Sensitivity analysis and steepest-ascent directions for degenerate-eigenvalue problems

Steven G. Johnson

August 2, 2008

1 Introduction

In this note, we review the techniques used to perform sensitivity analysis in design problems involving eigenvalues with multiplicity ≥ 1, which require special care [1] compared to the ordinary adjoint methods [2] typically used for sensitivity analysis. In particular, we consider the case where one is optimizing some function of an eigenvalue of an $N \times N$ generalized eigenvalue problem

$$A(\mathbf{p})\mathbf{x} = \lambda B(\mathbf{p})\mathbf{x} \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^N$ and $A(\mathbf{p}), B(\mathbf{p}) \in \mathbb{C}^{N\times N}$ are self-adjoint matrices (and B is positive-definite, leading to real λ) that are explicit functions of M "design parameters" $\mathbf{p} \in \mathbb{R}^M$. We wish to compute the *sensitivity* of λ to small changes in $\mathbf{p}$, i.e. the rate of change $\frac{d\lambda}{d\epsilon}$ when $\mathbf{p} \to \mathbf{p}+\epsilon\mathbf{d}$ for any given direction $\mathbf{d} \in \mathbb{R}^M$. We also wish to compute the *steepest-ascent direction* $\mathbf{g} \in \mathbb{R}^{\mathbb{M}}$: the direction in which to change $\mathbf{p}$ to make λ increase most quickly.

For a simple (non-degenerate) eigenvalue λ, both of these questions are answered by computing the gradient $\nabla_{\mathbf{p}}\lambda$, giving the sensitivity or directional derivative $\frac{d\lambda}{d\epsilon} = \mathbf{d} \cdot \nabla_{\mathbf{p}}\lambda$ and the steepest-ascent direction $\mathbf{g} \propto \nabla_{\mathbf{p}}\lambda$; the gradient can be computed easily from $\mathbf{x}$ via first-order perturbation theory [3] or (equivalently) by an adjoint method [2]. However, the problem is considerably complicated in cases involving degenerate eigenvalues (multiplicity > 1), in which case the gradient *per se* is not even well-defined [1]. Degenerate eigenvalues arise surprisingly often in cases involving optimization of eigenvalues, unfortunately. For example, a typical problem is the maximization of the minimum eigenvalue (e.g. the maximization of the minimum vibration frequency of some mechanical structure [1]), and in this case the optimization will often naturally push the system towards a case where the minimum eigenvalue is repeated. As another example, in maximization of band gaps (eigenvalue splittings) of periodic structures, the optimum often occurs at a degenerate case where the gap edge is tangent to more than one eigenvalue band [4].

Fortunately, in these cases a generalization of the gradient/adjoint analysis is possible that allows one to compute both sensitivities and steepest-ascent

directions very efficiently from the solution of the unperturbed problem (1). The sensitivity analysis is equivalent to what is called *degenerate perturbation theory* in quantum mechanics [3], while the determination of the steepest-ascent direction requires additional steps [1].

2 Sensitivity analysis & generalized gradients

Let us suppose that the eigenproblem (1) has K solutions $\mathbf{x}^{(k)}$, for $k = 1, \ldots, K$, with the same eigenvalue λ (i.e., λ has multiplicity K), and we have orthonormalized these so that $\mathbf{x}^{(k)} \cdot B\mathbf{x}^{(k')} = \delta_{kk'}$ (where $\delta_{kk'}$ is the Kronecker delta and the inner product is $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^\dagger \mathbf{y}$, with $\dagger$ the conjugate-transpose). The choice of eigenvectors is, of course, not unique, because any linear combination also gives an eigenvector for λ. Now, we perform a small perturbation $\mathbf{p} \to \mathbf{p} + \epsilon\mathbf{d}$ for some direction $\mathbf{d}$, and we wish to know the rate of change $d\lambda/d\epsilon$. The difficulty is that the perturbation will, in general, break the degeneracy, so that λ splits into multiple distinct eigenvalues, but we don't know which linear combination of the $\mathbf{x}^{(k)}$'s will go with which split eigenvalue. Fortunately, this problem can be solved via a small $K \times K$ eigenproblem as follows.

First, we form an arbitrary linear combination

$$\mathbf{x} = \sum_k u_k \mathbf{x}^{(k)} \tag{2}$$

of the degenerate eigenvectors, where $\mathbf{u} \in \mathbb{C}^K$ is an unknown to be determined. This $\mathbf{x}$ is itself an eigenvector, of course, and satisfies (1). Now, we simply differentiate both sides of (1) with respect to ϵ (assuming $\mathbf{u}$ will be chosen to yield an eigevector even after an infinitesimal perturbation), and then take the inner product of both sides with $\mathbf{x}^{(k)}$ for any k, to obtain:

$$\mathbf{x}^{(k)} \cdot \left(\frac{dA}{d\epsilon}\mathbf{x} + A\frac{d\mathbf{x}}{d\epsilon} \right) = \mathbf{x}^{(k)} \cdot \left(\frac{d\lambda}{d\epsilon} B\mathbf{x} + \lambda \frac{dB}{d\epsilon}\mathbf{x} + \lambda B \frac{d\mathbf{x}}{d\epsilon} \right). \tag{3}$$

Because A and B are self-adjoint and λ is real, however, we obtain $\mathbf{x}^{(k)} \cdot A\frac{d\mathbf{x}}{d\epsilon} = A\mathbf{x}^{(k)} \cdot \frac{d\mathbf{x}}{d\epsilon} = \lambda \mathbf{x}^{(k)} \cdot B\frac{d\mathbf{x}}{d\epsilon}$, which cancels the $\frac{d\mathbf{x}}{d\epsilon}$ term on the right-hand side of (3). The remaining terms can be solved for the sensitivity $\frac{d\lambda}{d\epsilon}$:

$$\begin{aligned} \mathbf{x}^{(k)} \cdot \left(\frac{dA}{d\epsilon} - \lambda \frac{dB}{d\epsilon} \right) \mathbf{x} &= \frac{d\lambda}{d\epsilon} \mathbf{x}^{(k)} \cdot B\mathbf{x} = \frac{d\lambda}{d\epsilon} c_k \\ &= \sum_{k'} C_{kk'} c_{k'}, \end{aligned}$$

where

$$C_{kk'} = \mathbf{x}^{(k)} \cdot \left(\frac{dA}{d\epsilon} - \lambda \frac{dB}{d\epsilon} \right) \mathbf{x}^{(k')} = \overline{C_{k'k}}, \tag{4}$$

where the overline denotes complex conjugation. But this is precisely a $K \times K$ self-adjoint eigenproblem

$$C\mathbf{u} = \frac{d\lambda}{d\epsilon}\mathbf{u}. \tag{5}$$

The eigenvalues of C in eq. (5) are the rates of change $\frac{d\lambda}{d\epsilon}$ of the K degenerate eigenvalues λ, and the K eigenvectors are the coefficients $\mathbf{c}$ of the resulting eigenvectors $\mathbf{x}$ (to zero-th order in ϵ) after the eigenvalue λ is split by the perturbation (if C has distinct eigenvalues).

Of course, the matrix C depends on the perturbation direction $\mathbf{d}$. For example, $\frac{dA_{nn'}}{d\epsilon} = \sum_m d_m \frac{\partial A_{nn'}}{\partial p_m} = \mathbf{d} \cdot \nabla_{\mathbf{p}} A_{nn'}$. Applying this to both A and B and substituting into eq. (4), we obtain

$$C_{kk'} = \mathbf{d} \cdot \mathbf{f}^{(kk')}, \tag{6}$$

where the vectors $\mathbf{f}^{(kk')} \in \mathbb{C}^M$ are the "generalized gradient" vectors [1] with components:

$$f_m^{(kk')} = \mathbf{x}^{(k)} \cdot \left(\frac{dA}{dp_m} - \lambda \frac{dB}{dp_m} \right) \mathbf{x}^{(k')} = \overline{f_m^{(k'k)}}. \tag{7}$$

We will use these vectors further in the next section to find the steepest-ascent direction.

3 Steepest-ascent direction

In this section, we derive the steepest-ascent direction, which is used as part of numerous gradient-based local-optimization algorithms, and is critical for efficient search of high-dimensional parameter spaces. What is presented here is a much-abbreviated version of the more complete analysis in Ref. [1], which also includes the possibility of a linear constraint on $\mathbf{x}$ in addition to the eigenproblem.

The key observation is that, in optimization problems involving a degenerate eigenvalue, we need to push *all* the degenerate eigenvalues *together*. It does no good to push one of the eigenvalues in the direction you want if another eigenvalue goes in the opposite direction (e.g. if you are trying to maximize the minimum eigenvalue). Nor is it ideal to push one eigenvalue by a lot if another one moves only a little—you want all of the eigenvalues to move by the same amount. This criterion, however, corresponds simply to the requirement that C be proportional to the identity matrix. In particular, we want to find a direction $\mathbf{g} \in \mathbf{R}^M$ so that

$$C_{kk'} = \mathbf{g} \cdot \mathbf{f}^{(kk')} = \delta_{kk'}. \tag{8}$$

Without loss of generality, we can restrict ourselves to considering $\mathbf{g}$ in the span of the $\mathbf{f}^{(kk')}$—any component of $\mathbf{g}$ in the orthogonal complement of the $\mathbf{f}^{(kk')}$ has no effect on C and hence is irrelevant to the sensitivity. That is, we can write

$$\mathbf{g} = \sum_{kk'} \gamma_{kk'} \mathbf{f}^{(kk')} \tag{9}$$

for some $\gamma_{kk'} \in \mathbb{C}$ satisfying $\gamma_{kk'} = \overline{\gamma_{k'k}}$ (so that $\mathbf{g}$ is real). Combining eqs. (8–9), we obtain the following equations for the unknown coefficients $\gamma_{kk'}$:

$$\sum_{kk'} \left[\overline{\mathbf{f}^{(\ell\ell')}} \cdot \mathbf{f}^{(kk')} \right] \gamma_{kk'} = \delta_{\ell\ell'}. \tag{10}$$

Because the $\ell\ell'$ and $\ell'\ell$ equations are complex conjugates, this yields exactly the right number of equations to solve for the unknowns $\gamma_{kk'}$. In the end, we have $K + K(K-1) = K^2$ equations for the K^2 distinct real and imaginary parts of $\gamma_{kk'}$ for complex-Hermitian matrices A and B, or $K+K(K-1)/2 = K(K+1)/2$ equations for the distinct real $\gamma_{kk'}$ for real-symmetric A and B. These equations must be nonsingular if such a C exists. More generally, they can be shown to be singular only when no direction exists that pushes all the eigenvalues in the same direction—this is equivalent to the optimality criterion in an optimization problem, because in that case it is impossible to push λ further in the direction that you want [1].

However, in practice, it is desirable to modify the $\ell = \ell'$ equations of (10) slightly, to account for the fact that in numerical computations the eigenvalues are never exactly degenerate. In particular, if two eigenvalues λ_k and $\lambda_{k'}$ are close enough to one another, say $|\lambda_k - \lambda_{k'}| < \delta$ for some small δ, you want to treat them as degenerate—otherwise, optimizing one eigenvalue may still push the other nearby eigenvalue in the other direction, so that they cross and your objective function is worsened rather than improved. Instead, if the eigenvalues are close enough, then you want your next optimization step to push them in the same direction. However, say you are maximizing/minimizing the minimum/maximum eigenvalue λ, respectively—in that case, you want to push the λ_k that are closer to λ by more than the λ_k that are farther from λ. This can be achieved by changing eq. (10) for $\ell = \ell'$ to, for example:

$$C_{\ell\ell} = \sum_{kk'} \left[\overline{\mathbf{f}^{(\ell\ell)}} \cdot \mathbf{f}^{(kk')}\right] \gamma_{kk'} = 1 - |\lambda - \lambda_\ell|. \tag{11}$$

The reason why this works is that the diagonals $C_{\ell\ell}$ are precisely proportional to the rates of change of each λ_k for distinct λ_k, because they correspond to the nondegenerate ($K = 1$) version of the sensitivity equation (5).

Now, the vector $\mathbf{g}$ is the steepest-ascent direction, but what should its magnitude be? Most optimization software will expect you to supply a steepest-ascent direction whose magnitude is equal to the directional derivative in that direction, just like the "real" gradient. That is, we want the "effective gradient" vector $\widetilde{\nabla_{\mathbf{p}}\lambda} = \alpha\mathbf{g}$, for some α, such that, for a change $\mathbf{p} \to \mathbf{p} + \epsilon(\mathbf{g}/\|\mathbf{g}\|)$, we obtain the directional derivative $\widetilde{\nabla_{\mathbf{p}}\lambda} \cdot (\mathbf{g}/\|\mathbf{g}\|) = d\lambda/d\epsilon$. From the sensitivity equation (5), since C is proportional to the identity for this direction, $d\lambda/d\epsilon = (\mathbf{g}/\|\mathbf{g}\|) \cdot \mathbf{f}^{(kk')} = 1/\|\mathbf{g}\| = (\alpha\mathbf{g}) \cdot (\mathbf{g}/\|\mathbf{g}\|) = \overline{\alpha}\|\mathbf{g}\|$, and hence $\alpha = 1/\|\mathbf{g}\|^2$. In short, our "effective gradient" vector for use in optimization is:

$$\widetilde{\nabla_{\mathbf{p}}\lambda} = \mathbf{g}/\|\mathbf{g}\|^2. \tag{12}$$

As we approach the optimality criterion on λ, the equations for $\gamma_{kk'}$ become singular and the length $\|\mathbf{g}\|$ diverges [1]—thus, the effective gradient $\widetilde{\nabla_{\mathbf{p}}\lambda}$ goes to zero at optimality as you might expect. (Note that this is *not* a true gradient, however, in that it gives the correct directional derivative only for directions parallel to $\mathbf{g}$!)

There is one important special case that arises in optimization of band structures [4]: the "multiplicity" of a given eigenvalue λ will often come from eigensolutions at different Bloch wavevectors in the band diagram. In this case, the proper inner product to compute $\mathbf{f}^{(kk')}$ between two eigensolutions at different wavevectors involves an integral over all space, and the difference in wavevectors will yield $\mathbf{f}^{(kk')} = 0$ for k and k' corresponding to different wavevectors (assuming the change in $\mathbf{p}$ preserves the underlying periodicity). We can then freely set $\gamma_{kk'} = 0$ for that (k, k') pair, reducing the number of unknowns to solve for (and requiring only inner products over the unit cell for degenerate eigensolutions at equal wavevectors). If *all* the degeneracies come from distinct wavevectors, then we will only have K equations to solve for the diagonal γ_{kk} coefficients.

References

[1] A. P. Seyranian, E. Lund, and N. Olhoff, "Multiple eigenvalues in structural optimization problems," *Structural Optimization*, vol. 8, pp. 207–227, 1994.

[2] S. G. Johnson, "Notes on adjoint methods for 18.336." Online at `http://math.mit.edu/ stevenj/18.336/adjoint.pdf`, October 2007.

[3] C. Cohen-Tannoudji, B. Din, and F. Laloë, *Quantum Mechanics*. Paris: Hermann, 1977. *Vol. One*, ch. 2; and *Vol. Two*, ch. 11 and 13.

[4] O. Sigmund and J. S. Jensen, "Systematic design of phononic band-gap materials and structures by topology optimization," *Phil. Trans. Roy. Soc. London A*, vol. 361, pp. 1001–1019, 2003.