Matrices $A(t)$ depending on a Parameter t

Math. 412, Fall 1995 Jerry L. Kazdan

If a square matrix $A(t)$ depends smoothly on a parameter t, are its eigenvalues and eigenvectors also smooth functions of t? The answer is "yes" most of the time, but not always. This story, while old, is interesting and elementary — and deserves to be better known.
One can also ask the same question for objects such as the Schrödinger operator whose potential depends on a parameter, where much of current understanding arose.

Warm-up Exercise

Given a polynomial $p(x,t) = x^n + a_{n-1}(t)x^{n-1} + \cdots + a_1(t)x + a_0(t)$ whose coefficients depend smoothly on a parameter t. Assume at $t = 0$ the number $x = c$ is a simple root of this polynomial, $p(c,0) = 0$. Show that for all t sufficiently near 0 there is a unique root $x(t)$ with $x(0) = c$ that depends smoothly on t. Moreover, if $p(x,t)$ is a real analytic function of t, that is, it has a convergent power series expansion in t near $t = 0$, then so does $x(t)$.

SOLUTION: Given that $p(c,0) = 0$ we want to solve $p(x,t) = 0$ for $x(t)$ with $x(0) = c$. The assertions are immediate from the implicit function theorem. Since $x(0) = c$ is a simple zero of $p(x,0) = 0$, then $p(x,0) = (x-c)g(x)$, where $g(c) \neq 0$. Thus the derivative $p_x(c,0) \neq 0$.

The example $p(x,t) := x^3 - t = 0$, so $x(t) = t^{1/3}$, shows $x(t)$ may not be a smooth function at a multiple root. In this case the best one can get is a Puiseux expansion in fractional powers of t (see [Kn, ¶15]).

The Generic Case: a simple eigenvalue

In the following, let λ be an eigenvalue and X a corresponding eigenvector of a matrix A. We say λ is a *simple eigenvalue* if λ is a simple root of the characteristic polynomial. We will use the equivalent version: *if* $(A-\lambda)^2 V = 0$*, then* $V = cX$ *for some constant* c. The point is to eliminate matrices such as the zero 2×2 matrix $A = 0$, where $\lambda = 0$ is a double eigenvalue and any vector $V \neq 0$ is an eigenvector, as well as the more complicated matrix $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ which has $\lambda = 0$ as an eigenvalue with geometric multiplicity one but algebraic multiplicity two.

Theorem *Given a square matrix* $A(t)$ *whose elements depend smoothly on a real parameter* t*, if* $\lambda = \lambda_0$ *is a* simple *eigenvalue at* $t = 0$ *with a corresponding unit eigenvector* X_0*, then for all* t *near* 0 *there is a corresponding eigenvalue and unique (normalized) eigenvector that depend smoothly on* t.
Also, if the elements of $A(t)$ *are real analytic function of* t*, then so are the eigenvalue and eigenvector.*

REMARK [JAN. 1998]: The only text I know of that has a proof of this is [Lax]. The proof there is different.

PROOF:* Although we won't use it, the eigenvalue part is immediate from the warm-up exercise above applied to the characteristic polynomial. It is the eigenvector aspect that is takes a bit more work.

Given $A(0)X_0 = \lambda_0 X_0$ for some vector X_0 with $\|X_0\| = 1$, we want a function $\lambda(t)$ and a vector $X(t)$ that depend smoothly on t with the properties

$$A(t)X(t) = \lambda(t)X(t), \quad \langle X_0, X(t)\rangle = 1, \qquad \text{and} \qquad \lambda(0) = \lambda_0, \quad X(0) = X_0.$$

Here, $\langle X, Y\rangle$ is the standard inner product. Of course we could also have used a different normalization, such as $\|X(t)\|^2 = 1$.

SOME BACKGROUND ON THE IMPLICIT FUNCTION THEOREM.

If $H : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N$, say we want to solve the equations $H(Z,t) = 0$ for $Z = Z(t)$. These are N equations for the N unknowns $Z(t)$. Assume that $Z = Z_0$ is a solution at $t = 0$, so $H(Z_0, 0) = 0$. Expanding H in a Taylor series in the variable Z near $Z = Z_0$ we get

$$H(Z,t) = H(Z_0,t) + H_Z(Z_0,t)(Z - Z_0) + \cdots,$$

where H_Z is the derivative matrix and $\cdots$ represent higher order terms. If these higher order terms were missing then the solution of $H(Z,t) = 0$ would be simply

$$Z - Z_0 = -[H_Z(Z_0,t)]^{-1} H(Z_0,t),$$

that is,

$$Z = Z_0 - [H_Z(Z_0,t)]^{-1} H(Z_0,t).$$

This assumes that the first derivative matrix $H_Z(Z_0, 0)$ is invertible (since it is then invertible for all t near zero). The *implicit function theorem* says that this is still true even if there are higher order terms. The key assumption is that the first derivative matrix $H_Z(Z_0, 0)$ is invertible. Although we may think of the special case where $t \in \mathbb{R}$, this works without change if the parameter $t \in \mathbb{R}^k$ is a vector.

(CONTINUATION OF THE PROOF) We may assume that $\lambda_0 = 0$. Write our equations as

$$F(X,\lambda,t) := \begin{pmatrix} f(X,\lambda,t) \\ g(X,\lambda,t) \end{pmatrix} := \begin{pmatrix} A(t)X - \lambda X \\ \langle X_0, X\rangle - 1 \end{pmatrix},$$

where we have written $f(X,\lambda,t) := A(t)X - \lambda X$ and $g(X,\lambda,t) := \langle X_0, X\rangle - 1$. We wish to solve: $F(X,\lambda,t) = 0$ for both $X(t)$ and $\lambda(t)$ near $t = 0$. In the notation of the previous paragraph, $Z = (X,\lambda)$ and $H(Z,t) = F(X,\lambda,t)$. Thus the derivative matrix H_Z involves differentiation with respect to both X and λ.

The derivative with respect to the parameters X and λ is the partitioned matrix

$$F'(X,\lambda,t) = \begin{pmatrix} f_X & f_\lambda \\ g_X & g_\lambda \end{pmatrix} = \begin{pmatrix} A(t) - \lambda & -X \\ X_0^T & 0 \end{pmatrix}.$$

Here we used $\langle X_0, X\rangle = X_0^T X$, where X_0^T is the transpose of the column vector X_0. Thus at $t = 0$

$$F'(X_0, 0, 0) = \begin{pmatrix} A(0) & -X_0 \\ X_0^T & 0 \end{pmatrix}.$$

*This was worked out at the blackboard with Dennis DeTurck.

For the implicit function theorem we check that the matrix on the right is invertible. It is enough to show its kernel is zero. Thus, say $F'(X_0,0,0)W = 0$, where $W = \binom{V}{r}$. Then $A(0)V - X_0 r = 0$ and $\langle X_0, V\rangle = 0$. From the first equation we find

$$A(0)^2 V = rA(0)X_0 = 0.$$

By assumption, the eigenvalue $\lambda_0 = 0$ is simple. Thus the only solutions of $A(0)^2 V = 0$ are $V = (\text{const})X_0$. But then $\langle X_0, V\rangle = 0$ gives $V = 0$. Consequently also $r = 0$ so $W = 0$.

Since the derivative matrix $F'(X_0,0,0)$ is invertible, by the implicit function theorem the equation $F(X,\lambda,t) = 0$ has the desired smooth solution $X = X(t)$, $\lambda = \lambda(t)$ near $t = 0$. If F is real analytic in t near $t = 0$ then so is this solution .

The General Case

If the eigenvalue $\lambda(t)$ is not simple, then the situation is more complicated – except if $A(t)$ is self-adjoint and depends analytically on t. Then both the eigenvalue and corresponding eigenvectors are analytic functions of t (see [Ka] and [Re]). However, it is obviously false that the *largest* eigenvalue is an analytic function of t — as one sees from the simple example $\left(\begin{smallmatrix} t & 0 \\ 0 & 0 \end{smallmatrix}\right)$ for t near 0.

We conclude with three examples showing that if either the self-adjoint or analyticity assumptions are deleted, the eigenvalue and/or eigenvector may not depend smoothly on t.

EXAMPLE 1 At $t = 0$ the matrix $A(t) = \begin{pmatrix} 0 & 1 \\ t & 0 \end{pmatrix}$ has 0 as a double eigenvalue. Since the characteristic polynomial is $p(t) := \lambda^2 - t$, the eigenvalues are not smooth functions of t for t near 0.

EXAMPLE 2a This is a symmetric matrix depending smoothly (but not analytically) on t. Near $t = 0$ the eigenvectors are not even continuous functions of t. This is from Rellich's nice book [Re, page 41]. Let

$$B(t) = \begin{pmatrix} a(t) & 0 \\ 0 & -a(t) \end{pmatrix} \qquad \text{and} \qquad R(t) = \begin{pmatrix} \cos\varphi(t) & -\sin\varphi(t) \\ \sin\varphi(t) & \cos\varphi(t) \end{pmatrix},$$

where $a(0) = 0$. For $t \neq 0$ we let $a(t) := \exp(-1/t^2)$ and $\varphi(t) := 1/t$.

The desired symmetric matrix is $A(t) = R(t)B(t)R^{-1}(t)$. It is similar to $B(t)$, but the new basis determined by the orthogonal matrix $R(t)$ is spinning quickly near $t = 0$. We find

$$A(t) = a(t)\begin{pmatrix} \cos 2\varphi(t) & \sin 2\varphi(t) \\ \sin 2\varphi(t) & -\cos 2\varphi(t) \end{pmatrix},$$

Its eigenvalues are $\lambda_\pm = \pm a(t)$. For $t \neq 0$ the orthonormal eigenvectors are

$$V_+(t) = \begin{pmatrix} \cos\varphi(t) \\ \sin\varphi(t) \end{pmatrix} \qquad \text{and} \qquad V_-(t) = \begin{pmatrix} -\sin\varphi(t) \\ \cos\varphi(t) \end{pmatrix}.$$

Since $a(t)$ goes to 0 so quickly near $t = 0$, even though $\varphi(t) = 1/t$ the matrix $A(t)$ is a C^∞ function of t. However the eigenvectors keep spinning and don't even converge as $t \to 0$.

EXAMPLE 2b Another example of the same phenomenon. Let M_+ and M_- be symmetric 2×2 matrices with different orthonormal eigenevctors V_1, V_2, and W_1, W_2,respectively. For instance

$$M_+ = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad M_- = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix},$$

so we can let $V_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, $V_2 = (\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})$, and $W_1 = (1,0)$, $W_2(0,1)$. With $a(t) := \exp(-1/t^2)$ as above, let

$$A(t) = \begin{cases} a(t)M_+ & \text{for } t > 0 \\ a(t)M_- & \text{for } t < 0 \end{cases}$$

with $A(0) = 0$. This matrix $A(t)$ depends smoothly on t and has the eigenvectors V_1 and V_2, for $t > 0$, but W_1 and W_2 for $t < 0$. The eigenvectors are not continuous in a neighborhood of $t = 0$.

EXAMPLE 3 In the previous example the eigenvalues were still smooth functions, but this was lucky. There are symmetric matrices depending smoothly on a parameter t whose eigenvalues are not C^2 functions of t. Since the eigenvalues are roots of the characteristic polynomial, this is just the situation of a polynomial whose coefficients depend on a parameeter and asking how smoothly the roots depend on the parameter. One instance is $x^2 - f(t) = 0$, where $f(t) \geq 0$ is smooth. The key observation (certainly known by Rellich in [Re]) is the perhaps surprising fact that this $f(t)$ may not have a smooth square root. This has been rediscovered many times. One example is

$$f(t) = \sin^2(1/t)e^{-1/t} + e^{-2/t} \quad \text{for } t > 0 \qquad \text{while} \qquad f(t) = 0 \quad \text{for } t \leq 0.$$

For a recent discussion with additional details and references, see [AKML].

References

[AKML] Alekseevsky, D., Kriegl, A., Michor, P., Losik, M., "Choosing Roots of Polynomials Smoothly, II" *Israel Journal of Mathematics* **139** 2004, 183–188.

[Ka] Kato, Tosio, *Perturbation theory for linear operators*, Grundlehren **132**, 1976, Springer-Verlag, Berlin, New York.

[Kn] Knopp, K., *Theory of Functions*, Vol. **II**, 1947, Dover, New York (translated from the German).

[Lax] Lax, Peter D., *Linear Algebra*,Wiley-Interscience, 1996 ISBN: 0471111112.

[Re] Rellich, Franz, *Perturbation Theory of Eigenvalue Problems*, 1953 New York University Lecture Notes reprinted by Gordon and Breach, 1968.