Stabilization of DAEs and invariant manifolds

Uri M. Ascher*
Department of Computer Science
University of British Columbia
Vancouver, British Columbia
Canada V6T 1Z2
ascher@cs.ubc.ca

Hongsheng Chin †
Department of Mathematics
University of British Columbia
Vancouver, British Columbia
Canada V6T 1Z2

Sebastian Reich ‡
Institut für Angewandte Mathematik und Stochastik
Mohrenstraße 39
Berlin, O–1086, Germany
na.reich@na-net.ornl.gov

November 15, 1994

Revised, May 1993

Abstract

Many methods have been proposed for the stabilization of higher index differential-algebraic equations (DAEs). Such methods often involve constraint differentiation and problem stabilization, thus obtaining a stabilized index reduction. A popular method is Baumgarte stabilization, but the choice of parameters to make it robust is unclear in practice.

Here we explain why the Baumgarte method may run into trouble. We then show how to improve it. We further develop a unifying theory for stabilization methods which includes many of the various techniques proposed in the literature. Our approach is to (i) consider stabilization of ODEs with invariants, (ii) discretize the stabilizing term in a simple way, generally different from the ODE discretization, and (iii) use orthogonal projections whenever possible.

The best methods thus obtained are related to methods of coordinate projection. We discuss them and make concrete algorithmic suggestions.

*The work of this author was partially supported under NSERC Canada Grant OGP0004306.

†The work of this author was partially supported under NSERC Canada Grants OGP0004306 and OGP0000236

‡The work of this author was partially supported by the German Science Foundation and was finished while the author was visiting Simon Fraser University, Burnaby.

1 Introduction

Many methods have been proposed for the stabilization of higher-index differential-algebraic equations (DAEs), see [10], [4] and references therein. Such methods often involve constraint differentiation and problem stabilization, thus obtaining a stabilized index reduction.

The basic reason for replacing the original problem by one with lower index is that the reformulated problem is presumably easier, or more convenient, to solve numerically. For instance, in the case of incompressible Navier-Stokes equations, which yield a semi-explicit, pure (Hessenberg) index-2 DAE in time, a staggered finite difference grid or some potentially inconvenient mixed finite element spaces are needed for the spatial discretization. If instead one differentiates the constraint of zero divergence, one obtains the pressure-Poisson equation (cf. [14]), and now a nonstaggered grid or a “normal” finite element discretization can be used. In the case of multibody systems with holonomic constraints (with closed loops), the trouble is simply that the DAE has index 3; generally, robust methods for DAEs of index > 2 (even in pure semi-explicit form) are not known (and with good reason: such problems are ill-posed, see [15, 4]). However, it has long been recognized that a direct constraint differentiation, especially when it is repeated more than once, leads to (mild) instabilities for long-time numerical integrations. The effect is often measured by the “drift” – the error in the original constraint (which is now part of an invariant of the integrated ODE but is not satisfied exactly by the discretization scheme) grows. Hence, some stabilization is required.

A popular stabilization technique is Baumgarte’s [7]. To be specific, consider the DAE of order m and pure index $m+1$

$$\begin{aligned} \mathbf{x}^{(m)} &= \mathbf{f}(\mathbf{x}, \mathbf{x}', \ldots, \mathbf{x}^{(m-1)}, t) - B(\mathbf{x}, t)\mathbf{y} && (1.1a) \\ \mathbf{0} &= \mathbf{g}(\mathbf{x}, t) && (1.1b) \end{aligned}$$

where $G = \mathbf{g_x}$ is generally rectangular and GB is nonsingular for all t, $0 \le t \le t_f$. (We will consider cases where $m = 1$ or 2.) A direct m-fold differentiation of the constraints (1.1b) yields

$$\begin{aligned} \mathbf{g}^{(m)} &= \frac{d^m \mathbf{g}(\mathbf{x}(t), t)}{dt^m} = \mathbf{0} && (1.2a) \\ \mathbf{g}(\mathbf{x}(0), 0) &= \frac{d}{dt}\mathbf{g}(\mathbf{x}(0), 0) = \ldots = \frac{d^{m-1}}{dt^{m-1}}\mathbf{g}(\mathbf{x}(0), 0) = \mathbf{0} && (1.2b) \end{aligned}$$

and the DAE (1.1a), (1.2a) now has index 1. The algebraic unknowns $\mathbf{y}$ can therefore be eliminated and an ODE

$$\mathbf{x}^{(m)} = \tilde{\mathbf{f}}(\mathbf{x}, \mathbf{x}', \ldots, \mathbf{x}^{(m-1)}, t) \qquad (1.3)$$

is obtained, for which the original constraint together with its first $m-1$ derivatives is an invariant. But this causes drift difficulties, so a generalization of Baumgarte’s method [7] replaces (1.2a) with the equation

$$\sum_{j=0}^{m} \alpha_j \frac{d^j}{dt^j}\mathbf{g}(\mathbf{x}(t), t) = \mathbf{0} \qquad (1.4)$$

where α_j are chosen so that $\alpha_m = 1$ and the roots of the polynomial

$$\sigma(\tau) = \sum_{j=0}^{m} \alpha_j \tau^j$$

are all negative. For instance, one may choose

$$\sigma(\tau) = (\tau + \gamma)^m \tag{1.5}$$

for some $\gamma > 0$. Essentially, what this does is turn the invariant manifold from being just stable, or even mildly unstable, to being asymptotically stable (attracting).

The apparent conceptual simplicity of the Baumgarte stabilization technique must be considered a major reason for its popularity in engineering applications. But the practical choice of parameters (γ in (1.5)) to make it robust is widely regarded as unknown, despite many attempts (see, e.g., [18]). We now give three indications to explain why this parameter choice is indeed inherently difficult and how the situation can be improved.

One reason that may have made the search for a good γ difficult is that the form (1.4),(1.5) suggests that γ should be independent of the discretization method and step size h. (Indeed, the numerical solution of the obtained ODE is often computed using standard software.) But such a conclusion is not clear in practice. In fact, the results of this paper suggest that the optimal γ may well depend on both the step size and the discretization method.

Another difficulty with Baumgarte's technique arises when applying it directly to a problem (1.1) with $m \geq 2$, as is the usual practice in multibody systems simulation (where $m = 2$). One would have hoped that the larger γ is the better the stabilization is, because the manifold becomes "more attractive". But when $\gamma \to \infty$ such that $\gamma h \gg 1$, the discretized problem is close to a discretization of the index-$(m+1)$ DAE and therefore numerical stability difficulties arise. In this paper we consider stabilizations which reduce in the limit to an index-2 DAE.

Finally, let us consider the simplest, index-2 case, i.e. let $m = 1$ in (1.1). It can be easily verified that the Baumgarte technique is equivalent to reformulating the original DAE as

$$\mathbf{x}' = \tilde{\mathbf{f}}(\mathbf{x}, t) - \gamma B(GB)^{-1}\mathbf{g}(\mathbf{x}, t) \tag{1.6}$$

i.e. we add a stabilizing term to the ODE (1.3) which vanishes on the constraint manifold. Again we expect optimal error damping for any γ large enough, i.e., we want no deterioration in the solution error when γ is taken larger and larger for a fixed h. However, this cannot always be guaranteed either: Example 2 in [4] demonstrates that taking γ too large may yield poor results as well, when $\|GB\| \ll \|G\|\|B\|$, i.e. when G and B are almost orthogonal. Much better results are obtained in such a case if we replace (1.6) by

$$\mathbf{x}' = \tilde{\mathbf{f}}(\mathbf{x}, t) - \gamma G^T(GG^T)^{-1}\mathbf{g}(\mathbf{x}, t) \tag{1.7}$$

Experiments with (1.7) for Example 2 of [4] show no deterioration in the error as γ is increased. [1]

Remarks

[1] The importance of a lower γ value, if it produces a sufficient stabilization effect, is that we may then solve the ODE using a nonstiff method if there is no other source of stiffness.

1. It is important to make a distinction between the stabilizing reformulations which we are considering here and general regularization methods. In the latter one perturbs the problem to be solved (e.g. by adding artificial viscosity or artificial compressibility to a fluid flow problem, etc.) to obtain a nearby problem which is easier to solve. The solution of the perturbed problem is not the same as that of the original one, hence the perturbation must be small (corresponding to γ being very small or very large above). The stabilizing reformulations considered here, on the other hand, have the same solution as the original problem before discretization. Thus, γ need not be restricted to very small or very large values. The conditioning of the stabilized problem does not necessarily depend on the perturbation parameter as it does in the regularization case.

2. The basic question whether an invariant should be imposed in the course of computing an approximate solution does not appear to have an immediate or unique answer in practice. Of course a growing drift cannot be tolerated, but if the drift remains reasonably small then the corresponding approximate solution is not necessarily less accurate than one which is projected onto the invariant manifold. Examples can be found in [5]. For instance, experiments with the method of characteristic strips for the shape-from-shading problem yield a similar conclusion for that application [8]. Also, a symplectic integrator for a Hamiltonian system may do a better job without constraint projection [20]. On the other hand, setting the drift in the holonomic constraints to 0 may be important for display purposes in vehicle simulation (more important than making the full solution error extremely small). Also, in [3] constraint (or coordinate-) projection improves the convergence order of the discretization scheme.

   Our concern in this paper is that of stability (and efficiency), however. The key question is then whether the ODE stability remains essentially the same around the manifold as it is on it. If the stability deteriorates once the solution is off the constraint manifold then there is ample reason to enforce its return (or at least getting closer) to the manifold, either by means of stabilization with a large γ or by outright projection.

Our first task in this paper is to study stabilization techniques in a general framework. We analyze nonlinear problems directly, unlike in [4]. For simplicity of exposition we will consider only autonomous problems. Thus, we reformulate (at least in principle) the higher index (now autonomous) DAE (1.1) as a first order ODE (cf. (1.3))

$$\mathbf{z}' = \hat{\mathbf{f}}(\mathbf{z}) \tag{1.8}$$

with an invariant

$$\mathbf{0} = \mathbf{h}(\mathbf{z}) \tag{1.9}$$

where

$$\mathbf{z} = \begin{pmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_{m-1} \\ \mathbf{z}_m \end{pmatrix} = \begin{pmatrix} \mathbf{x} \\ \vdots \\ \mathbf{x}^{(m-2)} \\ \mathbf{x}^{(m-1)} \end{pmatrix}, \qquad \hat{\mathbf{f}}(\mathbf{z}) = \begin{pmatrix} \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_m \\ \tilde{\mathbf{f}}(\mathbf{z}) \end{pmatrix}, \qquad \mathbf{h}(\mathbf{z}(t)) = \begin{pmatrix} \mathbf{g}(\mathbf{x}(t)) \\ \frac{d}{dt}\mathbf{g}(\mathbf{x}(t)) \\ \vdots \\ \frac{d^{(m-1)}}{dt^{(m-1)}}\mathbf{g}(\mathbf{x}(t)) \end{pmatrix} \tag{1.10}$$

and consider in the next section the stabilization families

$$\mathbf{z}' = \hat{\mathbf{f}}(\mathbf{z}) - \gamma F(\mathbf{z})\mathbf{h}(\mathbf{z}) \tag{1.11}$$

where $H = \mathbf{h}_{\mathbf{z}}$ (for $m = 1$ in (1.1), $H \equiv G$), and where

$$F = D(HD)^{-1} \tag{1.12}$$

or

$$F = D \tag{1.13}$$

with $D(\mathbf{z})$ smooth such that HD is nonsingular (indeed, $\|HD\|\|(HD)^{-1}\|$ should be nicely bounded) for each $\mathbf{z}$. If (1.13) is used then HD is further required to be uniformly positive definite. The best choice for D from the stability standpoint is often $D = H^T$, but Baumgarte's technique for $m = 1$ is obtained with $D = B$ in (1.12). For $m > 1$, Baumgarte's technique is not in the family (1.11). We obtain asymptotic stability results which include persistence under small perturbations.

In Section 3 we then consider the numerical discretization of (1.11). We make the simple but important observation that the stabilizing term need not be discretized by the same method as the ODE and show that simple forward and backward Euler schemes for the stabilizing term maintain the accuracy of a high order method applied to the underlying ODE part. So does a simple modification of both these schemes which turns out to be closely related to coordinate projection. Moreover, for the latter scheme and (1.12) the choice $\gamma = h^{-1}$ is then found to be close to optimal. We recommend this method (i.e. (3.7) with $\alpha = 1$) for practical use.

In Section 4 we then apply our results to DAEs of index 2 and 3, and in particular to constrained mechanical systems. Conclusions and discussion are offered in Section 5.

2 Stabilization of Invariants

In this section and in Section 3 we consider an ODE system

$$\mathbf{z}' = \hat{\mathbf{f}}(\mathbf{z}) \tag{2.1}$$

with an invariant set $\mathcal{M}$ given by

$$\mathbf{0} = \mathbf{h}(\mathbf{z}) \tag{2.2}$$

where both $\mathbf{h} : U \subseteq \mathcal{R}^n \to \mathcal{R}^{n_y}$ and $\hat{\mathbf{f}} : U \subseteq \mathcal{R}^n \to \mathcal{R}^n$ are assumed to be sufficiently smooth, and $H(\mathbf{z}) = \mathbf{h}_{\mathbf{z}}(\mathbf{z})$ has a full row rank.

We distinguish between the cases when (i) the mapping $\mathbf{h}$ in (2.2) is an integral invariant of (2.1), i.e., $H(\mathbf{z})\hat{\mathbf{f}}(\mathbf{z}) = \mathbf{0}$ for all $\mathbf{z} \in U$; and (ii) $\mathbf{h}$ is not an integral invariant: $H(\mathbf{z})\hat{\mathbf{f}}(\mathbf{z}) = \mathbf{0}$ holds only on $\mathcal{M}$. It can be easily verified that, upon using index reduction as previously described, the index-2 DAE (1.1) with $m = 1$ yields an integral invariant whereas higher index DAEs ((1.1) with $m \geq 2$) do not.

We consider the family of stabilization methods

$$\mathbf{z}' = \hat{\mathbf{f}}(\mathbf{z}) - \gamma F(\mathbf{z})\mathbf{h}(\mathbf{z}) \tag{2.3}$$

with F as described in (1.12) or (1.13).

Proposition 2.1 *Let the mapping* $\mathbf{h}$ *in* (2.2) *be an integral invariant of* (2.1). *Then the manifold* $\mathcal{M}$ *is an asymptotically stable invariant manifold of the ODE* (2.3) *for all* $\gamma > 0$. *The flow of* (2.3) *on* $\mathcal{M}$ *reduces to the flow of* (2.1) *restricted to* $\mathcal{M}$.

Proof. Multiply the ODE (2.3) by $H(\mathbf{z})$. Introduce the new variable $\mathbf{v} = \mathbf{h}(\mathbf{z})$. This yields the ODE $\mathbf{v}' = -\gamma \mathbf{v}$ for (1.12) and $\mathbf{v}' = -\gamma HD\mathbf{v}$ for (1.13). Both of these ODEs are uniformly asymptotically stable. □

In the general case the situation is a bit more complicated. We obtain

Proposition 2.2 *Let the manifold* $\mathcal{M}$ *be an invariant manifold of the ODE* (2.1), *and assume that there exist positive constants* γ_o *and* δ *such that*

$$\|H(\mathbf{z})\hat{\mathbf{f}}(\mathbf{z})\|_2 \leq \gamma_0 \|\mathbf{h}(\mathbf{z})\|_2 \tag{2.4}$$

for all $\mathbf{z}$ *in a* δ*-neighborhood of* $\mathcal{M}$. *In case of (1.13) assume also that* D *is scaled so that the smallest eigenvalue of* HD *is* ≥ 1. *Then the manifold* $\mathcal{M}$ *is an asymptotically stable invariant manifold of the ODE* (2.3) *for all* $\gamma > \gamma_o$. *The flow of* (2.3) *on* $\mathcal{M}$ *reduces to the flow of* (2.1) *restricted to* $\mathcal{M}$.

Proof. Introduce the Liapunov function $V(\mathbf{z}) = \mathbf{h}^T(\mathbf{z})\mathbf{h}(\mathbf{z})$. Then, using (2.3) and the proposition's assumption,

$$\begin{aligned}
V' &= 2\mathbf{h}(\mathbf{z})^T H(\mathbf{z})\mathbf{z}' \\
&= 2\mathbf{h}(\mathbf{z})^T H(\mathbf{z})[\hat{\mathbf{f}}(\mathbf{z}) - \gamma F(\mathbf{z})\mathbf{h}(\mathbf{z})] \\
&\leq 2(\gamma_0 \mathbf{h}^T\mathbf{h} - \gamma \mathbf{h}^T HF\mathbf{h}) \\
&\leq -2(\gamma - \gamma_0)V
\end{aligned}$$

This yields the claimed results. □

Remarks.

1. The assumption (2.4) is necessary. Consider, for example,

$$\hat{\mathbf{f}}(\mathbf{z}) = \begin{pmatrix} z_2 \\ -z_1 \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \frac{\mathbf{h}(\mathbf{z})^{1/3}}{r^2}, \qquad \mathbf{h}(\mathbf{z}) = r^2 - 1, \qquad r^2 = z_1^2 + z_2^2$$

   for which all the assumptions of Proposition 2.2 except for (2.4) hold, but the conclusion does not.

2. If $\mathbf{h}$ is not an integral invariant of (2.1) then we can make it so by modifying $\hat{\mathbf{f}}$. This can be done by considering in place of (2.1)

$$\mathbf{z}' = (I - F(\mathbf{z})H(\mathbf{z}))\hat{\mathbf{f}}(\mathbf{z}) \tag{2.5}$$

   where F satisfies (1.12), and applying the stabilization (2.3) to this ODE instead. For mechanical systems this corresponds to reducing the index to 2 by the technique of [13] before applying the Baumgarte-like stabilization (2.3).

3. Rewriting (2.3) as

$$\begin{aligned}
\mathbf{z}' &= \hat{\mathbf{f}}(\mathbf{z}) - D(\mathbf{z})\lambda \\
\mathbf{0} &= \mathbf{h}(\mathbf{z}) - \frac{1}{\gamma} HD(\mathbf{z})\lambda
\end{aligned} \tag{2.6}$$

in case of (1.12) and

$$\begin{aligned} \mathbf{z}' &= \hat{\mathbf{f}}(\mathbf{z}) - D(\mathbf{z})\lambda \\ \mathbf{0} &= \mathbf{h}(\mathbf{z}) - \frac{1}{\gamma}\lambda \end{aligned} \tag{2.7}$$

in case of (1.13), we see that as we let $\gamma \to \infty$ the method reduces to the index-2 DAE

$$\begin{aligned} \mathbf{z}' &= \hat{\mathbf{f}}(\mathbf{z}) - D(\mathbf{z})\lambda \\ \mathbf{0} &= \mathbf{h}(\mathbf{z}) \end{aligned} \tag{2.8}$$

This is the projected invariant method proposed by Gear [12] for the choice $D = H^T$; see also [4], [5].

4. As already mentioned at the beginning of this section, the invariant (1.9) is not an integral invariant of (1.8) for DAEs (1.1) with $m \geq 2$. Thus Proposition 2.2 applies in this situation. However, similar to Proposition 2.1, we obtain that $\mathcal{M}$ is asymptotically stable for all $\gamma > 0$ in this particular situation. We will come back to this fact in Section 4.

This gives us a unified picture of a large family of stabilization techniques. The conclusion that, at least before discretization, they all do act as stabilizers in the sense of this section agrees well with similar conclusions based on linear stability analysis proved in [4].

3 Discretization of the Stabilized ODE Formulation

As already observed in the introduction, the two terms on the right hand side of (2.3) differ substantially from each other, both in purpose ($-\gamma F\mathbf{h}$ is just a stabilization term) and in size. Hence it makes sense to apply different discretization schemes to them.

Let us consider the discretization of the ODE (2.1) by a one–step scheme which results in the time–h–map

$$\mathbf{z}_{n+1} = \phi_h^f(\mathbf{z}_n)\,. \tag{3.1}$$

This advances the solution from the approximate state $\mathbf{z}_n$ at $t = t_n$ to an approximate state $\mathbf{z}_{n+1}$ at $t_{n+1} = t_n + h$. We make the following assumptions on this discretization scheme:

Assumptions 3.1

1. ϕ_h^f is C^1
2. ϕ_h^f has order $p \geq 1$ on a bounded subset $K_2 \subset \mathcal{R}^n$.
3. $\phi_h^f(\mathbf{z}) = \mathbf{z}$ outside a bounded subset K_1, $K_2 \subset K_1 \subset \mathcal{R}^n$.

Of these assumptions, the first is self-explanatory. For the second and third, we will assume that a usual discretization scheme, say Runge-Kutta, is modified by a smooth cut–off function in such a way that $\phi_h^f(\mathbf{z}) = \mathbf{z}$ outside a bounded subset $K_1 \subset \mathcal{R}^n$ and $\mathbf{z}_{n+1} = \phi_h^f(\mathbf{z}_n)$ as above on a bounded subset $K_2 \subset K_1$. Note that this assumption is not severe since in numerical computations we are always interested in bounded solutions (K_2 can be defined by the overflow value in a given computational environment). However, this assumption allows us to restrict our considerations to the compact set $\bar{K}_1 = K_1 \cup \partial K_1$. We will need a compactness argument to apply certain persistence results for invariant manifolds in the theorems below.

Next we consider implicit and explicit discretizations of the stabilization term in (2.3).

1. First consider discretizing the ODE

$$\mathbf{z}' = -\gamma F(\mathbf{z})\mathbf{h}(\mathbf{z}) \tag{3.2}$$

by the stiffly stable backward Euler scheme, which results in:

$$\mathbf{z}_{n+1} = \mathbf{z}_n - \alpha F(\mathbf{z}_{n+1})\mathbf{h}(\mathbf{z}_{n+1}) \tag{3.3}$$

with $\alpha = h\gamma$. Bringing the two discretizations (3.1) and (3.3) together we obtain:

$$\mathbf{z}_{n+1} = \phi_h^f(\mathbf{z}_n) - \alpha F(\mathbf{z}_{n+1})\mathbf{h}(\mathbf{z}_{n+1}) \tag{3.4}$$

A simplistic analysis suggests that in this formulation best stabilization is obtained for α large. Indeed, by rewriting (3.4) as

$$\begin{aligned}
\tilde{\mathbf{z}}_{n+1} &= \phi_h^f(\mathbf{z}_n) \\
\mathbf{z}_{n+1} &= \tilde{\mathbf{z}}_{n+1} - D(\mathbf{z}_{n+1})\lambda \\
0 &= \mathbf{h}(\mathbf{z}_{n+1}) - \frac{1}{\alpha}H(\mathbf{z}_{n+1})D(\mathbf{z}_{n+1})\lambda
\end{aligned}$$

one sees that, for $\alpha \to \infty$, (3.4) becomes a coordinate projection method, i.e. at the end of each integration step the solution is projected onto the invariant manifold (see, e.g., [10], [3]). But even for $\infty > \alpha > 0$ we get the following convergence theorem:

Theorem 3.1 *Let α in* (3.4) *satisfy* $0 < \alpha < \infty$ *and let the method ϕ satisfy Assumptions 3.1. Then there exists an $h_o > 0$ (depending on α) such that for all h, $0 < h \leq h_o$, the scheme* (3.4) *possesses an invariant manifold $\mathcal{M}_h$ which is asymptotically stable. Moreover, the global error in $\mathbf{z}_{n+1}$ is $O(h^p)$ and $d(\mathcal{M}, \mathcal{M}_h) = O(h^{p+1})$. For $\alpha > 1$, $d(\mathcal{M}, \mathcal{M}_h) = O(h^{p+1}/\alpha)$.*

Proof. Note first that $\phi_h^f(\mathbf{z}) = \mathbf{z} + O(h)$. In particular, $\phi_0^f(\mathbf{z}) = \mathbf{z}$. For $h = 0$, linearization of (3.4) at $\mathbf{z}_n = \mathbf{z}_{n+1} \in \mathcal{M}$ yields

$$\hat{\mathbf{z}}_{n+1} = \hat{\mathbf{z}}_n - \alpha F(\mathbf{z}_{n+1})H(\mathbf{z}_{n+1})\hat{\mathbf{z}}_{n+1} \, .$$

Multiplying both sides by $H(\mathbf{z}_n)$ and introducing the new variable $\hat{\mathbf{v}}_n = H(\mathbf{z}_n)\hat{\mathbf{z}}_n$ results in the difference equation

$$\hat{\mathbf{v}}_{n+1} = \hat{\mathbf{v}}_n - \alpha\hat{\mathbf{v}}_{n+1}$$

for (1.12) and

$$\hat{\mathbf{v}}_{n+1} = \hat{\mathbf{v}}_n - \alpha H D(\mathbf{z}_{n+1}))\hat{\mathbf{v}}_{n+1}$$

for (1.13). This implies the asymptotic stability of $\mathcal{M}$ for $\infty > \alpha > 0$ at $h = 0$. On $\mathcal{M}$ the resulting map is the identity map. Thus $\mathcal{M}$ is normally hyperbolic in the sense of [19]. Furthermore, according to our assumptions on the modified ϕ_h^f, $\mathcal{M}_h = \mathcal{M}$ outside the bounded set K_1. The persistence of $\mathcal{M} \cap \bar{K}_1$ under small C^1–perturbations, which are in this case given by $\phi_h^f(\mathbf{z}_n) - \mathbf{z}_n$, follows now from Theorem 4.1 in [19]. This gives the first part of the theorem.

To prove the second part, let $\mathbf{e}_n = \mathbf{z}(t_n) - \mathbf{z}_n$. Then

$$\mathbf{e}_{n+1} = \phi_h^f(\mathbf{z}(t_n)) - \phi_h^f(\mathbf{z}_n) + O(h^{p+1}) - \alpha F(\mathbf{z}_{n+1})[\mathbf{h}(\mathbf{z}(t_{n+1})) - \mathbf{h}(\mathbf{z}_{n+1})] \tag{3.5}$$

But $\mathbf{h}(\mathbf{z}(t_{n+1})) - \mathbf{h}(\mathbf{z}_{n+1}) = (H(\mathbf{z}_{n+1}) + O(h))\mathbf{e}_{n+1}$. Substituting in (3.5) this gives a recursion for $\mathbf{e}_n$ which readily yields that $\mathbf{e}_n = O(h^p)$.

Now that we have the bound on the solution error $\mathbf{e}_n$, multiply (3.5) by $H(\mathbf{z}_{n+1})$ and define the projected error $\mathbf{s}_n = H(\mathbf{z}_n)\mathbf{e}_n$. Absorbing terms like $O(h)\mathbf{e}_n$ and $O(h)\mathbf{e}_{n+1}$ into $O(h^{p+1})$, we get

$$\mathbf{s}_{n+1} = (1+\alpha)^{-1}(\mathbf{s}_n + O(h^{p+1}))$$

with an obvious modification for (1.13). It follows that $\mathbf{s}_{n+1} = (1+\alpha)^{-1}O(h^{p+1})$, since the recursion is strictly contracting. This in turn yields the claimed result regarding $d(\mathcal{M}, \mathcal{M}_h)$. □

Remark. Numerical experiments indicate that the constant h_o in Theorem 3.1 satisfies $h_o \sim \frac{\alpha}{\gamma_o}$ where γ_o is the constant defined in Proposition 2.2. In particular, h_o becomes larger as α gets larger. This is also illustrated by the fact that in the limit (3.4) becomes a coordinate projection method. Furthermore, the computational expenses do not increase by making $\alpha = \infty$. This suggests to use the above coordinate projection method instead of (3.4) with $\alpha < \infty$ if α has to be large anyway. (This is the case if the constant γ_o in Proposition 2.2 is large.)

2. While the backward Euler stabilization (3.4), and in particular its coordinate projection limit, yield satisfactory stabilization schemes for appropriate choices of D, these are implicit schemes. Often, cheaper and still satisfactory stabilizations may be obtained by considering explicit schemes instead.

The simplest discretization of (3.2) is by forward Euler, which yields for (2.3) the scheme

$$\mathbf{z}_{n+1} = \phi_h^f(\mathbf{z}_n) - \alpha F(\mathbf{z}_n)\mathbf{h}(\mathbf{z}_n) \tag{3.6}$$

But a better explicit scheme is obtained by "marrying" forward Euler with the backward Euler stabilization, when viewed as a two-stage process as in the discussion preceding Theorem 3.1. This leads to the following explicit modification, which turns out to be *our method of choice*:

$$\begin{aligned} \tilde{\mathbf{z}}_{n+1} &= \phi_h^f(\mathbf{z}_n) && \text{(3.7a)} \\ \mathbf{z}_{n+1} &= \tilde{\mathbf{z}}_{n+1} - \alpha F(\tilde{\mathbf{z}}_{n+1})\mathbf{h}(\tilde{\mathbf{z}}_{n+1}) && \text{(3.7b)} \end{aligned}$$

This can be also viewed as a modification of the forward Euler stabilization, where the stabilizing term is evaluated at the predicted solution iterate $\tilde{\mathbf{z}}_{n+1}$ rather than at $\mathbf{z}_n$.

For both schemes (3.6) and (3.7) the manifold $\mathcal{M}$ is an asymptotically stable invariant manifold of the discretization scheme if and only if α is in the stability domain of the forward Euler method. This can be seen from the linearization of (3.7b) at $\tilde{\mathbf{z}}_{n+1} \in \mathcal{M}$. A procedure similar to that in the proof of Theorem 3.1 gives

$$\hat{\mathbf{v}}_{n+1} = (1 - \alpha)\hat{\mathbf{v}}_n \tag{3.8}$$

for (1.12) and

$$\hat{\mathbf{v}}_{n+1} = (I - \alpha HD(\tilde{\mathbf{z}}_{n+1}))\hat{\mathbf{v}}_n \tag{3.9}$$

for (1.13). These can be viewed as forward Euler discretizations for obvious simple ODEs. In particular, for the choice (1.12) with h small enough, we must require $0 < \alpha < 2$, and an excellent choice for γ is $\gamma = h^{-1}$, so that $\alpha = 1$. If (1.13) is used instead of (1.12) then one needs the largest and the smallest eigenvalues of HD to find a good α. (Note that α in this case might depend on $\tilde{\mathbf{z}}_{n+1}$ too.)

Theorem 3.2 *Let the method ϕ in* (3.1) *satisfy Assumptions 3.1 and let α in (3.7) (resp. (3.6)) be chosen uniformly from inside the absolute stability region of the forward Euler method, as described above. Then there exists an $h_o > 0$ such that for all h, $0 < h \le h_o$, the scheme (3.7) (resp. (3.6)) possesses an invariant manifold $\mathcal{M}_h$ which is asymptotically stable. Furthermore, the global error in $\mathbf{z}_{n+1}$ is $O(h^p)$ (i.e. the low order discretization of the stabilizing term does not reduce accuracy) and $d(\mathcal{M}, \mathcal{M}_h) = O(h^{p+1})$.*

Proof. The proof is the same as the proof of Theorem 3.1 except that the equation for $\hat{\mathbf{v}}_n$ is now (3.8) (or (3.9) in case of (1.13)). When α is chosen from inside the absolute stability region of the forward Euler method, this implies the asymptotic stability of $\mathcal{M}$ at $h = 0$. Similarly, the recursion for the projected error $\mathbf{s}_n$ is now

$$\mathbf{s}_{n+1} = (1 - \alpha)\mathbf{s}_n + O(h^{p+1})$$

(with an obvious modification for (1.13)), and this yields the claimed result $d(\mathcal{M}, \mathcal{M}_h) = O(h^{p+1})$, since the recursion is strictly contracting. □

Remarks.

1. While the choice (1.12) offers a simple procedure for choosing the best value for γ (namely, h^{-1}), which the simpler choice (1.13) does not offer, there remains the issue of the quick evaluation of the stabilization term. To that end note that it is possible (and sensible) to decompose HD once and use this for the approximate evaluation of $F(\mathbf{z}_n)$ over a few integration steps, as is customary in the modified Newton's method used in stiff ODE codes. More on this in Section 5.

2. Because of the equivalence of our stabilization approach with Baumgarte's method applied to index-2 DAEs with $D = B$, a discretization scheme similar to (3.6) can be derived for Baumgarte's formulation for the case $m = 1$ in (1.1). We leave the details to the reader. (Note again that the situation is different for index-3 problems.)

Let us now turn to our method of choice, (3.7). One reason for our claim that the stabilization (3.7) is particularly attractive is that it is very close to the method of coordinate projection. Indeed, rewriting the latter,

$$\tilde{\mathbf{z}}_{n+1} = \phi_h^f(\mathbf{z}_n) \tag{3.10a}$$

$$\mathbf{z}_{n+1} = \tilde{\mathbf{z}}_{n+1} - D(\tilde{\mathbf{z}}_{n+1})\lambda \tag{3.10b}$$

$$\mathbf{0} = \mathbf{h}(\mathbf{z}_{n+1}) \tag{3.10c}$$

we readily obtain

Corollary

1. *One Newton iteration for solving (3.10b),(3.10c) yields (3.7b) with $\alpha = 1$. A choice of $\alpha < 1$ corresponds to a damped Newton step.*

2. *In particular, if* $\mathbf{h}$ *is linear then the stabilization method (3.7) projects the solution at the end of each step back onto the invariant manifold $\mathcal{M}$.*

In Section 5 we relate our results to others about coordinate projection methods (e.g. [21], [9]). Here we note that numerical calculations for Examples 1 and 2 in [4] (with $D = H^T$) confirm that the stabilization (3.7) yields best results and zero drift for these (albeit linear) examples. This is a better stabilization performance than what is obtained with either the forward or the backward Euler stabilizations.

Example 1. Consider the ODE

$$z' = 3t^2 \tag{3.11}$$

with the invariant manifold $\mathcal{M}$ given by $z = t^3$. We discretize this ODE by the midpoint rule (this gives a method ϕ of order $p = 2$) and apply the forward Euler stabilization (3.6) with $D = H = 1$. This yields:

$$z_{n+1} = z_n + 3h(t_n + h/2)^2 - \alpha(z_n - t_n^3)$$

For $\alpha = 0$ (i.e. without stabilization) the solution z_n to the initial value problem with $z_o = 0$ is given by $z_n = t_n^3 - (n/4)h^3$. Obviously, the drift grows linearly in time (for a fixed h) and the accuracy in z is $O(h^2)$. For $\alpha = 1$ we get an invariant manifold $\mathcal{M}_h$ given by $z_n = t_n^3 - (1/4)h^3$, i.e. the drift does not grow in time and the solution is $O(h^3)$ accurate. Note that $z_n \in \mathcal{M}_h$ for n large enough (typically $n > 3$) even if $x_o \neq 0$. The invariant manifold $\mathcal{M}_h$ persists for all $h \geq 0$. Also, we caution the reader not to be misled by the fact that here the manifold defines the solution: In general, only the drift and not the solution itself gains a power of h in accuracy.

Applying our method (3.7) with $\alpha = 1$ yields the exact solution for this simple example, $z_n = t_n^3$ (note that $\mathcal{M} = \mathcal{M}_h$). Similarly, applying the backward Euler stabilization (3.4) gives $z_n = z(t_n) + O(h^3)$, but $z_n \to z(t_n)$ as $\alpha \to \infty$.

Let us add now a second equation to (3.11) of the form

$$y' = -(1 + \nu(z - t^3))y \tag{3.12}$$

where ν is a constant, $\nu \gg 1$. If we discretize and stabilize the ODE (3.11) as before, we know that with (3.6) $z_n = t_n^3 - .25h^3$, so (3.12) becomes

$$y' = -(1 - .25\nu h^3)y \ .$$

Thus, to maintain the stability of the equilibrium solution $y(t) = 0$ we have to restrict the step size h to

$$h < \sqrt[3]{\frac{1}{4\nu}}\ .$$

Now, if $\nu \gg 1$, this may result in a very small step size h. A restriction of a similar sort arises when using the backward Euler scheme (3.4), unless α is very large. No such restriction arises for a coordinate projection method, which in the linear case includes (3.7).

This example indicates that the stabilization methods (3.4) and (3.6) may result in a much smaller step size h compared to those for projection methods if the ODE (2.1) has a qualitatively different behavior away from the manifold $\mathcal{M}$. (See also Example 1 in [4].) □

Example 2. Kepler's problem [2], concerns motion in a central field with potential $U = -K/r$, where K is a constant and r is a radius. In Euclidean coordinates the equations of motion become

$$\begin{aligned}
p_1' &= v_1 \\
p_2' &= v_2 \\
v_1' &= -\frac{K}{r^3}p_1 \\
v_2' &= -\frac{K}{r^3}p_2
\end{aligned}$$

where $r = \sqrt{p_1^2 + p_2^2}$. For notational simplicity we abbreviate the right hand side of this ODE by $\hat{\mathbf{f}}(\mathbf{p}, \mathbf{v})$. We also consider only initial values $p_1(0) = c$, $p_2(0) = 0$, $v_1(0) = 0$, $v_2(0) = \sqrt{2.0/c - 1.0}$ with $1 > c > 0$ and $K = 1$. It can be shown that the analytic solutions of these initial value problems have period $T = 2\pi$. However, numerical discretization of this ODE results in general in a growth of the computed p_2 at $t_n = kT$, k a natural number, which is quadratic in k [16, 1]. (The exact solution would be $p_2(kT) = 0$.)

This difficulty can be avoided by stabilizing the energy e

$$e(\mathbf{p}, \mathbf{v}) = \frac{v_1^2 + v_2^2}{2} - \frac{K}{r}$$

which is an integral invariant of the problem [16]. Note that for the given initial values we have $e = -.5$. Thus our stabilization approach (2.3) with $F = E^T(EE^T)^{-1}$ results in the stabilized ODE

$$\begin{pmatrix} \mathbf{p}' \\ \mathbf{v}' \end{pmatrix} \quad = \quad \mathbf{f}(\mathbf{p}, \mathbf{v}) - \gamma F(\mathbf{p}, \mathbf{v})(e(\mathbf{p}, \mathbf{v}) + .5)$$

where $E(\mathbf{p}, \mathbf{v}) = (p_1/r^3, p_2/r^3, v_1, v_2)$. Numerical experiments show that for the stabilized formulation (using the discretization (3.7) with forward Euler as ϕ and $\alpha = 1$) the growth in the computed $p_2(kT)$ is linear in k and $p_2(kT) = O(h^2)$. The same linear growth was observed for the unstabilized original formulation in case a symplectic integrator (e.g. implicit midpoint) was used. Here the stabilization affects only the magnitude of the global error in p_2. Results for $c = .5$ are recorded in Table 3.1. Note that for this problem the period T depends only on the energy e. This helps to explain the dramatic improvement which our stabilization yields for the forward Euler discretization. □

discretization scheme	stabilization	h	$p_2(2\pi)$	$p_2(4\pi)$
forward Euler	no	$.001\pi$	-.63	-.91
forward Euler	yes	$.001\pi$	.12e-3	.24e-3
forward Euler	no	$.0005\pi$	-.35	-.88
forward Euler	yes	$.0005\pi$	.32e-4	.63e-4
midpoint	no	$.001\pi$	.47e-3	.94e-3
midpoint	yes	$.001\pi$	.27e-4	.55e-4

Table 3.1: Error in the variable p_2 for the stabilized and unstabilized discretizations of Kepler's problem

4 Stabilized DAEs and Euler–Lagrange Equations

For a semi-explicit, pure index-2 DAE

$$\begin{aligned} \mathbf{x}' &= \mathbf{f}(\mathbf{x}, t) - B(\mathbf{x}, t)\mathbf{y} && (4.1a)\\ \mathbf{0} &= \mathbf{g}(\mathbf{x}, t) && (4.1b) \end{aligned}$$

we have already essentially described the process: the constraints (4.1b) are differentiated once, and the obtained expression together with (4.1a) are equivalent to an ODE (2.1). The invariant manifold defined by (4.1b) is related to as (2.2) (with the usual formal conversion to an autonomous form, which of course we do not perform in practice). It is an integral invariant, as can be readily verified. The stabilization (3.7) may be applied. The whole integration process may, in fact, be accomplished efficiently by explicit discretization schemes if the ODE is not stiff.

For a semi-explicit, pure index-3 DAE, e.g.

$$\begin{aligned} \mathbf{x}'' &= \mathbf{f}(\mathbf{x}, \mathbf{x}', t) - B(\mathbf{x}, t)\mathbf{y} && (4.2a)\\ \mathbf{0} &= \mathbf{g}(\mathbf{x}, t) && (4.2b) \end{aligned}$$

we apply two differentiations to the constraints (4.2b). Again the resulting expression together with (4.2a) are equivalent to an ODE (of second order, which may of course be written as a first order system of twice the size as in (1.10)). For the invariant manifold (2.2) we may choose the set defined by (4.2b) and its derivative (this is not an integral invariant), or we may choose to consider only the derivative of (4.2b) as the invariant manifold. Such choices lead to different stabilizations.

Let us further consider the important class of index-3 DAEs arising in modeling the dynamics of constrained multibody systems. A Lagrangian formulation of the equations describing a constrained (autonomous) multibody system may be written as

$$\begin{aligned} \mathbf{p}' &= \mathbf{v} \\ M(\mathbf{p})\mathbf{v}' &= \mathbf{f}(\mathbf{p}, \mathbf{v}) - G(\mathbf{p})^T\lambda \\ \mathbf{0} &= \mathbf{g}(\mathbf{p}) \end{aligned} \qquad (4.3)$$

where $M(\mathbf{p})$ is the mass matix (assumed positive definite), $\mathbf{f}(\mathbf{p}, \mathbf{v})$ is the vector of applied forces, and λ represents the Lagrange multipliers coupled to the system by the matrix

$G(\mathbf{p}) = \mathbf{g_p}(\mathbf{p})$ which is assumed to have full row rank. We assume no explicit time dependence, for notational simplicity. If we differentiate the constraints of the problem with respect to time, we obtain the constraint equations on velocity level

$$\mathbf{0} = G(\mathbf{p})\mathbf{v}$$

and a further differentiation with respect to time results in the constraint equations on acceleration level

$$\mathbf{0} = G(\mathbf{p})\mathbf{v}' + L(\mathbf{p}, \mathbf{v})\mathbf{v}$$

where $L(\mathbf{p}, \mathbf{v}) = \mathbf{v}^T \mathbf{g_{pp}}(\mathbf{p})$ has the dimensions of G. From this equation and (4.3) it is possible to abtain λ as a function of $\mathbf{p}$ and $\mathbf{v}$:

$$\lambda = \Lambda(\mathbf{p}, \mathbf{v}) := (GM^{-1}G^T)^{-1}(GM^{-1}\mathbf{f} + L\mathbf{v})$$

This expression may then be reintroduced in (4.3), resulting in an ODE for $\mathbf{p}$ and $\mathbf{v}$, namely, the first two equations in (4.3) with λ replaced by $\Lambda(\mathbf{p}, \mathbf{v})$. This ODE has the manifold defined by

$$\begin{aligned} \mathbf{0} &= \mathbf{g}(\mathbf{p}) \\ \mathbf{0} &= G(\mathbf{p})\mathbf{v} \end{aligned}$$

as an invariant manifold. In accordance with our notation introduced in Section 2 we abbreviate the right hand side of the above ODE by $\hat{\mathbf{f}}$ and the right hand side of the above algebraic equations by $\mathbf{h}$. The dependent variable is $\mathbf{z} = (\mathbf{p}, \mathbf{v})$ and

$$H = \begin{pmatrix} G(\mathbf{p}) & 0 \\ L(\mathbf{p}, \mathbf{v}) & G(\mathbf{p}) \end{pmatrix}$$

Thus we obtain the ODE (2.1) with an invariant manifold $\mathcal{M}$ given by (2.2). Note that the mapping $\mathbf{h}$ is not an integral invariant of the ODE (2.1); i.e., we have

$$H(\mathbf{p}, \mathbf{v})\hat{\mathbf{f}}(\mathbf{p}, \mathbf{v}) = \begin{pmatrix} G(\mathbf{p})\mathbf{v} \\ \mathbf{0} \end{pmatrix}$$

Stabilization of the invariant manifold in the sense of (2.3) leads to the stabilized ODE

$$\begin{pmatrix} \mathbf{p}' \\ \mathbf{v}' \end{pmatrix} = \hat{\mathbf{f}}(\mathbf{p}, \mathbf{v}) - \gamma F(\mathbf{p}, \mathbf{v})\mathbf{h}(\mathbf{p}, \mathbf{v}) \tag{4.4}$$

where F is given by (1.12). With $\alpha = 1$ in the method (3.7) we then obtain the following two-stage discretization step:

1. Starting with $(\mathbf{p}_n, \mathbf{v}_n)$ at $t = t_n$, use a favourite ODE integration scheme ϕ_h^f (e.g. Runge-Kutta or multistep) to advance the system

$$\begin{aligned} \mathbf{p}' &= \mathbf{v} \\ M(\mathbf{p})\mathbf{v}' &= \mathbf{f}(\mathbf{p}, \mathbf{v}) - G^T(\mathbf{p})\lambda \\ \mathbf{0} &= G(\mathbf{p})\mathbf{v}' + L(\mathbf{p}, \mathbf{v})\mathbf{v} \end{aligned}$$

by one step. Denote the resulting values at $t_{n+1} = t_n + h$ by $(\tilde{\mathbf{p}}_{n+1}, \tilde{\mathbf{v}}_{n+1})$.

2. Stabilize:

$$\begin{pmatrix} \mathbf{p}_{n+1} \\ \mathbf{v}_{n+1} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{p}}_{n+1} \\ \tilde{\mathbf{v}}_{n+1} \end{pmatrix} - F(\tilde{\mathbf{p}}_{n+1}, \tilde{\mathbf{v}}_{n+1})\mathbf{h}(\tilde{\mathbf{p}}_{n+1}, \tilde{\mathbf{v}}_{n+1})$$

Recall that in order to apply Proposition 2.1 instead of 2.2 we could precede (4.4) by an index-2 reduction as in [13]. However, this turns out not to be necessary: Instead of applying Proposition 2.2 we introduce the new variables $\mathbf{s} = \mathbf{g}(\mathbf{p})$ and $\mathbf{r} = G(\mathbf{p})\mathbf{v}$. The corresponding differential equations for these variables can be obtained by premultiplying both sides of (4.4) by H:

$$\begin{aligned} \mathbf{s}' &= \mathbf{r} - \gamma \mathbf{s} \\ \mathbf{r}' &= -\gamma \mathbf{r} \end{aligned} \tag{4.5}$$

Thus the manifold $\mathcal{M}$ is an asymptotically stable invariant manifold of (4.4) for all $\gamma > 0$. This implies that the stabilizing term in (4.4) can be discretized as in Section 3.

This is to be contrasted with the Baumgarte technique for (4.3) which yields

$$\begin{aligned} \mathbf{s}' &= \mathbf{r} \\ \mathbf{r}' &= -\gamma_1 \mathbf{s} - \gamma_2 \mathbf{r} \end{aligned} \tag{4.6}$$

While this gives an asymptotically stable manifold $\mathcal{M}$ for, e.g., $\gamma_1 = \gamma^2$ and $\gamma_2 = 2\gamma$ with $\gamma > 0$, the system (4.5) is favoured over (4.6). To see this, consider discretization of (4.6), e.g., by forward Euler. (This is of course a simplification of the full picture, which involved discretization of the actual mechanical system — the discretization and the change of variables are operations which in general do not commute.) It results in

$$\begin{aligned} \mathbf{s}_{n+1} &= \mathbf{s}_n + h\mathbf{r}_n \\ \mathbf{r}_{n+1} &= \mathbf{r}_n - \alpha_1 \mathbf{s}_n - \alpha_2 \mathbf{r}_n \end{aligned}$$

with $\alpha_1 = h\gamma_1$ and $\alpha_2 = h\gamma_2$. Best stabilization is obtained for the choice $\alpha_1 = 1/h$ and $\alpha_2 = 2$ which yields $\mathbf{s}_n = \mathbf{r}_n = 0$ for $n \geq 2$ starting from arbitrary initial values $\mathbf{s}_o$, $\mathbf{r}_o$. Note however that $\mathbf{r}_1 \approx -\mathbf{s}_o/h$. In the full, nonlinear case, such a perturbation is undesirable. In contrast, a forward Euler discretization of (4.5), which with the choice $\gamma = 1/h$ also yields $\mathbf{s}_n = \mathbf{r}_n = 0$ for $n \geq 2$, gives $\mathbf{r}_1 = 0, \mathbf{s}_1 = h\mathbf{r}_0$, and no disturbing perturbations arise.

The stabilization involving $F = H^T(HH^T)^{-1}$ with the matrix H as given above is safe, but perhaps cumbersome. Other, cheaper choices for F are possible. One is using (1.12) and

$$D = \begin{pmatrix} G^T & 0 \\ 0 & G^T \end{pmatrix} \tag{4.7}$$

which has the advantage that only GG^T needs to be decomposed (or "inverted"). Another possibility, which avoids using L altogether, is to use (1.13) with

$$F(\mathbf{p}, \mathbf{v})\,(= D) \quad = \quad \begin{pmatrix} G^T(GG^T)^{-1} & 0 \\ 0 & G^T(GG^T)^{-1} \end{pmatrix} \tag{4.8}$$

Then $\mathcal{M}$ is again asymptotically stable for all $\gamma > 0$, but choosing α in (3.7) is trickier. This stabilization should not be used when L dominates G.

It is also possible, according to our theory, to stabilize the velocity constraints alone. Note that the velocity constraints form an invariant manifold for the ODE (that is the ODE obtained by eliminating λ from (4.3) as previously described). On the other hand, the position constraints alone do not form an invariant manifold, hence our theory does not cover a stabilization like (3.7) or coordinate projection using just the position constraints. It is not clear, however, whether the stabilization along velocity constraints alone should be recommended in the most general case, since it does not satisfy the "beauty requirement" of no drift in the position constraints, and in cases where L is large and cannot be dropped the cost is anyway comparable to that of the first alternative in the previous paragraph.

Remark. It is interesting to note that, despite the above remarks, a "projected invariant" method on the position constraints which was proposed in [4], [5] works rather well for many problems. □

Example 3. In this example we consider a slider–crank mechanism. Following [17, 1], the motion of this mechanism can be described by the following index-3 DAE:

$$
\begin{aligned}
J_1\theta'' &= -\lambda_1 r \sin\theta - \lambda_2 r\cos\theta + n_1\\
m_2 x_2'' &= -\lambda_1\\
m_2 y_2'' &= -\lambda_3 - m_2 g\\
J_2\psi'' &= -\lambda_1 l_1 \sin\psi + \lambda_2 l\cos\psi + \lambda_3(l-l_1)\cos\psi\\
0 &= x_2 - r\cos\theta - l_1\cos\psi\\
0 &= r\sin\theta - l\sin\psi\\
0 &= y_2 - (l-l_1)\sin\psi
\end{aligned}
$$

where J_1, J_2, m_2, g, r, l, l_1 are constants which we assume here to take the values $J_1 = 10$, $J_2 = 1$, $m_2 = 1$, $g = 9.81$, $r = 1$, $l = 3$, $l_1 = 2$. Furthermore, we consider the case where the torque n_1 is given by $n_1 = sin(t) - \theta'$ (the second term represents friction at the shaft) and take initial values $(\theta(0), x_2(0), y_2(0), \psi(0)) = (0,3,0,0)$, $(\theta'(0), x_2'(0), y_2'(0), \psi'(0)) = (-1, 0, -1/3, -1/3)$.

The resulting ODE was integrated using a second–order explicit midpoint scheme (step size h = .1), and (3.7) with $\alpha = 1$ and F as in (4.8) was applied for the stabilization of the coordinate and velocity constraints. The error in the position of the slider was computed by comparing the numerical results with those obtained for step size $h = .01$. Fig. 4.1 shows this error when using no stabilization (solid line), our stabilization as described above (dashed line), and Baumgarte's technique with $\gamma = .6/h$ in (1.5) (dotted line). (The chosen parameter value for γ is close to optimal for this Baumgarte-midpoint scheme.) It is clear from Fig. 4.1 that the error in the solution grows linearly for this range of time when no stabilization is applied, but does not grow when either of the stabilization techniques is used. The same observations were made with respect to the drift. The maximum value of the drift in velocity and position levels, resp., was $(2.4e-4,\ 8.0e-9)$ for our method and $(8.0e-3,\ 1.5e-3)$ for the Baumgarte method. The advantage of applying our simple stabilization is evident.

We also computed the solutions using the same ϕ and h and stabilizing either the velocity constraints or the position constraints (but not both). Both of these stabilizations prove

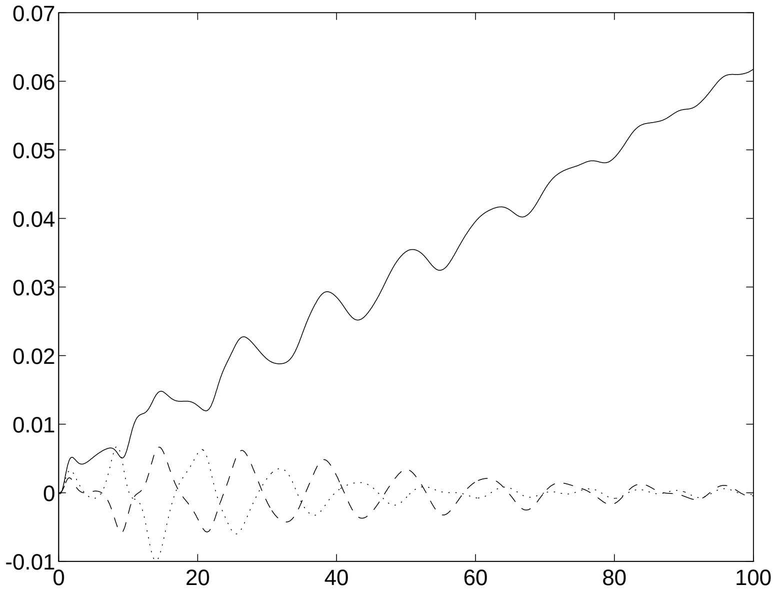

Figure 4.1: Error in the position of the slider: without stabilization (solid line), with our stabilization (dashed line) and with Baumgarte's stabilization (dotted line)

worthwhile for this example, although the best bounds on the drifts were obtained using (4.8). □

5 Summary and discussion

Starting with the Baumgarte stabilization technique, we have explained its inherent limitations, especially for index-3 DAEs. This suggests that a further experimental search for "optimal" values of Baumgarte parameters independent of the discretization may prove frustrating. We have then considered a series of improvements, each refining the next with respect to either stability or efficiency or both. Our first step was to consider stabilization (without discretization) of invariant manifolds. This gave us a unified view of a family of stabilization techniques, excluding those which in the limit lead to DAEs of index > 2. Our next step was to consider the discretization of such stabilization formulations. Simple, special purpose discretization of the stabilizing term which still maintains the high order of a (correspondingly high order) discretization of the unstabilized ODE, is possible and surprisingly affordable. Close to optimal choices for what corresponds to the Baumgarte parameter were also established along the way. The application of these ideas to high-index DAEs in general and to mechanical systems in particular were discussed and demonstrated.

This process has eventually led us to the stabilization method (3.7), which in turn can be interpreted (for $\alpha = 1$) as one Newton step of a coordinate projection method (3.10). To be precise, for the underdetermined algebraic system of equations $\mathbf{h}(\mathbf{z}_{n+1}) = \mathbf{0}$ one applies a Newton step starting with $\tilde{\mathbf{z}}_{n+1}$, where the search direction is restricted to be in range$\{D\}$.

If $D = H^T(\tilde{\mathbf{z}}_{n+1})$ then the obtained correction has minimum l_2 norm, as in [21]. But we stress that our derivation is entirely different, both in motivation and in results, from an approximate coordinate projection method as studied, e.g., in [21] and [9]: The method (3.7) stands alone and Theorem 3.2 applies for it without any approximation; values of $\alpha \neq 1$ (including $\alpha > 1$) make sense as well; we prove what is assumed in Remark 3.1 of [9].

In case that the invariant equations are linear in the dependent variable, the stabilization method (3.7) coincides with a coordinate projection method. Here it is important to consider the nonautonomous case, because we are allowing a non-constant $H = H(t)$. In the nonlinear case, similar conclusions arise if a quasilinearization approach is applied to a given DAE, i.e. a sequence of linearizations is considered, and the methods discussed here are applied to each linearized problem. (This is the standard technique for solving boundary value problems, implemented e.g. for projected collocation [3] in [6], and it corresponds to a waveform variant for initial value problems.) But within the usual approach to solving nonlinear initial value problems we do have a different method in (3.7) (and a satisfactory one at that, according to Theorem 3.2 and experiments).

We note that there are some special cases where the numerical method for discretizing the ODE (2.1) automatically satisfies the invariant (2.2) at the end of each step as well (assuming consistent initial values). This is the case for all reasonable Runge-Kutta schemes if $\mathbf{h}' = const.$ [21]. It also holds if each component of $\mathbf{h}$ is quadratic, i.e. $\mathbf{h}_j = \mathbf{z}^T P_j \mathbf{z}$ where P_j are constant matrices, for the Gauss–Legendre Runge–Kutta scheme. (The latter result, which is easy to see when viewing the method as a collocation scheme, has been noted a few times in the literature, including in [3], [20].) In such cases the stabilization techniques are deemed unnecessary – the stability of the ODE discretization scheme is sufficient. On the other hand, in [3] a coordinate projection method which coincides with (3.7) is *proved* and numerically demonstrated to *improve* the stability properties of a discretization scheme for (4.1) (which can be viewed as a discretization scheme for (2.1)).

Next we address the question of choosing D and that of choosing between (1.12) and (1.13). For an index-2 DAE, we have seen (Example 2 in [4]) that the Baumgarte choice $D = B$ (with F satisfying (1.12)) can be unfortunate when B and G^T are almost orthogonal and vary in t. The choice $D = G^T$, or more generally $D = H^T$ for (2.1),(2.2), yields an orthogonal projection in (1.7) or (2.3), and generally yields a better stabilization. However, there is a question of cost involved: Starting from a DAE (4.1) or (4.2) the elimination of the algebraic unknowns involves decomposing GB, not GG^T. While we stress (following Petzold) that the explicit form of (2.1) is not to be formulated – rather, the equivalent DAE with differentiated constraint is used to eliminate the algebraic unknowns $\mathbf{y}$ only when necessary – it may still be argued that a stabilization involving $D = B$ is cheaper than one involving $D = G^T$ under these circumstances. Of course, this extra expense in using the preferred G^T (or H^T in the notation of Sections 2 and 3) disappears if we use (1.13) instead of (1.12), but for the latter the choice of α (or γ) is trickier when the eigenvalues of HD are spread apart.

On the other hand, note that all that is required of D is to form a reasonably small angle with H^T: it does not have to be any of the choices above. For instance, as already mentioned before, we can form and decompose HH^T only once every few time steps, as is commonly done in stiff ODE solvers when applying a modified Newton method. Another possibility is to realize that (1.12) can be viewed as a preconditioned form of (1.13): essentially, the stabilization is effective (with an appropriate choice of α in (3.7)) if the eigenvalues of HF are closely clustered, so that for each eigenvalue μ of HF, $\mu\alpha \approx 1$. In many cases it is sufficient to simply use an unsophisticated preconditioner like an SOR iteration for an approximation of $(HD)^{-1}$ in (1.12) (which, of course, is never explicitly formed either). For Example 2 in [4], no preconditioning is needed, and the term $G^T(GG^T)^{-1}$ in (1.7) can be replaced by G^T. But in applications arising from partial differential equations there is less reason to expect a similar success. Still, in general an SOR preconditioning iteration at time $t_n + h$ starting from given values at time t_n can be very effective, unless a large discontinuous change takes place across the step.

Thus, the cost of using a good stabilizer can be reduced to a small portion of the cost of simply solving the ODE (2.1) (or the corresponding DAE (4.1) or (4.2)), even when the latter is not stiff.

Acknowledgement The authors would like to thank Dr. Linda Petzold and Dr. Erik Van Vleck for many fruitful discussions.

References

[1] T. Alishenas, *Zur numerischen Behandlung, Stabilisierung durch Projektion und Modellierung mechanischer Systeme mit Nebenbedingungen und Invarianten*, PhD thesis, Königliche Technische Hochschule Stockholm, 1992.

[2] V.I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer–Verlag, 1989.

[3] U. Ascher and L. Petzold, *Projected implicit Runge-Kutta methods for differential-algebraic equations*, SIAM J. Numer. Anal. 28 (1991), 1097-1120.

[4] U. Ascher and L. Petzold, *Stability of Computational Methods for Constrained Dynamics Systems*, SIAM J. Scient. Comp. 14 (1993), 95-120.

[5] U. Ascher and L. Petzold, *Projected collocation for higher-order higher-index differential-algebraic equations*, J. Comp. Appl. Math. 43 (1992), 243-259.

[6] U. Ascher and R. Spiteri, *Collocation software for boundary value differential-algebraic equations*, Tech. Rep. 92-18, Dept. Computer Science, Univ. of BC, 1992.

[7] J. Baumgarte, *Stabilization of constraints and integrals of motion in dynamical systems*, Comp. Math. Appl. Mech. Eng. 1 (1972), 1-16.

[8] P. Carter, PhD Thesis, Univ. of B.C., in preparation.

[9] E. Eich, *Convergence results for a coordinate projection method applied to mechanical systems with algebraic constraints*, SIAM J. Numer. Anal., to appear.

[10] E. Eich, K. Führer, B. Leimkuhler and S. Reich, *Stabilization and projection methods for multibody dynamics*, Research report, Inst. Math., Helsinki Univ. of Technology, 1990.

[11] K. Führer and B. Leimkuhler, *Numerical solution of differential-algebraic equations for constrained mechanical motion*, Numer. Math. 59 (1991), 55-69.

[12] C. W. Gear, *Maintaining solution invariants in the numerical solution of ODEs*, SIAM J. Sci. Stat. Comp., 7 (1986), 734-743.

[13] C.W. Gear, G. Gupta and B. Leimkuhler, *Automatic integration of the Euler-Lagrange equations with constraints*, J. Comput. Appl. Math. 12 (1985), 77-90.

[14] Gresho P.M. & Sani R.L., *On pressure boundary conditions for the incompressible Navier-Stokes equations*, Int. J. Numer. Methods Fluids 7 (1987), 1111-1145.

[15] E. Griepentrog and R. März, *Differential-Algebraic Equations and their Numerical Treatment*, Teubner-Texte Math. 88, Teubner, Leipzig, 1986.

[16] W. von Grünhagen, *Zur Stabilisierung der numerischen Integration von Bewegungsgleichungen*, PhD thesis, University Braunschweig, 1979.

[17] E. Haug, *Elements and methods of computational dynamics*, in: Computer aided analysis and optimization of mech. system dynamics, E. Haug (editor), Springer, 1984, 3-38.

[18] E. Haug and R. Deyo, *Real-Time Integration Methods for Mechanical Sustem Simulation*, NATO ASI Series, Springer 1991.

[19] M. Hirsch, C. Pugh, and M. Shub, *Invariant manifolds*, Lecture Notes in Math. No. 583, Springer–Verlag, 1976.

[20] B. Leimkuhler and S. Reich, *The numerical solution of constrained Hamiltonian systems*, Technical Report, Zuse Center Berlin, 1992.

[21] L.F. Shampine, *Conservation laws and the numerical solution of ODEs*, Comp. Maths. Appls. 12B (1986), 1287-1296.