Eigenvalue sensitivity analysis for a class of implicitly defined matrices

Alen Alexanderian* and Abhijit Chowdhary†

Abstract

We use a Lagrange multiplier approach to derive formulas for eigenvalue sensitivities for a class of implicitly defined matrices. The present study is motivated by eigenvalue sensitivity analysis for the data misfit Hessian in infinite-dimensional linear inverse problems. The derivations are meant to provide the readers some insight before transitioning to the more involved infinite-dimensional setting. We also provide a simple numerical example.

1 Introduction

We consider an implicitly defined matrix $\mathbf{H} \in \mathbb{R}^{N \times N}$ whose action on a vector $\boldsymbol{v}$ is defined as follows:

$$\mathbf{H}\boldsymbol{v} := \mathbf{C}^\top \boldsymbol{p}, \tag{1.1a}$$

where

$$\mathbf{A}\boldsymbol{u} + \mathbf{C}\boldsymbol{v} = \mathbf{0}, \tag{1.1b}$$

$$\mathbf{A}^\top \boldsymbol{p} + \mathbf{Q}^\top \mathbf{Q}\boldsymbol{u} = \mathbf{0}. \tag{1.1c}$$

Here, $\mathbf{A} \in \mathbb{R}^{M \times M}$ is nonsingular, $\mathbf{C} \in \mathbb{R}^{M \times N}$, and $\mathbf{Q} \in \mathbb{R}^{D \times M}$. The motivation to study this specific form of operator $\mathbf{H}$ comes from inverse problems. Namely, in a linear inverse problem the data misfit Hessian admits such a representation.[1] Note that we can collapse the above definition of $\mathbf{H}$ to show

$$\mathbf{H} = \mathbf{C}^\top \mathbf{A}^{-\top} \mathbf{Q}^\top \mathbf{Q} \mathbf{A}^{-1} \mathbf{C}. \tag{1.2}$$

Clearly, $\mathbf{H}$ is a symmetric positive semidefinite matrix. In large-scale problems, building this matrix is computationally prohibitive. Therefore, matrix free approaches that only the action of $\mathbf{H}$ to vectors are needed. This is facilitated by (1.1), which provides an efficient approach for computing matrix-vector products with $\mathbf{H}$. Namely, $\mathbf{H}\boldsymbol{v}$ can be computed at the cost of two linear solves.

Suppose $\mathbf{A}$ is parameterized by a parameter vector $\boldsymbol{\theta} \in \mathbb{R}^p$. We assume $\mathbf{A}(\boldsymbol{\theta})$ is non-singular for every $\boldsymbol{\theta}$.[2] Suppose further that $\mathbf{A}$ is a differentiable function of a parameter vector $\boldsymbol{\theta}$. We seek to compute the partial derivatives of the eigenvalues of $\mathbf{H} = \mathbf{H}(\boldsymbol{\theta})$ with respect to the components of $\boldsymbol{\theta}$. Specifically, let λ be a simple eigenvalue of $\mathbf{H}$ with a corresponding eigenvector $\boldsymbol{v}$.[3] We can describe the relation $\mathbf{H}\boldsymbol{v} = \lambda\boldsymbol{v}$ as follows,

$$\mathbf{A}(\boldsymbol{\theta})\boldsymbol{u} + \mathbf{C}\boldsymbol{v} = \mathbf{0}, \tag{1.3a}$$

$$\mathbf{A}(\boldsymbol{\theta})^\top \boldsymbol{p} + \mathbf{Q}^\top \mathbf{Q}\boldsymbol{u} = \mathbf{0}, \tag{1.3b}$$

$$\mathbf{C}^\top \boldsymbol{p} = \lambda\boldsymbol{v}. \tag{1.3c}$$

Below, we describe the process of computing the partial derivatives of $\lambda = \lambda(\boldsymbol{\theta})$.

[1] *This also describes the form of the Gauss–Newton Hessian in nonlinear inverse problems.*

[2] *In practice, the parameter vector $\boldsymbol{\theta}$ may take values in a set of admissible parameter vectors.*

[3] *We consider simple eigenvalues to ensure their differentiability; see [3, Page 130].*

*North Carolina State University, Raleigh, NC, USA. E-mail alexanderian@ncsu.edu

†North Carolina State University, Raleigh, NC, USA.

Last revised: October 22, 2024

2 Derivation

Assuming $\boldsymbol{v}$ has unit norm, we have $\lambda = \boldsymbol{v}^\top \mathbf{H}\boldsymbol{v}$. We can thus define the mapping $\boldsymbol{\theta} \mapsto \lambda(\boldsymbol{\theta})$ implicitly as follows,

$$\lambda(\boldsymbol{\theta}) = \boldsymbol{v}^\top \mathbf{C}^\top \boldsymbol{p}, \tag{2.1a}$$

where

$$\mathbf{A}(\boldsymbol{\theta})\boldsymbol{u} + \mathbf{C}\boldsymbol{v} = \mathbf{0}, \tag{2.1b}$$

$$\mathbf{A}(\boldsymbol{\theta})^\top \boldsymbol{p} + \mathbf{Q}^\top \mathbf{Q}\boldsymbol{u} = \mathbf{0}, \tag{2.1c}$$

$$\boldsymbol{v}^\top \boldsymbol{v} = 1. \tag{2.1d}$$

Note that we have also added the requirement that $\boldsymbol{v}$ be of unit norm.

We next use a Lagrange multiplier approach to compute the adjoint-based expressions for the partial derivatives of $\lambda(\boldsymbol{\theta})$.[4] Consider the Lagrangian

[4] *For an overview of adjoint based gradient computation, see the review [5].*

$$\mathcal{L} = \left\langle \boldsymbol{v}, \mathbf{C}^\top \boldsymbol{p} \right\rangle + \left\langle \boldsymbol{p}^\star, \mathbf{A}(\boldsymbol{\theta})\boldsymbol{u} + \mathbf{C}\boldsymbol{v} \right\rangle + \left\langle \boldsymbol{u}^\star, \mathbf{A}(\boldsymbol{\theta})^\top \boldsymbol{p} - \mathbf{Q}^\top \mathbf{Q}\boldsymbol{u} \right\rangle + \lambda^\star \left(1 - \boldsymbol{v}^\top \boldsymbol{v}\right). \tag{2.2}$$

Note that $\mathcal{L} = \mathcal{L}(\boldsymbol{u}, \boldsymbol{p}, \boldsymbol{v}, \boldsymbol{u}^\star, \boldsymbol{p}^\star, \lambda^\star)$. The arguments of $\mathcal{L}$ in (2.2) are suppressed for notational convenience. Also, here $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product.

The derivative of $\lambda(\boldsymbol{\theta})$ with respect to θ_j are given by $\mathcal{L}_{\theta_j}$, where the variations of $\mathcal{L}$ with respect to all other variables are set to zero. Setting the variations of $\mathcal{L}$ with respect to $\boldsymbol{p}^\star$, $\boldsymbol{u}^\star$, and $\lambda^\star$ equal to zero recovers equations (2.1b)–(2.1d). Moreover, we have

$$\mathcal{L}_{\boldsymbol{u}} = \mathbf{A}^\top \boldsymbol{p}^\star + \mathbf{Q}^\top \mathbf{Q}\boldsymbol{u}^\star, \tag{2.3a}$$

$$\mathcal{L}_{\boldsymbol{p}} = \mathbf{A}\boldsymbol{u}^\star + \mathbf{C}\boldsymbol{v}, \tag{2.3b}$$

Letting these variations vanish, we obtain equations for $\boldsymbol{u}^\star$ and $\boldsymbol{p}^\star$, which are identical to the equations for $\boldsymbol{u}$ and $\boldsymbol{p}$. Therefore, we have that $\boldsymbol{u}^\star = \boldsymbol{u}$ and $\boldsymbol{p}^\star = \boldsymbol{p}$. Finally, we note,

$$\mathcal{L}_{\theta_j} = \left\langle \boldsymbol{p}^\star, [\partial_j \mathbf{A}]\boldsymbol{u} \right\rangle + \left\langle \boldsymbol{u}^\star, [\partial_j \mathbf{A}]^\top \boldsymbol{p} \right\rangle = \left\langle \boldsymbol{p}, [\partial_j \mathbf{A}]\boldsymbol{u} \right\rangle + \left\langle \boldsymbol{u}, [\partial_j \mathbf{A}]^\top \boldsymbol{p} \right\rangle = 2 \left\langle \boldsymbol{p}, [\partial_j \mathbf{A}]\boldsymbol{u} \right\rangle,$$

where ∂_j is used a shorthand for $\frac{\partial}{\partial \theta_j}$. Therefore, to compute the sensitivity of λ with respect to θ_j, we need to perform the following calculations:

- solve (2.1b) for $\boldsymbol{u}$;
- solve (2.1c) for $\boldsymbol{p}$; and
- evaluate eigenvalue sensitivities

$$\partial_j \lambda(\boldsymbol{\theta}) = 2 \left\langle \boldsymbol{p}, [\partial_j \mathbf{A}]\boldsymbol{u} \right\rangle, \quad j \in \{1, \ldots, p\}. \tag{2.4}$$

3 Remarks

Note that it is possible to compute the eigenvalue sensitivities for $\mathbf{H}$ directly using the definition (1.2) and the well-known formula for eigenvalue sensitivities of symmetric matrices with simple eigenvalues [4], given by

$$\partial_j \lambda = \left\langle \boldsymbol{v}, [\partial_j \mathbf{H}]\boldsymbol{v} \right\rangle, \quad j \in \{1, \ldots, n\}.$$

The derivation in the previous section has the advantage that (i) it is self-contained for the present class of problems; (ii) it is matrix-free by construction; and (iii) it guides eigenvalue sensitivity analysis for infinite-dimensional inverse problems governed by PDEs. This also facilitates derivation of adjoint-based expressions for eigenvalue sensitivities in the cases of nonlinear inverse problems. In the infinite-dimensional setting, the equations (2.1b)–(2.1c) will be replaced by the weak forms of the so-called incremental state and adjoint equations. Such calculations can be found in [2] for the case of linear inverse problems and [1] for nonlinear inverse problems. The present derivation also facilitates eigenvalue sensitivity analysis in cases where it is otherwise unclear how one would form the action of $\partial_j \mathbf{H}$.

4 Numerical example

To numerically illustrate the approach discussed in the previous section, we consider a simple example.[5] Namely, we let

$$\mathbf{A} = \begin{bmatrix} -20(1+a_2\theta_2) & 3(1+a_3\theta_3) & 0 \\ 1+a_1\theta_1 & -20(1+a_2\theta_2) & 3(1+a_3\theta_3) \\ 0 & 1+a_1\theta_1 & -20(1+a_2\theta_2) \end{bmatrix} \quad \text{and} \quad \boldsymbol{a} = \begin{bmatrix} 0.08 \\ 0.06 \\ 0.09 \end{bmatrix}.$$

[5] For a more interesting example, involving a large-scale inverse problem, see [2].

Also, the remaining matrices for the present example are as follows:

$$\mathbf{Q} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}.$$

In this case, the matrix $\mathbf{H}$ is a 2×2 matrix. We consider the sensitivity of the largest eigenvalue, denoted generically by λ, to the components of $\boldsymbol{\theta} = \begin{bmatrix}\theta_1 & \theta_2 & \theta_3\end{bmatrix}^\top$, at the nominal parameter vector $\bar{\boldsymbol{\theta}} = [0.1\ 0.2\ 0.3]^\top$. In Figure 1, we report the difference between the finite-difference approximation $\partial_j^h \lambda$ and the (exact) derivative $\partial_j \lambda$ computed using (2.4). The finite-difference is computed using a forward difference formula.

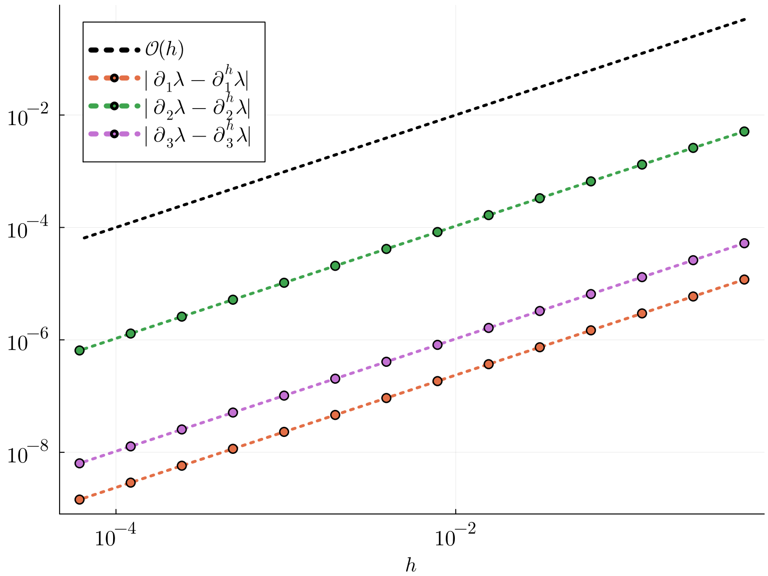

Figure 1: Finite difference check for the formula (2.4).

References

[1] A. Chowdhary, A. Attia, and A. Alexanderian. Robust optimal design of large-scale bayesian nonlinear inverse problems. *arXiv preprint arXiv:2409.09137*, 2024.

[2] A. Chowdhary, S. Tong, G. Stadler, and A. Alexanderian. Sensitivity analysis of the information gain in infinite-dimensional bayesian linear inverse problems. *International Journal for Uncertainty Quantification*, 14(6), 2024.

[3] P. D. Lax. *Linear algebra and its applications*. Pure and Applied Mathematics (Hoboken). Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2007.

[4] M. L. Overton and R. S. Womersley. Second derivatives for optimizing eigenvalues of symmetric matrices. *SIAM Journal on Matrix Analysis and Applications*, 16(3):697–718, 1995.

[5] R.-E. Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503, 2006.