation all right-hand factors including X have common leading columns. Hence it is convenient to write them as perturbations of the basis matrix X itself. For the columns with $2 \le j \le k = s - q$ this results in

$$\begin{aligned}(R^{\mathsf{T}}, \mathbb{1}, \frac{1}{2}(I+\sigma C)\mathbb{1}, \ldots, \frac{1}{k}(I+\sigma C)^{k-1}\mathbb{1})e_{q+j} =& Xe_{q+j} + W_1 e_{j-1},\\ (R^{\mathsf{T}}, \mathbb{1}, \frac{1}{2}C\mathbb{1}, \ldots, \frac{1}{k}C^{k-1}\mathbb{1})e_{q+j} =& Xe_{q+j} + W_2 e_{j-1},\\ (R^{\mathsf{T}}, \mathbb{1}, (I+\sigma C)\mathbb{1}, \ldots, (I+\sigma C)^{k-1}\mathbb{1})e_{q+j} =& Xe_{q+j} + W_3 e_{j-1}.\end{aligned}$$

Here, three rectangular matrices $W \in \mathbb{R}^{s\times(s-q-1)}$ have been introduced with the columns

$$W_1 e_{j-1} = \frac{1}{j}(I+\sigma C)^{j-1}\mathbb{1} - C^{j-1}\mathbb{1}, \tag{20}$$

$$W_2 e_{j-1} = -\frac{j-1}{j}C^{j-1}\mathbb{1}, \tag{21}$$

$$W_3 e_{j-1} = (I+\sigma C)^{j-1}\mathbb{1} - C^{j-1}\mathbb{1}. \tag{22}$$

According to (16) also B has a similar representation $BX = X + W_0(0, I_{s-q-1})$, where W_0 contains the last $s-q-1$ nontrivial columns of $X(\tilde{B} - I)$. Thus, we have

$$B = I + W_0 Z^{\mathsf{T}} \quad \text{with} \quad Z^{\mathsf{T}} = (0, I_{s-q-1})X^{-1}. \tag{23}$$

Due to (16) it holds that $Z^{\mathsf{T}}W_0 = \tilde{B}_{m,4} - I$ and we note that in designing peer methods it is convenient to deal with the elements of W_0 as free parameters under the restriction that the eigenvalues of $Z^{\mathsf{T}}W_0$ lie inside a small circle centered at -1. Putting these low-rank representations into (19) gives

$$\begin{aligned}\sigma A =& (I+\sigma C)(I + W_1 Z^{\mathsf{T}}) - BC(I + W_2 Z^{\mathsf{T}}) - \sigma\Gamma(I + W_3 Z^{\mathsf{T}})\\ =& (I+\sigma C)(I + W_1 Z^{\mathsf{T}}) - (I + W_0 Z^{\mathsf{T}})C(I + W_2 Z^{\mathsf{T}}) - \sigma\Gamma(I + W_3 Z^{\mathsf{T}})\\ =& I + (\sigma-1)C - \sigma\Gamma - W_0 Z^{\mathsf{T}} C\\ &+ \big((I+\sigma C)W_1 - CW_2 - W_0 Z^{\mathsf{T}} C W_2 - \sigma\Gamma W_3\big)Z^{\mathsf{T}}.\end{aligned} \tag{24}$$

These low-rank representations of B and A reveal some special structure of the method. With (23) and (24) the peer step (3) takes the form

$$\begin{aligned}Y_m - h_m\Gamma_m F_m =& Y_{m-1} + h_{m-1}\big(I - C + \sigma_m(C-\Gamma)\big)F_{m-1}\\ &+ W_0 Z^{\mathsf{T}}\big(Y_{m-1} - h_{m-1}CF_{m-1}\big) + h_{m-1}W(\sigma)Z^{\mathsf{T}}F_{m-1}.\end{aligned}$$

The first line shows a simple diagonal part without communication between the different peers and it is easily recognized as a collection of independent steps of the tau method. This decoupled part is complemented by some low rank corrections collected from the data $Y_{m-1,i} - h_{m-1}c_i f(Y_{m-1,i}) \cong y(t_{m-1}, \rho r_i)$ and F_{m-1}. Moreover, this information is gathered only from the projection onto one fixed subspace $Rg(Z)$. This subspace which determines the communication

pattern of the method is easily characterized. Since the last block row of the inverse X^{-1} is orthogonal to the first block column of X it holds that

$$Rg(Z) = Null(R). \tag{25}$$

Thus, the null space $Null(R)$ of the off-step matrix R is also the 'kernel' of the method from which all communication originates. In the satellite configuration of Section 2.2 this kernel simply consists of the last $s-q$ stages. However, if R has more than q nontrivial columns and the problem is nonlinear the situation may be more complex.

2.2 A satellite configuration

We describe a simple and interesting situation arising for a special choice of method parameters and confine the discussion to parallel methods with $\Gamma = \mathrm{diag}(\gamma_i)$. It is quite obvious and will also be shown in Section 3 that good accuracy of the parameter derivatives requires that $q+1$ stages are positioned at the same point in time. So, the first q stages are placed off the central trajectory $p=0$ on the same off-step point $c_j = c_s$, $j \le q$, in time and all the remaining stages starting with $Y_{m,q+1}$ are on the central trajectory, i.e.

$$c_1 = \ldots = c_q = c_s, \quad r_{q+1} = \ldots = r_s = 0. \tag{26}$$

The first q vectors r_j must form a basis for $\mathbb{R}^q$. This choice is called the satellite configuration since it corresponds to an accurate central trajectory at $p=0$ represented by the stages $Y_{m,q+1}, \ldots, Y_{ms}$ which are accompanied by q satellites $Y_{m1}, \ldots, Y_{mq}$. The last central stage Y_{ms} is the reference stage for the satellites due to $c_1 = c_s$. The choice (26) means that the image of R^{T} is also the image of $E_q := (I_q, 0)^{\mathsf{T}} \in \mathbb{R}^{s\times q}$ and the subspace spanned by E_q is invariant under the basis matrix X from (18). In fact, $R^{\mathsf{T}} = E_q \hat{R}^{\mathsf{T}}$ with a nonsingular matrix $\hat{R} \in \mathbb{R}^{q\times q}$ and $CR^{\mathsf{T}} = CE_q\hat{R}^{\mathsf{T}} = c_1 R^{\mathsf{T}}$. Now, for the central stages $i = q+1, \ldots, s$ the order conditions (13) are written in $\mathbb{R}^q$ and become

$$0 = \sum_{j=1}^{q} (c_1 b_{ij} + \sigma_m a_{ij}) r_j = c_1 r_i + \sigma_m \sum_{j=1}^{q} a_{ij} r_j = \sigma_m \sum_{j=1}^{q} a_{ij} r_j$$

due to (12). And for the satellites $i = 1, \ldots, q$ the nodes $c_i = c_1$ coincide giving

$$\begin{aligned} (1+\sigma_m c_1) r_i =& \sigma_m \gamma_i r_i + c_1 \sum_{j=1}^{q} b_{ij} r_j + \sigma_m \sum_{j=1}^{q} a_{ij} r_j \quad \Longleftrightarrow \\ (1+\sigma_m c_1 - c_1) r_i =& \sigma_m \big(\gamma_i r_i + \sum_{j=1}^{q} a_{ij} r_j\big) \end{aligned}$$

due to (12), too. Both equations combine to the condition $\sigma_m(A+\Gamma)R^{\mathsf{T}} = (1+\sigma_m c_1 - c_1)R^{\mathsf{T}}$ which is equivalent to

$$\sigma_m (A+\Gamma) E_q = (1+\sigma_m c_1 - c_1) E_q \tag{27}$$

due to the nonsingularity of $\hat{R}$. It shows that $A+\Gamma$ has an upper triangular block structure and a q-fold eigenvalue $c_1+(1-c_1)/\sigma_m$ which is equal to one for the convenient choice $c_1=1$. We note that the block sizes differ by one from those in (16). For the diagonal matrix Γ the image of E_q is also invariant by $\Gamma E_q = E_q\hat{\Gamma}$ where $\hat{\Gamma}$ contains the leading q diagonal entries of Γ and thus

$$AE_q = E_q(\alpha I - \hat{\Gamma}), \quad \alpha = c_1 + \frac{1-c_1}{\sigma_m}. \tag{28}$$

Combining with (12) it is seen that all coefficient matrices have a block triangular structure with leading blocks in diagonal form, $BE_q = E_q$, $\Gamma E_q = E_q\hat{\Gamma}$ and (28). This property means that none of the first q stages passes any information to other stages receiving information from the last stages $Y_{m,q+1},\ldots,Y_{ms}$ only. Hence the first q stages are indeed satellites escorting the central trajectory and being controlled by the central solution. A consequence of this structure is that the $\mathcal{O}(\rho^2)$ term from Lemma 2 does not appear in the local error, see §3.

Due to the block triangular structure absolute stability of the satellite stages is governed by a diagonal matrix. In fact,

$$M(z)E_q = E_q(I - z\hat{\Gamma})^{-1}\big(I + z(\alpha I - \hat{\Gamma})\big)$$

due to (28). Absolute stability for small z to the left of the imaginary axis requires $\alpha > 0$ which is easy to fulfill for all stepsize ratios $\sigma > 0$ under the conditions $0 \le c_1 \le 1$. However, for implicit methods satisfying $2\gamma_i \ge c_1 + (1-c_1)/\sigma_m$, $i=1,\ldots,q$, the satellite method is even A-stable since the leading block of $M(z)$ is bounded by one in the left complex halfplane, $\|(I-z\hat{\Gamma})^{-1}\big(I+z(\alpha I-\hat{\Gamma})\big)\| \le 1$. Since these parameter restrictions are obeyed very easily the stability of the complete scheme is essentially determined by that of the central stages $Y_{m,q+1},\ldots,Y_{ms}$. Although this satellite configuration may be quite simple it shows that the new order conditions (13) do not prohibit the construction of stable methods. For very stiff problems even fully implicit satellite stages with $a_{ii}=0$ may be possible with the choice $\gamma_i=\alpha$, $1\le i\le q$.

In the applications in Section 5 and 6 a further advantage of the satellite configuration will be seen. In shooting and parameter estimation the Newton step for nonlinear problems is usually implemented with a line search where no parameter derivatives are required and an approximation of the central trajectory $y(t,0)$ is needed only. Here, the satellites may be switched off and the computations be restricted to the few central stages $Y_{m,q+1},\ldots,Y_{ms}$.

2.3 A second order explicit satellite peer method

The results of the previous subsection are combined for the construction of a simple explicit satellite method with $s=q+2$ stages. It will have order 2 since from (11) only the h and h^2 conditions are satisfied. Here, with c_s=1 the term

$$Z^{\mathsf{T}} = e_s^{\mathsf{T}}X^{-1} = \frac{1}{1-c_{s-1}}(e_s - e_{s-1})^{\mathsf{T}}$$

is a row vector only. The low rank modification W_0 of $B = I + W_0Z^{\mathsf{T}}$ has to satisfy $Z^{\mathsf{T}}W_0 = \tilde{B}_{m,4} - I$. Now, with the choice $\tilde{B}_{m,4} = 0 \in \mathbb{R}^{1\times 1}$ we may

use the column vector $W_0 =: (\alpha_1, \ldots, \alpha_s)^\mathsf{T}$ as a free parameter with the side condition $Z^\mathsf{T}\alpha = \tilde{B}_{m,4} - I = -1$. This means that $\alpha_{s-1} = 1 - c_{s-1} + \alpha_s$. From (20) to (22) and $j = 2$ the identities follow

$$W_1 = \frac{1}{2}(I + \sigma C - 2C)\mathbb{1}, \quad W_2 = -\frac{1}{2}C\mathbb{1}, \quad W_3 = (I + (\sigma - 1)C)\mathbb{1}.$$

Furthermore,

$$Z^\mathsf{T}CW_2 = -\frac{1}{2}Z^\mathsf{T}C^2\mathbb{1} = -\frac{c_{s-1}+1}{2}.$$

Since we will not discuss stepsize control for this simple explicit scheme it is written down here only for the simple case of constant stepsizes with $\sigma = 1$.

$$\begin{aligned} Y_m =& Y_{m-1} + hF_{m-1} + \frac{h}{2}\mathbb{1}Z^\mathsf{T}F_{m-1} \\ &+ \alpha Z^\mathsf{T}\big(Y_{m-1} + h\big(\frac{c_{s-1}+1}{2}I - C\big)F_{m-1}\big). \end{aligned} \tag{29}$$

Evidently, $B = I + \alpha Z^\mathsf{T}$ satisfies the requirements of zero stability due to $Z^\mathsf{T}\alpha = -1$. However, if also rounding errors are considered a moderate norm $\|B\| \geq 1$ is an advantage. Since Z has only two non-zero components both the infinity norm $\|B\|_\infty$ and the spectral norm $\|B\|_2 = \|\alpha\|_2\|Z\|_2$ suggest the choice $\alpha_1 = \ldots = \alpha_{s-2} = 0$. Then, the infinity norm is optimal, $\|B\|_\infty = 1$, for $\beta := -\alpha_s/\delta \in [0,1]$, $\delta := 1 - c_{s-1}$, and the minimal spectral norm of B is obtained at $\beta = 1/2$. With the free parameter β the method (29) consists of the following stages with the abbreviation $\delta = 1 - c_{s-1}$,

$$\begin{aligned} Y_{m,i} =& \quad Y_{m-1,i} + hF_{m-1,i} + \frac{h}{2\delta}(F_{m-1,s} - F_{m-1,s-1}), \\ & \quad i = 1, \ldots, s-2, \\ Y_{m,s-1} =& \quad \beta Y_{m-1,s-1} + (1-\beta)Y_{m-1,s} \\ & \quad + \frac{h}{2\delta}(\beta\delta^2 - c_{s-1}^2)F_{m-1,s-1} + \frac{h}{2\delta}(\beta\delta^2 + 1 - \delta^2)F_{m-1,s}, \\ Y_{m,s} =& \quad \beta Y_{m-1,s-1} + (1-\beta)Y_{m-1,s} \\ & \quad + \frac{h}{2\delta}(\beta\delta^2 - 1)F_{m-1,s-1} + \frac{h}{2\delta}(\beta\delta^2 + 2\delta + 1)F_{m-1,s}. \end{aligned} \tag{30}$$

Obviously, the last two stages constitute an explicit two-stage peer method on the central trajectory at $p = 0$. We note that for the special parameter values $\beta = c_{s-1} = 0$ the stage $s-1$ reduces to $Y_{m,s-1} = Y_{m-1,s}$ and the central scheme is simply Adams-Bashforth-2. In the satellite stages $1, \ldots, s-2$ we have an explicit Euler scheme which is modified by a second order term from the central trajectory given by a difference quotient of first derivatives. It is quite convenient in applications that the coefficients in (30) are independent of the dimension q. This will be exploited in the next subsection.

2.4 Higher order satellite methods

Efficient time integration requires a variable stepsize implementation based on error estimates. Hence, we consider peer methods of order 3 with an error estimate relative to an embedded method of order 2. In (30) it could be noted that the coefficients are essentially identical for all satellite stages. This is a nice

property in practice since one may apply the peer method with any number of parameters q by using a cardinal basis in $\mathbb{R}^q$ and simply adding one satellite stage for each parameter with same coefficients. It also reduces the cost of the scheme, since the correction term $\sum_{j>q} a_{ij} F_{m-1,j}$, $i \le q$, may be computed once and added to all satellite stages once the central solution is accepted in the current time step. In the next lemma the observation on the satellites is generalized to higher order methods. We remind the fact that in designing methods the elements of W_0 from (23) may serve as parameters of the method and that zero entries in the first rows of W_0 lead to a small norm of B. In the following statement δ_{ij} denotes the Kronecker symbol and we recall the definition $E_q = (I_q, 0)^\mathsf{T}$.

Lemma 4 *Consider a satellite method* (26) *with $s > q$ stages satisfying the order conditions for $\varrho, h\varrho$ and $h^1, \dots, h^{s-q}$. If also $R = E_q^\mathsf{T}$, $E_q^\mathsf{T} B = E_q^\mathsf{T}$, $E_q^\mathsf{T}\Gamma = \gamma_1 E_q^\mathsf{T}$ and $E_q^\mathsf{T} W_0 = 0$ hold, then the method has essentially identical parameters in all satellite stages, $b_{ij} = \delta_{ij}$, $a_{ij} = (c_1 - \gamma_1 + (1-c_1)/\sigma)\delta_{ij}$ for $1 \le i \le n$, $1 \le j \le q$, and $b_{ij} = 0$, $a_{ij} = a_{1j}$ for $1 \le i \le q$, $q < j \le s$.*

Proof The stated properties of B are a simple consequence of the $h^0\rho^1$ condition (14) and the assumptions, where the basis matrix X has an upper block triangular structure. Hence, the first q columns of Z^T are zero and from (24) and the assumptions follows that

$$\begin{aligned}\sigma E_q^\mathsf{T} A =& (1 + (\sigma-1)c_1 - \sigma\gamma_1) E_q^\mathsf{T} \\ &+ \big((1+\sigma c_1) E_q^\mathsf{T} W_1 - c_1 E_q^\mathsf{T} W_2 - \sigma\gamma_1 E_q^\mathsf{T} W_3\big) Z^\mathsf{T}.\end{aligned}$$

This shows the assertion for the first q columns of A by (28). From the definitions (20,21,22) it follows that the first q rows of each of the matrices W_1, W_2, W_3 are identical since they depend only on σ and c_1. So, the matrices $E_q^\mathsf{T} W_k = E_q^\mathsf{T} \mathbb{1} e_1^\mathsf{T} W_k$, $k = 1, 2, 3$, actually have rank-1 structure and this completes the proof. □

In the satellite configuration the central stages $Y_{m,q+1}, \dots, Y_{ms}$ constitute a standard two-step peer method as discussed in [13] for the explicit case. Since no 3-stage methods were constructed in [13] we consider the following explicit method. According to Lemma 4 only the case $q = 1$ is considered with $s = q + 3 = 4$. The elements of W_0 are chosen in order to obtain a moderate norm $\|B\|_\infty$ and a nilpotent block $\tilde{B}_4$. The method uses the nodes $c_1 = c_4 = 1$, $c_2 = 0$, $c_3 = \frac{2}{5}$ and the following parameters

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -\frac{3}{32} & 0 & \frac{35}{32} \\ 0 & \frac{33}{800} & 0 & \frac{767}{800} \\ 0 & -\frac{3}{32} & 0 & \frac{35}{32} \end{pmatrix},$$

$$\sigma A = \begin{pmatrix} \sigma & \frac{3\sigma^2}{4} + \frac{5\sigma^3}{6} & -\frac{25\sigma^2}{12} - \frac{25\sigma^3}{18} & \frac{4\sigma^2}{3} + \frac{5\sigma^3}{9} \\ 0 & -\frac{1}{128} & -\frac{25}{384} & -\frac{1}{48} \\ 0 & \frac{11}{3200} + \frac{3\sigma^2}{25} + \frac{4\sigma^3}{75} & \frac{11}{384} - \frac{\sigma^2}{3} - \frac{4\sigma^3}{45} & \frac{11}{1200} + \frac{2\sigma}{5} + \frac{16\sigma^2}{75} + \frac{8\sigma^3}{225} \\ 0 & -\frac{1}{128} + \frac{3\sigma^2}{4} + \frac{5\sigma^3}{6} & -\frac{25}{384} - \frac{25\sigma^2}{12} - \frac{25\sigma^3}{18} & -\frac{1}{48} + \sigma + \frac{4\sigma^2}{3} + \frac{5\sigma^3}{9} \end{pmatrix},$$

for general stepsize ratios $\sigma > 0$. For $\sigma = 1$ the region of absolute stability of this method is contained in $[-1.081, 0] \times [-0.5613, 0.5613]$ and the error coefficient of the central method is

$$\lim_{z\to 0} z^{-4} \det(e^z I - M_c(z)) = \frac{1139}{5760} \doteq 0.1977,$$

where M_c means the lower 3×3 block of M. Hence, the leading local error of the method is $\cong \frac{1}{5}(\sigma h)^4 y^{(4)}$. No estimate for this expression is available easily. However, as in [13], Section 6, the last row of X^{-1} is a second order difference quotient and by

$$ee := h_{m-1}\frac{\sigma^3}{3}\Big(\frac{5}{2}F_{m-1,2} - \frac{25}{6}F_{m-1,3} + \frac{5}{3}F_{m-1,4}\Big) \cong \frac{h_m^3}{3!}y^{(3)}(t_m) \tag{31}$$

an estimate for the local error of an embedded method of order 2 is available for error control.

3 The global error

The error of the peer method depends on two stepsizes now, h and ρ. Still, there is a fundamental difference between both since decreasing the time stepsize h increases the computational effort by the greater number of time steps. The choice of ρ, however, has no influence on the effort and it may be chosen as small as rounding errors are negligible in the parameter derivatives (5). This point will be important in the following discussion since we may assume $\varrho = \mathcal{O}(h)$, at least. In the situation of Lemma 2 the local error has the form

$$h_m\Delta_m = \mathcal{O}(\rho^2 + h_{m-1}^2\rho + h_{m-1}^{k+1})$$

and it seems that the $\mathcal{O}(\rho^2)$ terms may accumulate during time integration. For a closer look at the local error we abbreviate $y_m = \big(y(t_{mi}, \rho r_i)^\mathsf{T}\big)_{i=1}^s$, $y'_m = \big(y'(t_{mi}, \rho r_i)^\mathsf{T}\big)_{i=1}^s$ and write the difference $y_m - B_m y_{m-1}$ as $(I - B_m)y_{m-1} + \mathcal{O}(h_m)$. Now, since B_m has a multiple eigenvalue one by (9), (14) the local error has the representation

$$\begin{aligned} h_m\Delta_m =& (I - B_m)(y_{m-1} - \mathbb{1}v^\mathsf{T} - R^\mathsf{T}U^\mathsf{T}) \\ &+ y_m - y_{m-1} - h_m\Gamma y'_m - h_m A y'_{m-1}. \end{aligned}$$

where arbitrary elements from the null space of $I - B_m$ may be subtracted in the first line with $v \in \mathbb{R}^n$ and $U \in \mathbb{R}^{n\times s}$. Obviously all terms in the second line are of order $\mathcal{O}(h)$ at least and the others are multiplied by $I - B_m$. In fact, by choosing appropriate Taylor coefficients for the terms v, U those leading terms in the Taylor expansion of $h_m\Delta_m$ without any h-factors turn out to be

$$h_m\Delta_m^{[\rho]} := (I - B_m)\Big(2\sum_{|\ell|=2} \frac{r_{j1}^{\ell_1}\cdots r_{jq}^{\ell_q}}{\ell!}\int_0^\rho (\rho - \tau)\frac{\partial^{|\ell|}y_i}{\partial^\ell p}(t_{m-1}, \tau r_j)\Big)_{j,i} d\tau \tag{32}$$

with a multi-index $\ell = (\ell_1, \ldots, \ell_q)$. Obviously, the integral term is of order $\mathcal{O}(\rho^2)$. In the satellite configuration all terms in (32) are actually zero since

$(I - B_m)E_{q+1} = 0$ and $r_{\nu j} = 0$, $\nu > q$, here. Thus, for satellite methods all local error terms contain one factor h, at least.

The global errors $\Theta_m = Y_m - y_m$ of the peer method satisfy the recursion

$$\begin{aligned}\Theta_m - h_m\Gamma_m\big(\Theta_{mj}^{\mathsf{T}}J_{mj}^{\mathsf{T}}\big)_{j=1}^{s} =& B_m\Theta_{m-1} + h_mA_m\big(\Theta_{m-1,j}^{\mathsf{T}}J_{m-1,j}^{\mathsf{T}}\big)_{j=1}^{s} \\ &+ h_m\Delta_m,\end{aligned} \tag{33}$$

where J_{mj} is some mean value of the Jacobian f_y. In principle (33) has the form $\Theta_m = B_m\Theta_{m-1} + h_m\Psi_m$, $m \geq 0$, where Ψ_m may also depend on errors Θ_j. This recursion may be solved by

$$\Theta_m = B_m \cdots B_1\Theta_0 + \sum_{j=1}^{m} h_jB_m \cdots B_{j+1}\Psi_j.$$

Now, under the assumptions of Lemma 3 it follows that

$$\begin{aligned}\|\Theta_m\| \leq& K'\|\Theta_0\| + \sum_{j=1}^{m} h_j\|B_m \cdots B_{j+1}\Psi_j\| \\ \leq& K'\big(\|\Theta_0\| + \sum_{j=1}^{m} h_j\|\Psi_j\|\big)\end{aligned} \tag{34}$$

where K' is a uniform bound for all B-products. The cases of satellite methods and more general ones are discussed separately. For satellite methods a first lemma follows along the lines of [8] (Th. 5.8).

Lemma 5 *Let the explicit peer method have a satellite configuration (26) and satisfy the preconsistency and order conditions (12), (13) and (11) for $1 \leq \ell \leq k$. If it also satisfies the stability assumptions of Lemma 3 and uses accurate starting values $Y_0 = y_0 + \mathcal{O}(H\rho + H^k)$ then its global error is bounded by*

$$\|Y_{mi} - y(t_{mi}, \rho r_i)\| = \mathcal{O}(H\rho + H^k), 1 \leq i \leq s,\ t_1 \leq t_m \leq t_{end}.$$

Proof From (33) follows $\Psi_m \leq const(\|\Theta_{m-1}\| + \|\Delta_m\|)$, $m \geq 1$. This leads to a relation of the form

$$\theta_m \leq \alpha \sum_{j=1}^{m} h_j\theta_{j-1} + \eta_m \tag{35}$$

for the errors $\theta_m = \|\Theta_m\| \geq 0$ with $\eta_0 = \theta_0$ and $\eta_m = \mathcal{O}(\theta_0 + \max_j \|\Delta_j\|)$, $m \geq 1$. Now, since $\alpha h \leq e^{\alpha h} - 1$ it is verified readily that the sequence θ_m is bounded by

$$\theta_m \leq \eta_m + \alpha \sum_{j=1}^{m} e^{\alpha(t_{m+1} - t_{j+1})}h_j\eta_{j-1} \leq const \cdot \max_{j=0}^{m} \eta_j \tag{36}$$

showing $\theta_m = \mathcal{O}(H\rho + H^k)$. $\square$

For more general configurations with $h_m\Delta_m^{[\rho]} \neq 0$ the terms in (34) and (36) have to be studied more closely. The critical point is the growth of the combined factors

$$B_m \cdots B_{j+1}(I - B_j), \quad m > j,$$

from several steps. In the case of a constant coefficient matrix $B = B_m$ these products have the following simple form

$$B^{m-j}(I - B) = X \begin{pmatrix} 0 & -\tilde{B}_2\tilde{B}_4^{m-j} \\ 0 & \tilde{B}_4^{m-j}(I - \tilde{B}_4) \end{pmatrix} X^{-1} \tag{37}$$

by (16). Thus, for $\|\tilde{B}_4\| < 1$ sums of these products are bounded and by choosing $\tilde{B}_4 = 0$ all $\mathcal{O}(\rho^2)$ terms are even canceled out within one single step due to $B(I - B) = 0$. The following theorem deals with the global error for explicit methods with a constant coefficient matrix B.

Theorem 6 *Let an explicit peer method satisfy the preconsistency and order conditions (12), (13) and (11) for $1 \leq \ell \leq k$ and use accurate starting values $Y_0 = y_0 + \mathcal{O}(\rho^2 + H\rho + H^k)$. If the coefficient matrix B is constant for all steps and satisfies $\|\tilde{B}_4\| \leq \beta < 1$ in (16) then the global error is bounded by*

$$\|Y_{mi} - y(t_{mi}, \rho r_i)\| = \mathcal{O}(\rho^2 + H\rho + H^k), \quad 1 \leq i \leq s,\ t_1 \leq t_m \leq t_{end}.$$

Proof Due to the assumption $\|\tilde{B}_4\| \leq \beta < 1$ the products in (37) go to zero as $\beta^{m-j} \to 0\,(m - j \to \infty)$. Hence, the terms in (34) are bounded by

$$h_j\|B_m \cdots B_{j+1}\Psi_j\| \leq const\big(\theta_{j-1} + \beta^{m-j}\rho^2 + h_{j-1}^2\rho + h_{j-1}^{k+1}\big),$$

$1 \leq j \leq m$, and the errors η_m in (35) now satisfy

$$\eta_m \leq const\big(\theta_0 + \sum_{j=1}^{m}(\beta^{m-j}\rho^2 + h_{j-1}^2\rho + h_{j-1}^{k+1})\big) = O\big(\rho^2 + H\rho + H^k\big).$$

Thus, from (36) follows the assertion. □

The parameter derivatives ϕ_i from (4) will be approximated from linear combinations of the stages Y_{mi}. Thus, appropriate constellations for the off-steps c_i, r_i have to be found. Simple choices for the off-step directions r_i are unit vectors, but more general choices may also have advantages. Since it is natural to have one reference stage for solution output at t_{m+1} (and $p = 0$) the last time off-step is usually chosen as $c_s = 1$, [12, 13]. It should also lie on the central path where $r_s = 0$ is appropriate. Looking now at the parameter derivatives, at all off-step points with $c_i = c_s$ the solution satisfies

$$y(t_{mi}, \rho r_i) = y(t_{ms}, 0) + \rho\sum_{j=1}^{q}\phi_j(t_{ms})^{\mathsf{T}} r_j + \mathcal{O}(\rho^2) \tag{38}$$

with the parameter derivatives ϕ_j from (4). At other points with $c_i \neq c_s$ additional terms $h(c_j - c_s)y'(t_{ms}, 0)$ and also error terms $\mathcal{O}(h_m^2 + h_m\rho)$ appear and may lead to a large error $\mathcal{O}(h_m^2/\rho)$ in (5) which should be avoided. Hence,

we introduce the matrix $Q \in \mathbb{R}^{s\times\ell}$, $Q^{\mathsf{T}}Q = I_\ell$, which selects only those $\ell \le s$ stages where $c_i = c_s = 1$, $i \ne s$, from the set of all s stages. Combining identity (38) for these stages and using again the notation $y_m := \big(y(t_{mi}, \rho r_i)^{\mathsf{T}}\big) \in \mathbb{R}^{s\times n}$ one gets

$$Q^{\mathsf{T}} y_m = \mathbb{1} y(t_{ms}, 0)^{\mathsf{T}} + \rho Q^{\mathsf{T}} R^{\mathsf{T}} (\nabla_p y)^{\mathsf{T}} + \mathcal{O}(\rho^2),$$

with the matrix $\nabla_p y = \big(\phi_1(t_{m+1}), \ldots, \phi_q(t_{m+1})\big)$. The exact solution y_m may be replaced by its numerical approximation Y_m and by Theorem 6 we get

$$Q^{\mathsf{T}}(Y_m - \mathbb{1} Y_{ms}^{\mathsf{T}}) = \rho (RQ)^{\mathsf{T}} (\nabla_p y)^{\mathsf{T}} + \mathcal{O}(\rho^2 + H\rho + H^k).$$

The identity may be solved for the matrix of parameter derivatives

$$\nabla_p y = \frac{1}{\rho}(Y_m^{\mathsf{T}} - Y_{ms}\mathbb{1}^{\mathsf{T}}) Q (RQ)^{+} + \mathcal{O}(\rho + H + \frac{H^k}{\rho}) \tag{39}$$

if $rank(RQ) \ge q$. This rank condition means that at least $\ell = q + 1$ stages have the same value in the time off-step c_i and the corresponding parameter off-steps r_i form a full-dimensional set in $\mathbb{R}^q$. Obviously, these conditions are satisfied for the satellite configuration specified in Lemma 4. In this case the computation of the approximation (39) is very simple if a cardinal basis for the parameter offsets is used. In fact, the matrices $R^{\mathsf{T}} = Q = E_q$ and $RQ = \hat{R} = I_q$ are used in our numerical computations.

Considering the choice of the parameter stepsize ϱ the error in the parameter derivatives (39) is smallest for $\rho \cong H^{k/2}$. Thus, for $k \ge 2$ the error in $\nabla_p y$ is of order $\mathcal{O}(H)$ only. This behavior has been verified numerically by simple tests not shown here. We note again that this simple approximation is obtained very cheaply with only $s = q + k$ stages which is much less effort than the standard approach with variational equations needing essentially $(q + 1)k$ stages.

Before going on to some applications in shooting methods we will analyze the influence of approximation errors on the Newton or Gauss-Newton iteration and will address the choice of ϱ, again. Later, in Sections 5 and 6 tests will demonstrate that the peer approximation is sufficiently accurate for some applications in different situations.

4 Inexact Gauss-Newton methods

Since peer methods compute approximations for solutions and derivatives with different levels of accuracy we discuss some implications on the convergence of inexact Newton methods. In the applications discussed later on we consider the solution of some nonlinear system $G(p) = 0$, where $G : D \to \mathbb{R}^q$, $D \subseteq R^q$ is a smooth map. A more general setting are overdetermined systems where the system may be solved in a least-squares sense only by

$$\min_{p\in D} \frac{1}{2}\|G(p)\|_2^2, \quad G : D \to \mathbb{R}^\mu, \ D \subseteq \mathbb{R}^q, \tag{40}$$

$\mu \ge q$. In a Newton-type iteration the nonlinear function G is replaced by its linearization at some actual guess $p^{(\ell)}$ and the update $\Delta p^{(\ell)} = p^{(\ell+1)} - p^{(\ell)}$ is

characterized by

$$\min_{\Delta p} \frac{1}{2}\|G(p^{(\ell)}) + J(p^{(\ell)})\Delta p\|_2^2 \;\Rightarrow\; \Delta p^{(\ell)} = -J(p^{(\ell)})^+ G(p^{(\ell)}), \tag{41}$$

where $J(p) = G'(p)$ is the Jacobian of G and J^+ the Moore-Penrose pseudoinverse. It is assumed that $J(p^{(\ell)})$ has full rank q. In an ODE application like boundary value problems, however, both G and J are approximated numerically. Thus, the iteration step is replaced by an approximate step

$$p^{(\ell+1)} = p^{(\ell)} + \widetilde{\Delta p}^{(\ell)}, \quad \widetilde{\Delta p}^{(\ell)} = \Delta p^{(\ell)} + \delta p^{(\ell)}, \tag{42}$$

where $\Delta p^{(\ell)}$ is the exact update from (41) and $\delta p^{(\ell)}$ its numerical error. This error is assumed to fulfill the inequality

$$\|\delta p^{(\ell)}\| \le \varepsilon \|\widetilde{\Delta p}^{(\ell)}\|, \quad \varepsilon < 1. \tag{43}$$

This inexact iteration is analyzed with modifications of some techniques from [4], [5], [6].

Theorem 7 *Assume that $J(p)^+$, $p \in D$, satisfies the following conditions for all $t \in [0,1]$, and $p, z = p + r \in D$:*

$$\left\|J(z)^+ \left(J(p+tr) - J(p)\right) r\right\| \le \omega_1 t \|r\|^2, \quad \omega_1 < \infty, \tag{44}$$

$$\left\|J(z)^+ \left(J(p+tr) - J(p)\right)\right\| \le \omega_2 t \|r\|, \quad \omega_1 \le \omega_2 < \infty, \tag{45}$$

$$\left\|J(z)^+ \Psi(p)\right\| \le \kappa \|r\|, \quad \kappa < 1, \tag{46}$$

where $\Psi(p) := G(p) + J(p)\Delta p = G(p) - J(p)J(p)^+ G(p)$ denotes the residual of the linearized system (41). Assume that the initial guess $p^{(0)} \in D$ is sufficiently close to a solution and satisfies

$$\eta_0 := \|\widetilde{\Delta p}^{(0)}\| \Big(\frac{\omega_1 + 2\omega_2 \varepsilon}{2(1-\varepsilon)} \Big) + \frac{\kappa + \varepsilon}{1 - \varepsilon} < 1, \tag{47}$$

and let the ball $D_0 := B\big(p^{(0)}, \|\widetilde{\Delta p}^{(0)}\|/(1-\eta_0)\big)$ be contained in D, $D_0 \subseteq D$. Then, the following holds
a) the sequence of iterates from (42) remains in D_0,
b) there exists $p^* \in D_0$ with $J(p^*)^+ G(p^*) = 0$ and $p^{(\ell)} \to p^*$ $(\ell \to \infty)$ with the a-priori estimate

$$\|p^{(\ell)} - p^*\| \le \eta_0^\ell \frac{\|\widetilde{\Delta p}^{(0)}\|}{1 - \eta_0},$$

c) the convergence is linear, $\|\widetilde{\Delta p}^{(\ell+1)}\| \le \eta_\ell \|\widetilde{\Delta p}^{(\ell)}\|$ with

$$\eta_\ell := \|\widetilde{\Delta p}^{(\ell)}\| \Big(\frac{\omega_1 + 2\omega_2 \varepsilon}{2(1-\varepsilon)} \Big) + \frac{\kappa + \varepsilon}{1 - \varepsilon} < 1.$$

Proof By (43) the exact and perturbed updates are related by

$$\|\Delta p^{(\ell)}\| = \|\widetilde{\Delta p}^{(\ell)} - \delta p^{(\ell)}\| \geq (1-\varepsilon)\|\widetilde{\Delta p}^{(\ell)}\|.$$

For $p^{(\ell)}$, $p^{(\ell+1)} \in D_0$ we obtain

$$\begin{aligned}
(1-\varepsilon)\|\widetilde{\Delta p}^{(\ell+1)}\| \leq & \|\Delta p^{(\ell+1)}\| = \|J(p^{(\ell+1)})^+ G(p^{(\ell+1)})\| \\
\leq & \|J(p^{(\ell+1)})^+ \big(G(p^{(\ell+1)}) - G(p^{(\ell)}) - J(p^{(\ell)})\widetilde{\Delta p}^{(\ell)}\big) \\
& + J(p^{(\ell+1)})^+ \big(G(p^{(\ell)}) + J(p^{(\ell)})\Delta p^{(\ell)}\big) \\
& + J(p^{(\ell+1)})^+ J(p^{(\ell)})\delta p^{(\ell)}\| \\
\leq & \Big\|J(p^{(\ell+1)})^+ \int_0^1 \big(J(p^{(\ell)} + t\widetilde{\Delta p}^{(\ell)}) - J(p^{(\ell)})\big)dt\, \widetilde{\Delta p}^{(\ell)}\Big\| \\
& + \|J(p^{(\ell+1)})^+ \Psi(p^{(\ell)})\| + \|\delta p^{(\ell)}\| \\
& + \|J(p^{(\ell+1)})^+ J(p^{(\ell)})\delta p^{(\ell)} - J(p^{(\ell+1)})^+ J(p^{(\ell+1)})\delta p^{(\ell)}\|.
\end{aligned}$$

Using the assumptions on the Jacobian leads to the estimates

$$\begin{aligned}
\|\widetilde{\Delta p}^{(\ell+1)}\| \leq & \frac{1}{2}\frac{\omega_1}{1-\varepsilon}\|\widetilde{\Delta p}^{(\ell)}\|^2 + \frac{\kappa}{1-\varepsilon}\|\widetilde{\Delta p}^{(\ell)}\| + \frac{1}{1-\varepsilon}\|\delta p^{(\ell)}\| \\
& + \frac{1}{1-\varepsilon}\Big\|J(p^{(\ell+1)})^+ (J(p^{(\ell)}) - J(p^{(\ell+1)}))\delta p^{(\ell)}\Big\| \\
\leq & \frac{1}{2}\frac{\omega_1}{1-\varepsilon}\|\widetilde{\Delta p}^{(\ell)}\|^2 + \frac{\kappa+\varepsilon}{1-\varepsilon}\|\widetilde{\Delta p}^{(\ell)}\| + \frac{\omega_2\varepsilon}{1-\varepsilon}\|\widetilde{\Delta p}^{(\ell)}\|^2 \\
= & \eta_\ell \|\widetilde{\Delta p}^{(\ell)}\|.
\end{aligned}$$

By induction follows $\eta_\ell \leq \eta_0$ as long as $p^{(\ell)}$ remains in D_0. Now, since

$$\|p^{(\ell+i)} - p^{(\ell)}\| \leq \sum_{j=0}^{i-1} \|\widetilde{\Delta p}^{(\ell+j)}\| \leq \eta_0^\ell \frac{\|\widetilde{\Delta p}^{(0)}\|}{1-\eta_0}$$

we see indeed that the sequence $(p^{(\ell)})$ does not leave D_0 and is a Cauchy sequence having a limit p^*. Finally, $J(p^*)^+ G(p^*) = \Delta p^* = 0$ from the continuity of $J(p)^+$ and of G, boundedness of $J(p)^+$ on D_0 and the inequality $\|\Delta p^{(\ell)}\| \leq (1+\varepsilon)\|\widetilde{\Delta p}^{(\ell)}\|$. $\square$

Remarks a) The assumptions (44), (45) imply a global bound on the "curvature" of G. Of course, they may be replaced by a uniform bound $\|J(z)^+\| \leq \beta_1$ and a Lipschitz condition $\|J(z) - J(p)\| \leq \beta_2\|z-p\|$ for J. However, the product $\beta_1\beta_2$ grossly overestimates the bounds ω_1, ω_2.

b) The constant κ in assumption (46) describes the incompatibility of the problem to the measurements. Indeed, since

$$J(p)^+\Psi(p) = J(p)^+(I - J(p)J(p)^+)G(p) = 0$$

we can rewrite condition (46) as follows:

$$\big(J(z)^+ - J(p)^+\big)\big(I - J(p)J(p)^+\big)G(p)\| \leq \kappa\|z-p\|,\ \kappa < 1.$$

This condition guarantees that the second-order part of the Hessian of nonlinear least-squares problems does not dominate the first-order part, [7]. For this condition to hold true the residual $G(p)$ should be small in a neighborhood of the solution and the pseudoinverse J^+ should satisfy a Lipschitz condition

$$\|J(z)^+ - J(p)^+\| \leq \beta\|z - p\|$$

with sufficiently small $\beta < \infty$.
c) The condition (47) may be rewritten as

$$\|\widetilde{\Delta p}^{(0)}\|(\frac{\omega_1}{2} + \omega_2\varepsilon) + \kappa + 2\varepsilon < 1 \tag{48}$$

and can be read in different ways. Obviously, a sufficiently accurate initial guess $p^{(0)}$ is required for convergence in the first place. However, it also requires that the incompatibility constant κ and the discretization error together are smaller than one. Hence, error control can only use the place left by κ, i.e. $2\varepsilon < 1 - \kappa$. And only for $\kappa + 2\varepsilon \cong 0$ fast convergence of the iteration is possible by statement c) of the theorem.

Convergence of the iteration (42) depends on the accuracy of the computed update $\widetilde{\Delta p}^{(k)}$ which we analyze now by considering a perturbed linearized problem

$$\min_{\widetilde{\Delta p}} \frac{1}{2}\|G(p) + \delta G(p) + \big(J(p) + \delta J(p)\big)\widetilde{\Delta p}\|_2^2, \tag{49}$$

where $\widetilde{\Delta p} = \Delta p + \delta p$. As before, $\Delta p = -J(p)^+ G(p)$ is the unperturbed solution and $\Psi(p) := G(p) + J(p)\Delta p$ its residual. The following Lemma decomposes the error along the lines of [4].

Lemma 8 *Let $\widetilde{J(p)} = J(p) + \delta J(p)$ have full column rank. Then, the solution of the perturbed problem* (49) *is given by $\widetilde{\Delta p} = \Delta p + \delta p$ with*

$$\delta p = \delta p_1 + \delta p_2 + \delta p_3,$$

where

$$\begin{aligned}
\delta p_1 &= -\widetilde{J(p)}^+ \delta G(p), \\
\delta p_2 &= -\widetilde{J(p)}^+ \widetilde{J}(p)^{+^{\mathsf{T}}} \delta J(p)^{\mathsf{T}} \Psi(p), \\
\delta p_3 &= -\widetilde{J(p)}^+ \delta J(p)\Delta p.
\end{aligned}$$

Proof Problem (49) is equivalent to

$$\min_{\delta p} \frac{1}{2}\|\Psi(p) + \delta G(p) + \delta J(p)\Delta p + \widetilde{J(p)}\delta p\|_2^2.$$

With the pseudoinverse $\widetilde{J(p)}^+$ the solution is

$$\begin{aligned}
\delta p =& -\widetilde{J(p)}^+ \big(\delta G(p) + \delta J(p)\Delta p + \Psi(p)\big) \\
=& \delta p_1 + \delta p_3 - \widetilde{J(p)}^+ \Psi(p).
\end{aligned}$$

Using the exact representation of $\widetilde{J(p)}^+$ and the identity $J(p)^\mathsf{T}\Psi(p) = J(p)^\mathsf{T}(G(p)+J(p)\Delta p) = J(p)^\mathsf{T}G(p) + J(p)^\mathsf{T}J(p)\Delta p = 0$ we rewrite

$$\begin{aligned}\widetilde{J(p)}^+\Psi(p) =& \Big(\widetilde{J(p)}^\mathsf{T}\widetilde{J(p)}\Big)^{-1}\widetilde{J(p)}^\mathsf{T}\Psi(p)\\ =& \Big(\widetilde{J(p)}^\mathsf{T}\widetilde{J(p)}\Big)^{-1}(J(p)^\mathsf{T}+\delta J(p)^\mathsf{T})\Psi(p)\\ =& \Big(\widetilde{J(p)}^\mathsf{T}\widetilde{J(p)}\Big)^{-1}\delta J(p)^\mathsf{T}\Psi(p) = -\delta p_2. \qquad \square\end{aligned}$$

The Lemma will be used to discuss practical consequences for error control in order to satisfy the assumption (43) of Theorem 7. Here, bounds of the form

$$\|\delta p_j\| \le \varepsilon_j \|\widetilde{\Delta p}\|,\ j = 1,2,3, \tag{50}$$

are required. For an easier discussion we make the additional assumptions $\|\widetilde{J}(p)^+\delta J(p)\| \le \alpha$ and $\|\delta J(p)\widetilde{J}(p)^+\| \le \alpha$ with $\alpha < 1$. Then, the proof of the Lemma can be easily modified to obtain the bound

$$\|\delta p\| \le \frac{1}{1-\alpha}\big(\|\delta p_1\| + \|\delta p_2\| + \alpha\|\widetilde{\Delta p}\|\big),$$

where only δp_3 has been replaced in order to obtain $\varepsilon_3 = \alpha/(1-\alpha)$. The error δp_1 is determined by the solution error in time integration. If the integration is performed by stepsize control with the relative error *tol* one may expect that $\|\delta G\| \sim tol\|G+\delta G\|$ and (50) holds for δp_1 with $\varepsilon_1 \le K\cdot tol \cong K_1(H^k + H\varrho)$ for a satellite peer method of order k. No easy estimate in the form (50) seems to exist for $\|\delta p_2\| \le \alpha\|\widetilde{J(p)}^+\|\|\Psi(p)\|$. Hence, the product $\alpha\|\Psi(p)\|$ of the residual Ψ and the approximation error α of the Jacobian defines the obtainable level of accuracy.

Neglecting δp_2 which is zero for nonsingular quadratic problems, anyway, the condition (48) shows that $\varepsilon = \varepsilon_1 + \varepsilon_3$ should be much smaller than one in order to leave room for some initial error. Using the result from Lemma 5 and from (39) we have $\varepsilon_1 = K_1(H^k + H\varrho)$ and $\varepsilon_3 = K_3(H + H^k/\varrho + \varrho)$ and the sum

$$\varepsilon_1 + \varepsilon_3 = (K_1H + K_3)\varrho + K_3\frac{H^k}{\varrho} + K_1H^k + K_3H$$

has a minimal value for $\varrho \sim H^{k/2}$. Since stepsize control will enforce stepsizes with $H^k \sim tol$ we obtain the simple rule $\varrho \sim \sqrt{tol}$ for choosing the parameter stepsize.

5 Application to shooting methods

The application of the peer method (2) is similar for the different problem types mentioned. We consider first the solution of an ordinary boundary value problem

$$y'(t) = \tilde{f}\big(y(t), \tilde{p}\big), \quad g(y(t_0), y(t_{end})) = 0, \tag{51}$$

with $\tilde{p} \in \mathbb{R}^{\tilde{q}}$ by single shooting. Generalization to multiple shooting is straightforward. The boundary conditions are represented by a smooth function $g(u, v)$, $u, v \in \mathbb{R}^n$. Shooting requires the variation of the initial values in some subspace of $\mathbb{R}^n$. In order to describe this variation in $y(t_0)$ we introduce an additional set of parameters $\hat{p} \in \mathbb{R}^{\hat{q}}$ where $\hat{q} \leq n$ since some of the initial values may be fixed. Hence, the peer method is applied to the following initial value problem

$$y'(t) = \tilde{f}\big(y(t), \tilde{p}\big), \quad y(t_0) = u(p) := y_0 + \hat{L}\hat{p} \in \mathbb{R}^n. \tag{52}$$

The variation of the initial values $u(p)$ is described by some guess y_0 and the matrix $\hat{L} \in \mathbb{R}^{s \times \hat{q}}$ spanning the subspace of variable initial conditions. This problem (52) fits in the original situation (1) by defining $q := \tilde{q}+\hat{q}$, $p^{\mathsf{T}} = (\tilde{p}^{\mathsf{T}}, \hat{p}^{\mathsf{T}})$ and $f(u, p) := \tilde{f}(u, \tilde{p})$. Setting $L = (0, \tilde{L})$ the initial conditions are described by $u(p) = y_0 + Lp$. Again, the solution of (52) is denoted by $y(t, p)$. This problem is well-posed only if q boundary conditions are given, i.e. $g \mapsto \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^q$, and $q > n$ is possible if $\tilde{q} > 0$.

The structure of (52) covers ordinary shooting only with initial values ($\tilde{q} = 0$) and shooting only with parameters ($\hat{q} = 0$) in a consistent way. The same is true for the following inexact Newton method for (51). It is described for the case of a satellite peer method satisfying

$$c_1 = \ldots = c_q = 1 = c_s, \; r_s = 0, \quad s > q,$$

and where the square matrix $\hat{R}$ is nonsingular, see §2.2. Obviously the two-step peer methods need starting values $Y_{0j} = y(t_0 + h_0 c_j, p + \rho r_j)$, $j = 1, \ldots, s$ in the first time interval which may be computed by some Runge-Kutta method as in [13] for stages with $c_j \neq 0$. In fact, for the peer method of order 2 from Section 2.3 explicit Euler steps are sufficient,

$$Y_{0j} = u(p + \rho r_j) + h_0 c_j f(u(p + \rho r_j), p + \rho r_j), \; 1 \leq j \leq s. \tag{53}$$

These starting approximations are used in step 2 of the following inexact full-step Newton method:

$$\begin{array}{lll}
1. & \text{choose} & h_0 > 0, \rho > 0, \text{ and initial guess } p, \\
2. & \text{set} & Y_{0j} \cong y(t_0 + h_0 c_j, p + \rho r_j),\ j = 1, \ldots, s, \\
3. & \text{apply peer method (2) until } t_{m+1} = t_{end}, & \\
4. & \text{compute} & DY := \frac{1}{\rho}(Y_m^{\mathsf{T}} E_q - Y_{ms} \mathbb{1}_q^{\mathsf{T}}), \\
 & \text{and} & J := g_u L\hat{R} + g_v \cdot DY, \\
5. & \text{solve} & Jd = -g(u(p), Y_{ms}) \text{ for } d, \\
6. & \text{update} & p := p + \hat{R}d \text{ and continue with step 2.}
\end{array} \tag{54}$$

In principle, step 4 corresponds to (39) with $Q = E_q$ and $\hat{R} = RE_q$. However, the inversion of the matrix $\hat{R}$ can be saved by a reformulation of the parameter update. The Jacobian of the function $p \mapsto g(y_0 + Lp, y(t_{end}, p)))$ is given by $g_u \cdot L + g_v \cdot \partial y/\partial p \cong g_u \cdot L + g_v \cdot DY \cdot \hat{R}^{-1} = J\hat{R}^{-1}$ with the notations of (54). However, since the inverse of the Jacobian $(J\hat{R}^{-1})^{-1} = \hat{R}J^{-1}$ is required the matrix $\hat{R}$ may be moved to the update step 6 by $\Delta p = Rd$. In the following subsections algorithm (54) is applied to some simple test problems.

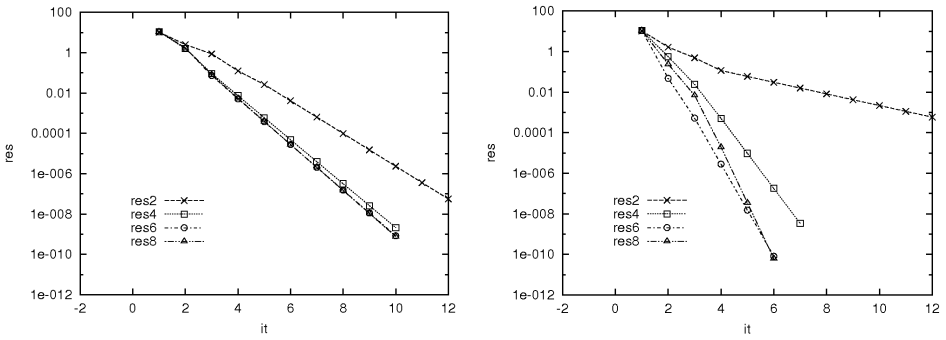

Figure 1: Iteration for pendulum with $\varrho = 10^{-1}$ (left) and $\varrho = 10^{-3}$ (right)

5.1 A pendulum problem

The physical pendulum exhibits periodic motions of arbitrarily large periods. As an example we consider the boundary value problem

$$y''(t) + \sin(y(t)) = 0, \quad y(0) - y'(0) = 1,\ y(T) + y'(T) = 0. \tag{55}$$

One of the boundary conditions is inhomogeneous to eliminate the trivial solution. For $T = 6$ the problem has a solution with $y(0) = 1 + y'(0) \doteq 1.57673$. The boundary conditions are separated but not in Dirichlet form. Thus, shooting with two parameters $y(0) = y_1(0) = p_1$ and $y'(0) = y_2(0) = p_2$ may be used. The shooting method (54) was applied to this problem with the satellite peer method of order 3 and $s = 5$ stages with stepsize control and a cardinal basis in parameter space, i.e. $\hat{R} = I_2$. Newton's iteration starts with $p = (1, 2)^{\mathsf{T}}$. All computations are performed on an Intel-i7 PC in a GNU Fortan90 implementation (gfortran).

In the first test we are looking for a sensible choice of the parameter stepsize ϱ in relation to the integration tolerance *tol*. Here, we display the residual in the boundary conditions in Newton step *it* for 4 different integration tolerances $tol \in \{10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}\}$ with fixed ϱ. The line labeled 'res$x$' belongs to $tol = 10^{-x}$. Figure 1 shows the results for $\varrho = 10^{-1}$ on the left and for $\varrho = 10^{-3}$ on the right. The first principal observation is that the approximation (39) of the parameter derivative is accurate enough for Newton's iteration to converge at all. Also, the speed of convergence improves with smaller ϱ for small tolerances, but suffers if ϱ is too small. From this and other tests we deduce that ϱ should decrease with *tol* but should not become too small. Hence, in the light of the results of Section 4 we used the following heuristic for choosing the parameter stepsize

$$\varrho = a\sqrt{tol} + 10^{-4}. \tag{56}$$

The constant a may be adapted to the actual problem. Figure 2 shows the behavior of the iteration (54) for the same tolerances with this choice (56) and $a = 0.5$. It is seen that the speed of convergences improves for smaller tolerances. In the test the iteration was performed down to quite small norms of the Newton update $\|d\|_2 \leq 10^{-9}$ in order to have a clear impression of the

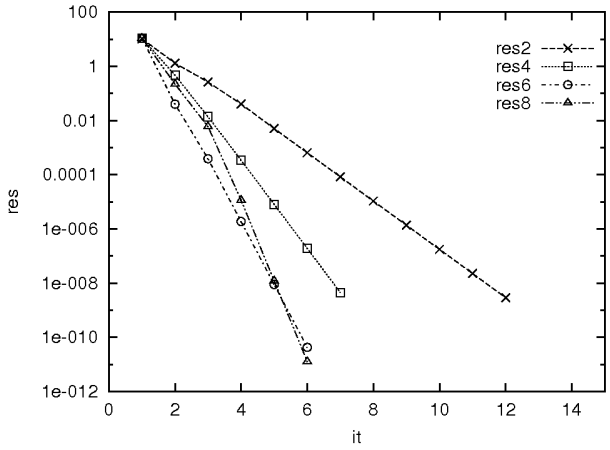

Figure 2: Iteration for pendulum with ϱ from (56), $a = 0.5$.

speed of convergence. In practice however, no stopping criterion much below the integration tolerance would be used. In general, for inexact Newton methods decreasing the integration tolerance with the progress of the iteration is an obvious option.

5.2 Time-periodic solutions of the Brusselator

A well-known problem with periodic solutions is the Brusselator which models oscillating chemical reactions with two species. It is discussed frequently in the literature as an ordinary differential equation wit $n = 2$ species or as a reaction-diffusion equation in one or several space dimensions. As an ODE ($d_1 = d_2 = 0$) or as a parabolic equation for $y_j(t, x)$, $j = 1, 2$, $x \in [0, 1]$, it is described by the system

$$\begin{aligned} y_1' =& d_1 \frac{\partial^2 y_1}{\partial x^2} + \alpha - (\beta + 1) y_1 + y_1^2 y_2, \\ y_2' =& d_2 \frac{\partial^2 y_2}{\partial x^2} + \beta y_1 - y_1^2 y_2, \end{aligned} \tag{57}$$

where (α, β) are parameters and $d_1, d_2 > 0$ diffusion constants. The Brusselator ODE has an equilibrium at $\bar{y} = (\alpha, \beta/\alpha)^\mathsf{T}$ which is unstable for $\beta > \alpha^2 + 1$. In the parabolic case these values will be used in Dirichlet boundary conditions

$$y_1(t, 0) = y_1(t, 1) = \alpha, \quad y_2(t, 0) = y_2(t, 1) = \frac{\beta}{\alpha}. \tag{58}$$

Since the periodic orbits are limit cycles and asymptotically stable they may simply be computed by solving the initial value problem over sufficiently long times. However, shooting is more efficient. Two different problem types may be considered.

1. Compute a periodic orbit through some given point $y_0 \in \mathbb{R}^2$. Here a solution of the boundary value problem with

$$y(0, p) = y_0, \quad y(T, p) = y_0,$$

is computed by adjusting the two parameters $p = (\alpha, \beta)^\mathsf{T}$ of the Brusselator itself. In the boundary value problem formulation (51) this means $\tilde{q} = 2$ and $g(u, v) = v - y_0$.

2. Compute an orbit for given parameters α, β by searching for the initial condition $y(0) = u$ such that the periodicity condition

$$y(T) - y(0) = 0$$

is satisfied. This problem corresponds to the definition $g(u, v) = -u+v$ in (51) and (52) with $p = y(0)-\tilde{y}_0$ and some guess $\tilde{y}_0$. Unfortunately, for autonomous problems the Jacobian of the boundary equation has a nontrivial kernel since the monodromy matrix has an eigenvalue one. However, the kernel vector is $f(y(T)) = y'(T)$ since each shifted solution $y(\tau + t)$ solves the problem, too, [15]. In order to avoid this singularity the system in the Newton step in (54) is complemented by a phase condition $f_0^\mathsf{T} p = 0$ [15] with a fixed vector $f_0 = f(T, \tilde{y}(T))$ from the endpoint of an initial trajectory. Then, also T is considered unknown and the Newton step for the system $y(T, u) - u = 0$, $u = \tilde{y}_0 + p$, is of the form

$$\begin{pmatrix} DY - I & f(y(T)) \\ f_0^\mathsf{T} & 0 \end{pmatrix} \begin{pmatrix} d \\ \delta \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix} \tag{59}$$

with a nonsingular Jacobian near the solution.

As the first application of the peer method an ODE orbit through the point $y_0 = (1.8, 1.8)$ is computed with prescribed period length $T = 7.16$. Figure 3 shows the convergence of the undamped iteration (54) with the peer method of order 3, $s = 5$, starting parameters $p_0 = (1, 3)^\mathsf{T}$, and for tolerances $tol = 10^{-2j}$, $j = 1, \ldots, 4$. The parameter off-step is (56) with $a = 0.2$. A closed orbit is found at $p \doteq (1.556, 3.973)$. The iteration converges for $tol \leq 10^{-4}$ and the speed of convergence improves for smaller tolerances. For the first few iterates convergence seems to be almost quadratic.

A more demanding test for the peer methods is the one-dimensional Brusselator (57) with boundary conditions (58). We use the data $d_1 = 0.008$, $d_2 = 0.004$ and difference discretization in space with only 31 interior points from [9] since we want to get only a mildly stiff ODE for the explicit peer methods and need some reference data. The problem has dimension $n = 62$ and shooting with initial values and fixed equation parameters $(\alpha, \beta) = (2, 5.45)$ uses the satellite method with 3 central stages and $q = 62$ satellites, i.e. $s = 65$ stages. The Newton iteration starts with constant functions $y_1(0, x) = 2.5$ and $y_2(0, x) = 3.2$ and period length $T = 3.4$. A more complete set of tolerances $tol = 10^{-j}$, $j = 3, \ldots, 8$ was used and ϱ from (56) with $a = 0.2$, again. Since the period length of the orbit depends on the accuracy of the integration the period T is updated by $T := T + \delta$ in the Newton step (59) to allow for small final residuals. The results are shown in Figure 4. For weak tolerances the residuals stagnate somewhere near the tolerance level but for $tol \leq 10^{-6}$ they converge fast. The final period length for $tol = 10^{-8}$ was $T = 3.434865839$.

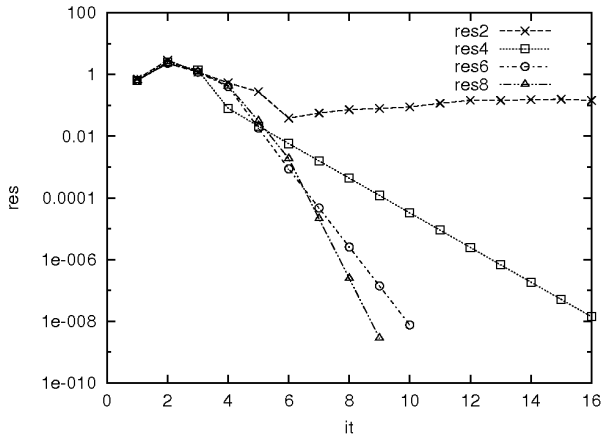

Figure 3: Iteration for Brusselator ODE, parameter search

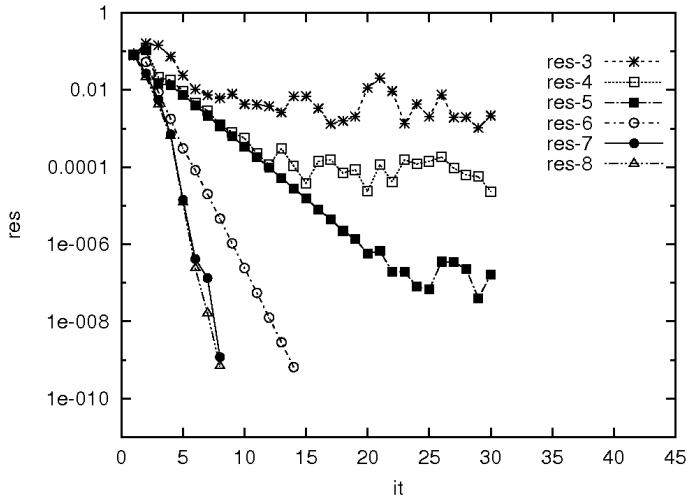

Figure 4: Iteration for 1D-Brusselator, initial value search

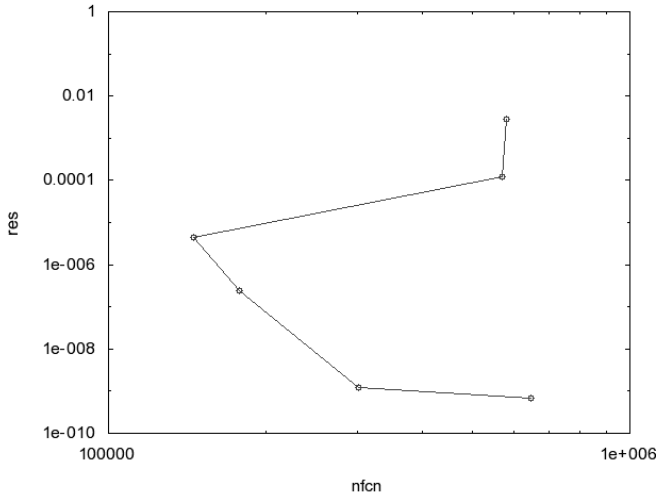

Figure 5: Efficiency for 1D-Brusselator

Now that the effectiveness of the scheme has been established it is also of interest from a practical point of view to ask for which integration tolerance some required accuracy of the solution is obtained most efficiently. To this end the Newton iteration is stopped if the norm of the update d is well below the integration tolerance and the number of function evaluations for all Newton steps is counted. With the same tolerances and stepsizes ϱ as before Newton's iteration is performed until $\|d\| \leq 0.1 \cdot tol$ or for at most 90 steps. Figure 5 shows the final accuracy. Obviously, for tolerances $tol \leq 10^{-5}$ the effort is an increasing function of tol^{-1}.

6 Parameter estimation

In practical applications the structure of the ODE system

$$y'(t,p) = \tilde{f}(y(t,p),\tilde{p}),\ t \in [t_0, t_{end}],$$

may be known but not the actual parameters $\tilde{p} \in \mathbb{R}^{\tilde{q}}$ of an observed solution. Here, the parameter values have to be estimated from measurements in a calibration phase. We assume that some incomplete data from the solution can be acquired at times $\tau_i \in [t_0, t_{end}]$, $i = 1, \ldots, \mu$ (e.g. positions but no velocity in the pendulum problem). For ease of implementation these metering points are a small subset of the integration points of the peer method (2) but $\tau_i = \tau_j$, $i \neq j$, is allowed. Then, the deviation of some given solution $y(t,p)$ from the measurements is described by scalar functions $g_i(y(\tau_i,p))$ like $g_i(v) = d_i^{\mathsf{T}}(v - \tilde{y}_i)$ where $d_i^{\mathsf{T}}\tilde{y}_i$ represents the incomplete measurement at τ_i. Nonlinear versions of these functions g_i are possible, of course, but have not been considered. Although a multiple shooting approach may be more appropriate here [6], for the sake of simplicity we consider again single shooting. As in (52) the unknown initial value $y(t_0) = \hat{p} \in \mathbb{R}^n$ may be considered as a set of additional parameters and with $p^{\mathsf{T}} = (\tilde{p}^{\mathsf{T}}, \hat{p}^{\mathsf{T}})$ we denote by $y(t,p)$ the solution of the ODE (1) with

initial value $\hat{p}$. In a least-squares approach the unknown parameters and initial values are characterized by the minimization problem

$$\min_p \sum_{i=1}^{\mu} \big\| g_i\big(y(\tau_i,p)\big)\big\|^2. \tag{60}$$

The Gauss-Newton step (41) at an actual guess p with solutions $y_i = y(\tau_i,p)$ is

$$\min_{\Delta p} \sum_{i=1}^{\mu} \big\| g_i(y_i) + g_i'(y_i)\frac{\partial y(\tau_i)}{\partial p}\Delta p\big\|^2.$$

Its minimizer Δp satisfies

$$\Big(\sum_{i=1}^{\mu} N_i^{\mathsf{T}} N_i\Big)\Delta p = -\sum_{i=1}^{\mu} N_i^{\mathsf{T}} g_i(y_i). \tag{61}$$

with $N_i := g_i'(y_i)\,\partial y(\tau_i)/\partial p$ and is the least-squares solution of the overdetermined system

$$\begin{pmatrix} N_1 \\ \vdots \\ N_\mu \end{pmatrix} \Delta p = -\begin{pmatrix} g_1(y_1) \\ \vdots \\ g_\mu(y_\mu) \end{pmatrix}, \quad \Delta p = \begin{pmatrix} \Delta\tilde{p} \\ \Delta\hat{p} \end{pmatrix},$$

if its matrix has full rank $q = \tilde{q} + n$. The number of measurements μ has to be sufficiently large to determine all parameters. With exactly one datum measured at any point τ_i the matrix in (61) is a sum of rank-1-matrices and $m \geq \tilde{q} + n = q$ is required for a nonsingular system. Still, convergence of the simple Gauss-Newton iteration may be erratic and should be enhanced by line-searches. In a simple line search only the residual function in (60) has to be computed and no derivatives. Here, the satellite peer methods have an advantage since the central stages $Y_{m,q+1},\ldots,Y_{ms}$, see (26), required for the evaluation of the residual do not depend on the satellite stages. So, during line search the q satellite stages can be spared. Summarizing, a simple inexact Gauss-Newton iteration with simple line search is described by:

1.	choose	$h_0 > 0, \rho > 0,$ and initial guess p,
2.	let	$\lambda := 0,\ d := 0,\ res0 := \infty,\ it := 1$
3.	let	$p^* := p + \lambda d,$
	and	$Y_{0j} \cong y(t_0 + h_0 c_j, p^* + \rho r_j),\ j = 1,\ldots,s,$
4.	for $i = 1,\ldots,\mu$ apply peer method (2) until $t_{m+1} = t_i$ and	
	4a. let	$res_i := -g_i(Y_{ms}),$
	4b. compute	$DY := \frac{1}{\rho}(Y_m^{\mathsf{T}} E_q - Y_{ms}\mathbb{1}_q^{\mathsf{T}}),$
	4c. let row	$e_i^{\mathsf{T}} J := g_i'(Y_{ms})\cdot DY,$
5.	if	$\|res\| < res0$ then
	5a. accept	$p := p^*,$
	5b. compute	$d := -J^+ res,$
	5c. set	$\lambda := 1, res0 =: \|res\|,$
6.	else	$\lambda := \lambda/2,$
7.	let	$it := it + 1,$ continue with step 3.

(62)

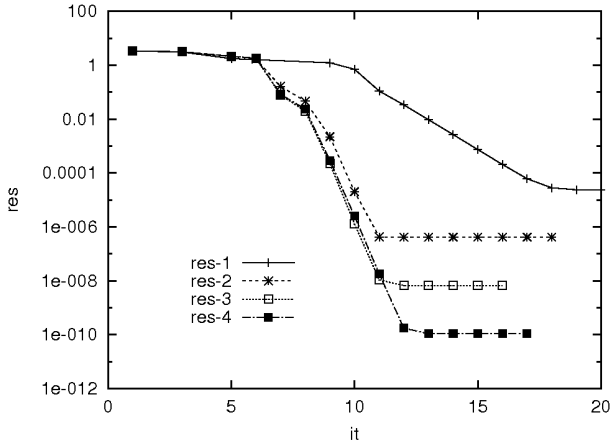

Figure 6: Convergence in parameter identification for Brusselator

The method is applied to a nonperiodic numerical solution of the Brusselator ODE (57) with exact parameters $\tilde{p} = (1, 3)^\mathsf{T}$, initial values $\hat{p} = y(0) = (1.8, 1.8)^\mathsf{T}$, and end time $t_{end} = 7.16$. Only the first component $e_1^\mathsf{T}\hat{y}(t_i)$ of a high-accuracy solution obtained by DOPRI5 [8] is saved at $\mu = 10$ points $\tau_i = t_{end}/10$, $i = 1, \ldots, \mu$, not including t_0. In order to hit the metering points time integration is performed with the order 3 method and fixed stepsizes. Thus, shooting uses $q = 4$ parameters and $s = 7$ stages. In Figure 6 the convergence of the iteration (62) is shown with 4 different time stepsizes $h = t_{end}4^{-j}/100$, $j = 1, \ldots, 4$. The iteration number it on the horizontal axis counts all time integrations, see (62). Damped iterations are sometimes used in the beginning and can be identified through missing labels. For stepsizes $h \leq 0.005$ convergence is almost independent of the stepsize. However, residuals level off at the different accuracy levels of time integration. In fact, the discretization error characterizes the noise level in the measurements $g_i(y)$ with respect to the high-accuracy solution $\hat{y}$.

7 Conclusion

It was demonstrated that peer two-step methods can be extended easily to obtain sensitivity information for parameter-dependent ordinary differential equations. By adding one stage for each parameter direction peer methods can approximate the solution manifold of such equations with high order on the central trajectory and with first-order accuracy in the parameter derivatives. Effort and memory requirements are much smaller than in the standard approach using the integration of variational equations. A flexible and efficient subclass, the satellite configuration, has been identified where an arbitrary number of parameter stages may be used at runtime and the computation of the satellite stages is quite cheap. A convergence analysis for an inexact Gauss-Newton iteration and several numerical tests on different problem types show that the sensitivity information from the peer method is accurate enough for good convergence of the iteration. This paper is an introduction to the design of peer methods for

parameter-dependent initial value problems with several application areas like shooting and parameter estimation, no realistic comparison with other existing approaches was intended.

References

[1] U.M. Ascher, R.M.M. Mattheij, R.D. Russell, *Numerical solution of boundary value problems for ordinary differential equations*, SIAM, Philadelphia, 1995.

[2] S. Beck, R. Weiner, H. Podhaisky, B.A. Schmitt, *Implicit peer methods for large stiff ODE systems*, to appear in J. Appl. Math. Comp.

[3] H. G. Bock, *Numerical treatment of inverse problems in chemical reaction kinetics*. In K.H. Ebert, P. Deuflhard, and W. Jäger, editors, Modelling of Chemical Reaction Systems, volume 18 of Springer Series in Chemical Physics, p. 102-125. Springer, 1981.

[4] H. G. Bock. *Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen.* Bonner Mathematische Schriften, **187**, Bonn, 1987.

[5] H. G. Bock, E. Kostina, and J.P. Schlöder, *On the Role of Natural Level Functions to Achieve Global Convergence for Damped Newton Methods*, in M.J.D. Powell and S. Scholtes, eds., *System Modelling and Optimization. Methods, Theory and Applications*, Kluwer, Boston, pp. 51–74, 2000.

[6] H.G. Bock, E.A. Kostina, and J.P. Schlöder, *Numerical Methods for Parameter Estimation in Nonlinear Differential Algebraic Equations*, *GAMM Mitteilungen* 30(**2**), pp. 376–408, 2007.

[7] H.G. Bock, E. Kostina, and J.P. Schlöder, *How Good are "Solutions" of "Large Residual" Parameter Estimation Problems?*, in preparation, 2011.

[8] E. Hairer, S.P. Norsett, G. Wanner, *Solving Ordinary Differential Equations I, Nonstiff Problems*, Springer, 1987.

[9] K. Lust, D. Roose, A. Spence, A.R. Champneys, *An adaptive Newton-Picard algorithm with subspace iteration for computing periodic solutions*, SIAM J. Sci. Comput. 19 (1998), 11888-1209.

[10] B.A. Schmitt, R. Weiner, *Parallel Two-Step W-Methods with Peer Variables*, SIAM J. Numer. Anal. 42 (2004), 265-282.

[11] B.A. Schmitt, R. Weiner, K. Erdmann, *Implicit parallel peer methods for stiff initial value problems*, Appl. Numer. Math. 53 (2005).

[12] B.A. Schmitt, R. Weiner, H. Podhaisky, *Multi-implicit peer two-step W-methods for parallel time integration*, BIT 45 (2005), 197-217.

[13] B.A. Schmitt, R. Weiner, S. Jebens, *Parameter optimization for explicit parallel peer two-step methods*, Appl. Numer. Math. 59 (2009), 769-782.

[14] R. Serban, A.C. Hindmarsh, CVODES, the sensitivity-enabled ODE solver in SUNDIALS, Proceedings of IDETC/CIE, ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference September 24-28, 2005, Long Beach, California, USA.

[15] R. Seydel, *Practical bifurcation and stability analysis*, 3rd ed., Springer, New York, 2010.

[16] J. Stoer, R. Bulirsch, *Numerical mathematics 2. An introduction*, 5th ed., Springer, 2005.

[17] R. Weiner, K. Biermann, B.A. Schmitt, H. Podhaisky, *Explicit two-step peer methods*, Comp. Math. Appl. 55 (2008), 609-619.