Annales Fennici Mathematici
Volumen 49, 2024, 621–629

On semi-orthogonal matrices with row vectors of equal lengths

Kalle Leppälä

Abstract. When does a rectangular matrix with an orthonormal set of column vectors have row vectors of equal lengths? The column spaces of such matrices are multidimensional generalizations of the projection plane used in isometric perspective. We show that in the absence of unexpected linear relations, any rectangular matrix can be row-scaled so that if we were to orthonormalize the column vectors, the row vectors would attain equal lengths in the process. We use Grassmann coordinates to reduce the question into an instance of the famous matrix scaling problem, and with the help of existing theory introduce simple numerical solutions.

Semi-ortogonaalisista matriiseista joiden rivivektorit ovat saman pituisia

Tiivistelmä. Milloin suorakulmaisella matriisilla on ortonormaalit sarakevektorit ja keskenään yhtä pitkät rivivektorit? Tällaisen matriisin sarakeavaruus yleistää isometrisessä perspektiivissä käytetyn projektiokuvauksen korkeampiin ulottuvuuksiin. Osoitamme että ilman yllättäviä lineaarisia riippuvuuksia minkä hyvänsä suorakulmaisen matriisin rivit voidaan skaalata siten, että mikäli ortonormeeraisimme sarakevektorit niin rivivektorit päätyisivät keskenään saman pituisiksi. Palautamme ongelman Grassmann-koordinaattien avulla erikoistapaukseksi tunnetusta matriisien skaalausongelmasta ("matrix scaling problem"), ja esittelemme alan teoriaan nojaten yksinkertaisia numeerisia ratkaisuja.

1. Introduction

An orthogonal (square) matrix has column vectors of unit length perpendicular to each other, forming what we call an *orthonormal set*. Because in a group a left inverse is also a right inverse, the row vectors form an orthonormal set as well. We will next observe that this property can't carry on to rectangular (non-square) matrices. Following the textbook definition of Abadir and Magnus [1], we set:

Definition. When $G \in \mathbb{R}^{m\times n}$, $m > n$ (resp. $m < n$), and $G^\top G = I$ (resp. $GG^\top = I$), we call the rectangular matrix G *semi-orthogonal*. That is, the column (resp. row) vectors form an orthonormal set.

Note that the term semi-orthogonality is sometimes used for other generalizations of orthogonality; when the scalar product of complex vectors has zero real part [6], or *G-matrices* whose transpose is their inverse up to a positive scaling of the row and column vectors [5]. Assuming $m > n$, there's not enough dimensions for the m row vectors to be perpendicular. As all the column vectors have unit length, the sum of all squared elements of G (the squared Frobenius norm) is n. Thereby the row vectors can't all have unit length or this value would have to be m instead—but they can all have an equal length of $\sqrt{n/m}$. The object of this study is the

https://doi.org/10.54330/afm.152122
2020 Mathematics Subject Classification: Primary 15B10; Secondary 15A75.
Key words: Orthogonal, semi-orthogonal, Grassmann coordinates, isometric perspective, axonometry.

semi-orthogonal matrix with row vectors of equal lengths, and our main result is the following theorem.

Theorem 1. *Let $G \in \mathbb{R}^{m\times n}$, $m > n$, and assume any set of n row vectors of G is linearly independent. There exists a diagonal matrix $D \in \mathbb{R}_+^{m\times m}$, unique up to a positive scalar factor, and an invertible matrix $X \in \mathbb{R}^{n\times n}$ such that DGX is semi-orthogonal and the Euclidean length of each row vector of DGX is $\sqrt{n/m}$.*

It's not clear that solutions must always exist. But it's a simple calculation to see that if (D, X) is a solution, then so is (D, XQ), where $Q \in \mathbb{R}^{n\times n}$ is orthogonal. This means that the requirement on the row vector lengths of the matrix DG is in fact a property of the column space $\mathrm{col}(DG)$. The Grassmann coordinate system is a way to characterize linear subspaces. In Section 2 we cover results on Grassmann coordinates necessary for the subsequent sections, including the connection to matrix row lengths (Lemma 3) and the effect of row scaling (Lemma 4). In Section 3 we examine the desired property of the linear subspace; that the row vectors have equal lengths in any matrix whose column vectors form an orthonormal basis. We make an analogue to isometric projection used in technical drawing, and are motivated to label the property *isometric axonometry*. Subspaces with isometric axonometry are multidimensional generalizations of the projection plane of the usual isometric perspective. Row scaling a matrix so that its column space has isometric axonometry can now be seen as an instance of generalized matrix scaling problem, on a multidimensional array of Grassmann coordinates. In Section 4 we point out this connection, and draw from the rich literature on matrix scaling problems to conclude that Theorem 1 is in fact an unconstrained convex optimization problem. Finally, in Section 5, we demonstrate how this convex optimization problem can be solved numerically by operating directly on matrix G, without ever explicitly working out the Grassmann coordinates—which for large m and n would be an unfeasible task.

2. Grassmann coordinates

An n-dimensional subspace W of $\mathbb{R}^m$ is uniquely described by its Grassmann (or Plücker) coordinates in a subset of the $\binom{m}{n}$-dimensional real projective space, see [7] for a deeper treatment than presented here.

Let W be an n-dimensional subspace of $\mathbb{R}^m$, and $G = [\, r_1^\top \ \cdots \ r_m^\top \,]^\top \in \mathbb{R}^{m\times n}$ any matrix whose column space $\mathrm{col}(G)$ is W. By the notation $G_{\mathcal{S}}$ we mean the submatrix $[\, r_{i_1}^\top \ \cdots \ r_{i_n}^\top \,]^\top \in \mathbb{R}^{n\times n}$, where $\mathcal{S} = \{i_1, \ldots, i_n\} \subseteq \{1, \ldots, m\}$, $i_1 < \cdots < i_n$. The Grassmann coordinate of W associated to a set $\mathcal{S} \subseteq \{1, \ldots, n\}$, $|\mathcal{S}| = n$ is defined as

$$p_{\mathcal{S}} = \det(G_{\mathcal{S}}). \tag{1}$$

Of course the matrix G is not uniquely determined by the subspace W, as any matrix GX, where $X \in \mathbb{R}^{n\times n}$ is invertible, has the same column space. Using GX instead of G in (1) scales each coordinate $p_{\mathcal{S}}$ by $\det(X)$. This is why two $\binom{m}{n}$-tuples of coordinates are considered equivalent if they only differ by a nonzero constant ratio; equivalence classes of nonzero $\binom{m}{n}$-tuples form the homogeneous coordinate system in the $\binom{m}{n}$-dimensional real projective space. The key fact about Grassmann coordinates is that no two distinct subspaces have the same homogeneous coordinates.

Typically homogeneous coordinates are standardized with a projection onto an affine plane (and points at infinity), for example by fixing one coordinate to one. Instead, we choose representatives of the equivalence classes using the (Euclidean) l^2-norm.

Definition. The Grassmann coordinates $p_{\mathcal{S}}$ of an n-dimensional subspace W of $\mathbb{R}^m$ are *l^2-standardized*, if

$$\sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,m\}\\|\mathcal{S}|=n}} p_{\mathcal{S}}^2 = 1.$$

The following two lemmas concern the Grassmann coordinates of the column space of a semi-orthogonal matrix with m rows and n orthonormal columns.

Lemma 2. *Let $m \geq n$ and $G \in \mathbb{R}^{m\times n}$ be an orthogonal ($m = n$) or a semi-orthogonal ($m > n$) matrix. Then the Grassmann coordinates of* $\mathrm{col}(G)$ *are l^2-standardized.*

Proof. First, the case $n = 1$ is trivial as then the sole column of G has unit length. Thus, let $n > 1$. Without loss of generality, we may also assume that

$$G = \begin{bmatrix} c & 0 \\ C & H \end{bmatrix},$$

where $c \in \mathbb{R}$, $0^\top \in \mathbb{R}^{n-1}$, $C \in \mathbb{R}^{m-1}$ and $H \in \mathbb{R}^{(m-1)\times(n-1)}$, because multiplying G by any $n \times n$ orthogonal matrix from the right preserves orthonormality of the column vectors and the Grassmann coordinates given by (1).

We use induction on m. When $m = n$ the claim is immediate because the determinant of an orthogonal matrix is either 1 or -1. For $m > n$, Laplace expansions along the first row when $1 \in \mathcal{S}$ give

$$\begin{aligned}\sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,m\}\\|\mathcal{S}|=n}} p_{\mathcal{S}}^2 &= \sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,m\}\\|\mathcal{S}|=n}} \det(G_{\mathcal{S}})^2 = \sum_{\substack{\mathcal{S}\subseteq\{2,\ldots,m\}\\|\mathcal{S}|=n}} \det(G_{\mathcal{S}})^2 + \sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,m\}\\|\mathcal{S}|=n\\1\in\mathcal{S}}} \det(G_{\mathcal{S}})^2 \\ &= \sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,m-1\}\\|\mathcal{S}|=n}} \det([\,C\ H\,]_{\mathcal{S}})^2 + c^2 \sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,m-1\}\\|\mathcal{S}|=n-1}} \det(H_{\mathcal{S}})^2 \\ &= \sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,m-1\}\\|\mathcal{S}|=n}} \det([\,C\ H\,]_{\mathcal{S}})^2 + c^2.\end{aligned}$$

In the last step we used induction hypothesis on the semi-orthogonal matrix H. Now if C is the zero vector then $\det([\,C\ H\,]_{\mathcal{S}}) = 0$ for all the sets $\mathcal{S}$. The claim follows from $c = 1$, implied by the semi-orthogonality of G. If the vector C has positive length, we denote $D = \mathrm{diag}(1/|C|, 1, \ldots, 1)$ and calculate

$$\begin{aligned}\frac{1}{|C|^2} \sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,m-1\}\\|\mathcal{S}|=n}} \det([\,C\ H\,]_{\mathcal{S}})^2 &= \det(D)^2 \sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,m-1\}\\|\mathcal{S}|=n}} \det([\,C\ H\,]_{\mathcal{S}})^2 \\ &= \sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,m-1\}\\|\mathcal{S}|=n}} \det(([\,C\ H\,]D)_{\mathcal{S}})^2 = 1.\end{aligned}$$

In the last step we used induction hypothesis on the semi-orthogonal matrix $[\,C\ H\,]D$. By the semi-orthogonality of G we have $|C|^2 + c^2 = 1$. □

The converse is not true: multiplying any matrix $G \in \mathbb{R}^{m\times n}$ with a scalar c scales each determinant $\det(G_{\mathcal{S}})$ by c^n, and so the l^2-standardization can be achieved without changing the angles between the column vectors of G.

Lemma 3. *Let $m \geq n$, $G = [\, r_1^\top \ \cdots \ r_m^\top \,]^\top \in \mathbb{R}^{m\times n}$ be a semi-orthogonal matrix and $p_{\mathcal{S}}$ be the Grassmann coordinates of* $\operatorname{col}(G)$*. Then*

$$\sum_{\substack{\mathcal{S}\subseteq\{1,\dots,m\}\\ |\mathcal{S}|=n\\ i\in\mathcal{S}}} p_{\mathcal{S}}^2 = |r_i|^2$$

for each $i \in \{1,\dots,m\}$.

Proof. It suffices to show the claim for $i = 1$ only, as changing the order of rows only switches the signs of some of the Grassmann coordinates, which are all squared in the formula. Furthermore, we may assume that $r_1 = [\, |r_1| \ 0 \ \cdots \ 0 \,]$, as multiplying G by any $n \times n$ orthogonal matrix from the right preserves the lengths of the row vectors, orthonormality of the column vectors, and the column space.

When $n = 1$, the claim is self evident. When $n > 1$, we denote by H the $(m-1)\times(n-1)$ submatrix obtained by removing the first row and the first column of G. Laplace expansions along r_1 give simply

$$\sum_{\substack{\mathcal{S}\subseteq\{1,\dots,m\}\\ |\mathcal{S}|=n\\ 1\in\mathcal{S}}} p_{\mathcal{S}}^2 = \sum_{\substack{\mathcal{S}\subseteq\{1,\dots,m\}\\ |\mathcal{S}|=n\\ 1\in\mathcal{S}}} \det(G_{\mathcal{S}})^2 = |r_1|^2 \sum_{\substack{\mathcal{S}\subseteq\{1,\dots,m-1\}\\ |\mathcal{S}|=n-1}} \det(H_{\mathcal{S}})^2 = |r_1|^2.$$

The last step used Lemma 2 and inductively the semi-orthogonality of H. □

The effect of multiplying a matrix $G \in \mathbb{R}^{m\times n}$ from the left by a diagonal matrix $D = \operatorname{diag}(d_1,\dots,d_m)$ on the Grassmann coordinates $p_{\mathcal{S}}$ of $\operatorname{col}(G)$ is easily described. The ith row of G gets multiplied by d_i, and so does any determinant that contains the ith row, in other words, any $p_{\mathcal{S}}$ with $i \in \mathcal{S}$. We have derived:

Lemma 4. *Let $G \in \mathbb{R}^{m\times n}$ and $D = \operatorname{diag}(d_1,\dots,d_m)$. If we denote the Grassmann coordinates of* $\operatorname{col}(G)$ *by $p_{\mathcal{S}}$, then the Grassmann coordinates of* $\operatorname{col}(DG)$ *are*

$$\pi_{\mathcal{S}} = p_{\mathcal{S}} \prod_{j\in\mathcal{S}} d_j.$$

3. Isometric axonometry

To emphasize the fact that the property we desire for a matrix is in fact a property of its column space, we make a new definition:

Definition. We say that an n-dimensional subspace W of $\mathbb{R}^m$ has *isometric axonometry* (with respect to the natural basis), if

$$|Pe_1| = \cdots = |Pe_m| = \sqrt{\frac{n}{m}},$$

where P is the orthogonal projection from $\mathbb{R}^m$ onto W and $e_1,\dots,e_m$ the natural basis vectors.

The word isometric means distance preserving and axonometry refers to measuring along the axes—each of which undergo a uniform rate of foreshortening when projected onto W. We have adapted the term from the discipline of technical drawing, where the isometric projection is one of the many commonly used parallel projections from $\mathbb{R}^3$ onto $\mathbb{R}^2$ [12]. Precisely, the isometric projection is the orthogonal projection from the three-dimensional space onto the plane $\{[x,y,z]^\top \in \mathbb{R}^3 \colon x+y+z = 0\}$. This is one of the four two-dimensional subspaces of $\mathbb{R}^3$ with isometric axonometry, others being $\{[x,y,z]^\top \in \mathbb{R}^3 \colon -x+y+z = 0\}$, $\{[x,y,z]^\top \in \mathbb{R}^3 \colon x-y+z = 0\}$

and $\{[x,y,z]^\top \in \mathbb{R}^3 \colon x+y-z=0\}$. Obviously a subspace has isometric axonometry if and only if its orthogonal complement has isometric axonometry, in this case the four lines $\{[ax,bx,cx]^\top \in \mathbb{R}^3\}$, $a,b,c \in \{\pm 1\}$. For $m=4$ the spaces $\{[w,x,y,z]^\top \in \mathbb{R}^4 \colon aw+bx+cy+dz=0\}$, $a,b,c,d \in \{\pm 1\}$ (where the projection of the hypercube is the rhombic dodecahedron) and $\{[ax,bx,cx,dx]^\top \in \mathbb{R}^4\}$, $a,b,c,d \in \{\pm 1\}$ are the eight subspaces with isometric axonometry of dimensions 3 and 1, respectively, while for $n=2$ there exists an infinite 2-dimensional manifold of such spaces. As a special case of a constructive theorem of Nisnevich and Bryzgalov [10] (English translation in [11]) subspaces with isometric axonometry exist whenever $n \le m$; typically there are infinitely many of them.

Theorem 5. *Let W be an n-dimensional subspace of $\mathbb{R}^m$. The following properties of W are equivalent:*

- *The subspace W has isometric axonometry.*
- *Each row of the matrix $G = [\,c_1 \ \cdots \ c_n\,]$ has Euclidean length $\sqrt{n/m}$ whenever $\{c_1,\dots,c_n\}$ is an orthonormal basis of W.*
- *The l^2-standardized Grassmann coordinates $p_{\mathcal{S}}$ of W satisfy*

$$\sum_{\substack{\mathcal{S}\subseteq\{1,\dots,m\}\\ |\mathcal{S}|=n\\ i\in\mathcal{S}}} p_{\mathcal{S}}^2 = \frac{n}{m}$$

for each $i \in \{1,\dots,m\}$.

Proof. Let $\{c_1,\dots,c_n\}$ be any orthonormal basis of W and let $[\,g_{i,1} \ \cdots \ g_{i,n}\,]$ be the ith row of the matrix $G = [\,c_1 \ \cdots \ c_n\,]$. Because the columns $\{c_1,\dots,c_n\}$ are orthonormal, the projection matrix from $\mathbb{R}^m$ onto $\operatorname{col}(G) = W$ is $P = GG^\top$. Now

$$Pe_i = GG^\top e_i = [\,c_1 \ \cdots \ c_n\,][\,g_{i,1} \ \cdots \ g_{i,n}\,]^T = g_{i,1}c_1 + \cdots + g_{i,n}c_n,$$

and by the orthonormality of the vectors $\{c_1,\dots,c_n\}$ we have

$$|Pe_1| = \sqrt{g_{i,1}^2 + \cdots + g_{i,n}^2} = |[\,g_{i,1} \ \cdots \ g_{i,n}\,]|.$$

Suppose this length is l for each $i \in \{1,\dots,m\}$. Then l has to be $\sqrt{n/m}$ because

$$n = \sum_{j=1}^{n} 1 = \sum_{j=1}^{n}\sum_{i=1}^{m} g_{i,j}^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} g_{i,j}^2 = \sum_{i=1}^{m} l^2 = ml^2.$$

The rest of the claim follows from Lemma 3. □

4. Generalized matrix scaling problem

From now on, keep in mind that no n row vectors of G are linearly dependent, in other words, all Grassmann coordinates are nonzero. Theorem 1 can be written as a problem of scaling the Grassmann coordinates $p_{\mathcal{S}}$ of $\operatorname{col}(G)$. Find $D = \operatorname{diag}(d_1,\dots,d_m) \in \mathbb{R}_+^{m\times m}$ such that

$$\sum_{\substack{\mathcal{S}\subseteq\{1,\dots,m\}\\ |\mathcal{S}|=n\\ i\in\mathcal{S}}} p_{\mathcal{S}}^2 \prod_{j\in\mathcal{S}} d_j^2 = \frac{n}{m} \tag{2}$$

for each $i \in \{1,\dots,m\}$. According to Lemma 4, the terms $p_{\mathcal{S}}^2 \prod_{j\in\mathcal{S}} d_j^2$ are squared Grassmann coordinates of $\operatorname{col}(DG)$. Orthonormalizing the columns of DG by multiplying with $X \in \mathbb{R}^{n\times n}$ from the right scales each sum (2) by $\det(X)^2$. It follows from

Theorem 5 that $\det(X) \in \{\pm 1\}$ i.e. that the Grassmann coordinates of $\operatorname{col}(DG)$ were already l^2-normalized, and that the Euclidean length of each row vector of DGX is $\sqrt{n/m}$.

When $n = 2$, this is an instance of D_1AD_1 scaling problem ([9] contains a survey of research), which again is a special case of the extensively studied matrix scaling problem (thorough review in [8]). To be precise, let $A \in \mathbb{R}^{\binom{m}{2}\times\binom{m}{2}}$ be the hollow (zero diagonal) symmetric matrix with strictly positive off-diagonal elements $a_{ij} = p^2_{\{i,j\}}$. Now (2) is equivalent to finding $D_1 = \operatorname{diag}(d_1^2, \dots, d_m^2)$ such that each row and column sum of D_1AD_1 is equal. No solution is known for general matrices, but it follows from independent works of Brualdi, Parter and Schneider [4] or Sinkhorn and Knopp [15] that the matrix D_1 exists when A is nonnegative and has total support (there's a sum S of permutation matrices that has zeros at exactly the same coordinates as A has) and the solution is unique when A is fully indecomposable (can't be transformed with row and column permutations into a 2×2 block form where the diagonal blocks are nonempty square matrices and at least one of the other blocks is zero)—both conditions are true in our case. (In fact those works are about the general matrix scaling problem D_1AD_2, but the existence of a solution for symmetric A with $D_1 = D_2$ is implied by uniqueness [8, 14].)

When $n > 2$, equation (2) becomes an instance of a generalized scaling problem. The following theorem is due to Rothblum [13] who interpreted a similar framework of Bapat and Raghavan [2] as an unconstrained convex optimization problem.

Theorem 6. (Rothblum [13]) *Let $a \in \mathbb{R}^N_+$, $b \in \mathbb{R}^M_+$ and $C \in \mathbb{R}^{M\times N}$. If $\mathcal{K} = \{x \in \mathbb{R}^N_+ : Cx = b\} \neq \emptyset$, then there exists a unique $d \in \mathbb{R}^M_+$ such that $Ca' = b$, where $a' \in \mathbb{R}^N_+$ is defined by*

$$a'_j = a_j \prod_{i=1}^{M} d_i^{C_{i,j}}.$$

To prove Theorem 1, we just use equation (2) and Theorem 6. Take $M = m$, $N = \binom{m}{n}$, $a = (p^2_{\mathcal{S}_1} \cdots p^2_{\mathcal{S}_{\binom{m}{n}}})^\top$, $b = (2n/m \cdots 2n/m)^\top$ and C defined by

$$C_{i,j} = \begin{cases} 2 & \text{when } i \in \mathcal{S}_j, \\ 0 & \text{otherwise.} \end{cases}$$

Each row of the matrix C contains exactly $\binom{m-1}{n-1}$ twos and zeros otherwise, so we can pick $x = (m(n-1)!(m-n)!/(2n(m-1)!) \cdots m(n-1)!(m-n)!/(2n(m-1)!))^\top$ to show that $\mathcal{K} \neq \emptyset$.

5. Numerical optimization

As mentioned, Rothblum proves [13] Theorem 6 by interpreting the problem as an unconstrained convex optimization problem with the objective function

$$\begin{aligned} f(\log d_1, \dots, \log d_m) &= \sum_{j=1}^{N} a_j \prod_{i=1}^{M} d_i^{C_{i,j}} - \sum_{k=1}^{M} b_k \log d_k \\ &= \sum_{\substack{\mathcal{S}\subseteq\{1,\dots,m\} \\ |\mathcal{S}|=n}} p^2_{\mathcal{S}} \prod_{j\in\mathcal{S}} d_j^2 - \sum_{i=1}^{m} \frac{2n}{m} \log d_i. \end{aligned} \tag{3}$$

While the Grassmann coordinates can be theoretically illuminating, the sheer quantity $\binom{m}{n}$ of them renders them impractical for directly solving (approximate) D in Theorem 1 by minimizing the objective (3). However, they are of use indirectly: suitable manipulation of the matrix G itself correspond to familiar optimization techniques on the objective (3), which we can then minimize without ever explicitly computing the Grassmann coordinates. Let us examine the partial derivatives of the objective function,

$$\frac{\partial}{\partial \log d_i} f(\log d_1, \ldots, \log d_m) = \sum_{\substack{\mathcal{S} \subseteq \{1,\ldots,m\} \\ |\mathcal{S}|=n \\ i \in \mathcal{S}}} 2p_{\mathcal{S}}^2 \prod_{j \in \mathcal{S}} d_j^2 - \frac{2n}{m}. \tag{4}$$

By Lemmas 3 and 4, this is simply twice the difference between the squared Euclidean length $|r_i|^2$ of the ith row of semi-orthogonalized G, and its desired value n/m.

An immediately obvious idea (Algorithm 1) for a numerical approach to Theorem 1 is to alternate between forcing the columns to be orthonormal by multiplying by an $n \times n$ matrix from the right (the Gram–Schmidt process or the rectangular QR-decomposition), and forcing the row lengths to $\sqrt{n/m}$ by multiplying by a scaling diagonal matrix from the left. Each of the two operations "ruins" the other.

Algorithm 1 Fully parallel coordinate descent

Require: $G = [\, r_1^\top \ \cdots \ r_m^\top \,]^\top \in \mathbb{R}^{m \times n}$, all sets of n rows linearly independent
Ensure: $D \in \mathbb{R}_+^{m \times m}$ diagonal, $\operatorname{col}(DG)$ has isometric axonometry
  $d_i \leftarrow 1$ for all $i \in \{1, \ldots, m\}$
  repeat
    $G \leftarrow Q$, where $G = QR$ is the rectangular QR-decomposition of G
    for $i \in \{1, \ldots, m\}$ **do**
      $r_i \leftarrow \sqrt{n} r_i / (\sqrt{m} |r_i|)$
      $d_i \leftarrow \sqrt{n} d_i / (\sqrt{m} |r_i|)$
    end for
  until convergence
  $D \leftarrow \operatorname{diag}(d_1, \ldots, d_m)$
  return D

In practical experiments (not presented) this method converged fine, but in light of our new understanding of the partial derivatives (4) it's theoretically naive. Forcing the ith row length to its desired value $\sqrt{n/m}$ is nothing but minimizing the objective (3) along the ith coordinate, assuming the matrix G is semi-orthogonal. But for that to hold for each coordinate, we'd have to perform the QR-step after every time scaling a row.

With this alteration, the algorithm (Algorithm 2) is just the familiar coordinate descent, which in our case (convex and differentiable objective in convex and closed domain, minimizing along coordinates always unique) is guaranteed [3] to converge. That guarantee comes with a high cost, as the computational complexity of each iterate increases from roughly $\mathcal{O}(mn^2)$ in Algorithm 1 to roughly $\mathcal{O}(m^2n^2)$ in Algorithm 2. We can view Algorithm 1 as a fully parallel version of the coordinate descent algorithm; our synchronization step, the QR-decomposition, is particularly heavy so we're highly incentivized for parallelization, but unfortunately no suitable convergence criterion for parallel coordinate descent is known despite active research [16].

Algorithm 2 Coordinate descent

Require: $G = [\, r_1^\top \ \cdots \ r_m^\top \,]^\top \in \mathbb{R}^{m\times n}$, all sets of n rows linearly independent
Ensure: $D \in \mathbb{R}_+^{m\times m}$ diagonal, $\operatorname{col}(DG)$ has isometric axonometry
  $d_i \leftarrow 1$ for all $i \in \{1, \dots, m\}$
  repeat
    for $i \in \{1, \dots, m\}$ **do**
      $G \leftarrow Q$, where $G = QR$ is the rectangular QR-decomposition of G
      $r_i \leftarrow \sqrt{n} r_i / (\sqrt{m} |r_i|)$
      $d_i \leftarrow \sqrt{n} d_i / (\sqrt{m} |r_i|)$
    end for
  until convergence
  $D \leftarrow \operatorname{diag}(d_1, \dots, d_m)$
  return D

Because we know the partial derivatives (4), we can also consider the gradient descent (Algorithm 3). The gradient is not globally Lipschitz-continuous, but inside any lower contour set of the objective (3) it is, and so Algorithm 3 converges as long as the step size is at most the inverse of the local Lipschitz-constant. Thus, Algorithm 3 is has a theoretical guarantee of convergence when the step size γ is small enough, while keeping the computational complexity of each iterate at roughly $\mathcal{O}(mn^2)$—the best of both worlds compared to Algorithms 1 and 2.

Algorithm 3 Gradient descent

Require: $G = [\, r_1^\top \ \cdots \ r_m^\top \,]^\top \in \mathbb{R}^{m\times n}$, all sets of n rows linearly independent
Ensure: $D \in \mathbb{R}_+^{m\times m}$ diagonal, $\operatorname{col}(DG)$ has isometric axonometry
  $d_i \leftarrow 1$ for all $i \in \{1, \dots, m\}$
  repeat
    choose a suitable step size γ
    $G \leftarrow Q$, where $G = QR$ is the QR-decomposition of G
    for $i \in \{1, \dots, m\}$ **do**
      $\nabla_i \leftarrow 2(|r_i|^2 - n/m)$
      $r_i \leftarrow r_i / \exp(\gamma \nabla_i)$
      $d_i \leftarrow d_i / \exp(\gamma \nabla_i)$
    end for
  until convergence
  $D \leftarrow \operatorname{diag}(d_1, \dots, d_m)$
  return D

Algorithms 1–3 only return the matrix D but not X of Theorem 1. This is because instead of keeping track on what transformations R^{-1} from the right we do to G during each Gram-Schmidt process, it's easier to just solve for X after the iteration is finished. We can take $X = R^{-1}$ where $DG = QR$ is the QR-decomposition of DG. Note that while D is unique up to a scalar factor, X is not. Because $\operatorname{col}(DG)$ has isometric axonometry, any map transforming the set of column vectors of DG into an orthonormal set would do.

Acknowledgements. The author is thankful to Dr. Ilkka Launonen for fruitful discussions on convex optimization. The work was supported by Academy of Finland Center of Excellence in Tree Biology (TreeBio AoF CoE 346139).

References

[1] Abadir, K. M., and J. R. Magnus: Matrix algebra. - Econom. Exerc. 1, Cambridge Univ. Press, 2005.

[2] Bapat, R. B., and T. E. S. Raghavan: An extension of a theorem of Darroch and Ratcliff in loglinear models and its application to scaling multidimensional matrices. - Linear Algebra Appl. 114, 1989, 705–715.

[3] Bertsekas, D.: Convex optimization algorithms. - Athena Scientific, 2015.

[4] Brualdi, R. A., S. V. Parter, and H. Schneider: The diagonal equivalence of a nonnegative matrix to a stochastic matrix. - J. Math. Anal. Appl. 16:1, 1966, 31–50.

[5] Fiedler, M., and F. J. Hall: G-matrices. Linear Algebra Appl. 436:3, 2012, 731–741.

[6] Groß, J., G. Trenkler, and S. O. Troschke: On semi-orthogonality and a special class of matrices. - Linear Algebra Appl. 289:1-3, 1999, 169–182.

[7] Hodge, W. V. D., and D. Pedoe: Methods of algebraic geometry: Volume 1. - Cambridge Univ. Press, 1953.

[8] Idel, M.: A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps. - arXiv:1609.06349 [math.RA], 2016.

[9] Johnson, C. R., and R. Reams: Scaling of symmetric matrices by positive diagonal congruence. - Linear Multilinear Algebra 57:2, 2009, 123–140.

[10] Nisnevich, L. B., and V. I. Bryzgalov: On a problem of n-dimensional geometry. - Uspekhi Mat. Nauk 8:4, 1953, 169–172.

[11] Prasolov, V. V.: Problems and theorems in linear algebra. - Transl. Math. Monogr. 134, Amer. Math. Soc., Providence, RI, 1994.

[12] Reddy, K. V.: Textbook of engineering drawing. - BS Publications, 2nd edition, 2008.

[13] Rothblum, U. G.: Generalized scalings satisfying linear equations. - Linear Algebra Appl. 114, 1989, 765–783.

[14] Sinkhorn, R.: A relationship between arbitrary positive matrices and doubly stochastic matrices. - Ann. Math. Statist. 35:2, 1964, 876–879.

[15] Sinkhorn, R., and P. Knopp: Concerning nonnegative matrices and doubly stochastic matrices. - Pacific J. Math. 21:2, 1967, 343–348.

[16] Wright, S. J.: Coordinate descent algorithms. - Math. Program. 151:1, 2015, 3–34.

Received 12 January 2024 • Revision received 25 October 2024 • Accepted 12 November 2024
Published online 13 November 2024

Kalle Leppälä
University of Helsinki
The Organismal and Evolutionary Biology Research Programme
kalle.m.leppala@gmail.com