

□

PHY 564

Advanced Accelerator Physics

Lecture 6

Matrices and Matrix function

Sylvester formulae

Vladimir N. Litvinenko

CENTER for ACCELERATOR SCIENCE AND EDUCATION
Department of Physics & Astronomy, Stony Brook University

Notations

- Accelerator curvilinear coordinate system

$$\vec{r} = \vec{r}_o(s) + x \cdot \vec{n}(s) + y \cdot \vec{b}(s)$$

$$\frac{d\vec{\tau}}{ds} = -K(s) \cdot \vec{n}; \quad \frac{d\vec{n}}{ds} = K(s) \cdot \vec{\tau} - \kappa(s) \cdot \vec{b}; \quad \frac{d\vec{b}}{ds} = \kappa(s) \cdot \vec{n};$$

$$q^1 = x; \quad q^2 = s, \quad q^3 = y.$$

$$P_1 = P_x; P_2 = (1 + Kx)P_s + \kappa(P_x y - P_y x); \quad P_3 = P_y;$$

Most general Hamiltonian of linearized motion in accelerator (Lecture 4)

$$\tilde{h} = \frac{P_1^2 + P_3^2}{2p_o} + F \frac{x^2}{2} + Nxy + G \frac{y^2}{2} + L(xP_3 - yP_1) + \frac{\delta^2}{2p_o} \cdot \frac{m^2 c^2}{p_o^2} + U \frac{\tau^2}{2} + g_x x \delta + g_y y \delta + F_x x \tau + F_y y \tau \quad (140)$$

$$\frac{F}{p_o} = \left[-K \cdot \frac{e}{p_o c} B_y - \frac{e}{p_o c} \frac{\partial B_y}{\partial x} + \left(\frac{e B_s}{2 p_o c} \right)^2 \right] - \frac{e}{p_o v_o} \frac{\partial E_x}{\partial x} - 2K \frac{e E_x}{p_o v_o} + \left(\frac{m e E_x}{p_o^2} \right)^2;$$

$$\frac{G}{p_o} = \left[\frac{e}{p_o c} \frac{\partial B_x}{\partial y} + \left(\frac{e B_s}{2 p_o c} \right)^2 \right] - \frac{e}{p_o v_o} \frac{\partial E_y}{\partial y} + \left(\frac{m e E_z}{p_o^2} \right)^2;$$

$$\frac{2N}{p_o} = \left[\frac{e}{p_o c} \frac{\partial B_x}{\partial x} - \frac{e}{p_o c} \frac{\partial B_y}{\partial y} \right] - K \cdot \frac{e}{p_o c} B_x - \frac{e}{p_o v_o} \left(\frac{\partial E_x}{\partial y} + \frac{\partial E_y}{\partial x} \right) - 2K \frac{e E_y}{p_o v_o} + \left(\frac{m e E_z}{p_o^2} \right) \left(\frac{m e E_x}{p_o^2} \right)$$

$$L = \kappa + \frac{e}{2 p_o c} B_s; \quad \frac{U}{p_o} = \frac{e}{p c^2} \frac{\partial E_s}{\partial t}; \quad g_x = \frac{(m c)^2 \cdot e E_x}{p_o^3} - K \frac{c}{v_o}; \quad g_y = \frac{(m c)^2 \cdot e E_y}{p_o^3}; \quad (141)$$

$$F_x = \frac{e}{c} \frac{\partial B_y}{\partial ct} + \frac{e}{v_o} \frac{\partial E_x}{\partial ct}; \quad F_y = -\frac{e}{c} \frac{\partial B_x}{\partial ct} + \frac{e}{v_o} \frac{\partial E_y}{\partial ct}.$$

Static magnetic “lattice”

$$\vec{E} = 0, \quad \frac{\partial \vec{B}}{\partial t} = 0.$$

$$\tilde{h} = \frac{P_1^2 + P_3^2}{2p_o} + F \frac{x^2}{2} + Nxy + G \frac{y^2}{2} + L(xP_3 - yP_1) + \frac{\delta^2}{2p_o} \cdot \frac{m^2 c^2}{p_o^2} + g_x x \delta; \quad (188)$$

with

$$\begin{aligned} \frac{F}{p_o} &= \left[\left(\frac{e}{p_o c} B_y \right)^2 - \frac{e}{p_o c} \frac{\partial B_y}{\partial x} + \left(\frac{e B_s}{2 p_o c} \right)^2 \right]; \quad \frac{G}{p_o} = \left[\frac{e}{p_o c} \frac{\partial B_x}{\partial y} + \left(\frac{e B_s}{2 p_o c} \right)^2 \right] \\ \frac{N}{p_o} &= \left[\frac{e}{p_o c} \frac{\partial B_x}{\partial x} \right]; \quad L = \kappa + \frac{e}{2 p_o c} B_s; \quad g_x = -K \frac{c}{v_o}; \\ \frac{\partial B_y}{\partial x} &= \frac{\partial B_x}{\partial y}; \quad \frac{\partial B_x}{\partial x} = -\frac{\partial B_y}{\partial y}; \end{aligned} \quad (189)$$

Since p_o is constant in magnetic field, we also can use (134) and rewrite Hamiltonian of the linearized motion as

$$\tilde{h}_n = \frac{\pi_1^2 + \pi_3^2}{2} + f \frac{x^2}{2} + n \cdot xy + g \frac{y^2}{2} + L(x\pi_3 - y\pi_1) + \frac{\pi_o^2}{2} \cdot \frac{m^2 c^2}{p_o^2} + g_x x \pi_o; \quad (188-n)$$

with

$$f = \frac{F}{p_o}; \quad n = \frac{N}{p_o}; \quad g = \frac{G}{p_o}; \quad ; \quad (189-n)$$

Focusing/defocusing in transverse direction can come from

(a) a dipole field B_y or in other words, from the curvature of trajectory. Note that it is always focusing.

(b) from quadrupole field $\frac{\partial B_y}{\partial x} = \frac{\partial B_x}{\partial y}$. Note that quadrupole is focusing in one direction and defocusing in the other.

(c) from solenoidal field, B_s . Note that it is always focusing.

The other terms, are responsible for x-y coupling:

(a) the transverse motion (x & y): solenoidal field, B_s and torsion κ as well as SQ-quadrupole $\frac{\partial B_x}{\partial x}$.

(b) or transverse and longitudinal motion: $g_x x \delta$ - it is responsible of dependence of the time of flight on transverse coordinate.

Finally, there is $\frac{\delta^2}{2p_o} \cdot \frac{m^2 c^2}{p_o^2}$ term which corresponds to the velocity dependence on the particle energy. It is frequently neglected at very high energies when $m^2 c^2 / p_o^2 \approx \gamma^{-2} \lll 1$. But it should be kept for many accelerators, including RHIC.

We should not forget one of the most common element in any accelerator lattice – an empty space, call drift.

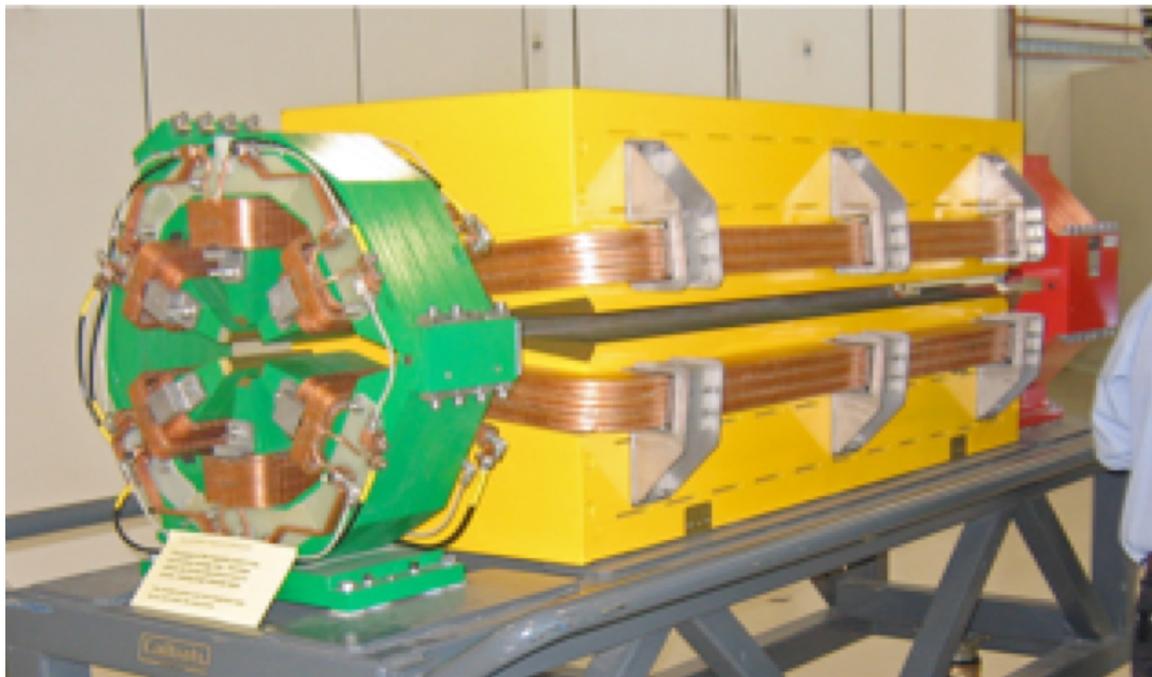
In standard accelerator physics book you will find solution (matrices) for various elements of the lattice: drift, bending magnet (with or with field gradient), quadrupole. Then, piecewise, you can see introduction of solenoids, SQ-quadrupoles.... Instead of solving dozen of second, fourth and sixth order differential equations... we will use matrix function approach to find all solutions at once.

Matrices and matrix functions. As a practical matter, when somebody wants to build an accelerator, she or he should use some approximations. One of VERY popular design approximation is called “an element (usually a magnet)” with nearly constant parameters. Then our Hamiltonian is s -independent on a part of the trajectory.

$$\mathbf{H} = \mathbf{H}_i(s); \quad \mathbf{H}_i(s) = \text{const}; \left\{ s_{i-1} < s < s_i \right\}, \quad i = 1, 2, 3, \dots, m;$$

$$M(s_o, s_m) = \prod_{i=1}^m \mathbf{M}_i; \quad \mathbf{M}_i(s_{i-1}, s_i) = \exp\left(\mathbf{S} \cdot \mathbf{H}_i(s_i - s_{i-1})\right);$$
(6-1)

e.g. we just need to learn how to calculate $\exp\left(\mathbf{S} \cdot \mathbf{H}_i(s_i - s_{i-1})\right)$. Finally, she or he then should try to build such elements. They are never ideal but can be relatively close to the ideal boxes...



Typical elements of accelerators are dipoles and quadrupoles (or their combination), sextupoles and octupoles (they are nonlinear), solenoids, wigglers.... Let's start from a linearized Hamiltonian (143) magnetic DC elements – this is typical accelerator beamline.

Order of matrix multiplication. Matrices, in general case, may not commute: .

$$\mathbf{M}_1\mathbf{M}_2 \neq \mathbf{M}_2\mathbf{M}_1$$

and order of multiplication is of critical importance for correct calculation of transport matrices! Let's consider sequence of two matrix transports:

$$X(s_1) = \mathbf{M}(s_o, s_1)X(s_o); \quad X(s_2) = \mathbf{M}(s_1, s_2)X(s_1)$$

tan we can conclude that

$$X(s_2) = \mathbf{M}(s_1, s_2)X(s_1) = \mathbf{M}(s_1, s_2)\mathbf{M}(s_o, s_1)X(s_o);$$

$$\mathbf{M}(s_o, s_2) = \mathbf{M}(s_1, s_2)\mathbf{M}(s_o, s_1)$$

which specifies order of the multiplication: first “element: is on the right, second is on the left, third is further left, etc... It means that we must rewrite (187) as orderly product:

$$\mathbf{H} = \mathbf{H}_i(s); \quad \mathbf{H}_i(s) = \text{const}; \{s_{i-1} < s < s_i\}; \quad \frac{d\mathbf{M}}{ds} = \mathbf{SH} \cdot \mathbf{M}; \quad \mathbf{D} = \mathbf{SH} \tag{187-orderd}$$

$$\mathbf{M}(s_o, s) = \prod_{i=1}^N \mathbf{M}_i = \mathbf{M}_N \cdot \mathbf{M}_{N-1} \cdots \mathbf{M}_2 \cdot \mathbf{M}_1; \quad \mathbf{M}_i = \exp(\mathbf{SH}_i(s_i - s_{i-1}))$$

Calculating matrices. Next, we focus on the question of how matrices are calculated. We already discussed general idea that they can be integrates piece-wise wherein the coefficients in the Hamiltonian expansion do not change significantly. In practice, accelerators are build from elements, which, to a certain extent, offers such conditions.

Since method of calculating 6x6 or 4x4 (or even some 2x2) matrices is very similar to that for 2nx2n, where n is arbitrary integer. Hence, initially we will explore a general way of calculating matrices, and then consider few examples. When the matrices **D** are piece-wise constant and the **D** from different elements do not commute, we can write

$$\mathbf{M}(s_o|s) = \prod_i \mathbf{M}(s_{i-1}|s_i); \mathbf{M}(s_{i-1}|s) = \prod_{elements} \exp[\mathbf{D}_i(s - s_{i-1})] \quad (193)$$

The definition of the matrix exponent is very simple

$$\exp[\mathbf{A}] = \mathbf{I} + \sum_{k=1}^{\infty} \frac{\mathbf{A}^k}{k!}; \quad \exp[\mathbf{D} \cdot s] = \mathbf{I} + \sum_{k=1}^{\infty} \frac{\mathbf{D}^k s^k}{k!} \quad (194)$$

According to the general theorem of Hamilton-Kelly, the matrix is a root of its characteristic equation:

$$d(\lambda) = \det[\mathbf{D} - \lambda \mathbf{I}]; \quad d(\lambda_k) = 0 \quad (195)$$

$$d(\mathbf{D}) \equiv 0 \quad (196)$$

i.e., a root of a polynomial of order $\leq 2n$. There is a theorem in theory of polynomials (rather easy to prove) that any polynomial $p_1(x)$ of power n can be expressed via any polynomial $p_2(x)$ of power $m < n$ as

$$p_1(x) = p_2(x) \cdot d(x) + r(x)$$

where r(x) is a polynomial of power less than m. Accordingly, series (194) can be always truncated to

$$\exp[\mathbf{D}] = \mathbf{I} + \sum_{k=1}^{2n-1} c_k \mathbf{D}^k, \quad (197)$$

with the remaining daunting task of finding coefficients c_k !

There are two ways of doing this; one is a general, and the other is case specific, but an easy one. Starting from a specific case when the matrix \mathbf{D} is nilpotent ($m < 2n+1$), i.e.,

$$\mathbf{D}^m = 0.$$

In this case, $\mathbf{D}^{m+j} = 0$ the truncation is trivial:

$$\exp[D] = I + \sum_{k=1}^{m-1} \frac{D^k}{k!}. \quad (198)$$

We lucky to have such a beautiful case in hand – a drift, where all fields are zero and $\mathbf{K}=0$ and $\kappa=0$:

$$\tilde{h} = \frac{\pi_1^2 + \pi_3^2}{2} + \frac{\pi_\delta^2}{2} \cdot \frac{m^2 c^2}{p_o^2};$$

$$\mathbf{D} = \begin{bmatrix} D_\perp & 0 & 0 \\ 0 & D_\perp & 0 \\ 0 & 0 & D_\parallel \end{bmatrix}; D_\perp = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}; D_\parallel = \begin{bmatrix} 0 & \frac{m^2 c^2}{p_o^2} \\ 0 & 0 \end{bmatrix}; \quad (199)$$

where it is easy to check: $\mathbf{D}^2 = 0$. Hence, the 6x6 matrix of drift with length l will be

$$\mathbf{M}_{drift} = \exp[\mathbf{D} \cdot l] = \mathbf{I} + \sum_{k=1}^{\infty} \frac{\mathbf{D}^k l^k}{k!} = \mathbf{I} + \mathbf{D} \cdot l = \begin{bmatrix} M_\perp & 0 & 0 \\ 0 & M_\perp & 0 \\ 0 & 0 & M_\parallel \end{bmatrix}; \quad (200)$$

$$M_\perp = \begin{bmatrix} 1 & l \\ 0 & 1 \end{bmatrix}; M_\parallel = \begin{bmatrix} 1 & l / (\beta_o \gamma_o)^2 \\ 0 & 1 \end{bmatrix};$$

In 1883, English mathematician James Joseph Sylvester derived his famous formula for function of matrices which can be diagonalized. A bit later another British mathematician, Arthur Buchheim, extended it for a general case of matrices reducible to Jordan form, e.g. those with some eigen values having multiplicity >1 .

Modern text related to matrix functions:

N.J. Higham, Functions of Matrices: Theory and Computation

<https://www.maths.manchester.ac.uk/~higham/fm/>

Classical text: F. R. Gantmacher, Theory of Matrices

We will start from simplest case when matrix can be diagonalized and finish with full blown general case...

XXXIX. *On the Equation to the Secular Inequalities in the Planetary Theory.* By J. J. SYLVESTER, F.R.S.*

A VERY long time ago I gave, in this Magazine, a proof of the reality of the roots in the above equation, in which I employed a certain property of the square of a symmetrical matrix which was left without demonstration. I will now state a more general theorem concerning the *product* of any two matrices of which that theorem is a particular case. In what follows it is of course to be understood that the product of two matrices means the matrix corresponding to the combination of two substitutions which those matrices represent.

It will be convenient to introduce here a notion (which plays a conspicuous part in my new theory of multiple algebra), viz. that of the *latent roots* of a matrix—latent in a somewhat similar sense as vapour may be said to be latent in water or smoke in a tobacco-leaf. If from each term in the diagonal of a given matrix, λ be subtracted, the determinant to the matrix so modified will be a rational integer function of λ ; the roots of that function are the latent roots of the matrix; and there results the important theorem that the latent roots of

On the Theory of Matrices. By Mr. A. BUCHHEIM, M.A.

[Read Nov. 13th, 1884.]

INTRODUCTION.

The methods used in the following paper are essentially, though not historically, an extension of Hamilton's theory of the linear function of a vector, and the simplest way to connect Grassmann's methods with the theory created by Cayley and Sylvester will be to connect them both with Hamilton's investigations.

It is, or ought to be, well known that the linear and vector function of a vector is simply the matrix of the third order. This is obvious from the definition: for, if ρ is any vector, $\sigma = \phi\rho$ is a vector whose constituents are linear functions of ρ 's constituents; that is, if

$$\rho = xi + yj + zk, \quad \sigma = x'i + y'j + z'k,$$

we must have the three equations

$$x' = ax + a'y + a''z,$$

$$y' = bx + b'y + b''z,$$

$$z' = cx + c'y + c''z,$$

that is, $(x'y'z') = \begin{pmatrix} a & a' & a'' \\ b & b' & b'' \\ c & c' & c'' \end{pmatrix} (x, y, z) \dots\dots\dots(A).$

That is to say, it is the same thing whether we say that $\sigma = \phi\rho$, or

The general evaluation of the matrix exponent in (193) is straightforward using the eigen values of the D-matrix:

$$\det[\mathbf{D} - \lambda \cdot \mathbf{I}] = \det[\mathbf{SH} - \lambda \cdot \mathbf{I}] = 0 \quad (201)$$

When the eigen values are all different (2n numerically different eigen values, $\lambda_i = \lambda_j \Rightarrow i = j$, no degeneration, i.e., D can be diagonalized),

$$\mathbf{D} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}; \mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & & 0 \\ 0 & \lambda_2 & & 0 \\ & & \dots & 0 \\ 0 & 0 & 0 & \lambda_{2n} \end{pmatrix}; \quad (202)$$

we can use Sylvester's formula that is correct for any analytical f(D), http://en.wikipedia.org/wiki/Sylvester's_formula for evaluating (193):

$$\exp[\mathbf{D}s] = \sum_{k=1}^{2n} e^{\lambda_k s} \prod_{j \neq k} \frac{\mathbf{D} - \lambda_j \mathbf{I}}{\lambda_k - \lambda_j} \quad (203)$$

Let's prove this very useful formula. First, let consider a polynomial function

$$f_N(x) = \sum_{k=0}^N a_k x^k \quad (204)$$

and apply it to (202)

$$f_N(D) = \sum_{k=0}^N a_k D^k = \sum_{k=0}^N a_k (\mathbf{U}\Lambda\mathbf{U}^{-1})^k = \mathbf{U} \left\{ \sum_{k=0}^N a_k \Lambda^k \right\} \mathbf{U}^{-1} = \mathbf{U} \cdot f_N(\Lambda) \cdot \mathbf{U}^{-1} \quad (205)$$

$$f_N(\Lambda) \equiv \begin{bmatrix} \dots & 0 & 0 \\ 0 & f_N(\lambda_i) & 0 \\ 0 & 0 & \dots \end{bmatrix}$$

e.g. function of diagonalizable matrix is a similarity transformation of the diagonal matrix with function of it eigen values. Goin to infinite series, we get

$$\exp(D) = \sum_{k=0}^{\infty} \frac{D^k}{k!} = \mathbf{U} \left\{ \sum_{k=0}^{\infty} \frac{1}{k!} (\Lambda)^k \right\} \mathbf{U}^{-1} = \mathbf{U} \exp(\Lambda) \mathbf{U}^{-1} \quad (206)$$

$$\exp(\Lambda) \equiv \begin{bmatrix} \dots & 0 & 0 \\ 0 & e^{\lambda_i} & 0 \\ 0 & 0 & \dots \end{bmatrix}$$

Now we start using our refresher on linear algebra. Each eigen value of diagonalizable matrix corresponds to an eigen vector

$$D \cdot Y_i = \lambda_i Y_i. \quad (207)$$

(existence comes from statement that $(D - \lambda_i I)Y_i = 0$ has non-trivial solution if $\det(D - \lambda_i I) = 0$). The set of eigen vectors is a full set of vectors, e.g. any arbitrary vector can be expanded as

$$X = \sum_i \alpha_i Y_i. \quad (208)$$

This eigen vectors are columns of the matrix used for similarity transform to its diagonal form:

$$\mathbf{U} = [Y_1, Y_2, \dots, Y_{2n}] \quad (209)$$

which is trivial to prove using (208) and (209) and comparing it with (202)

$$\begin{aligned} \mathbf{D}\mathbf{U} &= \mathbf{U}\Lambda; \quad \rightarrow \mathbf{D} = \mathbf{U}\Lambda\mathbf{U}^{-1} \\ \mathbf{U}\Lambda &\equiv [\lambda_1 Y_1, \lambda_2 Y_2, \dots, \lambda_{2n} Y_{2n}] \end{aligned} \quad (210)$$

Now, let's build a unit projection operator on Y_k :

$$P_k = \prod_{i \neq k} \frac{M - \lambda_i I}{\lambda_k - \lambda_i} \quad (211)$$

It is easy to show that

$$P_k Y_k = Y_k; \quad P_k Y_{i \neq k} = 0; \quad (212)$$

First, each of the elements of the product (211) is unit on Y_k

$$\frac{M - \lambda_i I}{\lambda_k - \lambda_i} \cdot Y_k = \frac{\lambda_k - \lambda_i}{\lambda_k - \lambda_i} Y_k = Y_k; \quad i \neq k \quad (213)$$

while it is a zero-operator for all other eigen vectors:

$$\frac{M - \lambda_i I}{\lambda_k - \lambda_i} \cdot Y_i = \frac{\lambda_i - \lambda_i}{\lambda_k - \lambda_i} \cdot Y_i = 0 \quad (214)$$

Hence

$$P_k \cdot Y_j = \prod_{i \neq k} \frac{\lambda_j - \lambda_i}{\lambda_k - \lambda_i} \cdot Y_j = \delta_{jk}$$

Now we write

$$P_k \mathbf{U} = [\dots 0, Y_k, 0 \dots] \quad (215)$$

and

$$\begin{aligned} f(D) &= \mathbf{U} \cdot f(\Lambda) \cdot \mathbf{U}^{-1} \\ \mathbf{U} \cdot f(\Lambda) &= \sum_{k=1}^{2n} f(\lambda_k) [\dots 0, Y_k, 0 \dots] = \sum_{k=1}^{2n} f(\lambda_k) \cdot P_k \cdot \mathbf{U} \end{aligned} \quad (216)$$

and finally

$$f(D) = \mathbf{U} \cdot f(\Lambda) \cdot \mathbf{U}^{-1} = \sum_{k=1}^{2n} f(\lambda_k) \cdot P_k \cdot \mathbf{U} \cdot \mathbf{U}^{-1} = \sum_{k=1}^{2n} f(\lambda_k) \cdot P_k \quad (217)$$

e.g.

$$f[\mathbf{D}] = \sum_{k=1}^{2n} f(\lambda_k) \prod_{j \neq k} \frac{\mathbf{D} - \lambda_j \mathbf{I}}{\lambda_k - \lambda_j} \quad (218)$$

equivalent to

$$f[\mathbf{D}_s] = \sum_{k=1}^{2n} f(\lambda_k s) \prod_{j \neq k} \frac{\mathbf{D} - \lambda_j \mathbf{I}}{\lambda_k - \lambda_j} \quad (219)$$

we got famous Sylvester formula.

We will use most of the time $f : \exp$ and Sylvester formula in form of (203). Naturally, (219) is comprised of power of matrix \mathbf{D} up to $2n-1$ – perfectly with agreement that \mathbf{D} is a root its characteristic equation (196).

Since \mathbf{D} is real matrix, any of its complex eigen values paired with their complex conjugates:

$$\mathbf{D}Y_k = \lambda_k Y_k \Leftrightarrow \mathbf{D}Y_k^* = \lambda_k^* Y_k^* \quad (220)$$

meanwhile real eigen values not always related. One more important ratio for accelerators: trace of \mathbf{D} is equal to zero, e.g. sum of it eigen values is also equal to zero:

$$\text{Trace}[\mathbf{D}] = \text{Trace}[\mathbf{U}\Lambda\mathbf{U}^{-1}] = \text{Trace}[\mathbf{U}^{-1}\mathbf{U}\Lambda] = \text{Trace}[\Lambda] = \sum_{k=1}^{2n} \lambda_k \quad (221)$$

It is especially useful for $n=1$ – you will see it in your home work.

Another easy case is when \mathbf{D} can be diagonalized, even though the number of different eigen values is $m < 2n$ (there is degeneration, i.e. some eigen values have multiplicity >1). We can use again simple Sylvester's formula (202) again, which just has fewer elements (m instead of $2n$):

$$\exp[\mathbf{D}s] = \sum_{k=1}^m e^{\lambda_k s} \prod_{\lambda_j \neq \lambda_k} \frac{\mathbf{D} - \lambda_j \mathbf{I}}{\lambda_k - \lambda_j} \quad (225)$$

We will use most of the time $f : \exp$ and Sylvester formula in form of (203). Naturally, (219) is comprised of power of matrix \mathbf{D} up to $2n-1$ – perfectly with agreement that \mathbf{D} is a root its characteristic equation (196).

Since \mathbf{D} is real matrix, any of its complex eigen values paired with their complex conjugates:

$$\mathbf{D}Y_k = \lambda_k Y_k \Leftrightarrow \mathbf{D}Y_k^* = \lambda_k^* Y_k^* \quad (220)$$

meanwhile real eigen values not always related. One more important ratio for accelerators: trace of \mathbf{D} is equal to zero, e.g. sum of it eigen values is also equal to zero:

$$\text{Trace}[\mathbf{D}] = \text{Trace}[\mathbf{U}\Lambda\mathbf{U}^{-1}] = \text{Trace}[\mathbf{U}^{-1}\mathbf{U}\Lambda] = \text{Trace}[\Lambda] = \sum_{k=1}^{2n} \lambda_k \quad (221)$$

It is especially useful for $n=1$ – you will see it in your home work.

Another easy case is when \mathbf{D} can be diagonalized, even though the number of different eigen values is $m < 2n$ (there is degeneration, i.e. some eigen values have multiplicity >1). We can use again simple Sylvester's formula (202) again, which just has fewer elements (m instead of $2n$):

$$\exp[\mathbf{D}s] = \sum_{k=1}^m e^{\lambda_k s} \prod_{\lambda_j \neq \lambda_k} \frac{\mathbf{D} - \lambda_j \mathbf{I}}{\lambda_k - \lambda_j} \quad (225)$$

General discussion

It is easy to show that polynomial form of matrix function is not unique. This is easy to show using the fact that every matrix \mathbf{A} has minimal polynomial* of which it is a root:

$$\psi(\lambda) = \sum_{k=0}^m a_k \lambda^k = \prod_{k=1}^l (\lambda - \lambda_k)^{h_k}; \det(A - \lambda_k I) = 0; a_m = 1.$$

$$\psi(A) = \sum_{k=0}^m a_k A^k = \prod_{k=1}^l (A - \lambda_k I)^{h_k} = 0; m = \sum_{k=1}^l h_k \leq 2n.$$

First, the Hamilton-Kelly theorem states that:

$$\psi_{HK}(\lambda) = \det(A - \lambda I) = \prod_{k=1}^l (A - \lambda_k I)^{h_k}; \sum_{k=1}^l h_k = 2n; \psi_{HK}(A) = 0$$

and such polynomials exist! Example: if matrix is diagonalized but some eigen values have multiplicity, the minimal polynomial has order less then

$$\psi_l(\lambda) = \prod_{k=1}^l (\lambda - \lambda_k I); \psi_l(A) = 0; l < 2n;$$

Let $\psi(A) = 0$ and we defined polinomial

$$p_f(\lambda) = \sum_{k=0}^m a_k \lambda^k \rightarrow f(A) = p_f(A);$$

then polynomial $p(\lambda) = p_f(\lambda) + r(\lambda) \cdot \psi(\lambda)$, where r is an arbitry polynomial, is also a valid polynomial expression for $f(A)$:

$$p(A) = p_f(A) + r(A) \cdot \psi(A) \downarrow_0 = p_f(A) \approx f(A);$$

There is nothing wrong with this – sometimes it is useful to have options.

*The minimal polynomial of \mathbf{A} is defined to be the unique monic polynomial ψ of lowest degree such that $\psi(\mathbf{A}) = 0$. It is unique, because if there are two minimal polynomials $\psi_1(\mathbf{A}) = 0$; $\psi_2(\mathbf{A}) = 0$, $\psi_3 = \psi_1 - \psi_2$ is lower order polinomila for which is , which contradicts definition of lowest degree!

Full consideration requires a bit more work. An arbitrary matrix \mathbf{M} can be reduced to an unique matrix, which in general case has a Jordan form: for a matrix with arbitrary height of eigen values the set of eigen values $\{\lambda_1, \dots, \lambda_m\}$ contains only unique eigen values, i.e. $\lambda_k \neq \lambda_j; \forall k \neq j$:

$$\begin{aligned} size[\mathbf{M}] = M; \quad \{\lambda_1, \dots, \lambda_m\}; \quad m \leq M; \quad \det[\lambda_k \mathbf{I} - \mathbf{M}] = 0; \\ \mathbf{M} = \mathbf{U} \mathbf{G} \mathbf{U}^{-1}; \quad \mathbf{G} = \sum_{\oplus k=1, m} \mathbf{G}_k = \mathbf{G}_1 \oplus \dots \oplus \mathbf{G}_m; \quad \sum size[\mathbf{G}_k] = M \end{aligned} \quad (226)$$

where \oplus means direct sum of block-diagonal square matrixes \mathbf{G}_k which correspond to the eigen vector sub-space adjacent to the eigen value λ_k . Size of \mathbf{G}_k , which we call l_k , is equal to the multiplicity of the root λ_k of the characteristic equation

$$\det[\lambda \mathbf{I} - \mathbf{M}] = \prod_{k=1, m} (\lambda - \lambda_k)^{l_k}.$$

In general case, \mathbf{G}_k is also a block diagonal matrix comprised of orthogonal sub-spaces belonging to the same eigen value

$$\mathbf{G}_k = \sum_{\oplus j=1, p_k} \mathbf{G}_k^j = \mathbf{G}_k^1 \oplus \dots \oplus \mathbf{G}_k^{p_k}; \quad \sum size[\mathbf{G}_k^j] = l_k \quad (227)$$

where we assume that we sorted the matrixes by increasing size: $size[\mathbf{G}_k^{j+1}] \geq size[\mathbf{G}_k^j]$, i.e. the

$$n_k = size[\mathbf{G}_k^{p_k}] \leq l_k \quad (228)$$

is the maximum size of the Jordan matrix belonging to the eigen value λ_k . General form of the Jordan matrix is:

$$\mathbf{G}_k^n = \begin{bmatrix} \lambda_k & 1 & 0 & 0 \\ 0 & \lambda_k & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_k \end{bmatrix} \quad (229)$$

An arbitrary analytical matrix function of \mathbf{M} can be expanded into Taylor series and reduced to the function of its Jordan matrix \mathbf{G} :

$$f(\mathbf{M}) = \sum_{i=1}^{\infty} f_i \mathbf{M}^i = \sum_{i=1}^{\infty} f_i (\mathbf{U} \mathbf{G} \mathbf{U}^{-1})^i \equiv \left(\sum_{i=1}^{\infty} f_i \mathbf{U} (\mathbf{G})^i \mathbf{U}^{-1} \right) = \mathbf{U} \left(\sum_{i=1}^{\infty} f_i (\mathbf{G})^i \right) \mathbf{U}^{-1} = \mathbf{U} f(\mathbf{G}) \mathbf{U}^{-1} \quad (230)$$

it is direct sum of the function of the Jordan blocks:

$$f(\mathbf{G}) = \sum_{i=0}^{\infty} f_i \mathbf{G}^i = \sum_{i=0}^{\infty} f_i \begin{bmatrix} \mathbf{G}_1^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_m^{p_m} \end{bmatrix}^i = \begin{bmatrix} \sum_{i=0}^{\infty} f_i (\mathbf{G}_1^1)^i & \mathbf{0} \\ \mathbf{0} & \dots \\ \mathbf{0} & \dots & \sum_{i=0}^{\infty} f_i (\mathbf{G}_m^{p_m})^i \end{bmatrix}$$

$$= \begin{bmatrix} f(\mathbf{G}_1^1) & \mathbf{0} \\ \mathbf{0} & \dots \\ & & f(\mathbf{G}_m^{p_m}) \end{bmatrix} = \sum_{\oplus k=1, m, j=1, p_k} f(\mathbf{G}_k^j) = f(\mathbf{G}_1^1) \oplus \dots \oplus f(\mathbf{G}_m^{p_m}); \quad (231)$$

Function of a Jordan block of size n contains not only the function of corresponding eigen value λ , but also its derivatives to (n-1)th order:

$$\mathbf{G} = \begin{bmatrix} \lambda & 1 & \dots & 0 \\ 0 & \lambda & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & \lambda \end{bmatrix}; f(\mathbf{G}) = \begin{bmatrix} f(\lambda) & f'(\lambda)/1! & \dots & f^{(k)}(\lambda)/k! & f^{(n-1)}(\lambda)/(n-1)! \\ 0 & f(\lambda) & \dots & \dots & f^{(n-2)}(\lambda)/(n-2)! \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & f'(\lambda)/1! & \dots \\ 0 & 0 & \dots & f(\lambda) & \dots \end{bmatrix} \quad (232)$$

The prove of Eq. 21 is your take-home task – use polynomial as a function.

It means that in general case Sylvester formula will include not only function of the eigen values, but also their derivative.

Formal definition of matrix functions.

The function f is said to be defined on the spectrum of matrix A if the values

$$f(\lambda_i), f'(\lambda_i) \dots f^{(l_i-1)}(\lambda_i), \quad i = 1, \dots, m \quad (233)$$

where $\lambda_1, \dots, \lambda_m$ are distinct eigen values and l_i is the order (size) of the largest Jordan block related to eigen value λ_i . They are called the values of function on the spectrum of matrix A . It is obvious that any polynomial is defined at the spectrum of A .

From definition of Jordan canonical form, we can clearly see that minimal polynomial

$$\psi(\lambda) = \prod_{i=1}^m (\lambda - \lambda_i)^{l_i} \quad (234)$$

is zero at spectrum of matrix A .

Now a theorem: *For polynomials p and q and A , $p(A) = q(A)$ if and only if p and q take the same values on the spectrum of A .*

Proof: Suppose that two polynomials p and q satisfy $p(A) = q(A)$. Then $d = p - q$ is zero at A so is divisible by the minimal polynomial ψ^* . In other words, d takes only the value zero on the spectrum of A , that is, p and q take the same values on the spectrum of A .

Conversely, suppose p and q take the same values on the spectrum of A . Then $d = p - q$ is zero on the spectrum of A and so must be divisible by the minimum polynomial ψ . Hence $d = \psi r$ for some polynomial r , and since $d(A) = \psi(A)r(A) = 0$, it follows that $p(A) = q(A)$.

* *A key property is that the minimal polynomial divides any other polynomial p for which $p(A) = 0$. Indeed, by polynomial long division any such p can be written $p = \psi q + r$, where the degree of the remainder r is less than that of ψ . But $0 = p(A) = \psi(A)q(A) + r(A) = r(A)$, and this contradicts the minimality of the degree of ψ unless $r = 0$. Hence $r = 0$ and ψ divides p .*

Definition of matrix function via Hermite interpolation:

Let f be defined on the spectrum of A and let ψ be the minimal polynomial of A . Then $f(A) := p(A)$, where p is the polynomial of degree less than that of ψ and satisfies the interpolations conditions:

$$p(\lambda_i) = f(\lambda_i), p'(\lambda_i) = f'(\lambda_i) \dots p^{(l_i-1)}(\lambda_i) = f^{(l_i-1)}(\lambda_i), \quad i = 1, \dots, m \quad (235)$$

Such polynomial is unique and is known as the Hermite interpolating polynomial

As we discussed before, polynomial expansion of matrix function is not unique. If q is a polynomial that satisfies the interpolation conditions (235) and some additional interpolation conditions (at the same or different λ_i) then q and the polynomial p take the same values on the spectrum of A . Proven theorem states: $q(A) = p(A) = f(A)$. Sometimes, in constructing a polynomial q for which $q(A) = f(A)$, it is convenient to impose more interpolation conditions than necessary—typically if the eigenvalues of A are known but the Jordan form is not. Doing so yields a polynomial of higher degree than necessary but does not affect the ability of the polynomial to produce $f(A)$. For example.

$$p(\lambda_i) = f(\lambda_i), p'(\lambda_i) = f'(\lambda_i) \dots p^{(N_i)}(\lambda_i) = f^{(N_i)}(\lambda_i), \quad i = 1, \dots, m \quad (235')$$

where N_i is multiplicity of λ_i , which can be larger than l_i .

Another example of such polynomials come from a direct derivation given in additional reading materials on the case website: **Generalization of Sylvester formula**. At the time I made this derivation, I was on the other side “Iron curtain” and did not had access to modern book about matrix function – hence, I had to derived it for myself from scratch... and it is a bit scratchy and bulky...

The Hermite interpolating polynomial p is given explicitly by the Lagrange–Hermite formula

$$p(\lambda) = \sum_{i=1}^m \left[\left(\sum_{j=0}^{l_i-1} \frac{\phi_i^{(j)}(\lambda_i)}{j!} (\lambda - \lambda_i)^j \right) \prod_{j \neq i} (\lambda - \lambda_j)^{l_j} \right]; \quad \phi_i(\lambda) = f(\lambda) / \prod_{j \neq i} (\lambda - \lambda_j)^{l_j} \quad (236)$$

Hence, general Sylvester formula for matrix function is

$$f(A) = \sum_{i=1}^m \left[\left(\sum_{j=0}^{l_i-1} \frac{\phi_i^{(j)}(\lambda_i)}{j!} (A - \lambda_i I)^j \right) \prod_{j \neq i} (A - \lambda_j I)^{l_j} \right]; \quad (237)$$

It is easy to show that for the case of diagonal Jordan form with all $l_i=1$, we derive to the already known formula:

$$p(\lambda) = \sum_{i=1}^m f(\lambda_i) \prod_{j \neq i} \frac{\lambda - \lambda_j}{\lambda_i - \lambda_j}; \quad f(A) = \sum_{i=1}^m f(\lambda_i) \prod_{j \neq i} \frac{A - \lambda_j I}{\lambda_i - \lambda_j}.$$

While in this course we will mostly use exponent as the function of interest, not you have a tool for your research and can find square or cubic root of matrix (with appropriate branch, take logarithm of matrix (again, with a proper branch), or do may other things. Thus – you have a new tool in your hands! Evaluating general Sylvester formula using Mathematica is a piece of case – try it, just for fun! Note, that while we will use even size $2n \times 2n$ matrices, Sylvester formula is derived for arbitrary square matrices.

Just one more step before we embark on specific cases. In many case you need to evaluate matrix function with a parameter , for example we will use eigen values of scaled matrix

$$A = \mathbf{D} \cdot s; \det[\mathbf{D} - \lambda_i \mathbf{I}] = 0 \Rightarrow \det[A - (\lambda_i s) \mathbf{I}] = 0 \quad (238)$$

Hence, general Sylvester formula for scaled matrix function is

$$f(\mathbf{D} \cdot s) = \sum_{i=1}^m \left[\left(\sum_{j=0}^{l_i-1} \frac{\phi_i^{(j)}(\lambda_i \cdot s)}{j!} (\mathbf{D} - \lambda_i I)^j \cdot s^j \right) \prod_{j \neq i} (\mathbf{D} - \lambda_j I)^{l_j} \cdot s^{l_j} \right];$$

$$\phi_i(\lambda s) = \frac{f(\lambda s)}{\prod_{j \neq i} (\lambda - \lambda_j)^{l_j}} \left(\prod_{j \neq i} s^{l_j} \right)^{-1};$$

$$\phi_i^{(j)}(\lambda s) = \frac{\partial^j \phi_i}{\partial (\lambda s)^j} = \frac{\partial^j \phi_i}{\partial \lambda^j} s^{-j} = \frac{s^{-j}}{\prod_{j \neq i} s^{l_j}} \frac{\partial^j}{\partial \lambda^j} \left(\frac{f(\lambda s)}{\prod_{j \neq i} (\lambda - \lambda_j)^{l_j}} \right);$$

$$f(\mathbf{D} \cdot s) = \sum_{i=1}^m \left[\left(\sum_{j=0}^{l_i-1} \frac{\bar{\phi}_i^{(j)}(\lambda_i)}{j!} (\mathbf{D} - \lambda_i I)^j \right) \prod_{j \neq i} (\mathbf{D} - \lambda_j I)^{l_j} \right];$$

$$\bar{\phi}_i(\lambda) = \frac{f(\lambda s)}{\prod_{j \neq i} (\lambda - \lambda_j)^{l_j}}; \bar{\phi}_i^{(j)}(\lambda) = \frac{\partial^j \bar{\phi}_i}{\partial \lambda^j}.$$
(239)

with much easier form for diagonal case:

$$f(\mathbf{D} \cdot s) = \sum_{i=1}^m f(\lambda_i \cdot s) \prod_{j \neq i} \frac{\mathbf{D} - \lambda_j \mathbf{I}}{\lambda_i - \lambda_j}. \quad (240)$$

Hamiltonian system

- We should expect that matrix representing Hamiltonian systems would have special features

$$\mathbf{M} = \exp(\mathbf{D} \cdot s); \mathbf{D} = \mathbf{S} \cdot \mathbf{H}$$

- First, \mathbf{D} are real
- Second, \mathbf{D} have zero trace
- And more

For Hamiltonian system, eigen values split into pairs with the opposite sign:

$$\begin{aligned} \det[\mathbf{SH} - \lambda \cdot \mathbf{I}] &= \det[\mathbf{SH} - \lambda \cdot \mathbf{I}]^T = \det[-\mathbf{HS} - \lambda \cdot \mathbf{I}] = \\ (-1)^{2n} \det[\mathbf{HS} + \lambda \cdot \mathbf{I}] &= \det(\mathbf{S}[\mathbf{HS} + \lambda \cdot \mathbf{I}]\mathbf{S}^{-1}) = \det[\mathbf{SH} + \lambda \cdot \mathbf{I}] \# \end{aligned} \quad (241)$$

First, it makes finding eigen values an easier problem, because characteristic equation is bi-quadratic:

$$\det[\mathbf{D} - \lambda \mathbf{I}] = \prod (\lambda_i - \lambda)(-\lambda_i - \lambda) = \prod (\lambda^2 - \lambda_i^2) = 0. \quad (241')$$

For accelerator elements it is of paramount importance, 1D case is reduces to trivial (243), 2D case is reduced to solution of quadratic equation and 3D case (6D phase space) required to solve cubic equation. For analytical work it gives analytical expressions – compare it with attempt to write analytical formula for roots of a generic polynomial of 6-order? It simply does not exist! Thus, we have an extra gift for accelerator physics – the roots can be written and studied! It is also allows us to simplify (202) into

$$\begin{aligned} \exp[\mathbf{D}s] &= \left\{ \sum_{k=1}^n e^{\lambda_k s} \frac{\mathbf{D} + \lambda_k \mathbf{I}}{2\lambda_k} \prod_{j \neq k} \left(\frac{\mathbf{D}^2 - \lambda_j^2 \mathbf{I}}{\lambda_k^2 - \lambda_j^2} \right) - e^{-\lambda_k s} \frac{\mathbf{D} - \lambda_k \mathbf{I}}{2\lambda_k} \prod_{j \neq k} \left(\frac{\mathbf{D}^2 - \lambda_j^2 \mathbf{I}}{\lambda_k^2 - \lambda_j^2} \right) \right\} \\ \exp[\mathbf{D}s] &= \sum_{k=1}^n \left(\frac{e^{\lambda_k s} + e^{-\lambda_k s}}{2} \mathbf{I} + \frac{e^{\lambda_k s} - e^{-\lambda_k s}}{2\lambda_k} \mathbf{D} \right) \prod_{j \neq k} \left(\frac{\mathbf{D}^2 - \lambda_j^2 \mathbf{I}}{\lambda_k^2 - \lambda_j^2} \right) \end{aligned} \quad (242)$$

where index k goes only through n pairs of $\{\lambda_k, -\lambda_k\}$. While (242) does not look simpler, it really makes it easier (4 times less calculations) when we do it by hands... For example we can look at 1D case. First, we can easily see that

$$\text{Trace} \mathbf{D} = \lambda_1 + \lambda_2 = 0 \rightarrow \lambda_1 = -\lambda_2 = \lambda; \quad \lambda^2 = -\det[\mathbf{D}] \quad (243)$$

Thus, it is non-degenerated case only when $\det[\mathbf{D}] \neq 0$. Regular Sylvester formula gives us a simple two-piece expression :

$$\exp[\mathbf{D}s] = e^{\lambda s} \frac{\mathbf{D} - \lambda \mathbf{I}}{2\lambda} - e^{-\lambda s} \frac{\mathbf{D} + \lambda \mathbf{I}}{2\lambda} \quad (244)$$

brining it home right away:

$$\begin{aligned} \exp[\mathbf{D}s] &= \mathbf{I} \cdot \frac{e^{\lambda s} + e^{-\lambda s}}{2} + \mathbf{D} \frac{e^{\lambda s} - e^{-\lambda s}}{2\lambda}; \\ \exp[\mathbf{D}s] &= \mathbf{I} \cdot \cosh|\lambda|s + \frac{\mathbf{D} \sinh|\lambda|s}{|\lambda|}; \quad \det[\mathbf{D}] < 0; \quad |\lambda| = \sqrt{-\det[\mathbf{D}]} \end{aligned} \quad (245)$$

$$\exp[\mathbf{D}s] = \mathbf{I} \cdot \cos|\lambda|s + \frac{\mathbf{D} \sin|\lambda|s}{|\lambda|}; \quad \det[\mathbf{D}] > 0; \quad |\lambda| = \sqrt{\det[\mathbf{D}]}$$

The case $\det[\mathbf{D}] = 0$ means in this case that \mathbf{D} is nilpotent:

$$\begin{aligned} \det \mathbf{D} = 0 &\Rightarrow \lambda_1 = -\lambda_2 = 0; \\ d(\lambda) = \det[\mathbf{D} - \lambda \mathbf{I}] &= (\lambda_1 - \lambda)(-\lambda_1 - \lambda) = \lambda^2 \Rightarrow \mathbf{D}^2 = 0 \end{aligned} \quad (246)$$

hence

$$\exp[\mathbf{D}s] = \mathbf{I} + \mathbf{D}s; \quad \det[\mathbf{D}] = 0; \quad (247)$$

Naturally, this is the same as result of full-blown degenerated case, but it also can be obtained as a limit case of (245) when $|\lambda| \rightarrow 0$:

$$\begin{aligned} \mathbf{I} \cdot \cosh|\lambda|s + \frac{\mathbf{D} \sinh|\lambda|s}{|\lambda|} &\xrightarrow{|\lambda| \rightarrow 0} \mathbf{I} + \mathbf{D}s; \\ \mathbf{I} \cdot \cos|\lambda|s + \frac{\mathbf{D} \sin|\lambda|s}{|\lambda|} &\xrightarrow{|\lambda| \rightarrow 0} \mathbf{I} + \mathbf{D}s. \end{aligned}$$

What we learned today?

- Linear ordinary equations with constant coefficients (\mathbf{D} -matrix) have a natural solution as $\exp(\mathbf{D}\cdot s)$
- We can use functions of matrices and built entire method have analytical expression of matrix function as soon as we know eigen values of matrix \mathbf{D}
- Matrix function have a very simple and elegant form – called Sylvester formula- when eigen values are unique (e.g. in non-degenerating case) and \mathbf{D} can be diagonalized
- But even in a most general case, we can write analytical expression for matrix function
- In linear Hamiltonian case, eigen values split in pair of $(\lambda, -\lambda)$ and the expression can be even further simplified
- The remaining task for linear matrices if accelerators is to find analytical expression for eigen values – the job for next class

If time allow, direct proof of generalized Sylvester formula

Direct proof of the General Sylvester formula

We had a very nice derivation of Sylvester formula for non-degenerated case when all \mathbf{n} eigen values are distinct:

$$f(\mathbf{D}) = \sum_{k=1}^n f(\lambda_k) \prod_{j \neq k} \frac{\mathbf{D} - \lambda_j \mathbf{I}}{\lambda_k - \lambda_j} \quad (1)$$

but generalized formular which works for degenerated case and is based on $\mathbf{m} < \mathbf{n}$ distinct eigen values use Hermite polynomial

$$f(\mathbf{D}) = \sum_{i=1}^m \left[\left(\sum_{j=0}^{l_i-1} \frac{\phi_i^{(j)}(\lambda_i)}{j!} (\mathbf{D} - \lambda_i \mathbf{I})^j \right) \prod_{j \neq i} (\mathbf{D} - \lambda_j \mathbf{I})^{l_j} \right]; \quad (2)$$

$$\phi_i(\lambda) = \frac{f(\lambda)}{\prod_{j \neq i} (\lambda - \lambda_j)^{l_j}}; \phi_i^{(j)}(\lambda) = \frac{\partial^j \phi_i}{\partial \lambda^j}.$$

where l_i is maximum size of the Jordan bloc related to eigen value λ_i :

$$\mathbf{D} = \mathbf{U} \mathbf{G} \mathbf{U}^{-1}; \mathbf{G} = \sum_{\oplus k=1}^m \mathbf{G}_k = \mathbf{G}_1 \oplus \dots \oplus \mathbf{G}_k; l_k = \max(n_{ik}) = \text{Max}[\text{size}(\mathbf{G}_{ki})]$$

$$\mathbf{G}_k = \sum_{\oplus i=1}^{m_i} \mathbf{G}_{ki} = \mathbf{G}_{k1} \oplus \dots \oplus \mathbf{G}_{km_i}; \mathbf{G}_{ki} = \begin{bmatrix} \lambda_k & 1 & 0 & 0 \\ \dots & \dots & \dots & 0 \\ 0 & 0 & \lambda_k & 1 \\ 0 & 0 & 0 & \lambda_k \end{bmatrix}_{n_{ik} \times n_{ik}}; \quad (3)$$

By definition:

$$\begin{aligned}
f(\mathbf{G}) &= \sum_{i=0}^{\infty} f_i \mathbf{G}^i = \sum_{i=0}^{\infty} f_i \begin{bmatrix} \mathbf{G}_1^1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_m^{p_m} \end{bmatrix}^i = \begin{bmatrix} \sum_{i=0}^{\infty} f_i (\mathbf{G}_1^1)^i & \mathbf{0} \\ \mathbf{0} & \dots \\ & & \sum_{i=0}^{\infty} f_i (\mathbf{G}_m^{p_m})^i \end{bmatrix} \\
&= \begin{bmatrix} f(\mathbf{G}_1^1) & \mathbf{0} \\ \mathbf{0} & \dots \\ & & f(\mathbf{G}_m^{p_m}) \end{bmatrix} = \sum_{\oplus k=1, m, j=1, p_k} f(\mathbf{G}_k^j) = f(\mathbf{G}_1^1) \oplus \dots \oplus f(\mathbf{G}_m^{p_m}); \quad (4)
\end{aligned}$$

and we need to prove that

$$\begin{aligned}
f(\mathbf{D}) &= \mathbf{U} \sum_{i=1}^m \left[\left(\sum_{j=0}^{l_i-1} \frac{\phi_i^{(j)}(\lambda_i)}{j!} (\mathbf{G} - \lambda_i \mathbf{I})^j \right) \prod_{j \neq i} (\mathbf{G} - \lambda_j \mathbf{I})^{l_j} \right] \mathbf{U}^{-1} = \\
&\mathbf{U} \cdot \sum_{\oplus k} \sum_{i=1}^m \left[\left(\sum_{j=0}^{l_i-1} \frac{\phi_i^{(j)}(\lambda_i)}{j!} (\mathbf{G}_k - \lambda_i \mathbf{I})^j \right) \prod_{j \neq i} (\mathbf{G}_k - \lambda_j \mathbf{I})^{l_j} \right] \mathbf{U}^{-1} \quad (5)
\end{aligned}$$

First, it is easy to prove that: Lemma 1

$$\prod_{j \neq i} (\mathbf{G}_k - \lambda_j \mathbf{I})^{l_j} = 0; k \neq i.$$

Since $k \neq i$, the product contains $(\mathbf{G}_k - \lambda_k \mathbf{I})^{l_k}$:

$$\mathbf{A} = \mathbf{G}_{ki} - \lambda_k \mathbf{I} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{j \times j}; \mathbf{A}^2 = \begin{bmatrix} 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \mathbf{A}^{j-1} = \begin{bmatrix} 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{A}^j = 0 \Rightarrow (\mathbf{G}_{ki} - \lambda_k \mathbf{I})^{l_k} = 0; (\mathbf{G}_k - \lambda_k \mathbf{I})^{l_k} = \sum_{\oplus} 0 = 0 \quad (6)$$

Hence it becomes direct sum of function of the Jordan blocks:

$$f(\mathbf{D}) = \mathbf{U} \cdot \sum_{\oplus i=1}^m \left[\left(\sum_{j=0}^{l_i-1} \frac{\phi_i^{(j)}(\lambda_i)}{j!} (\mathbf{G}_i - \lambda_i \mathbf{I})^j \right) \prod_{j \neq i} (\mathbf{G}_i - \lambda_j \mathbf{I})^{l_j} \right] \mathbf{U}^{-1} \quad (7)$$

Now we need for show that

$$\left(\sum_{j=0}^{l_i-1} \frac{\phi_i^{(j)}(\lambda_i)}{j!} (\mathbf{G}_{ik} - \lambda_i \mathbf{I})^j \right) \prod_{j \neq i} (\mathbf{G}_{ik} - \lambda_j \mathbf{I})^{l_j} = f(\mathbf{G}_{ik});$$

$$f(\mathbf{G}_{ik}) = \begin{bmatrix} f(\lambda_i) & f'(\lambda_i)/1! & \dots & f^{(N-2)}(\lambda_i)/(N-2)! & f^{(N-1)}(\lambda_i)/(N-1)! \\ 0 & f(\lambda_i) & \dots & & \\ 0 & 0 & \dots & f'(\lambda_i)/1! & \\ 0 & 0 & \dots & f(\lambda_i) & f'(\lambda_i)/1! \\ 0 & 0 & \dots & 0 & f(\lambda_i) \end{bmatrix}_{N \times N} \quad (8)$$

Let's define two functions: $f(\lambda)$ and $\psi_i(\lambda) = \prod_{j \neq i} (\lambda - \lambda_j)$, which immediately yields:

$$\prod_{j \neq i} (\mathbf{G}_{ik} - \lambda_j \mathbf{I})^{l_j} = \psi_i(\mathbf{G}_{ik}) =$$

$$\begin{bmatrix} \psi_i(\lambda_i) & \psi_i'(\lambda_i)/1! & \dots & \psi_i^{(N-2)}(\lambda_i)/(N-2)! & \psi_i^{(N-1)}(\lambda_i)/(N-1)! \\ 0 & \psi_i(\lambda_i) & \dots & & \\ 0 & 0 & \dots & \psi_i'(\lambda_i)/1! & \\ 0 & 0 & \dots & \psi_i(\lambda_i) & \psi_i'(\lambda_i)/1! \\ 0 & 0 & \dots & 0 & \psi_i(\lambda_i) \end{bmatrix};$$

$$(\mathbf{G}_{ik} - \lambda_i \mathbf{I}) \prod_{j \neq i} (\mathbf{G}_{ik} - \lambda_j \mathbf{I})^{l_j} = \begin{bmatrix} 0 & \psi_i(\lambda_i) & \dots & \psi_i^{(j-3)}(\lambda_i)/(j-3)! & \psi_i^{(j-2)}(\lambda_i)/(j-2)! \\ 0 & 0 & \dots & & \\ 0 & 0 & \dots & \psi_i(\lambda_i) & \\ 0 & 0 & \dots & 0 & \psi_i(\lambda_i) \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

$$(\mathbf{G}_{ik} - \lambda_i \mathbf{I})^{N-1} \prod_{j \neq i} (\mathbf{G}_{ik} - \lambda_j \mathbf{I})^{l_j} = \begin{bmatrix} 0 & 0 & \psi_i(\lambda_i) \\ 0 & \dots & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (9)$$

or to be exact

$$\left(\sum_{j=0}^{l_i-1} \frac{\phi_i^{(j)}(\lambda_i)}{j!} (\mathbf{G}_{ik} - \lambda_i \mathbf{I})^j \right) \prod_{j \neq i} (\mathbf{G}_{ik} - \lambda_j \mathbf{I})^{l_j} =$$

$$f(\mathbf{G}_{ik}) = \begin{bmatrix} \phi_i \psi_i & \phi_i \psi_i' / 1! + \phi_i^{(j)} \psi_i & \dots & \dots & \sum_{k=0}^{N-1} \frac{\psi_i^{(k)}}{k!} \phi_i^{(N-1-k)} \\ 0 & \phi_i \psi_i & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & \dots \\ 0 & 0 & \dots & \phi_i \psi_i & \phi_i \psi_i' / 1! + \phi_i^{(j)} \psi_i \\ 0 & 0 & \dots & 0 & \phi_i \psi_i \end{bmatrix}_{N \times N} \quad (10)$$

Diagonal terms are obviously correct:

$$\phi_i(\lambda) = f(\lambda) / \psi_i(\lambda) \rightarrow \phi_i(\lambda_i) \psi_i(\lambda_i) \equiv f(\lambda_i)$$

The rest is also straightforward:

$$f(\lambda) = \phi_i(\lambda) \cdot \psi_i(\lambda) \rightarrow f^{(k)} = \sum_{n=0}^k C_n^k \phi_i^{(n)}(\lambda) \cdot \psi_i^{(k-n)}(\lambda); C_n^k = \frac{n!}{k!(n-k)!};$$

$$\sum_{n=0}^k \frac{\phi_i^{(n)}(\lambda) \cdot \psi_i^{(k-n)}(\lambda)}{k!(n-k)!} = \frac{f^{(k)}}{n!} \# \quad (11)$$

This is direct proof, which does not require understanding of polynomials!