isid/ms/2009/02
February 21, 2009
http://www.isid.ac.in/~statmath/eprints

The operator equation $\sum_{i=0}^{n} A^{n-i} X B^{i} = Y$

Rajendra Bhatia
Mitsuru Uchiyama

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

The operator equation $\sum_{i=0}^{n} A^{n-i}XB^{i} = Y$

Rajendra Bhatia and Mitsuru Uchiyama

May 18, 2009

Indian Statistical Institute, New Delhi, 110016, India,
rbh@isid.ac.in
Dep. of Math. Shimane Univ., Matsue, 690-8504, Japan,
uchiyama@riko.shimane-u.ac.jp

Abstract

The solution of the linear operator equation:
$A^{n-1}X + A^{n-2}XB + \cdots + AXB^{n-2} + XB^{n-1} = Y$ is given by $X = \frac{\sin \pi/n}{\pi} \int_0^\infty (t + A^n)^{-1} Y (t + B^n)^{-1} t^{1/n} dt$ if the specta of A and B are in the sector $\{z : z \neq 0, -\pi/n < \arg z < \pi/n\}$.

2000 MSC: Primary15A06, Secondary 15A24, 47A50
Keyword: Linear operator equation, Fréchet derivative, Sylvester equation.

The object of this note is the equation

$$A^{n-1}X + A^{n-2}XB + \cdots + AXB^{n-2} + XB^{n-1} = Y, \tag{1}$$

where A, B, Y are $m \times m$ complex matrices, or linear operators in a Banach space, and an X satisfying (1) is to be found. The special case

$$AX - XB = Y \tag{2}$$

is the much studied Sylvester equation, of great interest in operator theory, numerical analysis, and engineering.

There is a long tradition of finding different expressions for the solution of (2) in the form of operator integrals, some prominent examples of which occur in the works of E. Heinz, M. G. Krein, M. Rosenblum, and R. Bhatia, C. Davis and A. McIntosh. A comprehensive summary of these is contained in the survey article [2]. Our main results continue this tradition.

Consider first the special case $A = B$. Then our equation is

$$A^{n-1}X + A^{n-2}XA + \cdots + AXA^{n-2} + XA^{n-1} = Y. \tag{3}$$

Assume that the spectrum of A is in the sector

$$\{z : z \neq 0, -\pi/n < \arg z < \pi/n\}.$$

The simplest situation arises when A is a diagonal matrix with distinct diagonal entries $\lambda_1, \cdots, \lambda_m$. The equation (3) then can be written as

$$(\lambda_i^{n-1} + \lambda_i^{n-2}\lambda_j + \cdots + \lambda_j^{n-1})x_{ij} = y_{ij} \quad (1 \leqq i, j \leqq m),$$

or,

$$\tfrac{\lambda_i^n - \lambda_j^n}{\lambda_i - \lambda_j} x_{ij} = y_{ij} \quad (1 \leqq i, j \leqq m). \tag{4}$$

When $i = j$ the quotient on the left hand side is interpreted to mean $n\lambda_i^{n-1}$. The solution of this equation is

$$x_{ij} = \frac{\lambda_i - \lambda_j}{\lambda_i^n - \lambda_j^n} y_{ij}.$$

With the substitution $\alpha_i = \lambda_i^n$ this becomes

$$x_{ij} = \frac{\alpha_i^{1/n} - \alpha_j^{1/n}}{\alpha_i - \alpha_j} y_{ij}. \tag{5}$$

The hypothesis on the location of λ_i ensures that α_i does not lie on the negative real axis and has a well-defined n-th root. There is a well-known formula (p.116 of [1]) from complex analysis that says

$$\alpha^r = \frac{\sin r\pi}{\pi} \int_0^\infty \frac{\alpha}{t + \alpha} t^{r-1} dt \tag{6}$$

for $0 < r < 1$. Using this we see that

$$\begin{aligned} \frac{\alpha^r - \beta^r}{\alpha - \beta} &= \frac{1}{\alpha - \beta} \frac{\sin r\pi}{\pi} \int_0^\infty \left(\frac{\alpha}{t + \alpha} - \frac{\beta}{t + \beta}\right) t^{r-1} dt \\ &= \frac{\sin r\pi}{\pi} \int_0^\infty \frac{t^r}{(t + \alpha)(t + \beta)} dt. \end{aligned} \tag{7}$$

Thus the solution (5) can be expressed as

$$x_{ij} = \frac{\sin \pi/n}{\pi} \int_0^\infty \frac{y_{ij}}{(t + \lambda_i^n)(t + \lambda_j^n)} t^{1/n} dt.$$

This looks more complicated than (5) but there is an advantage. We can get rid of the coordinates and write this in the matrix form

$$X = \frac{\sin \pi/n}{\pi} \int_0^\infty (t + A^n)^{-1} Y (t + A^n)^{-1} t^{1/n} dt. \tag{8}$$

If A is not diagonal but is similar to a diagonal matrix, then we can easily see that the same formula gives a solution of (3). Matrices similar to diagonal are dense in the space of all matrices. So this formula gives the solution in the general case as well. See Sec.VII.2 of [1] where this sort of argument is used. Instead of giving the details of this we extract from the discussion above a connection with Fréchet derivatives which makes the argument work equally well for operators in a Banach space, and even for an abstract Banach algebra. For simplicity we continue to talk only of matrices.

Let f be a differentiable map on the space of matrices and let $Df(A)$ be its deivative at A. We refer the reader to Chapter X of [1] for basic facts that we use here. Let $\varphi(A) = A^n$. If no point of the spectrum of A is on the negative real axis, then φ has a well defined inverse map $\psi(A) = A^{1/n}$. The derivative of φ at A is a linear map on matrices whose action is given as

$$\begin{aligned} D\varphi(A)(X) &= \frac{d}{dt}\Big|_{t=0} \varphi(A + tX) \\ &= A^{n-1}X + A^{n-2}XA + \cdots + AXA^{n-2} + XA^{n-1}. \end{aligned}$$

This is of interest to us because the equation (3) can be written as

$$D\varphi(A)(X) = Y.$$

Its solution is found by inverting the map $D\varphi(A)$. It is a standard fact that $D\varphi(A)^{-1} = D\psi(\varphi(A))$. So the problem is reduced to finding a good expression for the derivative $D\psi$. One such expression can be derived using the operator version of the formula (6) which says

$$\psi(A) = \frac{\sin \pi/n}{\pi} \int_0^\infty A(t + A)^{-1} t^{1/n-1} dt. \tag{9}$$

The advantage is that the integrand is more amenable to calculations involving derivatives. The derivative of the function $h(A) = A^{-1}$ is given by $Dh(A)(Y) = -A^{-1}YA^{-1}$. Hence that of $f(A) := A(t + A)^{-1} = I - t(t + A)^{-1}$ is

$$Df(A)(Y) = t(t + A)^{-1}Y(t + A)^{-1}.$$

So from (9) we obtain

$$D\psi(A)(Y) = \frac{\sin \pi/n}{\pi} \int_0^\infty (t + A)^{-1}Y(t + A)^{-1} t^{1/n} dt.$$

The differentiation under the integral sign can be justified with the usual arguments involving the dominated convergence theorem. Replacing A by $\varphi(A) = A^n$ we get

$$X = D\psi(\varphi(A))(Y) = \frac{\sin \pi/n}{\pi} \int_0^\infty (t + A^n)^{-1}Y(t + A^n)^{-1} t^{1/n} dt. \tag{10}$$

As explained above this represents the solution of the equation (3).

The passage to (1) is affected by the much used Berberian trick. Given $m \times m$ matrices A, B and Y, let

$$\tilde{A} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}, \qquad \tilde{Y} = \begin{pmatrix} 0 & Y \\ 0 & 0 \end{pmatrix}.$$

Consider the equation (3) in $2m \times 2m$ matrices with $\tilde{A}$ and $\tilde{Y}$ in place of A and Y. The solution is

$$\tilde{X} = \frac{\sin \pi/n}{\pi} \int_0^\infty (t + \tilde{A}^n)^{-1} \tilde{Y} (t + \tilde{A}^n)^{-1} t^{1/n} dt.$$

It is easy to see, by reading off the $(1,2)$ entries of the block matrices involved that the solution of (1) is

$$X = \frac{\sin \pi/n}{\pi} \int_0^\infty (t + A^n)^{-1} Y (t + B^n)^{-1} t^{1/n} dt. \tag{11}$$

Specialising to the case $n = 2$ we obtain yet another formula for the solution of the Sylvester equation not included among the ones given in [2]: If the spectrum of A is contained in the open right half plane and that of B in the open left half plane, then the solution of (2) can be represented as

$$X = \frac{1}{\pi} \int_0^\infty (t + A^2)^{-1} Y (t + B^2)^{-1} t^{1/2} dt. \tag{12}$$

Besides the representation (6) there is another that can be exploited in this context. We have

$$\alpha^r = \frac{r}{\Gamma(1-r)} \int_0^\infty (1 - e^{-t\alpha}) t^{-(r+1)} dt \tag{13}$$

for $0 < r < 1$. Formulas for derivatives of the exponential of a matrix, under the names of Duhamel, Feynman, Karpus, Schwinger, are used extensively by physicists and numerical analysts. We refer the reader to [5] and [6] for excellent surveys. If $g(A) = e^A$, then

$$Dg(A)(X) = \int_0^1 e^{tA} X e^{(1-t)A} dt. \tag{14}$$

We leave it to the reader to fill in the details of the following calculation. (If help is needed go to [3] or [6].)

Let $f(A) = A^r$, $0 < r < 1$. Using the formulas (13) and (14) we can obtain

$$Df(A)(Y) = \frac{r}{\Gamma(1-r)} \int_0^\infty [\int_0^t e^{-sA} Y e^{-(t-s)A} ds] t^{-(r+1)} dt. \tag{15}$$

So the solution of the equation (3) can be represented also as

$$X = \frac{1/n}{\Gamma(1-1/n)} \int_0^\infty [\int_0^t e^{-sA^n} Y e^{-(t-s)A^n} ds] t^{-(1/n+1)} dt. \tag{16}$$

From this one can get another form of the solution of the equations (1) and (2).

Some other equations can be solved using this technique. We need a pair of functions φ and ψ where $\psi = \varphi^{-1}$, and a good formula for the Fréchet derivative of ψ. The exponential function, the Zhukovsky function and their inverses are examples of such pairs, as are some cubic polynomials with solutions given by the Cardano formula.

This method of solving equation has been used by others. For example, it occurs in a paper of F. Hiai and H. Kosaki [4] where they obtain the formula (11). Their analysis is restricted to positive definite matrices. Incidentally, the formula (10) shows that if A and Y are positive semidefinite matices, then so is X. This is an analogue of one of the important facts for the Lyapunov equation.

References

[1] R. Bhatia, Matrix Analysis, Springer, 1996.

[2] R. Bhatia, P.Rosenthal, How and why to solve the operator equation $AX - XB = Y$, Bull. London Math. Soc., 29(1997)1–21.

[3] R. Bhatia, K. B. Sinha, Variation of real powers of positive operators, Indiana Univ. Math. J., 43(1994)913–925.

[4] F. Hiai, H. Kosaki, Means for matrices and comparison of their norms, Indiana Univ. Math. J.,48(1999)899–936.

[5] I. Najfeld, T. Havel, Derivatives of the matrix exponential and their computation, Adv. Appl. Math., 16(1995)321–375.

[6] C. Van Loan, The sensitivity of the matrix exponential, SIAM J. Numer. Anal., 14(1977)971–981.