Computing gradients and Hessians using the adjoint method

Alen Alexanderian* and Isaac Sunseri *

Abstract

We illustrate adjoint based gradient and Hessian computation for simple PDE-based model inverse problem. A formal Lagrange approach is used to derive the adjoint based expressions for gradient and Hessian.

1 Introduction

Adjoint-based gradient computation is essential for infinite-dimensional PDE-constrained optimization problems. Examples include PDE-based inverse problems, or control of systems governed by PDEs. The purpose of this note is to provide a quick summary of adjoint based gradient and Hessian computation within the context some simple examples. The techniques presented here are standard. See for instance [2]. The approach for computing the Hessian is adapted from [1]. We consider two examples in this note: a coefficient inversion problem governed by an elliptic PDE, and a source term inversion problem governed by the heat equation. The former is a nonlinear inverse problem, due to the nonlinearity of the parameter-to-observable map, and the latter is a linear inverse problem.

2 Coefficient inversion in an elliptic PDE

We consider the following inverse problem.

$$\min_m \mathcal{J}(m) := \frac{1}{2}\int_\Omega (u-u_d)^2\,d\mathbf{x} + \frac{\gamma}{2}\int_\Omega |\nabla m|^2\,d\mathbf{x},$$

where u solves the PDE

$$\begin{aligned} -\nabla\cdot(e^m\nabla u) &= f \qquad \text{in } \Omega,\\ u &= 0 \qquad \text{on } \partial\Omega. \end{aligned} \tag{2.1}$$

Here $\gamma > 0$ is a regularization parameter. The functional $\mathcal{J}$ is defined for $m \in H^1(\Omega)$. Because of the homogeneous Dirichlet boundary conditions, the solution space of the state equation is $V = H^1_0(\Omega)$. The weak form of (2.1) is as follows: find $u \in V$ such that

$$\int_\Omega e^m \nabla u\cdot\nabla v\,d\mathbf{x} = \int_\Omega fv\,d\mathbf{x}, \quad \text{for all } v \in V.$$

Gradient computation. We seek to describe the action of the directional derivative of $\mathcal{J}$ at a given $m \in H^1(\Omega)$ on a direction $\tilde{m} \in H^1(\Omega)$. We denote this by $g(m)(\tilde{m})$. Generally, for a functional $\Pi(u)$, we use the notation $\Pi_u(u)(\tilde{u})$ to denote the variational derivative of Π at u and in the direction $\tilde{u}$:

$$\Pi_u(u)(\tilde{u}) := \frac{d}{d\epsilon}\Pi(u+\epsilon\tilde{u})\Big|_{\epsilon=0}.$$

*North Carolina State University, Raleigh, NC. E-mail: {alexanderian,ipsunser}@ncsu.edu
Last revised: May 26, 2019.

Following a formal Lagrange approach, we consider the Lagrangian:

$$\mathcal{L}(u, m, p) = \mathcal{J}(m) + \int_\Omega e^m \nabla u \cdot \nabla p \, d\mathbf{x} - \int_\Omega f p \, d\mathbf{x}. \tag{2.2}$$

We have, for a given $m \in H^1(\Omega)$

$$g(m)(\tilde{m}) = \mathcal{L}_m(u, m, p)(\tilde{m}), \quad \tilde{m} \in H^1(\Omega), \tag{2.3a}$$

where u and p, respectively, satisfy:

$$\begin{aligned} \mathcal{L}_p(u, m, p)(\tilde{p}) &= 0, \quad \forall \tilde{p} \in V, \\ \mathcal{L}_u(u, m, p)(\tilde{u}) &= 0, \quad \forall \tilde{u} \in V. \end{aligned}$$

It is straightforward to compute

$$\begin{aligned} \mathcal{L}_p(u, m, p)(\tilde{p}) &= \int_\Omega e^m \nabla u \cdot \nabla \tilde{p} \, d\mathbf{x} - \int_\Omega f \tilde{p} \, d\mathbf{x}, \\ \mathcal{L}_u(u, m, p)(\tilde{u}) &= \int_\Omega (u - u_d) \tilde{u} \, d\mathbf{x} + \int_\Omega e^m \nabla p \cdot \nabla \tilde{u} \, d\mathbf{x}, \\ \mathcal{L}_m(u, m, p)(\tilde{m}) &= \gamma \int_\Omega \nabla m \cdot \nabla \tilde{m} \, d\mathbf{x} + \int_\Omega (\tilde{m} e^m) \nabla u \cdot \nabla p \, d\mathbf{x}. \end{aligned} \tag{2.4}$$

Note that letting $\mathcal{L}_p(u, m, p)(\tilde{p})$ equal to zero, for every $\tilde{p} \in V$, recovers the weak form of the state equation. Letting $\mathcal{L}_u(u, m, p)(\tilde{u})$ vanish for every $\tilde{u} \in V$, leads to the weak form of the adjoint equation.

To summarize, for computing $g(m)(\tilde{m})$, we proceed as follows:

1. Given m, solve the state equation: find $u \in V$ such that

$$\int_\Omega e^m \nabla u \cdot \nabla \tilde{p} \, d\mathbf{x} - \int_\Omega f \tilde{p} \, d\mathbf{x} = 0, \quad \text{for all } \tilde{p} \in V. \tag{2.5}$$

2. Given m and u, solve the adjoint equation: find $p \in V$ such that

$$\int_\Omega e^m \nabla p \cdot \nabla \tilde{u} \, d\mathbf{x} = -\int_\Omega (u - u_d) \tilde{u} \, d\mathbf{x} \quad \text{for all } \tilde{u} \in V. \tag{2.6}$$

3. Evaluate the gradient:

$$g(m)(\tilde{m}) = \gamma \int_\Omega \nabla m \cdot \nabla \tilde{m} \, d\mathbf{x} + \int_\Omega (\tilde{m} e^m) \nabla u \cdot \nabla p \, d\mathbf{x}. \tag{2.7}$$

It is also worth noting the strong form of the adjoint equation. Before doing so, we recall the following Green's identity, which is a multidimensional integration-by-parts formula: for $u, v \in H^1(\Omega)$,

$$\int_\Omega k \nabla u \cdot \nabla v \, d\mathbf{x} = -\int_\Omega v \nabla \cdot (k \nabla u) \, d\mathbf{x} + \int_{\partial\Omega} v \left(k \nabla u \cdot \mathbf{n} \right) \, ds.$$

Now, consider (2.6), using Green's identity and the fact that $\tilde{u} \in V = H_0^1(\Omega)$, we have

$$\int_\Omega \left[-\nabla \cdot (e^m \nabla p) + (u - u_d) \right] \tilde{u} \, d\mathbf{x} = 0, \quad \text{for all } \tilde{u} \in V.$$

Using this, and $p \in V$, we get the strong form of the adjoint equation:

$$\begin{aligned} -\nabla \cdot (e^m \nabla p) &= -(u - u_d) \qquad && \text{in } \Omega, \\ p &= 0 && \text{on } \partial\Omega. \end{aligned}$$

Hessian computation. Computing the action of the Hessian, in a given direction, can also be facilitated through an adjoint based formulation. Note that for fixed m and $\hat{m}$ in $H^1(\Omega)$, the directional derivative is given by

$$g(m)(\hat{m}) = \gamma \int_\Omega \nabla m \cdot \nabla \hat{m}\, d\mathbf{x} + \int_\Omega (\hat{m} e^m) \nabla u \cdot \nabla p\, d\mathbf{x},$$

where

$$\begin{aligned} -\nabla \cdot (e^m \nabla u) &= f && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega. \end{aligned} \tag{2.8}$$

And

$$\begin{aligned} -\nabla \cdot (e^m \nabla p) &= -(u - u_d) && \text{in } \Omega, \\ p &= 0 && \text{on } \partial\Omega. \end{aligned} \tag{2.9}$$

Notice that (2.8) is the strong form of the state equation and (2.9) is the strong form of adjoint equation derived above. We focus on differentiating, $g(m)(\hat{m})$ viewed as a PDE-based implicitly defined functional. Consider the (meta-) Lagrangian:

$$\begin{aligned} \mathcal{L}^H(u, m, p, \hat{u}, \hat{p}; \hat{m}) = \gamma \int_\Omega \nabla m \cdot \nabla \hat{m}\, d\mathbf{x} + \int_\Omega \hat{m} e^m \nabla u \cdot \nabla p\, d\mathbf{x} + \int_\Omega e^m \nabla u \cdot \nabla \hat{p}\, d\mathbf{x} \\ - \int_\Omega f \hat{p}\, d\mathbf{x} + \int_\Omega (u - u_d) \hat{u}\, d\mathbf{x} + \int_\Omega e^m \nabla p \cdot \nabla \hat{u}\, d\mathbf{x}. \end{aligned} \tag{2.10}$$

Letting $\mathcal{L}^H(\hat{p})(\tilde{p}) = 0$ for all $\tilde{p}$ yields the state equation and $\mathcal{L}^H(\hat{u})(\tilde{u}) = 0$ for all $\tilde{u}$ yields the adjoint equation. Note that for notational convenience we have suppressed the arguments $(u, m, p, \hat{u}, \hat{p}; \hat{m})$ of $\mathcal{L}^H$. The action of the Hessian, H on $\hat{m} \in H^1(\Omega)$, in a direction $\tilde{m} \in H^1(\Omega)$ is given by

$$\begin{aligned} H(m)(\hat{m}, \tilde{m}) &= \mathcal{L}^H_m(\tilde{m}) \\ &= \int_\Omega \tilde{m} e^m \nabla u \cdot \nabla \hat{p}\, d\mathbf{x} + \int_\Omega \tilde{m} e^m \nabla p \cdot \nabla \hat{u}\, d\mathbf{x} + \int_\Omega \hat{m} \tilde{m} e^m \nabla u \cdot \nabla p + \gamma \int_\Omega \nabla \tilde{m} \cdot \nabla \hat{m}\, d\mathbf{x}, \end{aligned}$$

where $\hat{u}$ and $\hat{p}$ satisfy

$$\mathcal{L}^H_p(\tilde{p}) = \int_\Omega \hat{m} e^m \nabla u \cdot \nabla \tilde{p}\, d\mathbf{x} + \int_\Omega e^m \nabla \tilde{p} \cdot \nabla \hat{u}\, d\mathbf{x} = 0, \quad \forall \tilde{p} \in V, \tag{2.11}$$

and

$$\mathcal{L}^H_u(\tilde{u}) = \int_\Omega \hat{m} e^m \nabla \tilde{u} \cdot \nabla p\, d\mathbf{x} + \int_\Omega e^m \nabla \tilde{u} \cdot \nabla \hat{p}\, d\mathbf{x} + \int_\Omega \tilde{u} \hat{u}\, dx = 0, \quad \tilde{u} \in V. \tag{2.12}$$

We point out that (2.11) and (2.12) are the weak forms of the so called incremental state and incremental adjoint equations. Their strong forms are easily found to be

$$\begin{aligned} -\nabla \cdot (e^m \nabla \hat{u}) &= \nabla \cdot (\hat{m} e^m \nabla u) && \text{in } \Omega, \\ \hat{u} &= 0 && \text{in } \partial\Omega, \end{aligned} \tag{2.13}$$

and

$$\begin{aligned} \nabla \cdot (e^m \nabla \hat{p}) &= -\hat{u} - \nabla \cdot (\hat{m} e^m \nabla p) && \text{in } \Omega, \\ \hat{p} &= 0 && \text{in } \partial\Omega. \end{aligned} \tag{2.14}$$

We can also write down the strong form of the Hessian application

$$H(m)(\hat{m}) = e^m \nabla u \cdot \nabla \hat{p} + e^m \nabla \hat{u} \cdot \nabla p + \hat{m} e^m \nabla u \cdot \nabla p - \gamma \Delta \hat{m}.$$

3 An inverse problem governed by a time-dependent PDE

Here we illustrate the basic ideas of adjoint-based gradient and Hessian computation within the context of a simple linear inverse problem governed by the heat equation. The model problem here concerns inversion of a coefficient function in the source term.

We consider the following inverse problem:

$$\min_m \mathcal{J}(m) := \frac{1}{2}\int_\Omega \big(u(T,\mathbf{x}) - u_d(\mathbf{x})\big)^2\, d\mathbf{x} + \frac{\alpha}{2}\int_0^T m(t)^2\, dt \tag{3.1a}$$

where

$$\begin{aligned} u_t - \kappa\Delta u &= f(\mathbf{x})m(t) && \text{in } \Omega\times(0,T),\\ u(0,\mathbf{x}) &= 0 && \text{in } \Omega,\\ \kappa\nabla u\cdot\mathbf{n} &= 0 && \text{on } \partial\Omega\times(0,T). \end{aligned} \tag{3.1b}$$

This assumes we have measurements of the temperature at the final time, using which we want to estimate the coefficient function in the right hand side. Note that as in the previous example, this problem formulation is an idealization, and in practice temprea-ture measurements will be available at discrete points in the domain. Below, we build the initial condition $u(0,\mathbf{x}) = 0$ in the test function space for the forward equation.

Gradient computation. As before, we follow a Lagrange multiplier approach for computing the gradient of $\mathcal{J}$. Consider the Lagrangian,

$$\begin{aligned} \mathcal{L}(u,m,p) = \frac{1}{2}\int_\Omega (u(T,\cdot) - u_d)^2\, d\mathbf{x} + \frac{\alpha}{2}\int_0^T m^2\, dt \\ + \int_0^T\int_\Omega u_t p d\mathbf{x}dt + \int_0^T\int_\Omega \kappa\nabla u\cdot\nabla p\, d\mathbf{x}dt - \int_0^T\int_\Omega f(\mathbf{x})m(t)p\, d\mathbf{x}dt. \end{aligned}$$

Note that $p = p(t,\mathbf{x})$ is the adjoint variable.

Setting $\mathcal{L}_p(\tilde{p}) = 0$, for every $\tilde{p}$, we recover the weak form of the state equation. The adjoint equation is derived by considering $\mathcal{L}_u(\tilde{u}) = 0$, for every $\tilde{u}$. This translates to

$$\int_\Omega \big(u(T,\cdot) - u_d\big)\tilde{u}(T,\cdot)\, d\mathbf{x} + \int_0^T\int_\Omega \tilde{u}_t p d\mathbf{x}dt + \int_0^T\int_\Omega \kappa\nabla\tilde{u}\cdot\nabla p\, d\mathbf{x}dt = 0,$$

for every $\tilde{u}$. Performing integration by parts in time and space, we have

$$\begin{aligned} \int_\Omega \big(u(T,\cdot) - u_d\big)\tilde{u}(T,\cdot)\, d\mathbf{x} + \int_\Omega p(T,\cdot)\tilde{u}(T,\cdot)\, d\mathbf{x} - \int_0^T\int_\Omega p_t\tilde{u} d\mathbf{x}dt \\ - \int_0^T\int_\Omega (\kappa\Delta p)\tilde{u}\, d\mathbf{x}dt + \int_0^T\int_{\partial\Omega} \kappa(\nabla p\cdot\mathbf{n})\tilde{u}\, d\mathbf{x}dt = 0, \end{aligned}$$

for all $\tilde{u}$. From this we obtain the strong form of the adjoint equation:

$$\begin{aligned} -p_t - \kappa\Delta p &= 0 && \text{in } \Omega\times(0,T),\\ p(T,\mathbf{x}) &= -(u(T,\cdot) - u_d)) && \text{in } \Omega,\\ \kappa\nabla p\cdot\mathbf{n} &= 0 && \text{on } \partial\Omega\times(0,T). \end{aligned}$$

As for the gradient,

$$g(m)(\tilde{m}) = \mathcal{L}_m(\tilde{m}) = \alpha\int_0^T m(t)\tilde{m}(t)\, dt - \int_0^T\int_\Omega f(\mathbf{x})\tilde{m}(t)p(t,\mathbf{x})\, d\mathbf{x}dt.$$

In strong form,

$$g(m) = \alpha m - \int_\Omega f(\mathbf{x})p(t,\mathbf{x})\, d\mathbf{x}.$$

Hessian computation. As done for the elliptic inverse problem, we consider differentiating through the directional derivative $g(m)(\hat{m})$, as constrained by the solutions of the forward and adjoint equations. To this end, we define the (meta-) Lagrangian:

$$\begin{aligned}\mathcal{L}^H(u,m,p,\hat{u},\hat{p};\hat{m}) = \alpha \int_\Omega m(t)\hat{m}(t)\,dt - \int_0^T \int_\Omega f(\mathbf{x})\hat{m}(t)p(t,\mathbf{x})\,d\mathbf{x}dt \\ + \int_0^T \int_\Omega u_t \hat{p} d\mathbf{x}dt + \int_0^T \int_\Omega \kappa\nabla u \cdot \nabla\hat{p}\,d\mathbf{x}dt - \int_0^T \int_\Omega f(\mathbf{x})m(t)\hat{p}\,d\mathbf{x}dt \\ + \int_\Omega \big(u(T,\cdot) - u_d\big)\hat{u}(T,\cdot)\,d\mathbf{x} - \int_0^T \int_\Omega p_t \hat{u} d\mathbf{x}dt + \int_0^T \int_\Omega \kappa\nabla\hat{u}\cdot\nabla p\,d\mathbf{x}dt.\end{aligned}$$

Letting variations of $\mathcal{L}^H$ with respect to $\hat{p}$ and $\hat{u}$ vanish recovers the forward and adjoint equations. We can derive the incremental state and adjoint equations by setting variations of $\mathcal{L}^H$ with respect to p and u equal to zero (along all test functions). This results in the incremental state and adjoint equations:

$$\begin{aligned}\hat{u}_t - \kappa\Delta\hat{u} &= f(\mathbf{x})\hat{m}(t) && \text{in } \Omega\times(0,T),\\ \hat{u}(0,\mathbf{x}) &= 0 && \text{in } \Omega, \quad \text{(incremental state equation)}\\ \kappa\nabla\hat{u}\cdot\mathbf{n} &= 0 && \text{on } \partial\Omega\times(0,T),\end{aligned}$$

and

$$\begin{aligned}-\hat{p}_t - \kappa\Delta\hat{p} &= 0 && \text{in } \Omega\times(0,T),\\ \hat{p}(T,\mathbf{x}) &= -\hat{u}(T,\mathbf{x}) && \text{in } \Omega, \quad \text{(incremental adjoint equation)}\\ \kappa\nabla\hat{p}\cdot\mathbf{n} &= 0 && \text{on } \partial\Omega\times(0,T).\end{aligned}$$

The Hessian apply $H(m)(\hat{m},\tilde{m})$ at $m \in L^2(\Omega)$ on $\hat{m}$, along test function $\tilde{m}$ is

$$H(m)(\hat{m},\tilde{m}) = \mathcal{L}^H_m(\tilde{m}) = \alpha \int_\Omega \hat{m}(t)\tilde{m}(t)\,dt - \int_0^T \int_\Omega f(\mathbf{x})\tilde{m}(t)\hat{p}\,d\mathbf{x}dt.$$

In strong form,

$$H(m)\hat{m} = \alpha\hat{m} - \int_\Omega f(\mathbf{x})\hat{p}(\cdot,\mathbf{x})\,d\mathbf{x}.$$

References

[1] Ghattas, Omar. Lecture notes. `http://users.ices.utexas.edu/~omar/inverse_problems/Sensitivity_Analysis_Notes.pdf`.

[2] Gunzburger, Max D. Perspectives in flow control and optimization. Vol. 5. Siam, 2003.