98-110

Proc. of the Mathematical Theory of Networks and Systems (MTNS98), Padova, Italy, Jul. 1998, pp.

Solving the Structured Total Least Squares problem using the Riemannian SVD

Cambre Johan
johan.cambre@esat.kuleuven.ac.be
phone: 32-(0)16-321926
fax : 32-(0)16-321970
Bart De Moor
bart.demoor@esat.kuleuven.ac.be
phone: 32-(0)16-321715 (personal)
phone: 32-(0)16-32.17.09 (secretary: Ida Tassens)
fax : 32-(0)16-321970
Dept. of Electr. Eng., ESAT-SISTA
Kardinaal Mercierlaan 94
3001 Heverlee
BELGIUM

Keywords: Structured Total Least squares, Riemannian SVD, continuous time flows, Hankel matrices, rank reduction

Abstract

The structured TLS problem arises in many engineering applications. The Riemannian Singular Value Decomposition is a generalized SVD which takes into account the structure imposed on the rank deficient approximation of the data matrix. We present a possible solution strategy for the Riemannian SVD equations using continuous time flows. We also derive a new min-max characterization of the minimal Riemannian singular value.

1 Structured Total Least Squares (STLS) problem formulation

We study the STLS problem, which can be formulated as follows: Given an affine set of matrices determined by an origin $B_0 \in \mathbb{R}^{p\times q}$ and a basis $B_1, \ldots, B_n \in \mathbb{R}^{p\times q}$.
Let A be a given matrix (structured or not). Find $B = B_0 + b_1B_1 + b_2B_2 + \ldots + b_nB_n$ such that:

- B is rank deficient,
- $\sum_k(\langle A, B_k\rangle - b_k)^2 w_k$ is minimal,

where w_i are user specified weights and $\langle A, B_k\rangle$ is the projection of A on the vector B_k.
Example: Given the set of 5×3 Hankel matrices. Let A be a matrix in this set, find the rank deficient Hankel matrix B such that $\|A - B\|_F$ is minimal.
It was shown in [1] that the Lagrange multiplier equations for the STLS problem can be rewritten as a set of nonlinear SVD-like equations, which will be called the Riemannian SVD equations.

$$\begin{aligned} Av &= D_v u\tau, &(1)\\ A^T u &= D_u v\tau, &(2)\end{aligned}$$

under the conditions

$$\begin{aligned} u^T D_v u &= 1, &(3)\\ v^T v &= 1, &(4)\end{aligned}$$

where D_u and D_v are constructed as

$$D_v = \sum_{i=1}^{n} \frac{1}{w_i} B_i v (B_i v)^T,$$

$$D_u = \sum_{i=1}^{n} \frac{1}{w_i} B_i^T u (B_i^T u)^T.$$

These are square matrices, quadratic in the elements of v and u respectively. In general these matrices are positive definite. The minimizing solution (u, v, τ) of the Riemannian SVD equations corresponds to $\tau = \|A - B\|_F$ and enables us to construct a rank deficient approximation of A in the following way: Suppose $A = \sum_{k=1}^{n} a_k B_k$ with B_k the basis of the imposed linear structure, the rank deficient approximation $B = \sum_{k=1:n} b_k B_k$ is computed as follows:

$$b_k = a_k - u^T B_k v\tau.$$

This means that the norm of the error introduced on the matrix A to get a rank deficient matrix is equal to τ. Therefore we will be looking for the triplet (u, v, τ) corresponding to the minimal τ that solve the Riemannian SVD equations.More details on the Riemannian SVD can be found in [1], [3], [5] and [7].

2 The characterization of the Riemannian Singular Value τ

We will mention two characterizations:

1. From equations (1), (2), (3) and (4) it is easy to derive that
$$\tau^2 = v^T A^T D_v^{-1} Av, \tag{5}$$
provided that D_v is invertible. At a local minimum for this characterization, the corresponding v will solve the Riemannian SVD equations if we take $u = D_v^{-1} Av$.
The minimization problem is not convex, except for some special cases, and requires the inversion of the matrix D_v at every function evaluation/gradient computation.

2. We can also characterize τ in terms of u and v as
$$\tau = \frac{u^T Av}{\sqrt{u^T D_v u}}. \tag{6}$$
At a saddle point for this characterization, the corresponding u and v will solve the Riemannian SVD equations, which leads us to the optimization problem
$$\min_{v \in \mathbb{R}^q} \max_{u \in \mathbb{R}^p} \frac{u^T Av}{\sqrt{u^T D_v u}}. \tag{7}$$

The proof consists of a series of Schur complement arguments, and will be sketched here. Starting from the positive definiteness of the matrix D_v and the positivity of $\tau^2 - v^T A^T D_v^{-1} Av$ in the neighborhood of a minimum v_{opt} for $v^T A^T D_v^{-1} Av$ we can conclude that the matrix

$$\begin{bmatrix} D_{v_{opt}} & Av_{opt} \\ v_{opt}^T A^T & \tau(v)^2 \end{bmatrix}$$

is positive definite, which is equivalent with the positive definiteness of

$$\begin{bmatrix} \tau(v)^2 & v_{opt}^T A^T \\ Av_{opt} & D_{v_{opt}} \end{bmatrix}$$

The Schur complement of this last matrix gives the positivity of $\tau(v)^2$ and the positive definiteness of the matrix

$$D_{v_{opt}} - Av_{opt} \frac{1}{\tau^2} (Av_{opt})^T,$$

from which the minmax criterion can be derived.

3 Finding the saddle point of a function

In this section we will develop some ideas on finding a saddle point for a function.

3.1 Characterization of a saddle point

Suppose we have a function f in two variables $u \in \mathbb{R}^p$ and $v \in \mathbb{R}^q$, we are looking for necessary conditions for a couple (u, v) to be a local maximum in u and a local minimum in v. First of all let's define what we are looking for:

Definition
Let f be a real-valued function in two sets of variables $u \in \mathbb{R}^p$ and $v \in \mathbb{R}^q$.Suppose we denote the tangent vectors to a point $(u_0, v_0) \in \mathbb{R}^{p+q}$ by

$$T_{(u_0,v_0)}\mathbb{R}^{p+q} = T_{u_0}\mathbb{R}^p \oplus T_{v_0}\mathbb{R}^q.$$

A point (u_0, v_0) is called a (local) minmaximum for f if we have that

$\forall \alpha \in T_{u_0}\mathbb{R}^p : f(u_0 + t\alpha) < f(u_0)$ for sufficiently small t,

$\forall \beta \in T_{u_0}\mathbb{R}^q : f(v_0 + t\beta) > f(v_0)$ for sufficiently small t.

The following theorem gives first and second order necessary conditions on f at a local minmaximum:
Theorem If at a point (u_0, v_0) we reach a local maximum in u_0 and a local minimum in v_0 we have that
i)$\forall \alpha \in T_{u_0}\mathbb{R}^p \ \frac{\partial f}{\partial u}\alpha = 0$; $\forall \beta \in T_{v_0}\mathbb{R}^q \ \frac{\partial f}{\partial v}\beta = 0$, this implies that the derivatives in u and v are zero.
ii)If we denote the tangent vectors at the point (u_0, v_0) by $T_{(u_0,v_0)}\mathbb{R}^{p+q} = T_{u_0}\mathbb{R}^p \oplus T_{v_0}\mathbb{R}^q$, we have that

$$\forall a \in T_{u_0}\mathbb{R}^p : \begin{bmatrix} a & 0 \end{bmatrix} \frac{\partial^2 f}{\partial (u,v)^2} \begin{bmatrix} a \\ 0 \end{bmatrix} < 0,$$

$$\forall b \in T_{v_0}\mathbb{R}^q : \begin{bmatrix} 0 & b \end{bmatrix} \frac{\partial^2 f}{\partial (u,v)^2} \begin{bmatrix} 0 \\ b \end{bmatrix} > 0.$$

We can prove that these conditions are also sufficient

3.2 Solving optimization problems with continuous time algorithms

In this section we will review the idea of using continuous time flows for optimization problems, see also [2], [4].

3.2.1 Solving minimization problems with continuous time algorithms

Under general conditions, the optimization problem

$$\min_{v \in \mathbb{R}^q} f(v),$$

is solved by the gradient flow:

$$\dot{v} = -\frac{\partial f}{\partial v}. \tag{8}$$

If we look at the evolution in time of the function value

$$\frac{d}{dt}f(v(t)) = (\frac{\partial f}{\partial t})^T\dot{v} = -\|\frac{\partial f}{\partial t}\|^2 < 0,$$

we can see that the function decreases along the trajectory of $v(t)$ until we reach a point where $\frac{\partial f}{\partial t} = 0$, which is a local minimum for the function. Discrete time algorithms for minimization of $f(v)$ can be interpreted as an interpolation for $v(t)$ in the case that the flow cannot be integrated explicitly.

It is tedious though straightforward to derive a gradient flow to minimize (5). The gradient of $\tau^2(v) = v^T A^T D_v^{-1} Av$ is given by

$$\dot{v} = -\left(2A^T D_v^{-1}Av - \begin{bmatrix} v^T A^T D_v^{-1}(\frac{d}{dv_1}D_v)D_v^{-1}Av \\ \vdots \\ v^T A^T D_v^{-1}(\frac{d}{dv_q}D_v)D_v^{-1}Av \end{bmatrix}\right),$$

where we have used the fact that for a matrix $A(v)$ we have that

$$(\frac{d}{dv}A(v)^{-1}) = -A(v)^{-1}(\frac{d}{dv}A(v))A(v)^{-1}.$$

3.2.2 Solving minmax problems with continuous time algorithms

Inspired by the steepest descent continuous time flow (8) we can suggest the following for the the ascent-descent continuous time flow for the minmax problem:

$$\min_{v\in\mathbb{R}^q}\max_{u\in\mathbb{R}^p} f(u,v),$$

which gives

$$\begin{aligned} \dot{u} &= \tfrac{\partial f}{\partial u}, \\ \dot{v} &= -\tfrac{\partial f}{\partial v}. \end{aligned}$$

In the neighborhood of a min-maximum this flow is guaranteed to converge to a local min-maximum, following the candidate Lyapunov function:

$$L = \|\frac{\partial f}{\partial u}\|^2 + \|\frac{\partial f}{\partial v}\|^2.$$

We find that

$$\dot{L} = \frac{\partial f}{\partial u}^T\frac{\partial^2 f}{\partial u^2}\frac{\partial f}{\partial u} - \frac{\partial f}{\partial v}^T\frac{\partial^2 f}{\partial v^2}\frac{\partial f}{\partial v},$$

which is strictly negative in the neighborhood of a min-maximum implying that the flow converges to a couple satisfying:

$$\begin{aligned} \tfrac{\partial f}{\partial u} &= 0, \\ \tfrac{\partial f}{\partial v} &= 0. \end{aligned}$$

The interested reader can apply the minmax flow to find the the saddle point of $f(u,v) = -u^2 + v^2$.

4 A min-max flow for τ

We apply the ideas of the previous section to the minmax problem related with the Riemannian SVD (7). For the function,

$$\tau = \frac{u^T Av}{\sqrt{u^T D_v u}},$$

the minmax flow is given by:

$$\begin{aligned} \dot{u} &= \tfrac{Av(u^t D_v u) - (u^t Av)D_v u}{\sqrt{u^t D_v u}^3}, \\ \dot{v} &= -\phi\tfrac{A^t u(u^t D_v u) - (u^t Av)D_v}{\sqrt{u^t D_v u}^3} \end{aligned}$$

with $0 < \phi < 1$ a user defined constant to slow down the flow in v. Further analysis is required to determine the role of ϕ for the convergence of the minmax flow. One can observe that for $\phi \approx 0$ the trajectory of the minmax flow in the v-space is similar to the trajectory of the continuous time flow for the minimization of $\tau^2 = v^T A^T D_v^{-1} Av$. The maximization of the minmax flow corresponds with an iterative approximation of $D_v^{-1}Av$ of the minimization flow.

5 Example

We will consider the set of 3×2 Hankel matrices which is a vector-subspace of $\mathbb{R}^{3\times 2}$.

The D_v matrix for the corresponding STLS problem is given by:

$$D_v = \begin{bmatrix} v_1^2 + \frac{1}{2}v_2^2 & \frac{1}{2}v_1v_2 & 0 \\ \frac{1}{2}v_1v_2 & \frac{1}{2}v_1^2 + \frac{1}{2}v_2^2 & \frac{1}{2}v_1v_2 \\ 0 & \frac{1}{2}v_1v_2 & \frac{1}{2}v_1^2 + v_2^2 \end{bmatrix}.$$

Let A be a Hankel matrix

$$A = \begin{bmatrix} -0.0780 & 0.2203 \\ 0.2203 & -0.1009 \\ -0.1009 & 0.1897 \end{bmatrix}.$$

The results of the integration of the minmax can be seen in figures 1 and 2.

Convergence of the minmax flow implies that the gradients of $\tau = \frac{u^T Av}{\sqrt{u^T D_v u}}$, must be equal to zero which means that the Riemannian SVD equations are solved at this point.

The rank deficient approximation is given by

$$B = \begin{bmatrix} -0.1184 & 0.1359 \\ 0.1359 & -0.1560 \\ -0.1560 & 0.1790 \end{bmatrix}.$$

More examples can be found in [2], [4].

6 Acknowledgments

Bart De Moor is a senior Research Associate with the F.W.O. (Fund for Scientific Research-Flanders).
Cambré Johan is a Research Assistant.
This work is supported by:
the Flemish Government (GOA-MIPS, FWO project G.0292.95 and G.0256.97 FWO-ICCoS, IWT-ISIS), the Belgian Federal Government (IUAP P4-02, IUAP P4-24, DWTC project MD/01/24) and the European Commission: SCIENCE-ERNSI, ALAPEDES, SYSIDENT-KIT124.

7 Figures

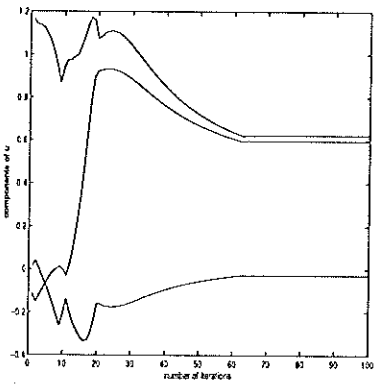

Figure 1: The convergence in the u coordinates

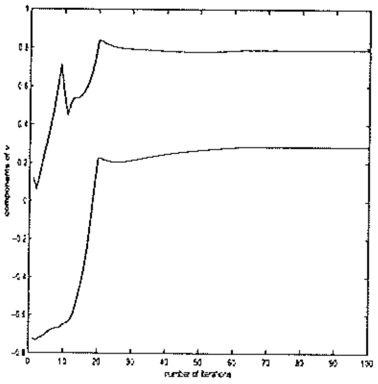

Figure 2: The convergence in the v coordinates

References

[1] De Moor B., "Structured total least squares and L2 approximation problems", Systems Control, Special Issue of Linear Algebra and its Applications on Numerical Linear Algebra, (Van Dooren, Ammar, Nichols, and Mehrmann, eds.), vol.188-189, Jul. 1993, pp. 163-207.

[2] De Moor B., "The Riemannian singular value decomposition", in SVD and Signal Processing, III: Algorithms, Architectures and Applications, (Moonen M., and De Moor B., eds.), Elsevier, 1995, pp. 61-78.

[3] De Moor B., "Structured total least squares and L2 approximation problems", Systems Control, Special Issue of Linear Algebra and its Applications on Numerical Linear Algebra, (Van Dooren, Ammar, Nichols, and Mehrmann, eds.), vol.188-189, Jul. 1993, pp. 163-207 (also IMA Preprint Series no.1036, Institute for Mathematics and its Applications, University of Minnesota, Sep. 1992).

[4] De Moor B., "Continuous-time algorithms for the Riemannian SVD", in Proc. of the 34th IEEE Conference on Decision and Control (CDC'95), New Orleans, Louisiana, Dec. 1995, pp. 1084-1090.

[5] De Moor B., "Total least squares for affinely structured matrices and the noisy realization problem", IEEE Transactions on Signal Processing, vol.42, no.11, Nov. 1994, pp. 3104-3113.

[6] De Moor B., "Linear system identification, structured total least squares and the Riemannian SVD", in Recent Advances in Total Least Squares Techniques and Errors-in-Variables modeling SIAM, Philiadelphia, 1997, pp. 225-238.

[7] De Moor B., "Structured total least squares for Hankel matrices", in 13 Communications, Computation, Control and Signal Processing, (Paulray A., Roychowdhury V., and Schaper C.D., eds.), A tribute to Thomas Kailath, Kluwer Academic Publishers, 1997, pp. 243-258.

[8] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

[9] Roorda B. *Algorithms for Global Total Least Squares Modelling of Finite Multi-variable Time Series.* Automatica, Vol.31, no.3, 1995.

[10] Roorda B., Heij C. *Global Total Least Squares Modelling of Multivariable Time Series.* IEEE Transcations on Automatic Control, Vol.40, no.1, pp.50-63, 1995.

[11] J.C. Willems, From time series to linear system, part I: Finite dimensional linear time invariant systems; part II: Exact modelling; part III: Approximate modelling, *Automatica*, 22/23 (1986/7), pp. 561-580, 675-694, 87-115.