Algebraic Fundamentals in Artificial Intelligence for the Purpose of Undergraduate Education and Training

Silan Zhang\textsuperscript{1,2} and Jingbo Xia\textsuperscript{2,3,*}

1. College of Science, Huazhong Agricultural University, Hubei Province, P.R. China.
2. Institute of Applied Mathematics, Huazhong Agricultural University, Hubei Province, P.R. China.
3. Hubei Key Lab of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Hubei Province, P.R. China.
* Correspondence should be addressed to: xiajingbo.math@gmail.com, xjb@mail.hzau.edu.cn

Abstract. College algebra syllabus setting is of utmost importance for the higher education developments and students education and training, especially for the newly opened bachelor majors. In this research, we investigated the traditional syllabus of college algebra in a nationwide manner, and discussed the algebraic fundamentals in artificial intelligence. Three case studies were proposed to address the fundamentals of abstract logic thinking, data observation and knowledge inference, and data-driven thinking. Eventually, a suggested syllabus for “Advanced Algebra and its Applications in AI” was introduced subsequently for the new major of “artificial intelligence”.

1 Introduction

Prior research indicated that students who took more and higher levels of mathematics had been more positively effected on career-related outcomes in life [1]. Among basic mathematical courses, algebra is an irreplaceable fundamental one for college students, since algebra education is known for cultivation of logic thinking in the form of notation system and logic inference. The syllabus of the college algebra was organized to unveil the computational rules for transforming and solving equations, and along the way to introduce basic notations including vectors and matrix which played important role historically [2]. In addition, advance algebra cultivated typical ideology of logic thinking via discussions on linear space and polynomial algebra [3].

Though it seemed hard to find a nationwide guideline book for college algebra syllabus setting, the contents were traceable from the guideline of national entrance exam for graduates, released by the national education examination authority, (Official document http://yankao.neea.edu.cn/html1/report/16103/1384-1.htm). Currently, it was a nationwide practice to give linear algebra course for freshmen with 30 to 40 teaching hours. Basically, the scope of linear algebra covered necessary definitions, computation and proof requirements for solving system of linear equations, including determinants, matrices, vectors and their combination, Gaussian...
elimination. In the course with more teaching hours, binary form and linear space were introduced. However, they were not among popular syllabus for general majors. It is clear that syllabus and textbook selection were diversified among various majors, and the discussion of syllabus setting for specific major under a typical background was inevitably important.

In recent years, big data and artificial intelligence arose intense attention nationwide [4]. It was considered that big data growth in China will be faster than the other economies and it will achieve quick expansion of big data market in the near future [5]. Emerging new major of artificial intelligence calls on education research upon syllabus setting of classic mathematics courses [6]. Started from 2016, the amount of university which newly opened “big data” major was 3 (Feb, 2016), 32 (Mar, 2017), 250 (Feb, 2018), 196 (Mar, 2019). Meanwhile, 35 China university newly opened “artificial intelligence” in Mar 2019, and it was predicted that the amount will be increasing in the following years as well. The tendency of artificial intelligence education in the era of big data made it of most importance to college course reform including college algebra education.

College algebra has long been regarded as basic mathematical prerequisite for artificial learning. Russell and Norvig suggested that the mathematical foundation of artificial intelligence included complexity analysis, e.g., $O()$ notation, linear algebra in the forms of vectors and matrices, and probability theory [7]. Meanwhile, an early work of Bender [8] as well mentioned algebra as part of mathematical fundamentals and added multivariate calculus. The popularity of college algebra fundamentals were also clearly presented in textbooks like Zhou’s “Machine Learning” [9] or Han’s “Data mining: concepts and techniques” [10], both of which were regarded as classic textbook among artificial intelligence (AI) community in China.

However, it was lacking of generally accepted syllabus for the newly opened AI-related undergraduate majors. The research of this paper is to discuss the fundamentals and syllabus setting for the new majors like “artificial intelligence” in China with the background of “big data science”. Algebraic fundamentals in artificial intelligence were discussed and suggestions were given accordingly, for the purpose of syllabus setting.

2 The Fundamentals of College Algebra Education

2.1 Algebra cultivates abstract logic — a case study of the system of linear equations in the linear space

College algebra consisted of abundant materials for abstract logic training. A typical case study of system of linear equations was collected here, which was cited from a textbook widely used in HZAU campus in the previous 20 years [11].
Case study 1. To solve the system of linear equations:

\[
\begin{align*}
  x_1 + x_2 - x_3 - x_4 &= 0 \\
  2x_1 + 2x_2 + x_4 &= 0 \\
  x_1 + 2x_2 + x_3 + 2x_4 &= 0
\end{align*}
\]  \quad (1)

In this case study, four views with ascended abstract understanding level were introduced.

i) The view of “System of linear equations solving”. Gaussian elimination directly obtained the general solutions of the system of linear equations (1):

\[
\begin{pmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4 
\end{pmatrix} = k_1 \begin{pmatrix}
  -1 \\
  1 \\
  0 \\
  0 
\end{pmatrix} + k_2 \begin{pmatrix}
  -\frac{1}{2} \\
  0 \\
  -\frac{3}{2} \\
  1 
\end{pmatrix}, \quad \text{with } k_1, k_2 \in \mathbb{R}.
\]

ii) The view of “Fundamental system of solutions”. Based on the teaching syllabus after Gaussian elimination, the vector pair \( \eta_1 := (-1, 1, 0, 0)^T \) and \( \eta_2 := (-\frac{1}{2}, 0, -\frac{3}{2}, 1) \) was subsequently defined as fundamental system of the solutions of system of linear equations (1). And this notation led to a general structural understanding of linear combination of solutions for representing the whole system of linear equations.

iii) The view of “Maximal linearly independent set”. Regarded \( I = \{X_1, X_2, X_3, \cdots \} \) as an infinite-long vector set which satisfies the linear equation system (1), it was discovered that \( \{\eta_1, \eta_2\} \) was the maximal linearly independent set of the \( I \).

iv) The view of “Basis of linear space”. Denote \( V = \{X | AX = 0 \} \), with

\[
A = \begin{pmatrix}
  1 & 1 & -1 & -1 \\
  2 & 2 & 0 & 1 \\
  1 & 1 & 1 & 2 
\end{pmatrix}, \quad X = \begin{pmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4 
\end{pmatrix}, \quad \text{and } 0 = \begin{pmatrix}
  0 \\
  0 \\
  0 \\
  0 
\end{pmatrix}, \quad (2)
\]

and it was unveiled that \( V = L(\eta_1, \eta_2) \), where \( V \) was regarded as a vector space with dimension equals to 2, while \( \eta_1, \eta_2 \) formed basis of the linear space \( V \).

2.2 Algebra represents a view for data observation and knowledge inference—a case study from matrix decomposition

Algebra provided a view for structured data with the form of matrices, and it led to knowledge inference directly. Matrix decomposition, sometimes called as matrix factorization, was an advanced part of college algebra. Usually, it was excluded in a 32-hour-long college algebra course, but appeared in a “Matrix theory” course for junior or senior college students with engineering or science majors. A case study was introduced from singular value decomposition (SVD) and was extent to the observation of logic flow to knowledge inference.
Case study 2. Knowledge inference in singular value decomposition: For an \( m \times n \) matrix \( A \) of rank \( r \), there exists a SVD factorization as follows:

\[
A = U \Sigma V^T,
\]

where \( U \) and \( V \) are orthogonal matrices, \( \Sigma = \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix} \), \( \Delta = \text{diag}(\sigma_1, \cdots, \sigma_r) \), \( \sigma_i = \sqrt{\lambda_i} \) (which is called as singular value of \( A \)), \( i = 1, 2, \cdots, r \), \( r = \text{Rank}(A) \), \( \lambda_i \) is the eigenvalue of \( A A^T \).

i) The view of “Low rank approximation in SVD”. SVD itself reflected basic consideration of low-rank approximation, while Eckart-Young-Mirsky theorem [12, 13] showed that for an approximation of \( A \) as \( \tilde{A} = U \tilde{\Delta} V^T \), where \( \tilde{\Sigma} = \begin{pmatrix} \tilde{\Delta} & 0 \\ 0 & 0 \end{pmatrix} \), \( \tilde{\Delta} = \text{diag}(\sigma_1, \cdots, \sigma_l) \), the Frobenius norm of the approximation satisfied

\[
\min_{\text{rank}(A) \leq l} ||A - \tilde{A}||^2_F = \sigma^2_1 + \cdots + \sigma^2_r.
\]

ii) The view of “Principle component analysis (PCA) under matrix decomposition”. PCA aims to use orthogonal transformation to derive a set of linear uncorrelated variable (principal components) out of the set of observations of possibly correlated variables (https://en.wikipedia.org/wiki/Principal_component_analysis). SVD and principle component analysis have close relevance. Assume the matrix \( A \in \mathbb{R}^{m \times n} \) contains \( m \) “\( n \)-tuple” sample vectors \( A_i \) \( (i = 1, 2, \cdots, m) \), i.e., \( A = (A_1^T, \cdots, A_m^T)^T \). The purpose of PCA is to find several ”\( n \)-tuple” principle components \( V_1^T, \cdots, V_r^T \), (here \( V_j \) \( (j = 1, 2, \cdots, r) \) are by chance the eigenvectors of co-variant matrix \( A^T A \)) s.t., \( A_i = b_{i1} V_1^T + b_{i2} V_2^T + \cdots + b_{ir} V_r^T \), in another word, \( A = BV^T \). Coincidentally, SVD offered solution to \( A = U \Sigma V^T \). As a good visualization of PCA based on SVD, an online teaching material was recommended, http://www.infoq.com/cn/articles/matrix-decomposition-of-recommend-system.

iii) The view of “Hidden factor or latent factor” from PCA and its expansion to latent semantic indexing. It was thought that the rows of \( B \) in equation \( (??) \) represented coordinates for samples \( A_i \)s in the latent space. In the case when \( A \) was defined as a document-term matrix in the context of topic modeling, the row of \( B \) were defined as hidden factor or latent factor [14], and were expanded to the idea of latent semantic index. The latter achieved a good amount of successes [15] in health [16], social science [17], online rating [18], and so on.

iv) The view of “Hidden factors serving for data fusion with jointly tensor and matrix decomposition”. Both matrix and tensor represented flexible data structure while the latter was treated the higher order form of matrix. Tensor decomposition achieved a low rank approximation and served well in novel link discovery [19], which had various applications in NLP [20], human behavior [21], image processing [22], bioinformatics [23], and so on.
Nimishakavi [24] made good use of hidden factors of matrix decomposition and integrated it with a RESCAL-based tensor decomposition method [25]. The original data of triple (entity, predicate, entity) was filled in the tensor $X$, while side information with hidden factor encoding of each entity was filled in the matrix $A$. Therefore, a joint decomposition of tensor and matrix, i.e., $X \leftarrow BRB^T$, $A \leftarrow BV$ led to a hybrid strategy for data fusion. The loss function of this coupled factorization was proposed as:

$$
\min_{B,V,R} \sum_{k=1}^{m} f(X_k, B, R_k) + f_{np}(A, B, V) + f_{rel}(S, R)
$$

where $f(X_k, B, R_k)$ mainly contained $||X_k - BR_kB^T||_F^2$ that reflected the approximation of tensors, $f_{np}(A, B, V)$ contained $||A - BV||_F^2$ that sufficed to trace the approximation of matrix, $f_{rel}(S, R)$ is a regularizer for relation clustering. Since the row vector of $B$ contained consistent latent factor in entity semantic space, the above joint decomposition achieved both knowledge inference in tensor decomposition and entity association in matrix decomposition.

2.3 Algebra lays the mathematical basis for artificial intelligence. Algebra, combined with matrix calculation, forms the theorem basis of neural networks.

It was noted that the burst of deep learning in recent 10 years dramatically changed the research pattern among AI communities. The success of deep learning rooted from both the availability of big data and fast-implementation computational unit. Deep layers neural network well integrated the complicated data structure and made good use of big data training ideology. Unfortunately, the lack of mathematics research made the deep learning applications fragile in theorem basis [26, 27, 28]. Algebra was considered part of mathematics basis for neural network efficiency analysis like gradient descending and the loss function optimization. In addition, there were general prerequisite requirements for most artificial intelligence courses and that made it possible to list several advanced algebra contents into the syllabus of college algebra for AI.

i) The computation demands for “Matrix differentials”. The understanding of the computation part in deep learning and most artificial intelligence algorithms usually required preliminaries in computational mathematics such as norm, convexity, and matrix differentials. Moreover, the above preliminaries were widely required in all sorts of following courses in machine learning or data mining. For instance, the computation procedure of PCA also represented a series of matrix differentials computation, https://en.wikipedia.org/wiki/Principal_component_analysis. Another example came from a most popular natural language processing (NLP) course, Stanford university CS224d — deep learning for NLP, http://cs229.stanford.edu/section/cs229-linalg.pdf, where matrix calculus were introduced as prerequisite for deep learning based text mining. The contents included the Gradient, the Hessian, gradients and Hessians of quadratic and linear functions, least Squares, gradients of the determinant, and eigenvalues as optimization.
ii) The computation demands for “Gradient calculation in neural networks”. First, various framework of neural networks designed various loss function with multivariate vector representations. The understanding of neural network parameter searching relied heavily on gradient calculation. Second, to investigate the effectiveness of network structure required various gradient computation estimation for gradient explosion or gradient vanishing [29]. Currently, convergence of traditional BP neural networks were widely investigated [30], the convergence analysis of deep neural network was still in its infancy [31, 32]. Therefore, the algebraic fundamentals for gradient analysis was considered part of syllabus in the course of college algebra for AI.

3 Conclusion

3.1 Algebraic fundamentals for artificial intelligence

To conclude, there are three kinds of algebraic fundamentals for artificial intelligence. First, algebra cultivates abstract logic; second, it represents a view for data observation and knowledge inference; third, it lays the mathematical basis for artificial intelligence.

3.2 Syllabus Setting of College Algebra for Artificial Intelligence in the Era of Big Data

To wrap up, the case studies in this research formed the outline of syllabus for the course of “Advanced Algebra and its Applications in AI”. The slides of the courses are under preparation. And the sample chapters are available in GitHub, https://github.com/JingboXia/JingboSlidesLinearRegression.git.

| Syllabus of “Advanced Algebra and its Applications in AI” (32 hours) |
|---------------------------------------------------------------------|
| **Chapter1: Abstract Core of Algebra**                               |
| 1.1 From Fundamental System of Solutions to Linear Space            |
| 1.2 Matrix Computation and Abstract Addition in a Set               |
| 1.3 Concepts of Algebraic Structure                                 |
| **Chapter2: Algebraic view of data in \( \mathbb{R}^n \)**           |
| 2.1 From Geometric 2D or 3D Space to Euclidean Space, \( \mathbb{R}^n \) |
| 2.2 Dot and Line Segments                                           |
| 2.3 Norm, Norm Ball and Convexity                                   |
| 2.4 Hyper Plane                                                     |
| 2.5 Application: Optimal Hyper Plane — Algebraic View of Data in Classification |
| 2.6 Application: Linear Regression and LASSO — Algebraic View of Data in Regression |
Chapter 3: Matrix Calculus

3.1 Multivariate Differentials
3.2 Matrix Calculus Skill in Trace and Norm
3.3 Lagrangian Multiplier
3.4 Application: Activation Function and Loss Function — Algebraic View of Structured Data in Neural Network

Chapter 4: Matrix Decomposition

4.1 Singular Value Decomposition
4.2 Principal Component Analysis
4.3 Application: Latent Semantic Indexing
4.4 Application: RESCAL in Knowledge Graph

Chapter 5: Gradient Computation

5.1 Gradient
5.2 Gradient Descending
5.3 Application: Error propagation in BP neural networks
5.4 Application: Gradient Explosion and Gradient Vanishing in RNN

It is noted that this syllabus is for AI-majored students instead of math-majored ones. Therefore, neither modern algebra nor matrix theory is in the whole curriculum, hence the contents in Chapter 1 and 4 are a set of good addition. The purpose of Chapter 2 is to lays basis for understanding of classification and regression in data mining, while Chapter 3 and 5 are ready for mathematical analysis on neural network.

4 Author Contributions

Both of the authors taught advanced algebra and linear algebra for years to undergraduates. Silan Zhang also taught number theory and abstract algebra, while Jingbo Xia taught text mining and data mining. The two authors both contributed their teaching and research experiences when forming the main idea of this paper.

5 Acknowledgements

The authors would like to express their gratitude for students and audiences in their algebra class for the last dozen years. This research was funded by the Fundamental Research Funds for the Central Universities of China, grant number 2662018PY096.
6 References

[1] Gaertner M N, Kim J, DesJardins S L and McClarty K L 2014 Research in Higher Education 55 143–165
[2] Bednarz N, Kieran C and Lee L 1996 Approaches to algebra: Perspectives for research and teaching Approaches to algebra (Springer) pp 3–12
[3] Li X, Zhao P and Li H 2018 Thoughts on mathematical ideology of linear space 2018 4th International Conference on Humanities and Social Science Research (Atlantis Press)
[4] Song M, Cen L, Zheng Z, Fisher R, Liang X, Wang Y and Huisingh D 2017 Journal of Cleaner Production 142 489–500
[5] Liu Y, He J, Guo M, Yang Q and Zhang X 2014 China Communications 11 1–10
[6] Gui J, Zhang Z, Guo K and et al 2018 Computer education 27–31
[7] Russell S J and Norvig P 2016 Artificial intelligence: a modern approach (Malaysia; Pearson Education Limited.)
[8] Bender E A 1996 Mathematical methods in artificial intelligence (IEEE Computer Society Press Los Alamitos)
[9] Zhou Z 2016 Machine Learning (in Chinese) (Tsinghua University Press)
[10] Han J, Pei J and Kamber M 2011 Data mining: concepts and techniques (Elsevier)
[11] Deng Z 2000 Linear algebra and its application (in Chinese) (Higher Education Press)
[12] Nie F, Zhu W and Li X 2017 Unsupervised large graph embedding Thirty-first AAAI conference on artificial intelligence
[13] Eckart C and Young G 1936 Psychometrika 1 211–218
[14] Hofmann T 1999 Probabilistic latent semantic analysis Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence (Morgan Kaufmann Publishers Inc.) pp 289–296
[15] Landauer T K, McNamara D S, Dennis S and Kintsch W 2013 Handbook of latent semantic analysis (Psychology Press)
[16] Khan M R and Kannan R 2018 Inter. Jour. of Scientific Research in Science and Technology 4
[17] Valdez D, Pickett A C and Goodson P 2018 Social Science Quarterly 99 1665–1679
[18] Guo Y, Barnes S J and Jia Q 2017 Tourism Management 59 467–483
[19] Kolda T G and Bader B W 2009 Siam Review 51 455–500
[20] Bahargam S and Papalexakis E E 2018 Constrained coupled matrix-tensor factorization and its application in pattern and topic detection 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)
[21] Khokher M R, Bouzerdoum A and Phung S L 2017 Human interaction recognition using low-rank matrix approximation and super descriptor tensor decomposition IEEE International Conference on Acoustics
[22] Du B, Zhang M, Zhang L, Hu R and Tao D 2016 IEEE Transactions on Multimedia 19 67–79
[23] Zhou K, Zhang Sheng G M, Kim J D, Luo Q, Cohen K B and Xia J 2019 Mathematical Biosciences and Engineering 16 1376–1391
[24] Nimishakavi M, Saini U S and Talukdar P 2016 arXiv preprint arXiv:1605.04227
[25] Nickel M, Tresp V and Kriegel H P 2011 A three-way model for collective learning on multi-relational data. ICML vol 11 pp 809–816
[26] Shield M and Dole S 2013 Educational Studies in Mathematics 82 183–199
[27] Tygert M, Bruna J, Chintala S, LeCun Y, Piantino S and Szlam A 2016 Neural computation 28 815–825
[28] Vidal R, Bruna J, Giryes R and Soatto S 2017 arXiv preprint arXiv:1712.04741
[29] Hochreiter S 1998 International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 6 107–116
[30] Wu W, Wang J, Cheng M and Li Z 2011 Neural Networks 24 91–98
[31] Arora S, Cohen N, Golovich N and Hu W 2018 arXiv preprint arXiv:1810.02281
[32] Du S S, Zhai X, Poczos B and Singh A 2018 arXiv preprint arXiv:1810.02054