Sylvester's Influence on Applied Mathematics

Higham, Nicholas J.

2014

MIMS EPrint: **2014.26**

ISSN 1749-9097

Sylvester's Influence on Applied Mathematics*

Nicholas J. Higham†

Abstract

James Joseph Sylvester coined the term "matrix" and contributed much to the early development of matrix theory. To mark the 200th anniversary of his birth I show how Sylvester's work on matrices continues to influence applied mathematics today.

1 Introduction

This year is the 200th anniversary of the birth of James Joseph Sylvester (September 3, 1814–March 15, 1897), FRS. Sylvester was a prolific mathematician, the four volumes of his collected works totalling almost 3000 pages. He also led an eventful life, holding positions at five academic institutions, two of them in the USA. For several years he was an actuary by day and did his mathematical research at night, and indeed he was one of the founders of the Institute of Actuaries. During his second stay in the USA he founded the American Journal of Mathematics and formed the first research school in the country. He was a controversial figure, being prone to rows and to disputes over the priority of research.

Since the centenary of his death, Sylvester's life and work has been the subject of renewed interest. In this article I describe some of Sylvester's mathematical contributions and show that they are still very much in use in applied mathematics today, especially in the areas of linear algebra and numerical analysis.

For details of Sylvester's life I recommend the masterly biography by Parshall [25] (which I reviewed in [15]) or, for a shorter summary, the article by James [22].

2 Sylvester the Neologist

One of Sylvester's greatest influences on mathematics is relatively little known: he introduced many mathematical terms that are still used today. In 1850 he coined the term "matrix", writing [29]

> We must commence, not with a square, but with an oblong arrangement of terms consisting, suppose, of m lines and n columns. This will not in itself represent a determinant, but is, as it were, a Matrix out of which we may form various systems of determinants by fixing upon a number p, and selecting at will p lines and p columns, the squares corresponding to which may be termed determinants of the pth order.

As the quote indicates, determinants were in common use at that time and preceded the notion of matrix.

It was Cayley who took the first steps to develop matrix algebra, in his 1858 memoir [5]. Sylvester did not return to matrices for another thirty years, and when he did he reinvented the subject under the name "universal algebra", claiming to have been unaware of Cayley's paper [34].

In Table 1 we list some other terms that the Oxford English Dictionary credits to Sylvester. Sylvester himself annotated[1] a 27 term "index to definitions" in his copy of Salmon's textbook on algebra [27] with the statement "With the exception of the words 'Eliminant' and 'Quantic' all the above terms are of Mr Sylvester's creation." However, it is as well to be a little skeptical about this claim, since Sylvester was notorious for what his biographer Parshall calls "an impatience with bibliographic research" [25, p. 59].

In a long 1853 article [32] Sylvester included a six page glossary of "new or unusual Terms,

*This is a reprint of N. J. Higham, Sylvester's Influence on Applied Mathematics, *Mathematics Today* 50 (4), pp. 202–206, August 2014 and contains additional historical references.

†School of Mathematics, University of Manchester, Manchester, M13 9PL, UK (nick.higham@manchester.ac.uk, http://www.maths.manchester.ac.uk/~higham)

[1] See the plates following page 224 in [25].

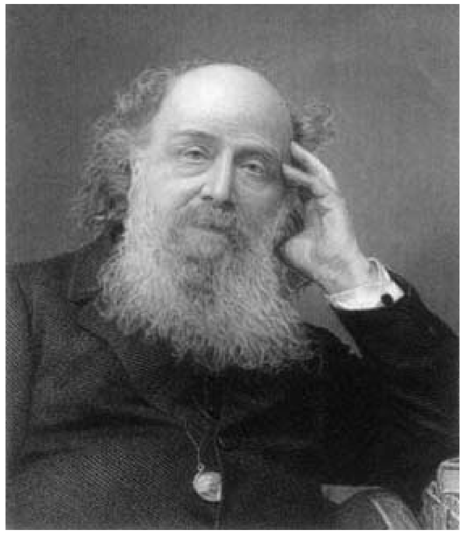

Figure 1: James Joseph Sylvester, "sometime after his arrival in Oxford in 1884" [25, plate following p. 224]. Source: http://www-history.mcs.st-and.ac.uk/history/PictDisplay/Sylvester.html.

or of Terms used in a new or unusual sense in the preceding Memoir." The glossary contains more than just technical description: the entry for "Hessian" is "named after Dr. Otto Hesse, of Konigsberg (the worthy pupil of his illustrious master, Jacobi, but who, to the scandal of the mathematical world, remains still without a Chair in the University which he adorns with his presence and his name)." The Oxford English Dictionary credits "Hessian" to Cayley in 1856, but Sylvester used the term as early as 1851 [7, p. 473], [25, p. 100], [30].

One term that has fallen out of use is "latent root", introduced by Sylvester in 1883 [33] with two charming similes:

> "It will be convenient to introduce here a notion (which plays a conspicuous part in my new theory of multiple algebra), namely 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."

This term was in use up until the 1970s [10] but has now been supplanted by "eigenvalue", though some authors complain about the latter term's incomplete translation from the German *eigenwert*. McIntyre [23] credits Sylvester with being the first to use the symbol λ to denote an eigenvalue of a matrix, in 1852 [31].

Sylvester introduced the adjective "derogatory" for a matrix whose minimal polynomial has degree less than the characteristic polynomial [38]. He also called such a matrix "privileged", suggesting that this property has both good and bad features. The property arose in the context of finding all matrices that commute with a given matrix, a goal that was completely attained not by him but by his German contemporaries using a more sophisticated line of attack exploiting canonical forms.

3 Sylvester's Equation

The Sylvester equation is the linear matrix equation

$$AX + XB = C, \tag{1}$$

where A is $m \times m$, B is $n \times n$, and X is an unknown $m \times n$ matrix. In 1884 Sylvester [37] considered the homogeneous version of the equation and thereby showed that the condition for (1) to have a unique solution is that A and $-B$ have no eigenvalues in common.

Since the equation is linear in X it must be possible to write it in the more usual form "$Ax = b$." Indeed if we denote by $\mathrm{vec}(X)$ the vector comprising the columns of X stacked one on top of the other from first to last then (1) can be written

$$\underbrace{(I_n \otimes A + B^T \otimes I_m)}_{nm\times nm} \mathrm{vec}(X) = \mathrm{vec}(C), \tag{2}$$

where for $F \in \mathbb{R}^{p\times q}$ and $G \in \mathbb{R}^{r\times s}$, $F \otimes G := (f_{ij}G) \in \mathbb{R}^{pr\times qs}$ is the Kronecker product. Interestingly, the Kronecker product predates the Sylvester equation: in 1858 Zehfuss gave the result $\det(A \otimes B) = \det(A)^n \det(B)^m$

Table 1: Terms coined by Sylvester, as credited in the Oxford English Dictionary [24].

Term	Year	Term	Year
matrix	1850	invariant	1851
minor	1850	Jacobian	1852
syzygy	1850	covariant	1853
canonical form	1851	latent root	1883
discriminant	1851	nullity	1884

[14]. Sylvester called the matrix in (2) the nivellateur (French for "leveller") [36].

The coefficient matrix of (2) is highly structured but it is not easy to take advantage of the structure. Therefore a great deal of research has been directed at analyzing and solving (1) directly. Of the various formulas that have been obtained for a solution we mention just one: if the integral $\int_0^\infty e^{At}Ce^{Bt}\,dt$ exists then minus this integral is a solution of the Sylvester equation.

One way in which the Sylvester equation arises is in block diagonalization. Suppose we wish to find a similarity transformation that introduces zeros into the $(1,2)$ block of the block upper triangular matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}.$$

This is a useful step if we wish to compute eigenvalues of A or matrix functions $f(A)$. It is easy to verify that

$$\begin{bmatrix} I & -X \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} I & -X \\ 0 & I \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}$$

if and only if X satisfies

$$A_{11}X - XA_{22} = A_{12}. \quad (3)$$

Hence block-diagonalizing A reduces to solving the Sylvester equation (3), which we know is possible if and only if the eigenvalues of A_{11} are distinct from those of A_{22}. This is an unsurprising restriction, as if it were not present we could diagonalize a 2×2 Jordan block, which of course is impossible.

For another way in which Sylvester equations arise consider the expansion $(X+E)^2 = X^2 + XE + EX + E^2$ for square matrices X and E, from which it follows that $XE + EX$ is the Fréchet derivative of the function x^2 at X in the direction E, written $L_{x^2}(X,E)$. We can find the Fréchet derivative of $x^{1/2}$ by applying the chain rule to $(x^{1/2})^2 = x$, which gives $L_{x^2}(X^{1/2}, L_{x^{1/2}}(X,E)) = E$. Therefore $Z = L_{x^{1/2}}(X,E)$ is the solution to the Sylvester equation $X^{1/2}Z + ZX^{1/2} = E$. The need to compute $L_{x^{1/2}}$ arises in the computation of the Fréchet derivatives of the matrix logarithm and of matrix powers [1], [16], [20].

In recent years research has focused particularly on solving Sylvester equations in which A and B are large and sparse and C has low rank, which arise in applications in control theory and model reduction, for example. In this case it is usually possible to find good low rank approximations to X and iterative methods based on Krylov subspaces have been very successful.

The Sylvester equation has many variations and special cases, including the Lyapunov equation $AX + XA^* = C$ (where "$*$" denotes conjugate transpose), the discrete Sylvester equation $X + AXB = C$, and versions of all these for operators [2]. It has also been generalized to multiple terms and with coefficient matrices on both sides of X, yielding

$$\sum_{i=1}^{k} A_i X B_i = C. \quad (4)$$

For $k \le 2$ and $m = n$ this equation can be solved in $O(n^3)$ operations. For $k > 2$, no $O(n^3)$ algorithm is known and deriving efficient numerical methods remains an open problem. The system (4) arises in stochastic finite element discretizations of partial differential equations with random inputs. The matrices A_i and B_i are large and sparse and, depending on the statistical properties of the random inputs, k can be arbitrarily large. In recent research efficient iterative solvers and preconditioners for such systems have been developed [26].

It is notable, and I think purely coincidental, that in this anniversary year MATLAB release 2014a provides a new function `sylvester` for solving the Sylvester equation, finally removing the need for MATLAB users to write their own solver.

4 The Quadratic Matrix Equation

Sylvester also considered nonlinear matrix equations, specifically the quadratic equation in $n \times n$ matrices

$$AX^2 + BX + C = 0. \tag{5}$$

A solution X is called a solvent. Note that because of the noncommutativity of matrices this is not the only possible quadratic equation: others include $X^2A + XB + C = 0$, for which an analogous treatment is possible, and the algebraic Riccati equation $XAX + BX + XC + D = 0$, which is important in control theory.

It is natural to wonder how properties of (5) generalize from the scalar case. For example, when $A = I$ can we write

$$X = -\frac{1}{2}B + \frac{1}{2}(B^2 - 4C)^{1/2}\,?$$

The answer is yes if B commutes with C and the square root exists. However, (5) may have no solution at all, as is clear from that fact that a special case is the matrix square root equation $-X^2 + C = 0$, which has no solution for $C = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$, for example.

We will just touch on just one aspect of the theory of (5). Consider the associated quadratic eigenvalue problem

$$(\lambda_i^2 A + \lambda_i B + C)v_i = 0, \tag{6}$$

where λ_i is an eigenvalue and the nonzero vector v_i a corresponding eigenvector. Suppose we can find n eigenpairs (λ_i, v_i) and that $V = [v_1, \dots, v_n]$ forms a nonsingular matrix. Then, with $\Lambda = \mathrm{diag}(\lambda_i)$, we have

$$AV\Lambda^2 + BV\Lambda + CV = 0,$$

and postmultiplying by V^{-1} we find that $X = V\Lambda V^{-1}$ is a solution of (5). Since the quadratic eigenvalue problem has $2n$ finite eigenvalues when A is nonsingular [40], this argument yields up to $\binom{2n}{n}$ choices of X. This number of solvents was identified by Sylvester [35], [39] for the case $A = I$. However, the existence and classification of solvents is more complicated than this discussion might indicate. One reason is that eigenvectors of (6) corresponding to distinct eigenvalues need not be linearly independent, in constrast to the situation for the eigensystem of a single matrix.

Today the quadratic matrix equation (5) is of interest because of its appearance in quasi birth-death processes, a form of Markov chain used in stochastic models in telecommunications, computer performance, and modeling of ecological systems. For more on the theory, applications, and numerical solution of the quadratic matrix equation see [19].

5 Law of Inertia

Undergraduate students may first come across Sylvester's name in connection with his law of inertia. Recall that the inertia of a Hermitian matrix is the triple of integers (ν, ζ, π), where ν is the number of negative eigenvalues, ζ is the number of zero eigenvalues, and π is the number of positive eigenvalues. Sylvester's law of inertia (1852) [31] says that for any Hermitian A and nonsingular matrix X the inertia of A is the same as that of X^*AX. (In fact, Sylvester stated and proved the result in the language of quadratic forms rather than matrix theory.) A transformation of the form X^*AX is called a congruence, so Sylvester's law says that the number of negative, zero, and positive eigenvalues does not change under congruence transformations.

Sylvester's law of inertia has many applications, of which we mention just the computation of the eigenvalues of Hermitian tridiagonal matrices T. Suppose we want to compute the kth smallest eigenvalue of T. Let $N(x)$ be the number of eigenvalues of T that are less than x. We need to find the point where $N(x)$ jumps from $k-1$ to k. It is feasible to do this by the bisection method if we can cheaply compute $N(x)$. Suppose we factorize $T - xI = LDL^*$, where D is diagonal and L is unit lower bidiagonal. This factorization can be computed in just $O(n)$ operations and Sylvester's law of inertia tells us that $T - xI$ and D have the same inertia, so the number of negative diagonal elements of D equals the number of eigenvalues of $T - xI$ less than 0, which is the number of eigenvalues of T less than x, that is, $N(x)$. As there is no pivoting in the factorization it might be thought that this approach would be numerically unstable (and it would be unstable if our aim was to solve a linear system with the factorization), but as a means of determining the diagonal of D it can be shown to be perfectly stable [9, Lem. 5.3].

Sylvester's law says nothing about the magnitudes of the eigenvalues after a congruence transformation. Ostrowski showed that [21, Thm. 4.5.9]

$$\lambda_k(X^*AX) = \theta_k \lambda_k(A),$$

where $\lambda_n(X^*X) \le \theta_k \le \lambda_1(X^*X)$, where the eigenvalues are ordered $\lambda_n \le \lambda_{n-1} \le \cdots \le \lambda_1$. This quantitative result is useful for developing minimal perturbations of a matrix that change its inertia in a specified way [18].

6 Functions of Matrices

In his 1858 memoir [5] Cayley treated square roots of 2×2 and 3×3 matrices, and he later revisited these cases in 1872 [6]. The first formula for a general function of a matrix was given by Sylvester in 1883 [33], for an $n\times n$ matrix A with distinct eigenvalues λ_i:

$$f(A) = \sum_{i=1}^{n} f(\lambda_i) \prod_{j\ne i} \frac{A-\lambda_j I}{\lambda_i - \lambda_j}. \qquad (7)$$

Today we call this the interpolating polynomial definition of $f(A)$, as the formula says that $f(A) = p(A)$ where p is the unique polynomial of degree at most $n-1$ that interpolates to f at the eigenvalues of A. The particular expression in (7) for p is called the Lagrange interpolating polynomial, or sometimes the Sylvester interpolating polynomial, although Lagrange's 1795 publication of the formula predates Sylvester's.

Buchheim[2] gave a derivation of (7) [3] and then generalized it to multiple eigenvalues using Hermite interpolation [4], thereby giving the first completely general definition of a matrix function.

Sylvester's interpolating polynomial definition of $f(A)$ is useful theoretically, but it is rarely used for computation. However, the best method for computing a general function of an $n \times n$ matrix, the Schur–Parlett method [8] (implemented in MATLAB as `funm`), has a strong Sylvester connection. The method begins by computing the Schur decomposition $A = QTQ^*$, where Q is unitary and T upper triangular. Then it carries out some further unitary similarity transformations to produce a new triangular matrix U with a partitioning $U = (U_{ij})$ in which the different diagonal blocks U_{ii} have no eigenvalues in common. The matrix $G = f(U)$ will be triangular, like U, and the diagonal blocks are $G_{ii} = f(U_{ii})$, which are computed by a truncated Taylor series (or any available method specific to f). The off-diagonal blocks G_{ij} are computed by solving a sequence of Sylvester equations

$$U_{ii}G_{ij} - G_{ij}U_{jj} = G_{ii}U_{ij} - U_{ij}G_{jj} + \sum_{k=i+1}^{j-1} (G_{ik}U_{kj} - U_{ik}G_{kj}), \quad i<j,$$

which are derived from the relation $GU = UG$. Finally, $F = f(A)$ is obtained by undoing the unitary transformations. Given its reliance on Sylvester equations this method should perhaps be more accurately called the "Schur–Parlett–Sylvester" method.

7 The Sylvester Resultant Matrix

An important problem arising in computational geometry is to find the intersection of algebraic and parametric (e.g., B-spline and Bézier) curves, and this reduces to determining whether two polynomials have a common root. Suppose the polynomials are

$$p(x) = a_n x^n + \cdots + a_1 x + a_0 = a_n \prod_{i=1}^{n} (x-\alpha_i),$$

$$q(x) = b_m x^m + \cdots + b_1 x + b_0 = b_m \prod_{i=1}^{m} (x-\beta_i),$$

with $a_n \ne 0$ and $b_m \ne 0$. A resultant is a scalar function of the coefficients a_i and b_i that is zero if and only if p and q have a common root. A resultant matrix is a matrix whose determinant is a resultant.

The Sylvester resultant matrix [28] is the matrix of dimension $m+n$,

$$S(p,q) = \left[\begin{array}{cccccc} a_n & a_{n-1} & \dots & a_1 & a_0 & \\ & a_n & a_{n-1} & \dots & a_1 & a_0 \\ & & \ddots & \ddots & \ddots & \ddots \\ & & & a_n & a_{n-1} \dots & a_1 \; a_0 \\ \hline b_m & b_{m-1} & \dots & b_1 & b_0 & \\ & b_m & b_{m-1} & \dots & b_1 & b_0 \\ & & \ddots & \ddots & \ddots & \ddots \\ & & & b_m & b_{m-1} \dots & b_1 \; b_0 \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c}a\\a\\a\\a\end{array}}\right\} m \text{ rows} \\ \left.\vphantom{\begin{array}{c}a\\a\\a\\a\end{array}}\right\} n \text{ rows} \end{array}.$$

It can be shown that

$$\det(S(p,q)) = a_n^m b_m^n \prod_{i=1}^{n} \prod_{j=1}^{m} (\alpha_i - \beta_j), \qquad (8)$$

[2] Buchheim (1859–1888), who had studied with Henry Smith at Oxford and Felix Klein at Leipzig, was a teacher at Manchester Grammar School [17].

which confirms that the Sylvester matrix is indeed a resultant matrix. Moreover, it turns out that the dimension of the null space of S is equal to the number of common zeros of p and q and the greatest common divisor of p and q can be read off from the echelon form of S.

Sylvester's is not the only resultant matrix. Another is that of Bézout, called the Bezoutian matrix or simply the Bezoutian (the accent on the "e" is usually omitted), and was so-named by Sylvester [32].

Functions to generate the Sylvester matrix can be found in software such as MATLAB (`linalg::sylvester` in MuPAD in the Symbolic Math Toolbox) and Maple (`SylvesterMatrix`).

8 Coda

Sylvester has left a large legacy in linear algebra and numerical analysis. Much of his terminology is still used. His inertia result is fundamental to undergraduate linear algebra courses, his expression for a matrix function is one of the standard definitions, the Sylvester matrix is widely used in computer algebra, and his linear and quadratic matrix equations and their variants arise in many applications and are the subject of ongoing research.

Sylvester's friend Cayley initiated the study of matrix theory and discovered the famous Cayley–Hamilton theorem. But he published little on the subject after that and Sylvester's name is much more commonly encountered today in matrix analysis and linear algebra.

Sylvester's style of mathematics was less rigorous than that of the Berlin school of Frobenius, Kronecker, and Weierstrass, and this has led some to downplay his role in the development of matrix theory (see, for example [11], [12], and the recent book [13]). However, as this article shows, Sylvester had a remarkable knack for identifying and naming key concepts and for making discoveries that would turn out to have lasting mathematical importance and practical relevance.

A theme running through a 1991 article by McIntyre on the programming language APL [23] is that APL provides a tool that enables us to think and work with matrices instead of scalars, just "as J. J. Sylvester so eloquently urged us to do a century ago." Over the last thirty years there has been a trend towards matrix-based computation, with languages and problem solving environments such as Fortran 90, Julia, MATLAB, Maple, Mathematica, Python (with SciPy and SymPy), and R all exploiting the power of matrices. Sylvester's influence therefore lives on not only in mathematics but also in the tools we use for computation.

Finally, I note that the four volumes of Sylvester's collected works are available in PDF form at https://archive.org, specifically https://archive.org/details/SylvesterCollected2 and related URLs, and these are invaluable for anyone who wishes to consult his original papers.

Bibliographic Note

In the electronic (PDF) version of this article the reference list contains hyperlinks from the paper title to the paper itself whenever the paper has a digital object identifier (DOI). Some of the link targets require a subscription for access.

References

[1] Awad H. Al-Mohy, Nicholas J. Higham, and Samuel D. Relton. Computing the Fréchet derivative of the matrix logarithm and estimating the condition number. *SIAM J. Sci. Comput.*, 35(4):C394–C410, 2013.

[2] Rajendra Bhatia and Peter Rosenthal. How and why to solve the operator equation $AX - XB = Y$. *Bull. London Math. Soc.*, 29: 1–21, 1997.

[3] A. Buchheim. On the theory of matrices. *Proc. London Math. Soc.*, 16:63–82, 1884.

[4] A. Buchheim. An extension of a theorem of Professor Sylvester's relating to matrices. *Phil. Mag.*, 22(135):173–174, 1886. Fifth series.

[5] Arthur Cayley. A memoir on the theory of matrices. *Philos. Trans. Roy. Soc. London*, 148:17–37, 1858.

[6] Arthur Cayley. On the extraction of the square root of a matrix of the third order. *Proc. Roy. Soc. Edinburgh*, 7:675–682, 1872.

[7] Tony Crilly. *Arthur Cayley: Mathematician Laureate of the Victorian Age*. Johns Hopkins University Press, Baltimore, MD, USA, 2006. xxi+610 pp. ISBN 0-8018-8011-4.

[8] Philip I. Davies and Nicholas J. Higham. A Schur–Parlett algorithm for computing matrix functions. *SIAM J. Matrix Anal. Appl.*, 25(2):464–485, 2003.

[9] James W. Demmel. *Applied Numerical Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1997. xi+419 pp. ISBN 0-89871-389-7.

[10] Sven Hammarling. *Latent Roots and Latent Vectors*. The University of Toronto Press, Toronto, Canada, 1970. x+172 pp. ISBN 0-8020-1709-6.

[11] Thomas Hawkins. Another look at Cayley and the theory of matrices. *Arch. Internat. Histoire Sci.*, 27(100):82–112, 1977.

[12] Thomas Hawkins. Weierstrass and the theory of matrices. *Archive for History of Exact Sciences*, 12(2):119–163, 1977.

[13] Thomas Hawkins. *The Mathematics of Frobenius in Context. A Journey Through 18th to 20th Century Mathematics*. Springer-Verlag, New York, 2013. xiii+699 pp. ISBN 978-1-4614-6332-0.

[14] Harold V. Henderson, Friedrich Pukelsheim, and Shayle R. Searle. On the history of the Kronecker product. *Linear and Multilinear Algebra*, 14(2): 113–120, 1983.

[15] Nicholas J. Higham. Cayley, Sylvester, and early matrix theory. *Linear Algebra Appl.*, 428:39–43, 2008.

[16] Nicholas J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008. xx+425 pp. ISBN 978-0-898716-46-7.

[17] Nicholas J. Higham. Arthur Buchheim (1859-1888). `http://nickhigham.wordpress.com/2013/01/31/arthur-buchheim`, 2013.

[18] Nicholas J. Higham and Sheung Hun Cheng. Modifying the inertia of matrices arising in optimization. *Linear Algebra Appl.*, 275–276:261–279, 1998.

[19] Nicholas J. Higham and Hyun-Min Kim. Numerical analysis of a quadratic matrix equation. *IMA J. Numer. Anal.*, 20(4):499–519, 2000.

[20] Nicholas J. Higham and Lijing Lin. An improved Schur–Padé algorithm for fractional powers of a matrix and their Fréchet derivatives. *SIAM J. Matrix Anal. Appl.*, 34(3): 1341–1360, 2013.

[21] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Second edition, Cambridge University Press, Cambridge, UK, 2013. xviii+643 pp. ISBN 978-0-521-83940-2.

[22] I. M. James. James Joseph Sylvester, F.R.S. (1814–1897). *Notes and Records Roy. Soc. London*, 51(2):247–261, 1997.

[23] D. B. McIntyre. Language as an intellectual tool: from hieroglyphics to APL. *IBM Syst. J.*, 30(4):554–581, 1991.

[24] The Oxford English dictionary. `http://www.oed.com`. Accessed March 22, 2014.

[25] Karen Hunger Parshall. *James Joseph Sylvester. Jewish Mathematician in a Victorian World*. Johns Hopkins University Press, Baltimore, MD, USA, 2006. xiii+461 pp. ISBN 0-8018-8291-5.

[26] Catherine E. Powell and Howard C. Elman. Block-diagonal preconditioning for spectral stochastic finite-element systems. *IMA J. Numer. Anal.*, 29(2):350–375, 2009.

[27] George Salmon. *Lessons Introductory to the Modern Algebra*. Hodges, Smith, and Co., Dublin, 1859. xi+147 pp.

[28] J. J. Sylvester. A method of determining by mere inspection the derivatives from two equations of any degree. *Philosophical Magazine Series 3*, 16(101):132–135, 1840.

[29] J. J. Sylvester. Additions to the articles, "On a New Class of Theorems," and "On Pascal's Theorem". *Philosophical Magazine*, 37:363–370, 1850.

[30] J. J. Sylvester. Sketch of a memoir on elimination, transformation, and canonical forms. *Cambridge and Dublin Mathematical Journal*, 6:186–200, 1851.

[31] J. J. Sylvester. A demonstration of the theorem that every homogeneous quadratic polynomial is reducible by real orthogonal substitutions to the form of a sum of positive and negative squares. *Philosophical Magazine*, (Fourth Series) 4:138–142, 1852.

[32] J. J. Sylvester. On a theory of the syzygetic relations of two rational integral functions, comprising an application to the theory of Sturm's functions, and that of the greatest algebraical common measure. *Phil. Trans. R. Soc. Lond.*, 143:407–548, 1853.

[33] J. J. Sylvester. On the equation to the secular inequalities in the planetary theory. *Philosophical Magazine*, 16:267–269, 1883.

[34] J. J. Sylvester. Lectures on the principles of universal algebra. *Amer. J. Math.*, 6(1):270–286, 1883–1884.

[35] J. J. Sylvester. On Hamilton's quadratic equation and the general unilateral equation in matrices. *Philosophical Magazine*, 18:454–458, 1884.

[36] J. J. Sylvester. Sur la résolution générale de l'équation linéaire en matrices d'un ordre quelconque. *Comptes Rendus de l'Académie des Sciences*, 99:409–412, 1884.

[37] J. J. Sylvester. Sur l'equation en matrices $px = xq$. *Comptes Rendus de l'Académie des Sciences*, 99:67–71 and 115–116, 1884.

[38] J. J. Sylvester. Sur les quantités formant un groupe de nonions analogues aux quaternions de Hamilton. *Comptes Rendus de l'Académie des Sciences*, 98:273–276, 471–475, 1884.

[39] J. J. Sylvester. On the trinomial unilateral quadratic equation in matrices of the second order. *Quarterly Journal of Mathematics*, 20:305–312, 1885.

[40] Françoise Tisseur and Karl Meerbergen. The quadratic eigenvalue problem. *SIAM Rev.*, 43(2):235–286, 2001.