ALGEBRAIC EQUIVALENCE OF CONJUGATE DIRECTION AND MULTISTAGE WIENER FILTERS

Louis L. Scharf, L. Todd McWhorter, Edwin K. P. Chong, J. Scott Goldstein, and Michael D. Zoltowski

ABSTRACT

We consider* iterative subspace Wiener filters *for solving minimum mean-squared error and minimum variance unbiased estimation problems in low-dimensional subspaces. In this class of subspace filters, the* conjugate direction *and* multistage *Wiener filters comprise two large subclasses, and within each of these the conjugate gradient and orthogonal multistage Wiener filters are the most prominent. We establish very general equivalences between conjugate direction and multistage Wiener filters, wherein the direction vectors of a conjugate direction filter and the stagewise vectors of a multistage filter are related through a one-term autoregressive recursion. By virtue of this recursion, the expanding subspaces of the two filters are identical, even though their bases for them are different. As a consequence, their subspace Wiener filters, gradients, and mean-squared errors are identical at each stage of the subspace iteration. If the conjugate direction filter is a conjugate gradient filter, then the equivalent stagewise filter is an orthogonal multistage filter, and vice-versa. If either the conjugate gradient filter or the orthogonal multistage filter is initialized at the cross-covariance vector between the signal and the measurement, then each of the subspace filters iteratively turns out a basis for a Krylov subspace.

This work was supported by the Office of Naval Research under contract N00014-01-1-1019-P0001 and by the National Science Foundation under grants CCR-0112573 and 0098089-ECS

L. L. Scharf is with the Departments of Electrical and Computer Engineering and Statistics, Colorado State University, Ft. Collins, CO 80523, e-mail: scharf@engr.colostate.edu

L. T. McWhorter is with the Mission Research Corporation, 3665 JFK Parkway, Bldg 1, Suite 206, Ft. Collins, CO 80525, e-mail: mcwhorter@aster.com

E. K. P. Chong is with the Departments of Electrical and Computer Engineering and Mathematics, Colorado State University, Ft. Collins, CO 80523, e-mail: echong@engr.colostate.edu

J. S. Goldstein is with SAIC International, 4501 Daly Drive, Suite 400, Chantilly, VA, 20151, e-mail: jay.s.goldstein@saic.com

M. D. Zoltowski is with the School of Electrical and Computer Engineering, Purdue University, W. Lafayette, IN, e-mail: mikedz@ecn.purdue.edu

1. INTRODUCTION

Rank-reduction is now an important element of many signal processing algorithms that aim to reduce the dimension of a measurement space in advance of detection, estimation, classification, beamforming, or spectrum analysis. The idea, in the *analysis* stage, is to linearly resolve a measurement onto a sequence of expanding subspaces, until a satisfactory trade-off has been achieved between the increased bias and reduced variance associated with rank reduction. If the linearly resolved measurements have a simple correlation structure, then the implementation of the reduced dimension filter in the *synthesis* stage will be efficient.

This paper is addressed to *iterative subspace Wiener filters* for solving minimum mean-squared error and minimum variance unbiased estimation problems in low-dimensional subspaces. In this class of subspace filters, the *conjugate direction* and *multistage* Wiener filters comprise two large subclasses, and within each of these the conjugate gradient and orthogonal multistage Wiener filters are the most prominent.

The conjugate gradient algorithm for iteratively solving quadratic minimization problems was discovered by Hestenes and Stiefel in 1952 [1], and the multistage Wiener filter was discovered by Goldstein in 1998 [2]. In spite of the fact that there is no essential difference between quadratic minimization, the solving of linear systems, and Wiener filtering, the connection between conjugate gradient algorithms and multistage filtering was not made until Weippert and Goldstein recognized in 2001 a connection between conjugate gradients and orthogonal multistage Wiener filters [3, 4]. This insight was based on work by Honig, Xiao, Yohan, Sun, Zoltowski, and Goldstein, subsequently published in [5, 6, 7, 8], regarding the Krylov subspace underlying the multistage Wiener filter. Further connections were reported by Zoltowski in his 2002 tutorial [9].

Our first aim in this paper is to further refine what is currently known about connections between conjugate gradient and multistage Wiener filters, by appealing to fundamental insights in least squares filtering. In so doing we further clarify the sense in which conjugate gradient and multistage filters are compelling, rather than ad hoc, realizations of iterative filters.

Our second aim is to enlarge our understanding of iterative subspace filtering by showing that there is an even larger class of equivalences, of which the equivalence between conjugate gradients and orthogonal multistage Wiener filters is only a special case. That is, for every conjugate *direction* filter there is an infinite family of equivalent multistage filters, and conversely, for every multistage filter there is an infinite family of equivalent conjugate direction filters. The fact that a multistage Wiener filter has a conjugate direction implementation, or vice-versa, is just a consequence of the autoregressive connection between stagewise filters and conjugate directions. If the stagewise filters are orthogonal, and they are initialized at the cross-covariance vector between the signal and the measurement, then they are an orthogonal basis for a Krylov subspace. The equivalent conjugate direction filter is then a conjugate gradient filter whose direction vectors are a nonorthogonal (but conjugate with respect to the covariance matrix) basis for the same subspace. Conversely, if the conjugate direction filter is a conjugate gradient filter, then its equivalent multistage filter is orthogonal. Both turn out a basis for the Krylov subspace, assuming the conjugate gradient filter is initialized at the cross-covariance.

It is worth emphasizing that our arguments are based entirely on the algebraic structure of iterative subspace Wiener filters, with no appeal to specific algorithms for computing conjugate direction or multistage vectors.

2. FILTERING AND QUADRATIC MINIMIZATION

The problem of minimum mean-squared error estimation is this. The complex scalar $x \in \mathbb{C}$ is a random variable to be estimated from the complex vector of measurements, $y \in \mathbb{C}^n$. We shall assume that the signal x and measurement y share the composite covariance matrix

$$R_{zz} = \begin{bmatrix} r_{xx} & r^* \\ r & R \end{bmatrix},$$

where $r_{xx} = E[xx^*] \in R^+$ is the scalar variance of x, $r = E[yx^*] \in \mathbb{C}^{n\times 1}$ is the vector covariance of (x, y), and $R = E[yy^*] \in \mathbb{C}^{n\times n}$ is the matrix covariance of y. Throughout, we shall use superscript $*$ to denote Hermitian transpose.

From this second-order description we may write the mean-squared error of estimating x from w^*y as

$$\begin{aligned} 2Q(w) &= E[|x - w^*y|^2] \\ &= r_{xx} - r^*w - w^*r + w^*Rw \\ &= r_{xx} - r^*R^{-1}r \\ &\quad + [w^* - r^*R^{-1}]R[w - R^{-1}r], \end{aligned}$$

with minimum value $r_{xx} - r^*R^{-1}r$ at the Wiener solution $Rw = r$. This is a consequence of the principle of orthogonality, which says $E[y(x - w^*y)^*] = r - Rw = 0$. This is the way a filtering theorist describes the filtering experiment. An optimization theorist would say the problem is to minimize the quadratic function $Q(w)$, whose gradient is $\gamma(w) = \nabla Q(w) = Rw - r$, and whose Hessian is $\nabla^2 Q(w) = R > 0$, meaning the minimum is achieved where the gradient is 0, namely $\gamma(w) = Rw - r = 0$. So a zero gradient plays the same role in optimization theory as the principle of orthogonality plays in filtering theory.

3. ITERATIVE SUBSPACE WIENER FILTERS

The framework for our discussion of *iterative subspace Wiener filters* is this. The complex scalar x is a random variable to be estimated from the complex vector of measurements, $y \in \mathbb{C}^n$. These measurements are a basis for the subspace $\langle y\rangle = \langle y_1, y_2, \ldots, y_n\rangle$, and any linear estimate of x will lie in this subspace. Let us suppose that we approach the filtering or optimization problem by first resolving the measurements onto a k-dimensional basis $A_k = [a_1, a_2, \ldots, a_k]$ to produce the coordinates $u = A_k^*y = [u_1, u_2, \ldots, u_k]^T$. The trick is to determine an algorithm for expanding this subspace in such a way that $\langle u\rangle$ is near to $\langle y\rangle$ and the covariance matrix of $u \in \mathbb{C}^k$ is easy to invert.

The variables (x, u) share the composite covariance matrix

$$R_{ww} = \begin{bmatrix} r_{xx} & r^*A_k \\ A_k^*r & A_k^*RA_k \end{bmatrix}$$

From this second-order description of the composite vector we may write the mean-squared error of estimating x from v^*u as

$$\begin{aligned} 2Q(v) &= E[|x - v^*u|^2] \\ &= r_{xx} - r^*A_kv - v^*A_k^*r + v^*A_k^*RA_kv \\ &= r_{xx} - r^*A_k(A_k^*RA_k)^{-1}A_k^*r \\ &\quad + [v^* - r^*A_k(A_k^*RA_k)^{-1}]A_k^*RA_k \\ &\quad\quad [v - (A_k^*RA_k)^{-1}A_k^*r], \end{aligned}$$

with minimum value $Q_k = r_{xx} - r^*A_k(A_k^*RA_k)^{-1}A_k^*r$ at the Wiener solution $(A_k^*RA_k)v_k = A_k^*r$. This is a consequence of the principle of orthogonality, which says $E[u(x - v_k^*u)^*] = A_k^*r - A_k^*RA_kv_k = 0$. But A_kv_k is just w_k, meaning this orthogonality condition may also be written as

$$A_k^*(r - Rw_k) = -A_k^*\gamma_{k+1} = 0,$$

where $\gamma_{k+1} = \gamma(w_k) = Rw_k - r = RA_kv_k - r$ is the gradient of $Q(w_k)$. This orthogonality may be iterated to derive the important result that

$$A_k^*\Gamma_k = C_k,$$

where $\Gamma_k = [\gamma_1, \gamma_2, \ldots, \gamma_k]$ is the matrix of gradients and C_k is a lower triangular matrix.

The importance of this formula is the following: no matter how the coordinate system A_k is chosen, its basis will be orthogonal to the current gradient $\gamma_{k+1} = \gamma(w_k)$ at $w_k = A_k v_k$, and all future gradients. Thus, as long as the current gradient is non-zero, the subspace $\langle [A_k, \gamma_{k+1}] \rangle$ will continue to expand in dimension, as a direct sum of the subspaces $\langle A_k \rangle$ and $\langle \gamma_{k+1} \rangle$. Thus, there is a clear suggestion that as new coordinates of the form $A_{k+1} = [A_k, a_{k+1}]$ are built, the old coordinates A_k and the gradient γ_{k+1} are compelling candidates in this expansion. Moreover, if A_k is a basis for the Krylov subspace $\langle K_k \rangle$, where $K_k = [r, Rr, \ldots, R^{k-1}r]$ is the Krylov matrix of Krylov vectors, then $\gamma_{k+1} = RA_k v_k - r$ is a linear combination of the Krylov vectors $[K_k, R^k r]$ and $[A_k, \gamma_{k+1}]$ is a basis for the Krylov subspace $\langle K_{k+1} \rangle$. By choosing $v_0 = 0$, in which case $w_0 = 0$, $\gamma_1 = r$, and $a_1 = \gamma_1$, then $A_1 = r$ and the induction argument is initialized. Moreover, the basis Γ_k will be an *orthogonal* basis for the Krylov subspace $\langle K_k \rangle$. What makes this argument go is the fact that $w_k = A_k v_k$, the suboptimum Wiener filter in the original coordinate system, is computed from the optimum Wiener filter v_k in the coordinate system A_k.

Thus, every subspace Wiener filter that is initialized at the one-dimensional subspace $\langle r \rangle$ and expanded using a linear combination of the current subspace and current gradient, will turn out an orthogonal basis for the Krylov subspace. This conclusion has nothing whatsoever to do with the actual algorithm used. Thus, this Krylov property is not peculiar to any particular algorithm for designing subspace Wiener filters.

To this point we have required nothing special about the coordinate system A_k, and its corresponding filters $\{a_i\}_1^k$. Consequently, nothing special can be said about the Wiener filters $w_k = A_k v_k = A_k (A_k^* R A_k)^{-1} A_k^* r$, the gradients $\gamma_k = R w_k - r$, or the mean-squared errors $Q_k = r_{xx} - r^* A_k (A_k^* R A_k)^{-1} A_k^* r$. Moreover, there is, as yet, no indication of how these may be computed iteratively. It is worth emphasizing, however, that any two iterative subspace Wiener filters whose analysis filters span the same subspace are identical, meaning their gradients, and mean-squared errors are identical.

4. ALGEBRAIC EQUIVALENCES

When we talk about a conjugate direction filter, then we say the analysis stage consists of the *direction vectors* $A_k = D_k = [d_1, d_2, \ldots, d_k]$ where $d_i \in \mathbb{C}^n$. A conjugate direction Wiener filter is an iterative subspace Wiener filter in which the coordinates of the original measurement in the direction subspace are *uncorrelated*. That is, $E[uu^*] = E[D_k^* y y^* D_k] = D_k^* R D_k = \Sigma_k^2$, and we say the directions are R-conjugate. The resulting iterative Wiener filter may then be synthesized in a forward synthesis, and error formulas are continued sums, a fact that is obvious, but not derived in this paper.

When we talk about a multistage filter we say the analysis stage consists of the *multistage vectors* $A_k = G_k = [g_1, g_2, \ldots, g_k]$, where $g_i \in \mathbb{C}^n$. A multistage Wiener filter is an iterative subspace Wiener filter in which the coordinates of the original measurement in the multistage subspace are *tridiagonally correlated*. That is, $E[uu^*] = E[G_k^* y y^* G_k] = G_k^* R G_k = T_k$, where T_k is tridiagonal.[1] The resulting iterative Wiener filter may then be synthesized in a backward synthesis, and error formulas are continued fractions, a fact that is almost obvious, but not derived in this paper.

Our aim is to show that for every conjugate direction Wiener filter there is a family of equivalent multistage Wiener filters, and vice-versa. Because of the autoregressive recursion that will connect the conjugate direction vectors to the stagewise vectors, the two implementations will share a common expanding subspace. Thus neither can outperform the other. Generally the underlying subspace is not Krylov, so the motivation to design something other than a standard conjugate gradient or orthogonal multistage filter is to construct an expanding subspace that is better matched, for $k < n$, to the optimum Wiener filter than is the Krylov space. If the conjugate direction filter is a conjugate gradient Wiener filter, and its corresponding multistage Wiener filter is orthogonal, then both are stuck in the Krylov subspace.

4.1. Conjugate direction and multistage Wiener filters

The idea behind any conjugate direction algorithm is to force the coordinates u to be uncorrelated. That is,

$$\begin{aligned} D_k^* R D_k &= \Sigma_k^2\text{: diagonal,} \\ D_k^* \Gamma_k &= C_k\text{: lower triangular,} \end{aligned}$$

where Γ_k is the matrix of gradients and C_k is a lower triangular matrix. Now suppose the filters G_k are defined according to a first-order autoregressive recursion $D_k B_k^* = G_k$, for some arbitrary upper bidiagonal B_k^*. These filters define the multistage Wiener filter

$$\begin{aligned} G_k^* R G_k &= B_k \Sigma_k^2 B_k^* = T_k\text{: tridiagonal,} \\ G_k^* \Gamma_k &= B_k C_k\text{: lower triangular.} \end{aligned}$$

[1] The perceptive reader will have noticed that all iterative subspace Wiener filters are "multistage," meaning that the language here is imprecise. Ideally we would prefer to use the term "multistage Wiener filter" for the entire class of iterative filters, and reserve the terms "diagonal direction" and "tridiagonal direction" for what are currently called conjugate direction (R-diagonal) and multistage (R-tridiagonal) Wiener filters. But this would require a rewriting of history.

The arbitrary choice of B_k indicates that there are infinitely many multistage Wiener filters that are equivalent to the given conjugate direction algorithm.

Conversely, suppose we start with a multistage Wiener filter

$$\begin{aligned} G_k^* R G_k &= T_k = B_k \Sigma_k^2 B_k^*\text{: tridiagonal,} \\ G_k^* \Gamma_k &= B_k C_k\text{: lower triangular,} \end{aligned}$$

where $B_k \Sigma_k^2 B_k^* = T_k$ is a bidiagonal–diagonal–upper-bidiagonal factorization of the tridiagonal matrix T_k. Define the filters D_k according to the first-order autoregressive recursion $D_k B_k^* = G_k$. These filters define a conjugate direction Wiener filter

$$\begin{aligned} D_k^* R D_k &= \Sigma_k^2\text{: diagonal,} \\ D_k^* \Gamma_k &= C_k\text{: lower triangular.} \end{aligned}$$

This conjugate direction Wiener filter is in turn equivalent to an infinite family of filters related through scaling of each multistage vector.

In summary, for every conjugate direction Wiener filter there is a family of equivalent multistage Wiener filter, and vice-versa. They share common filters, gradients, and mean-squared errors, but nothing special may be said about these gradients or subspaces, except that they are common.

4.2. Conjugate gradient and orthogonal multistage Wiener filters

If the direction vectors D_k are computed according to the conjugate gradient algorithm[2] $D_k B_k^* = \Gamma_k$, where B^* is an upper bidiagonal matrix, then the gradients are orthogonal and they tridiagonalize R:

$$\begin{aligned} \Gamma_k^* R \Gamma_k &= B_k \Sigma_k^2 B_k^* = T_k\text{: tridiagonal,} \\ \Gamma_k^* \Gamma_k &= B_k D_k^* \Gamma_k\text{: diagonal.} \end{aligned}$$

The orthogonality of the gradient matrix Γ_k follows from the fact that B_k and $D_k^* \Gamma_k$ are lower triangular, as is their product, and the only Hermitian lower triangular matrix is diagonal. We may follow the argument of [11] to show that Γ_k is an orthogonal basis for the Krylov subspace $\langle K_k \rangle$.

In summary, any conjugate gradient Wiener filter defines an equivalent orthogonal multistage Wiener filter whose filters are $G_k = \Gamma_k$, whose gradients are Γ_k, and whose mean-squared error is the same as for the conjugate gradient Wiener filter. Both filters turn out a basis for the Krylov subspace K_k, provided the conjugate gradient Wiener filter is initialized at the cross-covariance r.

[2] We call a conjugate direction algorithm a *conjugate gradient* algorithm if the current direction vector is a linear combination of the current gradient and the previous direction vector. This slightly generalizes the standard definition (e.g., [10]).

Conversely, suppose we start with an orthogonal multistage Wiener filter

$$\begin{aligned} G_k^* R G_k &= T_k = B_k \Sigma_k^2 B_k^*\text{: tridiagonal,} \\ & G_k^* G_k\text{: diagonal,} \\ G_k^* \Gamma_k &= B_k C_k\text{: lower triangular.} \end{aligned}$$

and define the direction vectors $D_k B_k^* = G_k$. Then these directions define a conjugate direction algorithm:

$$\begin{aligned} D_k^* R D_k &= \Sigma_k^2\text{: diagonal,} \\ D_k^* \Gamma_k &= C_k\text{: lower triangular.} \end{aligned}$$

Both G_k and Γ_k are a causal basis for $\langle K_k \rangle$, meaning that $G_k = \Gamma_k U_k$, where U_k is upper triangular. Thus,

$$\begin{aligned} G_k^* G_k\text{: diagonal} &= G_k^* \Gamma_k U_k \\ &= B_k C_k U_k\text{: lower-times-upper.} \end{aligned}$$

We deduce from this that U_k and $B_k C_k$ are diagonal. This means $G_k = \Gamma_k \Lambda_k$, where Λ_k is diagonal. Thus the conjugate direction algorithm is a conjugate gradient algorithm.

In summary, any orthogonal multistage Wiener filter determines an equivalent conjugate gradient Wiener filter whose filters, gradients, and mean-squared errors are the same as for the orthogonal Wiener filter. The gradients are within a diagonal matrix of the orthogonal multistage vectors (i.e., each multistage vector is a scaled gradient). Both filters turn out a basis for the Krylov subspace K_k, provided the orthogonal multistage Wiener filter is initialized at the cross-covariance r.

5. CONCLUSIONS

Our conclusions are the following:

- For any iterative subspace Wiener filter, the subspace A_k of dimension k is orthogonal to the current and future gradients $[\gamma_{k+1}, \gamma_{k+2}, \ldots, \gamma_n]$. This is the orthogonality principle at work.
- Any subspace expansion in an iterative subspace Wiener filter that uses the direct sum of the past subspace and the current gradient turns out an orthogonal basis for the Krylov subspace, assuming the algorithm is initialized at the cross-covariance vector.
- For every multistage Wiener filter there is a family of equivalent conjugate direction Wiener filters, with identical subspaces, gradients, and mean-squared errors. And vice-versa.
- For every orthogonal multistage Wiener filter there is a family of equivalent conjugate gradient Wiener filters, with identical subspaces, gradients, and mean-squared errors. And vice-versa.

- If an orthogonal multistage Wiener is initialized at the cross-covariance, then it and its equivalent conjugate gradient Wiener filter turn out the Krylov subspace. And vice-versa.

All of these conclusions extend to matrix iterative subspace Wiener filters for estimating vectors $x \in \mathbb{C}^{m\times 1}$. They should extend to iterative submanifold Wiener filters for estimating with respect to norms on manifolds, and they should extend to (true) iterative subspace Wiener filters for estimating time- or space-series, rather than scalars or vectors.

6. REFERENCES

[1] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Res. Nat. Bur. Stand.*, vol. 49, pp. 409–436, 1952.

[2] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inform. Th.*, vol. 44, no. 7, pp. 2943–2959, Nov. 1998.

[3] M. E. Weippert and J. S. Goldstein, "Relationship between the multistage Wiener filter and the method of conjugate gradients," Internal SAIC Memorandum, August 31, 2001 (Patent pending).

[4] M. E. Weippert, J. D. Hiemstra, J. S. Goldstein, and M. D. Zoltowski, "Insights from the relationship between the multistage Wiener filter and the method of conjugate gradients," *Proc. IEEE Workshop on Sensor Array and Multichannel Signal Processing*, 4–6 August, 2002.

[5] M. D. Zoltowski and J. S. Goldstein, "Reduced-rank adaptive filtering and applications," Tutorial, *ICASSP 2001*, Salt Lake City, May 2001.

[6] M. L. Honig and W. Xiao, "Performance of reduced-rank linear interference suppression," *IEEE Trans. Inform. Th.*, vol. 47, pp. 1928–1946, July 2001.

[7] M. Joham, Y. Sun, M. D. Zoltowski, M. L. Honig and J. S. Goldstein, "A new backward recursion for the multistage nested Wiener filter employing Krylov subspace methods," *Proc. IEEE MILCOM*, October 2001.

[8] G. Dietl, M. D. Zoltowski, and M. Joham, "Recursive reduced-rank adaptive equalization for wireless communications," *Proc. SPIE*, vol. 4395, April 2001.

[9] M. D. Zoltowski, "Conjugate gradient adaptive filtering with application to space-time processing for wireless digital communications," Plenary Lecture Viewgraphs, published in *Proc. IEEE Workshop on Sensor Array and Multichannel Signal Processing*, 4–6 August, 2002.

[10] E. K. P. Chong and S. H. Żak, *An Introduction to Optimization, Second Edition*. John Wiley and Sons, Inc., New York, NY, 2001.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, 1983, p. 324.