LOCALLY REDUCED-RANK OPTIMAL FILTERING AND ITS APPROXIMATION BY SUCCESSIVE ALTERNATING MINIMIZATION

Isao Yamada and Jamal Elbadraoui

Dept. of Communications and Integrated Systems,
Tokyo Institute of Technology, Tokyo, 152-8552 Japan
E-mails: {isao,jamal}@comm.ss.titech.ac.jp

ABSTRACT

In this paper, we introduce a *locally reduced-rank optimal filtering* that is a generalization of the *globally reduced-rank optimal filtering* studied extensively as a fundamental tool in signal processing applications. After formulating the problem of the *locally reduced-rank optimal filtering*, we present a closed form solution to the problem in terms of the SVD. Moreover, in a way similar to the techniques shown recently by Hua et al, we deduce a numerical algorithm converging globally and exponentially to the solution without passing any computation of the eigen value decomposition (or SVD). Numerical example shows that the proposed algorithm converges efficiently to the locally reduced-rank optimal filter that realizes ideal trade-off between the rank-reduction and the estimation accuracy.

1. INTRODUCTION

In this paper, we address a *locally reduced-rank optimal filtering* that is a generalization of the *globally reduced-rank optimal filtering* studied extensively as a fundamental tool in signal processing applications asking for (i) data or model reduction, (ii) robustness against noise or model errors, or (iii) high computational efficiency [1–10] (A unified treatment of the (globally) reduced-rank filtering is presented in [10]). The current strong demand for reduced-rank filtering arises from the growing disparity between the large number of degrees of freedom in the next generation of wireless communication systems, radar systems, sonar systems, etc., and limitations on sample support size due to high mobility, high sensitivity to small movements/perturbations, etc [11, 12].

Consider the two random processes $\boldsymbol{x}_k \in \mathbb{R}^n$ and $\boldsymbol{y}_k \in \mathbb{R}^m$. Define, for $r \leq \min\{m,n\}$,

$$\mathbb{R}^{m\times n}(r) := \{T \in \mathbb{R}^{m\times n} \mid rank(T) \leq r\}.$$

Then the *(globally) reduced-rank optimal filter* [10] is defined as $T_{Gopt} \in \mathbb{R}^{m\times n}(r)$ that minimizes

$$J_G(T) := E\left[tr\left\{W(\boldsymbol{y}_k - T\boldsymbol{x}_k)(\boldsymbol{y}_k - T\boldsymbol{x}_k)^t\right\}\right], \quad (1)$$

$\forall T \in \mathbb{R}^{m\times n}(r)$, where E denotes statistical expectation (or sample averaging) and $W \in \mathbb{R}^{m\times m}$ is a given positive definite (weighting) matrix.

Let $R_{xx} := E\{\boldsymbol{x}_k\boldsymbol{x}_k^t\} \in \mathbb{R}^{n\times n}$,$R_{yy} := E\{\boldsymbol{y}_k\boldsymbol{y}_k^t\} \in \mathbb{R}^{m\times m}$,and $R_{yx} := E\{\boldsymbol{y}_k\boldsymbol{x}_k^t\} \in \mathbb{R}^{m\times n}$. Then we have

$$T_{Gopt} = W^{-1/2}trun_r\left\{W^{1/2}R_{yx}(R_{xx}^{1/2})^{\dagger}\right\}(R_{xx}^{1/2})^{\dagger}. \quad (2)$$

[Note: (i) The minimizer of J_G with $W = R_{y,y}^{-1}$ agrees with that of $J_{det}(T) := \det\left[E\left\{(\boldsymbol{y}_k - T\boldsymbol{x}_k)(\boldsymbol{y}_k - T\boldsymbol{x}_k)^t\right\}\right]$ (See [10]). (ii) The above result (2) can be derived simply based on Lemma 1 in Sec.2. (iii) For simplicity, in this paper, we only discuss the real filtering although the all discussion can be extended without any difficulty to the corresponding complex filtering.] The reduced-rank constraint and the *(globally) reduced-rank optimal filter* T_{Gopt} are quite attractive not only because finding suitable subspace is a key in a wide range of signal processing [13–16] but also because $T_{Gopt} \in \mathbb{R}^{m\times n}(r)$ well suites to applications including wireless communications and telephone networks [17] where the effects of multi-path signals and cross-signal interference are in general modelled, with multiple sensors and multiple transmitters, by reduced rank matrices [8]. Moreover several successive approximation techniques [9, 10] can generate, without computing any eigenvalue decomposition (or SVD), matrix sequences converging globally to T_{Gopt} under a condition where R_{xx} is nonsingular and $rank(T_{Gopt}) = r$.

On the other hand, by defining (i) submatrices $T^{(j)} \in \mathbb{R}^{m\times n_j}$ $(j = 1,\ldots,p)$ of T by $T = [T^{(1)}T^{(2)}\cdots T^{(p)}] \in \mathbb{R}^{m\times n}$ (hence $\sum_{j=1}^{p} n_j = n$) and (ii) random processes $\boldsymbol{x}_k^{(j)}$ $(j = 1,\ldots,p)$ by $\boldsymbol{x}_k = \left((\boldsymbol{x}_k^{(1)})^t(\boldsymbol{x}_k^{(2)})^t\cdots(\boldsymbol{x}_k^{(p)})^t\right)^t$, we obtain the equivalent expression [of (1)]:

$$J_L(T^{(1)},\ldots,T^{(p)}) := E\left[tr\left\{W\left(\boldsymbol{z}_k\boldsymbol{z}_k^t\right)\right\}\right], \quad (3)$$

where $\boldsymbol{z}_k := \boldsymbol{y}_k - \sum_{j=1}^{p} T^{(j)}\boldsymbol{x}_k^{(j)}$. Each random process $\boldsymbol{x}_k^{(j)}$ $(j = 1, \ldots, p)$ contributes respectively, through $T^{(j)}$, to approximate $\boldsymbol{y}_k$. Obviously this interpretation has substantial meaning in many applications including higher order statistical signal processing. For example, the vector valued version [18] of the Volterra filtering (i.e., optimal polynomial filtering) [19] is a natural higher order generalization, of the (vector valued) Wiener filtering, where $\boldsymbol{x}_k^{(j)}$ can be a vector valued process of which components are monomials, having specified total degrees, of an input random process. Because (i) the major burden of the application of polynomial filtering is in general its huge computational complexity and (ii) the rank-reduction usually yields computationally efficient structure, it would be of great interest in the achievable performance of the polynomial filter using $T^{(j)}$ $(j = 1, \ldots, p)$ satisfying the pre-specified conditions $rank\left(T^{(j)}\right) \leq r_j$. However the results on the globally reduced-rank optimal filtering do not answer to this primitive question because we need analyses in the case where the rank constraints are imposed independently on each $T^{(j)}$.

In this paper, after formulating the problem of the *locally reduced-rank optimal filtering*, we present its solution for $p = 2$ and $r_1 = \min\{m, n_1\}$ in terms of the SVD. Moreover, in a way similar to the techniques in [9, 10], we present a numerical algorithm converging globally to the solution without passing any computation of the eigen value decomposition. Numerical example shows the effectiveness of the proposed algorithm. (Note: All proofs are omitted due to lack of space.)

2. PRELIMINARIES

Throughout this paper, we use the following notations:

$\mathbb{R}$:	the set of all real numbers,
A^t :	the transposition of $A \in \mathbb{R}^{m\times n}$,
$\mathcal{R}(A)$:	the range space of $A \in \mathbb{R}^{m\times n}$,
$A^\dagger$:	the Moore-Penrose pseudoinverse of $A \in \mathbb{R}^{n\times n}$,
$A^{1/2}$:	the positive semi-definite square root of positive semi-definite $A \in \mathbb{R}^{n\times n}$,
$\|A\|$:	the Frobenius norm of $A \in \mathbb{R}^{m\times n}$.

Let $A = U\Sigma V^T$ be the singular value decomposition (SVD) of $A \in \mathbb{R}^{m\times n}$, where $U := [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m] \in \mathbb{R}^{m\times m}$ and $V := [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n] \in \mathbb{R}^{n\times n}$ are unitary matrices, and $\Sigma := \operatorname{diag}\{\sigma_1, \ldots, \sigma_{\min\{m,n\}}\} \in \mathbb{R}^{m\times n}$ is a diagonal matrix with nonnegative diagonal elements arranged in a nonincreasing order: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{m,n\}} \geq 0$. For $r \in \{0, 1, \ldots, \min\{m, n\}\}$, $trun_r(A) := \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$ satisfies $rank\,(trun_r(A)) \leq r$. (For details on the SVD and its numerical computation, see for example [20].)

The following fundamental result is useful in the analyses of the locally reduced-rank optimal filtering.

Lemma 1 *For a given symmetric matrix $A \in \mathbb{R}^{n\times n}$ and a matrix $B \in \mathbb{R}^{m\times n}$, define a function $\Theta : \mathbb{R}^{m\times n} \to \mathbb{R}$ by*

$$\Theta(X) := tr\{W(XAX^t - 2BX^t)\},$$

where $W \in \mathbb{R}^{m\times m}$ is a positive definite (weighting) matrix.

(a) *Suppose that Θ is bounded below. Then A is a positive semi-definite matrix and $\forall X \in \mathbb{R}^{m\times n}$,*

$$\Theta(X) \geq \Theta(BA^\dagger) = -\|W^{1/2}B(A^{1/2})^\dagger\|^2.$$

(The minimizer is unique if A is nonsingular.)

(b) *Suppose that Θ is bounded below. Then for any $r \in \{0, 1, \ldots, \min(m, n)\}$,*

$$X_r^* := W^{-1/2} trun_r\{W^{1/2}B(A^{1/2})^\dagger\}(A^{1/2})^\dagger$$

satisfies $X_r^* \in \mathbb{R}^{m\times n}(r)$ and

$$\begin{aligned}\Theta(X_r^*) &= \Theta(BA^\dagger) + \sum_{i=r+1}^{\min\{m,n\}} \sigma_i^2 \\ &\leq \Theta(X), \forall X \in \mathbb{R}^{m\times n}(r),\end{aligned}$$

where $\{\sigma_i\}_{i=1}^{\min\{m,n\}}$ are singular values of $W^{1/2}B(A^{1/2})^\dagger$, arranged in a nonincreasing order: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{m,n\}} \geq 0$.

3. LOCALLY REDUCED-RANK OPTIMAL FILTER

As announced in Sec.1, we consider the following problem that is a generalization of the globally reduced-rank optimal filtering problem [1–10].

Problem 1 *(Locally reduced-rank optimal filter) Let $\boldsymbol{x}_k^{(j)} \in \mathbb{R}^{n_j}$ $(j = 1, 2, \ldots, p)$ and $\boldsymbol{y}_k \in \mathbb{R}^m$ be random processes. Suppose that we have $R_{ij} := E\left\{\boldsymbol{x}_k^{(i)}(\boldsymbol{x}_k^{(j)})^t\right\} \in \mathbb{R}^{n_i\times n_j}$, $R_{yj} := E\left\{\boldsymbol{y}_k(\boldsymbol{x}_k^{(j)})^t\right\} \in \mathbb{R}^{m\times n_j}$ and $R_{yy} := E\{\boldsymbol{y}_k\boldsymbol{y}_k^t\} \in \mathbb{R}^{m\times m}$, $\forall (i,j) \in \{1, 2, \ldots, p\}^2$. Then the problem is to find $T_{Lopt}^{(j)} \in \mathbb{R}^{m\times n_j}(r_j)$ $(j = 1, 2, \ldots, p)$ satisfying*

$$\begin{aligned}&J_L(T_{Lopt}^{(1)}, \ldots, T_{Lopt}^{(p)}) \leq J_L(T^{(1)}, \ldots, T^{(p)}),\\ &\forall T^{(j)} \in \mathbb{R}^{m\times n_j}(r_j), j = 1, 2, \ldots, p, \qquad (4)\end{aligned}$$

where J_L is defined in (3) and $0 \le r_j \le \min\{m, n_j\}$ $(j = 1, \ldots, p)$ are pre-specified ranks. We call $\{T^{(j)}_{Lopt}\}_{j=1}^{p}$ the locally reduced-rank filter.

Next proposition presents a closed form expression of a locally reduced rank optimal filter for $p = 2$ and $r_1 = \min\{m, n_1\}$.

Proposition 1 *For $p = 2$, $r_1 = \min\{m, n_1\}$ and $r_2 \le \min\{m, n_2\}$, we have a closed form solution to Problem 1 as*

$$\begin{aligned} T^{(1)}_{Lopt} &= (R_{y1} - T^{(2)}_{Lopt} R_{2,1}) {R_{1,1}}^{\dagger}, \\ T^{(2)}_{Lopt} &= W^{-1/2} trun_{r_2} \Big[W^{1/2} (R_{y,2} - R_{y,1} {R_{1,1}}^{\dagger} R_{1,2}) \\ &\qquad \{(R_{2,2} - R_{2,1} {R_{1,1}}^{\dagger} R_{1,2})^{1/2}\}^{\dagger} \Big] \\ &\qquad\quad \{(R_{2,2} - R_{2,1} {R_{1,1}}^{\dagger} R_{1,2})^{1/2}\}^{\dagger}. \end{aligned}$$

(Note: Lemma 1(a) guarantees the positive semi definiteness of $R_{2,2} - R_{2,1} {R_{1,1}}^{\dagger} R_{1,2}$.)

Thanks to Proposition 1, we deduce the next algorithmic solution to Problem 1. The algorithm shown in Proposition 2 can approximate iteratively, without passing any computation of eigen value decomposition or SVD, the locally reduced-rank optimal filter $\{T^{(j)}_{Lopt}\}_{j=1}^{2}$ in Proposition 1 . In particular, Proposition 2 for $R_{i,1} = 0$ $(i = 1, 2)$ reproduces the algorithms in [9, 10].

Proposition 2 *(Successive alternating minimization) Suppose that $R_{2,2} - R_{2,1} {R_{1,1}}^{\dagger} R_{1,2}$ is nonsingular and $rank(T^{(2)}_{Lopt}) = r_2$. Define matrix sequences $\{F(k)\}_{k=1}^{\infty} \subset \mathbb{R}^{m \times r_2}(r_2) \backslash \mathbb{R}^{m \times r_2}(r_2 - 1)$ and $\{G(k)\}_{k=1}^{\infty} \subset \mathbb{R}^{r_2 \times n_2}(r_2) \backslash \mathbb{R}^{r_2 \times n_2}(r_2 - 1)$, with arbitrarily given $G(0) \in \mathbb{R}^{r_2 \times n_2}(r_2) \backslash \mathbb{R}^{r_2 \times n_2}(r_2 - 1)$, by*

$$\begin{aligned} & F(k+1) G(k) \left(R_{2,2} - R_{2,1} {R_{1,1}}^{\dagger} R_{1,2} \right) G^t(k) \\ & \quad = \left(R_{y,2} - R_{y,1} {R_{1,1}}^{\dagger} R_{1,2} \right) G^t(k), \\ & \left(F^t(k+1) W F(k+1) \right) G(k+1) \left(R_{2,2} - R_{2,1} {R_{1,1}}^{\dagger} R_{1,2} \right) \\ & \quad = F^t(k+1) W \left(R_{y,2} - R_{y,1} {R_{1,1}}^{\dagger} R_{1,2} \right). \end{aligned}$$

Then the matrix sequences $\{\mathcal{T}^{(j)}(k)\}_{k=1}^{\infty}$ $(j = 1, 2)$ defined by $\mathcal{T}^{(2)}(k) := F(k) G(k)$ and $\mathcal{T}^{(1)}(k) := (R_{y1} - F(k) G(k) R_{2,1}) {R_{1,1}}^{\dagger}$ satisfy

$$\lim_{k \to \infty} \left\| \mathcal{T}^{(j)}(k) - T^{(j)}_{Lopt} \right\| = 0, (j = 1, 2). \qquad (5)$$

(Note: In (5), $\{\mathcal{T}^{(j)}(k)\}_{k=1}^{\infty}$ converges exponentially.)

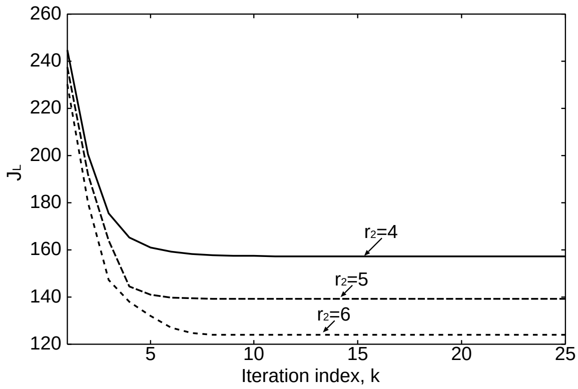

Fig. 1. Convergence of J_L by the formula in Proposition 2

4. NUMERICAL EXAMPLES AND REMARKS

To examine the performance of the proposed locally reduced-rank optimal filtering and its successive approximation technique (in Proposition 2), we consider the approximation, by the locally reduced-rank filter, of the ARMA(2,2) process $(y_k)_{k \in \mathbb{Z}}$ [1, 13] characterized by the rational function $\frac{1 - z^{-2}}{1 - 1.70223 z^{-1} + 0.71902 z^{-2}}$ (This ARMA process corresponds to the USASI signal used vastly to simulate speech signals. Its generation routine can be found in `http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/txt/usasi.txt`). The vector-valued test processes are simply defined as $\boldsymbol{y}_k := (y_k, \ldots, y_{k+m-1})^t$ and $\boldsymbol{x}_k := (x_k, \ldots, x_{k+n-1})^t$, where $(x_k)_{k \in \mathbb{Z}}$ is the noise process of zero mean, i.i.d. Gaussian random variable $\mathcal{N}(0, 1)$ and fed to the ARMA(2,2) to produce $(y_k)_{k \in \mathbb{Z}}$. We set $(m, n) := (50, 40)$ and $(n_1, n_2, r_1) = (10, 30, 10)$, and employed $W := I_{50} \in \mathbb{R}^{50 \times 50}$ (the identity matrix) as the weighting matrix for all filters. We applied the formula in Proposition 2 to obtain the locally reduced-rank optimal filters for $r_2 = 4, 5, 6$. As a fair starting condition, $G(0) = \mathrm{diag}\{1, 1, \ldots, 1\} \in \mathbb{R}^{r_2 \times 30}$ is employed in all cases. Fig.1 dipicts the convergence behaviors of $J_L(k) := J_L\left(\mathcal{T}^{(1)}(k), \mathcal{T}^{(2)}(k)\right)$. Fig.2 shows the convergence behaviors of $Dist(k) := \{\|\mathcal{T}^{(1)}(k) - T^{(1)}_{Lopt}\|^2 + \|\mathcal{T}^{(2)}(k) - T^{(2)}_{Lopt}\|^2\}^{1/2}$. As expected, the proposed algorithm converges efficiently to the locally reduced-rank optimal filter that realizes reasonable trade-off between the rank-reduction and the estimation accuracy.

Finally, we remark that the idea of the algorithm shown in Proposition 2 could be extended naturally to more general cases of Problem 1. Further consideration in this direction will be reported elsewhere.

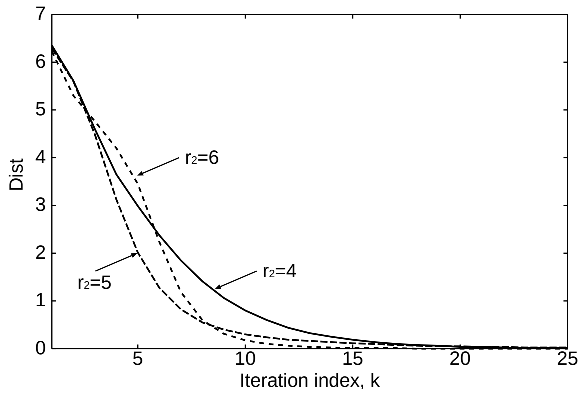

Fig. 2. Convergence of $Dist$ by the formula in Proposition 2

Acknowledgement: The authors wish to thank Prof. K. Sakaniwa of Tokyo Institute of Technology for helpful discussion.

5. REFERENCES

[1] D.R. Brillinger, *Time Series: Data Analysis and Theory*, Holt, Rinehart and Winston, 1975.

[2] L.L. Scharf, *Statistical Signal Processing*, Addison-Wesley, 1991.

[3] L.L. Scharf, "The SVD and reduced rank signal processing," *Signal processing* **25**, pp.113-133, 1991.

[4] L.L. Scharf and J.K. Thomas, "Wiener filter in canonical coordinates for transform coding, filtering, and quantizing," *IEEE Trans. on Signal Processing* **46** pp.647-654, 1998.

[5] L.L. Scharf and C.T. Mullis, "Canonical coordinates and the geometry of inference, rate and capacity," *IEEE Trans. on Signal Processing* **48** pp.824-831, 2000.

[6] K.I. Diamantara and S.Y. Kung, "Multivariate neural networks for reduced rank approximation," *IEEE Trans. Neural Network* **5**, pp.684-697, 1994.

[7] Y. Yamashita and H. Ogawa, "Relative Karhunen-Loeve transform," *IEEE Trans. on Signal Processing* **44**, pp.371-378, 1996.

[8] P. Stoica and M. Viberg, "Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions," *IEEE Trans. on Signal Processing* **44**, pp.3069-3078, 1996.

[9] Y. Hua and M. Nikpour, "Computing the reduced rank Wiener filter by IQMD," *IEEE Signal Processing Letters* **6**, pp.240-242, 1999.

[10] Y. Hua, M. Nikpour, and P. Stoica, "Optimal reduced-rank estimation and filtering," *IEEE Trans. on Signal Processing* **49**, pp.457-469, 2001.

[11] J.S. Goldstein, I.S. Reed and L.L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Information Theory* **44**, pp.2943-2959, 1998.

[12] M.D. Zoltowski, M. Joham, and S. Chowdhury, "Recent advances in reduced-rank adaptive filtering with application to high-speed wireless communications," *Proceedings of SPIE2001*, 2001.

[13] D.G. Manolakis, V.K. Ingle, and S.M. Kogan, *Statistical and adaptive signal processing*, McGraw-Hill, 2000.

[14] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. on Signal Processing* **43**, pp.95-107, 1995.

[15] I. Yamada, K. Slavakis, and K. Yamada, "An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques," *IEEE Trans. on Signal Processing* **50**, pp.1091-1101, 2002.

[16] I. Yamada, N. Ogura and N. Shirakawa, "A numerically robust hybrid steepest descent method for the convexly constrained generalized inverse problems," in *Inverse Problems, Image Analysis, and Medical Imaging,* (Z.Nashed and O. Scherzer, Eds.) *Contemporary Mathematics, Amer. Math. Soc.*, 2002 (in press).

[17] A.J. Paulraj and C.B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Mag.* **14**, pp.49-83, 1997.

[18] I. Yamada, T. Sekiguchi and K. Sakaniwa, "Reduced rank Volterra filter for robust identification of nonlinear systems," *Proceedings of the 2nd International Workshop on Multidimensional (nD) Systems*, pp.171-175, Poland, 2000.

[19] V.J. Mathews and G.L. Sicuranza, *Polynomial Signal Processing*, John Wiley & Sons, 2000.

[20] G.H. Golub and C.F. Van Loan, *Matrix Computations, 3rd ed.* The Johns Hopkins University Press, 1996.