Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 4 (2011) 1771–1781

Procedia Computer Science

On the adaptive solution of space–time inverse problems with the adjoint method

Mihai Alexe[1], Adrian Sandu[1]

Computational Science Laboratory,
Virginia Polytechnic and State University,
Blacksburg, VA, 24060, USA

Abstract

Adaptivity in space and time is ubiquitous in modern numerical simulations. The large number of unknowns associated with today's typical inverse problem may run in the millions, or more. To capture small scale phenomena in regions of interest, adaptive mesh and temporal step refinements are required, since uniform refinements quickly make the problem computationally intractable. To date, there is still a considerable gap between the state–of–the–art techniques used in direct (forward) simulations, and those employed in the solution of inverse problems, which have traditionally relied on fixed meshes and time steps. This paper describes a framework for building a space-time consistent adjoint discretization for a general discrete forward problem, in the context of adaptive mesh, adaptive time step models. The discretize–then–differentiate approach to optimization is a very attractive approach in practice, because the adjoint model code may be generated using automatic differentiation (AD). However, several challenges are introduced when using an adaptive forward solver. First, one may have consistency problems with the adjoint of the forward numerical scheme. Similarly, intergrid transfer operators may reduce the accuracy of the discrete adjoint sensitivities. The optimization algorithm may need to be specifically tailored to handle variations in the state and gradient vector sizes. This work shows that several of these potential issues can be avoided when using the Runge–Kutta discontinuous Galerkin (DG) method, an excellent candidate method for h/p-adaptive parallel simulations. Selective application of automatic differentiation on individual numerical algorithms may simplify considerably the adjoint code development. A numerical data assimilation example illustrates the effectiveness of the primal/dual RK–DG methods when used in inverse simulations.

Keywords: Inverse problems, discrete adjoint method, adaptive mesh refinement, automatic differentiation, discontinuous Galerkin method.

Email addresses: malexe@cs.vt.edu (Mihai Alexe), sandu@cs.vt.edu (Adrian Sandu)

[1]This work was supported in part by the US National Science Foundation, through the following awards: NSF-CCF-0635194, NSF OCI-0904397, NSF CCF-0916493, and NSF DMS-0915047.

doi:10.1016/j.procs.2011.04.192

1. Introduction and background

1.1. Background and objectives

Inverse problems (IPs) [1] are defined as estimation problems for unknown parameters that appear in models of various engineering or physical processes. One infers the *best* values for the unknown parameters (where optimality is defined in problem-dependent statistical terms), based on known *a priori* observations available, e.g., from direct measurements, or indirectly through forward simulation results. The mathematical formulation of IPs reduces to a PDE- or ODE-constrained numerical optimization problem. We focus on the *discretize–then–optimize* approach, which calculates the optimal values of the model parameters starting from the discretization of the primal set of equations, i.e., optimality of the solution holds over our choice of discrete spaces. One major advantage of this approach is that the numerical gradients needed in the optimization process can be calculated efficiently using the adjoint method, and automatic differentiation. Moreover, we consider adjoints of models that have been discretized on adaptive meshes and with adaptive time steps. Space-time adaptivity introduces several challenges in the inversion process, which have delayed the adoption of state-of-the-art techniques used in forward numerical simulations. Reviews on recent literature show a growing trend of research into the use of adaptive inverse problem solvers (see, e.g., [2, 3, 4, 5, 6], and references therein).

This paper discusses a framework for the derivation of adjoint models in both function spaces (the continuous setting), as well as for discrete models (that may be implemented using AD). To obtain a consistent adjoint model, the cost functional, as well as the differential volume and boundary operators in the forward model, need to satisfy a set of *compatibility conditions*. Extending previous work by Giles [7], we derive more general compatibility conditions applicable for target functionals that contain both volume and boundary terms. Hartmann [8] gives a framework for the derivation of dual consistent discontinuous Galerkin discretizations for elliptic problems; see also the earlier work of Lu [9]. The addition of the time dimension precludes a simple extension of Hartmann's dual consistency theory to time dependent problems, unless the time dimension is also discretized with DG. Leveraging the results in [8, 10], we show that Runge–Kutta DG discretizations that are dual consistent in space, are also adjoint consistent in time. Moreover, their adjoints retain the temporal accuracy of the forward Runge-Kutta methods.

The adjoints of intergrid operators are also important components in the inversion process. Inconsistent solution transfer operators may lead to a loss of accuracy in the discrete adjoint solution. For discontinuous Galerkin and other types of methods that use truncated polynomial expansions to approximate the exact solution, we show that the by transposing the mesh transfer operators used in the forward simulation, we retrieve exactly the intergrid operators that would be used in a discretization of the continuous adjoint equations. This result (see [11] for general proofs), derived from the $\mathcal{L}^2$-orthogonality of element-wise h/p-mesh operators, allows AD to be used transparently for the generation of all mesh transfers in the adjoint simulation. This reduces the possibility of errors, and shortens the development effort required to build the adjoint framework.

1.2. Organization

This paper is organized as follows. Section 2 outlines the adjoint framework for space-time PDE problems. We discuss both the function space formulation, as well as the discrete setting, and highlight the requirements for consistence in both optimize–then–discretize, and discretize–then–optimize approaches. Section 3 considers the consistency and accuracy discrete adjoints of intergrid transfer operators for discontinuous Galerkin methods. Numerical experiments that

make use of the theoretical framework are presented in section 4. Finally, section 5 summarizes the conclusions, and discusses extensions that are the subject of current research.

2. The adaptive inverse problem framework

Consider a dynamical system whose space and time evolution is described by the following partial differential equation:

$$\begin{aligned} \mathbf{u}_t &= N(\mathbf{u}) + \mathbf{f}\ ,\ \mathbf{x} \in \Omega\ ,\ t \in [0,T] \\ B(\mathbf{u}) &= \mathbf{g}\ ,\ \mathbf{x} \in \Gamma\ ,\ t \in [0,T] \\ \mathbf{u}(t=0,\mathbf{x}) &= \mathbf{u}^0(\mathbf{x})\ ,\ \mathbf{x} \in \Omega\ . \end{aligned} \tag{1}$$

The PDE system (2) admits solutions $\mathbf{u} : [0,T] \to \mathcal{U}$, such that $\mathbf{u} \in \mathcal{L}^2([0,T];\mathcal{U})$, and $\mathbf{u}_t \in \mathcal{L}^2([0,T];\mathcal{U})$, where $\mathcal{U}$ is a space of sufficiently smooth functions. Here N and B are continuously Frechét differentiable, nonlinear differential operators, containing spatial and boundary derivative terms. In the following we consider the initial condition $\mathbf{u}^0$ as the inversion variable. Let the Frèchet derivatives of N and B in the direction $\mathbf{w} \in \mathcal{U}$, be denoted by $L\,\mathbf{w} = N'[\mathbf{u}]\,\mathbf{w}$, and $B'\,\mathbf{w} = B'[\mathbf{u}]\,\mathbf{w}$, respectively. Here the brackets denote the state about which the linearization is performed, while the prime notation indicates a Fréchet derivative. Consider a nonlinear cost functional of the form

$$\mathcal{J}(\mathbf{u}) = \int_0^T \int_\Omega J_\Omega[C_\Omega\,\mathbf{u}]\,\mathrm{d}\mathbf{x}\,\mathrm{d}t + \int_0^T \int_\Gamma J_\Gamma[C_\Gamma\,\mathbf{u}]\,\mathrm{d}s\,\mathrm{d}t + \int_\Omega K_\Omega[E_\Omega\,\mathbf{u}\,]_{t=T}\,\mathrm{d}\mathbf{x}\ . \tag{2}$$

The differential operators C_Ω and E_Ω act on the domain Ω, while C_Γ is a boundary operator (all are assumed to be Frechét differentiable). Their Frèchet derivatives are denoted by C'_Ω, E'_Ω, and C'_Γ, respectively. Also, let $j_\Omega := \left(J'_\Omega[C_\Omega\,\mathbf{u}]\right)^T$, $j_\Gamma := \left(J'_\Gamma[C_\Gamma\,\mathbf{u}]\right)^T$, and $k_\Omega := \left(K'_\Omega[E_\Omega\,\mathbf{u}\,]\right)^T$. To fix notation, we define the following inner products on the space-time domains $[0,\,T]\times\Omega$ and $[0,\,T]\times\Gamma$:

$$\langle \mathbf{u},\,\mathbf{v}\rangle_{[0,\,T]\times\Omega} := \int_0^T \int_\Omega \mathbf{u}\,\mathbf{v}\,\mathrm{d}\mathbf{x}\ ,\ \ \langle \mathbf{u},\,\mathbf{v}\rangle_{[0,\,T]\times\Gamma} := \int_0^T \int_\Gamma \mathbf{u}\,\mathbf{v}\,\mathrm{d}\mathbf{s}\ ,\ \forall\ \mathbf{u},\,\mathbf{v} \in \mathcal{U}\ .$$

Similar inner products are defined for the discrete variables.

An inverse problem is typically formulated as a constrained minimization problem for (2). In the continuous approach, also called *differentiate-then-discretize*, one defines the minimization problem in terms of the analytical model formulation:

$$\text{Find}\ \ \mathbf{u}^0_* = \arg\min_{\mathbf{u}^0 \in \mathcal{U}} \mathcal{J}(\mathbf{u})\ ,\ \ \text{subject to}\ \ (2)\ . \tag{3}$$

We instead focus on the discrete approach, also called *discretize-then-optimize*. The optimization problem is formulated in terms of the Runge-Kutta discontinuous Galerkin (RK–DG) space-time discretization of the continuous equations (2). For simplicity, we assume that the discrete domain Ω^h, and its boundary Γ^h cover exactly the analytical domains Ω and Γ. A discretization in space of the continuous primal problem (1) leads to the semi-discrete model:

Find $\mathbf{u}^h \in \mathcal{L}^2([0,\,T]\,;\,\mathcal{U}_h)$ such that $(\mathbf{u}^h)_t \in \mathcal{L}^2([0,\,T]\,;\,\mathcal{U}_h)$ and

$$\begin{aligned} \left\langle \frac{\partial \mathbf{u}^h}{\partial t},\,\mathbf{v}^h(t)\right\rangle_\Omega &= \mathcal{N}\left(t;\mathbf{u}^h,\,\mathbf{v}^h\right) + \left\langle \mathbf{f}^h,\,\mathbf{v}^h\right\rangle_\Omega + \mathcal{B}\left(\mathbf{g}^h,\,\mathbf{v}^h\right)\ , \\ &\qquad \forall\ \mathbf{v}^h \in \mathcal{L}^2([0,T];\mathcal{U}_h)\ ,\ \text{a.a.}\ t \in [0,T]\ . \end{aligned} \tag{4}$$

Here the semi-linear form $\mathcal{N}$ is nonlinear in $\mathbf{u}^h$, and linear in the test functions $\mathbf{v}^h$. $\mathcal{B}(\cdot,\cdot)$ is a bilinear form defined on the boundary Γ, which depends on the prescribed boundary data $\mathbf{g}^h$. The semi-discrete cost functional

$$\mathcal{J}_h(\mathbf{u}^h) = \int_0^T \int_\Omega J_\Omega[C_\Omega\,\mathbf{u}^h]\,\mathrm{d}\mathbf{x}\,\mathrm{d}t + \int_0^T \int_\Gamma J_\Gamma[C_\Gamma\,\mathbf{u}^h]\,\mathbf{ds}\,\mathrm{d}t + \int_\Omega K_\Omega[E_\Omega\,\mathbf{u}^h\,]_{t=T}\,\mathrm{d}\mathbf{x}\;, \quad (5)$$

is a discretization in space of the continuous functional (2). The time integrals in (5) will be evaluated using Runge–Kutta quadratures of order p.

A full discretization of the PDE is obtained by discretizing the time derivative in (4) using an s-stage, pth order Runge–Kutta method [12]. Let $\mathbf{U}_h$ be a Sobolev space of broken piecewise polynomial functions. In the following, $\mathbf{u}^n \in \mathcal{U}_h \subset \mathcal{U}$ is the fully discrete solution at t^n, $\mathbf{U}_i^n \in \mathcal{U}_h$ is the i-th stage vector at time step n, and $T_i^n = t^n + c_i\,h^{n+1}$ is the stage time moment. The time grid has $N+1$ points: from $t^0 = 0$, up to $t^N = T$, and $t^{n+1} = t^n + \tau^{n+1}$. For simplicity of notation, we omit the discrete space superscripts in the following discussion. The Runge–Kutta discretization of (4) reads:

$$\begin{aligned}
\langle \mathbf{U}_i^n\,,\,\mathbf{v}\rangle_\Omega &= \langle \mathbf{u}^n, \mathbf{v}\rangle_\Omega + \tau^{n+1}\sum_{j=1}^{s} a_{i,j}\left[\,\mathcal{N}\left(T_j^n; \mathbf{U}_j^n,\,\mathbf{v}\right) + \left\langle \mathbf{f}_j^n,\,\mathbf{v}\right\rangle + \mathcal{B}\left(\mathbf{g}_j^n,\,\mathbf{v}\right)\right]\;,\;\forall\,\mathbf{v}\in\mathcal{U}_h \quad (6)\\
\left\langle \mathbf{u}^{n+1},\mathbf{v}\right\rangle_\Omega &= \langle \mathbf{u}^n, \mathbf{v}\rangle_\Omega + \tau^{n+1}\sum_{i=1}^{s} b_i\,\left[\,\mathcal{N}\left(T_i^n; \mathbf{U}_i^n,\,\mathbf{v}\right) + \left\langle \mathbf{f}_i^n,\,\mathbf{v}\right\rangle + \mathcal{B}\left(\mathbf{g}_i^n,\,\mathbf{v}\right)\right]\;,\; n = 0\ldots N-1\;.
\end{aligned}$$

Thus, the discrete inverse problem formulation reads:

$$\text{Find } \mathbf{u}_*^{h,0} = \arg\min_{\mathbf{u}^h\in\mathcal{U}_h} \mathcal{J}_h(\mathbf{u}^{h,0:N})\;,\;\text{ subject to }\;(6)\;. \quad (7)$$

We now briefly present the adjoint framework for both continuous and discrete problems. For the full details of the derivations, the reader can refer to [11].

2.1. The adjoint framework in function spaces

The concept of adjoint consistency, together with its implications in optimization, have been investigated for steady-state problems by Lu [9], Harriman, Gavaghan and Süli [13], Hartmann and Houston [14, 15], and Oliver and Darmofal [16]. We leverage previous results on dual consistency for temporal [17], and spatial discretizations [14], to give a unified framework for the analysis of adjoint consistency of space-time DG discretizations. This section discusses space-time duality relations for continuous model formulations. We outline a general strategy for the derivation of the adjoint problem. This is applicable whenever the cost functional, and the associated model differential operators, satisfy a set of compatibility conditions [11]. The next section will discuss dual consistency of the time quadratures for Runge–Kutta DG discretizations (assumed to be dual consistent in space).

The variation of the cost functional (2) is

$$\delta\mathcal{J} = \mathcal{J}'[\mathbf{u}]\,\mathbf{w} := \left\langle C_\Omega'\,\mathbf{w}\;,\;j_\Omega\right\rangle_{[0,T]\times\Omega} + \left\langle C_\Gamma'\,\mathbf{w}\;,\;j_\Gamma\right\rangle_{[0,T]\times\Gamma} + \left\langle E_\Omega'\,\mathbf{w}\;,\;k_\Omega\right\rangle_\Omega\Big|_{t=T}\;. \quad (8)$$

We wish to express the variation (8) as

$$\delta\mathcal{J} \;=\; \left\langle C_\Omega^{\mathrm{adj}}\,\boldsymbol{\lambda}\;,\;\delta\mathbf{f}\right\rangle_{[0,T]\times\Omega} + \left\langle C_\Gamma^{\mathrm{adj}}\,\boldsymbol{\lambda}\;,\;\delta\mathbf{g}\right\rangle_{[0,T]\times\Gamma} + \left\langle E_\Omega^{\mathrm{adj}}\,\boldsymbol{\lambda}|_{t=0}\;,\;\delta\mathbf{u}_0\right\rangle_\Omega\;, \quad (9)$$

for any perturbations $\delta\mathbf{u}^0$, $\delta\mathbf{f}$, and $\delta\mathbf{g}$. The adjoint variables $\boldsymbol{\lambda}$ are obtained by solving the dual problem

$$\begin{aligned} -\boldsymbol{\lambda}_t &= L^* \boldsymbol{\lambda} + \mathbf{f}^{\text{adj}} \ , \ \mathbf{x} \in \Omega \ , \ t \in [0,T] \\ B^{\text{adj}} \boldsymbol{\lambda} &= \mathbf{g}^{\text{adj}} \ , \ \mathbf{x} \in \Gamma \ , \ t \in [0,T] \\ \boldsymbol{\lambda}(t=T,\mathbf{x}) &= E_\Omega^{\text{adj}} k_\Omega \ , \ \mathbf{x} \in \Omega \ . \end{aligned} \tag{10}$$

Here the * superscript denotes an adjoint operator. The differential operators B^{adj} and C_Γ^{adj} need to be chosen such that (8) and (9) are equivalent. To determine these adjoint volume and boundary differential operators, one needs to impose several *compatibility conditions* on L, C'_Ω, and E'_Ω [11]. If these conditions are satisfied, then we say that the cost functional and the primal PDE are *compatible*. For $\mathcal{J}$ defined by (2), the adjoint system (10) becomes:

$$\begin{aligned} -\boldsymbol{\lambda}_t &= L^* \boldsymbol{\lambda} + C_\Omega'^* \, j_\Omega \ , \ \mathbf{x} \in \Omega \ , \ t \in [0,T] \\ B^{\text{adj}} \boldsymbol{\lambda} &= j_\Gamma \ , \ \mathbf{x} \in \Gamma \ , \ t \in [0,T] \\ \boldsymbol{\lambda}(t=T,\mathbf{x}) &= E'^* k_\Omega \ . \end{aligned} \tag{11}$$

Assume that

$$\langle L\,\mathbf{w},\, \mathbf{v}\rangle_\Omega = \langle \mathbf{w},\, L^*\,\mathbf{v}\rangle_\Omega + \sum_i \left\langle F_i^L\,\mathbf{w},\, G_i^L \mathbf{v} \right\rangle_\Gamma \ , \ \forall\ \mathbf{w},\, \mathbf{v} \in \mathcal{U} \ ,$$

where the operators F_i^L, and G_i^L are determined through integration by parts. If the adjoint boundary condition is defined by the compatibility condition [11]:

$$\left\langle B^{\text{adj}}\,\boldsymbol{\lambda} \ ,\ C'_\Gamma \mathbf{w} \right\rangle_{[0,T]\times\Gamma} = \left\langle C_\Gamma^{\text{adj}}\,\boldsymbol{\lambda} \ ,\ B'\,\mathbf{w} \right\rangle_{[0,T]\times\Gamma} - \sum_i \left\langle F_i^L\,\mathbf{w}\,, G_i^L\,\boldsymbol{\lambda} \right\rangle_{[0,T]\times\Gamma} \ , \tag{12}$$

then the variation of the cost functional $\mathcal{J}$ can be expressed in terms of the dual variables:

$$\mathcal{J}'[\mathbf{u}]\,\mathbf{w} = \left\langle \delta\mathbf{u}^0 \ ,\ \boldsymbol{\lambda} \right\rangle_\Omega \Big|_{t=0} + \langle \delta\mathbf{f} \ ,\ \boldsymbol{\lambda}\rangle_{[0,T]\times\Omega} + \left\langle C_\Gamma^{\text{adj}}\,\boldsymbol{\lambda} \ ,\ \delta\mathbf{g} \right\rangle_{[0,T]\times\Gamma} \ . \tag{13}$$

Thus the adjoint variable $\boldsymbol{\lambda}$ can be interpreted as the sensitivities to changes in the initial condition $\mathbf{u}^0$, right-hand side $\mathbf{f}$, or boundary forcing $\mathbf{g}$, depending on the particular choice of direction of differentiation $\mathbf{w}$.

2.2. Duality relations for space-time discretizations

Due to the linearity of the Runge–Kutta procedure, the tangent linear model (TLM) of the fully discrete system (6) reads:

$$\begin{aligned} \langle \mathbf{W}_i^n \ ,\ \mathbf{v}\rangle_\Omega &= (\mathbf{w}^n, \mathbf{v})_\Omega + \tau^{n+1} \sum_{j=1}^s a_{i,j} \left[\mathcal{N}'[\mathbf{U}_j^n]\left(T_j^n; \mathbf{W}_j^n \ ,\ \mathbf{v}\right) + \left\langle \delta\mathbf{f}_j^n,\ \mathbf{v} \right\rangle + \mathcal{B}\left(\delta\mathbf{g}_j^n,\ \mathbf{v}\right) \right] \\ \left\langle \mathbf{w}^{n+1},\ \mathbf{v} \right\rangle_\Omega &= \langle \mathbf{w}^n,\ \mathbf{v}\rangle_\Omega + \tau^{n+1} \sum_{i=1}^s b_i \left[\mathcal{N}'[\mathbf{U}_i^n]\,(T_i^n; \mathbf{W}_i^n,\ \mathbf{v}) \ +\ \langle \delta\mathbf{f}_i^n,\ \mathbf{v}\rangle + \mathcal{B}\,(\delta\mathbf{g}_i^n,\ \mathbf{v}) \right] \ , \ \forall\, \mathbf{v} \in \mathcal{U}_h \ . \end{aligned}$$

Rewriting the tangent linear model of the fully discrete system to use different test functions $\boldsymbol{\lambda}^n \in \mathcal{U}_h$ (later recast as the adjoint variables), summing up over all time steps $n = 0 \ldots N-1$, and

finally identifying the terms involving the same test functions on the left and right hand sides, we can define the fully discrete adjoint system as:

$$\begin{aligned}
\langle \mathbf{w},\, \theta_i^n \rangle_\Omega &= \tau^{n+1}\, \mathcal{N}'[\mathbf{U}_i^n] \left(T_i^n; \mathbf{w}\,,\, b_i \lambda^{n+1} + \sum_{\ell=1}^{s} a_{\ell,i}\, \theta_\ell^n \right) - \tau^{n+1} \sum_{i=1}^{s} b_i \left\langle j_\Omega[C_\Omega\, \mathbf{U}_i^n]\,,\, C'_\Omega\, \mathbf{w} \right\rangle_\Omega \\
&\quad - \tau^{n+1} \sum_{i=1}^{s} b_i \left\langle j_\Gamma[C_\Gamma\, \mathbf{U}_i^n]\,,\, C'_\Gamma\, \mathbf{w} \right\rangle_\Gamma \;,\; \forall\, \mathbf{w} \in \mathcal{U}_h \\
\langle \mathbf{w},\, \lambda^n \rangle_\Omega &= \left\langle \mathbf{w},\, \lambda^{n+1} \right\rangle_\Omega + \sum_{i=1}^{s} \langle \mathbf{w},\, \theta_i^n \rangle_\Omega \;,\; \forall\, \mathbf{w} \in \mathcal{U}_h \,. \qquad (14)
\end{aligned}$$

Sum up the TLM relations for $n = 0, \ldots, N-1$, to get:

$$\left\langle \mathbf{w}^N,\, \lambda^N \right\rangle_\Omega = \left\langle \delta\mathbf{u}^0,\, \lambda^0 \right\rangle_\Omega - \mathcal{J}'_h\, \mathbf{w}^0 + \left\langle \left(k'_\Omega[E_\Omega\, \mathbf{u}^N\,] \right)^T \;,\; E'_\Omega\, \mathbf{w}^N \right\rangle_\Omega + S_\mathbf{f} + S_\mathbf{g} \,. \qquad (15)$$

with

$$S_\mathbf{f} = \sum_{n=0}^{N-1} \tau^{n+1} \sum_{i,j=1}^{s} a_{i,j} \left\langle \delta\mathbf{f}_j^n,\, \theta_i^n \right\rangle_\Omega + \sum_{n=0}^{N-1} \tau^{n+1} \sum_{i=1}^{s} b_i \left\langle \delta\mathbf{f}_i^n,\, \boldsymbol{\lambda}^{n+1} \right\rangle_\Omega \,, \qquad (16)$$

and

$$S_\mathbf{g} = \sum_{n=0}^{N-1} \tau^{n+1} \sum_{i,j=1}^{s} a_{i,j}\, \mathcal{B}\left(\delta\mathbf{g}^{n,j},\, \theta_i^n \right) + \sum_{n=0}^{N-1} \tau^{n+1} \sum_{i=1}^{s} b_i\, \mathcal{B}\left(\delta\mathbf{g}_i^n,\, \lambda^{n+1} \right) \,. \qquad (17)$$

Using the theory of formal Runge–Kutta adjoints in optimal control [18, 19], it can be shown [11] that, as $\tau := \max_n \tau^n \to 0$, and h is kept constant,

$$\begin{aligned}
S_\mathbf{f} &= \langle\, \boldsymbol{\lambda},\, \delta\mathbf{f}\, \rangle_{[0,T]\times\Omega} + O(\tau^p) \\
S_\mathbf{g} &= \int_0^T \mathcal{B}(\delta\mathbf{g},\, \boldsymbol{\lambda})\, \mathrm{d}t + O(\tau^p) \,.
\end{aligned}$$

where p is the temporal order of accuracy of the primal Runge-Kutta method (6).

We define the final adjoint condition by

$$\left\langle \lambda^N,\, \mathbf{w} \right\rangle_\Omega = \left\langle k'_\Omega[E_\Omega\, \mathbf{u}^N\,] \,,\, E'_\Omega\, \mathbf{w} \right\rangle_\Omega \;,\; \forall\, \mathbf{w} \in \mathcal{U}_h \,. \qquad (18)$$

Then, (15) becomes:

$$\mathcal{J}'_h\, \mathbf{w} \approx \left\langle \delta\mathbf{u}^0,\, \lambda^0 \right\rangle_{[0,T]\times\Omega} + \langle \delta\mathbf{f},\, \boldsymbol{\lambda} \rangle_{[0,T]\times\Omega} + \int_0^T \mathcal{B}\,(\delta\mathbf{g},\, \boldsymbol{\lambda})\; \mathrm{d}t \,.$$

The discrete adjoint variables $\lambda^{h,n}$ can yield different sensitivities, depending on the direction in which the Fréchet derivative of $\mathcal{J}_h$ is computed. The gradients of $\mathcal{J}_h$ with respect to the initial state $\mathbf{u}^0$, and volume forcing $\mathbf{f}$, can be computed by differentiation along $(\delta\mathbf{u}^0,\, 0,\, 0)$, and $(0,\, \delta\mathbf{f},\, 0)$, respectively. However, consistency of the boundary sensitivities, obtained by differentiation along $(0,\, 0,\, \delta\mathbf{g})$, is not guaranteed unless the boundary discretization $\mathcal{B}$ is dual consistent [11].

The discrete adjoint model (14)–(18) is obtained by applying the discrete Runge–Kutta adjoint numerical method to the semi-discrete adjoint system [11]. According to [17] the discrete adjoint Runge Kutta method has the same order of accuracy as the forward Runge Kutta method.

3. Discrete adjoints of intergrid transfer operators

The primal-dual solution procedure for (6)–(14) can be written as follows:

$$\begin{aligned} \mathbf{u}^{h,0} &= \mathbf{u}^{h,0}(\mathbf{x}^h) \\ \mathbf{u}^{h,n+1} &= \mathcal{I}_{n\to n+1}\left(\mathcal{S}_{n\to n+1}\,\mathbf{u}^{h,n}\right)\ ,\ 0 \le n \le N-1 \\ \boldsymbol{\lambda}^{h,N} &= \boldsymbol{\lambda}^{h,N}(\mathbf{x}^h) \\ \boldsymbol{\lambda}^{h,n} &= \mathcal{S}'^{*}_{n+1\to n}\left(\mathcal{I}^{T}_{n\to n+1}\,\boldsymbol{\lambda}^{h,n+1}\right)\ ,\ N-1 \ge n \ge 0\ . \end{aligned} \tag{19}$$

Here $\mathcal{S}_{n\to n+1}$ is the discrete solution operator that advances the discrete solution from t^n to t^{n+1}. $\mathcal{S}'^{*}_{n+1\to n}$ is the discrete dual (transpose) of the linearization of $\mathcal{S}$, and is used in the backward time evolution of the discrete adjoint solution $\boldsymbol{\lambda}^{h,n}$. The final values of the adjoint variables depend on the specific form of the cost functional $\mathcal{J}_h$ under consideration. The derivation and consistency properties of $\mathcal{S}^*$ have been discussed in the previous section; we are now concerned with the dual consistency of the intergrid transfer operators $\mathcal{I}_{n\to n+1}$. More precisely, due to the orthogonality property of $\mathcal{L}^2$-interpolation and restriction for polynomial expansion approximations used with discontinuous Galerkin, we have that

$$\mathcal{I}^{T}_{n\to n+1} = \mathcal{I}_{n+1\to n}\ . \tag{20}$$

This transpose relationship holds for both hierarchical refinement, and on general (unstructured) grids. See [11] for a proof of each case. Equation (20) implies that the discrete adjoints of grid interpolation and restriction operations are identical to the coarsening and interpolation operators used in the discretization of the continuous adjoint problem. This is important in practice, since $\mathcal{I}^{T}_{n\to n+1}$ can be easily generated from $\mathcal{I}_{n\to n+1}$ using automatic differentiation. Automatic code generation makes coding less error-prone (as opposed to hand-coded derivatives and adjoints), and reduces the total development effort.

We note that (20) does not hold for finite volume discretizations [11]. For this particular type of discretization method [20], quadratic (or higher order) interpolation operators, reduce to first order averaging when transposed in the discrete adjoint model. Moreover, due to the stencil structure of the interpolants, first-order perturbations are added to the solution averages of neighboring cells. Post-processing is required to cancel the effect of these undesired perturbations (see [11] for more details).

4. A one-dimensional test problem

The one-dimensional data assimilation problem is formulated as:

$$\min_{\mathbf{u}^0} \mathcal{J}^h\left(\mathbf{u}^0\right)\ , \tag{21}$$

subject to

$$\mathbf{u}_t + 2\,\mathbf{u}_x = f(x)\ ,\ x \in \Omega = [-3\pi,\ 3\pi]\ ,\ t \in [0, T]\ , \tag{22}$$

with $\mathbf{u}^0(x) = \exp(-|x|)\sin(x)\cos(x)$, periodic boundary conditions, and $T = 0.5$. Equation (22) is discretized with a upwind discontinuous Galerkin method, coupled with a fourth order TVD Runge-Kutta scheme for time marching [21]. The space-time dual consistency for this discretization can be easily proven (see, e.g., [8] for the spatial discretization).

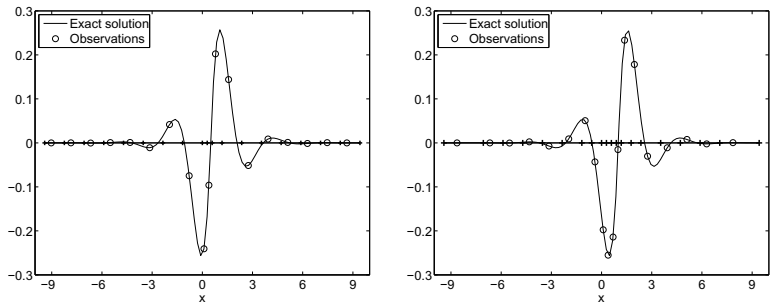

Figure 1: Reference observations (circles), and the exact solution (continuous line) of (22) at $t^1 = 0.25$ (left) and $t^2 = 0.5$ (right). The adaptive spatial mesh is marked on the x-axis. It is locally refined in areas of high variations in the primal solution $\mathbf{u}^{h,n}$. The refinement is done using an element-wise error estimator based on a finite-difference approximation to the solution gradient $\mathbf{u}_x$.

The discrete cost functional reads:

$$\mathcal{J}_h = \mathcal{J}_B^h + \mathcal{J}_O^h = \frac{1}{2}\left(\mathbf{u}^0 - \mathbf{u}^B\right)^T B^{-1}\left(\mathbf{u}^0 - \mathbf{u}^B\right) + \frac{1}{2}\sum_{k=1}^{K}\left(\mathcal{H}_k\,\mathbf{u}^k - y^k\right)^T R_k^{-1}\left(\mathcal{H}_k\,\mathbf{u}^k - y^k\right) . \quad (23)$$

The background term $\mathcal{J}_h^B$ quantifies the departure of the inverse solution from a background state $\mathbf{u}_B^{h,0}$. It also acts as a regularization term that makes the inverse problem well-posed. $\mathcal{J}_h^O$ quantifies the mismatch between the model predictions and a set of *a priori* available observations y^k at selected grid locations and time points (see Figure 1). B and R_k are the error covariance matrices for the background state $\mathbf{u}_B^{h,0}$, and the observation values at t_k, respectively. Finally, $\mathcal{H}_k$ is a linear observation operator that maps the discrete model state to the observation space.

The setup is that of a classical *twin* experiment: the observations are recorded during a reference run of the model, starting from the reference solution $\mathbf{u}_{\text{ref}}^{h,0}$. The background state $\mathbf{u}_B^{h,0} = 1.4\,\mathbf{u}_{\text{ref}}^{h,0}$ is the initial guess for the optimization routine. We use a 5th order discontinuous basis for the reference run, whereas the inversion is done using only cubic approximations, to avoid the "inverse crime" [22]. The goal of the inversion process is to retrieve a good approximation to the reference initial condition as the *a posteriori* analysis state: $\mathbf{u}_A^{h,0} \approx \mathbf{u}_{\text{ref}}^{h,0}$.

4.1. Numerical results

The data assimilation procedure follows the description in [10]. The observation times are $t^1 = 0.25$, and $t^2 = 0.5$ (figure 1). Our tests were run with a C++ implementation [23] of the limited memory BFGS method by Nocedal et al. [24, 25]. Figure 2 shows the results of the assimilation experiment. It can be seen that the discrete adjoint gradient leads to a considerable decrease in the cost function value, as well as in the RMS error of the computed *a posteriori* analysis $\mathbf{u}_A^{h,0}$. We computed the analysis root-mean square error (RMSE) ϵ using the formula:

$$\epsilon := \frac{\left\| \mathbf{u}_A^{h,0} - \mathbf{u}_{\text{ref}}^{h,0} \right\|_{L^2(\Omega)}}{\left\| \mathbf{u}_{\text{ref}}^{h,0} \right\|_{L^2(\Omega)}} . \quad (24)$$

Figure 3 shows that the quality of the analysis obtained through the 4D-Var process is much better than that of the initial guess (the background $\mathbf{u}_B^{h,0}$).

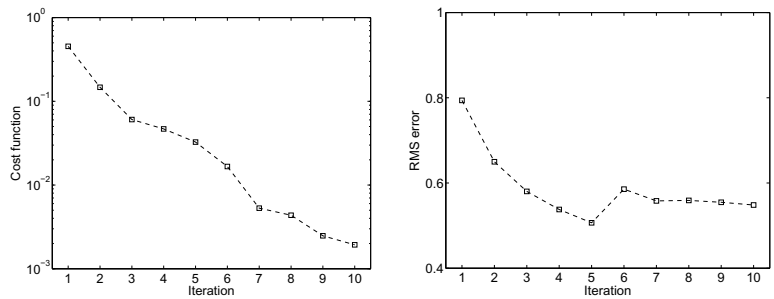

Figure 2: Relative decrease (i.e. ratio to the background values) of the cost function $\mathcal{J}_h$ (a), and of the RMSE (b).

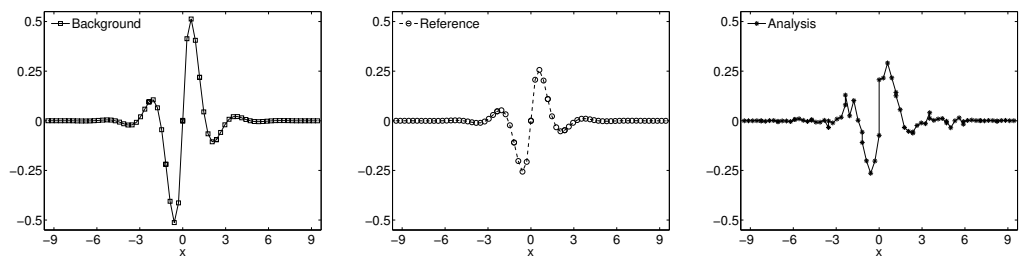

Figure 3: Background, reference, and analysis states at $t = 0$ for the problem (21)–(22).

5. Conclusions and future research directions

Time step and space mesh adaptivity have become necessary features in the solution of large scale inverse problems. This paper develops a framework for the construction of discrete adjoints of discontinuous Galerkin discretizations for time-dependent problems. We discuss dual consistency for space-time problems, a concept that has, to the authors' best knowledge, not been considered before in the literature. A straightforward extension of existing frameworks [9, 8] to include the time dimension, is possible only if the time integrals are discretized using a DG quadrature. If the more common Runge–Kutta methods are used for time discretization, we show that dual consistency is preserved. Moreover, the order of accuracy of the discrete adjoint time discretization is identical to the one of the primal method. This is a very useful property in practice, since discrete adjoints of RK-DG discretizations can be generated with low effort using AD.

Solution intergrid transfer operators (interpolation and restriction) play an important role in the inversion process. Equation (20) implies that the adjoints of the primal interpolation and restriction operators are valid intergrid operators in the discrete adjoint solution process. As noted before, the use of AD will reduce the complexity of adjoint code generation for mesh transfer operators.

We demonstrate the discrete-adjoint based inversion technique for a space-time adaptive, dual consistent DG discretization of the advection equation. The gradient-based optimization procedure yields a good quality analysis.

Error-driven adaptation of the optimal solution mesh, guided by primal/dual *a posteriori* error estimates, is the subject of on-going research. A forthcoming paper will analyze in detail the space-time optimality system for the discrete problem (6)–(7), and its relationship with the continuous optimality equations. The aim is to obtain general error estimates for the discrete primal and dual solutions that will guarantee convergence of the inverse problem solution.

References

[1] A. Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation, 2nd Edition, SIAM, Philadelphia, PA, USA, 2005.

[2] W. Bangerth, A framework for the adaptive finite element solution of large-scale inverse problems, SIAM J. Sci. Comput. 30 (6) (2008) 2965–2989.

[3] S. Li, L. Petzold, Adjoint sensitivity analysis for time-dependent partial differential equations with adaptive mesh refinement, J. Comput. Phys. 198 (1) (2004) 310–325.

[4] F. Fang, C. Pain, M. Piggott, G. Gorman, A. Goddard, An adaptive mesh adjoint data assimilation method applied to free surface flows, International Journal for Numerical Methods in Fluids 47 (8-9) (2005) 995–1001.

[5] F. Fang, M. Piggott, C. Pain, G. Gorman, A. Goddard, An adaptive mesh adjoint data assimilation method, Ocean Modelling 15 (1-2) (2006) 39–55, the Third International Workshop on Unstructured Mesh Numerical Modelling of Coastal, Shelf and Ocean Flows.

[6] F. Fang, C. C. Pain, M. D. Piggott, G. J. Gorman, A. J. H. Goddard, Data assimilation of three-dimensional free surface flow using an adaptive adjoint method. part 1: Adjoint model formulation, Tech. rep., Imperial College London, London, UK (2007).

[7] M. B. Giles, N. Pierce, Adjoint equations in CFD: duality, boundary conditions and solution behavior, in: Proceedings of the 13th Computational Fluid Dynamics Conference, no. AIAA Paper 97-1850, 1997.

[8] R. Hartmann, Adjoint consistency analysis of discontinuous Galerkin discretizations, SIAM J. Numer. Anal. 45 (6) (2007) 2671–2696.

[9] J. Lu, An a posteriori error control framework for adaptive precision optimization using the discontinuous Galerkin finite element method, Ph.D. thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, Cambridge, MA (2005).

[10] A. Sandu, L. Zhang, Discrete second order adjoints in atmospheric chemical transport modeling, J. Comput. Phys. 227 (12) (2008) 5949–5983. doi:http://dx.doi.org/10.1016/j.jcp.2008.02.011.

[11] M. Alexe, A. Sandu, Space-time adaptive solution of inverse problems with the discrete adjoint method, Tech. Rep. TR-10-14, Blacksburg, VA, USA (2010).

[12] E. Hairer, G. Wanner, Solving Ordinary Differential Equations: Stiff and Differential-Algebraic Problems, Vol. II of Computational Mathematics, Springer-Verlag, 1994.

[13] K. Harriman, D. Gavaghan, E. Süli, The importance of adjoint consistency in the approximation of linear functionals using the discontinuous Galerkin finite element method, Tech. Rep. 04/18, Oxford University Computing Lab, Numerical Analysis Group, Oxford, UK (2004).

[14] R. Hartmann, Error estimation and adjoint based refinement for an adjoint consistent dg discretisation of the compressible euler equations, Int. J. Comput. Sci. Math. 1 (2-4) (2007) 207–220.

[15] R. Hartmann, P. Houston, An optimal order interior penalty discontinuous Galerkin discretization of the compressible Navier-Stokes equations, Journal of Computational Physics 227 (22) (2008) 9670–9685.

[16] T. Oliver, D. L. Darmofal, Analysis of dual consistency for discontinuous Galerkin discretizations of source terms, SIAM Journal on Numerical Analysis 47 (5) (2009) 3507–3525. doi:10.1137/080721467.

[17] A. Sandu, On the properties of Runge - Kutta discrete adjoints, in: International Conference on Computational Science (4), 2006, pp. 550–557.

[18] W. W. Hager, Runge-Kutta methods in optimal control and the transformed adjoint system, Numerische Mathematik 87 (2) (2000) 247–282.

[19] A. Sandu, P. Miehe, Forward, tangent linear, and adjoint Runge-Kutta methods in KPP–2.2 for efficient chemical kinetic simulations, Tech. Rep. TR-06-17, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA (2006).

[20] R. J. Leveque, Finite volume methods for hyperbolic methods, Vol. 21 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, UK, 2002.

[21] S. Gottlieb, L.-A. J. Gottlieb, Strong stability preserving properties of Runge–Kutta time discretization methods for linear constant coefficient operators, J. Sci. Comput. 18 (1) (2003) 83–109.

[22] J. Kaipio, E. Somersalo, Statistical inverse problems: Discretization, model reduction and inverse crimes, Journal of Computational and Applied Mathematics 198 (2) (2007) 493–504, applied Computational Inverse Problems.

[23] libLBFGS: a library of limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) (2009).
URL `http://www.chokkan.org/software/liblbfgs/`
[24] R. H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM Journal on Scientific Computing 16 (5) (1995) 1190–1208.
[25] C. Zhu, R. H. Byrd, P. Lu, J. Nocedal, Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization, ACM Transactions on Mathematical Software 23 (4) (1997) 550–560.