Computing the Reduced Rank Wiener Filter by IQMD

Yingbo Hua, *Senior Member, IEEE,* and Maziar Nikpour

***Abstract*—The reduced-rank Wiener filter (RRWF) is a generic tool for data compression and filtering. This letter presents an iterative quadratic minimum distance (IQMD) algorithm for computing the RRWF. Although it is iterative in nature, the IQMD algorithm is shown to be globally and exponentially convergent under some weak conditions. While the conventional algorithms for computing the RRWF require an order of n^3 flops, the IQMD algorithm requires only an order of n^2 flops at each iteration where n is the dimension of data. The number of iterations required in practice is often small due to the exponential convergence rate of the IQMD.**

***Index Terms*— Data compression, data filtering, generalized Karhunen–Loeve transform, Karhunen–Loeve transform, rank reduction, reduced rank (low rank) Wiener filter, singular value decomposition (SVD), subspace decomposition, Wiener filter.**

I. Introduction

The reduced rank Wiener filter (RRWF) was developed by Scharf [1]–[2] as a tool primarily for estimation and filtering. The RRWF was recently rediscovered via the relative Karhunen-Loeve transform (RKLT) [3] and the generalized Karhunen–Loeve transform (GKLT) [4] in the context of data compression. The GKLT was developed (without the knowledge of the RRWF) as a generalized version of the RKLT. The GKLT, developed for both singular and nonsingular cases, turns out to be identical to the RRWF shown in [1]–[2] for the nonsingular case. In this letter, the RRWF, RKLT, and GKLT are all referred to as RRWF.

The RRWF is an $m \times n$ matrix $\mathbf{T}_{\mathrm{RRWF}}$ of a pre-specified rank r that minimizes the following distance between a reference signal $\mathbf{y}(m \times 1)$ and the reduced-rank transform[1] of another signal $\mathbf{x}(n \times 1)$:

$$J_{\mathrm{RRWF}}(\mathbf{T}) = E\{\|\mathbf{y} - \mathbf{T}\mathbf{x}\|_2^2\} \tag{1}$$

where E denotes statistical expectation (or sample averaging), $\|\ldots\|_2$ the 2-norm, and $r < \min(m,n)$. As shown in [1]–[2], [4], such an optimum transform (with minimum F-norm) is given by

$$\mathbf{T}_{\mathrm{RRWF}} = \mathrm{trun}_r\{\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1/2^H}\}\mathbf{R}_{xx}^{-1/2} \qquad (2)^2$$

where the superscript ^{H} denotes the conjugate transpose, $\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^H\}$, $\mathbf{R}_{yx} = E\{\mathbf{y}\mathbf{x}^H\}$, trun_r denotes rank-r SVD truncation, and $\mathbf{R}_{xx}^{1/2}$ denotes a (symmetric or asymmetric) square root of $\mathbf{R}_{xx}$, and $\mathbf{R}_{xx}^{-1/2}$ is the Moore–Penrose pseudoinverse of $\mathbf{R}_{xx}^{1/2}$. In this letter, $\mathbf{R}_{xx}$ is assumed to be nonsingular, which is what we call the nonsingular case.

While the expression (2) is compact, explicit and fundamental, the computation of $\mathbf{T}_{\mathrm{RRWF}}$ by any conventional algorithm requires $O(n^3)$ flops (assuming $m = n$). It is clear that the SVD truncation may be easily avoided by computing the rank-r principal subspace of $\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1}\mathbf{R}_{yx}^H$ and then pre-multiplying $\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1/2^H}$ by a corresponding projection matrix. But the product of the three matrices, $\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1}\mathbf{R}_{yx}^H$, would require $O(n^3)$ flops in the first place, and the square-root matrix $\mathbf{R}_{xx}^{-1/2}$ would also require $O(n^3)$ flops in general. What is shown next is an iterative algorithm for computing $\mathbf{T}_{\mathrm{RRWF}}$, which requires only $O(n^2)$ flops at each iteration. The number of iterations is small in practice due to the exponential convergence rate of the algorithm.

Note that in some situations the desired rank r needs to be estimated from the data. This can be achieved via the SVD of $\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1/2^H}$ as in the original RRWF [1]–[2]. If the rank r is known apriori, then the SVD is not necessary and the RRWF can be computed as shown next.

II. IQMD Algorithm

Let us reformulate the criterion (1) as follows. Since any rank-r matrix $\mathbf{T}$ can be written as $\mathbf{A}\mathbf{B}^H$ where $\mathbf{A}(m \times r)$ is the reconstruction matrix and $\mathbf{B}(n \times r)$ is the compression matrix, (1) can be rewritten as

$$J_{\mathrm{RRWF}}(\mathbf{A},\mathbf{B}) = E\{\|\mathbf{y} - \mathbf{A}\mathbf{B}^H\mathbf{x}\|_2^2\}. \tag{3}$$

Although the optimal minimizers $\mathbf{A}_{\mathrm{opt}}$ and $\mathbf{B}_{\mathrm{opt}}$ of $J_{\mathrm{RRWF}}(\mathbf{A},\mathbf{B})$ are clearly not unique, the desired product of the optimal minimizers is given by (2), i.e., $\mathbf{T}_{\mathrm{RRWF}} = \mathbf{A}_{\mathrm{opt}}\mathbf{B}_{\mathrm{opt}}^H$.

We now consider minimizing J_{RRWF} by $\mathbf{A}$ and $\mathbf{B}$ alternately as follows. Given a previous estimate $\mathbf{B}(k)$ of $\mathbf{B}_{\mathrm{opt}}$, the new estimate $\mathbf{A}(k+1)$ of $\mathbf{A}_{\mathrm{opt}}$ is obtained by minimizing $J_{\mathrm{RRWF}}(\mathbf{A},\mathbf{B})$ with respect to $\mathbf{A}$ only. Then, given the new estimate $\mathbf{A}(k+1)$, the new estimate $\mathbf{B}(k+1)$ is obtained by minimizing $J_{\mathrm{RRWF}}(\mathbf{A},\mathbf{B})$ with respect to $\mathbf{B}$ only. Since each of the above steps is a quadratic minimization problem, the closed-form expressions of the new estimates in terms

Manuscript received April 3, 1999. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. P. Milanfar.

The authors are with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3052, Australia (e-mail: yhua@ee.mu.oz.au).

Publisher Item Identifier S 1070-9908(99)06745-0.

[1] Or, equivalently, compressed and reconstructed signal.

[2] The superscript H is missing from the expression (10) in [1].

of the given estimates can be easily obtained. After some simple manipulations, the above algorithm can be put into the following form:

$$\mathbf{A}(k+1)(\mathbf{B}^H(k)\mathbf{R}_{xx}\mathbf{B}(k)) = \mathbf{R}_{yx}\mathbf{B}(k) \tag{4}$$
$$(\mathbf{A}^H(k+1)\mathbf{A}(k+1))\mathbf{B}^H(k+1)\mathbf{R}_{xx} = \mathbf{A}^H(k+1)\mathbf{R}_{yx} \tag{5}$$

where the initial estimate $\mathbf{B}(0)$ is randomly selected. The linear equation (4) is solved for $\mathbf{A}(k+1)$ with fixed $\mathbf{B}(k)$ and then the linear equation (5) is solved for $\mathbf{B}(k+1)$ with fixed $\mathbf{A}(k+1)$. By repeating the above process, we have the iterative-quadratic-minimum-distance (IQMD) algorithm. Since $\mathbf{A}$ and $\mathbf{B}$ are "tall" matrices (i.e., assuming $r \ll \min(n, m)$), it is not difficult to see that at each iteration, (4) and (5) can be implemented using only $O(n^2)$ flops (assuming $n = m$). (For examples of numerical solutions of linear equations, see [5]).

Although the IQMD algorithm represented by (4) and (5) is intuitively motivated, it will be shown next that the IQMD algorithm is globally and exponentially convergent to the RRWF.

III. Global and Exponential Convergence

This section proves that the IQMD algorithm globally and exponentially converges to the RRWF under two very weak conditions that are both satisfied in practice with probability one. The two conditions will be highlighted in italics. Denote the SVD of $\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1/2}$ by

$$\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1/2} = \mathbf{U}\Sigma\mathbf{V}^H \tag{6}$$

where $\mathbf{R}_{xx}^{1/2}$ is chosen here to be conjugate symmetric for convenience (without loss of generality), both $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, and Σ is $m \times n$ "diagonal" matrix of singular values. We will use the standard partition $\mathbf{U}\Sigma\mathbf{V}^H = [\mathbf{U}_1 \;\; \mathbf{U}_2]\,\mathrm{diag}\{\Sigma_1 \;\; \Sigma_2\}[\mathbf{V}_1 \;\; \mathbf{V}_2]^H$ where Σ_1 is $r \times r$ and contains the r largest singular values which are *assumed to be strictly larger than the rest*. Note that $\mathrm{trun}_r\{\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1/2}\} = \mathbf{U}_1\Sigma_1\mathbf{V}_1^H$. Then, we know that there exists some $\mathbf{T}(0)$ such that the randomly (hence, "globally") selected $\mathbf{B}(0)$ can be written as

$$\mathbf{B}(0) = \mathbf{R}_{xx}^{-1/2}\mathbf{V}\mathbf{T}(0) \tag{7}$$

where the top $r \times r$ submatrix of $\mathbf{T}(0)$ is *assumed to have full rank*. To observe the result from the first iteration of the IQMD algorithm, we first consider the following:

$$\mathbf{R}_{yx}\mathbf{B}(0) = \mathbf{R}_{yx}\mathbf{R}_{xx}^{-1/2}\mathbf{V}\mathbf{T}(0) = \mathbf{U}\Sigma\mathbf{T}(0) \tag{8}$$
$$\mathbf{B}^H(0)\mathbf{R}_{xx}\mathbf{B}(0) = T^H(0)\mathbf{T}(0). \tag{9}$$

Using (8)–(9) in (4) leads to

$$\mathbf{A}(1) = \mathbf{U}\Sigma\mathbf{T}(0)(\mathbf{T}^H(0)\mathbf{T}(0))^{-1}. \tag{10}$$

Then, we obtain

$$(\mathbf{A}^H(1)\mathbf{A}(1))^{-1} = (\mathbf{T}^H(0)\mathbf{T}(0))(\mathbf{T}(0)\Sigma^2\mathbf{T}(0))^{-1} \times (\mathbf{T}^H(0)\mathbf{T}(0)) \tag{11}$$

$$\mathbf{A}^H(1)\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1} = (\mathbf{T}^H(0)\mathbf{T}(0))^{-1}\mathbf{T}^H(0)\Sigma^2\mathbf{V}^H\mathbf{R}_{xx}^{-1/2} \tag{12}$$

where $\Sigma^2 = \Sigma^H\Sigma$. Using (11)–(12) in (5) leads to

$$\mathbf{B}^H(1) = \mathbf{T}^H(1)\mathbf{V}^H\mathbf{R}_{xx}^{-1/2} \tag{13}$$
$$\mathbf{T}^H(1) = (\mathbf{T}^H(0)\mathbf{T}(0))(\mathbf{T}^H(0)\Sigma^2\mathbf{T}(0))^{-1}\mathbf{T}^H(0)\Sigma^2. \tag{14}$$

Note that given the assumption on $\mathbf{T}(0)$, the top $r \times r$ submatrix of $\mathbf{T}(1)$ is also full rank. Combining (10) and (13) yields

$$\mathbf{A}(1)\mathbf{B}^H(1) = \mathbf{U}\Sigma\mathbf{T}(0)(\mathbf{T}^H(0)\Sigma^2\mathbf{T}(0))^{-1} \times \mathbf{T}^H(0)\Sigma^2\mathbf{V}^H\mathbf{R}_{xx}^{-1/2}. \tag{15}$$

Note that (13)–(15) are results from the first iteration of the IQMD algorithm. More generally, (13)–(15) imply the following:

$$\mathbf{B}^H(k) = \mathbf{T}^H(k)\mathbf{V}^H\mathbf{R}_{xx}^{-1/2} \tag{16}$$
$$\mathbf{A}(k+1)\mathbf{B}^H(k+1) = \mathbf{U}\Sigma\mathbf{T}(k)(\mathbf{T}^H(k)\Sigma^2\mathbf{T}(k))^{-1} \times \mathbf{T}^H(k)\Sigma^2\mathbf{V}^H\mathbf{R}_{xx}^{-1/2} \tag{17}$$
$$\mathbf{T}(k) = \Sigma^2\mathbf{T}(k-1)(\mathbf{T}^H(k-1)\Sigma^2 \times \mathbf{T}(k-1))^{-1}\mathbf{T}^H(k-1)\mathbf{T}(k-1). \tag{18}$$

By iterating (18) k times from the initial $\mathbf{T}(0)$, one can verify (with some work) that

$$\mathbf{T}(k) = \mathbf{H}(k)\mathbf{H}(k-1)\cdots\mathbf{H}(1)\mathbf{T}(0) \tag{19}$$
$$\mathbf{H}(k) = \Sigma^{2k}\mathbf{T}(0)(\mathbf{T}^H(0)\Sigma^{4k-2}\mathbf{T}(0))^{-1}\mathbf{T}^H(0)\Sigma^{2k-2}. \tag{20}$$

What we need to show next is that the norm of $\mathbf{B}(k)$ for all k is both lower and upper bounded, i.e., $0 < \|\mathbf{B}(k)\|_2 < \infty$ for all k, and furthermore $\mathbf{A}(k+1)\mathbf{B}^H(k+1)$ approaches T_{RRWF} exponentially as k increases. Given (16), however, the boundedness of $\mathbf{B}(k)$ is equivalent to that of $\mathbf{T}(k)$. From (18), we have

$$\mathbf{T}^H(k-1)\mathbf{T}(k) = \mathbf{T}^H(k-1)\mathbf{T}(k-1). \tag{21}$$

Since

$$\|\mathbf{T}^H(k-1)\mathbf{T}(k)\|_2 \le \|\mathbf{T}(k-1)\|_2\|\mathbf{T}(k)\|_2 \quad \text{and}$$
$$\|\mathbf{T}^H(k-1)\mathbf{T}(k-1)\|_2 = \|\mathbf{T}(k-1)\|_2^2,$$

(21) implies $\|\mathbf{T}(k)\|_2 \ge \|\mathbf{T}(k-1)\|_2$ and hence

$$\|\mathbf{T}(\infty)\|_2 \ge \|\mathbf{T}(k)\|_2 \ge \|\mathbf{T}(0)\|_2. \tag{22}$$

To show that the upper bound $\|\mathbf{T}(\infty)\|_2$ is finite, we first let $\mathbf{T}(k) = \left[\begin{smallmatrix}\mathbf{T}_1(k)\\ \mathbf{T}_2(k)\end{smallmatrix}\right]$ and $\Sigma^2 = \mathrm{diag}\{\Sigma_1^2 \;\; \Sigma_2^2\}$ where $\mathbf{T}_1(k)$ and

Σ_1^2 are both of $r \times r$. Then, (20) becomes

$$\mathbf{H}(k) = \begin{bmatrix} \Sigma_1^{2k}\mathbf{T}_1(0) \\ \Sigma_2^{2k}\mathbf{T}_2(0) \end{bmatrix} \big(\mathbf{T}_1^H(0)\Sigma_1^{4k-2}\mathbf{T}_1(0) + \mathbf{T}_2^H(0)\Sigma_2^{4k-2}\mathbf{T}_2(0)\big)^{-1} \begin{bmatrix} \mathbf{T}_1^H(0)\Sigma_1^{2k-2} \\ \mathbf{T}_2^H(0)\Sigma_2^{2k-2} \end{bmatrix}$$

Applying the matrix identity $(\mathbf{C}+\mathbf{D})^{-1} = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{D}(\mathbf{I}+\mathbf{C}^{-1}\mathbf{D})^{-1}\mathbf{C}^{-1}$ to the inverse factor in the above, one can verify (with some work) that for large k

$$\mathbf{H}(k) = \mathrm{diag}(1 \quad \cdots \quad 1 \quad 0 \quad \cdots \quad 0) + O\{\varepsilon(k)\} \tag{23}$$

where the diagonal matrix has r ones; $\varepsilon(k) = (\sigma_{r+1}/\sigma_r)^{2k}$; σ_r and σ_{r+1} are the rth and $(r+1)$th largest singular values of $\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1/2}$. It follows from (23) that for some positive constant c, $||\mathbf{H}(k)||_2 \leq 1 + c\varepsilon(k)$. From (19), we know that

$$\begin{aligned} ||\mathbf{T}(\infty)||_2 &\leq ||\mathbf{H}(\infty)||_2 \cdots ||\mathbf{H}(2)||_2 ||\mathbf{H}(1)||_2 ||\mathbf{T}(0)||_2 \\ &\leq ||\mathbf{T}(0)||_2 \prod_{i=1}^{\infty}(1 + c\varepsilon(i)) \\ &= ||\mathbf{T}(0)||_2 \exp\left\{\sum_{i=1}^{\infty} \log\{1 + c\varepsilon(i)\}\right\} \\ &\leq ||\mathbf{T}(0)||_2 \exp\left\{\sum_{i=1}^{\infty} c\varepsilon(i)\right\} \\ &= ||\mathbf{T}(0)||_2 \exp\left\{\frac{c}{1 - (\sigma_{r+1}/\sigma_r)^2}\right\}. \end{aligned}$$

The norm of $\mathbf{T}(k)$ is now proven to be both lower and upper bounded. We now further consider (18), which leads to

$$\begin{aligned} \mathbf{T}_1(k) &= \Sigma_1^2\mathbf{T}_1(k-1)(\mathbf{T}^H(k-1)\Sigma^2\mathbf{T}(k-1))^{-1} \\ &\quad \times \mathbf{T}^H(k-1)\mathbf{T}(k-1) \\ \mathbf{T}_2(k) &= \Sigma_2^2\mathbf{T}_2(k-1)(\mathbf{T}^H(k-1)\Sigma^2\mathbf{T}(k-1))^{-1} \\ &\quad \times \mathbf{T}^H(k-1)\mathbf{T}(k-1) \end{aligned}$$

Then, it follows that $\mathbf{T}_2(k)\mathbf{T}_1^{-1}(k) = \Sigma_2^2\mathbf{T}_2(k-1)\mathbf{T}_1^{-1}(k-1)\Sigma_1^{-2}$ and hence

$$\mathbf{T}_2(k)\mathbf{T}_1^{-1}(k) = \Sigma_2^{2k}\mathbf{T}_2(0)\mathbf{T}_1^{-1}(0)\Sigma_1^{-2k}. \tag{24}$$

From (24), we know that for large k, each element of $\mathbf{T}_2(k)\mathbf{T}_1^{-1}(k)$ is no larger than $\varepsilon(k)$ multiplied by a constant, and hence converges to zero exponentially. Since the norm of $\mathbf{T}(k)$ is upper bounded, (24) also implies that $\mathbf{T}_2(k)$ converges to zero exponentially as k increases. With this property, (17) implies that for large k

$$\begin{aligned} \mathbf{A}(k+1)\mathbf{B}^H(k+1) &= \mathbf{U}_1\Sigma_1\mathbf{T}_1(k)\big(\mathbf{T}_1^H(k)\Sigma_1^2\mathbf{T}_1(k)\big)^{-1} \\ &\quad \times \mathbf{T}_1^H(k)\Sigma_1^2\mathbf{V}_1^H\mathbf{R}_{xx}^{-1/2} + O(\varepsilon(k)) \\ &= \mathbf{U}_1\Sigma_1\mathbf{V}_1^H\mathbf{R}_{xx}^{-1/2} + O(\varepsilon(k)) \\ &= \mathbf{T}_{\mathrm{RRWF}} + O(\varepsilon(k)). \end{aligned}$$

The proof is now completed.

IV. Final Remarks

The IQMD algorithm presented in this letter has also been verified by all our simulation examples (where n varies in tens and hundreds and r is about one tenth of n). With typically five to ten iterations (depending on accuracy measures), it rapidly converged to the RRWF. The IQMD algorithm can be easily adopted for adaptive computation of the RRWF by letting the correlation matrices be updated with new data at each iteration (or after a few iterations). This adaptive IQMD algorithm is also globally convergent if the updated correlation matrices converge to some constant matrices. It is also important to note that the insight shown in this letter is instrumental to many other possible further developments. One such development currently under way is to consider the computation of the reduced rank maximum likelihood estimator of a multivariate regression system [6] and the computation of the canonical coordinates elaborated in [7].

Acknowledgment

The first author thanks Dr. C. Zhang for a brief but useful discussion on the convergence proof.

References

[1] L. L. Scharf, *Statistical Signal Processing—Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991, p. 330.

[2] L. L. Scharf, "The SVD and reduced rank signal processing," *Signal Process.*, vol. 25, pp. 113–133, 1991.

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

[4] Y. Hua and W. Liu, "Generalized Karhunen-Loeve transform," *IEEE Signal Processing Lett.*, vol. 5, pp. 141–142, June 1998.

[5] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1983.

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

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