Second Order Adjoints in Optimization

Noémi Petra and Ekkehard W. Sachs

Abstract Second order, Newton-like algorithms exhibit convergence properties superior to gradient-based or derivative-free optimization algorithms. However, deriving and computing second order derivatives—-needed for the Hessian-vector product in a Krylov iteration for the Newton step—often is not trivial. Second order adjoints provide a systematic and efficient tool to derive second derivative information. In this paper, we consider equality constrained optimization problems in an infinite-dimensional setting. We phrase the optimization problem in a general Banach space framework and derive second order sensitivities and second order adjoints in a rigorous and general way. We apply the developed framework to a partial differential equation-constrained optimization problem.

Keywords Adjoint-based methods · Second order adjoints · Optimization in infinite dimensions · Newton method · PDE-constrained optimization

1 Introduction

We consider second order methods for equality constrained optimization problems in an infinite-dimensional setting. These methods are locally very fast, but on the downside require substantial computational effort for each iteration. To alleviate this disadvantage, variants of Newton's methods have been developed, e.g., Newton-Shamanskii, where the Jacobian is fixed throughout iterations, or inexact Newton methods, where the Newton step is computed inexactly. In this paper we devote ourselves to an adjoint-based framework that extends from the usual gradient computation all the way to an efficient way to compute Hessian-vector products. This opens up a venue to use iterative solvers for the computation of the Newton step.

N. Petra (✉)
Department of Applied Mathematics, University of California, Merced, CA 95340, USA
e-mail: npetra@ucmerced.edu

E. W. Sachs
Department of Mathematics, Trier University, 54286 Trier, Germany
e-mail: sachs@uni-trier.de

M. Al-Baali et al. (eds.), *Numerical Analysis and Optimization*, Springer Proceedings in Mathematics & Statistics 354, https://doi.org/10.1007/978-3-030-72040-7_10

Unlike in the case of approximating the Hessian-vector product by a finite-difference quotient, which carries issues about its accuracy, the use of second order adjoints do not exhibit this problem and can be computed to required accuracy, if needed.

Related work. The concept of second order adjoints can be found in different fields in the literature, also under different names, for example incremental adjoints and adjoint sensitivities. The earliest work using second order adjoints goes back to papers in the sixties [29, 32]. Second order adjoints have been widely applied especially to inverse problems governed by differential equations. Here we give a sample of the literature on applications to various optimization problems constrained by differential equations: seismology [41], structural optimization [20, 21, 27], unconstrained discrete-time and continuous-time optimal control problems with Bolza objective functions [13], ordinary differential equations (ODEs) and partial differential equations (PDEs)-constrained optimization problems in the context of air traffic flow [38], optimization problems governed by PDEs with inequality constraints [16, 17] or parabolic PDEs [6], optimal semiconductor design based on the standard drift diffusioxn model [26], open loop optimal control problems governed by the two-dimensional stationary Navier-Stokes equations [25], calculation of directional derivatives of stiff ODE embedded functionals [35], data assimilation for numerical weather prediction and ocean models [2, 3, 11, 12, 31, 40, 42], inversion of the initial concentration of the airborne contaminant in a convection-diffusion transport model [1], full wave form or global seismic inversion [7, 8, 14, 15, 33], inverse ice sheet modeling [28, 34, 37, 43], in the context of optimal experimental design [4], and optimal control of systems governed by PDEs with random parameter fields [5, 10], and inexact Hessian-vector products computed using approximate second order adjoints [24]. It is also worth mentioning the following very useful technical reports targeting model (academic) problems [19, 22, 36].

Contributions. The derivation of second order adjoints in infinite-dimensions are often motivated using ad-hoc analytic arguments, hence there is a need for a rigorous in-depth investigation of such derivation and adjoint expressions. The main goal of this paper is therefore to present a general and rigorous theory for second order adjoints which then can be applied to various applications. In addition, when second order optimization algorithms (e.g., Newton's method or variants) are applied to solve large-scale optimization problems often iterative solvers are used for the solution of the systems of linearized equations. Therefore, motivated by the need to avoid the computation of explicit Hessian information, we show in the general framework how the adjoint calculus can be applied to compute Hessian-vector products. We note that since we have two well studied pathways to compute first-order derivative information, e.g., via a sensitivity approach or via an adjoint approach, there are four pathways to obtain the second order derivatives. In this paper, we present all these four pathways and show that in fact three of them are the same. Finally, we apply our framework to an inverse problem that seeks to reconstruct a coefficient field in an elliptic PDE from observational data. This problem is formulated as a nonlinear least squares optimization problem governed by the Poisson problem. In this paper, we use the proposed second order adjoint derivation and Hessian-vector product expression to derive derivative information that can be used by fast optimization methods to solve complex optimization problems. We note that in [1, 5–8, 10, 14, 15, 28, 34, 36, 37, 43] second order adjoints are also considered in an infinite-dimensional setting, however, using the Lagrangian approach. Here, we recast the problem as an unconstrained optimization problem and derive the Lagrange multipliers, i.e., adjoint variables, directly.

Problem formulation. To set the stage, we choose a general framework: Let Y, U, Z be Banach spaces, e.g., Y is the space variable or dependent variable, U the space of control or design variables and Z the range space of the equality constraint. The distinction of the variable into y and u is essential for our approach, but can be found in many applications, see e.g., PDE-constrained optimization. The optimization problem is formulated as follows:

Problem 1

$$\begin{array}{ll} \min & \phi(y,u), \quad (y,u) \in Y \times U, \\ \text{s.t.} & g(y,u) = 0, \\ \text{where} & \phi : Y \times U \to I\!R, \quad g : Y \times U \to Z. \end{array} \tag{1}$$

If we assume that for each control variable u we have a unique solution $y = s(u)$ of the equality constraint $g(s(u), u) = 0$, then we can rewrite the constrained optimization problem as an unconstrained optimization problem:

$$\min \quad \Phi(u) = \phi(s(u), u), \quad u \in U, \quad \Phi : U \to I\!R.$$

In Newton's method, the correction step is defined as the solution of

$$\Phi''(u)v = -\Phi'(u), \quad v \in U, \tag{2}$$

where $\Phi'(u) \in U^*$ and $\Phi''(u) \in L(U, U^*)$, the space of linear operators mapping U into U^*. In many applications where the second derivative $\Phi''(u)$ is prohibitively expensive to compute in an explicit manner, the Eq. (2) is solved by an iterative technique, see for example Krylov methods [23]. In order to implement these methods efficiently, one needs a fast evaluation of the Hessian-vector product, in our notation $\Phi''(u)v$ for some $v \in U$. One way to achieve this is in the inexact Newton framework, where one uses the approximation

$$\Phi''(u)v \approx (\Phi'(u + hv) - \Phi'(u))/h. \tag{3}$$

The use of this approximation introduces an error at each iteration of the Krylov method which has to be handled with care, see e.g., [30].

However, if the optimization problem is of the type like shown in (1) there is another way of computing the Hessian-vector product. The reason for this lies in the fact that the variables are grouped into two groups y and u and the introduction of another adjoint variable, which we call the *second order adjoint*. This means that we have to solve two adjoint equations, i.e., equations of the type

$$g_y^* p = r_1, \qquad g_y^* \pi = r_2$$

for the first and second order adjoint variables p and π with different right hand sides r_1, r_2. The advantage of such an approach lies in the fact that we do not introduce an additional error by computing the Hessian-vector product exactly. Furthermore, the computational cost for the second order adjoint is not higher than an additional evaluation of the derivative $\Phi'(u + hv)$ as in (3).

It is well known that the derivative of the function $\phi(y, u)$ can be expressed in two ways, namely

– the sensitivity equations or
– the adjoint equations.

The first approach is considered reasonable for a small number of variables, whereas the second one requires a bit more analysis in the derivation, but shows to be highly efficient for large-scale problems [9].

If we turn to the second derivative applied to a vector as this required for iterative solvers like CG or GMRES, we are free to choose for this purpose again either the adjoint or sensitivity approach. Hence we have four different routes available which we could follow, namely

– first the sensitivity, then the adjoint approach or
– first the sensitivity, then the sensitivity approach or
– first the adjoint, then the sensitivity approach or
– first the adjoint, then the adjoint approach.

Content. In this paper we carefully analyze these four approaches and prove rigorously the results following these venues. It turns out that not four, but two different ways exist to compute the Hessian-vector products for this optimization problem. One approach is based on the sensitivity framework which is amenable for a small number of variables. The second one relies on the adjoint approach which leads to the concept of a second order adjoint that has to be computed for a Hessian-vector product. This approach is usually much more efficient, especially for problems with a large number of variables. The effort per iteration is comparable to if not lower than that of an inexact Newton's method where the matrix-vector multiplication is approximated by a finite difference quotient, yet it gives the precise result rather than an approximation.

The remaining sections of this paper are organized as follows. We begin by providing two venues to obtain representations of the first-order derivatives in Sect. 2. Next, in Sect. 3 we set the stage for the second order derivatives followed by the fourth section containing the results including proofs of the four approaches mentioned above. Section 5 is devoted to an application in PDE-constrained optimization, where we illustrate some of the theoretical results.

2 Representation of First-Order Fréchet-Derivative

For notational purposes we recall that a map $g : X \to Z$ from a Banach space X to another Banach space Z is called Fréchet-differentiable at $x \in X$, if there exists a linear operator denoted by $g'(x) : X \to Z$ such that

$$\|g(x+h) - g(x) - g'(x)h\|_Z \leq \alpha(\|h\|_X)\|h\|_X,$$

with a function $\alpha(r)$ satisfying $\alpha(r) \to 0$ for $r \to 0$. The partial Fréchet-derivatives of e.g., $g(y,u)$ is denoted by $g_y(y,u)$ or $g_u(y,u)$ with a subscript indicating the variable with respect to which the derivative is taken. The adjoint operator of $g'(x)$ is denoted by $g'(x)^* : Y^* \to X^*$. We note that Fréchet-derivatives of second order like $g_{yu}(y,u)$ are linear operators in the spaces $L(U, L(Y,Z)) = L(U \times Y, Z)$.

In what follows, we impose the following smoothness assumptions on the functions in the problem formulation.

Assumption 1 *Let the function ϕ and the mapping g be continuously Fréchet-differentiable on $Y \times U$.*

Furthermore we assume the following constraint qualification to hold at a later to be specified point $(y,u) \in Y \times U$.

Assumption 2 *For $(y,u) \in Y \times U$ let the partial Fréchet-derivative $g_y(y,u) : Y \to Z$ be surjective and invertible.*

With these assumptions we can apply the implicit function theorem [39].

Theorem 1 *Let Assumptions 1 and 2 hold at $(y_*, u_*) \in Y \times U$. Then there exist neighborhoods $B_Y \subset Y$ at y_* and $B_U \subset U$ at u_* and a Fréchet-differentiable map $s : B_U \to B_Y$ such that*

$$g(s(u),u) = 0 \quad \text{and} \quad g_y(s(u),u)s'(u) = -g_u(s(u),u). \tag{4}$$

This theorem can be used to reformulate the constrained optimization problem from above as an unconstrained optimization problem in a neighborhood around a local minimizer.

Corollary 1 *Let $(y_*, u_*) \in Y \times U$ be a local minimizer of optimization problem 1 and let Assumptions 1 and 2 hold at (y_*, u_*). Then u_* is also a local minimizer of the unconstrained optimization problem*

$$\min_{u \in B_U} \Phi(u), \quad \Phi(u) := \phi(s(u),u), \tag{5}$$

where $\Phi : B_U \to I\!R$, with $B_U \in U$ a neighborhood of u_*, and $s(u)$ given by the implicit function theorem.

In some applications the computation of $s(u)$ is theoretically possible, but numerically only feasible within a certain error tolerance. This happens, for example, for a constraint described by partial differential equations where $s(u)$ is the solution of a PDE. This would introduce an additional error into $s(u)$ and $s'(u)$ as well. The purpose of this paper is to derive in a rigorous way the expression for second derivatives without this additional complexity. This aspect, however, opens interesting venues for future research.

The necessary optimality conditions of first-order require various derivatives which are well defined under the statements above. Therefore the first derivative of the objective function of the unconstrained problem can be computed as follows. We note that this approach is usually denoted as the approach using the sensitivity equations.

Theorem 2 *Let Assumptions 1 and 2 hold at $(y, u) \in Y \times U$. Then the Fréchet-derivative of $\Phi(u)$ applied to $\Delta u \in U$ is given by*

$$\Phi'(u)\Delta u = \phi_y(s(u), u)\xi + \phi_u(s(u), u)\Delta u, \tag{6}$$

where $\xi = s'(u)\Delta u \in Y$ is the unique solution of the sensitivity equation

$$g_y(s(u), u)\xi = -g_u(s(u), u)\Delta u. \tag{7}$$

Proof The proof follows from the implicit function theorem, an application of the chain rule and

$$\begin{aligned} \Phi'(u)\Delta u &= \phi_y(s(u), u)s'(u)\Delta u + \phi_u(s(u), u)\Delta u \\ &= \phi_y(s(u), u)\xi + \phi_u(s(u), u)\Delta u. \end{aligned} \tag{8}$$

If U is not a Banach but a Hilbert space, one would expect for the Fréchet-derivative $\Phi'(u)\Delta u$ a gradient representation. In order to achieve this one would need an explicit representation of the derivative in a $\Phi(u)'\Delta u = \langle \nabla\Phi(u), \Delta u\rangle$ in the proper duality pairing. Such a representation clearly cannot be derived from Eq. (6) since ξ as shown in (7) depends in an implicit way on Δu. The only possibility to obtain this consists in computing the whole sensitivity operator $s'(u) \in L(U, Y)$ which in finite dimensions results in the computation of the sensitivity matrix. This requires repeated solves of the sensitivity Eq. (7), which for high dimensions is not a feasible approach. For completeness we formulate this in the following theorem.

Theorem 3 *Let Assumptions 1 and 2 hold at $(y, u) \in Y \times U$. The linear map $s'(u) : U \to Y$ is well defined by the equation*

$$g_y(s(u), u)s'(u) = -g_u(s(u), u).$$

Then we obtain

$$\Phi'(u) = s'(u)^*\phi_y(s(u), u) + \phi_u(s(u), u) \in U^*. \tag{9}$$

We can avoid this difficulty by using the adjoint operator $s'(u)^*$ of $s'(u) \in L(U, Y)$ and a so-called adjoint equation. The adjoint approach is outlined in the following theorem.

Theorem 4 *Let Assumptions 1 and 2 hold at $(y, u) \in Y \times U$. Then we obtain*

$$\Phi'(u) = g_u(s(u), u)^* p + \phi_u(s(u), u) \in U^*, \tag{10}$$

where $p \in Z^*$ is defined as the solution of the adjoint equation

$$g_y(s(u), u)^* p = -\phi_y(s(u), u) \in Y^*. \tag{11}$$

Alternatively, the action of the adjoint variable $p \in Z^*$ is given by

$$p(z) = -\phi_y(s(u), u) g_y(s(u), u)^{-1} z \quad \forall z \in Z. \tag{12}$$

Proof We know from (4)

$$s'(u) = -g_y(s(u), u)^{-1} g_u(s(u), u) \in L(U, Y)$$

and hence for $s'(u)^* : Y^* \to U^*$

$$s'(u)^* = -g_u(s(u), u)^* (g_y(s(u), u)^*)^{-1}.$$

Then the first term in (9) of the derivative of Φ can be rewritten as

$$s'(u)^* \phi_y(s(u), u) = -g_u(s(u), u)^* (g_y(s(u), u)^*)^{-1} \phi_y(s(u), u) = g_u(s(u), u)^* p,$$

where $p \in Z^*$ solves the adjoint equation (11).

Furthermore, we have for an arbitrary $z \in Z$ with $\Delta v := g_y(s(u), u)^{-1} z$ using (11)

$$\begin{aligned} -\phi_y(s(u), u) g_y(s(u), u)^{-1} z &= -\phi_y(s(u), u) \Delta v = [g_y(s(u), u)^* p] \Delta v \\ &= p(g_y(s(u), u) \Delta v) = p(z), \end{aligned}$$

which shows (12). Here we used the definition of the adjoint, namely

$$< g_y(s(u), u)^* p, \Delta v >_{Y^*, Y} = < p, g_y(s(u), u) \Delta v >_{Z^*, Z} = p(g_y(s(u), u) \Delta v).$$

To synchronize the representation with the sensitivity approach we give also a version where the Fréchet-derivative is applied to a vector Δv also for the adjoint version.

Corollary 2 *Let Assumptions 1 and 2 hold at $(y, u) \in Y \times U$. Then*

$$\Phi'(u) \Delta u = p(g_u(s(u), u) \Delta u) + \phi_u(s(u), u) \Delta u$$

with the adjoint functional p from Theorem 4 defined in (12).

The main advantage of this approach is that we do not need to solve for the sensitivity operator or matrix but rather need to solve only one Eq. (11) in order to compute the adjoint variable p.

3 Representation of Second Order Fréchet-Derivative

In this section we devote our efforts to a rigorous derivation of the second derivative of the function Φ. Since in many algorithmic applications, e.g., the use of iterative solvers for the Newton step, the complete Hessian information $\Phi''(u)$ is not needed, we concentrate on the computation of the Hessian-vector product $\Phi''(u)\Delta u$. Obviously we have to strengthen Assumption 1 as follows.

Assumption 3 *Let the functional ϕ and the map g be twice continuously Fréchet-differentiable on $Y \times U$.*

As noted before, the notation using second derivatives can be somewhat complex, since the e.g., the second partial derivative $g_{yu}(y, u)$ can be interpreted as a map in $L(U, L(Y, Z))$ or $L(U \times Y, Z)$, etc.

In order to compute the second derivative we proceed in the following way: For fixed $\Delta v \in U$ consider $\Phi'(u)\Delta v$ as a function of u. Then we differentiate this new functional with respect to u which gives us the second derivative. Since $\Phi(u) = \phi(s(u), u)$ contains variables $y = s(u)$ that are implicitly defined, this will also be the case for $\Phi'(u)\Delta v$. However, there are even more implicitly defined variables like ξ in the sensitivity approach or p in the adjoint approach. Both ξ and p also depend on u in an implicit way. All this has to be kept in mind for a careful computation of the second derivatives.

In some applications, the derivative $\Phi'(u)$ is no longer differentiable in a classical sense, but only in a generalized sense where generalized derivatives come into play. Those problems can sometimes be solved efficiently using semi-smooth Newton methods. In that case at each iteration, one has to select an element from the possibly set-valued generalized second derivative of Φ. The theory of this section can be extended in such a case, if this representor exhibits a certain structure close to the problem in this paper. A similar question arises when one wants to use a Gauss-Newton method, where several terms in the second derivative operator are omitted. These issues of generalization will be addressed in a forthcoming paper.

Let us outline the approach we take here for the next sections. In the previous section we found two routes to obtain a representation of the first derivative, i.e., via the sensitivity equation or the adjoint equation. If we calculate the second derivative as outlined above, we need to decide to use either the adjoint or sensitivity approach in the computation of the derivative of $\Phi'(\cdot)\Delta v$, i.e., the second derivative of $\Phi(u)$. Therefore, we have four different routes available which we could follow, as outlined at the end of Sect. 1.

In the following, we proceed along these routes in rigorous mathematical terms. Recall that in (6) the first derivative is given by $\Phi'(u)\Delta v = \phi_y(y,u)\xi + \phi_u(y,u)\Delta v$. These terms contain two variables y and ξ which depend on u and which are defined by the original equality constraint (1) and the sensitivity Eq. 7. Hence we expand the y-variables to $\tilde{y} := (y, \xi) \in Y \times Y$ and define in analogy to $\phi(y,u)$ the function

$$\tilde{\phi}(\tilde{y}, u) = \Phi'(u)\Delta v = \phi_y(y,u)\xi + \phi_u(y,u)\Delta v. \tag{13}$$

The vector $\tilde{y} := (y, \xi) \in Y \times Y$ is the solution of the state and sensitivity equation which we combine into

$$\tilde{g}(\tilde{y}, u) := \begin{pmatrix} g(y,u) \\ g_y(y,u)\xi + g_u(y,u)\Delta v \end{pmatrix} = 0. \tag{14}$$

If we want to proceed similar to the case of the first-order derivatives, i.e., eliminate the equality constraint due to an application of the implicit function theorem, then we need to check the invertibility of $\tilde{g}_{\tilde{y}}$.

Lemma 1 *We define $\tilde{\phi} : Y \times Y \times U \to I\!R$ by (13) and $\tilde{g} : Y \times Y \times U \to Z \times Z$ by (14). Then its Fréchet-derivatives are given by*

$$\tilde{g}_{\tilde{y}}(\tilde{y}, u) = \begin{pmatrix} g_y(y,u) & 0 \\ g_{yy}(y,u)\xi + g_{uy}(y,u)\Delta v & g_y(y,u) \end{pmatrix} \in L(Y \times Y, Z \times Z), \tag{15}$$

which is invertible if Assumption 2 holds. Furthermore,

$$\tilde{g}_u(\tilde{y}, u) = \begin{pmatrix} g_u(y,u) \\ g_{yu}(y,u)\xi + g_{uu}(y,u)\Delta v \end{pmatrix} \in L(U, Z \times Z) \tag{16}$$

and for the objective function $\tilde{\phi}$ we obtain

$$\tilde{\phi}_{\tilde{y}}(\tilde{y}, u) = \begin{pmatrix} \phi_{yy}(y,u)\xi + \phi_{uy}(y,u)\Delta v \\ \phi_y(y,u) \end{pmatrix} \in Y^* \times Y^* \tag{17}$$

and

$$\tilde{\phi}_u(\tilde{y}, u) = \phi_{yu}(y,u)\xi + \phi_{uu}(y,u)\Delta v \in U^*. \tag{18}$$

The statements of the lemma can be easily obtained from the definitions (13) and (14).

Since $\tilde{g}_{\tilde{y}}(\tilde{y}, u)$ from (15) is invertible under Assumption 2 we can apply the implicit function theorem, Theorem 1, to derive in the same way as in Theorem 1 the following:

Theorem 5 *Let the Assumptions 2 and 3 hold at $(\tilde{y}_*, u_*) \in Y \times Y \times U$ with $\tilde{g}(\tilde{y}_*, u_*) = 0$. Then there exist neighborhoods $\tilde{B}_U \subset U$ of u_* and $\tilde{B}_{Y\times Y} \subset Y \times Y$ of $\tilde{y}^*$ and a map*

$$\tilde{s}(\cdot) : \tilde{B}_U \to \tilde{B}_{Y\times Y} \quad \textit{such that} \quad \tilde{g}(\tilde{s}(u), u) = 0 \quad \textit{on} \quad \tilde{B}_U$$

with its derivative defined by

$$\tilde{g}_{\tilde{y}}(\tilde{s}(u), u)\tilde{s}'(u) = -\tilde{g}_u(\tilde{s}(u), u).$$

In the case of the adjoint approach we proceed in a similar fashion by augmenting the variable y to $\hat{y} = (y, p) \in Y \times Z^*$. Hence Eq. (10) becomes

$$\hat{\phi}(\hat{y}, u) := \Phi'(u)\Delta v = [g_u(y, u)^* p + \phi_u(y, u)]\Delta v. \tag{19}$$

The equality constraints then are defined as follows using (11)

$$\hat{g}(\hat{y}, u) := \hat{g}(y, p, u) = \begin{pmatrix} g(y, u) \\ g_y(y, u)^* p + \phi_y(y, u) \end{pmatrix} = 0. \tag{20}$$

The derivatives for these mappings are computed as follows:

Lemma 2 *We define $\hat{\phi} : Y \times Z^* \times U \to I\!R$ by (19) and $\hat{g} : Y \times Z^* \times U \to Z \times Y^*$ by (20). Then the Fréchet-derivatives for $\hat{\phi}$ are given by*

$$\hat{\phi}_u(\hat{y}, u) = (g_{uu}(y, u)\Delta v)^* p + \phi_{uu}(y, u)\Delta v \in U^* \tag{21}$$

$$\hat{\phi}_{\hat{y}}(\hat{y}, u) = \begin{pmatrix} (g_{uy}(y, u)\Delta v)^* p + \phi_{uy}(y, u)\Delta v \\ g_u(y, u)\Delta v \end{pmatrix} \in Y^* \times Z \tag{22}$$

and for $\hat{g}$ we have

$$\hat{g}_u(\hat{y}, u) = \begin{pmatrix} g_u(y, u) \\ (g_{yu}(y, u)(\cdot))^* p + \phi_{yu}(y, u) \end{pmatrix} \in L(U, Z \times Y^*) \tag{23}$$

$$\hat{g}_{\hat{y}}(\hat{y}, u) = \begin{pmatrix} g_y(y, u) & 0 \\ (g_{yy}(y, u)(\cdot))^* p + \phi_{yy}(y, u) & g_y(y, u)^* \end{pmatrix} \in L(Y \times Z^*, Z \times Y^*). \tag{24}$$

Note, that also in this case Assumption 2 implies that $\hat{g}_{\hat{y}}$ is invertible. Therefore we can apply the implicit function theorem also to this setting.

Theorem 6 *For fixed Δv let the Assumptions 2 and 3 hold at $(\hat{y}_*, u_*) \in Y \times Z^* \times U$ with $\hat{g}(\hat{y}_*, u_*) = 0$. Then there exist neighborhoods $\hat{B}_U \in U$ of u_* and $\hat{B}_{Y\times Z^*} \subset Y \times Z^*$ of $\hat{y}$ and a map*

$$\hat{s} : \hat{B}_U \to \hat{B}_{Y\times Z^*} \ \textit{ with } \ \hat{g}(\hat{s}(u), u) = 0 \textit{ on } \hat{B}_U$$

with its derivative defined by

$$\hat{g}_{\hat{y}}(\hat{s}(u), u)\hat{s}'(u) = -\hat{g}_u(\hat{s}(u), u).$$

With these technicalities resolved we turn to the computation of the second derivative using the four strategies outlined before.

4 Second Order Sensitivities and Second Order Adjoints

4.1 Sensitivity-Sensitivity Approach

If we apply the sensitivity equation approach to both the first and the second derivative, then we obtain the following theorem.

Theorem 7 *Let Assumptions 2 and 3 hold at* (y, u) *with* $g(y, u) = 0$*. Then for* $\Delta v, \Delta w \in U$

$$\begin{aligned}\Phi''(u)(\Delta v, \Delta w) = \phi_{uu}(y, u)(\Delta v, \Delta w) + \phi_{uy}(y, u)(\Delta v, \eta)\\ +\phi_{yu}(y, u)(\xi, \Delta w) + \phi_{yy}(y, u)(\xi, \eta) + \phi_y(y, u)\rho,\end{aligned}$$

where $\xi, \eta \in Y$ *solve the following first-order sensitivity equations*

$$\begin{aligned}g_y(y, u)\xi &= -g_u(y, u)\Delta v\\ g_y(y, u)\eta &= -g_u(y, u)\Delta w,\end{aligned}$$

and $\rho \in Y$ *solves the second order sensitivity equation*

$$\begin{aligned}g_y(y, u)\rho = -g_{yy}(y, u)(\xi, \eta) - g_{uy}(y, u)(\Delta v, \eta)\\ -g_{yu}(y, u)(\xi, \Delta w) - g_{uu}(y, u)(\Delta v, \Delta w).\end{aligned}$$

Proof Theorem 2 applied to the problem formulation in (13) and (14) yields

$$\tilde{\Phi}'(u)\Delta w = \tilde{\phi}_{\tilde{y}}(\tilde{s}(u), u)\tilde{\xi} + \tilde{\phi}_u(\tilde{s}(u), u)\Delta w,$$

where $\tilde{\xi} = (\eta, \rho)^T$ is the unique solution of the sensitivity equation

$$\tilde{g}_{\tilde{y}}(\tilde{s}(u), u)\tilde{\xi} = -\tilde{g}_u(\tilde{s}(u), u)\Delta w.$$

Using (17)–(18) implies

$$\begin{aligned}\tilde{\Phi}'(u)\Delta w &= \tilde{\phi}_{\tilde{y}}(\tilde{s}(u), u)\tilde{\xi} + \tilde{\phi}_u(\tilde{s}(u), u)\Delta w\\ &= \phi_{yy}(y, u)(\xi, \eta) + \phi_{uy}(y, u)(\Delta v, \eta)\\ &\quad +\phi_y(y, u)\rho + \phi_{yu}(y, u)(\xi, \Delta w) + \phi_{uu}(y, u)(\Delta v, \Delta w),\end{aligned}$$

where $\tilde{\xi} = (\eta, \rho)^T$ solve

$$\tilde{g}_{\tilde{y}}(\tilde{s}(u), u)\tilde{\xi} = -\tilde{g}_u(\tilde{s}(u), u)\Delta w.$$

With (15)–(16) this implies

$$\begin{pmatrix} g_y(y, u)\eta \\ g_{yy}(y, u)(\xi, \eta) + g_{uy}(y, u)(\Delta v, \eta) + g_y(y, u)\rho \end{pmatrix}$$
$$= -\begin{pmatrix} g_u(y, u)\Delta w \\ g_{yu}(y, u)(\xi, \Delta w) + g_{uu}(y, u)(\Delta v, \Delta w) \end{pmatrix}.$$

Rearranging the terms leads to the formulation in the theorem.

This theorem shows that in order to perform a full Hessian evaluation one needs three solves (25) of the linearized equality constraint equation, i.e. sensitivity equation. This is an interesting observation by itself, especially since only the right hand side of the sensitivity equation is changed. However, note that this is true only for an evaluation of the second order term in a Taylor expansion of the objective function, i.e. the Hessian applied to two arguments, here Δv and Δw. The computation of a Hessian vector product resulting into a vector, as it is required for a Newton step, is still problematic. Here we would have to scan the whole space U, i.e. let Δw run through all basis vectors, which is quite expensive.

In a similar way, one can derive a representation of the Hessian operator itself, without any application to arguments Δv, Δw. This can be obtained easily with mappings $s'(u) \in L(U, Y)$ and $\sigma(u) \in L(U, Y)$ and

$$\xi = s'(u)\Delta v, \quad \eta = s'(u)\Delta w, \quad \rho = \sigma(u)\Delta v.$$

Theorem 8 *Let Assumptions 2 and 3 hold at (y, u) with $g(y, u) = 0$. Then $\Phi''(u) : U \to U^*$ can be represented as*

$$\begin{aligned}\Phi''(u) = {} & \phi_{uu}(s(u), u) + \phi_{uy}(s(u), u)s'(u) \\ & + s'(u)^*\phi_{yu}(s(u), u) + s'(u)^*\phi_{yy}(s(u), u)s'(u) + \sigma(u)^*\phi_y(s(u), u),\end{aligned}$$

where $s(u) \in Y$ solves the system equation $g(s(u), u) = 0$. The operators $s'(u) \in L(U, Y)$ and $\sigma(u) \in L(U, Y)$ are the solutions to the following first and second order sensitivity equations

$$\begin{aligned}g_y(s(u), u)s'(u) &= -g_u(s(u), u) \\ g_y(s(u), u)\sigma(u) &= -s'(u)^*g_{yy}(y, u)s'(u) - g_{yu}(y, u)s'(u) \\ &\quad -s'(u)^*g_{uy}(y, u) - g_{uu}(y, u).\end{aligned}$$

Also for this theorem, one would need knowledge of the full operator $s'(u)$ which is computationally not available. We can compute an evaluation of $s'(u)\Delta v$ by the solve of one sensitivity equation, but to get information of the full operator $s'(u)$ one would need to solve it for all basis vectors Δv.

4.2 Sensitivity-Adjoint Approach

We start with the first-order derivative in sensitivity form as written up in Lemma 1 and used in the previous subsection. However, here we apply the adjoint approach for the calculation of the second derivative as outlined in Theorem 4.

In this theorem one sees that two adjoint equation equations need to be solved, both with the system matrix or operator $g_y(y,u)^*$. The first solution p comes from the adjoint equation (25) which we know from the computation of the gradient or first derivative, therefore we call it the first-order adjoint. The other solution π needs to be computed as a solution of (26) in order to obtain the information about the second derivative, therefore we call it the second order adjoint.

Theorem 9 *Let Assumptions 2 and 3 hold at (y,u) with $g(y,u)=0$. Then $\Phi''(u)\Delta v \in U^*$ can be represented as*

$$\Phi''(u)\Delta v = g_u(y,u)^*\pi + (g_{uy}(y,u)\xi)^* p + (g_{uu}(y,u)\Delta v)^* p \\ + \phi_{uy}(y,u)\xi + \phi_{uu}(y,u)\Delta v.$$

Here $\xi \in Y$ solves the first-order sensitivity equation

$$g_y(y,u)\xi = -g_u(y,u)\Delta v \in Z,$$

$p \in Z^*$ is a solution of the first-order adjoint equation

$$g_y(y,u)^* p = -\phi_y(y,u) \in Y^*, \tag{25}$$

and $\pi \in Z^*$ solves the second order adjoint equation

$$g_y(y,u)^*\pi = -(g_{yy}(y,u)\xi)^* p - (g_{yu}(y,u)\Delta v)^* p - \phi_{yy}(y,u)\xi - \phi_{yu}(y,u)\Delta v \in Y^*. \tag{26}$$

Proof Due to the definition of $\tilde{g}$ in (14) the corresponding multiplier is of the form $\tilde{p} := (\pi, p) \in Z^* \times Z^*$. Then inserting (16) and (18) into (10) gives

$$\begin{aligned}\tilde{\Phi}'(u) &= \tilde{g}_u(\tilde{s}(u),u)^*\tilde{p} + \tilde{\phi}_u(\tilde{s}(u),u) \\ &= \begin{pmatrix} g_u(y,u) \\ g_{yu}(y,u)\xi + g_{uu}(y,u)\Delta v \end{pmatrix}^* \begin{pmatrix} \pi \\ p \end{pmatrix} + \phi_{yu}(y,u)\xi + \phi_{uu}(y,u)\Delta v \\ &= g_u(y,u)^*\pi + (g_{yu}(y,u)\xi)^* p + (g_{uu}(y,u)\Delta v)^* p + \phi_{yu}(y,u)\xi + \phi_{uu}(y,u)\Delta v.\end{aligned}$$

The adjoint equation for the second derivative is obtained by inserting (15) and (17) into Eq. (11) which reads as

$$\tilde{g}_{\tilde{y}}(\tilde{s}(u),u)^*\tilde{p} = -\tilde{\phi}_{\tilde{y}}(\tilde{s}(u),u)$$

or

$$\begin{pmatrix} g_y(y,u) & 0 \\ g_{yy}(y,u)\xi + g_{yu}(y,u)\Delta v & g_y(y,u) \end{pmatrix}^* \begin{pmatrix} \pi \\ p \end{pmatrix} = -\begin{pmatrix} \phi_{yy}(y,u)\xi + \phi_{yu}(y,u)\Delta v \\ \phi_y(y,u) \end{pmatrix}$$

and finally

$$g_y(y,u)^*\pi + (g_{yy}(y,u)\xi)^* p + (g_{yu}(y,u)\Delta v)^* p = -\phi_{yy}((y,u)\xi - \phi_{uy}(y,u)\Delta v$$

and

$$g_y(y,u)^* p = -\phi_y(y,u),$$

which yields the statements of the theorem.

This result shows that it is possible to compute the Hessian-vector product by simply solving the (first-order) sensitivity equation for ξ and the second order adjoint equation (26) for π. The operator or matrix for the second order adjoint solve is the same as for the first-order adjoint. In contrast to the sensitivity approach in the previous section we obtain the full information of the vector that represents the Hessian-vector product. This is an highly efficient way to compute the information needed in each step of a Krylov method to solve for the Newton step.

4.3 Adjoint-Sensitivity Approach

In this subsection we start with the first derivative represented by the adjoint equation. For the computation of the second derivative we apply the sensitivity approach. In what follows, we use the notation $\hat{y} = (p, u)$ as outlined in (19) and (20). By Theorem 2 we have from Eq. (20) that

$$\hat{g}_{\hat{y}}(\hat{y}, u)\hat{\xi} = -\hat{g}_u(\hat{y}, u)\Delta w$$

or with $\hat{\xi} = (\xi, \pi)$ using (23) and (24)

$$\begin{pmatrix} g_y(y,u) & 0 \\ (g_{yy}(y,u)(\cdot))^* p + \phi_{yy}(y,u) & g_y(y,u)^* \end{pmatrix} \begin{pmatrix} \xi \\ \pi \end{pmatrix}$$

$$= -\begin{pmatrix} g_u(y,u) \\ (g_{yu}(y,u)(\cdot))^* p + \phi_{yu}(y,u) \end{pmatrix} \Delta w,$$

which leads to the first-order sensitivity equation for ξ

$$g_y(y,u)\xi = -g_u(y,u)\Delta w$$

and the second order adjoint equation for π

$$(g_{yy}(y,u)\xi)^* p + \phi_{yy}(y,u)\xi + g_y(y,u)^*\pi = -(g_{yu}(y,u)\Delta w)^* p - \phi_{yu}(y,u)\Delta w.$$

Furthermore, the Hessian-vector product can be obtained from (6)

$$\begin{aligned}\hat{\phi}_{\hat{y}}(\hat{y},u)\hat{\xi} + \hat{\phi}_u(\hat{y},u)\Delta w &= \begin{pmatrix} (g_{uy}(y,u)\Delta v)^* p + \phi_{uy}(y,u)\Delta v \\ g_u(y,u)\Delta v \end{pmatrix} \begin{pmatrix} \xi \\ \pi \end{pmatrix} \\ &\quad + [(g_{uu}(y,u)\Delta v)^* p + \phi_{uu}(y,u)\Delta v]\Delta w \\ &= [(g_{yu}(y,u)\xi)^* p + \phi_{yu}(y,u))\xi + (g_{uu}(y,u)\Delta w)^* p \\ &\quad + \phi_{uu}(y,u))\Delta w + g_u(y,u)^*\pi]\Delta v,\end{aligned}$$

which is the same expression as in Theorem 9.

In summary, we obtain the following remark.

Remark 1 For fixed Δv let the Assumptions 2 and 3 hold at $(\hat{y}_*, u_*)$. Then

$$\begin{aligned}\Phi''(u)(\Delta v, \Delta w) = [&(g_{yu}(y,u)\xi)^* p + \phi_{yu}(y,u))\xi + (g_{uu}(y,u)\Delta w)^* p \\ &+ \phi_{uu}(y,u))\Delta w + g_u(y,u)^*\pi]\Delta v,\end{aligned} \tag{27}$$

where (y,u) solve $g(y,u) = 0$, p solves the adjoint equation and ξ solves the first-order sensitivity equation

$$g_y(y,u)\xi = -g_u(y,u)\Delta w,$$

and π solves the second order adjoint equation

$$g_y(y,u)\pi = -(g_{yy}(y,u)\xi)^* p - \phi_{yy}(y,u)\xi - (g_{yu}(y,u)\Delta w)^* p - \phi_{yu}(y,u)\Delta w.$$

Since in (27) the vector Δv is separated outside of the parentheses, we have an expression for the Hessian-vector product. This result is identical with the findings in Theorem 9.

4.4 Adjoint-Adjoint Approach

Finally we apply the adjoint approach to compute the second derivative when the first derivative is also calculated by the adjoint approach.

The adjoint equation for the extended system reads as in (11)

$$\hat{g}_{\hat{y}}(y,u)^*\hat{p} = -\hat{\phi}_{\hat{y}}(\hat{y},u) \quad \hat{p} \in Z^* \times Y$$

and $\hat{g}_{\hat{y}}(y,u)^* \in L(Z^* \times Y, Y^* \times Z)$. Inserting the terms from (22) and (24) we obtain for $\hat{p} = (\pi, \xi) \in Z^* \times Y$

$$\begin{pmatrix} g_y(y,u) & 0 \\ (g_{yy}(y,u)(\cdot))^* p + \phi_{yy}(y,u) & g_y(y,u)^* \end{pmatrix}^* \begin{pmatrix} \pi \\ \xi \end{pmatrix}$$
$$= - \begin{pmatrix} (g_{uy}(y,u)\Delta v)^* p + \phi_{uy}(y,u)\Delta v \\ g_u(y,u)\Delta v \end{pmatrix} \in Y^* \times Z,$$

which results in

$$g_y(y,u)^*\pi + (g_{yy}(y,u)\xi)^* p + \phi_{yy}(y,u)\xi = -(g_{uy}(y,u)\Delta v)^* p - \phi_{uy}(y,u)\Delta v$$

and

$$g_y(y,u)^*\xi = -g_u(y,u)\Delta v.$$

The Hessian-vector product can be computed from Eq. (10) as

$$\hat{g}_u(\hat{y},u)^*\hat{p} + \hat{\phi}_u(\hat{y},u)$$

and inserting (21) and (23)

$$\begin{pmatrix} g_u(y,u) \\ (g_{yu}(y,u)(\cdot))^* p + \phi_{yu}(y,u) \end{pmatrix}^* \begin{pmatrix} \pi \\ \xi \end{pmatrix} + (g_{uu}(y,u)\Delta v)^* p + \phi_{uu}(y,u)\Delta v$$

and

$$g_u(y,u)^*\pi + (g_{yu}(y,u)\xi)^* p + \phi_{yu}(y,u)\xi + (g_{uu}(y,u)\Delta v)^* p + \phi_{uu}(y,u)\Delta v.$$

Therefore, we have the following remark.

Remark 2 The second derivative reads

$$\begin{aligned} \Phi''(u)\Delta v = {} & g_u(y,u)^*\pi + (g_{yu}(y,u)\xi)^* p + \phi_{yu}(y,u)\xi + (g_{uu}(y,u)\Delta v)^* p \\ & + \phi_{uu}(y,u)\Delta v, \end{aligned}$$

where y, u satisfy $g(y,u) = 0$ and p solves the adjoint equation. Furthermore ξ solves the sensitivity equation of first-order

$$g_y(y,u)\xi = -g_u(y,u)\Delta v,$$

and π the second order adjoint equation

$$g_y(y,u)^*\pi = -(g_{yy}(y,u)\xi)^* p - \phi_{yy}(y,u)\xi - (g_{uy}(y,u)\Delta v)^* p - \phi_{uy}(y,u)\Delta v.$$

From the results of the theorems above we realize that the four different approaches outlined in the beginning of Sect. 3 lead to two different results: one where a second order adjoint equation comes into play and another one where a second sensitivity equation has to be solved. We can also see that this can be derived in a fairly general setting for infinite-dimensional spaces and hence can be applied to all kinds of optimization problems. In this paper we will concentrate on two applications in the context of Newton's method.

5 PDE-Constrained Optimization

We consider as an application an optimization problem with a partial differential equation and control in the coefficient. In particular, we consider an inverse problem where the partial differential equation is given by

$$-\nabla \cdot ((\exp(u) + \epsilon)\nabla y) = f \quad \text{in } \Omega, \qquad y = 0 \quad \text{on } \partial\Omega \tag{28}$$

on a bounded and closed domain $\Omega \subset I\!R^2$ for a given right-hand side $f \in L^2(\Omega)$ and some small $\epsilon > 0$. The inverse problem consists of finding a proper $u \in L^\infty(\Omega)$ such that the corresponding solution $y \in H_0^1(\Omega)$ is close to a given observed output $w^{obs} \in H_0^1(\Omega_1)$, with $\Omega_1 \subseteq \Omega$.

To formulate this PDE-constrained optimization problem in the form problem 1 was posed, we set

$$Y = H_0^1(\Omega),\ U = L^\infty(\Omega),\ W = L^2(\Omega_1),\ Z = L^2(\Omega).$$

Next we define the constraint as

$$g(y, u) = -\nabla \cdot ((\exp(u) + \epsilon)\nabla y) - f, \quad g : Y \times U \to Z,$$

and the objective function including a regularization term with $\alpha > 0$,

$$\phi(y, u) = \frac{1}{2}\int_{\Omega_1} (\mathcal{B}y(x) - w^{obs}(x))^2 dx + \frac{\alpha}{2}\int_{\Omega} u(x)^2 dx,$$

where $\mathcal{B} : L^2(\Omega) \to L^2(\Omega_1)$ is a linear observation operator that extracts measurements from y. One can show that the Fréchet-derivatives up to second order exist and have the following form for the objective function

$$\phi_y(y, u)\bar{y} = \int_{\Omega} [\mathcal{B}^*(\mathcal{B}y(x) - w^{obs}(x))]\bar{y}(x)dx,$$

$$\phi_u(y, u)\bar{u} = \alpha \int_{\Omega} u(x)\bar{u}(x)dx,$$

$$\phi_{yy}(y, u)(\bar{y}, \bar{z}) = \int_{\Omega} \mathcal{B}^*\mathcal{B}\bar{y}(x)\bar{z}(x)dx,$$

$$\phi_{yu}(y,u)(\bar{y},\bar{u}) = \phi_{uy}(y,u)(\bar{u},\bar{y}) = 0,$$
$$\phi_{uu}(y,u)(\bar{u},\bar{v}) = \alpha \int_{\Omega} \bar{u}(x)\bar{v}(x)dx.$$

Similarly we obtain for the Fréchet-derivatives of the constraint

$$\begin{aligned} g_y(y,u)\bar{y} &= -\nabla\cdot((\exp(u)+\epsilon)\nabla\bar{y}),\\ g_u(y,u)\bar{u} &= -\nabla\cdot(\exp(u)\bar{u}\nabla y),\\ g_{yy}(y,u)(\bar{y},\bar{z}) &= 0,\\ g_{yu}(y,u)(\bar{y},\bar{u}) &= -\nabla\cdot(\exp(u)\bar{u}\nabla\bar{y}),\\ g_{uu}(y,u)(\bar{u},\bar{v}) &= -\nabla\cdot(\exp(u)\bar{u}\bar{v}\nabla y). \end{aligned}$$

Since the two adjoints are computed by an application of the operator g_y^*, it is straightforward to derive a representation for the solution $p \in Z^*$ of the equation $g_y^* p = r$ with given right hand side $r \in Z$ along the following lines.

Lemma 3 *The solution $p \in Z^*$ of $g_y^* p = r$ for given $r \in Y^* = H^{-1}(\Omega)$ is represented by $p(\zeta) = \langle \bar{p}, \zeta\rangle_Z$, $\zeta \in Z = L^2(\Omega)$, where $\bar{p} \in Y = H_0^1(\Omega)$ is a weak solution of*

$$-\nabla\cdot(\exp(u)+\epsilon)\nabla\bar{p} = r.$$

Proof Equivalently, $g_y^* p = r$ can be written as

$$(g_y^* p)(\eta) = p(g_y\eta) = \langle r, \eta\rangle_Z \quad \forall\, \eta \in Y.$$

Let us make an ansatz for the solution $p \in Z^* = L^2(\Omega)^*$, i.e., assume the linear functional p is represented by a function $\bar{p} \in H_0^1(\Omega)$ such that $p(\zeta) = \langle \bar{p}, \zeta\rangle_Z$ for all $\zeta \in Z$. Then

$$\begin{aligned} (g_y^* p)(\eta) = p(g_y\eta) &= -\langle \bar{p}, \nabla\cdot(\exp(u)+\epsilon)\nabla\eta)\rangle_Z = \langle \nabla\bar{p}, (\exp(u)+\epsilon)\nabla\eta)\rangle_Z\\ &= \langle (\exp(u)+\epsilon)\nabla\bar{p}, \nabla\eta)\rangle_Z \quad \forall\eta \in Y, \end{aligned}$$

and

$$(g_y^* p)(\eta) = \langle r, \eta\rangle_Z \iff \langle (\exp(u)+\epsilon)\nabla\bar{p}, \nabla\eta)\rangle_Z = \langle r, \eta\rangle_Z \quad \forall\eta \in Y.$$

This is the definition in weak form of a solution of the PDE

$$-\nabla\cdot(\exp(u)+\epsilon)\nabla\bar{p} = r.$$

Therefore, the first-order adjoint $p \in H_0^1(\Omega)$ according to (11) is given as the solution of

$$-\nabla\cdot(\exp(u)+\epsilon)\nabla p) = -\mathcal{B}^*(\mathcal{B}y - w^{obs}). \tag{29}$$

The second order adjoint $\pi \in H_0^1(\Omega)$ according to (26) is the solution of

$$-\nabla \cdot ((\exp(u) + \epsilon)\nabla \pi) = \nabla \cdot (\exp(u)\bar{u}\nabla p) - \mathcal{B}^*\mathcal{B}\xi. \tag{30}$$

The solution $\xi \in H_0^1(\Omega)$ of the first-order sensitivity equation can be obtained following (7) by solving

$$-\nabla \cdot ((\exp(u) + \epsilon)\nabla \xi) = \nabla \cdot (\exp(u)\bar{u}\nabla y). \tag{31}$$

Given all the derivatives above, we can apply Theorem 9 to see which partial differential equations need to be solved for an evaluation of a Hessian of Φ applied to a vector.

Theorem 10 *For $\Phi(u) = \phi(s(u), u)$ the application of the Hessian of ϕ to a vector $\bar{u}$, $\Phi''(u)\bar{u}$, can be obtained with $y = s(u)$ from*

$$\Phi''(u)\bar{u} = -\exp(u)[(\nabla \pi)^T \nabla y + (\nabla p)^T \nabla \xi] + [-\exp(u)(\nabla p)^T \nabla y + \alpha]\bar{u},$$

where $y \in H_0^1(\Omega)$ is the solution of the state Eq. 28, $p \in H_0^1(\Omega)$ is the solution of the first-order adjoint Eq. 29, $\xi \in H_0^1(\Omega)$ the solution of the first-order sensitivity Eq. 31, and $\pi \in H_0^1(\Omega)$ is the solution of the second order adjoint Eq. 30.

6 Summary and Conclusions

In this paper we derived rigorously second order adjoints for equality constrained optimization problems in a general, infinite-dimensional setting which are used for the Hessian-vector products in Newton's method. We showed that while there are four routes to arrive to the second order adjoints (e.g., via combinations of sensitivity or adjoint equations), except the sensitivity-sensitivity approach, all the other three of these coincide, i.e., they give the same Hessian-apply expression. This finding suggests that one can choose whichever route is more convenient without compromising the underlying computational effort. However, as discussed the sensitivity-sensitivity approach is feasible only in the case of small number parameters.

We have applied this general framework to a PDE-constrained optimization problem formulated as a nonlinear least squares problem governed by an elliptic PDE. This application revealed the ability to derive the second order adjoints (and Hessian-vector apply) in a straightforward manner when following the general framework established in this paper.

In this paper we chose to derive the expressions for the second order adjoints at the infinite-dimensional level for a number of reasons. First, the derivation and final results do not depend on any particular discretization of the underlying PDEs. Second, the derivation is clean and reveal similar structure. Third, the form of the boundary conditions for the adjoints falls out cleanly from the infinite-dimensional expressions. However, in certain cases working in finite-dimensions is beneficial, for instance in the case when the optimize-then-discretize (OTD) and discretize-then-optimize (DTO) approaches do not commute [18]. In the case of Hilbert spaces, the role of the adjoint solves could be simplified when the domain and range space are identical and the operator g_y turns out to be self-adjoint. Furthermore, the standard formulation of a Gauss-Newton method for a nonlinear least squares problem can be used in a Hilbert space setting and it can be shown that the adjoint solve is in fact a second order adjoint as defined in this context. The derivation of these equations in finite-dimension and the framework for Gauss-Newton methods is the subject of future work.

Acknowledgements This work was partially supported by the US National Science Foundation grant CAREER-1654311. E. S. acknowledges partial support from the University of California, Merced, and Lawrence Livermore National Laboratory.

References

1. Akçelik, V., Biros, G., Drăgănescu, A., Ghattas, O., Hill, J., van Bloeman Waanders, B.: Dynamic data-driven inversion for terascale simulations: Real-time identification of airborne contaminants. In: Proceedings of SC2005. Seattle (2005)
2. Alekseev, A.K., Navon, I.M.: The analysis of an ill-posed problem using multi-scale resolution and second-order adjoint techniques. Computer Methods in Applied Mechanics and Engineering **190**, 1937–1953 (2001)
3. Alekseev, A.K., Navon, I.M., Steward, J.: Comparison of advanced large-scale minimization algorithms for the solution of inverse ill-posed problems. Optimization Methods & Software **24**(1), 63–87 (2009)
4. Alexanderian, A., Petra, N., Stadler, G., Ghattas, O.: A fast and scalable method for A-optimal design of experiments for infinite-dimensional Bayesian nonlinear inverse problems. SIAM Journal on Scientific Computing **38**(1), A243–A272 (2016)
5. Alexanderian, A., Petra, N., Stadler, G., Ghattas, O.: Mean-variance risk-averse optimal control of systems governed by PDEs with random parameter fields using quadratic approximations. SIAM/ASA Journal on Uncertainty Quantification **5**(1), 1166–1192 (2017)
6. Becker, R., Meidner, D., Vexler, B.: Efficient numerical solution of parabolic optimization problems by finite element methods. Optimization Methods Software **22**, 813–833 (2007)
7. Bui-Thanh, T., Burstedde, C., Ghattas, O., Martin, J., Stadler, G., Wilcox, L.C.: Extreme-scale UQ for Bayesian inverse problems governed by PDEs. In: SC12: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (2012). Gordon Bell Prize finalist
8. Bui-Thanh, T., Ghattas, O., Martin, J., Stadler, G.: A computational framework for infinite-dimensional Bayesian inverse problems: Part I. The linearized case, with application to global seismic inversion. SIAM Journal on Scientific Computing **35**(6), A2494–A2523 (2013)
9. Cao, Y., Li, S., Petzold, L., Serban, R.: Adjoint sensitivity analysis for differential-algebraic equations: the adjoint DAE system and its numerical solution. SIAM Journal on Scientific Computing **24**(3), 1076–1089 (electronic) (2002)
10. Chen, P., Villa, U., Ghattas, O.: Taylor approximation and variance reduction for PDE-constrained optimal control under uncertainty. Journal of Computational Physics **385**, 163–186 (2019)
11. Cioaca, A., Alexe, M., Sandu, A.: Second-order adjoints for solving PDE-constrained optimization problems. Optimization Methods and Software **27**(4-5), 625–653 (2012)

12. Daescu, D.N., Navon, I.M.: An analysis of a hybrid optimization method for variational data assimilation. International Journal of Computational Fluid Dynamics **17**(4), 299–306 (2003).
13. Dunn, J.C., Bertsekas, D.P.: Efficient dynamic programming implementations of Newton's method for unconstrained optimal control problems. Journal of Optimization Theory and Applications **63**(1), 23–38 (1989)
14. Epanomeritakis, I., Akçelik, V., Ghattas, O., Bielak, J.: A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion. Inverse Problems **24**(3), 034015 (26pp) (2008)
15. Fichtner, A., Trampert, J.: Hessian kernels of seismic data functionals based upon adjoint techniques. Geophysical Journal International **185**(2), 775–798 (2011)
16. Griesse, R.: Parametric sensitivity analysis in optimal control of a reaction-diffusion system–part II: practical methods and examples. Optimization Methods and Software **19**(2), 217–242 (2004)
17. Griesse, R., Vexler, B.: Numerical sensitivity analysis for the quantity of interest in PDE-constrained optimization. SIAM Journal on Scientific Computing **29**(1), 22–48 (2007)
18. Gunzburger, M.D.: Perspectives in Flow Control and Optimization. SIAM, Philadelphia (2003)
19. Haber, E., Hanson, L.: Model problems in PDE-constrained optimization. Tech. Rep. TR-2007-009, Emory University (2007)
20. Haftka, R.T., Mróz, Z.: First- and second-order sensitivity analysis of linear and nonlinear structures. AIAA journal **24**(7), 1187–1192 (1986)
21. Haug, E.J.: Second-order design sensitivity analysis of structural systems. AIAA Journal **19**(8), 1087–1088 (1981)
22. Heinkenschloss, M.: Numerical solution of implicitly constrained optimization problems. Tech. Rep. TR08-05, Department of Computational and Applied Mathematics, Rice University (2008)
23. Herzog, R., Sachs, E.: Preconditioned conjugate gradient method for optimal control problems with control and state constraints. SIAM Journal on Matrix Analysis and Applications **31**(5), 2291–2317 (2010)
24. Hicken, J.E.: Inexact Hessian-vector products in reduced-space differential-equation constrained optimization. Optimization and Engineering **15**(3), 575–608 (2014)
25. Hinze, M., Kunisch, K.: Second order methods for optimal control of time–dependent fluid flow. SIAM Journal on Control and Optimization **40**, 925–946 (2001)
26. Hinze, M., Pinnau, R.: Second-order approach to optimal semiconductor design. Journal of Optimization Theory and Applications **133**(2), 179–199 (2007)
27. Hou, G.J.W., Sheen, J.: Numerical methods for second-order shape sensitivity analysis with applications to heat conduction problems. International Journal for Numerical Methods in Engineering **36**(3), 417–435 (1993)
28. Isaac, T., Petra, N., Stadler, G., Ghattas, O.: Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet. Journal of Computational Physics **296**, 348–368 (2015)
29. Jacobson, D.H.: Second-order and second-variation methods for determining optimal control: A comparative study using differential dynamic programming. International Journal of Control **7**(2), 175–196 (1968)
30. Kelley, C.T.: Iterative Methods for Optimization. SIAM, Philadelphia (1999)
31. Le Dimet, F.X., Navon, I.M., Daescu, D.N.: Second-order information in data assimilation. Monthly Weather Review **130**(3), 629–648 (2002)
32. Mayne, D.: A second-order gradient method for determining optimal trajectories of non-linear discrete-time systems. International Journal of Control **3**(1), 85–95 (1966)
33. Métivier, L., Brossier, R., Operto, S., Virieux, J.: Second-order adjoint state methods for full waveform inversion. In: EAGE 2012-74th European Association of Geoscientists and Engineers Conference and Exhibition (2012)
34. Nicholson, R., Petra, N., Kaipio, J.P.: Estimation of the Robin coefficient field in a Poisson problem with uncertain conductivity field. Inverse Problems **34**(11), 115005 (2018)

35. Özyurt, D.B., Barton, P.I.: Cheap second order directional derivatives of stiff ODE embedded functionals. SIAM Journal on Scientific Computing **26**(5), 1725–1743 (2005)
36. Petra, N., Stadler, G.: Model variational inverse problems governed by partial differential equations. Tech. Rep. 11-05, The Institute for Computational Engineering and Sciences, The University of Texas at Austin (2011)
37. Petra, N., Zhu, H., Stadler, G., Hughes, T.J.R., Ghattas, O.: An inexact Gauss-Newton method for inversion of basal sliding and rheology parameters in a nonlinear Stokes ice sheet model. Journal of Glaciology **58**(211), 889–903 (2012)
38. Raffard, R.L., Tomlin, C.J.: Second order adjoint-based optimization of ordinary and partial differential equations with application to air traffic flow. In: American Control Conference, pp. 798–803. IEEE (2005)
39. Rudin, W.: Principles of mathematical analysis, third edn. McGraw-Hill , Inc., New York (1976)
40. Sandu, A., Zhang, L.: Discrete second order adjoints in atmospheric chemical transport modeling. Journal of Computational Physics **227**(12), 5949–5983 (2008)
41. Santosa, F., Symes, W.W.: An analysis of least squares velocity inversion. Society of Exploration Geophysicists (1989)
42. Wang, Z., Navon, I.M., Le Dimet, F.X., Zou, X.: The second order adjoint analysis: theory and applications. Meteorology and Atmospheric Physics **50**(1-3), 3–20 (1992)
43. Zhu, H., Petra, N., Stadler, G., Isaac, T., Hughes, T.J.R., Ghattas, O.: Inversion of geothermal heat flux in a thermomechanically coupled nonlinear Stokes ice sheet model. The Cryosphere **10**, 1477–1494 (2016)