CMST 26(1) 15–20 (2020) DOI:10.12921/cmst.2019.0000048

On the Singular Value Decomposition and Ranking Techniques

P. Zizler[1], **P. Thangarajah**[2], **M. Sobhanzadeh**[3]

Mount Royal University
4825 Mt Royal Gate SW
Calgary, AB T3E 6K6
[1]*E-mail: pzizler@mtroyal.ca*
[2]*E-mail: pthangarajah@mtroyal.ca*
[3]*E-mail: msobhanzadeh@mtroyal.ca*

Received: 13 November 2019; revised: 18 March 2020; accepted: 18 March 2020; published online: 21 March 2020

Abstract: Let A be a positive non-singular $n \times n$ matrix. An approximation for a positive eigenvector for A^*A corresponding to the dominant singular value of A was suggested as the normalized version of a weighted sum of the rows of A with weights being the euclidean norms of the rows of A. In our paper we give a justification for this approach via the iteration of the power method and we show numerically that choosing the l^1 norm yields better results. Applications of our results are given to ranking techniques.

Key words: singular value decomposition, positive matrices, ranking

I. MAIN RESULTS

Singular value decomposition is a fundamental result in matrix theory with vast applications across the sciences, see [5]. In our paper we will consider applications to social networks as well as ranking algorithms of sport teams. For matrices with positive entries the dominant singular vectors have also positive entries and yield the desired applications. For example, in the context of sport team rankings, one dominant singular vector captures the offensive prowess of the teams where as the other one captures their defensive ability. Ranking of sport teams requires knowledge of both and therefore the singular value decomposition of a matrix is of interest.

While common sense might dictate that an excellent offense and excellent defense often go together, there are many instances of teams in sports that are great in offense and yet are weak on defense or conversely. This is most evident in the NBA basketball or NHL hockey. That is why the singular value decomposition of a matrix is needed, yielding these two dominant singular vectors.

As a result, fast approximation techniques for these vectors are of great interest. We provide theoretical results on these approximations and deliver numerical results as well for random matrices that can be reproduced later on. Our simulations are performed in MATLAB ® and are readily reproducible in different computational environments.

A natural question arises as to why there is a need to approximate the dominant singular vector when fast singular vector approximation algorithms are readily available, embedded in the numerical singular value decomposition algorithm. The answer can be two fold. The straightforward approximation techniques given in our paper have low complexity and therefore are readily to be implemented for larger matrices. Second, and more importantly, the approximation techniques presented in our paper are quite intuitive approximations for the dominant singular vectors. Therefore, they can be easily understood by an undergraduate student and thus it is of interest to see how these perform numerically.

We will use three norms with integral values as exponents, namely the l^1, l^2 and l^∞ norms. These norms are commonly used in applications and using non integral values for the norm exponents would yield unnecessary computational complexity. In particular, the euclidean norm l^2 is defined for a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T \in \mathbf{R}^n$ as $||x||_2^2 = = x_1^2 + x_2^2 + \cdots + x_n^2$. Similarly, the norm l^1 is defined for a vector $\mathbf{x} \in \mathbf{R}^n$ as $||x||_1 = |x_1|+|x_2|+\cdots+|x_n|$. The norm l^∞ is defined for a vector $\mathbf{x} \in \mathbf{R}^n$ as $||\mathbf{x}||_\infty = \max\{|x_i|\}$.

Given any square matrix, a physical matrix from a given specific application setting, it is easy to test numerically whether the ranking due to the approximated dominant singular vectors, arising from any norm used, is the same as the ranking due to the actual dominant singular vectors. Moreover, it is straightforward to measure the difference between the approximated dominant singular vector and the actual one, done any any norm of interest. To determine which norm performs better overall many matrices have to be tested as we do not provide closed form descriptions for which matrices do better in which norm. To this end, we rely on Matlab ® built in algorithms to generate matrices with random entries in the relevant range. For each matrix thus generated we compare which norm performs better.

As a result our main focus here is not a theoretical approach to provide results, rather it is based on numerical simulation models that are easily attainable for undergraduate students, where faculty and students can work together and the results given can be readily understood. Having said that, our paper does have some theoretical aspect, mainly the power method.

We would like to make a connection of our results here to the study of dynamical systems in analyzing Lyapunov instability. We direct the reader to some references on this topic [1] and [7].

We now present some needed nomenclature. A square matrix A is said to be non-singular if there exists a square matrix A^{-1} so that $AA^{-1} = A^{-1}A = I$, where I denotes the identity matrix, a matrix with ones on the diagonal and zeros elsewhere. A square matrix A is said to be singular if A fails to be non-singular.

Let A be a $n \times n$ non-singular positive matrix (all entries are positive real numbers). Consider the singular value decomposition of A (SVD)

$$A = UDV^*.$$

The column vectors $\{u_i\}_{i=1}^n$ of the matrix U (an orthonormal set of vectors) and the column vectors $\{v_i\}_{i=1}^n$ of the matrix V (an orthonormal set of vectors) are referred to as the singular vectors. In particular, we have $Av_i = \sigma_i u_i$, where σ_i is the (i,i) entry in the diagonal matrix D, referred to as the singular value of the matrix A. The first singular vector v_1 corresponds to the singular value σ_1, the largest singular value of A called the dominant singular value. We refer to v_1 as the dominant singular vector, the same terminology is used for the corresponding u_1. The Perron-Frobenius Theorem ensures the existence of these with v_1 being unique (unless two or more of the singular values tie for dominance) and moreover, according to the Theorem, the entries in v_1 and u_1 must be positive.

In [4] it was suggested that a good approximation $\widehat{v}_1$ for v_1 is given by the normalized version of

$$\widehat{v}_1 = \sum w_i \mathbf{a}_i, \tag{1}$$

where $\mathbf{a}_i$ is the ith row of A and w_i is its length. In the paper the length was understood as the euclidean l^2 norm and it was asserted the approximation performed well numerically. The following result gives a partial justification for this approximation.

Theorem 1. The sum

$$\widehat{v}_1 = \sum w_i \mathbf{a}_i,$$

where $w_i = ||\mathbf{a}_i||_2$ is the one step iteration of the power method for the eigenvector corresponding to the largest eigenvalue of A^*A with the initial vector $\mathbf{x}_0$ where the vector $\mathbf{x}_0$ has the same angle θ with all the vectors $\mathbf{a}_i$ and $||\mathbf{x}_0||_2 = \sec(\theta)$.

Proof: Consider

$$\widehat{v}_1 = \sum w_i \mathbf{a}_i,$$

where $w_i = ||\mathbf{a}_i||_2$. Rewriting we have

$$\widehat{v}_1 = A^* D_2 \mathbf{e},$$

where

$$D_2 = \begin{pmatrix} ||\mathbf{a}_1||_2 & 0 & 0 & \cdots & 0 \\ 0 & ||\mathbf{a}_2||_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & ||\mathbf{a}_n||_2 \end{pmatrix},$$

and $\mathbf{e} = (1, 1, \ldots, 1)^T$. We have

$$\widehat{v}_1 = A^* D_2 \mathbf{e} = A^* A \mathbf{x}_0.$$

Therefore we have to solve

$$A\mathbf{x}_0 = D_2 \mathbf{e},$$

which translates to

$$\begin{aligned} \operatorname{diag}[\mathbf{a}_i \cdot \mathbf{x}_0] &= \operatorname{diag}[||\mathbf{a}_i||_2 ||\mathbf{x}_0||_2 \cos(\theta_i)] = \\ &= \operatorname{diag}[||\mathbf{a}_i||_2], \end{aligned}$$

where θ_i is the angle between the vector $\mathbf{a}_i$ and the vector $\mathbf{x}_0$. Thus we have

$$\cos(\theta_i) = \frac{1}{||\mathbf{x}_0||_2},$$

which is constant for all i.

Naturally we can change the norm w_i in the equation 1. In particular we can set $w_i = ||\mathbf{a}_i||_1$, the l^1 norm of $\mathbf{a}_i$, which is the sum of its entries. We get the following result.

Theorem 2. The approximation

$$\widehat{v}_1 = \sum w_i \mathbf{a}_i,$$

where $w_i = ||\mathbf{a}_i||_1$ is the one step iteration of the power method for the eigenvector corresponding to the largest eigenvalue of A^*A with the initial vector $\mathbf{x}_0 = (1, 1, \ldots 1)^T$.

Proof: Consider

$$\widehat{v}_1 = A^* D_2 \mathbf{e} = A^* A \mathbf{x}_0,$$

with

$$D_1 = \begin{pmatrix} ||\mathbf{a}_1||_1 & 0 & 0 & \cdots & 0 \\ 0 & ||\mathbf{a}_2||_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & ||\mathbf{a}_n||_1 \end{pmatrix}.$$

We have to solve

$$A\mathbf{x}_0 = D_2 \mathbf{e},$$

yielding $\mathbf{x}_0 = \mathbf{e}$ and the result follows.

We can generalize to further l^p norms, in particular l^∞, where the norm of a vector is defined by the maximum of the absolute values of its entries. We can think of the equation 1 as a one step iteration of the power method for A^*A starting with some vector $\mathbf{x}_0$. We will show some Banach space geometry for the various choices of the norms.

The unweighted option approximation consists of implemented column sums (unweighted) in the matrix A. Thus the jth entry in the approximation vector is just the unweighted jth column sum of A. This vector can also be used in the approximation for the vector v_1. The given MATLAB ® simulations below will show its performance for certain random matrices.

We will consider the normalized vector $\widehat{v}_1$ from now on for our approximation results. To continue with the needed nomenclature, a Banach space is a complete vector space which is equipped with a norm. If the norm is induced by an inner product, then the Banach space in question is called a Hilbert space. Normalization in the l^2 norm is a standard procedure, however normalization in the l^1 norm is shown here to yield better results overall, in particular better approximation results for the dominant singular vectors. The results obtained are versatile and applications extend to any relevant areas such as sport team rankings or ranking of social group interactions. We do not provide a description for the matrices where the l^1 norm yields better approximation results, however, we provide numerical, empirical and probabilistic results, where we simulate with random matrices. Employing the three norms l^1, l^2 and l^∞ is done due to practical considerations, as the norm $l^{3/2}$ for example would be less intuitive and more numerically intense.

II. GEOMETRY OF SOLUTIONS

We will explain the geometry how the initial vector $\mathbf{x}_0$, for the power method iteration, is obtained from the row vectors $\mathbf{a}_i$.

- **The case of l^2.** Consider the row vector $\mathbf{a}_i$. Normalize in the l^2 norm and obtain the vector $\overline{\mathbf{a}}_\mathbf{i}$. The vector $\overline{\mathbf{a}}_i$ can be thought of as the supporting functional for the vector $\mathbf{a}_i$. Consider the hyperplane passing the row point $\overline{\mathbf{a}}_i$ whose normal vector is the vector $\overline{\mathbf{a}}_i$ (or equivalently $\mathbf{a}_i$). Implement this for all row vectors $\mathbf{a}_i$ and an intersection point of all the hyperplanes, as a vector, is the desired vector $\mathbf{x}_0$. The vector $\mathbf{x}_0$ has the same angle θ with the vectors $\mathbf{a}_i$ and its norm is $\sec(\theta)$.
- **The case of l^1.** Consider the row vector $\mathbf{a}_i$. Choose the vector $(1, 1 \ldots, 1)^T$. This vector can be thought of as the supporting functional to the vector $\mathbf{a}_i$ in the Banach space with the l^1 norm. Consider the hyperplanes passing through the point $(1, 1, \ldots, 1)^T$ whose normal vectors are the vectors $\mathbf{a}_i$ for each i. Clearly an intersection point is the point $(1, 1, \ldots, 1)^T$.
- **The case of l^∞.** Consider the row vector $\mathbf{a}_i$. Locate the maximum entry in the location j and consider the vector $\mathbf{e}_j$ with all zero entries except one in the location j. If we have multiple entries for the maximum entry then we can choose any value. Consider the hyperplane, for each row vector $\mathbf{a}_i$, passing through the point $\mathbf{e}_j$ whose normal is the row vector $\mathbf{a}_i$. Implement this for all row vectors $\mathbf{a}_i$ and an intersection point of all the hyperplanes is the desired vector $\mathbf{x}_0$.

In general, in all of these norm cases the solution vector $\mathbf{x}_0$ lies in the first quadrant. The maximum norm performs the worst similar to the unweighted column case. For more on introduction on Banach spaces and supporting functionals we refer the reader to [6].

III. MATLAB ® SIMULATIONS

We choose random matrices of various sizes with natural numbers as entries drawn from $[1, 100]$ uniformly distributed. Entry in Tab. 1 is a pair consisting of the mean and the standard deviation on the maximum norm discrepancy between the v_1 and $\widehat{v}_1$, $||v_1 - \widehat{v}_1||_\infty$. We average out over $10,000$ random matrices.

For larger matrices the discrepancy is smaller, however it is interesting to note that the 3×3 matrices perform better than the 5×5 counterparts for all the three norms in question, but not the unweighted option. Fig. 1 is the graph for the distribution of the maximum norm discrepancy between v_1 and $\widehat{v}_1$ for million random matrices drawn as above. The vector $\widehat{v}_1$ is obtained via the l^1 norm and it is implemented for 10×10 random matrices described as above.

Tab. 1. The mean and the standard deviation on the maximum norm discrepancy between the v_1 and $\widehat{v}_1$, $||v_1 - \widehat{v}_1||_\infty$

	unweighted	l^∞	l^2	l^1
3×3	{0.0266 0.0204}	{0.0128 0.0107}	{0.0036 0.0049}	{0.0057 0.0067}
5×5	{0.0256 0.0139}	{0.0192 0.0120}	{0.0069 0.0058}	{0.0064 0.0047}
10×10	{0.0150 0.0055}	{0.0130 0.0050}	{0.0045 0.0020}	{0.0030 0.0014}
20×20	{0.0071 0.0019}	{0.0064 0.0018}	{0.0023 0.0007}	{0.0012 0.0004}
50×50	{0.0022 0.0004}	{0.0022 0.0004}	{0.0007052 0.000139}	{0.0002384 0.0000518}

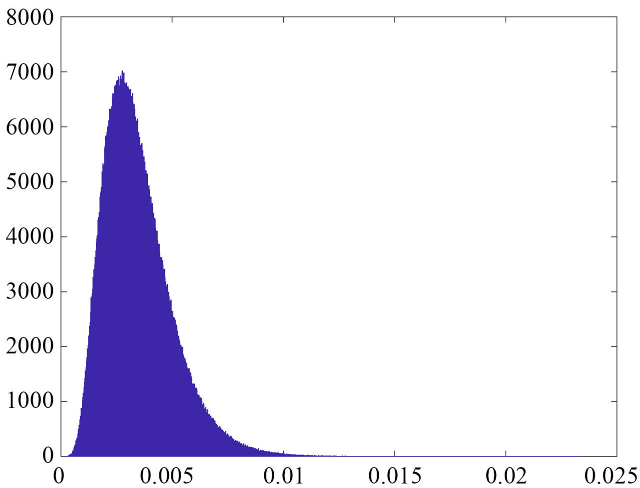

Fig. 1. Distribution of the maximum norm discrepancy between v_1 and $\widehat{v}_1$ for million random matrices

IV. RANK ONE MATRICES

Let $\mathbf{x}$ and $\mathbf{y}$ be unit vectors in $\mathbf{R}^n$. Consider the rank one matrix

$$A = \mathbf{x}\mathbf{y}^T.$$

We implement the SVD decomposition as

$$\left[\mathbf{x}\mathbf{x}^\perp\right]^T \left[\mathbf{x}\mathbf{y}^\perp\right] \left[\mathbf{y}\mathbf{y}^\perp\right] = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

The dominant singular value is $\sigma_1 = 1$ and $v_1 = \mathbf{y}$. We have

$$Av_1 = Ay = (xy^\perp)y = x = u_1.$$

Note that the approximation for both v_1 by $\widehat{v}_1$ and u_1 by $\widehat{u}_1$ is exact, regardless of the norm. In particular we have

$$\widehat{v}_1 = v_1 \text{ and } \widehat{u}_1 = u_1.$$

For a very nice exposition on this topic we refer the reader to [8].

V. SOCIAL NETWORKS

Suppose we have a working group of n people interacting in a group. The interaction quality of the (i, j) interaction is assessed on a quantitative scale $[1, 5]$ with the assessment being a natural number. The (i, j) entry measures how much the i individual enriched the j individual in the interaction. As a result we obtain a $n \times n$ matrix A with positive entries. Note that A need not be a symmetric matrix. The vector v_1 ranks the participants by the amount of benefit they receive from all the group interactions they had, higher the value the more benefit the participant received from the group. On the other hand, the vector u_1 ranks the participants by the amount of benefit they give to the other participants, higher the value the more benefit the participant gave to all the group members. The diagonal entry measures the participant's ability to perform in the group as an individual. Recall we have

$$Av_1 = u_1.$$

We give an example with $n = 4$

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

Here the entry $(2, 3)$ is 2 and it indicates how much the second member enriched the third member in the group interaction. We obtain

$$v_1 = [0.6049\ 0.3671\ 0.3622\ 0.6067],$$

$$u_1 = [0.5388\ 0.4794\ 0.4895\ 0.4901],$$

and

$$\widehat{v}_1 = [0.6015\ 0.3696\ 0.3629\ 0.6082],$$

$$\widehat{u}_1 = [0.5386\ 0.4799\ 0.4897\ 0.4897].$$

We see that in this example the rankings of participants induced by v_1 and $\widehat{v}_1$ are the same, highest to lowest, 4, 1, 2 and 3. Similarly, the rankings of participants induced by u_1 and $\widehat{u}_1$ are the same as well, highest to lowest, 1, 4, 3 and 2.

We choose $10,000$ random 4×4 matrices with entries natural numbers in $[1,5]$ uniformly distributed. Note we have 5^{16} of all such matrices. We record the number of times we have erroneous ranking via the approximation vectors (either for v_1 or u_1), see Tab. 2. By an erroneous ranking we mean a ranking of participants given by $\widehat{v}_1$ and $\widehat{u}_1$ that is different from the ranking given by v_1 and u_1.

Tab. 2. Number of erroneous rankings for 4×4 matrices with entries natural numbers in $[1,5]$ uniformly distributed

	unweighted	l^∞	l^2	l^1
4×4	4040	2725	913	657

For more reading on network analysis we refer the reader to [3].

VI. SPORT TEAM RANKINGS

Consider a matrix A where the entries in A are positive values reflecting a certain sport team ability. For example the entry (i,j) might indicate the number of runs a team i accomplished against the team j in a baseball game. The vector v_1 can be thought of as a vector that ranks the teams defensive ability (higher the value weaker on the defense) and the vector u_1 can be thought of as a vector that ranks the team offensive ability (higher the value stronger on the offense). The diagonal entry (i,i) is now the average of the ith row and the ith column as the teams do not play themselves. Recall we have

$$Av_1 = u_1.$$

We refer to reader to [2] for ideas on rankings of sport teams using linear algebra techniques. For more discussion of ranking techniques in the context of the singular value decomposition see [4].

We choose $10,000$ random matrices of various sizes with entries natural numbers in $[1,100]$ uniformly distributed. We record the number of times we have erroneous ranking via the approximation vectors (either for v_1 or u_1), see Tab. 3. By an erroneous ranking we mean a ranking of sport teams given by $\widehat{v}_1$ and $\widehat{u}_1$ that is different from the ranking given by v_1 and u_1.

Tab. 3 indicates the results for ranking using the approximation vectors are not stellar, the l_1 norm performing the best. The ranking using the approximation vectors performs better for smaller size matrices. We now choose $10,000$ random matrices of various sizes with entries natural numbers in $[50,100]$ uniformly distributed, which is more realistic in sport performance. We record the number of times we have erroneous ranking via either one of the approximation vectors (either for v_1 or u_1), see Tab. 4.

Tab. 3. Number of erroneous rankings for matrices of various sizes with entries natural numbers in $[1,100]$ uniformly distributed

	unweighted	l^∞	l^2	l^1
3×3	2508	972	377	325
5×5	6651	5183	2421	1539
10×10	9877	9708	7056	4620
20×20	10000	10000	9853	8033
50×50	10000	10000	10000	9913

Tab. 4. Number of erroneous rankings for matrices of various sizes with entries natural numbers in $[50,100]$ uniformly distributed

	unweighted	l^∞	l^2	l^1
3×3	916	434	36	31
5×5	3067	2528	272	182
10×10	7759	7340	1244	690
20×20	9921	9898	3627	1750
50×50	10000	10000	8504	4206

The approximation results are better now, especially for smaller size matrices. For example for a 5×5 matrix the chance of an erroneous ranking in this setting appears to be less than 2%.

VII. THE l^1 NORM VS. l^2 NORM

Putting ranking aside let us consider the question as to how well the vector $\widehat{v}_1$ approximates the vector v_1 in the l^∞ norm. It turns out in larger size matrices the l^1 norm performs better than the l^2 norm. The fact that the l^1 norm performs better than the l^2 norm is equivalent to asserting that the one step power method for the largest eigenvalue of A^*A starting with the vector $(1,1,\ldots,1)^T$ approximates the vector v_1 better than starting the one step power method with the vector x_0, a vector that has the same angle θ with all the row vectors $\mathbf{a}_i$ of A. The vector x_0 has magnitude $\sec(\theta)$.

Fig. 2 plots the proportion of random matrices with entries natural numbers uniformly distributed in $[1,5]$ of various sizes for which the l^1 norm performs better than the l^2 norm. The horizontal axis plots the sizes of the matrices and the vertical axis plots the proportion. We have simulated our results on $100,000$ random matrices. For matrices of sizes $2,3$ and 4 the l^1 norm performs worse overall than the l^2 norm and for sizes $n=5$ and larger it is the reverse.

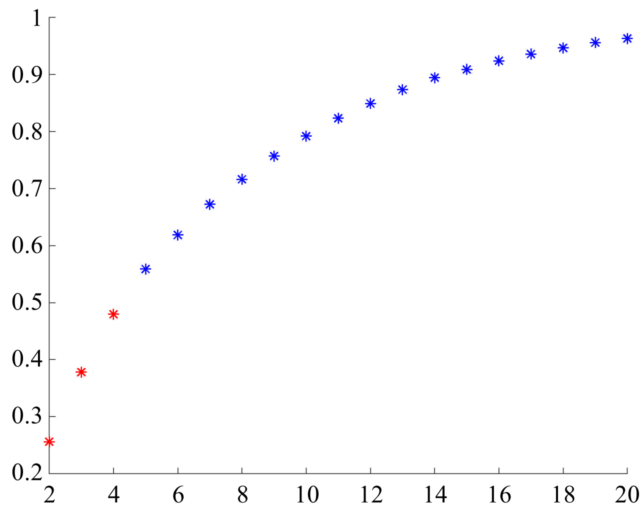

Fig. 2. Proportion of random matrices with entries natural numbers uniformly distributed in $[1, 5]$ for which the l^1 norm performs better than the l^2 norm. The horizontal axis plots the sizes of the matrices

References

[1] G. Benettin, L. Galgani, A. Giorgilli, J.M. Strelcyn, *Lyapunov Characteristic Exponents for smooth dynamical systems and for hamiltonian systems; A method for computing all of them. Part 2: Numerical application*, Meccanica **15**, 21–30 (1980).

[2] C. Hepler, P. Thangarajah, P. Zizler, *Ranking in Professional Sports: An Application of Linear Algebra for Computer Science Students*, 21st Western Canadian Conference on Computing Education, Kamloops, BC, Canada (2016).

[3] R.A. Hanneman, M. Riddle, *Concepts and Measures for Basic Network Analysis*, The Sage Handbook of Social Network Analysis, SAGE, 346–347 (2011).

[4] D. James, C. Botteron, *Understanding Singular Vectors*, The College Mathematics Journal **44**(3), 220–226 (2013).

[5] P. Lancaster, M. Tismenetsky, *The Theory of Matrices*, Academic Press (1985).

[6] S. Montesinos, P. Zizler, V. Zizler, *An Introduction to Modern Analysis*, Springer (2015).

[7] I. Shimada, T. Nagashima, *A Numerical Approach to Ergodic Problem of Dissipative Dynamical Systems*, Progress of Theoretical Physics **61**(6), 1605–1616 (1979).

[8] G. Strang, *Linear Algebra and Its Applications*, Cengage (previously Brooks/Cole), 4th edition (2006).

Peter Zizler is an associate professor in the Department of Mathematics and Computing at Mount Royal University in Calgary. His research work is in linear algebra, applied statistics and mathematics education. He likes to teach a variety of courses including calculus, statistics, linear algebra and courses in general education. In his spare time he likes to ride his Harley Davidson motorcycle as well as his electric bike.

Pamini Thangarajah is a full professor in the Department of Mathematics and Computing at Mount Royal University in Calgary. Her research interests are in the areas of algebra, representation theory, invariant theory, mathematics education and data envelopment analysis. She is interested in contributing to and adopting open educational resources (OER) in her classes. She is a recent recipient of the NSERC PromoScience grant and Provost's Teaching and Learning Enhancement grant.

Mandana Sobhanzadeh is an assistant professor in the Department of General Education at Mount Royal University in Calgary. Her research work is in science, technology, engineering and mathematics (STEM) education. Currently her research interests lie in the area of general education and the pedagogical research work associated. She is involved in many outreach activities to promote physics and mathematics in general education and society at large.

 DOI:10.12921/cmst.2019.0000048