Optimal phase-space projection for noise reduction

Xiaodong Luo,* Jie Zhang, and Michael Small

Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hom Hung, Hong Kong, China

(Received 7 June 2005; published 20 October 2005)

In this communication we will reexamine the widely studied technique of phase-space projection. By imposing a time-domain constraint on the residual noise, we deduce a more general version of the optimal projector, which includes those appearing in previous literature as subcases but does not assume the independence between the clean signal and the noise. As an application, we will apply this technique for noise reduction. Numerical results show that our algorithm has succeeded in augmenting the signal-to-noise ratio for simulated data from the Rössler system and experimental speech record.

DOI: 10.1103/PhysRevE.72.046710 PACS number(s): 05.10.−a

I. INTRODUCTION

Due to its simplicity in implementation and efficiency in computation, noise reduction based on phase-space projection has been widely studied in previous literature. For example, Broomhead and King [1] advocated that, in case of white noise, via singular value decomposition (SVD), one could extract qualitative dynamics from experimental (noisy) time series by removing the empirical orthogonal functions (EOF's) [2] of the trajectory matrix which correspond to the noise components. To deal with the case of colored noise, Allen and Smith [3] proposed a more general method, which would statistically prewhiten colored noise by introducing a transformation to the covariance matrix of noise. In general, phase-space projection based on these methods would not operate on the EOF's that span the signal-plus-noise subspace; therefore, those operations could achieve a lowest possible distortion for the clean signal, but at the price of a highest possible residual noise level [4]. To obtain an optimal trade-off between signal distortion and residual noise so as to minimize the overall distortion, Ephraim and Trees proposed the time-domain constraint (TDC) projector [4], which improves the performance of the existing methods by imposing a constraint on the residual noise and which also includes the existing methods as its subcases. As a generalization, some authors also extended the TDC projector to the cases with colored noise [5,6].

Usually, these authors will make two assumptions concerning the experimental time series. The first assumption is that the time series is stationary and ergodic, and the second one is that the noise components are independent of the clean signal. In this communication we will reexamine the idea of the TDC projector and deduce a more universal version. We will also show that, with the first assumption, the second is not necessary in general.

The remainder of this article will go as follows: In the second section we will introduce the idea of the TDC projector. Based on the assumption that the noisy time series is stationary and ergodic, we will obtain the optimal TDC projector for a trajectory matrix in the sense of minimizing signal distortion subject to a permissible noise level. In the third section we will apply the optimal TDC projector to simulated data from the Rössler system and experimental speech data. We will also compare the performance of the projectors under different TDC's. Finally, a conclusion is available to summarize the whole article.

*Electronic address: enxdluo@eie.polyu.edu.hk

II. MATHEMATICAL DEDUCTION

Given a noisy time series $s=\{s_i\}_{i=1}^M$, we suppose that the corresponding clean signal and the additive noise components are $d=\{d_i\}_{i=1}^M$ and $n=\{n_i\}_{i=1}^M$, respectively; thus for each noisy data point s_i, we have $s_i=d_i+n_i$. In addition, we assume $\{s_i\}_{i=1}^M$ is (weakly) stationary and ergodic so that its expectation exists and its variance is finite, while its (auto) covariances only depend on the time difference between the subsets.

Following the definition in [1], we could construct a $(M-m+1)\times m$ trajectory matrix $\mathbf{S}$ from $\{s_i\}_{i=1}^M$ by letting

$$\mathbf{S}=\begin{pmatrix} s_1 & s_2 & \cdots & s_m \\ s_2 & s_3 & \cdots & s_{m+1} \\ \vdots & & \ddots & \\ s_{M-m+1} & s_{M-m+2} & \cdots & s_M \end{pmatrix}_{(M-m+1)\times m},$$

with $M-m+1>m$. Similarly, we could also obtain the corresponding trajectory matrices $\mathbf{D}$ and $\mathbf{N}$ for components $\{d_i\}_{i=1}^M$ and $\{n_i\}_{i=1}^M$, respectively, and we have $\mathbf{S}=\mathbf{D}+\mathbf{N}$.

For the purpose of noise reduction, we introduce a projection operator $\mathbf{H}$ on the trajectory matrix $\mathbf{S}$ of the noisy signal, through which we could obtain a matrix $\mathbf{Z}=\mathbf{SH}$. We define $\mathbf{R}_0=\mathbf{Z}-\mathbf{D}=\mathbf{D}(\mathbf{H}-\mathbf{I}_m)+\mathbf{NH}$ as the matrix of the residual signal, where the term $\mathbf{D}(\mathbf{H}-\mathbf{I}_m)$ means signal distortion and the term $\mathbf{NH}$ is residual noise. With the intention of data augmentation, we would require the achievement of as small a signal distortion as possible. Thus $\mathbf{H}=\mathbf{I}_m$ would be an intuitive choice. However, in situations such as speech communication, one would also require a permissible residual noise level of the noisy signal, and the objective becomes to minimize signal distortion subject to achieving a permissible residual noise level. Thus, if the initial data does not fulfil this requirement, one has to reduce the initial noise level at the price of introducing possible signal distortion. Similar to the idea proposed in [4], here we impose a TDC μ on the term of residual noise $\mathbf{NH}$ and treat $\mathbf{R}=\mathbf{D}(\mathbf{H}-\mathbf{I}_m)+\mu\mathbf{NH}$ as the part that requires a minimal distortion, where $\mu^2\in[0,+\infty)$ is the Lagrange multiplier determined by the permissible noise level from the practical demand [see Eq. (33) of [4] and the related discussions therein]. Thus our objective will be to minimize the average energy $\Xi=(\Sigma_{i=1}^{M}r_i^2)/M$ of the data set $r=\{r_i\}_{i=1}^{M}$ that (approximately) corresponds to the matrix $\mathbf{R}$. If $M\gg m$, then

$$\Xi \approx \frac{1}{(M-m+1)m}\mathrm{tr}(\mathbf{R}^T\mathbf{R}), \tag{1}$$

where $\mathrm{tr}(\cdot)$ means the trace of a square matrix and $\mathbf{R}^T$ denotes the transpose of the matrix $\mathbf{R}$.

Discarding the constant term $\mathrm{tr}(\mathbf{D}^T\mathbf{D})$ in $\mathrm{tr}(\mathbf{R}^T\mathbf{R})$, we have

$$\mathrm{tr}(\mathbf{R}^T\mathbf{R}) = \mathrm{tr}[\mathbf{H}^T(\mathbf{D}+\mu\mathbf{N})^T(\mathbf{D}+\mu\mathbf{N})\mathbf{H}] - 2\mathrm{tr}[\mathbf{H}^T(\mathbf{D}+\mu\mathbf{N})^T\mathbf{D}]. \tag{2}$$

Taking m as a constant [7], for the minimization problem, by requiring $\partial\mathrm{tr}(\mathbf{R}^T\mathbf{R})/\partial\mathbf{H}=\mathbf{0}$, we would have $(\mathbf{D}+\mu\mathbf{N})^T(\mathbf{D}+\mu\mathbf{N})\mathbf{H}-(\mathbf{D}+\mu\mathbf{N})^T\mathbf{D}=\mathbf{0}$ according to the differential rules in, for example, [9] (p. 472). Therefore the optimal projector

$$\mathbf{H}_{\min} = \{(\mathbf{D}+\mu\mathbf{N})^T(\mathbf{D}+\mu\mathbf{N})\}^{-1}(\mathbf{D}+\mu\mathbf{N})^T\mathbf{D}. \tag{3}$$

With the noise components, $\partial\mathrm{tr}(\mathbf{R}^T\mathbf{R})/\partial\mathbf{H}^2=2(\mathbf{D}+\mu\mathbf{N})^T(\mathbf{D}+\mu\mathbf{N})$ is positive definite, which confirms that the extremum taken at $\mathbf{H}_{\min}$ is a minimum. The corresponding minimal value

$$\mathrm{tr}_{\min}(\mathbf{R}^T\mathbf{R}) = \mathrm{tr}(\mathbf{D}^T\mathbf{D}) - \mathrm{tr}[\mathbf{D}^T(\mathbf{D}+\mu\mathbf{N})\mathbf{H}_{\min}]. \tag{4}$$

But note that $(\mathbf{D}+\mu\mathbf{N})^T$ is not a square matrix; its (ordinary) inverse matrix usually is not defined, and thus we could not cancel the terms of $(\mathbf{D}+\mu\mathbf{N})^T$ in Eq. (3).

Since $\mathbf{S}=\mathbf{D}+\mathbf{N}$, we could also write Eq. (3) in the form of

$$\mathbf{H}_{\min} = \{[\mathbf{S}+(\mu-1)\mathbf{N}]^T[\mathbf{S}+(\mu-1)\mathbf{N}]\}^{-1} \times [\mathbf{S}+(\mu-1)\mathbf{N}]^T(\mathbf{S}-\mathbf{N}). \tag{5}$$

If we assume that the clean signal and the noise components are independent, statistically we have $\mathbf{D}^T\mathbf{N}=\mathbf{N}^T\mathbf{D}=0$ as $M\to\infty$, hence $\mathbf{S}^T\mathbf{S}=\mathbf{D}^T\mathbf{D}+\mathbf{N}^T\mathbf{N}$, and Eq. (5) reduces to

$$\mathbf{H}_{\min} = \{\mathbf{S}^T\mathbf{S}+(\mu^2-1)\mathbf{N}^T\mathbf{N}\}^{-1}(\mathbf{S}^T\mathbf{S}-\mathbf{N}^\mathbf{T}\mathbf{N}). \tag{6}$$

Let $\mathbf{C}_S$ and $\mathbf{C}_N$ denote the covariance matrices of $\{s_i\}_{i=1}^{M}$ and $\{n_i\}_{i=1}^{M}$, respectively, by assuming the expectation values $E(s)=E(n)=0$. We have $\mathbf{C}_S=\mathbf{S}^T\mathbf{S}/(M-m+1)$ and $\mathbf{C}_N=\mathbf{N}^T\mathbf{N}/(M-m+1)$ as $M\to\infty$. Thus Eq. (6) would be expressed as

$$\mathbf{H}_{\min} = \{\mathbf{C}_S+(\mu^2-1)\mathbf{C}_N\}^{-1}(\mathbf{C}_S-\mathbf{C}_\mathbf{N}), \tag{7}$$

which is consistent with the result in, for example, Eq. (3) of [6]. But note that here we use μ^2 to substitute for the multiplier μ in Eq. (3) of [6]. Also note that $\mathbf{H}_{\min}$ in our work is the transpose of that in Eq. (3) of [6]; this is because the trajectory matrices in our work are essentially the transpose of those in [4–6].

In many situations, although the noise components are theoretically uncorrelated to the clean signal, numerical calculations often indicate that the assumption $\mathbf{D}^T\mathbf{N}=\mathbf{N}^T\mathbf{D}=\mathbf{0}$ does not hold strictly for finite data sets. As a more rigorous form, Eq. (5) needs no independence assumption between the noise components and the clean signal. Thus this expression is a further generalization of previous studies.

III. NUMERICAL RESULTS

We note that the trajectory matrices previously introduced are all Hankel matrices. Take the trajectory matrix $\mathbf{S}$ of the noisy signal as an example. Its entries satisfy $\mathbf{S}(i,j)=\mathbf{S}(k,l)$ if $i+j=k+l$, where $\mathbf{S}(i,j)$ denote the element of matrix $\mathbf{S}$ on the ith row and jth column. However, the matrix $\mathbf{Z}=\mathbf{SH}$ usually will not be a Hankel matrix, and we may have many ways to obtain the filtered (or projected) signal $\{z_i\}_{i=1}^{M}$. In our work we use the method of secondary diagonal averaging to extract signal from the matrix $\mathbf{Z}$, which takes the average of the elements along the secondary diagonals of matrix $\mathbf{Z}$ as the filtered signal $\{z_i\}_{i=1}^{M}$ [for details, see [8] (p. 24)], and thus can form a new trajectory (Hankel) matrix $\mathbf{Z}^H$ from $\{z_i\}_{i=1}^{M}$. Golyandina *et al.* prove that this method is optimal among all Hankelization procedures in the sense that the matrix difference $\mathbf{Z}^H-\mathbf{Z}$ has a minimal Frobenius norm ([8], pp. 24 and 266).

We adopt the signal-to-noise ratio R_{SN} as the metric to evaluate the performance of our noise reduction scheme, which is defined (in dB) as [4,10]

$$R_{\mathrm{SN}} = 10\log_{10}\frac{\|d\|^2}{\|z-d\|^2}, \tag{8}$$

where $\|d\|^2=\Sigma_{i=1}^{M}d_i^2$ and $\|z-d\|^2=\Sigma_{i=1}^{M}(z_i-d_i)^2$.

We first apply our algorithm to a simulated data set, which is generated from the x component of the Rössler system:

$$\begin{cases} \dot{x} = -(y+z), \\ \dot{y} = x+ay, \\ \dot{z} = b+(x-c)z, \end{cases} \tag{9}$$

with parameters $a=0.15$, $b=0.2$, and $c=10$. The data are evenly sampled for every 0.1 time units. We generate 10 000 data points and discard the first 1000 to avoid transitions. To construct the trajectory matrices, we will set the window size $m=20$.

Let $\{s_i\}_{i=1}^{M}$ and $\{d_i\}_{i=1}^{M}$ again denote the noisy and clean signals, respectively. We consider adding three types of noise contamination to the clean data. The first one is additive white noise $\{\xi_i\}_{i=1}^{M}$ (so that $s_i=d_i+\xi_i$), which follows the normal Gaussian distribution $N(0,1)$. The second one is additive colored noise $\{\eta_i\}_{i=1}^{M}$ (so that $s_i=d_i+\eta_i$), which, as an example, is produced from a third-order autoregressive process [$P_{\mathrm{AR}}(3)$] in the form of $\eta_i=0.8\eta_{i-1}-0.5\eta_{i-2}+0.6\eta_{i-3}+\xi_i$, where the variable ξ follows the normal distribution $N(0,1)$. The last one is multiplicative noise $\{\zeta_i d_i\}_{i=1}^{M}$ [so that $s_i=(1+\zeta_i)d_i$]. As an example, we let $\zeta_i=\eta_i^2$, where $\{\eta_i\}_{i=1}^{M}$ is from the previous $P_{\mathrm{AR}}(3)$ process; then, the noise component $\{\zeta_i d_i\}_{i=1}^{M}$ is correlated to the clean data $\{d_i\}_{i=1}^{M}$.

TABLE I. Performance of TDC projectors for the Rössler system (in unit of dB).

TDC μ	Additive white noise	Additive colored noise	Multiplicative noise
	$20 \rightarrow 25.50 \pm 0.09$	$20 \rightarrow 20.92 \pm 0.05$	$20 \rightarrow 21.11 \pm 0.10$
0.0	$10 \rightarrow 15.95 \pm 0.11$	$10 \rightarrow 10.89 \pm 0.04$	$10 \rightarrow 11.14 \pm 0.09$
	$0 \rightarrow 5.88 \pm 0.06$	$0 \rightarrow 0.87 \pm 0.04$	$0 \rightarrow 1.16 \pm 0.10$
	$20 \rightarrow 25.80 \pm 0.10$	$20 \rightarrow 21.07 \pm 0.05$	$20 \rightarrow 23.23 \pm 0.24$
0.5	$10 \rightarrow 17.74 \pm 0.15$	$10 \rightarrow 11.64 \pm 0.06$	$10 \rightarrow 14.42 \pm 0.23$
	$0 \rightarrow 9.71 \pm 0.10$	$0 \rightarrow 3.12 \pm 0.08$	$0 \rightarrow 6.97 \pm 0.22$
	$20 \rightarrow 26.27 \pm 0.11$	$20 \rightarrow 21.17 \pm 0.05$	$20 \rightarrow 24.44 \pm 0.34$
1.0	$10 \rightarrow 18.29 \pm 0.16$	$10 \rightarrow 11.89 \pm 0.07$	$10 \rightarrow 16.12 \pm 0.32$
	$0 \rightarrow 10.10 \pm 0.09$	$0 \rightarrow 4.15 \pm 0.09$	$0 \rightarrow 9.56 \pm 0.33$

By varying the magnitude of the introduced noise, we have the initial noise level be 20 dB, 10 dB, and 0 dB, respectively, and for each noise level, we will include ten different noise samples from the same process in calculation. We will also study the performance of the projectors under different constraints. As examples, we let TDC $\mu=0$, 0.5, and 1 separately. TDC $\mu=0$ will lead to the least-squares (LS) projector based on the SVD technique, which appeared in, for example, [1–3]. We would need to specify the dimension of the signal-plus-noise subspace so as to group the EOF's and eigenvalues that correspond to the noisy signal and remove the complementary noise subspace, which is essentially related to the problem of choosing the embedding dimension for embedding reconstruction from a scalar time series (see the discussion in [10]). Thus here we adopt the criterion of false nearest neighbor [11], a method proposed for selection of appropriate embedding dimensions. To apply this criterion in calculations, we utilized the codes implemented in the TISEAN package [12] and found that the proper dimension size K of the signal-plus-noise subspace is 5 in our cases. For $\mu=1$, we will obtain the well-known linear minimum mean-squared-error (LMMSE) projector (detailed introductions are available in, e.g., [13]). After all of the calculations, we finally list the performance of these TDC

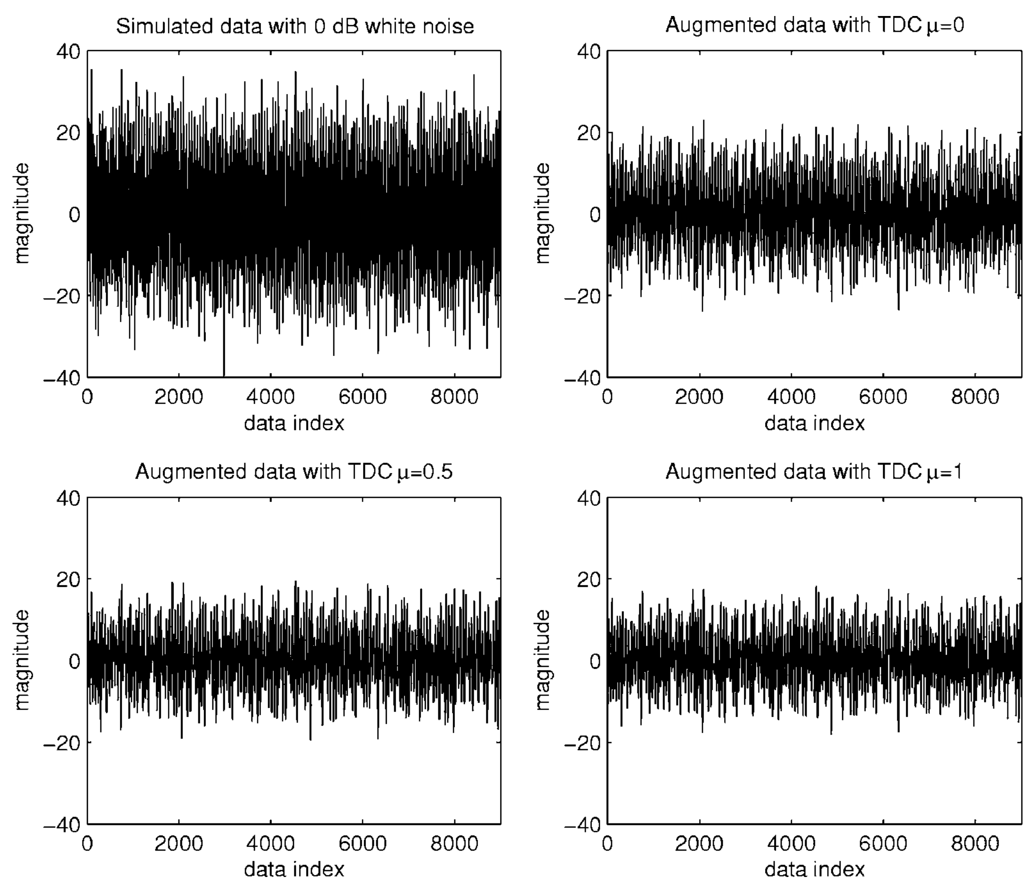

FIG. 1. (a) Time series from the Rössler system contaminated with 0 dB additive white noise. (b), (c), and (d) Augmented time series by TDC projectors with $\mu=0$, 0.5, and 1 separately.

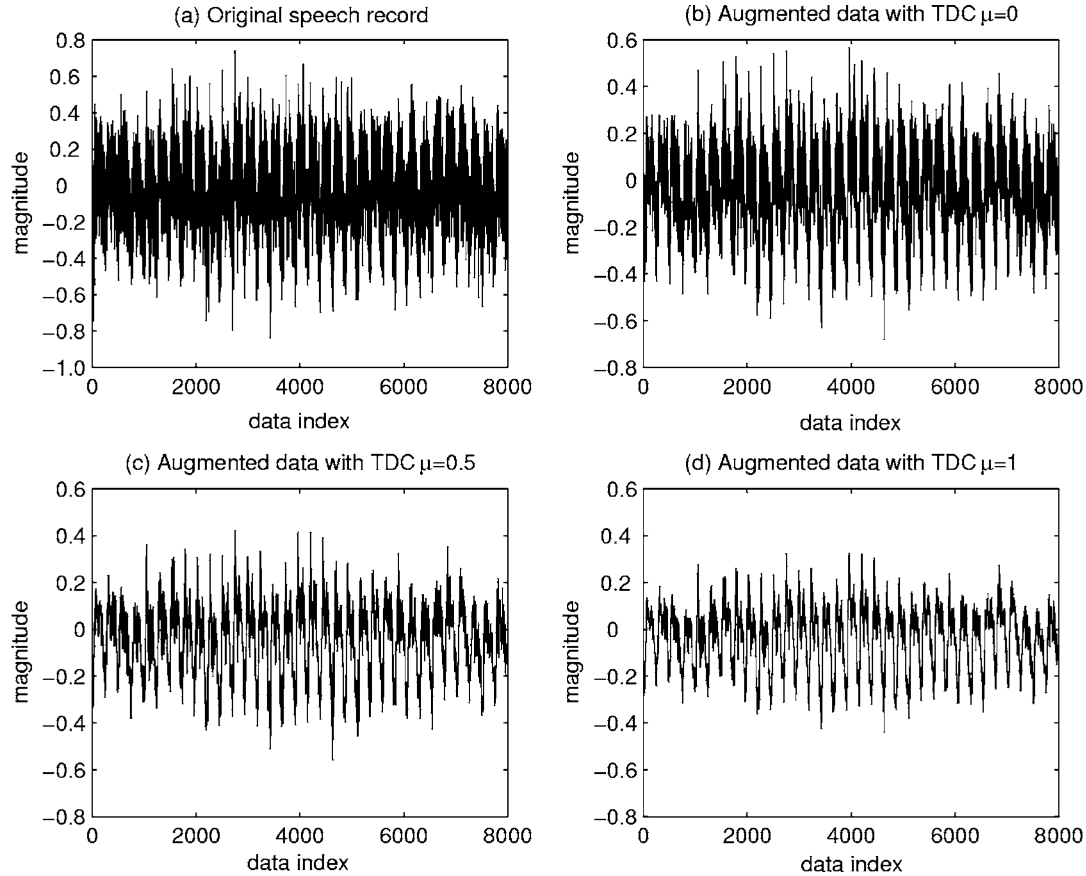

FIG. 2. (a) Original speech record. (b), (c), and (d) Speech data output from TDC projectors with $\mu=0$, 0.5, and 1 separately.

projectors in Table I. For better comprehension of the presented results, we provide the wave forms of all of the data listed in Table I as supplementary material [14]. To keep our presentation concise, here we only take out the raw data contaminated with 0 dB additive white noise as an example and depict its wave form of in panel (a) of Fig. 1. For comparison, we also plot the augmented data with TDC=0, 0.5, and 1 in panels (b), (c), and (d), whose mean noise levels are 5.88, 9.71, and 10.10 dB correspondingly.

From Table I, we see that for the Rössler system, our algorithm works for all of three types of contamination. But the data augmentation for additive colored noise is not as obvious as those for additive white noise and multiplicative noise (the possible explanation is explored in the Appendix). We also see that, in general, the LMMSE projector has better performance than that of the LS projector in the sense that it can achieve a better SNR as defined in Eq. (8).

We then apply our algorithm to a very noisy speech (vowel) data (with 8000 data points), which is sampled at 44 kHz and quantized to 16 bits. In this case we only know the background noise measured in the period without the signal. It would be preferred if we could produce a set of samples that mimic the behavior of the underlying noise. Here we adopt the pseudoperiodic surrogate (PPS) algorithm [15] to generate nine surrogates based on the original background noise. With these data sets, the initial signal-to-noise ratio (SNR) of the speech data is estimated to be -0.32 ± 0.18 dB via Eq. (8). To introduce phase-space projection to the speech data, we let the window size $m=30$ and set the dimension size of the signal-plus-noise subspace to be $K=8$, and then apply the TDC projectors $\mathbf{H}$ to its trajectory matrix. For the LS projector ($\mu=0$), the augmented $R_{\mathrm{SN}}=4.36\pm0.41$ dB. While for TDC $\mu=0.5$ and 1, the corresponding SNR's increase to 6.28 ± 0.61 dB and 6.97 ± 0.66 dB, respectively. As an illustration, we plot the wave forms of the original speech record and three projected data under different TDC's in Fig. 2, from which we can see that the LMMSE projector ($\mu=1$) would lead to a smoother speech wave form [panel (d)] than that of the LS projector [panel (b)]. Although the speech data output from the LMMSE projector has lower (signal) magnitudes than those of the speech record from the LS projector, it is still preferred to its rival in speech communication since a smoother data will usually bring better communication quality.

IV. CONCLUSION

In this communication we reexamined the noise reduction technique based on phase-space projection. By imposing a constraint on the residual noise, we deduced the optimal time-domain constrained projector in the sense of minimizing the signal distortion subject to a permissible noise level. We also showed that, in general, we need not assume independence between clean signal and noise components as was previously done. This viewpoint was confirmed by our numerical results (see the third column of the calculation results in Table I).

ACKNOWLEDGMENT

This research was supported by Hong Kong University Grants Council Competitive Earmarked Research Grant (CERG) No. PolyU 5216/04E.

APPENDIX

Here let us examine the metric of signal-to-noise ratio in more detail. According to the definition in Eq. (8), $R_{SN} = 10\log_{10}\|d\|^2/\|z-d\|^2$, where $\|d\|^2=\Sigma_{i=1}^M d_i^2$ and $\|z-d\|^2 = \Sigma_{i=1}^M (z_i-d_i)^2$. Note that $\|d\|^2=\mathrm{tr}(\mathbf{D}^T\mathbf{D})/m$ and $\|z-d\|^2 = \mathrm{tr}[(\mathbf{Z}-\mathbf{D})^T(\mathbf{Z}-\mathbf{D})]/m$ as $M\to\infty$; thus,

$$R_{SN} = 10\log_{10}\mathrm{tr}(\mathbf{D}^T\mathbf{D}) - 10\log_{10}\mathrm{tr}[(\mathbf{Z}-\mathbf{D})^T(\mathbf{Z}-\mathbf{D})].$$

Since $\mathbf{Z}=\mathbf{SH}$, we have $\mathrm{tr}[(\mathbf{Z}-\mathbf{D})^T(\mathbf{Z}-\mathbf{D})]=\mathrm{tr}(\mathbf{H}^T\mathbf{S}^T\mathbf{SH}) - 2\,\mathrm{tr}(\mathbf{D}^T\mathbf{SH})+\mathrm{tr}(\mathbf{D}^T\mathbf{D})$. For the case that the noise and the clean signal are independent, substituting the optimal projector $\mathbf{H}_{\min}$ into the expression, it can be shown that $\mathrm{tr}_{\min}[(\mathbf{Z}-\mathbf{D})^T(\mathbf{Z}-\mathbf{D})]=\mathrm{tr}(\mathbf{D}^T\mathbf{D})-\mathrm{tr}(\mathbf{H}_{\min}\mathbf{D}^T\mathbf{D})$. For simplicity, we assume the expectation values $E(d)=E(n)=0$; then, $\mathbf{C}_D = \mathbf{D}^T\mathbf{D}/(M-m+1)$ and $\mathbf{C}_N=\mathbf{N}^T\mathbf{N}/(M-m+1)$ as $M\to\infty$, where $\mathbf{C}_D$ and $\mathbf{C}_N$ are the covariance matrix of the clean signal and the noise, respectively, and $\mathbf{H}_{\min}$ can be expressed in the form of Eq. (7) or, equivalently, $\mathbf{H}_{\min}=\{\mathbf{C}_D + \mu^2\mathbf{C}_N\}^{-1}\mathbf{C}_D$. Therefore, in this case, we have $\mathrm{tr}_{\min}[(\mathbf{Z}-\mathbf{D})^T(\mathbf{Z}-\mathbf{D})]=\mathrm{tr}(\mathbf{C}_D)-\mathrm{tr}(\mathbf{H}_{\min}\mathbf{C}_D)$, and thus the maximal SNR can be expressed by

$$R_{SN\,\max} = 10\log_{10}\mathrm{tr}(\mathbf{C}_D) - 10\log_{10}[\mathrm{tr}(\mathbf{C}_D) - \mathrm{tr}(\{\mathbf{C}_D + \mu^2\mathbf{C}_N\}^{-1}\mathbf{C}_D^2)]. \quad \text{(A1)}$$

Through the SVD technique [1], $\mathbf{C}_D$ can be written as $\mathbf{C}_D=\mathbf{V}_D\mathbf{\Lambda}_D\mathbf{V}_D^T$, where $\mathbf{V}_D$ is the normalized eigenvector matrix of $\mathbf{C}_D$ and $\mathbf{\Lambda}_D$ is a diagonal matrix whose nonzero elements are the eigenvalues of $\mathbf{C}_D$ (in fact $\mathbf{V}_D^T\mathbf{V}_D=\mathbf{I}_m$ and $\mathbf{C}_D\mathbf{V}_D=\mathbf{V}_D\mathbf{\Lambda}_D$). Similarly, we have $\mathbf{C}_N=\mathbf{V}_N\mathbf{\Lambda}_N\mathbf{V}_N^T$. Let $\mathbf{V}_N = \mathbf{V}_D\mathbf{P}_{DN}$ (for better comprehension, $\mathbf{P}_{DN}$ can be thought as a kind of projection from $\mathbf{V}_N$ on $\mathbf{V}_D$); then, $\mathbf{C}_N = \mathbf{V}_D\mathbf{P}_{DN}\mathbf{\Lambda}_N\mathbf{P}_{DN}^T\mathbf{V}_D^T$. Substituting it into Eq. (10), we have

$$R_{SN\,\max} = 10\log_{10}\mathrm{tr}(\mathbf{\Lambda}_D) - 10\log_{10}[\mathrm{tr}(\mathbf{\Lambda}_D) - \mathrm{tr}(\{\mathbf{\Lambda}_D + \mu^2\mathbf{P}_{DN}\mathbf{\Lambda}_N\mathbf{P}_{DN}^T\}^{-1}\mathbf{\Lambda}_D^2)].$$

If the noise components are white, we have $\mathbf{\Lambda}_N=\sigma^2\mathbf{I}_m$ (with σ being the standard deviation of the noise process) and $\mathbf{V}_N=\mathbf{V}_D$ (i.e., $\mathbf{P}_{DN}=\mathbf{I}_m$) [3]. However, for the case of colored noise, usually $\mathbf{P}_{DN}\neq\mathbf{I}_m$. Instead it is possible that the absolute values of the elements in $\mathbf{P}_{DN}$ are relatively small. Thus, even for the same clean signal $\{d_i\}_i^M$, the $R_{SN\,\max}$ performance of the colored noise might be much worse than that of the white noise. This fact might explain the observation that the results in Table I are not that promising for additive colored noise.

[1] D. S. Broomhead and G. P. King, Physica D **20**, 217 (1986).
[2] R. Vautard and M. Ghil, Physica D **35**, 395 (1989).
[3] M. R. Allen and L. A. Smith, Phys. Lett. A **234**, 419 (1997).
[4] Y. Ephraim and H. L. V. Trees, IEEE Trans. Speech Audio Process. **3**, 251 (1995).
[5] S. Doclo and M. Moonen, IEEE Trans. Speech Audio Process. **13**, 53 (2005).
[6] Y. Hu and P. C. Loizou, IEEE Trans. Speech Audio Process. **11**, 334 (2003).
[7] We share the viewpoint with Golyandina *et al.* that (currently) there is no universal rule for the optimization problem of *m*. For general suggestions on choosing *m*, readers are referred to [5,8] (p. 53). In practice it is also possible that *m* is a constant—for instance, the signal length in a frame of speech record.
[8] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques* (Chapman and Hall/CRC, London, 2001).
[9] H. Lütkepohl, *Introduction to Multiple Time Series Analysis* (Springer-Verlag, Berlin, 1991).
[10] M. T. Johnson and R. J. Povinelli, Physica D **201**, 306 (2005).
[11] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, Phys. Rev. A **45**, 3403 (1992).
[12] R. Hegger, H. Kantz, and T. Schreiber, Chaos **9**, 413 (1999).
[13] H. L. V. Trees, *Detection, Estimation, and Modulation Theory* (Wiley, New York, 1968), Vol. 1.
[14] See EPAPS Document No. E-PLEEE8-72-167510 for the waveforms of all of the data listed in Table I. The document may also be reached via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html) or from ftp.aip.org in the directory /epaps/. See the EPAPS homepage for more information.
[15] M. Small, D. Yu, and R. G. Harrison, Phys. Rev. Lett. **87**, 188101 (2001).