ROTATION TO PHYSIOLOGICAL FACTORS REVISED

Miroslav Kárný, Martin Šámal and Josef Böhm

Reconstruction of underlying physiological structures from a sequence of images is a long-standing problem which has been solved by factor analysis with a success. This paper tries to return to roots of the problem, to exploit the available findings and to propose an improved paradigm.

1. INTRODUCTION

Analysis of image series is a frequent task in nuclear medicine. Reconstruction of underlying physiological structures represents an important class of problems addressed in this research/application area.

Analysis of a dynamic scintigraphic study represents a prototype of the task addressed. The patient is administered by a radioactively marked tracer and a sequence of planar images is taken above the inspected part of his body. These images are supposed to represent projections of a few compartments each characterized by its specific dynamics of the tracer. The compartments (factors) should be separated for the medical inspection in spite of the fact that their projections overlap.

The start up of the processing by factor analysis is relatively simple. The counts registered above ith pixel of tth picture are put into the entry D_{it} of the data matrix D. Using principal component analysis or singular value decomposition, this data matrix is well approximable by a matrix whose rank n_f coincides with the number of the underlying compartments. Thus it is much lower than dimensions of D. This approximator can be written as a product of two full rank rectangular matrices. The left-hand one M should ideally contain factor images: the brightness above each entry is proportional to the volume of the individual compartment projected to the corresponding pixel. The right-hand matrix multiplier C should ideally contain factor curves: time responses of respective compartments to the applied tracer.

There is, however, infinitely many decompositions with the identical ability to approximate the given data as the joint multiplication of M from right by a regular (rotation) matrix T and C by T^{-1} from left keeps the product unchanged. Majority of the "equivalent" factors is useless for medical interpretation. The choice of a proper T (called rotation problem) is the key obstacle in the discussed applications.

A substantial progress can be recorded in this respect, e.g. [4, 5, 6]. However,

as it happens with any outcome of long-term research, the number of layers forming the current solution is relatively high and mixed from various concepts, ideas and algorithmic steps. As it should be done with any (successful) outcome of such research, a time moment comes to stop, to revise results and to get rid off unnecessary layers. This paper tries to make this in the addressed area and brings a simple and efficient algorithm for solving the important problem of rotation to so called physiological factors.

2. DATA AND BACKGROUND SUBTRACTION

A sequence of images is organized in an (n, m)-matrix D of measured data, consisting of counts above particular pixels, i. e. $D_{it} \in \{0, 1, 2, \ldots\}$. Its tth column corresponds to tth image (tth time) and ith row contains data collected above ith pixel of the linearly ordered images.

Discrete data values measured above particular pixels have the Poisson distribution truncated by a finite sensitivity window of the recording apparatus. The expected value $\overline{E}$ of the data D is unknown and non-negative ($\overline{E} \geq 0$: this and similar matrix inequalities are understood entry-wise). For a given $\overline{E}$, the individual entries of D are conditionally independent.

The overall estimation algorithm to be designed analyses a finer structure of the expected value. Here, we deal with the background subtraction only. It means that the following form of the expected value is assumed.

$$\overline{E}_{it} = E_{it} + B_t, \quad i = 1, \ldots, n; \ t = 1, \ldots, m; \ E \geq 0, \ B \geq 0. \tag{1}$$

This decomposition is not unique. We are, however, searching for the decomposition in which the highest possible part of $\overline{E}$ is attributed to the background (curve) B.

Applied doses are kept as low as possible so that we can assume that the truncation effect can be neglected. Due to the assumed conditional independence of the observed data and form of the expected value we can treat the problem column-wise (image-wise). Thus, for each fixed t, we observe n-vector with entries $d_i = D_{it}$ with the expected value $\overline{e}_i = e_i + b = E_{it} + B_t$, $i = 1, \ldots, n$. The conditional probability of the (conditionally independent) entries is

$$p(d_i|\overline{e}) = \frac{\overline{e}_i^{d_i}}{\Gamma(d_i + 1)} \exp[-\overline{e}_i] \tag{2}$$

where Γ is Euler gamma function.

We need the following simple proposition.

Proposition 1. (Bayesian prediction of Poisson data) Let $\overline{e}_i = \mu$ for $i = 1, \ldots, k$ for some $k \leq n$ and the conjugate (self-reproducing) prior probability density function (pdf) assigned to μ be chosen. It has the form

$$p(\mu) = \frac{\kappa(0)^{d(0)+1}}{\Gamma(d(0) + 1)} \mu^{d(0)} \exp[-\kappa(0)\mu]. \tag{3}$$

It is given by prior "statistics" $\kappa(0) > 0$, $d(0) \geq 0$.

Then, the predictive probability of d_k has the form

$$p(d_k|d_1,\ldots,d_{k-1}) = \frac{[\kappa(k-1)]^{d(k-1)+1}}{[\kappa(k)]^{d(k)+1}} \frac{\Gamma(d(k)+1)}{\Gamma(d(k-1)+1)} \qquad (4)$$

with

$$\kappa(k) = \kappa(k-1) + 1, \quad d(k) = d(k-1) + d_k. \qquad (5)$$

For $k = 1$, the formula (4) gives a prior prediction of data determined by the prior "statistics" $\kappa(0) > 0$, $d(0) \geq 0$ that otherwise initiate the recursions (5).

P r o o f . A direct application of the chain rule and of the formula for predictive pdfs [3]. □

Let us consider a pair of data d_1, d_2 from a single image. They have a common expectation if they reflect just background, cf. (1). Their expectations differ if just one of them reflects background. This simple observation lies in the root of the proposed test. We formulate two hypothesis:

$\mathcal{H}$: expected values of d_1, d_2 coincide with the prior probability 0.5
$\overline{\mathcal{H}}$: expected values of d_1, d_2 differ with the prior probability 0.5.

It holds:

Proposition 2. (Bayesian test on equality of expected values of Poisson data) Let the prior pdf assigned to all considered expected values be in the conjugate form (3).

Then, the test ratio ζ is given by

$$\zeta \equiv \frac{p(\mathcal{H}|d_1, d_2)}{p(\overline{\mathcal{H}}|d_1, d_2)} = \left[\frac{\kappa(1)}{\kappa(2)}\right]^{d(2)+1} \left[\frac{\kappa(1)}{\kappa(0)}\right]^{d(0)+1} \frac{\Gamma(d(2)+1)\Gamma(d(0)+1)}{\Gamma(d(1)+1)\Gamma(d(0)+d_2+1)}. \qquad (6)$$

P r o o f . A direct application of Bayesian identification [3] and the formula (4). □

The Proposition 2 gives a direct hint how to estimate the background value b:

1. Take the smallest value among observations as d_1.

   It is expected to have background as its mean value.

2. Find such a value d_2 for which the ratio ζ in (6) is small enough.

   Often, $\zeta = 1$ suffices.

3. Take the mean value of the data in the range $[d_1, d_2]$ as the point estimate of the background.

   By construction, all data in this range have (with high probability influenced by ζ) a common expectation equal to the background. Then, the problem reduces to estimation of it from conditionally independent data with the Poisson distribution

Remarks.

1. The described algorithm is applicable to typical dimensions ($n \approx 4000$ $m \approx 100$) of the considered application.

   More sophisticated hypothesis could be formulated at the prize of the increased computational burden. The possible gain is conjectured as negligible.

2. Let our prior guess of $\min_i d_i$ be some $\bar{b} > 0$ and we admit that 100 % uncertainty of this guess. With this knowledge we should choose the following prior values of statistics

$$\kappa(0) = 1/\bar{b}, \quad d(0) = 0. \tag{7}$$

3. The solution of the equation for d_2 attaining the critical value can be made before-hand and tabulated. At this moment, a simple straight line $d_2 = 5 + 1.5d_1$ seems to be sufficient for $\zeta \approx 1$.

3. DATA AND THEIR DENOISING

For the background removal, the Poisson character of observed data is important as we have to analyze pixels with low expected values. Hereafter, we concentrate on pixels with higher expected values. Thus, we can approximate the data distribution by a normal one. Recall that entries of the data matrix are assumed to be independent when conditioned on parameters of their distribution. We assume, moreover, that the variance of data is constant. Quality of this approximation can be influenced by a proper normalization of data which is not elaborated here. The simplest initial balance can be gained by normalizing columns of D to a common value. In the context of this paper, it is important that it corresponds to a multiplication of the data matrix from right.

For a notational simplicity, the symbols D, E are preserved even after background removal and the discussed normalization. Thus, the observed data can be written in the form

$$D = E + \mathcal{N} \Leftrightarrow D_{it} = E_{it} + \mathcal{N}_{it}, \quad i = 1, \ldots, n;\ t = 1, \ldots, m \tag{8}$$

where $\mathcal{N}$ is a zero mean (matrix) noise with independent (approximately) normal entries with a constant variance. Let the matrix of expected values E have rank n_f where $n_f < \min(n, m)$ is the number of different dynamic structures (factors) considered. The estimation of the number of factors is discussed in [2]. Here, this number is assumed to be known.

It is well known, e.g. [1], that the maximum-likelihood estimate of a low-rank E can be gained from singular value decomposition of the data matrix $D = S\mathcal{D}V$ where S is an orthogonal (n, n)-matrix, $\mathcal{D}$ an (n, m)-matrix containing only nonzero (non-negative) values on its main diagonal and V is an orthogonal (m, m)-matrix. The constructed estimate is $\widehat{E} = S\widehat{\mathcal{D}}V$ where $\widehat{\mathcal{D}}$ keeps just n_f largest values of $\mathcal{D}$. Practically it is reasonable to compute this estimate without an explicit use of the singular value decomposition as follows.

1. Compute the positive semidefinite symmetric (m, m)-matrix $D'D$.

2. Find orthogonal eigenvectors corresponding to its n_f largest eigenvalues and store them into (m, n_f)-matrix, say $\mathcal{V}$.

3. Evaluate the desired estimate according to the following formula

$$\widehat{E} = D\mathcal{V}\mathcal{V}'. \tag{9}$$

For further considerations, it is important that this procedure multiplies the original data from right only. Let us assume that the noise is suppressed sufficiently by this procedure, i. e. we can practically assume

$$\widetilde{E} \equiv E\mathcal{V}\mathcal{V}' \approx \widehat{E} = D\mathcal{V}\mathcal{V}'. \tag{10}$$

Possible negative entries of $\widetilde{E}$ may be attributed to the imprecisely determined background. It may be and will be corrected in this stage by subtracting the minimum value from each column. With this step, we have $\widetilde{E} \geq 0$.

4. THE KEY PROBLEM

The structure of the expected value E makes the addressed problem specific. We assume that E has rank $n_f < n, m$ as it is a superposition of n_f compartment. Both underlying volumes (factor images) and flows of the traces (factor curves) are non-negative by their physical nature. This explains why we assume (similarly as other authors inspecting the problem) that

$$E = MC, \quad M \geq 0, \quad C \geq 0, \quad \text{rank}(M) = \text{rank}(C) = n_f \tag{11}$$

where $M = [M_{ik}]$ is the (n, n_f)-matrix of non-negative factor images; $C = [C_{kt}]$ is the (n_f, m)-matrix of non-negative factor curves.

The decomposition (11) is not unique. The non-unicity represents the key problem addressed within the research reported. Formally, the problem is caused by the fact that any regular rotation (n_f, n_f)-matrix T, such that $\overline{M} = MT \geq 0$, $\overline{C} = T^{-1}C \geq 0$ defines the decomposition indistinguishable from (11).

We want:

1. To propose conditions with a clear and physiologically acceptable meaning under which the decomposition (11) is unique.

2. To propose a computationally feasible algorithm that provides estimates of M, C in a reasonable computational time for realistic dimensions $n \approx 4000$, $m \approx 100$, $n_f \approx 5$.

5. UNICITY

The proposed solution relies on the widely (but not universally) acceptable assumption that the *factor images M do not overlap completely*. In other words, there are pixels in which just a single factor image manifests itself. It means that we adopt:

Assumption 1. The matrix M in (11) contains a regular diagonal sub-matrix.

Proposition 3. (Rotational Unicity) The matrix M in (11) fulfilling Assumption 1 is unique up to a positive scaling and permutation of its columns.

P r o o f. Let $\overline{M} = MT = \widetilde{M}\widetilde{T} \geq 0$ where $M, \widetilde{M} \geq 0$ contain diagonal matrices. Thus $T, \widetilde{T} \geq 0$. We can assume that the diagonal matrix in M is unit matrix I. This special case can be reached by the same permutation and scaling of both matrices $M, \widetilde{M}$. Let $\widetilde{A} \geq 0$ be the (square) submatrix of $\widetilde{M}$ at the position where the unit matrix is in M and let $A \geq 0$ be the submatrix of M at the position where the diagonal submatrix $\widetilde{D}$ is in $\widetilde{M}$. Then, $T = \widetilde{A}\widetilde{T}$ and $AT = \widetilde{D}\widetilde{T}$. A combination of these equalities leads to $(\widetilde{D} - A\widetilde{A})\widetilde{T} = 0$. This identity and regularity of $\widetilde{T}$ imply that the non-negative matrix $\widetilde{A}\widetilde{D}^{-1} = A^{-1}$. Thus both $\widetilde{A}$ and A have to be diagonal matrices as the non-negative matrix with non-negative inversion has to be diagonal (see below). It implies the claimed unicity $M = \widetilde{M}\widetilde{T}T^{-1} = \widetilde{M}\widetilde{A}^{-1}$, i. e. equality up to normalization.

It remains to show that the non-negative regular matrix A with a non-negative inversion has to be diagonal. Let us make a complete induction over the dimension l of the matrix A.

For $l = 2$,

$$A^{-1} = |A|^{-1} \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}.$$

If the determinant $|A|$ is positive then $A_{12} = A_{21} = 0$ in order to guarantee the non-negativity of both A and A^{-1}. If $|A| < 0$ then $A_{11} = A_{22} = 0$ and A becomes diagonal after a suitable permutation.

For a general l, the matrix A and its inversion are split into blocks

$$A = \begin{bmatrix} \mathcal{A} & c \\ b' & d \end{bmatrix}, \ A^{-1} = \begin{bmatrix} \mathcal{E} & g \\ f' & e \end{bmatrix}$$

where d, e are non-negative scalars and $'$ denotes transposition. The formula for inversion of the block-split matrix implies that

$$e = (d - b'\mathcal{A}^{-1}c) > 0, \ f' = -b'\mathcal{E}/d \geq 0, \ g = -\mathcal{A}^{-1}c/e.$$

Consequently, $f = 0$ and thus $\mathcal{E} = \mathcal{A}^{-1}$ is a diagonal matrix by induction and $g = 0$. $\square$

In the rest of the paper, the assumed diagonal matrix is normalized to unit matrix.

6. SOLUTION

First, we describe the algorithm DIAMAX whose complexity is well within the acceptable range. Then we prove that it provides information needed for obtaining the desired solution.

Algorithm DIAMAX

1. Drop non-positive rows of $\widetilde{E}$.

2. Normalize rows of the gained matrix so that sum of their elements equals to unity.

   The resulting matrix is denoted G.

3. Define the row vector with n_f entries

$$C^0 = \frac{1}{n} \sum_{i=1}^{m} G_{i*}$$

   where G_{i*} denotes ith row of G.

4. For $l = 1, \ldots, n_f$ select

$$C^l \in \mathrm{Arg} \max_{\{G_{i*}\}_{i=1}^{n}, G_{i*} \notin \{C^j\}_{j=1}^{l-1}} \sum_{j=1}^{l} ||G_{i*} - C^{j-1}||$$

   where $||x|| = \max_{j \in \{1, \ldots, n_f\}} |x_j|$.

5. Define estimate $\widehat{C}$ of the normalized filtered curves $\widetilde{C} = C \mathcal{V} \mathcal{V}'$

$$\widehat{C}' = \begin{bmatrix} C^1 \\ \vdots \\ C^{n_f} \end{bmatrix}. \tag{12}$$

Note that the algorithm is finite and requires $\approx n * m * n_f$ operations. Only the step 2 contains operations corresponding with (diagonal) multiplication of data from left, i. e. up to this minor and recoverable distortion, the factor images M are untouched during the processing proposed.

With the gained results, it is straightforward to find the desired estimates:

1. Define the estimate $\widehat{M}$ of factor images M by the formula

$$\widehat{M} = \widetilde{E}\widehat{C}(\widehat{C}\widehat{C}')^{-1}. \tag{13}$$

2. Find the final estimate $\widehat{B}$ of the background B as the mean of the difference (the *original* data $-$ estimate of the expectation E), i. e. mean of the rows in the matrix $D - \widehat{M}\widehat{C}'$.

3. Compute the final estimates of C from the regression on *original data without background*

$$D - \mathbf{1}_n \widehat{B}' = \widehat{M}C' + \mathcal{N}. \tag{14}$$

   where $\mathbf{1}_n$ denotes n-vector of units.

Proposition 4. (Properties of DIAMAX algorithm) Under assumptions of Proposition 3, DIAMAX determines the filtered curves $\widetilde{C}$ corresponding (uniquely) to the matrix of the factor images M that is searched for.

P r o o f. The definition of G and non-negativity of M imply that rows of G are convex combinations of the rows of the matrix $\widetilde{C}$ normalized to unit row sums. These normalized rows we denote C^l, $l = 1, \ldots, n_f$. Moreover, the vector C^0 is by construction strictly in the non-degenerated (due to the full rank of $\widetilde{C}$) convex hull spanned over $\widetilde{C}$.

For any G_{i*}, $i = 1, \ldots, m$, it holds

$$\|G_{i*} - C^0\| = \left\|\sum_{j=1}^{n_f} \alpha_j^i C^j - C^0\right\| = \left\|\sum_{j=1}^{n_f} \alpha_j^i (C^j - C_0)\right\|$$

$$\leq \sum_{j=1}^{n_f} \alpha_j^i \|(C^j - C^0)\| \leq \max_{k=1,\ldots,n_f} \|C^k - C^0\|$$

where α_j^i are the weights of the convex combination creating G_{i*} from C^j, $j = 1, \ldots, n_f$. Their non-negativity and unit sum are exploited in the second equality and the both inequalities. Otherwise just the triangle inequality is used which holds for the norm chosen.

If we take such G_{i*} that coincides with the maximizer in the last inequality we see that the upper bound is reached for this vector. Thus, the first step of DIAMAX selects a row in the normalized $\widetilde{C}$.

For a generic l and any G_{i*}, $i = 1, \ldots, m$, $G_{i*} \notin \{C^0, \ldots, C^{l-1}\}$ let us consider the quantity

$$\sum_{k=1}^{l} \|G_{i*} - C^{k-1}\| \leq \sum_{k=1}^{l}\sum_{j=1}^{n_f} \alpha_j^i \|C^j - C^{k-1}\| \leq \max_{C^j \notin \{C^i\}_{i=1}^{l-1},\, j=1,\ldots,n_f} \sum_{k=1}^{l} \|C^j - C^{k-1}\|.$$

The same properties of the convex weights and of the norm are used as above. And again, the upper bound is reached for a still unselected vertex for a still unselected row of the normalized matrix $\widetilde{C}$. This determines the position of the filtered curves in the normalized data (the position where M contains diagonal matrix) as the applied normalization does not change position of diagonals in M. The rest of the computation consists just of least squares solutions of a set of linear equations. The full rank of the involved matrices makes this step simple.   □

7. CONCLUSIONS

This paper brings a new view on the problems of background removal and of rotation in factor analysis that provides an improvement of contemporary solutions.

An additional improvement can be reached by normalizing data before using the singular value decomposition. It should help to overcome the weak point we

see: the real noise variance depends on the level of the signal. The solution of the rotation problem proposed in the paper can be enriched by any scaling (even iterative) procedure whenever it keeps factor images M untouched, whenever it reduces to right-matrix multiplication.

ACKNOWLEDGEMENT

REFERENCES

[1] G. H. Golub and C. F. VanLoan: Matrix Computations. The John Hopkins University Press, Baltimore – London 1989.

[2] M. Kárný and M. Šámal: Bayesian rank estimation with application to factor analysis. Kybernetika *30* (1994), 4, 433–443.

[3] V. Peterka: Bayesian system identification. In: Trends and Progress in System Identification (P. Eykhoff, ed.), Pergamon Press, Oxford 1981, pp. 239–304.

[4] M. Šámal, M. Kárný, W. Backfrieder, K. Kletter and H. Bergmann: Bayesian identification of compartment structures in dynamic scintigraphic data. In: Radioactive Isotopes in Clinical Medicine and Research (H. Bergmann and H. Sinzinger, eds.), Birkhäuser Verlag, Basel 1995, pp. 123–128.

[5] M. Šámal, M. Kárný, H. Sůrová and Z. Dienstbier: Theoretical and experimental basis of the clinical use of factor analysis. In: Radioactive Isotopes in Clinical Medicine and Research (R. Höffer, H. Bergmann, and H. Sinzinger, eds.), Schattauer Verlag, Stuttgart, New York 1991, pp. 200–203.

[6] M. Šámal, M. Kárný and D. Zahálka: Bayesian identification of a physiological model in dynamic scintigraphic data. In: Information Processing in Medical Imaging, Springer 1993, pp. 422–437.

Ing. Miroslav Kárný, DrSc., Ing. Josef Böhm, CSc., Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod vodárenskou veží 4, 182 08 Praha 8, Czech Republic.
e-mails: karny@utia.cas.cz, bohm@utia.cas.cz

MUDr. Martin Šámal, DrSc., Institute of Nuclear Medicine, 1st Faculty of Medicine, Charles University Prague, Salmovská 3, 120 00 Praha 2. Czech Republic.
e-mail: samal@aba.cesnet.cz