ΓΡΑΜΜΙΚΗ ΑΛΓΕΒΡΑ

Διάλεξη 12

Η Ορίζουσα

Ε. Γαλλόπουλος & Ε. Στεφανόπουλος

Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Πατρών

9 Δεκεμβρίου 2022

Υπενθύμιση και πρόγραμμα διάλεξης

1. Τελεστές προβολής και βέλτιστη προσέγγιση από υπόχωρο
2. Το πρόβλημα των ελαχίστων τετραγώνων
3. Παραδείγματα προσεγγίσεων ελαχίστων τετραγώνων.
4. Ελάχιστα τετράγωνα μέσω διαφορικού λογισμού.
5. Από την ΟΚ Gram-Schmidt στην παραγοντοποίηση QR και επίλυση του πρβλ. των ΕλΤετρ.

Σήμερα θα συζητήσουμε:

1. Ορισμένες ενδιαφέρουσες βαθμωτές συναρτήσεις μητρώων,
2. όπως το ίχνος
3. με έμφαση στην ορίζουσα,
4. αξιωματικός ορισμό της οριζουσας,
5. τύποι και τρόποι υπολογισμού της,
6. ιδιότητες και γεωμετρική ερμηνεία,
7. επίλυση συστήματος και αντιστροφή μητρώου μέσω οριζουσών.

Ορθοκανονικοποιήσεις & παραγοντοποιήσεις: Από τη Gram-Schmidt στην QR I

Υπενθυμίζουμε την GS:

$$\begin{aligned}
\rho_{11} q_1 &= q_1, \quad \rho_{11} = \|a_1\| \\
\rho_{22} q_2 &= a_2 - P_{q_1} a_2 = a_2 - q_1(q_1^\top a_2), \quad \rho_{12} = q_1^\top a_2, \rho_{22} \text{ τ.ώ. } \|q_2\| = 1, \\
\rho_{33} q_3 &= a_3 - P_{q_1} a_3 - P_{q_2} q_3 = a_2 - q_1(q_1^\top a_3) - q_2(q_2^\top a_3), \\
&\qquad \rho_{13} = q_1^\top a_3, \rho_{23} = q_2^\top a_3, \rho_{33} \text{ τ.ώ. } \|q_3\| = 1.
\end{aligned}$$

$$\begin{aligned}
[a_1, a_2, a_3] &= [q_1, q_2, q_3] \begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ 0 & \rho_{22} & \rho_{23} \\ 0 & 0 & \rho_{33} \end{pmatrix} \\
A &= QR. \text{ Το } R \text{ είναι άνω τριγωνικό. Το } Q \text{ έχει ορθογώνιες στήλες, δηλ. } Q^\top Q = I.
\end{aligned}$$

Υπενθυμίζουμε την GS:

$$\begin{aligned}
\rho_{11} q_1 &= a_1, \quad \boxed{\rho_{11} = \|a_1\|} \\
\rho_{22} q_2 &= a_2 - \rho_{12} q_1, \quad \boxed{\rho_{12} = q_1^\top a_2}, \boxed{\rho_{22} \text{ τ.ώ. } \|q_2\| = 1}, \\
\rho_{33} q_3 &= a_2 - q_1 \rho_{13} - q_2 \rho_{23}, \\
& \qquad \boxed{\rho_{13} = q_1^\top a_3}, \boxed{\rho_{23} = q_2^\top a_3}, \boxed{\rho_{33} \text{ τ.ώ. } \|q_3\| = 1}.
\end{aligned}$$

$$\begin{aligned}
[a_1, a_2, a_3] &= [q_1, q_2, q_3] \begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ 0 & \rho_{22} & \rho_{23} \\ 0 & 0 & \rho_{33} \end{pmatrix} \\
A &= QR.\ \text{Το } R \text{ είναι άνω τριγωνικό. Το } Q \text{ έχει ορθογώνιες στήλες, δηλ. } Q^\top Q = I.
\end{aligned}$$

```matlab
function [Q,R]=QR_GS (A);
% Simple "economy QR" factorization by means of Gram-Schmidt
[m,n]=size(A);
R(1,1) = norm(A(:,1)); Q(:,1)= A(:,1)/R(1,1);
for j=2:n
      Q(:,j)=A(:,j);
      for i=1:j-1
            R(i,j)=Q(:,i)'*A(:,j);
            Q(:,j)=Q(:,j)-R(i,j)*Q(:,i);
      end
R(j,j)=norm(Q(:,j));
Q(:,j)=Q(:,j)/R(j,j);
end
```

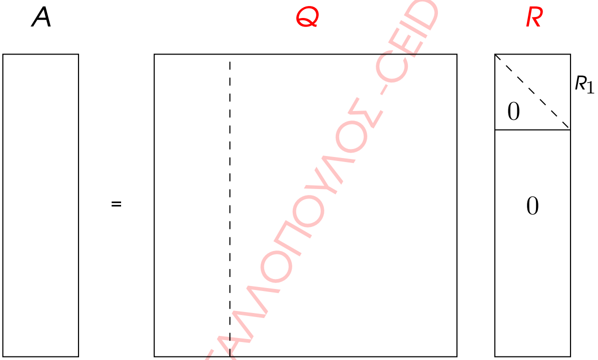
```
A =
3       1       3
2       1       3
2       4       2
4       3       2
(Q,R)= QR_GS(A);
>> Q
Q =
0.5222    -0.4483    0.2404
0.3482    -0.1814    0.6772
0.3482     0.8752    0.2556
0.6963    -0.0107   -0.6467
R =
5.7446     4.3519    4.7001
0      2.8391    -0.1601
0              0     1.9706
ans =
1.0e-15 *
0              0              0
0              0              0
0              0              0
0              0     0.2220
```

Παραγοντοποίηση QR

Θεώρημα

Κάθε μητρώο $A \in \mathbb{R}^{m\times n}$, $m \geq n$, μπορεί να μετασχηματιστεί σε άνω τριγωνική μορφή μέσω ορθογώνιου μετασχηματισμού. Ειδικότερα, υπάρχει παραγοντοποίηση $A = Q\begin{pmatrix} R_1 \\ 0 \end{pmatrix}$ όπου $Q \in \mathbb{R}^{m\times m}$ ορθογώνιο και $R = \begin{pmatrix} R_1 \\ 0 \end{pmatrix} \in \mathbb{R}^{m\times n}$ άνω τριγωνικό. Αν οι στήλες του A είναι γραμμικά ανεξάρτητες, τότε το $R_1 \in \mathbb{R}^{n\times n}$ είναι αντιστρέψιμο. Αν επιλέξουμε τα διαγώνια στοιχεία του R να είναι θετικά, τότε οι παράγοντες Q, R είναι μοναδικοί.

- Μια από τις πιο σημαντικές παραγοντοποιήσεις μητρώου, με πολλές εφαρμογές.
- Υπολογιστικός πυρήνας για
  1. τη λύση του προβλήματος ελαχίστων τετραγώνων.
  2. την προσέγγιση ιδιοτιμών/ιδιοδιανυσμάτων (μέθοδος QR)



A
=
Q
R
R_1
0
0

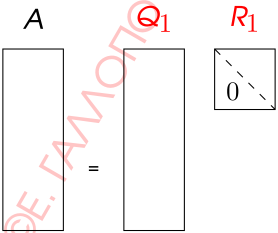
Παράδειγμα

```
A =

    1.0000    0.5000
    0.5000    0.3333
    0.3333    0.2500
    0.2500    0.2000

>> [Q,R]=qr(A); [Q,R]

ans =

   -0.8381    0.5226    0.0876    0.1293   -1.1932   -0.6705   R1
   -0.4191   -0.4417   -0.5883   -0.5322         0   -0.1185
   -0.2794   -0.5288    0.7763   -0.1992         0         0
   -0.2095   -0.5021   -0.2089    0.8126         0         0   0
```

Λεπτή (ή οικονομική) παραγοντοποίηση QR

$$A = [Q_1, Q_2] \begin{pmatrix} R_1 \\ 0 \end{pmatrix} = Q_1 R_1$$

Η λεπτή QR υπολογίζει $A = Q_1 R_1$ όπου $Q_1 \in \mathbb{R}^{m \times n}$ έχει ορθοκανονικές στήλες και το R_1 είναι τετραγωνικό, άνω τριγωνικό.



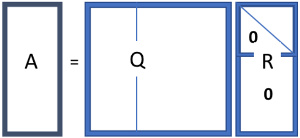
- Οι στήλες του Q_1 είναι ΟΚ βάση του χώρου στηλών ($\mathrm{range}(A)$).
- Το $R_1^\top$ είναι ο παράγοντας Cholesky του $A^\top A$.

Παράδειγμα (το A όπως πριν)

```
>> [Q,R]=qr(A,0)

Q =

   -0.8381    0.5226
   -0.4191   -0.4417
   -0.2794   -0.5288
   -0.2095   -0.5021


R =

   -1.1932   -0.6705
         0   -0.1185
```

Σύνοψη για την QR

WHAT



QR factorization

HOW

- Householder reflections
- Givens rotations
- Gram-Schmidt orthogonalization

- QR with pivoting
- QR without square roots
- Rank revealing QR
- QR for matrices with special structure
- Updating (recursive least squares)

- Mixed precision
- square-root-less computations

WHY

- stable parallel linear system solver
- least squares solvers, linear regression
- eigenvalue computations (QR algorithm)
- obtaining orthogonal bases for fundamental subspaces
- dimensionality reduction
- Low rank factorization
- cheaper alternative to SVD
- linear discriminant analysis
- leverage scores computation
- filtering

Χρήσεις και εφαρμογές

Applications of Matrix Decompositions for Machine Learning

Prince Grover Follow
Jul 25, 2018 · 9 min read

In this post, we'll learn how we can solve a lot of machine learning problems using our old math friend: matrix decompositions.

LINEAR REGRESSION, WITH MAP-REDUCE

18/06/2018 ARTHUR CHARPENTIER 4 COMMENTS

Sometimes, with big data, matrices are too big to handle, and it is possible to use tricks to numerically still do the map. Map-Reduce is one of those. With several cores, it is possible to split the problem, tc map on each machine, and then to agregate it back at the end.

Consider the case of the linear regression, $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ (with classical matrix notations). The OLS estimate of β is $\widehat{\beta} = [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{y}$. To illustrate, consider a not too big dataset, and run some regression.

Direct QR factorizations for tall-and-skinny matrices in MapReduce architectures

3 Author(s) Austin R. Benson ; David F. Gleich ; James Demmel View All Authors

15 Paper Citations 250 Full Text Views

Abstract

Document Sections

I. Introduction

II. Indirect QR factorizations in MapReduce

III. Direct QR Factorizations in MapReduce

IV. Performance Experiments

V. Conclusion

Abstract:
The QR factorization and the SVD are two fundamental matrix decompositions with applications throughout scientific computing and data analysis. For matrices with many more rows than columns, so-called "tall-and-skinr for computing the Ql computing environm decomposition use ε best case, these me compute a stable tal 2 passes over the da performance. We pr TSQR methods that implementation for M unstable methods fo performance model

The QR Decomposition For Regression Models

Michael Betancourt
July 2017

A common problem in regression modeling is correlation amongst the covariates which can induce strong posterior correlations that frustrate accurate computation. In this case study I will review the *QR decomposition*, a technique for decorrelating covariates and, consequently, the resulting posterior distribution.

Neural network algorithms based on the QR decomposition method of least squares

2 Author(s) T. Ogunfunmi ; Z. Chen View All Authors

71 Full Text Views

Abstract
Authors
References
Keywords
Metrics

Abstract:
We present a set of algorithms for feed-forward multilayer neural networks based on the QR and the inverse-QR recursive least-squares algorithms. These algorithms possess excellent numerical stability, fast convergence characteristics compared to the backpropagation algorithm and require much fewer iterations to train the neural networks. We apply these algorithms to practical problems of pattern recognition of different patterns and also for optimization with excellent results. We compare these algorithms with the previously reported ones which are also based on the least squares method and found the one based on the inverse QR method to be superior to the others. The computational complexity comparison of these algorithms is also presented.<>

UNIVERSITY OF OXFORD

QR Decomposition of Normalised Relational Data

Bas A.M. van Geffen
Kellogg College

IDR/QR: an incremental dimension reduction algorithm via QR decomposition.

January 2004
DOI: 10.1145/1014052.1014093
Source · DBLP
Conference: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004

Jieping Ye · Qi Li · Hui Xiong · Show all 6 authors · Vipin Kumar

ods and matrix decompositions have been studied as a tra complexity, stability and accuracy. In the first part of this t of state-of-the-art QR Decomposition methods applied to m is done. Stability and accuracy of these methods are analyz

KTH Electrical Engineering

IEEE TRANSACTIONS ON COMPUTERS, VOL. C-26, NO. 2, FEBRUARY 1977 147

A Parallel *QR* Algorithm for Symmetric Tridiagonal Matrices

AHMED H. SAMEH AND DAVID J. KUCK, MEMBER, IEEE

Algorithm 844: Computing Sparse Reduced-Rank Approximations to Sparse Matrices

MICHAEL W. BERRY and SHAKHINA A. PULATOVA
University of Tennessee, Knoxville
and
G. W. STEWART
University of Maryland, College Park

In many applications—latent semantic indexing, for example—it is required to obtain a reduced rank approximation to a sparse matrix A. Unfortunately, the approximations based on traditional decompositions, like the singular value and QR decompositions, are not in general sparse. Stewart [(1999), 313–323] has shown how to use a variant of the classical Gram–Schmidt algorithm, called the quasi–Gram-Schmidt–algorithm, to obtain two kinds of low-rank approximations. The first, the SPQR, approximation, is a pivoted, Q-less QR approximation of the form $(XR_{11}^{-1})(R_{11}\ R_{12})$, where X consists of columns of A. The second, the SCR approximation, is of the form the form $A \cong XTY^T$, where X and Y consist of columns and rows A and T, is small. In this article we treat the computational details of these algorithms and describe a MATLAB implementation.

On Stable Parallel Linear System Solvers

A. H. SAMEH AND D. J. KUCK

University of Illinois, Urbana, Illinois

ABSTRACT In this paper three stable parallel algorithms for solving dense and tridiagonal systems of linear equations are discussed The algorithms are based on Givens' reduction of a matrix to the upper triangular form The algorithm for the dense case requires $O(n)$ time steps compared to $O(n \log n)$ steps for Gaussian elimination with pivoting (in the absence of certain features of machine logic and hardware) For the tridiagonal case, one of the algorithms presented here is superior to the best previous algorithm in that with a modest increase in time it does not fail if any of the leading principal submatrices is singular, the probability of over- or underflow is minimized, and the error bound does not grow exponentially Furthermore, it is most suitable when only a limited number of processors is available.

A study of QR decomposition and Kalman filter implementations

ılman Filter implementations are discussed. The EKF is known to be numerically unstable and various methods have been proposed in the literature improve the performance of the filter. These methods include square-root d unscented versions of the filter that make use of numerical methods such QR, LDL and Cholesky Decomposition. At the end of the analysis, the dience/reader will get some idea about best implementation of the filter ven some specifications.

Χρήσεις και εφαρμογές

Recycling Givens rotations for the efficient approximation of pseudospectra of band-dominated operators

Marko Lindner* and Torge Schmidt†

Contents lists available at ScienceDirect
Parallel Computing
journal homepage: www.elsevier.com/locate/parco

A direct tridiagonal solver based on Givens rotations for GPU architectures

I.E. Venetis[a,*], A. Kouris[a], A. Sobczyk[a], E. Gallopoulos[a], A.H. Sameh[b]
[a] Computer Engineering and Informatics Department, University of Patras, Greece
[b] Computer Science Department, Purdue University, West Lafayette, IN 47907, USA

A 3D Parallel Algorithm for QR Decomposition

Grey Ballard, Wake Forest University, Winston Salem, NC, USA, ballard@wfu.edu
James Demmel, University of California, Berkeley, CA, USA, demmel@berkeley.edu
Laura Grigori, INRIA Paris-Rocquencourt, Paris, France, laura.grigori@inria.fr
Mathias Jacquelin, Lawrence Berkeley Natl. Lab., Berkeley, CA, USA, mjacquelin@lbl.gov
Nicholas Knight, New York University, New York, NY, USA, nknight@nyu.edu

A SQUARE ROOT AND DIVISION FREE GIVENS ROTATION FOR SOLVING LEAST SQUARES PROBLEMS ON SYSTOLIC ARRAYS*

J. GÖTZE† AND U. SCHWIEGELSHOHN‡

Grey Ballard, James Demmel, Laura Grigori, Mathias Jacquelin, Nicholas Knight, H. D. Nguyen:
Reconstructing Householder vectors from Tall-Skinny QR. J. Parallel Distrib. Comput. 85: 3-31 (2015)

Efficient Realization of Givens Rotation through Algorithm-Architecture Co-design for Acceleration of QR Factorization

Farhad Merchant, Tarun Vatwani, Anupam Chattopadhyay, *Senior Member, IEEE,* Soumyendu Raha, S K Nandy, *Senior Member, IEEE,*, Ranjani Narayan, and Rainer Leupers

Abstract—We present efficient realization of Generalized Givens Rotation (GGR) based QR factorization that achieves 3-100x better performance in terms of Gflops/watt over state-of-the-art realizations on multicore, and General Purpose Graphics Processing Units (GPGPUs). GGR is an improvement over classical Givens Rotation (GR) operation that can annihilate multiple elements of rows and columns of an input matrix simultaneously. GGR takes 33% lesser multiplications compared to GR. For custom implementation of GGR, we identify macro operations in GGR and realize them on a Reconfigurable Data-path (RDP) tightly coupled to pipeline of a Processing Element (PE). In PE, GGR attains speed-up of 1.1x over Modified Householder Transform (MHT) presented in the literature. For parallel realization of GGR, we use REDEFINE, a scalable massively parallel Coarse-grained Reconfigurable Architecture, and show that the speed-up attained is commensurate with the hardware resources in REDEFINE. GGR also outperforms General Matrix Multiplication (gemm) by 10% in-terms of Gflops/watt which is counter-intuitive.

A GPU-Accelerated SVD Algorithm, Based on QR Factorization and Givens Rotations, for DWI Denoising

2 Author(s) Livia Marcellino ; Guglielmo Navarra View

QRD-BASED SQUARE ROOT FREE AND DIVISION FREE ALGORITHMS AND ARCHITECTURES

K.J.R. Liu E. Frantzeskakis
Electrical Engr. Dept.
Institute of Systems Research
University of Maryland
College Park, Maryland

Parallel Computing
Volume 40, Issue 7, July 2014, Pages 161-172

Implementing QR factorization updating algorithms on GPUs

CUDA Pro Tip: Fast and Robust Computation of Givens Rotations

Επίλυση ελαχίστων τετραγώνων με οικονομική QR

Αν $A \in \mathbb{R}^{m\times n}$ όπου $m > n$ τότε η οικονομική παραγοντοποίηση QR μας επιστρέφει $Q_1 \in \mathbb{R}^{m\times n}$ και $R_1 \in \mathbb{R}^{n\times n}$.

Υποθέτουμε ότι A είναι πλήρους τάξης, δηλ. $\mathrm{rank}(A) = n$.

Τότε το R_1 θα είναι αντιστρέψιμο.

Απευθείας από την οικονομική παραγοντοποίηση $A = Q_1 R_1$ προκύπτει ότι η βέλτιστη λύση ελαχίστων τετραγώνων είναι

$$\begin{aligned} x_{\mathrm{LS}} &= (A^\top A)^{-1} A^\top b = (R_1^\top Q_1^\top Q_1 R_1)^{-1} R_1^\top Q_1^\top b \\ &= R_1^{-1} Q_1^\top b \end{aligned}$$

και μπορεί να υπολογιστεί άμεσα από τους παράγοντες Q_1, R_1.

Επίσης Q_1 Q_1^T (b-Ax_LS) = 0 καθώς το υπόλοιπο είναι κάθετο στο range(A)=range(Q_1)

Επομένως, ακολουθεί άμεσα η παραπάνω σχέση για το x_LS (υ

Ανάλυση με βάση την παραγοντοποίηση QR

Αν η πλήρης παραγοντοποίηση είναι $A = QR$, τότε επειδή το $Q = \left[Q_1, Q_2\right]$ είναι ορθογώνιο,

$$\begin{aligned} \|b - Ax\|_2 &= \|Q^\top (b - Ax)\|_2 \\ &= \|Q^\top b - Q^\top Ax\|_2 \\ &= \|Q^\top b - Rx\|_2 \end{aligned}$$

Έστω ότι $Q^\top b = \begin{pmatrix} Q_1^\top b \\ Q_2^\top b \end{pmatrix}$ και εφόσον $Rx = \begin{pmatrix} R_1 x \\ 0 \end{pmatrix}$ τότε

$$\|Ax - b\|_2^2 \quad = \quad \|Q^\top Ax - Q^\top b\|_2^2 = \|\begin{pmatrix} R_1 x - Q_1^\top b \\ Q_2^\top b \end{pmatrix}\|_2 = \|R_1 x - Q_1^\top b\|_2^2 + \|Q_2^\top b\|_2^2$$

Η βέλτιστη λύση είναι το διάνυσμα x_{LS} που προκύπτει από την επίλυση του άνω τριγωνικού συστήματος $R_1 x = Q_1^\top b$. Ο όρος $\|Q_2^\top b\|_2$ δεν μπορεί να μειωθεί περισσότερο και είναι το σφάλμα της προσέγγισης (κατάλοιπο), δηλ. $\min_{x \in \mathbb{R}^n} \|Ax_{\mathrm{LS}} - b\|_2 = \|Q_2^\top b\|_2$.

Σχετικά με την επίλυση του προβλήματος των ελαχίστων τετραγώνων μέσω QR και Gram-Schmidt I

Η βέλτιστη λύση αντιστοιχεί σε ορθογώνια προβολή επί του χώρου στηλών του .
Όμως $\text{Range}(A) = \text{Range}(Q_1)$ επομένως $(b - Ax) \perp \text{Range}(Q_1)$ επομένως

$$Q_1 Q_1^\top (b - Ax_{\text{LS}}) = 0 \Rightarrow Q_1^\top (b - Ax_{\text{LS}}) = 0 \Rightarrow Q_1^\top b = Q_1 R x_{\text{LS}}$$

Ειδικές τιμές μητρώων

Παρόλο που ένα μητρώο $A \in \mathbb{R}^{m\times n}$ περιέχει $m \times n$ αριθμούς, σε αυτό αντιστοιχούν επίσης ορισμένες ειδικές τιμές συνδεδεμένες με αυτό. Αυτές είναι βαθμωτές συναρτήσεις μητρώου ($f : \mathbb{R}^{n\times n}\mathbb{R}$), και αποκαλύπτουν σημαντικές πληροφορίες για το μητρώο.

είδος	*συμβολισμός*	*παρατηρήσεις*
διαστάσεις	m, n	αριθμ. γραμμών και αριθμ. στηλών
αραιότητα	$\text{nnz}(A)$	πλήθος μη μηδενικών
τάξη	$\text{rank}(A)$	μέγιστος αριθμ. γ.α. γραμμών (ή στηλών)
νόρμα	$\|A\|$	μετρική ((βάρους)) , π.χ. $\sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}$
ίχνος	$\text{trace}(A)$	$(m = n)$???
ορίζουσα	$\det(A)$	$(m = n)$???
ιδιοτιμές	???	$(m = n)$???
ιδιάζουσες τιμές	???	???

- Χαρακτηριστική τιμή για κάθε τετραγωνικό μητρώο ίση με το άθροισμα των διαγώνιων στοιχείων του.
- Βαθμωτή συνάρτηση

$$\mathrm{trace} : \mathbb{R}^{n\times n} \to \mathbb{R}$$

τέτοια ώστε

$$\mathrm{trace}(A) := \alpha_{1,1} + \alpha_{2,2} + \cdots + \alpha_{n,n}$$

Ιδιότητες

1. $\mathrm{trace}(\gamma A + \delta B) = \gamma\mathrm{trace}(A) + \delta\mathrm{trace}(B)$
2. $\mathrm{trace}(AB) = \mathrm{trace}(BA)$

Σημαντικό επακόλουθο: $\mathrm{trace}(X^{-1}AX) = \mathrm{trace}(A)$

- Μοναδικό χαρακτηριστικό μέγεθος (βαθμωτός) που υπάρχει για κάθε τετραγωνικό μητρώο.

$$\det : \mathbb{R}^{n\times n} \to \mathbb{R}$$

- Συμπυκνώνει πληροφορίες σχετικά με το μητρώο.
- Η πιο γνωστή: Ένα μητρώο A είναι αντιστρέψιμο αν και μόνον αν $\det(A) \neq 0$.
- Υπολογιστικά θέματα: Ο πιο πρακτικός τρόπος υπολογισμού του για μέτριο ή μεγάλο n είναι ως παραπροϊόν της παραγοντοποίησης LU με το αντίστοιχο κόστος.
- Γεωμετρική ερμηνεία: όγκος στερεού στο n-διάστατο χώρο.

Ιστορικά

- Η μελέτη των οριζουσών προηγήθηκε της Γραμμικής Άλγεβρας: Ο Gauss (1801) χρησιμοποίησε τον όρο εννοώντας την διακρίνουσα πολυωνύμων 4ου βαθμού. Ο όρος με τη σημερινή έννοια εισήχθη από τον Cauchy (1812).
- Ο Sylvester ((βάπτισε)) τις "μήτρες" για να αναδείξει ότι ((γεννούν)) ορίζουσες.

150 *On a new Class of Theorems.* [25

its form of greatest generality. For this purpose 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

37] *Linearly Equivalent Quadratic Functions.* 247

I have in previous papers defined a "Matrix" as a rectangular array of terms, out of which different systems of determinants may be engendered, as from the womb of a common parent; these cognate determinants being by no means isolated in their relations to one another, but subject to certain simple laws of mutual dependence and simultaneous deperition. The con-

$\underline{n = 2}$

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \Rightarrow \det(A) = \begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix} := \alpha_{11}\alpha_{22} - \alpha_{21}\alpha_{12}.$$

Στη συνέχεια θα χρησιμοποιήσουμε έναν αξιωματικό ορισμό της ορίζουσας από τον οποίο θα προκύψει αυτός και άλλο τύποι υπολογισμού.

<u>Παρατήρηση</u>: Να επαληθεύσετε την παρακάτω παραγοντοποίηση.

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \alpha_{21}\alpha_{11}^{-1} & 1 \end{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ 0 & \alpha_{22} - \alpha_{21}\alpha_{11}^{-1}\alpha_{12} \end{pmatrix}$$

- Έστω δ.χ. $U_1, \ldots, U_n$ και W επί του ιδίου σώματος $\mathcal{F}$ και έστω η απεικόνιση

$$f : U_1 \times U_2 \times \cdots \times U_n \to W$$

Η απεικόνιση ονομάζεται πλειογραμμική (ή n-γραμμική, multilinear) αν είναι γραμμική ως πρός κάθε όρισμα αν τα υπόλοιπα ορίσματα μείνουν αμετάβλητα. Στην επόμενη διαφάνεια θα δείτε ένα παράδειγμα πλειογραμμικής συνάρτησης στην επόμενη διαφάνεια Ο ορισμός θα γίνει καλύτερα αντιληπτός στην χρήση του στην πλειογραμμική συνάρτηση det στην επόμενη διαφάνεια.

- Έστω μια μετάθεση $\pi = (\pi_1, \ldots, \pi_n)$ των αριθμών 1 ως n. Ονομάζουμε φυσική διάταξη την n-άδα $(1, 2, \ldots, n)$. Τότε πρόσημο της μετάθεσης ονομάζεται ο αριθμός

$$\sigma(\pi) = \begin{cases} +1 & \text{αν η } \pi \text{ επανέρχεται στη φυσική διάταξη} \\ & \text{με άρτιο πλήθος εναλλαγών} \\ -1 & \text{αν η } \pi \text{ επανέρχεται στη φυσική διάταξη} \\ & \text{με περιττό πλήθος εναλλαγών.} \end{cases}$$

Ορίζοντας την ορίζουσα (αξιωματικά)

Η ορίζουσα $\det(A)$ ενός τετραγωνικού μητρώου $A \in \mathbb{R}^{n\times n}$ είναι μία συνάρτηση

$$\det : \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}$$

που έχει τις εξής 3 ιδιότητες για ένα μητρώο[α΄] $= [a_1, \cdots, a_n] \in \mathbb{R}^{n\times n}$:

1. Αν $a_i = a_j$ για $i \neq j$ τότε $\det([a_1, \ldots, a_n]) = 0$ (εκφυλισμός).
2. $\det([e_1, \ldots, e_n]) = 1$ (κανονικότητα).
3. Η συνάρτηση είναι γραμμική ως προς τις n στήλες:

$$\begin{aligned}\det([a_1, \ldots, a_{j-1}, \gamma a + \delta b, a_{j+1}, \ldots, a_n]) &= \gamma\det([a_1, \ldots, a_{j-1}, a, a_{j+1}, \ldots, a_n]) \\ &+ \delta\det([a_1, \ldots, a_{j-1}, b, a_{j+1}, \ldots, a_n])\end{aligned}$$

Συναρτήσεις με αυτή την ιδιότητα λέγονται πλειογραμμικές.

[α΄] Αναφερόμαστε στις στήλες, αντίστοιχες διατυπώσεις υπάρχουν και ως προς τις γραμμές).

Σημαντικά επακόλουθα

Έστω το μητρώο (σε μορφή στηλών) $= [a_1, \cdots, a_n] \in \mathbb{R}^{n \times n}$.

- Αν εναλλάξουμε 2 στήλες ενός μητρώου, η ορίζουσα του νέου μητρώου έχει ίδιο μέτρο αλλά αντίθετο πρόσημο. Γι΄ αυτό η ορίζουσα χαρακτηρίζεται και ως εναλλάσσουσα πλειογραμμική απεικόνιση (alternating multilinear map).
- δείγμα απόδειξης

$$
\begin{aligned}
0 &= \det([a_1 + a_2, a_1 + a_2]) = \det([a_1, a_1]) + \det([a_1, a_2]) + \det([a_2, a_1]) + \det([a_2, a_2]) \\
0 &= \det([a_1, a_2]) + \det([a_2, a_1])
\end{aligned}
$$

- Αν οι στήλες μητρώου είναι γραμμικά εξαρτημένες, η ορίζουσα είναι 0.
- δείγμα απόδειξης

$$
\det([a_1, a_2, \gamma a_1 + \delta a_2]) = \gamma \det([a_1, a_2, a_1]) + \delta \det([a_1, a_2, a_2]) = 0
$$

- Αν προσθέσουμε δύο στήλες η ορίζουσα παραμένει αμετάβλητη.
- δείγμα απόδειξης

$$
\det([a_1, a_1 + a_2]) = \det([a_1, a_1]) + \det([a_1, a_2]) = \det([a_1, a_2])
$$

Μπορούμε να την υπολογίσουμε βάσει των παραπάνω ιδιοτήτων:

$$\begin{aligned}\det([a_1, a_2]) &= \det([\alpha_{11}e_1 + \alpha_{21}e_2, \alpha_{12}e_1 + \alpha_{22}e_2]) \\ &= \alpha_{11}\det([e_1, \alpha_{12}e_1 + \alpha_{22}e_2]) + \alpha_{21}\det([e_2, \alpha_{12}e_1 + \alpha_{22}e_2])\end{aligned}$$

$$= \alpha_{11}\alpha_{12}\det([e_1, e_1]) + \alpha_{11}\alpha_{22}\det([e_1, e_2]) + \alpha_{21}\alpha_{12}\det([e_2, e_1]) + \alpha_{21}\alpha_{22}\det([e_2, e_2])$$

$$\det([a_1, a_2]) = \alpha_{11}\alpha_{22} + \alpha_{21}\alpha_{12}\det([e_2, e_1]) = \boldsymbol{\alpha_{11}\alpha_{22} - \alpha_{21}\alpha_{12}}$$

Προσοχή: πολλοί όροι μηδενίζονται (λόγω ((εκφυλισμού))).

- Όταν $n = 2$, αντί για $4 = 2^2$ όρους, έχουμε μόνον 2.
- με $n = 3$, αντί για $27 = 3^3$ όρους έχουμε 6 ...
- ... γενικά, αντί για για n^n όρους, έχουμε $n!$.

Υπολογισμός ορίζουσας από αλγεβρικά συμπληρώματα

Ανάπτυγμα Laplace

$$\det(A) = \alpha_{11}\det(A_{11}) + (-1)^{1+2}\alpha_{12}\det(A_{12}) + \cdots + (-1)^{1+n}\alpha_{1n}\det(A_{1n})$$

- Προσωρινός συμβολισμός Με $A_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ συμβολίζεται το μητρώο που προκύπτει αν διαγράψουμε από το A τη γραμμή i και τη στήλη j. Αποκαλείται ελάσσον (minor) του στοιχείου α_{ij}.
- Το $\psi_{i,j} := (-1)^{i+j}\det(A_{ij})$ λέγεται συμπαράγοντας (cofactor) του στοιχείου α_{ij}.
- Αν συμβολίσουμε με $\Psi = [\psi_{ij}]$ το μητρώο των συμπαραγόντων, τότε το ανάστροφό του $\Psi^\top$, ονομάζεται προσαρτημένο μητρώο (adjoint ή adjugate) και συμβολίζεται $\operatorname{adj}(A) = \Psi^\top$.

Γενικά ισχύει ότι μπορούμε να εκφράσουμε την ορίζουσα ως προς οποιαδήποτε γραμμή (ή στήλη):

$$\det(A) = \alpha_{i,1}\psi_{i,1} + \alpha_{i,2}\psi_{i,2} + \cdots + \alpha_{i,n}\psi_{i,n}$$

Προσοχή: Στον παραπάνω τύπο εκφράσαμε την ορίζουσα βάσει οριζουσών μεγέθους $(n-1)\times(n-1)$.
Σύσταση: Στην εφαρμογή του τύπου, αξίζει να επιλέγουμε τη γραμμή ή στήλη με τα περισσότερα μηδενικά!

$$\begin{vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{vmatrix} := \alpha_{11} \begin{vmatrix} \alpha_{22} & \alpha_{23} \\ \alpha_{32} & \alpha_{33} \end{vmatrix} - \alpha_{12} \begin{vmatrix} \alpha_{21} & \alpha_{23} \\ \alpha_{31} & \alpha_{33} \end{vmatrix} + \alpha_{13} \begin{vmatrix} \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{vmatrix}$$

Άρα

$$\det(A) = \alpha_{11}\alpha_{22}\alpha_{33} - \alpha_{11}\alpha_{23}\alpha_{32} - \alpha_{12}\alpha_{21}\alpha_{33} + \alpha_{12}\alpha_{23}\alpha_{31} + \alpha_{13}\alpha_{21}\alpha_{32} - \alpha_{13}\alpha_{22}\alpha_{31}$$

Προσέξτε:

$$\det(A) = \alpha_{11}\alpha_{22}\alpha_{33} - \alpha_{11}\alpha_{23}\alpha_{32} - \alpha_{12}\alpha_{21}\alpha_{33} + \alpha_{12}\alpha_{23}\alpha_{31} + \alpha_{13}\alpha_{21}\alpha_{32} - \alpha_{13}\alpha_{22}\alpha_{31}$$

Μεγάλος τύπος ορίζουσας (Τύπος Leibnitz)

Υπενθύμιση: Για μία μετάθεση π των στοιχείων $(1, \ldots, n)$, δηλ. $(\pi(1), \ldots, \pi(n))$, ορίζουμε ως πρόσημο της μετάθεσης, $\sigma(\pi)$, μία τιμή ± 1. Η τιμή θα είναι +1 αν η μετάθεση π προέρχεται από άρτιο αριθμό εναλλαγών ή -1 αν προέρχεται από περιττό αριθμό εναλλαγών. Το αντίστοιχο μητρώο μετάθεσης P_π θα είναι εκείνο για το οποίο $P_\pi \begin{pmatrix} 1 \\ 2 \\ \vdots \\ n \end{pmatrix} = \begin{pmatrix} \pi(1) \\ \pi(2) \\ \cdots \\ \pi(n) \end{pmatrix}$.

Παρατηρήστε ότι το P_π είναι μετάθεση του ταυτοτικού, επομένως $\det(P_\pi) = \pm 1$. Η τιμή του συμπίπτει με το πρόσημο της μετάθεσης.

Μεγάλος τύπος (γιατί έχει $n!$ όρους)

$\det(A) = \sum_{\pi \in \mathcal{S}} \sigma(\pi) \alpha_{1\pi(1)} \cdots \alpha_{n\pi(n)}$ όπου $\mathcal{S}$ είναι το σύνολο των $n!$ μεταθέσεων των στοιχείων $(1, \ldots, n)$ και $\sigma(\pi)$ είναι το πρόσημο της μετάθεσης π.

Ορίζουσες μητρώων με ειδική δομή

Τριγωνικά: Η ορίζουσα κάθε τριγωνικού μητρώου είναι ίση με το γινόμενο όλων των στοιχείων της διαγωνίου του.

Κατά πλοκάδες τριγωνικά Αν A, D τετραγωνικά μητρώα, τότε

$$\det\begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \det(A)\det(D)$$

Ειδική περίπτωση αν $\hat{A} \in \mathbb{R}^{(n-1)\times(n-1)}$,

$$\det\left(\begin{array}{c|ccc} 1 & \star & \cdots & \star \\ \hline 0_{n-1,1} & \hat{A} & & \end{array}\right) = \det(\hat{A})$$

Προσοχή: Για γενικά μητρώα:

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} \neq \det(A)\det(D) \text{ καθώς επίσης}$$

$$\neq \det(A)\det(D) - \det(B)\det(C)$$

Άλλες χρήσιμες ιδιότητες

Έστω ότι $A, B \in \mathbb{R}^{n \times n}$.

- Το είναι αντιστρέψιμο αν και μόνον αν $\det(A) \neq 0$.
- $\det(A^{-1}) = 1/\det(A)$ αν $\det(A) \neq 0$.
- $\det(A) = \det(A^\top)$
- $\det(AB) = \det(A)\det(B)$
- $\det(\gamma A) = \gamma^n \det(A)$ για κάθε $\gamma \in \mathbb{R}^n$

<u>ΠΡΟΣΟΧΗ</u> Η τέταρτη ιδιότητα αποκαλύπτει έναν τρόπο υπολογισμού του $\det(A)$ που έχει περίπου όσες πράξεις χρειάζεται η παραγοντοποίηση LU.

<u>Πώς?</u> Αν $A = PLU$ τότε

$$\begin{aligned} \det(A) &= \det(P) \underbrace{\det(L)}_{=1} \det(U) \\ &= (-1)^\kappa \det(U) = (-1)^\kappa \upsilon_{11}\upsilon_{22}\cdots\upsilon_{nn}. \end{aligned}$$

όπου κ είναι ο αριθμός των εναλλαγών που μετατρέπουν το I σε P. Το $(-1)^\kappa$ θα είναι το πρόσημο της αντίστοιχης μετάθεσης.

- Ο υπολογισμός της ορίζουσας με τους παραπάνω τύπους είναι ακριβός
- ... εκτός αν την υπολογίσουμε μέσω LU
- ... οπότε δεν υπάρχει λόγος να χρησιμοποιήσουμε τον κανόνα
- ... γιατί από την LU βρίσκουμε τη λύση με μπρος και πίσω αντικατάσταση, πιο φθηνά!
- Εκτός αν το μητρώο έχει αξιοποιήσιμη δομή (π.χ. είναι κάτω τριγωνικό).

Ορίζουσες και τάξη μητρώου

Ένα μητρώο $A \in \mathbb{R}^{m \times n}$ έχει τάξη ακριβώς r ανν και μόνον αν το μεγαλύτερο τετραγωνικό (υπο)μητρώο με μη μηδενική ορίζουσα είναι $r \times r$.

Αν $Ax = b$ τότε κάθε στοιχείο ξ_i του x μπορεί να υπολογιστεί με τον κανόνα Cramer:

Εύρεση συνιστωσών του διανύσματος λύσης

Έστω οι στήλες $a_1, \ldots, a_n$ τ.ώ. $\det([a_1, \ldots, a_n]) \neq 0$. Έστω επίσης ότι η λύση γράφεται ως $x = \xi_1 e_1 + \cdots + \xi_n e_n$. Τότε για κάθε $j = 1, \ldots, n$ ισχύει ότι

$$\xi_j = \frac{\det([a_1, \ldots, a_{j-1}, b, a_{j+1}, \ldots, a_n])}{\det([a_1, \ldots, a_n])}$$

Εύρεση στοιχείων του αντιστρόφου

Αν το A είναι αντιστρέψιμο, τότε

$$\begin{aligned}(A^{-1})_{i,j} &= (-1)^{i+j}\frac{\det(A_{j,i})}{\det(A)} = \frac{\psi_{j,i}}{\det(A)}\\ A^{-1} &= \frac{\operatorname{adj}(A)}{\det(A)} \text{ τύπος για το αντίστροφο}\end{aligned}$$

Μητρώο τάξης 1 Έστω ότι από τα διανύσματα (στήλες) u, v κατασκευάζουμε το μητρώο $A = uv^\top \in \mathbb{R}^{n\times n}$. Ποιά θα είναι η ορίζουσα?

Απάντηση: 0 (γιατί?)

Μητρώο προβολής σε (γνήσιο) υπόχωρο?

Απάντηση: 0 (γιατί?)

Ορθογώνιο μητρώο?

Απάντηση: $\det(A) = \pm 1$ (γιατί?)

Ταυτοτικό συν μητρώο "τάξης-1" $A = I + uv^\top$. Ποιά θα είναι η ορίζουσα? Απάντηση: $\det(A) = 1 + v^\top u$. (γιατί?)

$$B = \begin{pmatrix} 1 & -v^\top \\ u & I \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ u & I \end{pmatrix}\begin{pmatrix} 1 & -v^\top \\ 0 & I + uv^\top \end{pmatrix}, \text{ και } C = \begin{pmatrix} I & u \\ -v^\top & 1 \end{pmatrix} = \begin{pmatrix} I & 0 \\ -v^\top & 1 \end{pmatrix}\begin{pmatrix} I & u \\ 0 & 1 + v^\top u \end{pmatrix}$$

και αν $P = \begin{pmatrix} 0_{1,n} & 1 \\ I_n & 0_{n,1} \end{pmatrix}$ τότε $\det(P) = (-1)^{n+2}$ και $P^\top BP = C$ άρα $\det(P^\top BP) = \det(B) = \det(C)$. Από τις παραγοντοποιήσεις, $\det(I + uv^\top) = 1 + v^\top u$.

- **Ταυτότητα Sylvester** Γενικότερα ισχύει ότι αν $C \in \mathbb{R}^{m\times n}$, $D \in \mathbb{R}^{n\times m}$

$$\det(I_m + CD) = \det(I_n + DC)$$

- Συνιστάται να την αποδείξετε!

Παράδειγμα: Να υπολογίσετε την ορίζουσα του

$$= \begin{pmatrix} 4 & 2 & 2 & 2 \\ 2 & 4 & 2 & 2 \\ 2 & 2 & 4 & 2 \\ 2 & 2 & 2 & 4 \end{pmatrix}$$

Επίλυση: Παρατηρούμε ότι το μητρώο μπορεί να γραφτεί ως $A = 2I + 2ee^\top$ όπου $e = (1,1,1,1)^\top$. Επομένως

$$\det(A) = \det(2(I + ee^\top)) = 2^4\det(I + ee^\top) = 2^4(1 + e^\top e) = 16\cdot 5 = 80.$$

☺ Θα ήταν αρκετά πιο χρονοβόρο αν δουλεύατε απευθείας με το A.

How to find a good submatrix

S. A. Goreinov, I. V. Oseledets, D. V. Savostyanov,
E. E. Tyrtyshnikov, N. L. Zamarashkin

October 17, 2008

Abstract

Pseudoskeleton approximation and some other problems require the knowledge of sufficiently well-conditioned submatrix in a large-scale matrix. The quality of a submatrix can be measured by modulus of its determinant, also known as volume. In this paper we discuss a search algorithm for the maximum-volume submatrix which already proved to be useful in several matrix and tensor approximation algorithms. We investigate the behavior of this algorithm on random matrices and present some its applications, including maximization of a bivariate functional.

$$\mathrm{perm}(A) := \sum_{\pi \in \mathcal{S}} \alpha_{1\pi(1)} \cdots \alpha_{n\pi(n)}$$

όπου $\mathcal{S}$ είναι το σύνολο των $n!$ μεταθέσεων των στοιχείων $(1, \ldots, n)$.

1. Ενώ η ορίζουσα μπορεί να υπολογιστεί με $O(n^3)$ πράξεις, μέσω της παραγοντοποίησης *LU*, ο υπολογισμός της $\mathrm{perm}(A)$ απαιτεί πολύ μεγάλο πλήθος πράξεων (μεγαλύτερο από πολυωνυμικό) καθώς αφορά την άθροιση εκθετικού πλήθους ακεραίων. Θυμηθείτε ότι $n! \approx \sqrt{2\pi n}(\frac{n}{e})^n$ (προσέγγιση Stirling).
2. Ο πιο γνωστός αλγόριθμος (Ryser) για τον ακριβή υπολογισμό απαιτεί $O(n2^n)$ πράξεις.
3. Έχει αποδειχθεί ότι ο υπολογισμός είναι #P-hard (sharp-P hard). Για μητρώα στο GF(2) (δηλ με στοιχεία $\{0, 1\}$) ο υπολογισμός είναι #P-complete.

Ερευνητικό πρόβλημα: Επινόηση γρήγορων αλγορίθμου για την αποτελεσματική προσέγγιση της permanent σε πολυωνυμικό χρόνο.

Γεωμετρική ερμηνεία: Η ορίζουσα ως ((προσημασμένος όγκος)) I

Δείτε και την πολύ ωραία παρουσίαση των Margalit & Rabinoff (Georgia Tech) εδώ.

Τεταγμένο άπλοκο στον $\mathbb{R}^n$ (ordered simplex) πολύεδρο με $n+1$ κορυφές στον $\mathbb{R}^n$.

Γεωμετρική θεώρηση οριζουσών Αν θεωρήσουμε ως $n+1$ σημεία το 0 και τα n σημεία που ορίζονται από τις στήλες του $A = (0, a_1, \cdots, a_n)$ με αυτή τη σειρά, σε κάθε τετραγωνικό μητρώο αντιστοιχεί ένα τεταγμένο άπλοκο.

Όγκος απλόκου

$$\mathrm{Vol}(A) = \frac{1}{n}\mathrm{Vol}_{n-1}\big((\text{βάση})\big) \times (\text{ύψος})$$

Προσημασμένος όγκος

$$\text{Προσημασμένος όγκος}() = \text{Προσανατολισμός απλόκου } (A) \times \mathrm{Vol}(A)$$

Παράδειγμα Η ορίζουσα κάθε $A = [a_1, a_2] \in \mathbb{R}^{2\times 2}$ είναι (2 φορές) το προσημασμένο εμβαδόν του τριγώνου με κορυφές $(0, a_1, a_2)$. Η ορίζουσα κάθε $A = [a_1, a_2, a_3] \in \mathbb{R}^{3\times 3}$ είναι (6 φορές) ο προσημασμένος όγκος του τετραέδρου με κορυφές $(0, a_1, a_2, a_3)$, κ.ο.κ.

Αποτελέσματα

1. Δίνονται τα γραμμικά ανεξάρτητα διανύσματα $a_1, \ldots, a_n \in \mathbb{R}^m$ και το μητρώο $A = [a_1, \ldots, a_n]$. Τότε ο όγκος του n-διάστατου στερεού στον $\mathbb{R}^m$ που ορίζεται ως $\mathcal{S} = \{\sum_{i=1}^{n} \gamma_i a_i | \gamma_i \in [0, 1]\}$ είναι

$$V = \sqrt{\det(A^\top A)}$$

2. Αν $\ : \mathbb{R}^n \to \mathbb{R}^m$ είναι γραμμικός μετασχηματισμός και $A \in \mathbb{R}^{m \times n}$ το μητρωο που τον αναπαριστά και $\mathcal{S}$ χωρίο στον $\mathbb{R}^n$ όγκου $V_{\mathcal{S}}$, τότε ο όγκος της εικόνας του $\mathcal{S}$ υπό τον T είναι

$$V = \sqrt{\det(A^\top A)} V_{\mathcal{S}}.$$

ΠΡΟΑΙΡΕΤΙΚΟ Χρειάζεται η έννοια του εκθετικού μητρώου!

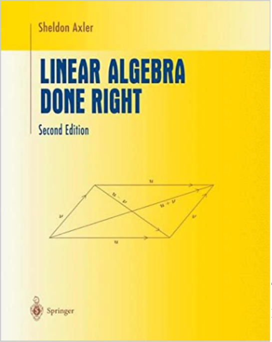
Για οποιοδήποτε $A \in \mathbb{R}^{n\times n}$, το εκθετικό του ορίζεται από την δυναμοσειρά μητρώου

$$\exp(A) \quad = \quad I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

Όπως και με τους βαθμωτούς, η σειρά συγκλίνει (τι σημαίνει αυτό για μητρώα?)

Μάλιστα ισχύει ότι ((η ορίζουσα του εκθετικού είναι ίση με το εκθετικό του ίχνους!))

$$\det(\exp(A)) \quad = \quad \exp(\mathrm{trace}(A))$$



Down with Determinants!

Sheldon Axler

This paper was published in the *American Mathematical Monthly* 102 (1995), 139-154.

In 1996 this paper received the Lester R. Ford Award for expository writing from the Mathematical Association of America.

Abstract: This paper shows how linear algebra can be done better without determinants. The standard proof that a square matrix

complexification discussed earlier.

Now we are ready for the formal definition. The *determinant* of T, denoted $\det T$, is defined to be the product of the eigenvalues of T, counting multiplicity. This definition would not be possible with the traditional approach to eigenvalues, because that method uses determinants to prove that eigenvalues exist. With the techniques used here, we already know (by Theorem 3.11(a)) that T has $\dim V$ eigenvalues, counting multiplicity. Thus our simple definition makes sense.

This paper focuses on showing that determinants should be banished from much of the theoretical part of linear algebra. Determinants are also useless in the computational part of linear algebra. For example, Cramer's rule for solving systems of linear equations is already worthless for 10×10 systems, not to mention the much larger systems often encountered in the real world. Many computer programs efficiently calculate eigenvalues numerically—none of them uses determinants. To emphasize the point, let me quote a numerical analyst. Henry Thacher, in a review (*SIAM News*, September 1988) of the *Turbo Pascal Numerical Methods Toolbox*, writes,

> I find it hard to conceive of a situation in which the numerical value of a determinant is needed: Cramer's rule, because of its inefficiency, is completely impractical, while the magnitude of the determinant is an indication of neither the condition of the matrix nor the accuracy of the solution.

Log-Determinant Function and Properties

The *log-determinant* function is a function from the set of symmetric matrices in $\mathbf{R}^{n\times n}$, with domain the set of positive definite matrices, and with values

$$f(X) = \begin{cases} \log\det X & \text{if } X \succ 0, \\ +\infty & \text{otherwise.} \end{cases}$$

The function can be expressed in terms of the (real, positive) eigenvalues of the argument matrix X; it does not depend on its eigenvectors.

This function provides a measure of the volume of an ellipsoid. Precisely, the volume of the ellipsoid

$$\mathbf{E} = \{x : x^T X^{-1} x \leq 1\}$$

is given by \bf vol$(\mathbf{E}) = C_n \prod_{i=1}^n \sqrt{\lambda_i(X)}$, where C_n is a constant (given by the volume of the unit sphere in $\mathbf{R}^n$). Thus, log \bf vol$(\mathbf{E}) = \frac{1}{2} f(X) +$ constant.

This means that the volume of the ellipsoid is a function of the product of the eigenvalues of the matrix X.

Proof of the concavity of the log-determinant function: We use the fact that a function is convex if and only if its restriction to a line is.

Entropic Trace Estimates for Log Determinants

Jack Fitzsimons[1], Diego Granziol[1], Kurt Cutajar[2]
Michael Osborne[1], Maurizio Filippone[2], and Stephen Roberts[1]

[1] Department of Engineering, University of Oxford, UK
{jack,diego,mosb,sjrob}@robots.ox.ac.uk
[2] Department of Data Science, EURECOM, France
{kurt.cutajar,maurizio.filippone}@eurecom.fr

Abstract. The scalable calculation of matrix determinants has been a bottleneck to the widespread application of many machine learning methods such as determinantal point processes, Gaussian processes, generalised Markov random fields, graph models and many others. In this work, we estimate log determinants under the framework of maximum entropy, given information in the form of moment constraints from stochastic trace estimation. The estimates demonstrate a significant improvement on state-of-the-art alternative methods, as shown on a wide variety of matrices from the SparseSuite Matrix Collection. By taking the example of a general Markov random field, we also demonstrate how this approach can significantly accelerate inference in large-scale learning methods involving the log determinant.

ELSEVIER

Linear Algebra and its Applications
Volume 533, 15 November 2017, Pages 95-117

A randomized algorithm for approximating the log determinant of a symmetric positive definite matrix

Christos Boutsidis, Petros Drineas [a], Prabhanjan Kambadur [b], Eugenia-Maria Kontopoulou [a], Anastasios Zouzias

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.laa.2017.07.004
Get rights and content

Abstract

We introduce a novel algorithm for approximating the logarithm of the determinant of a symmetric positive definite (SPD) matrix. The algorithm is randomized and approximates the traces of a small number of matrix powers of a specially constructed matrix, using the method of Avron and Toledo [1]. From a theoretical perspective, we present additive and relative error bounds for our algorithm. Our additive error bound works for any SPD matrix, whereas our relative error bound works for SPD matrices whose eigenvalues lie in the interval $(\theta_1, 1)$, with $0 < \theta_1 < 1$; the latter setting was proposed in [16]. From an empirical perspective, we demonstrate that a C++ implementation of our algorithm can approximate the logarithm of the determinant of large matrices very accurately in a matter of seconds.

Towards some RandNLA Techniques for Determinant Approximation, Sparse PCA and Analysis of Krylov Subspace Methods

Eugenia-Maria Kontopoulou & Petros Drineas

Computer Science, Purdue University

LogDet Problem

Given an SPD matrix $A \in \mathbb{R}^{n \times n}$, compute (exactly or approximately) $\log\det(A)$.

Additive Error Approximation

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be an SPD matrix. For any α with $\lambda_1(\mathbf{A}) < \alpha$, define $\mathbf{B} = \mathbf{A}/\alpha$ and $\mathbf{C} = \mathbf{I}_n - B$. Then,

$$\log\det(A) = n\log(\alpha) - \sum_{k=1}^{\infty} \frac{\mathbf{Tr}\left(\log\left(\mathbf{C}^k\right)\right)}{k}.$$

Algorithm 1

Input: $\mathbf{A} \in \mathbb{R}^{n \times n}$, accuracy parameter $\epsilon > 0$, integer $m > 0$.

1. Compute an estimate to the largest eigenvalue of $\mathbf{A}$,$\tilde{\lambda_1}(\mathbf{A})$, using the Power Method.
2. $\mathbf{C} = \mathbf{I}_n - \mathbf{A}/(7\tilde{\lambda_1}(\mathbf{A}))$
3. Create $p = \lceil 20\log(2/\delta)/\epsilon^2 \rceil$ i.i.d random Gaussian vectors, $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_p$.
4. Estimate $\sum_{k=1}^{\infty} \frac{\mathbf{Tr}(\log(\mathbf{C}^k))}{k}$ with a truncated Taylor Series type randomized trace estimator that computes $\sum_{k=1}^{m}\left(\frac{1}{p}\sum_{i=1}^{p}\mathbf{g}_i^\top \mathbf{C}^k \mathbf{g}_i\right)$

Let $\widehat{\log\det}(\mathbf{A})$ be the $\log\det$ approximation of the above procedure. Then, we **prove** that with probability at least $1 - 2\delta$,

$$|\widehat{\log\det}(\mathbf{A}) - \log\det(A)| \le 2\epsilon\Gamma$$

where $\Gamma = \sum_{i=1}^{n}\log\left(7 \cdot \frac{\lambda_1(\mathbf{A})}{\lambda_i(\mathbf{A})}\right)$ and $m \ge \lceil 7\kappa(\mathbf{A})\log(\frac{1}{\epsilon})\rceil$.

Relative Error Approximation

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be an SPD matrix whose eigenvalues lie in the interval $(\theta_1, 1)$, for some $0 < \theta_1 < 1$. Let $\mathbf{C} = \mathbf{I}_n - A$. Then,

$$\log\det(A) = -\sum_{k=1}^{\infty} \frac{\mathbf{Tr}\left(\log\left(\mathbf{C}^k\right)\right)}{k}.$$

Algorithm 2

Input: $\mathbf{A} \in \mathbb{R}^{n \times n}$, accuracy parameter $\epsilon > 0$, integer $m > 0$.

Sparse PCA

Given a centered matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ (the mean of its columns is zero), we seek for a vector $\mathbf{w}_{opt}$ that solves the optimization problem:

$$\underset{\mathbf{w}}{\text{maximize}} \;\; \mathbf{w}^\top \mathbf{X}^\top \mathbf{X}\, \mathbf{w}$$
$$\text{subject to } \|\mathbf{w}\|_0 \le k,\; \|\mathbf{w}\|_2 \le 1,\; \mathbf{w} \in \mathbb{R}^n.$$

This problem is **NP-hard** $\rightarrow$ relax to a **problem with convex constraints (but non-convex objective)**:

$$\underset{\mathbf{w}}{\text{maximize}} \;\; \mathbf{w}^\top \mathbf{X}^\top \mathbf{X}\, \mathbf{w}$$
$$\text{subject to } \|\mathbf{w}\|_1 \le \sqrt{k},\; \|\mathbf{w}\|_2 \le 1,\; \mathbf{w} \in \mathbb{R}^n.$$

Algorithm

Phase 1: Compute a stationary point $\tilde{\mathbf{w}}_{opt}$

1. Compute the gradient and make a gradient step.
2. Project onto the l_1 ball with radius $\sqrt{k}$ ($\|\mathbf{w}\|_1$).
3. Repeat until a threshold for the relative error is exceeded.

Phase 2: Invoke a **randomized rounding strategy**.

1. Create a Bernoulli distribution and randomly round the entries of $\mathbf{w}$.
2. Repeat the experiment 10 times and keep the best sparsification.

We prove the following:

Let $\mathbf{w}_{opt}$ be the optimal solution of the Sparse PCA problem (1) satisfying $\|\mathbf{w}_{opt}\|_2 = 1$ and $\|\mathbf{w}_{opt}\|_0 \le k$. Let $\hat{\mathbf{w}}_{opt}$ be the vector returned when the rounding sparsification strategy is applied on the optimal solution $\tilde{\mathbf{w}}_{opt}$ of the optimization problem (1), with $s = 200k/\epsilon^2$, where $\epsilon \in (0, 1]$ is an accuracy parameter. Then, $\hat{\mathbf{w}}_{opt}$ has the following properties:

1. $\mathbb{E}\|\hat{\mathbf{w}}_{opt}\|_0 \le s$.
2. With probability at least 3/4,
$$\|\hat{\mathbf{w}}_{opt}\|_2 \le 1 + 0.15\epsilon.$$
3. With probability at least 3/4,
$$\hat{\mathbf{w}}_{opt}^\top \mathbf{A}\hat{\mathbf{w}}_{opt} \ge \mathbf{w}_{opt}^\top \mathbf{A}\mathbf{w}_{opt} - \epsilon.$$

Krylov Methods

Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a starting guess matrix $\mathbf{X} \in \mathbb{R}^{n \times s}$, we want to use the block Krylov space $\mathcal{K}_q(\mathbf{A}\mathbf{A}^\top, \mathbf{A}\mathbf{X})$ to approximate the **left singular vector space** of $\mathbf{A}$.

We prove:

- Spectral & Frobenius bounds for the distance between the approximate and the actual space.
- Quality measurements of the bounds relative to the best low-rank approximation.

Theorem

Let $\phi(x)$ be a polynomial of degree $2q+1$ with odd powers only, such that $\phi(\mathbf{\Sigma}_k)$ is nonsingular. If $\mathrm{rank}(\mathbf{V}_k^\top \mathbf{X}) = k$ then

$$\|\sin\mathbf{\Theta}(\mathcal{K}_q, \mathbf{U}_k)\|_{2,F} \le \|\phi(\mathbf{\Sigma}_{k,\perp})\|_2 \|\phi(\mathbf{\Sigma}_k)^{-1}\|_2 \|\mathbf{V}_{k,\perp}^\top \mathbf{X}(\mathbf{V}_k^\top \mathbf{X})^\dagger\|_{2,F}.$$

If, in addition, $\mathbf{X}$ has orthonormal or linearly independent columns, then

$$\|\mathbf{V}_{k,\perp}^\top \mathbf{X}(\mathbf{V}_k^\top \mathbf{X})^\dagger\|_{2,F} = \|\tan\mathbf{\Theta}(\mathbf{X}, \mathbf{V}_k)\|_{2,F}$$

and

$$\|\sin\mathbf{\Theta}(\mathcal{K}_q, \mathbf{U}_k)\|_{2,F} \le \|\phi(\mathbf{\Sigma}_{k,\perp})\|_2 \|\phi(\mathbf{\Sigma}_k)^{-1}\|_2 \|\tan\mathbf{\Theta}(\mathbf{X}, \mathbf{V}_k)\|_{2,F}.$$

where $\mathbf{\Theta}(\mathcal{K}_q, \mathbf{U}_k) \in \mathbb{R}^{k \times k}$ is the diagonal matrix of principal angles between $\mathcal{K}_q$ and $\mathrm{range}(\mathbf{U}_k)$.

Theorem

Let $\phi(x)$ be a polynomial of degree $2q+1$ with odd powers only, such that $\phi(\mathbf{\Sigma}_k)$ is nonsingular and $\phi(\sigma_i) \ge \sigma_i$, for $1 \le i \le k$. If $\mathrm{rank}(\mathbf{V}_k^\top \mathbf{X}) = k$ then for $1 \le i \le k$,

$$\|A - \hat{\mathbf{U}}_i \hat{\mathbf{U}}_i^\top A\|_F \le \|\mathbf{A} - \mathbf{A}_i\|_F + \Delta$$
$$\|A - \hat{\mathbf{U}}_i \hat{\mathbf{U}}_i^\top A\|_2 \le \|\mathbf{A} - \mathbf{A}_i\|_2 + \Delta$$
$$\sigma_i - \Delta \le \|\hat{u}_i^\top A\|_2 \le \sigma_i.$$

If, in addition, $\mathbf{X}$ has orthonormal columns, then:

$$\Delta = \|\phi(\mathbf{\Sigma}_{k,\perp})\|_2 \|\tan\mathbf{\Theta}(\mathbf{X}, \mathbf{V}_k)\|_F$$

Citation

I. Ipsen, P. Drineas, E. Kontopoulou, M. Magdon-Ismail

Συμπληρωματικές Διαφάνειες

Την ορίζουσα του A συμβολίζουμε και με $|A|$.

Είδαμε ότι η ορίζουσα ενός 2×2 μητρώου A παίζει έναν ιδιαίτερο ρόλο και δίνει διάφορες πληροφορίες για το μητρώο. Για παράδειγμα, είδαμε ότι αν $\det A \neq 0$, τότε υπάρχει το αντίστροφο μητρώο A^{-1}, επιπλέον

$$A^{-1} = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Συμπεραίνουμε λοιπόν ότι η ορίζουσα ενός 2×2, αλλά και γενικότερα ενός $n\times n$ μητρώου όπως θα δούμε, δίνει πληροφορία για

1. την αντιστρεψιμότητα του μητρώου, ισοδύναμα
2. την γραμμική ανεξαρτησία των στηλών του μητρώου, ισοδύναμα
3. την γραμμική ανεξαρτησία των γραμμών του μητρώου, ισοδύναμα
4. την τάξη του μητρώου.

1.1. Ιδιότητες της ορίζουσας

1. Η ορίζουσα του ταυτοτικού μητρώου είναι ίση με 1.
2. Η ορίζουσα μητρώου αλλάζει πρόσημο εάν εναλλάξουμε τις γραμμές του μητρώου.
$$\begin{vmatrix} c & d \\ a & b \end{vmatrix} = cb - da = -(ad - bc) = -\begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$
3. Η ορίζουσα εξαρτάται γραμμικά από την πρώτη γραμμή του μητρώου. Πράγματι
$$\begin{aligned} \begin{vmatrix} a+p & b+q \\ c & d \end{vmatrix} &= (a+p)d - (b+q)c \\ &= (ad - bc) + (pd - qc) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} p & q \\ c & d \end{vmatrix} \end{aligned}$$
και
$$\begin{vmatrix} \lambda a & \lambda b \\ c & d \end{vmatrix} = \lambda ad - \lambda bc = \lambda(ad - bc) = \lambda \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

o4 Η ορίζουσα μητρώου είναι ίση με την ορίζουσα του ανάστροφου μητρώου. Πράγματι

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow \det A^{\mathrm{T}} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \det A.$$

Παρατήρηση (1)

Από τις ιδιότητες (ο2) και (ο3) έπεται ότι η ορίζουσα εξαρτάται γραμμικά από την δεύτερη γραμμή του μητρώου, αφού

$$\begin{vmatrix} a & b \\ c+p & d+q \end{vmatrix} = -\begin{vmatrix} c+p & d+q \\ a & b \end{vmatrix} = -\left(\begin{vmatrix} c & d \\ a & b \end{vmatrix} + \begin{vmatrix} p & q \\ a & b \end{vmatrix}\right) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a & b \\ p & q \end{vmatrix}$$

και

$$\begin{vmatrix} a & b \\ \lambda c & \lambda d \end{vmatrix} = -\begin{vmatrix} \lambda c & \lambda d \\ a & b \end{vmatrix} = -\lambda \begin{vmatrix} c & d \\ a & b \end{vmatrix} = \lambda \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

Συνεπώς η ορίζουσα εξαρτάται γραμμικά από την πρώτη ή δεύτερη γραμμή του μητρώου κρατώντας την άλλη γραμμή σταθερή.

Παρατήρηση (2)

Επίσης από την ιδιότητα (ο4) έπεται ότι οποιοδήποτε συμπέρασμα αφορά στις γραμές μητρώου ισχύει και για τις στήλες. Για παράδειγμα εναλλάσοντας δύο στήλες αλλάζει το πρόσημο

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} = -\begin{vmatrix} b & d \\ a & c \end{vmatrix} = -\begin{vmatrix} b & a \\ d & c \end{vmatrix}.$$

και όσον αφορά στη γραμμικότητα

$$\begin{aligned} \begin{vmatrix} a+\lambda p & b \\ c+\lambda q & d \end{vmatrix} &= \begin{vmatrix} a+\lambda p & c+\lambda q \\ b & d \end{vmatrix} \\ &= \begin{vmatrix} a & c \\ b & d \end{vmatrix} + \lambda \begin{vmatrix} p & q \\ b & d \end{vmatrix} \\ &= \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \lambda \begin{vmatrix} p & b \\ q & d \end{vmatrix}. \end{aligned}$$

Ένα 2×2 μητρώο A μπορούμε να το γράψουμε ως

$$A = \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{r}_1^{\mathrm{T}} \\ \mathbf{r}_2^{\mathrm{T}} \end{pmatrix},$$

όπου $\mathbf{c}_1, \mathbf{c}_2$ ειναι οι στήλες του A ως διανύσματα, και $\mathbf{r}_1, \mathbf{r}_2$ οι γραμμές του A ως διανύσματα (στήλες). Σχετικά με την Παρατήρηση 2 αν $\mathbf{a}$, $\mathbf{b}$ και $\mathbf{c}$ είναι διανύσματα στο $\mathbb{R}^2$ και γράψουμε $r_a = \mathbf{a}^{\mathrm{T}}$, $r_b = \mathbf{b}^{\mathrm{T}}$, και $r_c = \mathbf{c}^{\mathrm{T}}$, τότε

$$\det\begin{pmatrix} \mathbf{a} + \lambda \mathbf{b} & \mathbf{c} \end{pmatrix} = \det\begin{pmatrix} \mathbf{a} & \mathbf{c} \end{pmatrix} + \lambda \det\begin{pmatrix} \mathbf{b} & \mathbf{c} \end{pmatrix} \tag{1}$$

$$\det\begin{pmatrix} \mathbf{a} & \mathbf{b} + \lambda \mathbf{c} \end{pmatrix} = \det\begin{pmatrix} \mathbf{a} & \mathbf{b} \end{pmatrix} + \lambda \det\begin{pmatrix} \mathbf{a} & \mathbf{c} \end{pmatrix}, \tag{2}$$

$$\det\begin{pmatrix} r_a + \lambda r_b \\ r_c \end{pmatrix} = \det\begin{pmatrix} r_a \\ r_c \end{pmatrix} + \lambda \det\begin{pmatrix} r_b \\ r_c \end{pmatrix} \tag{3}$$

$$\det\begin{pmatrix} r_a \\ r_b + \lambda r_c \end{pmatrix} = \det\begin{pmatrix} r_a \\ r_b \end{pmatrix} + \lambda \det\begin{pmatrix} r_a \\ r_c \end{pmatrix}, \tag{4}$$

γεγονός που εκφράζει ότι η ορίζουσα είναι γραμμική ως προς κάθε στήλη ξεχωριστά, ή ως προς κάθε γραμμή ξεχωριστά.

Παρατήρηση (3)

Αναρωτιόμαστε αν είναι σωστό, και προκύπτει από τις ιδιότητες της ορίζουσας που αναφέραμε, ότι $\det(A+B) = \det A + \det B$. Ας το ελέγξουμε. Για

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

και λ μια σταθερά, από τη γραμμικότητα της ορίζουσας ως προς κάθε γραμμή ξεχωριστά, υπολογίζουμε

$$\det(\lambda A) = \begin{vmatrix} \lambda a & \lambda b \\ \lambda c & \lambda d \end{vmatrix} = \lambda \begin{vmatrix} a & b \\ \lambda c & \lambda d \end{vmatrix} = \lambda^2 \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \lambda^2 \det A. \tag{5}$$

Έτσι για $\lambda = 2$ έχουμε

$$\det(A+A) = \det(2A) = 4\det A \neq 2\det A = \det A + \det A,$$

κατά συνέπεια

$$\det(A+B) \neq \det A + \det B.$$

Θεώρημα (1 Ιδιότητες της ορίζουσας)

Έστω A ένα 2×2 μητρώο.

- O1 *Η ορίζουσα του ταυτοτικού μητρώου είναι ίση με 1.*
- O2 *Εάν το μητρώο A' προκύπτει από το A με εναλλαγή των γραμμών του, τότε η ορίζουσα αλλάζει πρόσημο, δηλαδή $\det A' = -\det A$.*
- O3 *Η ορίζουσα μητρώου ως συνάρτηση είναι γραμμική ξεχωριστά ως προς κάθε γραμμή κρατώντας την άλλη γραμμή σταθερή.*
- O4 *Η ορίζουσα μητρώου είναι ίση με την ορίζουσα του ανάστροφου μητρώου.*
- O5 *Εάν οι δύο γραμμές του A είναι ίσες, τότε $\det A = 0$.*
- O6 *Εάν οι δύο γραμμές του A είναι γραμμικά εξαρτημένες, τότε $\det A = 0$.*
- O7 *Εάν σε μια γραμμή του A προστεθεί το πολλαπλάσιο άλλης γραμμής, η τιμή της ορίζουσας δεν αλλάζει.*

Η αντίστοιχη εκδοχή για στήλη ή στήλες καθεμιάς από τις ιδιότητες $(\mathrm{O}2)-(\mathrm{O}3)$ και $(\mathrm{O}5)-(\mathrm{O}7)$ ισχύει επίσης.

Απόδειξη

$(O1) - (O4)$ είναι οι $(o1) - (o4)$.

$(O5)$ Εάν οι γραμμές i και j του A, $i \neq j$, είναι ίσες, τότε εναλλάσοντάς τες το μητρώο παραμένει το ίδιο αλλά από την $(O2)$ η ορίζουσα αλλάζει πρόσημο, έτσι θα έχουμε

$$\det A = -\det A \Rightarrow \det A = 0.$$

$(O6)$ Εάν η γραμμή j του A είναι πολλαπλάσιο της γραμμής i, $i \neq j$, έστω $r_j = \lambda r_i$, και A' είναι το μητρώο A στο οποίο η j γραμμή έχει αντικατασταθεί από την i, τότε το A' έχει δύο γραμμές ίσες, άρα $\det A' = 0$. Επίσης από την $(O3)$, παίρνουμε

$$\det A = \lambda \det A' = 0.$$

$(O7)$ Έστω A'' να είναι το μητρώο A στο οποίο η j γραμμή έχει αντικατασταθεί από την $r_j + \lambda r_i$ και έστω A' να είναι όπως στο $(O6)$, τότε και πάλι από τη γραμμικότητα της ορίζουζας βρίσκουμε

$$\det A'' = \det A + \lambda \det A' = \det A,$$

αφού $\det A' = 0$, που είναι ό,τι θέλουμε να αποδείξουμε.

Συνέπεια της $(O4)$ είναι ότι οτιδήποτε ισχύει για γραμμές ισχύει και για στήλες, ειδικά η για στήλη εκδοχή της $(O3)$ είναι οι (1) και (2). ■

Παράδειγμα (1) Αν

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = k$$

να βρεθούν, δίχως να εκτελεστούν πράξεις, οι τιμές των οριζουσών

(α´) $\begin{vmatrix} \lambda c & \lambda d \\ \lambda a & \lambda b \end{vmatrix}$ (β´) $\begin{vmatrix} a & b \\ \lambda a + \mu c & \lambda b + \mu d \end{vmatrix}$ (γ´) $\begin{vmatrix} a - \lambda c & b - \lambda d \\ \lambda a + c & \lambda b + d \end{vmatrix}$

όπου λ και μ είναι μια πραγματικές σταθερές.

(α´) Από τις ιδιότητες της ορίζουσας υπολογίζουμε

$$\begin{aligned} \begin{vmatrix} \lambda c & \lambda d \\ \lambda a & \lambda b \end{vmatrix} &= \lambda \begin{vmatrix} c & d \\ \lambda a & \lambda b \end{vmatrix} && \text{(από την (O3))} \\ &= \lambda^2 \begin{vmatrix} c & d \\ a & b \end{vmatrix} && \text{(από την (O3))} \\ &= \lambda^2(-1) \begin{vmatrix} a & b \\ c & d \end{vmatrix} && \text{(από την (O2))} \\ &= -\lambda^2 k. \end{aligned}$$

(β')

$$\begin{vmatrix} a & b \\ \lambda a+\mu c & \lambda b+\mu d \end{vmatrix} = \begin{vmatrix} a & b \\ \lambda a & \lambda b \end{vmatrix} + \begin{vmatrix} a & b \\ \mu c & \mu d \end{vmatrix} \qquad (\text{από την } (O2))$$

$$= \lambda \begin{vmatrix} a & b \\ a & b \end{vmatrix} + \mu \begin{vmatrix} a & b \\ c & d \end{vmatrix} \qquad (\text{από την } (O2))$$

$$= \lambda 0 + \mu k \qquad (\text{από την } (O6))$$

$$= \mu k.$$

(γ')

$$\begin{vmatrix} a-\lambda c & b-\lambda d \\ \lambda a+c & \lambda b+d \end{vmatrix} = \begin{vmatrix} a & b \\ \lambda a+c & \lambda b+d \end{vmatrix} - \lambda \begin{vmatrix} c & d \\ \lambda a+c & \lambda b+d \end{vmatrix}$$

$$= \lambda \begin{vmatrix} a & b \\ a & b \end{vmatrix} + \begin{vmatrix} a & b \\ c & d \end{vmatrix} - \lambda^2 \begin{vmatrix} c & d \\ a & b \end{vmatrix} - \lambda \begin{vmatrix} c & d \\ c & d \end{vmatrix}$$

$$= \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \lambda^2 \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

$$= (1+\lambda^2)k.$$

Ορισμός (2)

Αν $A = \left(a_{ij}\right)$ είναι ένα $n \times n$ μητρώο, με A_{ij} συμβολίζουμε το $\left(n-1\right) \times \left(n-1\right)$ μητρώο το οποίο προκύπτει από το A διαγράφοντας την i γραμμή και την j στήλη.

Παράδειγμα (2)

Αν

$$A = \begin{pmatrix} 2 & -1 & 3 \\ 1 & 4 & -2 \\ 0 & -3 & 2 \end{pmatrix}$$

τότε

$$A_{11} = \begin{pmatrix} 4 & -2 \\ -3 & 2 \end{pmatrix}, \quad A_{13} = \begin{pmatrix} 1 & 4 \\ 0 & -3 \end{pmatrix}, \quad A_{22} = \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix},$$

$$A_{23} = \begin{pmatrix} 2 & -1 \\ 0 & -3 \end{pmatrix}, \quad A_{31} = \begin{pmatrix} -1 & 3 \\ 4 & -2 \end{pmatrix}, \quad A_{32} = \begin{pmatrix} 2 & 3 \\ 1 & -2 \end{pmatrix}.$$

Την ορίζουσα ενός $n \times n$ μητρώου A συμβολίζουμε και με $|A|$, όπως ακριβώς και στην περίπτωση 2×2, έτσι αν

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \qquad \text{τότε} \qquad \det A = |A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

Ορισμός (3)

Αν $n \in \mathbb{N}$ ορίζουμε τη συνάρτηση $\det : \mathbb{M}^{n,n}(\mathbb{K}) \to \mathbb{K}$ ως

1. $n = 1$ και $A = \big(a_{11}\big)$, τότε
$$\det A = a_{11}.$$
2. $n > 1$ και $A = \big(a_{ij}\big)$, τότε
$$\det A = (-1)^{1+1} a_{11} \det A_{11} + (-1)^{1+2} a_{12} \det A_{12} + \cdots + (-1)^{1+n} a_{1n} \det A_{1n}. \quad (6)$$

Από τον ορισμό λοιπόν έπεται ότι

1. Αν $n = 2$, οπότε $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, από το (1) υπολογίζουμε

$$\det A = (-1)^{1+1} a_{11} \det(a_{22}) + (-1)^{1+2} a_{12} \det(a_{21}) = a_{11}a_{22} - a_{12}a_{21}$$

δηλαδή στην περίπτωση $n = 2$ η συνάρτηση det συμφωνεί με την ορίζουσα τάξης δύο.

2. Αν $n = 3$, οπότε $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, υπολογίζουμε

$$\begin{aligned} \det A &= (-1)^{1+1} a_{11} \det A_{11} + (-1)^{1+2} a_{12} \det A_{12} + (-1)^{1+3} a_{13} \det A_{13} \\ &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}, \end{aligned}$$

και το τελικό αποτέλεσμα προκύπτει από την περίπτωση $n = 2$. Κατά συνέπεια η έκφραση στην (6) γενικεύει την ορίζουσα τάξης δύο και μοιάζει να είναι ένας λογικός ορισμός για την ορίζουσα κάθε τάξης. Θέλουμε η ορίζουσα να ικανοποιεί τις ιδιότητες που ικανοποιεί η ορίζουσα τάξης δύο.

Θεώρημα (2)

Η συνάρτηση $\det : \mathbb{M}^{n,n}(\mathbb{K}) \to \mathbb{K}$ *στον Ορισμό 3 ικανοποιεί τις ιδιότητες*

1. *Αν* I *είναι το ταυτοτικό μητρώο, τότε* $\det I = 1$.
2. *Η συνάρτηση* $\det$ *είναι γραμμική ως προς κάθε γραμμή ξεχωριστά.*
3. *Αν το μητρώο* A' *προκύπτει από το* A *με εναλλαγή δύο γραμμών, τότε* $\det A' = -\det A$.

Επιπλέον η $\det$ *είναι η* **μοναδική** *συνάρτηση* $\mathbb{M}^{n,n}(\mathbb{K}) \mapsto \mathbb{K}$ *η οποία χαρακτηρίζεται από τις παραπάνω ιδιότητες.*

Άσκηση 1.

Δείξτε ότι η ιδιότητα (3) στο Θεώρημα 2, δεδομένης της (2), είναι, για την ακρίβεια ισοδύναμη με κάθε μια από τις $(3a)$, $(3b)$, $(3c)$

$(3a)$ Αν το μητρώο A έχει δύο ίσες γραμμές, τότε $\det A = 0$.

$(3b)$ Αν $\operatorname{rank} A < n$, τότε $\det A = 0$.

$(3c)$ Αν το μητρώο A' προκύπτει από το A προσθέτοντας το πολλαπλάσιο μιας γραμμής σε άλλη γραμμή, τότε $\det A = \det A'$.

Στην ιδιότητα (2) εννοούμε ότι αν με r_j συμβολίσουμε την j-γραμμή του μητρώου A, ως διάνυσμα γραμμή, ώστε $A = \begin{pmatrix} r_1 & r_2 & \ldots & r_n \end{pmatrix}^{\mathrm{T}}$, τότε

$$\det\begin{pmatrix} r_1 \\ \vdots \\ \lambda r + \mu r' \\ \vdots \\ r_n \end{pmatrix} = \lambda \det\begin{pmatrix} r_1 \\ \vdots \\ r \\ \vdots \\ r_n \end{pmatrix} + \mu \det\begin{pmatrix} r_1 \\ \vdots \\ r' \\ \vdots \\ r_n \end{pmatrix}$$

Μια άμεση συνέπεια της ιδιότητας αυτής είναι ότι αν μια γραμμή ενός μητρώου περιέχει μόνο το μηδέν, τότε η ορίζουσα του μητρώου είναι ίση με μηδέν. Πράγματι αν συμβολίσουμε με 0 τη μηδενική γραμμή, τότε

$$\det\begin{pmatrix} r_1 \\ \vdots \\ 0 \\ \vdots \\ r_n \end{pmatrix} = \det\begin{pmatrix} r_1 \\ \vdots \\ r - r \\ \vdots \\ r_n \end{pmatrix} = \det\begin{pmatrix} r_1 \\ \vdots \\ r \\ \vdots \\ r_n \end{pmatrix} - \det\begin{pmatrix} r_1 \\ \vdots \\ r \\ \vdots \\ r_n \end{pmatrix} = 0.$$

Θεώρημα (3 το ανάπτυγμα της ορίζουσας ως προς γραμμή)

Εάν $A = (a_{ij})$ είναι ένα $n \times n$ μητρώο με $n \geq 2$, τότε για $1 \leq i \leq n$ η ορίζουσα του A δίνεται από τη σχέση

$$\det A = (-1)^{i+1} a_{i1} \det A_{i1} + (-1)^{i+2} a_{i2} \det A_{i2} + \cdots + (-1)^{i+n} a_{in} \det A_{in}. \tag{7}$$

Θεώρημα (4 το ανάπτυγμα της ορίζουσας ως προς στήλη)

Εάν $A = (a_{ij})$ είναι ένα $n \times n$ μητρώο με $n \geq 2$, τότε για $1 \leq j \leq n$ η ορίζουσα του A δίνεται από τη σχέση

$$\det A = (-1)^{1+j} a_{1j} \det A_{1j} + (-1)^{2+j} a_{2j} \det A_{2j} + \cdots + (-1)^{n+j} a_{nj} \det A_{nj}. \tag{8}$$

Ορισμός (4)

Για το μητρώο $A = (a_{ij})$, τον όρο $(-1)^{i+j} \det A_{ij}$ στο ανάπτυγμα της ορίζουσας λέμε **συμπαράγοντα** (cofactor) του στοιχείου a_{ij}. Την έκφραση (7) λέμε ανάπτυγμα σε συμπαράγοντες ως προς την i-γραμμή του A και εκείνη στην (8) λέμε ανάπτυγμα ως προς την j-στήλη του A.

Παράδειγμα (3)

Να υπολογισθεί η ορίζουσα του μητρώου

$$A = \begin{pmatrix} 2 & -1 & 3 & 1 \\ 1 & 0 & 2 & -2 \\ 4 & 0 & 1 & 3 \\ 2 & 0 & -1 & 3 \end{pmatrix}.$$

Επιλέγοντας λογικά την δεύτερη στήλη υπολογίζουμε

$$\begin{aligned} \det A &= (-1)^{1+2}(-1)|A_{12}| + (-1)^{2+2}0|A_{22}| + (-1)^{3+2}0|A_{32}| + (-1)^{4+2}0|A_{42}| \\ &= \det A_{12} = \det \begin{pmatrix} 1 & 2 & -2 \\ 4 & 1 & 3 \\ 2 & -1 & 3 \end{pmatrix} \\ &= (-1)^{1+1}\det\begin{pmatrix} 1 & 3 \\ -1 & 3 \end{pmatrix} + (-1)^{1+2}(2)\det\begin{pmatrix} 4 & 3 \\ 2 & 3 \end{pmatrix} + (-1)^{1+3}(-2)\det\begin{pmatrix} 4 & 1 \\ 2 & -1 \end{pmatrix} \\ &= 6 - 2(6) - 2(-6) = 6. \end{aligned}$$

Παράδειγμα (4)

Να υπολογισθεί η ορίζουσα του μητρώου

$$A = \begin{pmatrix} 2 & 2 & 3 & 1 \\ 0 & 3 & 2 & 1 \\ 0 & 0 & 4 & 3 \\ 0 & 0 & 0 & 5 \end{pmatrix}.$$

Αναπτύσοντας ως προς την πρώτη στήλη υπολογίζουμε

$$\begin{aligned} \det A &= (-1)^{1+1}(2)\det A_{11} \\ &= 2\det\begin{pmatrix} 3 & 2 & 1 \\ 0 & 4 & 3 \\ 0 & 0 & 5 \end{pmatrix} = (2)(3)\det\begin{pmatrix} 4 & 3 \\ 0 & 5 \end{pmatrix} = (2)(3)(4)(5). \end{aligned}$$

Κατά συνέπεια η ορίζουσα του άνω τριγωνικού μητρώου είναι ίση με το γινόμενο των στοιχείων της (κύριας) διαγωνίου.

Πρόταση (1)

Η ορίζουσα ενός άνω, ή κάτω τριγωνικού, ή διαγωνίου μητρώου είναι ίση με το γινόμενο των στοιχείων της διαγωνίου.

3. Ιδιότητες της ορίζουσας

Θεώρημα (5)

Εάν A είναι ένα τετραγωνικό μητρώο, τότε $\det A = \det A^{\mathrm{T}}$.

Θεώρημα (6)

Εάν A και B είναι $n \times n$ μητρώα, τότε $\det(AB) = (\det A)(\det B)$.

Πόρισμα (1)

Εάν A είναι ένα $n \times n$ αντιστρέψιμο μητρώο τότε $\det A \neq 0$, *επιπλέον*

$$\det(A^{-1}) = \frac{1}{\det A}.$$

Παράδειγμα (5)

Υπολογίζουμε την 2×2 ορίζουσα με χρήση των ιδιοτήτων που την χαρακτηρίζουν μοναδικά (Θεώρημα 2) και τις ιδιότητες που απορρέουν από αυτές.

Αν

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{και} \qquad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

τότε

$$\begin{aligned}
\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \det\left(a\mathbf{e}_1 + c\mathbf{e}_2 \;\; b\mathbf{e}_1 + d\mathbf{e}_2\right) \\
&= a\det\left(\mathbf{e}_1 \;\; b\mathbf{e}_1 + d\mathbf{e}_2\right) + c\det\left(\mathbf{e}_2 \;\; b\mathbf{e}_1 + d\mathbf{e}_2\right) \\
&= ab\det\left(\mathbf{e}_1 \;\; \mathbf{e}_1\right) + ad\det\left(\mathbf{e}_1 \;\; \mathbf{e}_2\right) + cb\det\left(\mathbf{e}_2 \;\; \mathbf{e}_1\right) + cd\det\left(\mathbf{e}_2 \;\; \mathbf{e}_2\right) \\
&= ad\det\left(\mathbf{e}_1 \;\; \mathbf{e}_2\right) - bc\det\left(\mathbf{e}_1 \;\; \mathbf{e}_2\right) \\
&= ad\det I - bc\det I \\
&= ad - bc.
\end{aligned}$$

Έστω $A\mathbf{x} = \mathbf{b}$ ένα τετραγωνικό σύστημα, αν το μητρώο A είναι αντιστρέψιμο, τότε η μοναδική λύση του συστήματος δίνεται από $\mathbf{x} = A^{-1}\mathbf{b}$.

Θεώρημα (7 Ο κανόνας του Cramer)

Έστω A ένα $n \times n$ αντιστρέψιμο μητρώο και έστω $\mathbf{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}^{\mathrm{T}}$ η λύση του συστήματος $A\mathbf{x} = \mathbf{b}$. Για $k = 1, 2, \ldots, n$ συμβολίζουμε με A_k το μητρώο που προκύπτει από το A αν αντικαταστήσουμε την k-στήλη του με το διάνυσμα $\mathbf{b}$, τότε οι λύση του συστήματος δίνεται από τη σχέση

$$x_k = \frac{\det A_k}{\det A}, \qquad k = 1, 2, \ldots, n. \tag{9}$$

Απόδειξη

Αν $A = \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{pmatrix}$, συμβολίζουμε με X_k το μητρώο που προκύπτει από το ταυτοτικό μητρώο I αν αντικαταστήσουμε την k-στήλη του με $\mathbf{x}$, δηλαδή

$$X_k = \begin{pmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_{k-1} & \mathbf{x} & \mathbf{e}_{k+1} & \cdots & \mathbf{e}_n \end{pmatrix} \tag{10}$$

όπου $\mathbf{e}_k$, $k = 1, \ldots, n$ είναι οι στήλες του I, τότε

$$AX_k = \begin{pmatrix} A\mathbf{e}_1 & \cdots & A\mathbf{e}_{k-1} & A\mathbf{x} & A\mathbf{e}_{k+1} & \cdots & A\mathbf{e}_n \end{pmatrix}$$

ή

$$AX_k = \begin{pmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_{k-1} & \mathbf{b} & \mathbf{c}_{k+1} & \cdots & \mathbf{c}_n \end{pmatrix} = A_k.$$

Έτσι υπολογίζουμε

$$\det A_k = \det(AX_k) = (\det A)(\det X_k) \tag{11}$$

Από την (10) το τυπικό μητρώο X_k γράφεται ως

$$X_k = \begin{pmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_{k-1} & \underbrace{x_1\mathbf{e}_1 + \cdots + x_k\mathbf{e}_k + \cdots + x_n\mathbf{e}_n}_{k\text{-στήλη}} & \mathbf{e}_{k+1} & \cdots & \mathbf{e}_n \end{pmatrix} \tag{12}$$

οπότε από τη γραμμικότητα της ορίζουσας και τη γραμμική εξάρτηση των στηλών του X_k υπολογίζουμε

$$\det X_k = x_k \det I = x_k. \tag{13}$$

Από τις (11), (13) έπεται η ζητούμενη σχέση. ■

4. Ένας τύπος για το αντίστροφο μητρώο

Στην περίπτωση ενός 2×2 αντιστρέψιμου μητρώου

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \Rightarrow \quad A^{-1} = \frac{1}{\det A} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

Αν με $\left(A^{-1}\right)_{ij}$ συμβολίζουμε το στοιχείο στην i-γραμμή και j-στήλη του A^{-1}, και με $\psi_{ij} = (-1)^{i+j} \det A_{ij}$ τον συμπαράγοντα του a_{ij}, τότε

$$\begin{aligned}
\left(A^{-1}\right)_{11} &= \frac{a_{22}}{\det A} = (-1)^{1+1} \frac{\det A_{11}}{\det A} = \frac{\psi_{11}}{\det A}, \\
\left(A^{-1}\right)_{12} &= \frac{-a_{12}}{\det A} = (-1)^{1+2} \frac{\det A_{21}}{\det A} = \frac{\psi_{21}}{\det A}, \\
\left(A^{-1}\right)_{21} &= \frac{-a_{21}}{\det A} = (-1)^{2+1} \frac{\det A_{12}}{\det A} = \frac{\psi_{12}}{\det A}, \\
\left(A^{-1}\right)_{22} &= \frac{a_{11}}{\det A} = (-1)^{2+2} \frac{\det A_{22}}{\det A} = \frac{\psi_{22}}{\det A}.
\end{aligned}$$

Έτσι αν Ψ είναι το μητρώο των συμπαραγόντων, τότε

$$A^{-1} = \frac{1}{\det A} \Psi^{\mathrm{T}}.$$

Αποδεικνύεται ότι ο τύπος αυτός ισχύει και στη γενική $n \times n$ περίπτωση.

Θεώρημα (8)

Εάν A είναι ένα $n \times n$ αντιστρέψιμο μητρώο και Ψ είναι το μητρώο των συμπαραγόντων, τότε

$$A^{-1} = \frac{1}{\det A} \Psi^{\mathrm{T}}. \tag{14}$$

5. Ο τύπος του Leibniz

Ένας διαφορετικός τρόπος υπολογισμού της ορίζουζας επιτυγχάνεται με εργαλεία της Συνδυαστικής Ανάλυσης. Αναπτύσσοντας μια ορίζουσα, για παράδειγμα, τάξης 3 βρίσκουμε

$$\begin{aligned}
\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
&= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \\
&= \sum_{k=1}^{3!} (-1)^{n_k} a_{1\sigma_k(1)} a_{2\sigma_k(2)} a_{3\sigma_k(3)},
\end{aligned}$$

αφού κάθε όρος του αθροίσματος είναι της μορφής $a_{1i}a_{2j}a_{3k}$ με $i, j, k \in \{1,2,3\}$ όπου $i \neq j$, $i \neq k$, και $j \neq k$ σε όλους τους δυνατούς συνδυασμούς, όπου $n_k \in \{1,2\}$ και σ_k είναι μετάθεση του $\{1,2,3\}$, $k = 1,2,\ldots,6$. Μια μετάθεση του $\{1,\ldots,n\}$ είναι μια ένα-προς-ένα αντιστοιχία $\sigma : \{1,\ldots,n\} \to \{1,\ldots,n\}$. Οι μεταθέσεις αυτές αποτελούν τη συμμετρική ομάδα S_n με $n!$ στοιχεία. Λέμε ότι οι τιμές της ταυτοτικής μετάθεσης έχουν τη φυσική διάταξη. Λέμε ότι μια μια μετάθεση είναι **άρτια/περιττή** αν οι τιμές της επανέρχονται στη φυσική διάταξη με άρτιο/περιττό πλήθος εναλλαγών. Για παράδειγμα η 312 είναι άρτια μετάθεση ενώ η 132 είναι περιττή μετάθεση. Ορίζουμε το **πρόσημο** μιας μετάθεσης σ, $\pi(s)$, με τη σχέση

$$\pi(\sigma) = +1, \quad \text{αν η } \sigma \text{ είναι άρτια} \quad \& \quad \pi(\sigma) = -1, \quad \text{αν η } \sigma \text{ είναι περιττή}$$

Θεώρημα (9)

Εάν A είναι ένα $n \times n$ μητρώο και S_n είναι η ομάδα των μεταθέσεων του $\{1,\ldots,n\}$, τότε η ορίζουσα του A είναι

$$\det A = \sum_{\sigma \in S_n} \pi(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)},$$

όπου $\pi(\sigma)$ είναι το πρόσημο της μετάθεσης σ.

6. Το εξωτερικό γινόμενο

Αν $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ είναι η κανονική βάση στον $\mathbb{R}^3$ μπορούμε να γράψουμε, όπως συνηθίζεται στον Απειροστικό Λογισμό, ή στη Φυσική,

$$\mathbf{i} = \mathbf{e}_1, \qquad \mathbf{j} = \mathbf{e}_2, \qquad \mathbf{k} = \mathbf{e}_3.$$

Ορισμός (5)

Αν

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix},$$

είναι διανύσματα στον $\mathbb{R}^3$, ορίζουμε το **εξωτερικό γινόμενο** των $\mathbf{u}$ και $\mathbf{v}$, και το συμβολίζουμε με $\mathbf{u} \times \mathbf{v}$, το διάνυσμα

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \mathbf{k}. \tag{15}$$

Από το ανάπτυγμα της ορίζουσας βλέπουμε αμέσως ότι μπορούμε να παραστήσουμε το εξωτερικό γινόμενο ως ορίζουσα στην ευκολομνημόνευτη μορφή

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}.$$

Το εξωτερικό γινόμενο το λέμε και **σταυρωτό γινόμενο** (cross product), ή **διανυσματικό γινόμενο** (vector product).

Παρατήρηση (4)

Από την (15) και τις ιδιότητες της ορίζουσας έπεται αμέσως ότι

1. Αν τα $\mathbf{u}$ και $\mathbf{v}$ είναι γραμμικά εξαρτημένα, τότε $\mathbf{u} \times \mathbf{v} = \mathbf{0}$.
2. $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$.
3. $\mathbf{u} \times \mathbf{0} = \mathbf{0}$.

Επίσης άμμεση συνέπεια του ορισμού των διανυσμάτων $\mathbf{i}$, $\mathbf{j}$ και $\mathbf{k}$ και του εξωτερικού γινομένου είναι το αποτέλεσμα

Πρόταση (2)

Εάν, όπως ορίσαμε,

$$\mathbf{i} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{j} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \mathbf{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

τότε

$$\mathbf{i} \times \mathbf{j} = \mathbf{k}, \qquad \mathbf{j} \times \mathbf{k} = \mathbf{i}, \qquad \mathbf{k} \times \mathbf{i} = \mathbf{j}.$$

Πρόταση (3)

Εάν $\mathbf{u}$ *και* $\mathbf{v}$ *είναι διανύσματα στον* $\mathbb{R}^3$, *τότε*

$$\mathbf{u} \times \mathbf{v} \perp \mathbf{u} \qquad \textit{και} \qquad \mathbf{u} \times \mathbf{v} \perp \mathbf{v},$$

δηλαδή το $\mathbf{u} \times \mathbf{v}$ *είναι ορθογώνιο στο* $\mathbf{u}$ *και στο* $\mathbf{v}$, *κατά συνέπεια είναι ορθογώνιο στο επίπεδο που παράγεται από τα* $\mathbf{u}$ *και* $\mathbf{v}$ *οποτεδήποτε τα* $\mathbf{u}$ *και* $\mathbf{v}$ *είναι γραμμικά ανεξάρτητα.*

Απόδειξη.

Δείχνουμε ότι $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{u}=0$. Πράγματι από την (15) και τον ορισμό του εσωτερικού γινομένου υπολογίζουμε

$$\begin{aligned}(\mathbf{u}\times\mathbf{v})\cdot\mathbf{u} &= \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} u_1 - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} u_2 + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} u_3 \\ &= \begin{vmatrix} u_1 & u_2 & u_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix},\end{aligned}$$

από τον ορισμό της ορίζουσας, συνεπώς $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{u}=0$ αφού η ορίζουσα έχει δυο γραμμές ίσες. Η απόδειξη ότι $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{v}=0$ είναι ανάλογη, ενώ από τη γραμμικότητα του εσωτερικού γινομένου έπεται ότι

$$(\mathbf{u}\times\mathbf{v})\cdot(\lambda\mathbf{u}+\mu\mathbf{v})=\lambda(\mathbf{u}\times\mathbf{v})\cdot\mathbf{u}+\mu(\mathbf{u}\times\mathbf{v})\cdot\mathbf{v}=0$$

για κάθε $\lambda\in\mathbb{R}$ και για κάθε $\mu\in\mathbb{R}$. □

Εάν $\mathbf{w} \in \mathbb{R}^3$ παρατηρούμε ότι

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = \begin{vmatrix} w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = - \begin{vmatrix} u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

Ορισμός (6)

Εάν **u**, **v** και **w** είναι διανύσματα στον $\mathbb{R}^3$ το $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$ λέμε **τριπλό γινόμενο** (triple product) των **u**, **v** και **w**, έτσι

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = \begin{vmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{vmatrix}.$$

Από τον ορισμό του τριπλού γινομένου απορρέει αμέσως ότι τα διανύσματα **u**, **v** και **w** είναι γραμμικά εξαρτημένα, ισοδύναμα είναι συνεπίπεδα, περιέχονται δηλαδή στο ίδιο επίπεδο, αν και μόνο αν

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = 0.$$

Άσκηση 2.

Το εξωτερικό γινόμενο προσδίδει στον $\mathbb{R}^3$ επιπλέον δομή.

- (α') Εάν **a**, **b**, και **c** είναι διανύσματα στον $\mathbb{R}^3$ δείξτε ότι

$$\mathbf{a} \times (\mathbf{b}+\mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}, \qquad \text{και} \qquad (\mathbf{a}+\mathbf{b}) \times \mathbf{c} = \mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$$

  ισχύει δηλαδή ο επιμεριστικός νόμος.

- (β') Δείξτε ότι ο χώρος $\mathbb{R}^3$ εφοδιασμένος με το εξωτερικό γινόμενο, $(\mathbb{R}^3, +, \cdot, \times)$, είναι μια μη προσεταιριστική άλγεβρα, δηλαδή δεν ισχύει πάντα η ισότητα

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \times \mathbf{c}.$$

- (γ') Δείξτε ότι ικανοποιείται η **ταυτότητα του Jacobi**

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) + \mathbf{b} \times (\mathbf{c} \times \mathbf{a}) + \mathbf{c} \times (\mathbf{a} \times \mathbf{b}) = \mathbf{0}$$

  για όλα τα διανύσματα **a**, **b**, και **c**.

7. Όγκος και ορίζουσα

Αν τρία διανύσματα δεν είναι συνεπίπεδα ορίζουν ένα παραλληλεπίπεδο στο $\mathbb{R}^3$. Το βαθμωτό μέγεθος $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}$ σχετίζεται με τον όγκο του παραλληλεπιπέδου. Ας δούμε κάποιες απλές περιπτώσεις.

- Αν $\mathbf{u}=\mathbf{e}_1$, $\mathbf{v}=\mathbf{e}_2$, $\mathbf{w}=\mathbf{e}_3$, τότε

$$(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}=|I|=1$$

  όπου I είναι το ταυτοτικό μητρώο, και το $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}$ δίνει ακριβώς τον όγκο του μοναδιαίου κύβου.

- Αν λ,μ,ν είναι θετικές σταθερές και $\mathbf{u}=\lambda\mathbf{e}_1$, $\mathbf{v}=\mu\mathbf{e}_2$, $\mathbf{w}=\nu\mathbf{e}_3$, τότε

$$(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}=\begin{vmatrix}\lambda & 0 & 0\\ 0 & \mu & 0\\ 0 & 0 & \nu\end{vmatrix}=\lambda\mu\nu$$

  που και πάλι είναι ο όγκος του ορθογωνίου παραλληλεπιπέδου που ορίζουν τα τρία διανύσματα.

Πρόταση (3)

Εάν $\mathbf{u}$ *και* $\mathbf{v}$ *είναι διανύσματα στον* $\mathbb{R}^3$, *τότε*

$$\|\mathbf{u}\times\mathbf{v}\|^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - \langle\mathbf{u},\mathbf{v}\rangle^2, \tag{16}$$

όπου $\langle\cdot,\cdot\rangle$ *είναι το σύνηθες εσωτερικό γινόμενο.*

Απόδειξη.

Από την (15) υπολογίζουμε

$$\begin{aligned}\|\mathbf{u}\times\mathbf{v}\|^2 &= \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix}^2 + \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix}^2 + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}^2 \\ &= (u_2v_3 - u_3v_3)^2 + (u_1v_3 - u_3v_1)^2 + (u_1u_2 - u_2v_1)^2 \\ &= (u_1^2+u_2^2+u_3^2)(v_1^2+v_2^2+v_3^2) - (u_1v_1+u_2v_2+u_3v_3)^2\end{aligned}$$

που είναι το ζητούμενο. $\square$

Πόρισμα (2)

Εάν $\mathbf{u}$ *και* $\mathbf{v}$ *είναι διανύσματα στον* $\mathbb{R}^3$*, τότε*

$$\|\mathbf{u}\times\mathbf{v}\| = \|\mathbf{u}\|\,\|\mathbf{v}\|\sin\theta,$$

όπου θ *είναι η γωνία μεταξύ των* $\mathbf{u}$ *και* $\mathbf{v}$*,* $0\le\theta\le\pi$*.*

Απόδειξη.

Από την (16) και τον ορισμό του εσωτερικού γινομένου υπολογίζουμε

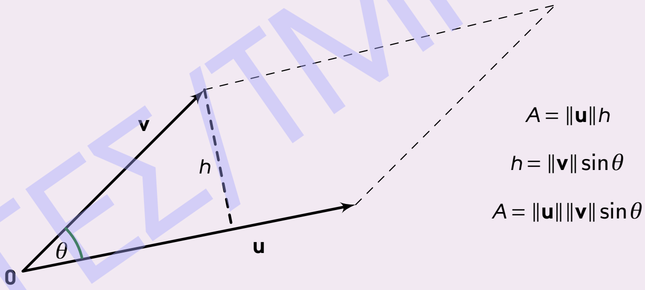
$$\begin{aligned}\|\mathbf{u}\times\mathbf{v}\|^2 &= \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - \|\mathbf{u}\|^2\|\mathbf{v}\|^2\cos^2\theta\\ &= \|\mathbf{u}\|^2\|\mathbf{v}\|^2\sin^2\theta,\end{aligned}$$

απ' όπου έπεται το ζητούμενο παίρνοντας τετραγωνικές ρίζες αφού $\sin\theta\ge 0$. □

Παρατήρηση (4 Εμβαδόν και εξωτερικό γινόμενο)

Εάν **u** και **v** είναι γραμμικά ανεξάρτητα διανύσματα στον $\mathbb{R}^3$ και με $A(\mathbf{u},\mathbf{v})$ συμβολίσουμε το εμβαδόν του παραλληλογράμμου με πλευρές τα **u** και **v**, τότε, όπως βλέπουμε και στο Σχήμα

$$A(\mathbf{u},\mathbf{v}) = \|\mathbf{u}\times\mathbf{v}\|. \tag{17}$$



Σχήμα: Εμβαδόν ορθογωνίου

Ορισμός (7)

Θα λέμε ότι η τριάδα **u**-**v**-**w** διανυσμάτων του $\mathbb{R}^3$ αποτελεί ένα **θετικά προσανατολισμένο σύστημα** αν

$$\det(\mathbf{u}\ \mathbf{v}\ \mathbf{w}) > 0.$$

Στην περίπτωση αυτή το σύστημα **u**-**v**-**w** έχει τον προσανατολισμό του τρισορθογωνίου συστήματος $\mathbf{e}_1$-$\mathbf{e}_2$-$\mathbf{e}_3$.

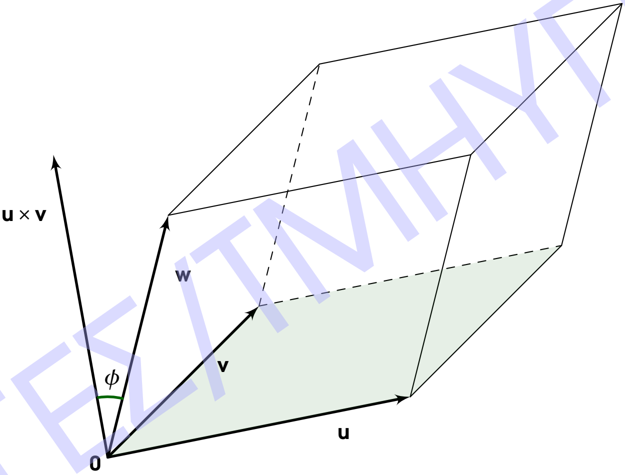
Παρατήρηση (5)

Αν τα διανύσματα **u** και **v** είναι γραμμικά ανεξάρτητα, τότε το σύστημα **u**-**v**-$\mathbf{u} \times \mathbf{v}$ είναι θετικά προσανατολισμένο.
Πράγματι γράφοντας $\mathbf{u} \times \mathbf{v} = \begin{pmatrix} p_1 & p_2 & p_3 \end{pmatrix}^{\mathrm{T}}$ έχουμε

$$\begin{aligned}
\det\left(\mathbf{u}\ \mathbf{v}\ \mathbf{u} \times \mathbf{v}\right) &= \begin{vmatrix} u_1 & v_1 & p_1 \\ u_2 & v_2 & p_2 \\ u_3 & v_3 & p_3 \end{vmatrix} \\
&= p_1 \begin{vmatrix} u_2 & v_2 \\ u_3 & v_3 \end{vmatrix} - p_2 \begin{vmatrix} u_1 & v_1 \\ u_3 & v_3 \end{vmatrix} + p_3 \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix} \\
&= \begin{vmatrix} u_2 & v_2 \\ u_3 & v_3 \end{vmatrix}^2 + \begin{vmatrix} u_1 & v_1 \\ u_3 & v_3 \end{vmatrix}^2 + \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix}^2
\end{aligned}$$

από την (15).

Έστω ότι το σύστημα **u**-**v**-**w** είναι θετικά προσανατολισμένο. Αν ϕ είναι η γωνία μεταξύ των **w** και $\mathbf{u} \times \mathbf{v}$, βλέπε Σχήμα, τότε $0 \le \phi < \pi/2$ (γιατί;).



Σχήμα: Το παραλληλεπίπεδο που παράγεται από τα διανύσματα **u**, **v**, και **w**

Αν $V(\mathbf{u},\mathbf{v},\mathbf{w})$ είναι ο όγκος του θετικά προσανατολισμένου στερεού, τότε

$$\begin{aligned} V(\mathbf{u},\mathbf{v},\mathbf{w}) &= A(\mathbf{u},\mathbf{v})\|\mathbf{w}\|\cos\phi \\ &= \|\mathbf{u}\times\mathbf{v}\|\,\|\mathbf{w}\|\cos\phi \\ &= (\mathbf{u}\times\mathbf{v})\cdot\mathbf{w} \\ &= \det(\mathbf{u}\ \mathbf{v}\ \mathbf{w}). \end{aligned}$$

Επειδή $\det(\mathbf{v}\ \mathbf{u}\ \mathbf{w}) = -\det(\mathbf{u}\ \mathbf{v}\ \mathbf{w})$, αν το σύστημα **u**-**v**-**w** δεν είναι θετικά προσανατολισμένο, τότε το **v**-**u**-**w** είναι θετικά προσανατολισμένο και $V(\mathbf{u},\mathbf{v},\mathbf{w}) = V(\mathbf{v},\mathbf{u},\mathbf{w})$. Δείξαμε λοιπόν το

Θεώρημα (9)

Αν **u**, **v**, **w** *είναι διανύσματα στον* $\mathbb{R}^3$, *τότε ο όγκος* $V(\mathbf{u},\mathbf{v},\mathbf{w})$ *του στερεού που παράγεται από τα τρία αυτά διανύσματα είναι*

$$V(\mathbf{u},\mathbf{v},\mathbf{w}) = |\det(\mathbf{u}\ \mathbf{v}\ \mathbf{w})|$$

ειδικά αν το σύστημα **u**-**v**-**w** *είναι θετικά προσανατολισμένο, τότε* $V(\mathbf{u},\mathbf{v},\mathbf{w}) = \det(\mathbf{u}\ \mathbf{v}\ \mathbf{w})$.