21-238, Math Studies Algebra 2, Department of Mathematical Sciences, Carnegie Mellon University
Spring 2012: Monday, Wednesday, Friday, 10:30 am, Doherty Hall 1211.
Luc TARTAR, University Professor of Mathematics, Wean Hall 6212, tartar@cmu.edu

12- Friday February 10, 2012.

Remark 12.1: $\mathbb{R}^n$ having its usual Euclidean structure, if $B \in L(\mathbb{R}^n, \mathbb{R}^n)$ is skew symmetric, one may consider that $B \in L(\mathbb{C}^n, \mathbb{C}^n)$ is skew Hermitian, so that $B^* = -B$ commutes with B, hence B is diagonal on an orthonormal basis of $\mathbb{C}^n$, and each eigenvalue λ_j satisfies $\overline{\lambda_j} = -\lambda_j$, i.e. λ_j is purely imaginary. The only possible real eigenvalue is 0, and corresponds to an eigen-space which is $ker(B)$, and one then restricts attention to $V = ker(B)^\perp$, and B maps V into V and is skew symmetric; if $\lambda = i\,a$ is an eigenvalue of B on V (with a non-zero $a \in \mathbb{R}$), then an eigenvector on $V_\mathbb{C}$ has the form $v + i\,w$ with non-zero $v, w \in V$, and $B\,(v + i\,w) = i\,a\,(v + i\,w)$ means $B\,v = -a\,w, B\,w = a\,v$, and since $(B\,v, v) = (B\,w, w) = 0$ because B is skew symmetric, one deduces that $(v, w) = 0$; then $a\,||w||^2 = -(B\,v, w) = (v, B\,w) = a\,||v||^2$ shows that $||v|| = ||w||$, and one rescales v and w to have norm 1; on the two-dimensional space spanned by the orthonormal basis $\{v, w\}$, one has $B = \begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix} = a\,J$, where J is the rotation of $-\frac{\pi}{2}$, which satisfies $J^2 = -I$, and one sees easily that the power series giving $e^{t\,B}$ is (on this particular two-dimensional space) $e^{t\,B} = \cos(a\,t)\,I + \sin(a\,t)\,J = \begin{pmatrix} \cos(a\,t) & \sin(a\,t) \\ -\sin(a\,t) & \cos(a\,t) \end{pmatrix}$, i.e. a rotation of $-a\,t$, which is an element of $S\mathbb{O}(2)$. One deduces that the dimension of $ker(B)^\perp$ is even, and that the matrix for $e^{t\,B}$ has 1s in the diagonal for the basis vectors in $ker(B)$, and then a few 2×2 diagonal blocks which are rotations, and this form implies that $e^{t\,B} \in S\mathbb{O}(n)$.

Remark 12.2: Conversely, if $P \in S\mathbb{O}(n)$, one may consider that $P \in L(\mathbb{C}^n, \mathbb{C}^n)$ is unitary, so that $P^* = P^{-1}$ commutes with P, hence P is diagonal on an orthonormal basis of $\mathbb{C}^n$, and each eigenvalue λ_j satisfies $\overline{\lambda_j} = \frac{1}{\lambda_j}$, i.e. $|\lambda_j| = 1$. The only possible real eigenvalues are $+1$ and -1, and -1 must have an even multiplicity, so that one may consider 2×2 diagonal blocks $-I$, which means rotations of π; if $\lambda = \cos\theta + i\,\sin\theta$ is an eigenvalue of P with $\theta \neq k\,\pi$, then an eigenvector in $\mathbb{C}^n$ has the form $v + i\,w$ with non-zero $v, w \in \mathbb{R}^n$, and $P\,(v + i\,w) = (\cos\theta + i\,\sin\theta)\,(v + i\,w)$ means $P\,v = \cos\theta\,v - \sin\theta\,w, P\,w = \sin\theta\,v + \cos\theta\,w$, so that $P\,(v - i\,w) = (\cos\theta - i\,\sin\theta)\,(v - i\,w)$, hence $v + i\,w$ and $v - i\,w$ must be orthogonal, which implies $||v||^2 = ||w||^2$ and $(v, w) = 0$, and one rescales v and w to have norm 1; on the two-dimensional space spanned by the orthonormal basis $\{v, w\}$, one has $P = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = R(\theta)$, a rotation of θ. The construction of Remark 12.1 then permits to find a skew symmetric B such that $P = e^B$. Of course, if $P \neq I$ there is more than one solution, since $e^{2k\,\pi\,J} = I$ for all $k \in \mathbb{Z}$.

Remark 12.3: Of course, the definition of e^A is valid for $A \in L(\mathbb{C}^n, \mathbb{C}^n)$.

Since a property of the exponential on $\mathbb{C}$ is that $e^a e^b = e^{a+b}$, it is important to observe that if $A, B \in L(\mathbb{C}^n, \mathbb{C}^n)$ commute, then one has $e^A e^B = e^B e^A = e^{A+B}$, but the situation is different if A and B do not commute, and one has $e^{s\,A} e^{t\,B} - e^{t\,B} e^{s\,A} = s\,t\,[A, B] + o(s^2 + t^2)$ for $|s|, |t|$ small, where $[A, B]$ denotes the *commutator* $A\,B - B\,A$. Indeed, $e^{s\,A} = I + s\,A + \frac{s^2}{2}\,A^2 + o(s^2)$, and $e^{t\,B} = I + t\,B + \frac{t^2}{2}\,B^2 + o(t^2)$, so that $e^{s\,A} e^{t\,B} = I + s\,A + t\,B + \frac{s^2}{2}\,A^2 + s\,t\,A\,B + \frac{t^2}{2}\,B^2 + o(s^2 + t^2)$, and $e^{t\,B} e^{s\,A} = I + s\,A + t\,B + \frac{s^2}{2}\,A^2 + s\,t\,B\,A + \frac{t^2}{2}\,B^2 + o(s^2 + t^2)$, hence the first term in $e^{s\,A} e^{t\,B} - e^{t\,B} e^{s\,A}$ is $s\,t\,(A\,B - B\,A)$.

In the case where $A, B \in L(\mathbb{C}^n, \mathbb{C}^n)$ commute, then for every polynomials $P, Q \in \mathbb{C}[x]$, $P(A)$ and $Q(B)$ commute (and it is true if $A, B \in L(\mathbb{R}^n, \mathbb{R}^n)$ commute, and $P, Q \in \mathbb{R}[x]$, of course), so that since e^A is the limit of $P_k(A)$ when k tends to ∞, with $P_k = 1 + x + \frac{x^2}{2!} + \ldots + \frac{x^k}{k!}$, and e^B is the limit of $P_k(B)$, the equality $P_k(A)\,P_k(B) = P_k(B)\,P_k(A)$ implies $e^A e^B = e^B e^A$ by letting k tend to ∞.

Remark 12.4: If one has a few evaluations of polynomials (or rational functions, or smooth enough functions) of an $n \times n$ matrix A with entries in E, it is useful to understand the general structure of $P(A)$ for all the polynomial $P \in E[x]$. In what follows, the field E is arbitrary, but the computations are done in any field extension F of E where the characteristic polynomial P_{char} of A (defined by $P_{char}(\lambda) = det(A - \lambda\,I)$) splits.

With V isomorphic to E^n, let us begin by the case where $A \in L(V,V)$ is diagonalizable, so that there is a basis $e_1,\ldots,e_n$ of V such that $A\,e_i = \lambda_i e_i$ for $i=1,\ldots,n$, hence $P(A)\,e_i = P(\lambda_i)\,e_i$ for $i=1,\ldots,n$. Let $e^1,\ldots,e^n$ be the dual basis of V^*, so that if $x \in V$, one has $x = \sum_i x^i e_i$, with $x^i = e^i(x)$ for $i=1,\ldots,n$, so that $P(A)\,x = \sum_i x^i P(A)\,e_i = \sum_i e^i(x)\,P(\lambda_i)\,e_i$, hence $P(A) = \sum_i P(\lambda_i)\,e_i \otimes e^i$. If $\lambda_1,\ldots,\lambda_r$ are the distinct eigenvalues, one has $P(A) = \sum_{i=1}^r P(\lambda_i)\,Z_i$ for some $Z_1,\ldots,Z_r \in L(V,V)$, but although the columns of Z_i are eigenvectors for the eigenvalue λ_i, one only needs to know the distinct eigenvalues, and then $Z_i = \pi_i(A)$ for $i=1,\ldots,r$, where one uses the interpolation polynomials (of degree $\le r-1$), defined by $\pi_i(\lambda_j) = \delta_{i,j}$ for $i,j=1,\ldots,r$.

If $f \in E(x)$ has no λ_i as pole, i.e. $f = \frac{P}{Q}$ with $Q(\lambda_i) \ne 0$ for $i=1,\ldots,r$, then $f(A) = P(A)\,\big(Q(A)\big)^{-1}$ is given by the formula $f(A) = \sum_{i=1}^r f(\lambda_i)\,Z_i$: indeed, let $R \in E[x]$ be any interpolation polynomial satisfying $R(\lambda_i) = \big(Q(\lambda_i)\big)^{-1}$ for $i=1,\ldots,r$, so that $R(A)\,Q(A) = \sum_{i=1}^r R(\lambda_i)\,Q(\lambda_i)\,Z_i = \sum_{i=1}^r Z_i = I$, hence $R(A) = \big(Q(A)\big)^{-1}$, and then $P(A)\,\big(Q(A)\big)^{-1} = P(A)\,R(A) = \sum_{i=1}^r P(\lambda_i)\,R(\lambda_i)\,Z_i = \sum_{i=1}^r f(\lambda_i)\,Z_i$.

Remark 12.5: If A is not diagonalizable, its minimum polynomial is $(x-\lambda_1)^{k_1}\cdots(x-\lambda_r)^{k_r}$, and for $i=1,\ldots,r$, $k_i \ge 1$ is the largest size of a Jordan block for the eigenvalue λ_i (and at least one $k_i \ge 2$). A Jordan block has the form $\lambda\,I + K$ with $K^{k-1} \ne 0, K^k = 0$, and one uses the binomial formula $(\lambda\,I+K)^m = \sum_{j=0}^m \binom{m}{j}\,\lambda^{m-j} K^j$ for all $m \ge 1$: if $P = \sum_m p_m x^m \in E[x]$, then $P(\lambda\,I+K) = \sum_{j=0}^{k-1}\big(\sum_m p_m \binom{m}{j}\,\lambda^{m-j}\big)\,K^j$, and it is usual to write $\sum_m p_m \binom{m}{j}\,\lambda^{m-j}$ as $\frac{1}{j!}\,P^{(j)}(\lambda)$, although if E has finite characteristic one should use the first form since dividing by $j!$ might have no meaning in E,[1] and with this understanding of the notation, one then has shown the existence of matrices $Z_{i,j}$ for $i=1,\ldots,r$ and $0 \le j \le k_i-1$ such that $P(A) = \sum_{i,j} \frac{1}{j!}\,P^{(j)}(\lambda_i)\,Z_{i,j}$ for all $P \in E[x]$. Of course, once the distinct eigenvalues $\lambda_1,\ldots,\lambda_r$ are known, one may compute the matrices $Z_{i,j}$ as $P_{i,j}(A)$ for a particular Hermite interpolation polynomial, and one may actually do such a computation for $j=0,\ldots,a_i-1$ where a_i is the algebraic multiplicity of λ_i, and one will find $Z_{i,j} = 0$ for $j > k_i-1$.

Remark 12.6: If $E = \mathbb{R}$ or $E = \mathbb{C}$, one may extend the preceding results to $f(A)$ in the case where f can be approximated uniformly in a neighbourhood of the eigenvalues of A (as well as some of its derivatives in the case of Jordan blocks) by a sequence of polynomials.

For example, it is the case for e^A, so that if A is diagonalizable one has $e^{t\,A} = \sum_i e^{t\,\lambda_i} Z_i$, and in the case of Jordan blocks one finds terms in $e^{t\,\lambda_i}$ multiplied by polynomials in t. This permits to improve the bound $e^{|t|\,||A||}$ for the norm of $e^{t\,A}$: if $\Re(\lambda_i) < -\alpha < 0$ for all the eigenvalues of A, then for any polynomial Q, one sees that $|e^{t\,\lambda_i} Q(t)|$ tends to 0 faster than $e^{-\alpha\,t}$ as t tends to $+\infty$, and one deduces that $||e^{t\,A}||$ tends to 0 faster than $e^{-\alpha\,t}$ as t tends to $+\infty$.[2]

Remark 12.7: Since $\frac{d(e^{t\,A})}{dt} = A\,e^{t\,A} = e^{t\,A} A$, one deduces that $X(t) = e^{t\,A^T} M\,e^{t\,A}$ satisfies $\frac{dX}{dt} = A^T X + X\,A$, and $X(0) = M$; if $\Re(\lambda_i) < -\alpha < 0$ for all the eigenvalues of A, which are also the eigenvalues of A^T, one deduces that $\int_0^{+\infty} ||X(t)||\,dt < \infty$, so that one can define $Y = \int_0^{+\infty} X(t)\,dt \in L(\mathbb{R}^n, \mathbb{R}^n)$, and then $-M = \int_0^{+\infty} \frac{dX}{dt}\,dt = A^T Y + Y\,A$; choosing $M = I$,[3] one finds that Y is symmetric positive definite, and satisfies $A^T Y + Y\,A = -I$.

This permits to prove the *asymptotic stability* of the stationary solution 0 of $\frac{dx}{dt} = F(x)$, where F is a C^1 mapping from R^n to itself with $F(0) = 0$ and $DF(0) = A$, by using the Lyapunov function $\psi(x) = (Y\,x, x)$:[4] since $\frac{d}{dt}\big[\psi\big(x(t)\big)\big] = \big(Y\,\frac{dx}{dt}, x\big) + \big(Y\,x, \frac{dx}{dt}\big) = \big(Y\,F(x), x\big) + \big(Y\,x, F(x)\big) = (Y\,A\,x, x) + (Y\,x, A\,x) + o(||x||^2) = -||x||^2 + o(||x||^2)$, so that any solution of $\frac{dx}{dt} = F(x)$ with $||x(0)||$ small enough has $x(t) \to 0$ (exponentially fast) as $t \to +\infty$.

---

[1] In other words, in $\sum_m p_m \binom{m}{j}\,\lambda^{m-j}$ the division by $j!$ is done on an integer (and the result $\binom{m}{j}$ is an integer) and not in E.

[2] If $\Re(\lambda_i) > \beta > 0$ for all the eigenvalues of A, then $||e^{t\,A}||$ tends to 0 when t tends to $-\infty$, faster than $e^{-\beta\,|t|}$.

[3] Choosing for M any symmetric positive definite matrix gives a symmetric positive definite Y, solution of $A^T Y + Y\,A = -M$.

[4] Aleksandr Mikhailovich LYAPUNOV, Russian mathematician, 1857–1918. He worked in Kharkov, St Petersburg, Russia, and Odessa (then in Russia, now in Ukraine).