Extension of the Signal Subspace Speech Enhancement Approach to Colored Noise

Hanoch Lev-Ari, *Senior Member, IEEE*, and Yariv Ephraim, *Fellow, IEEE*

***Abstract*—The signal subspace approach for speech enhancement is extended to colored-noise processes. Explicit forms for the linear time-domain- and spectral-domain-constrained estimators are presented. These estimators minimize the average signal distortion power for given constraints on the residual noise power in the time and spectral domains, respectively. Equivalent implementations of the two estimators using the whitening approach are described.**

***Index Terms*—Colored noise, signal subspace, speech enhancement.**

I. Introduction

A SIGNAL subspace approach for enhancing noisy speech signals was developed in [2]. Linear estimation of each vector of the signal was proposed using two possible optimality criteria. The first estimator was designed by minimizing the average signal distortion power for a given upper bound on the average power of the residual noise. The second estimator was also designed by minimizing the average signal distortion power, but now the average power of each spectral component of the residual noise was constrained not to exceed a certain threshold. The two estimators were referred to as the *time-domain-constrained* estimator and the *spectral-domain-constrained* estimator, respectively. For white noise, the two estimators were implemented by first applying the Karhunen–Loève transform (KLT) to the noisy signal. Then, KLT components corresponding to the signal subspace were modified by one of the two estimators, and the remaining KLT components corresponding to the noise subspace were nulled. The enhanced signal was obtained from inverse KLT of the modified components.

In [2], the noise was assumed to be white. For colored noise, it was suggested that the noise could be whitened, but no further elaboration on how this might be done was provided. Mittal and Phamdo [9] argued that the whitening approach is not desirable for the spectral-domain-constrained estimator, since then the constraints are imposed on the whitened signal rather than on the original signal. They proposed to classify vectors of the noisy signal as being dominated by either the clean speech signal or the noise process, and to apply the KLT corresponding to the dominating process in each frame. Rezayee and Gazor [10] proposed to implement the linear time-domain-constrained estimator using a diagonal rather than an identity matrix for the colored-noise power spectrum.

Hu and Loizou [6] proposed joint diagonalization of the covariance matrices of the clean signal and noise process. The linear time-domain-constrained estimator was implemented using the common diagonalizing matrix. The resulting estimator resembles the signal subspace estimator [2, eqs. (28) and (29)] with two main differences. First, the orthogonal KLT used when the noise is white is now replaced by a nonsingular nonorthogonal transformation. Second, the eigenvalues of the clean signal are essentially replaced by the eigenvalues of the whitened signal. No attempt was made in [6] to relate the nonsingular transformation to the KLT of the whitened signal. In [6], the linear matrix equation of the form $AX + XB = C$ that the spectral-domain-constrained estimator must satisfy was transformed using the common diagonalizing nonsingular matrix, but no explicit solution to this equation was presented. The equation was solved numerically using the iterative algorithm of Bartels and Stewart [1]. This algorithm is applicable to any real matrices A, B, and C of suitable dimensions and, hence, aims at a more general problem than we have here.

In this letter, we generalize the original spectral-domain-constrained estimation approach, as discussed in [2], [6], and [9], by allowing an arbitrary number of spectral constraints, any orthogonal spectral transformation of the residual noise, and a possibly unitary transformation such as that related to the discrete Fourier transform (DFT). We develop general explicit solutions to the linear time-domain- and spectral-domain-constrained estimators, and we show their equivalent implementations using the whitening approach.

II. Main Results

Let Y and W be k-dimensional random vectors in a Euclidean space $\mathcal{R}^k$ representing the clean signal and noise, respectively. Assume that the expected value of each vector is zero in an appropriately defined probability space. Let $Z = Y + W$ denote the noisy vector. Let R_y and R_w denote the covariance matrices of the clean signal and noise process, respectively. Assume that R_w is positive definite. Let H denote a $k \times k$ real matrix in the linear space $\mathcal{R}^{k \times k}$, and let $\hat{Y} = HZ$ denote the linear estimator of Y given Z. The residual signal in this estimation is given by

$$Y - \hat{Y} = (I - H)Y - HW \tag{1}$$

Manuscript received June 27, 2002; revised August 30, 2002. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Darren B. Ward.

H. Lev-Ari is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: levari@ece.neu.edu).

Y. Ephraim is with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA 22030 USA (e-mail: yephraim@gmu.edu).

Digital Object Identifier 10.1109/LSP.2003.808544

where I denotes, as usual, the identity matrix. To simplify notation, we shall not explicitly indicate the dimensions of the identity matrix. These dimensions should be clear from the context. In (1), $D = (I - H)Y$ is the signal distortion, and $N = HW$ is the residual noise in the linear estimation. Let $(\cdot)'$ denote the transpose of a real matrix or the conjugate transpose of a complex matrix. Let

$$\overline{\epsilon_d^2} = \frac{1}{k}\mathrm{tr}E\{DD'\} = \frac{1}{k}\mathrm{tr}\{(I - H)R_y(I - H)'\} \tag{2}$$

denote the average signal distortion power. Similarly, let

$$\overline{\epsilon_n^2} = \frac{1}{k}\mathrm{tr}E\{NN'\} = \frac{1}{k}\mathrm{tr}\{HR_wH'\} \tag{3}$$

denote the average residual noise power.

The matrix H for the linear time-domain-constrained estimator is obtained from

$$\min_H \overline{\epsilon_d^2} \quad \text{subject to: } \overline{\epsilon_n^2} \le \alpha \tag{4}$$

for some given α. Let $\mu \ge 0$ denote the Lagrange multiplier of the inequality constraint. The optimal matrix, say $H = H_1$, is given by [2], [6], [10]

$$H_1 = R_y(R_y + \mu R_w)^{-1}. \tag{5}$$

The matrix H_1 can be implemented as follows (e.g., see [3, eq. (15)]). Let $R_w^{1/2}$ denote the positive definite square root of R_w, and let $R_w^{-1/2} = (R_w^{1/2})^{-1}$ [7, p. 181]. Let U denote an orthogonal matrix of eigenvectors of the symmetric matrix $R_w^{-1/2}R_yR_w^{-1/2}$ [4, p. 308]. Let $\Lambda = \mathrm{diag}[\lambda_1, \ldots, \lambda_k]$ denote the diagonal matrix of nonnegative eigenvalues of $R_w^{-1/2}R_yR_w^{-1/2}$. Then

$$H_1 = R_w^{\frac{1}{2}}U\Lambda(\Lambda + \mu I)^{-1}U'R_w^{-\frac{1}{2}}. \tag{6}$$

When H_1 in (6) is applied to Z, it first whitens the input noise by applying $R_w^{-1/2}$ to Z. Then, the orthogonal transformation U' corresponding to the covariance matrix of the whitened clean signal is applied, and the transformed signal is modified by a diagonal Wiener-type gain matrix. Components of the whitened noisy signal that contain noise only are nulled as advocated by the signal subspace approach. The implementation of (5), developed in [6, eq. (12)], uses eigendecomposition of the nonsymmetric matrix $R_w^{-1}R_y$. This approach resulted in a nonsingular whitening transformation matrix for the noisy signal, but the decomposition of that matrix into $U'R_w^{-1/2}$ was not evident. The Wiener-type gain matrix is the same in both implementations, since the two matrices $R_w^{-1/2}R_yR_w^{-1/2}$ and $R_w^{-1}R_y$ are similar and hence have the same set of eigenvalues.

We turn now to the linear spectral-domain-constrained estimator. For this estimator, a set $\{v_i, i = 1, \ldots, m\}$, $m \le k$, of k-dimensional real or complex orthonormal vectors and a set $\{\alpha_i, i = 1, \ldots, m\}$ of nonnegative constants are chosen. The matrix H is obtained from

$$\min_H \overline{\epsilon_d^2} \quad \text{subject to: } E\left\{|v_i'N|^2\right\} \le \alpha_i, \qquad i = 1, \ldots, m. \tag{7}$$

This estimator minimizes the average signal distortion power for a given set of constraints on the power spectrum of the residual noise defined using $\{v_i\}$. The vectors $\{v_i\}$ are assumed orthonormal in order to preserve the norm of the residual noise vector N when $m = k$. In [2], there were $m = k$ constraints; $\{v_i\}$ were the eigenvectors of R_y; and some of the $\{\alpha_i\}$ corresponding to $\{v_i\}$ in the noise subspace were assumed zero. Another important choice for $\{v_i\}$ is given by the set of orthonormal vectors related to the DFT. In this case, $v_i' = k^{-1/2}(1, e^{-j(2\pi/k)(i-1)\cdot 1}, \ldots, e^{-j(2\pi/k)(i-1)\cdot(k-1)})$, and we must choose $\alpha_i = \alpha_{k-i+2}, i = 2, \ldots, k/2$, assuming k is even, for the residual noise power spectrum to be symmetric. This implies that at most $k/2 + 1$ constraints can be imposed. The DFT-related $\{v_i\}$ enable the use of constraints that are consistent with auditory perception of the residual noise.

Let $\{\mu_1, \ldots, \mu_m\}$ denote the Lagrange multipliers corresponding to the inequality constraints. Set $\mu_i = 0$ for $m < i \le k$. Let $M = \mathrm{diag}[k\mu_1, \ldots, k\mu_k]$. Extend the set $\{v_i, i = 1, \ldots, m\}$ to an orthogonal or unitary matrix $V = [v_1, \ldots, v_k]$. Let $L = VMV'$. By the Kuhn–Tucker necessary conditions [8, p. 314], any relative minimum that is a regular point of the constraints in (7) must satisfy the following gradient equation:

$$LHR_w + HR_y - R_y = 0 \tag{8}$$

with some $\{\mu_l \ge 0, l = 1, \ldots, m\}$. Equation (8) was originally derived in [2] for white noise and in [6] for colored noise. Under certain conditions, which we specify shortly, this equation has a unique solution that provides the desired spectral-domain-constrained estimator. The possible solutions of (8) lie in $\mathcal{R}^{k\times k}$ whenever L is a real matrix. This condition is met when V is real and possibly also when V is complex. For instance, if V is the DFT-related matrix, and $\alpha_i = \alpha_{k-i+2}, i = 2, \ldots, k/2$, then the Lagrange multipliers bear the same symmetry, and L is a (real) circulant matrix [5].

Let $Q = R_w^{-1/2}U$ and note that $Q'R_yQ = \Lambda$ and $Q'R_wQ = I$ [7, p. 185, Th. 2]. The matrix Q' is a nonsingular whitening matrix that appeared earlier in (6) and was also used in [6]. Let $\tilde{H} = Q'H(Q')^{-1}$ and $\tilde{L} = Q'L(Q')^{-1}$. Substituting these relations in (8) we have

$$\tilde{L}\tilde{H} + \tilde{H}\Lambda = \Lambda. \tag{9}$$

Let $\tilde{h}_l$ denote the lth column of $\tilde{H}$. Let e_l denote a unit vector in $\mathcal{R}^k$ for which the lth component is one, and all other components are zero. Now observe from (9) that $\tilde{L}\tilde{h}_l + \tilde{h}_l\lambda_l = \lambda_l e_l$ for $l = 1, \ldots, k$. Using the definitions of $\tilde{L}$ and L, and denoting $T = Q'V$, this equation can be rewritten as

$$T(M + \lambda_l I)T^{-1}\tilde{h}_l = \lambda_l e_l. \tag{10}$$

The latter equation has a unique solution if and only if $M + \lambda_l I$ is nonsingular. Thus, (8) has a unique solution if and only if $k\mu_i \ne -\lambda_l$ for all $i, l = 1, \ldots, k$. This condition is satisfied, regardless of the values of $\{\lambda_l\}$, if $\mu_i > 0$ for each i, since $\{\lambda_l\}$ are eigenvalues of a positive-semidefinite matrix. If R_y is positive definite, then the condition is relaxed to $\mu_i \ge 0$ for each $i = 1, \ldots, k$.

The explicit form of the optimal linear spectral-domain-constrained estimator is now readily available. Suppose that all Lagrange multipliers $\{\mu_i\}$ are nonnegative and that $k\mu_i \neq -\lambda_l$ for all i, l. Then, the optimal estimation matrix, say $H = H_2$, in the sense of (7), is given by

$$H_2 = (Q')^{-1}\tilde{H}_2 Q' = R_w^{\frac{1}{2}} U \tilde{H}_2 U' R_w^{-\frac{1}{2}} \tag{11}$$

where the columns of $\tilde{H}_2$ are obtained from (10) and are given by

$$\tilde{h}_l = T\lambda_l (M + \lambda_l I)^{-1} T^{-1} e_l, \qquad l = 1, \ldots, k. \tag{12}$$

The matrix inversion in (12) is straightforward since M is diagonal.

It is worthwhile noting that matrix equations of the form of (8) are usually studied using matrix vectorization and Kronecker products [7, Sec. 12.3]. For a $k \times k$ matrix, say $A = [a_1, \ldots, a_k]$, let $\mathrm{vec}A$ denote the k^2-dimensional vector obtained by stacking the columns of A on the top of each other, first a_1, then a_2, etc. [7, p. 409]. For two $k \times k$ matrices, say A and B, let $A \otimes B$ denote their *Kronecker product* [7, p. 409]. This product is obtained by substituting each element of A, say a_{il}, by the matrix $a_{il}B$. Hence, $A \otimes B$ is a $k^2 \times k^2$ matrix. The relation $\mathrm{vec}(AXB) = (B' \otimes A)\mathrm{vec}X$ for any $k \times k$ matrix X is useful here [7, p. 410]. Applying this relation to (8) and denoting $S = R_w^{-1} R_y$ gives

$$(I \otimes L + S \otimes I)\mathrm{vec}H = \mathrm{vec}S'. \tag{13}$$

The matrix $(I \otimes L + S \otimes I)$ is often called the *Kronecker sum* of L and S [7, p. 412]. From [7, Theorem 2, p. 414], (8) has a unique solution if and only if L and $-S$ have no eigenvalues in common. Since the eigenvalues of S and of $R_w^{-1/2} R_y R_w^{-1/2}$ coincide, this condition is the same as that stated earlier. Application of this approach to (9) provides an alternative proof of (12).

Note from (11) that the optimal linear spectral-domain-constrained estimator first whitens the noise, then applies the orthogonal transformation U' obtained from eigendecomposition of the covariance matrix of the whitened signal, and then modifies the resulting components using the matrix $\tilde{H}_2$. This is analogous to the operation of the optimal linear time-domain-constrained estimator. The matrix $\tilde{H}_2$, however, is not diagonal when the input noise is colored. The explicit solution specified by (11) and (12), the conditions for its uniqueness, and the whitening interpretation are new and were not discussed in [6]. In addition, we allow here a variable number of spectral constraints, any orthogonal matrix V not necessarily related to U, and possibly a unitary matrix V such as that related to the DFT. In [6], the maximal number of constraints $m = k$ was assumed, and the nonorthogonal matrix $V = Q$ was chosen.

When the input noise is white, say $R_w = I$, then $Q = U$. If $V = U$ is chosen, as was suggested in [2], then $T = I$, and $\tilde{H}_2 = \Lambda(\Lambda + M)^{-1}$ as was shown in [2, eq. (38)]. If V is different from U, then the general solution (11) and (12) must be used.

III. Comments

We have extended the signal subspace approach of [2] to colored-noise processes and general spectral constraints, not necessarily those associated with the signal or noise vector subspaces. We provided conditions for uniqueness of the solution of the gradient equation for the optimal linear spectral-domain-constrained estimator and explicitly solved this equation. We showed how the optimal linear time-domain- and spectral-domain-constrained estimators can be implemented using whitening of the input noise. Equation (8) and its solution, derived in this letter for real signals, hold for complex signals as well.

The enhancement algorithm may be implemented in a manner similar to that described in [2]. The particular choices of V and of $\{\alpha_i\}$ are of great practical significance. These aspects, however, were not discussed here. The estimator (12) could be used in two possible ways. The set of Lagrange multipliers $\{\mu_l\}$ may be heuristically adjusted to provide the desired filter. Alternatively, the constraints' equations in (7) could be solved for a given set of $\{\alpha_i\}$. We have assumed here that the noise covariance matrix R_w is positive definite. If this is not the case, then R_w could be regularized by substituting it with $R_w + \sigma I$ for some small positive constant σ. Alternatively, the zero eigenvalues of R_w could be replaced by a small value σ. When the eigensystem used for the explicit solution of (8) is ill-conditioned, then the iterative solution of Bartels and Stewart [1] may be used. The latter algorithm utilizes the Schur–Toeplitz theorem, which ascertains that any square matrix is unitarily similar to an upper triangular matrix [7, p. 176, Th. 2].

Acknowledgment

The authors thank the anonymous referees for their useful comments.

References

[1] R. H. Bartels and G. W. Stewart, "Algorithm 432: Solution of the matrix equation $AX + XB = C$," *Commun. ACM*, vol. 15, no. 9, pp. 820–826, Sept. 1972.

[2] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 251–266, July 1995.

[3] Y. Ephraim and W. J. J. Roberts, "On preprocessing for mismatched classification of Gaussian signals," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1251–1256, Mar. 2001.

[4] F. R. Gantmacher, *The Theory of Matrices*. New York: Chelsea, 1977.

[5] R. M. Gray, "Toeplitz and circulant matrices: II," Stanford Electron. Lab., Stanford, CA, Tech. Rep. 6504-1, Apr. 1977.

[6] Y. Hu and P. C. Loizou, "A subspace approach for enhancing speech corrupted by colored noise," in *Proc. ICASSP*, vol. I, May 2002, pp. 573–576.

[7] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, 2nd ed. New York: Academic, 1985.

[8] D. G. Luenberger, *Linear and Nonlinear Programming*. Reading, MA: Adison-Wesley, 1984.

[9] U. Mittal and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 159–167, Mar. 2000.

[10] A. Rezayee and S. Gazor, "An adaptive KLT approach for speech enhancement," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 87–95, Feb. 2001.