A Robust Past Algorithm for Subspace Tracking in Impulsive Noise

Shing-Chow Chan, *Member, IEEE*, Yu Wen, and Ka-Leung Ho, *Senior Member, IEEE*

Abstract—The PAST algorithm is an effective and low complexity method for adaptive subspace tracking. However, due to the use of the recursive least squares (RLS) algorithm in estimating the conventional correlation matrix, like other RLS algorithms, it is very sensitive to impulsive noise and the performance can be degraded substantially. To overcome this problem, a new robust correlation matrix estimate, based on robust statistics concept, is proposed in this paper. It is derived from the maximum-likelihood (ML) estimate of a multivariate Gaussian process in contaminated Gaussian noise (CG) similar to the M-estimates in robust statistics. This new estimator is incorporated into the PAST algorithm for robust subspace tracking in impulsive noise. Furthermore, a new restoring mechanism is proposed to combat the hostile effect of long burst of impulses, which sporadically occur in communications systems. The convergence of this new algorithm is analyzed by extending a previous ordinary differential equation (ODE)-based method for PAST. Both theoretical and simulation results show that the proposed algorithm offers improved robustness against impulsive noise over the PAST algorithm. The performance of the new algorithm in nominal Gaussian noise is very close to that of the PAST algorithm.

Index Terms—DOA estimation, impulsive noise, PAST algorithm, robust statistics, subspace tracking.

I. INTRODUCTION

MANY signal processing applications involve the computation of eigenvalues and eigen-basis of symmetric or Hermitian matrices. In some applications, only part of the eigen-structure needs to be updated. Instead of updating the whole eigen-structure, a subspace-tracking algorithm only works with the signal or noise subspace. The computation and storage requirements can therefore be significantly reduced. This advantage makes subspace-tracking algorithm very attractive and it has emerged recently as a valuable tool in array signal processing [12], [32], [42], blind system identification [22], [24], [34], [38], and a variety of applications in communications such as multiuser detection [35], [37], [39] and coherent combining in space-time processing [5]. A number of fast subspace tracking algorithms have been proposed [1], [3], [4], [7]–[9], [16], [17], [30], [32], [41], [42]. One very attractive subspace tracking algorithm is the PAST algorithm [42], which continuously estimates the signal subspace by minimizing the least square errors between the observation and an estimate

derived from the subspace estimated. Using projection approximation, the subspace vectors can be determined by minimizing an approximation of this distortion measure using the RLS algorithm.[1] Unfortunately, the RLS algorithm is extremely vulnerable to impulsive noise in nature. Such interference, which is either man-made or occurring naturally [2], significantly affects the performance of RLS-based subspace tracking algorithm. Simulation results, to be presented in Section V, show that the estimation error of RLS-based PAST algorithm increases significantly and becomes very large when the ambient noise exhibits impulsive characteristics. Any other RLS-based subspace tracking algorithms are therefore likely to suffer from the same problem. The reason is that the conventional auto-correlation matrix estimate: $\hat{C}_{xx}(n) = \sum_{i=1}^{n} \lambda^{n-i} \underline{x}(i)\underline{x}^T(i)$, where λ is the forgetting factor and $\underline{x}(i) = \underline{\tilde{x}}(i) + \underline{n}(i)$ is the input signal vector, is not a robust estimate of the underlying autocorrelation $C_{xx} = E[\underline{\tilde{x}}(i)\underline{\tilde{x}}^T(i)]$, if the "noise free" signal vector $\underline{\tilde{x}}(i)$ is corrupted by noise $\underline{n}(i)$ with impulsive characteristics [36]–[39]. It is therefore not surprising that RLS-based subspace tracking and other least squares-based algorithms are sensitive to impulsive or non-Gaussian noise [6], [18], [19], [36]–[39], [46], as we shall see later from the simulation results. In fact, this problem has been studied in the area of robust statistics [15] and the minimum volume ellipsoid (MVE) or other more robust estimators should be used. However, their computational complexities are usually prohibitive for real-time applications. We notice that, in recursive subspace estimation, a rough prior knowledge of the subspace estimate is available from previous iterations. Therefore, it is easier to detect whether the incoming signal vector is potentially corrupted by impulsive noise or not. This idea happens to coincide with the M-estimators or Maximum likelihood-like estimators of the correlation matrix [15]. This motivates us to consider in this paper the problem of robust subspace tracking under impulsive noise. First of all, a new robust estimate of the correlation matrix in contaminated Gaussian (CG) noise is proposed. It is derived from the maximum-likelihood (ML)[2] estimate of a multivariate Gaussian process in CG noise and the M-estimates in robust statistics. This new estimator is incorporated into the PAST algorithm to obtain a new robust PAST algorithm for robust subspace tracking in impulsive noise. More precisely, a robust statistic-based adaptive filters [6], [46]–[49], called the recursive least M-estimate (RLM) algorithm for matrix parameters, is derived in Appendix B for the efficient implementation of the

[1]Readers are referred to the textbook [14] for more details of the RLS algorithm.

[2]Readers are referred to textbooks such as [27] for an introduction to ML estimation.

robust PAST algorithm. The impulse-corrupted data vectors are detected using the robust M-estimator and are prevented from corrupting the subspace estimate. Furthermore, to handle long burst of impulsive interference, a restoring mechanism is also devised so that the tracking algorithm can recover more quickly from the hostile effect of the impulses. These new mechanisms prevent the impulsive noise from spoiling the fragile subspace tracking process. The convergence of the proposed algorithm is analyzed by extending a previous ordinary differential equation (ODE)-based method for PAST [43]. Both theoretical and simulation results show that the proposed algorithm offers improved robustness against impulsive noise over the PAST algorithm. On the other hand, the performance of the new algorithm in nominal Gaussian noise is very close to that of the PAST algorithm. The layout of the paper is as follows: Section II is a brief introduction to subspace tracking and the PAST algorithm. The proposed robust subspace tracking algorithm is then discussed in Section III. Section IV is devoted to the convergence analysis of the robust PAST, followed by simulation results in Section V. Finally, conclusions are drawn in Section VI.

II. SUBSPACE TRACKING AND THE PAST ALGORITHM

Subspace estimation plays an important role in a wide variety of signal processing applications. Two famous and successful examples are the multiple signal classification algorithm (MUSIC) [28] and the ESPRIT algorithm [26], [33]. They are used to estimate directions of arrival (DOA) or frequencies of sinusoidal plane waves from the sample data vectors of an antenna array [31]. Since the implementation of these techniques, which is based on classic eigenvalue decomposition (ED) or singular value decomposition (SVD), is computationally very expensive, adaptive subspace tracking algorithms have been proposed recently to reduce the computational complexity. One of them is the PAST (and PASTd) algorithm [42], which is based on RLS technique. We assume that there are r narrow-band incoherent complex sinusoidal signals impinging an array of N sensors, thus $\underline{x}(i) \in C^N$ is the data vector observed at the i-th snapshot. $\underline{x}(i) = [x_1(i), x_2(i), \ldots, x_N(i)]^T$ consists of the samples of the N sensors. Taking into account the additive noise, we have

$$\underline{x}(i) = \sum_{k=1}^{r} s_k(i)\underline{a}(\omega_k) + \underline{n}(i) = \mathbf{A}\underline{s}(i) + \underline{n}(i) \qquad (1)$$

where $\mathbf{A} = [\underline{a}(\omega_1), \underline{a}(\omega_2), \ldots, \underline{a}(\omega_r)]$ is a deterministic $n \times r$ matrix of the steering vectors, $\underline{s}(i) = [s_1(i), s_2(i), \ldots, s_r(i)]^T$ is a random source vector, and $\underline{n}(i)$ is a zero-mean spatially-white noise vector which is uncorrelated with $\underline{s}(i)$ and has a covariance matrix of $\sigma^2 \mathbf{I}$. In the case of a uniform linear array, the steering vector takes the special form of $\underline{a}(\omega_k) = [1, e^{j\omega_k}, \ldots, e^{j(N-1)\omega_k}]^T$, where ω_k is the angular frequency of the kth sinusoid. It can be shown that $\underline{x}(i)$ is a complex-valued random vector process with autocorrelation matrix $C_{xx} = E[\underline{x}(i)\underline{x}^H(i)] = \mathbf{A}C_{ss}\mathbf{A}^H + \sigma^2 \mathbf{I}$, where $\mathbf{C}_{ss} = E[\underline{s}(i) \cdot \underline{s}^H(i)]$ is the auto-correlation matrix of $\underline{s}(i)$. Let λ_i and $\underline{u}_i$ be respectively the i-th largest eigenvalue of C_{xx} and its corresponding eigenvector, then C_{xx} can be written in matrix notation as: $C_{xx} = \mathbf{U}\Sigma\mathbf{U}^H$, where

$\Sigma = \text{diag}(\lambda_1, \ldots, \lambda_N)$ and $\mathbf{U} = [\underline{u}_1, \ldots, \underline{u}_N]$. If r is less than N, then $\lambda_1 \geq \cdots \geq \lambda_r \geq \lambda_{r+1} = \cdots = \lambda_N = \sigma^2$, and the corresponding column span of eigenvectors: $\mathbf{U}_s = [\underline{u}_1, \ldots, \underline{u}_r]$ and $\mathbf{U}_n = [\underline{u}_{r+1}, \ldots, \underline{u}_N]$ are called, respectively, the signal subspace and noise subspace. The PAST algorithm [1] continuously estimates the signal subspace by minimizing the following cost function of $\mathbf{W}(i)$

$$J_P(\mathbf{W}(i)) = \sum_{n=1}^{i} \beta^{i-n} \|\underline{x}(n) - \mathbf{W}(i)\underline{y}(n)\|_2^2 \qquad (2)$$

where $0 < \beta < 1$ is a forgetting factor, and $\underline{y}(n) = \mathbf{W}^H(n-1)\underline{x}(n)$ is called the projection approximation. When $\mathbf{W}(n-1)$ is close to $\mathbf{W}(i)$, $J_P(\mathbf{W}(i)) \approx \sum_{n=1}^{i} \beta^{i-n} \|\underline{x}(n) - \mathbf{W}(i)\mathbf{W}^H(i)\underline{x}(n)\|_2^2 = J(\mathbf{W}(i))$. It has been proved in [42] that $J(\mathbf{W}(i))$ has a unique global minimum at which the column span of $\mathbf{W}(i)$ equals the signal subspace and there are no other local minima. Therefore, the signal subspace of C_{xx} can be reliably estimated by minimizing $J(\mathbf{W}(i))$, say using some iterative methods. Meanwhile, the minimization of $J(\mathbf{W}(i))$ will automatically result in a solution of $\mathbf{W}(i)$ with orthonormal columns. Due to the use of the projection approximation, (2) can be solved recursively using the RLS algorithm and it leads to the PAST algorithm in Table I [42]. The superscript ^{H} denotes Hermitian transpose and the operator $\text{Tri}\{\cdot\}$ indicates that only the upper (or lower) triangular part of the matrix argument is calculated and its Hermitian transposed version is copied to the lower (or upper) triangular part.

For each input vector $\underline{x}(i)$, the algorithm computes a new estimate of the signal subspace $\mathbf{W}(i)(\hat{\mathbf{U}}_s)$ from the previous estimate $\mathbf{W}(i-1)$. As mentioned earlier, the performance of this algorithm, like the RLS algorithm, is extremely sensitive to impulsive noise. Suppose that $n(i)$ is modeled as a contaminated Gaussian noise given by $\underline{n}(i) = \underline{n}_g(i) + b(i) \cdot \underline{n}_i(i)$, where $\underline{n}_g(i)$ and $\underline{n}_i(i)$ are uncorrelated zero-mean white Gaussian processes with covariance matrices $\sigma^2 \mathbf{I}_N$ and $\sigma_i^2 \mathbf{I}_N$, respectively. $\underline{n}_i(i)$ represents the impulsive component with $\sigma_i \gg \sigma$. $b(i) \in \{0, 1\}$ is a random binary sequence independent of $\underline{n}_i(i)$, which indicates the presence (absence) of an impulse at time i if $b(i) = 1(0)$. It can be shown that the correlation matrix C_{xx} becomes $C_{xx} = \mathbf{A}C_{ss}\mathbf{A}^H + \sigma^2 \mathbf{I}_N + E[b^2(i)]\sigma_i^2 \mathbf{I}_N$. Any subspace tracking or eigen-decomposition methods for estimating the subspaces from $C_{xx} = E[\underline{x}(i)\underline{x}^H(i)]$ will be significantly affected by the impulsive component $E[b^2(i)]\sigma_i^2 \mathbf{I}_N$. Here, we define the robust correlation matrix to be $C_{\rho\_xx} = E[\rho_r \cdot \underline{x}(i) \cdot \underline{x}^T(i)]$, where ρ_r is a weight function which should ideally be zero when an impulse is detected in vector $\underline{x}(i)$ and 1 otherwise. This definition of C_{ρ_xx} can be justified more formally using maximum likelihood (ML) estimation. In Appendix A, the ML estimate of the mean and covariance of a multivariate Gaussian process under contaminated Gaussian (CG) noise are derived. It was found that the corresponding ML estimate in (A-7) of Appendix A has the same form as C_{ρ_xx} defined above, except that the weighting function ρ_r becomes a rather complicated function of the underlying processes. Since impulsive noise is usually of short time duration and time varying, its statistics are rather difficult

TABLE I
PAST AND THE ROBUST PAST ALGORITHMS

PAST Algorithm	Robust PAST Algorithm
Initialize $\boldsymbol{P}(0)$ and $\boldsymbol{W}(0)$	Initialize $\boldsymbol{P}(0)$, $\boldsymbol{W}(0)$, $\hat{\sigma}^2(0)$, and $\hat{\mu}(0)$
FOR $i = 1, 2, \dots$ **DO**	**FOR** $i = 1, 2, \dots$ **DO**
$\quad \underline{y}(i) = \boldsymbol{W}^H(i-1)\underline{x}(i)$, $\underline{h}(i) = \boldsymbol{P}(i-1)\underline{y}(i)$,	$\quad \underline{y}(i) = \boldsymbol{W}^H(i-1)\underline{x}(i)$, $\underline{h}(i) = \boldsymbol{P}(i-1)\underline{y}(i)$,
$\quad \underline{g}(i) = \underline{h}(i)/[\beta + \underline{y}^H(i)\underline{h}(i)]$,	$\quad \underline{g}(i) = \underline{h}(i)/[\beta + \underline{y}^H(i)\underline{h}(i)]$,
$\quad \boldsymbol{P}(i) = \dfrac{1}{\beta}Tri\{\boldsymbol{P}(i-1) - \underline{g}(i)\underline{h}^H(i)\}$	$\quad \underline{e}(i) = \underline{x}(i) - \boldsymbol{W}(i-1)\underline{y}(i)$,
$\quad \underline{e}(i) = \underline{x}(i) - \boldsymbol{W}(i-1)\underline{y}(i)$,	$\quad$ update $\hat{\sigma}^2(i)$ and $\hat{\mu}(i)$ using (5),
$\quad \boldsymbol{W}(i) = \boldsymbol{W}(i-1) + \underline{e}(i)\underline{g}^H(i)$.	$\quad \Gamma(i) = 1.96 \cdot \hat{\sigma}(i)$,
END	$\quad \boldsymbol{P}(i) = (1 - \rho_r(|\Delta e_\mu(i)|))\boldsymbol{P}(i-1)$
	$\quad\quad + \rho_r(|\Delta e_\mu(i)|)\dfrac{1}{\beta}Tri\{\boldsymbol{P}(i-1) - \underline{g}(i)\underline{h}^H(i)\}$,
	$\quad \boldsymbol{W}(i) = \boldsymbol{W}(i-1) + \rho_r(|\Delta e_\mu(i)|)\underline{e}(i)\underline{g}^H(i)$,
	END

to estimate accurately. Instead of estimating these quantities in real-time, the basic idea of our robust statistics-based estimator is to choose ρ_r as a function of the residual error of the PAST algorithm, so that a more robust algorithm against impulsive noise can be developed [36].

III. THE ROBUST PAST SUBSPACE TRACKING ALGORITHM

We'll see from the simulation results to be presented in Section V that the conventional correlation matrix and hence the PAST algorithm is extremely sensitive to impulsive noise in the data vector $\underline{x}(i)$. This is also apparent by examining the PAST algorithm given in Table I. If $\underline{x}(i)$ is corrupted by additive impulsive noise, then $\underline{y}(i), \underline{h}(i), \underline{g}(i), \boldsymbol{P}(i), \underline{e}(i)$, and $\boldsymbol{W}(i)$ will be affected in turn by the impulse in $\underline{x}(i)$. The corrupted matrices, $\boldsymbol{P}(i)$ and $\boldsymbol{W}(i)$, will be used to compute the new $\boldsymbol{P}(i)$'s and $\boldsymbol{W}(i)$'s, causing hostile effects on the subspace estimate, which require many iterations to recover (see Figs. 3 and 4). We now consider the proposed robust PAST algorithm using the concept of robust statistics. First of all, we note that the purpose of ρ_r in the robust correlation matrix estimate $\boldsymbol{C}_{\rho\_xx}$ is to deemphasis the impulse-corrupted observation $\underline{x}(i)$. A similar approach can be applied to the PAST algorithm by defining the following robust distortion measure:

$$J_P(\boldsymbol{W}(i)) \approx \sum_{n=1}^{i} \beta^{i-n}\rho_r(\|\underline{e}(n)\|_F - \mu_e) \cdot \|\underline{e}(n)\|_F^2,$$
$$\text{where } \underline{e}(n) = \underline{x}(n) - \boldsymbol{W}(i)\underline{y}(n) \quad (3)$$

and $\|\underline{e}(n)\|_F (= \|\underline{e}(n)\|_2)$ is the Frobenious norm of $\underline{e}(n)$. If the process is ergodic, we are minimizing $J_P(\boldsymbol{W}) = E[\rho_r(\|\underline{e}(i)\|_F - \mu_e) \cdot \|\underline{e}(i)\|_F^2]$ and the weight function $\rho_r(\cdot)$ is chosen as the derivative of an M-estimate function [15]. The principle of the proposed robust measure is detailed in Appendix B. In particular, the nonzero mean of the Frobenious norm of the residual error vector when the PAST algorithm

is still converging is taken into account by including μ_e in $\rho_r(\|\underline{e}(i)\|_F - \mu_e)$. For the modified Huber M-estimate that will be used in this paper, $\rho_r(v) = 1$ when $|v| < \Gamma$ and 0 otherwise, where Γ is a threshold to be estimated continuously. μ_e is the robust location or mean estimator of $\|\underline{e}(i)\|_F$. The reason for choosing the modified Huber M-estimate function is because of its reasonably good performance and simplicity in implementation. Other M-estimate functions can also be used. Simulation results in [49] show that the Hampel three-part redescending function [13] gives slightly better results than the modified Huber function in CG and alpha stable noises. The latter however is simpler to analyze [6]. Note, (3) is a nonlinear system of equations, because $\underline{e}(n)$ in $J_P(\boldsymbol{W}(i))$ is also a function of $\boldsymbol{W}(i)$, and it should be solved iteratively. To reduce the arithmetic complexity, $\rho_r(\|\underline{e}(i)\|_F - \mu_e)$ is assumed to depend weakly on $\boldsymbol{W}(i)$. By treating it as a constant and using the multivariate recursive least M-estimate algorithm (RLM) derived in Appendix B, a robust PAST algorithm for approximately minimizing (3) is obtained in Table I. Furthermore, it will be shown later in Section IV that this approximated algorithm also converges to the robust covariance matrix $\boldsymbol{C}_{\rho\_xx}$. The principle of the robust distortion measure can be seen more clearly by considering the situation where $\underline{x}(i)$ is corrupted by impulses. Under these circumstances, the Frobenius norm of the error vector $\underline{e}(i), \|\underline{e}(i)\|_F$, will become very large (and likely to exceed the threshold Γ). $\rho_r(\|\underline{e}(i)\|_F - \mu_e)$ will become zero and the impulse-corrupted measurement is prevented from entering into the minimization. A similar approach has been successfully applied to develop robust adaptive filters under impulsive noise [6], [46]–[49]. We now consider the estimation of the threshold Γ and the robust mean estimator μ_e (for simplicity, the subscript e in μ_e will be dropped in subsequent discussion). Though the exact distribution of $\|\underline{e}(i)\|_F$ is unknown, for simplicity, it is assumed to be Gaussian distributed but corrupted by additive impulsive noise (note also that $\|\underline{e}(i)\|_F$ is always positive). Under this approximation, the probability

that the deviation of the "impulse free error" from its mean, $|\Delta e_\mu(i)| = |\|\underline{e}(i)\|_F - \hat{\mu}(i)|$, is greater than a given threshold $\Gamma(i)$ is

$$p_\Gamma = P_r\{|\Delta e_\mu(i)| > \Gamma(i)\} = \mathrm{erfc}(\Gamma(i)/\hat{\sigma}(i)) \qquad (4)$$

where $\mathrm{erfc}(r) = (2/\sqrt{\pi})\int_r^\infty e^{-x^2}\,dx$ is the complementary error function. $\hat{\mu}(i)$ and $\hat{\sigma}(i)$ are the estimated mean and standard deviation of the Frobenius norm of the "impulse free" error vector. Using different threshold parameter $\Gamma(i)$, we can detect the presence of the impulsive noise with different degrees of confidence. In this work, p_Γ is chosen to be 0.05 so that we have approximately 95% confidence in saying that the current error vector is corrupted by impulsive noise. The corresponding threshold parameter $\Gamma(i) = \xi \cdot \hat{\sigma}(i)$ is determined to be $\Gamma(i) = 1.96 \cdot \hat{\sigma}(i)$. A commonly used estimate for $\hat{\sigma}^2(i)$ and $\hat{\mu}(i)$ are respectively: $\hat{\sigma}^2(i) = \lambda_\sigma \hat{\sigma}^2(i-1) + (1-\lambda_\sigma)(\Delta e_\mu(i))^2$ and $\hat{\mu}(i) = \lambda_\mu \hat{\mu}(i-1) + (1-\lambda_\mu)\|\underline{e}(i)\|_F$, where λ_μ and λ_σ are some forgetting factors. It is, however, not robust to impulsive noise. In fact, a single impulse with large amplitude can substantially increase the value of $\hat{\sigma}(i)$ and $\hat{\mu}(i)$, and hence the values of $\Gamma(i)$. Better estimates for $\hat{\sigma}^2(i)$ and $\hat{\mu}(i)$ are [47][3]

$$\hat{\sigma}^2(i) = \lambda_\sigma \hat{\sigma}^2(i-1) + 1.483\left(1 + \frac{5}{N_w - 1}\right)$$
$$\times (1 - \lambda_\sigma)\mathrm{med}(A((\Delta e_\mu(i))^2)) \quad (5a)$$

and

$$\hat{\mu}(i) = \lambda_\mu \hat{\mu}(i-1) + (1-\lambda_\mu)\mathrm{med}(A(\|\underline{e}(i)\|_F)) \qquad (5b)$$

where $A(x(i)) = \{x(i), \ldots, x(i - N_w + 1)\}$, N_w is the length of the estimation window, and $\mathrm{med}(\cdot)$ is the median operator. λ_μ and λ_σ are the forgetting factors. In practice, the value of N_w can be chosen to lie between 5 and 11 in order to reduce the number of operations required by the median filter. For large values of N_w, the pseudo median [25] instead of the median can be computed to reduce the arithmetic complexity. Therefore, the arithmetic complexity of the proposed robust PAST algorithm is comparable to that of the conventional PAST algorithm. Our robust PAST algorithm updates $\Gamma(i) = 1.96 \cdot \hat{\sigma}(i)$, $\hat{\sigma}^2(i)$ and $\hat{\mu}(i)$ at each iteration. If $|\Delta e_\mu(i)| > \Gamma(i)$, both the signal subspace $\boldsymbol{W}(i)$ and the intermediate matrix $\boldsymbol{P}(i)$ will not be updated, preventing the impulse from affecting the subspace estimate. Using the weight function: $\rho_r(|\Delta e_\mu(i)|) = 1$ when $|\Delta e_\mu(i)| < \Gamma(i)$, and 0 otherwise, the robust PAST algorithm in Table I is obtained.

The choice of the threshold parameter ξ for $\Gamma(i) = \xi \cdot \hat{\sigma}(i)$ and N_w has been studied in [48]. It was found that the performances of the robust algorithms are not sensitive to the selection of ξ, provided they are not at the tail part of the distribution. Their values can however be adjusted to provide different tradeoff between robustness and accuracy, as mentioned previously. In addition, a value of N_w between 5 to 11 usually gives little degradation in tracking and sufficient robustness to individual

and consecutive impulses of limited duration. There is however one problem that remains unsolved, which occurs when a long burst of impulses is encountered. In this case, due to the finite length of the median filter N_w, the system might misinterpret the series of error vectors with large Frobenius norm as being created from a sudden system change in the signal subspace, e.g., sudden DOA change. To solve this problem, the differences in statistical properties of $\underline{e}(i)$ during sudden system change and a series of impulses are exploited. For the former case, if the system continues to adapt, the Frobenius norm of the error vector will continue to decrease, reaching a steady state when the algorithm converges. While for a long burst of consecutive impulsive noise, the impulses will also produce a sequence of error vector $\underline{e}(i)$ with large Frobenius norm. However, it remains at a certain level without a deterministic trend of decreasing in its magnitude. Therefore, the following buffering mechanism is adopted to distinguish between the two different situations of sudden system changes and corruption by a series of consecutive impulsive noise.

Suppose that at $i = i_0, |\Delta e_\mu(i)| > \Gamma(i_0)$, which indicates that the input vector might be corrupted by an impulse. $\boldsymbol{P}(i_0 - 1)$, $\boldsymbol{W}(i_0 - 1)$ and $\hat{\mu}(i_0 - 1)$ will be buffered, and the system continues to adapt. After an observation window of length L_w, which is chosen as a certain fraction of the initial convergence time of the tracking system to provide a sufficient decrease in $\|\underline{e}(i)\|_F$ in case of a system change, $\hat{\mu}(i_0 + L_w - 1)$ is compared to $\hat{\mu}(i_0 - 1)$. If $\hat{\mu}(i_0 + L_w - 1)$ is close to $\hat{\mu}(i_0 - 1)$, this means that there is a system change or the system has started to recover from the impulses. The restoring mechanism will not be invoked and the system will continue to adapt as normal. On the other hand, if $\hat{\mu}(i_0 + L_w - 1)$ is much greater than $\hat{\mu}(i_0 - 1)$, consecutive impulsive noise is expected and $\boldsymbol{P}(i_0 + L_w)$ and $\boldsymbol{W}(i_0 + L_w)$ will be reinitialized to $\boldsymbol{P}(i_0 - 1)$ and $\boldsymbol{W}(i_0 - 1)$, respectively. It might happen, though very rarely, that many system changes suddenly happen during the observation window, after a series of impulses, and give rise to a relatively high $\|\underline{e}(i)\|_F$. To avoid the restoring mechanism from disturbing this normal adaptation, we suggest to disable the restoring mechanism for a certain period of time, say 100 symbols, after its last activation. The robustness of the system to very long burst of impulse is therefore weakened. But simulation result shows that this scheme causes very little degradation in sudden system change scenarios and is able to suppress the adverse effect of long burst of impulses by periodic reinitialization. To differentiate the two situations at the end of the observation window, the relative discrepancy $\chi = \frac{\hat{\mu}(i_0 + L_w - 1) - \hat{\mu}(i_0 - 1)}{\hat{\mu}(i_0 - 1)}$ is adopted as a measure. If $\chi < \chi_s$, a certain threshold, it is recognized as a system change. Otherwise, it will be treated as the consecutive noise case. χ_s is chosen as 2 in the simulation section,[4] which means that the restoring mechanism will be invoked if $\hat{\mu}(i_0 + L_w - 1) > 3\hat{\mu}(i_0 - 1)$. More sophisticated system change detection algorithms are available and interested readers are referred to [11] for more details. The proposed algorithm was chosen because of its implementation simplicity and reasonable reliability.

---

[3]The constant 1.483 is a correction factor, which ensures that $\hat{\sigma}^2(i)$ in (5a) is identical to the variance of the input, if it is Gaussian distributed.

[4]This value is experimentally determined to combat hostile effects of long burst of impulses, while avoiding excessive interruption to the adaptation process. Slightly different values can be used to provide different sensitivity to impulse train.

IV. CONVERGENCE ANALYSIS

The convergence analysis of the PAST algorithm was first studied by B. Yang [43]–[45] using the ODE approach [21], [23]. The basic idea is to associate a continuous time deterministic ordinary differential equation with the discrete time stochastic approximation algorithm. Our analysis is an extension of the work in [43] to the robust statistics framework. Due to page limitation, only the key results will be outlined. Assume that the observed signal vector is corrupted by additive noise, which is modeled as a contaminated Gaussian process with $\underline{n}(i) = \underline{n}_g(i) + b(i) \cdot \underline{n}_i(i)$. Hence, $\underline{x}(i) = \underline{x}_S(t) + \underline{n}(t)$. Further, for simplicity, $\underline{x}(i)$ and $\underline{n}(i)$ are assumed to be real-valued. The robust PAST algorithm in Table I can also be written as follows:

Choose $\boldsymbol{P}(0)$ and $\boldsymbol{W}(0)$ suitably.

For $i = 1, 2, \ldots$ Do

$$\underline{y}(i) = \boldsymbol{W}^T(i-1)\underline{x}(i) \tag{6a}$$

$$\underline{e}(i) = \underline{x}(i) - \boldsymbol{W}(i-1)\underline{y}(i) \tag{6b}$$

$$\boldsymbol{C}_{yy}(i) = \beta \cdot \boldsymbol{C}_{yy}(i-1) + \rho_r \cdot \underline{y}(i)\underline{y}^T(i) \tag{6c}$$

$$\boldsymbol{W}(i) = \boldsymbol{W}(i-1) + \rho_r \cdot \underline{e}(i)\underline{y}^T(i)\boldsymbol{C}_{yy}^{-1}(i) \tag{6d}$$

where $\rho(|\Delta e_\mu(i)|) = \rho_r$ is a weighting function, which is equal to the derivative of an M-estimate distortion function (See Appendix B). $\boldsymbol{C}_{yy}(i)$ is the inverse of $\boldsymbol{P}(i)$ in Table I. If $\rho_r = 1$, we obtain the conventional PAST algorithm. Let $\gamma(i) = 1/(i+1)$ and $\boldsymbol{R}_{yy}(i) = \gamma(i)\boldsymbol{C}_{yy}(i)$. Multiplying both sides of (6c) by $\gamma(i)$ and choosing $\beta = 1$, we have

$$\boldsymbol{R}_{yy}(i) = \boldsymbol{R}_{yy}(i-1) + \gamma(i) \cdot \{\rho_r \underline{y}(i)\underline{y}^T(i) - \boldsymbol{R}_{yy}(i-1)\}. \tag{7}$$

Similarly, we can rewrite (6d) as

$$\boldsymbol{W}(i) = \boldsymbol{W}(i-1) + \gamma(i) \cdot \rho_r \underline{e}(i)\underline{y}^T(i)\boldsymbol{R}_{yy}^{-1}(i). \tag{8}$$

Following [43] and the ODE approach in [21], the asymptotic behavior of the asymptotic behavior of the robust PAST algorithm given by (7) and (8) can be described by the following ODEs:

$$\dot{\boldsymbol{R}}(t) = \boldsymbol{W}(t)^T \boldsymbol{C} \boldsymbol{W}(t) - \boldsymbol{R}(t) \tag{9a}$$

$$\dot{\boldsymbol{W}}(t) = (\boldsymbol{I} - \boldsymbol{W}(t)\boldsymbol{W}^T(t))\boldsymbol{C}\boldsymbol{W}(t)\boldsymbol{R}^{-1}(t) \tag{9b}$$

where $\boldsymbol{C} = \boldsymbol{C}_{\rho\_xx} = E[\rho_r \cdot \underline{x} \cdot \underline{x}^T]$, $\boldsymbol{R}(t)$ and $\boldsymbol{W}(t)$ are continuous time version of the discrete time estimate $\boldsymbol{R}_{yy}(i)$, and $\boldsymbol{W}(i)$. For the PAST algorithm, $\boldsymbol{C}$ is given by the conventional covariance estimate $\boldsymbol{C}_{xx} = E[\underline{x} \cdot \underline{x}^T]$. Using the result in [43], the asymptotic convergence of the robust PAST algorithm to the subspace spanned by $\boldsymbol{C} = \boldsymbol{C}_{\rho\_xx}$ can then be established. Note, the robust PAST algorithm will converge to the eigensubspace spanned by the r dominant eigenvectors of $\boldsymbol{C} = \boldsymbol{C}_{\rho\_xx}$. In case of Gaussian noise, there is a small penalty in using $\boldsymbol{C}_{\rho\_xx}$ instead of $\boldsymbol{C}_{xx}$, since the trailing part of the distribution is removed through the choice of the threshold $\Gamma = \xi \cdot \hat{\sigma}$. The value ξ, which is chosen as 1.96 in this work, determines the tradeoff between accuracy of estimation and immunity to impulsive noise. Fortunately, from the simulation results to be presented in Section V, it is observed that such penalty is indeed very small (usually the error norm of the matrix is within 1 to 2% of $\boldsymbol{C}_{xx}$ without the impulsive noise). On the other hand, if the input is corrupted by impulsive noise, the proposed algorithm will converge to $\boldsymbol{C}_{\rho\_xx}$, instead of $\boldsymbol{C}_{xx}$ for the PAST algorithm. Since $\boldsymbol{C}_{\rho\_xx}$ is a better estimator of $\boldsymbol{C}_{ss} = E[\underline{x}_s \underline{x}_s^H]$ than $\boldsymbol{C}_{xx}$ under impulsive noise, the robust PAST algorithm is expected to be less sensitive to contaminated Gaussian noise. Although the above ODE analysis yields the subspaces to which the PAST and robust PAST algorithms will converge, it does not provide us the convergence speed and the error covariance of the algorithms. In [43], [44], two convergence measures are proposed to evaluate the convergence rate of the PAST algorithm

$$f_{\text{ortho}}(i) = \|\boldsymbol{W}^H(i)\boldsymbol{W}(i) - \boldsymbol{I}\|_F^2 \approx i^{-2} \tag{10}$$

$$f_{\text{proj}}(i) = \|\boldsymbol{W}(i)\boldsymbol{W}^H(i) - \boldsymbol{U}_S\boldsymbol{U}_S^H\|_F^2$$

$$\approx \frac{2}{i}\sum_{k=1}^{r}\sum_{j+r+1}^{n} \upsilon(\lambda_j/\lambda_k) \tag{11}$$

where $\boldsymbol{U}_S = [\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_r]$ is the true signal subspace and $\upsilon(x) = x/(1-2x)$ for $\lambda_r > 2\lambda_{r+1}$. $f_{\text{ortho}}(i)$ is a measurement of deviation of $\boldsymbol{W}(i)$ from orthonormality. $f_{\text{proj}}(i)$ measures the difference between the projected estimate and the true signal subspaces. Apparently, one would expect that the rate of convergence r_ρ of robust PAST algorithm is $r_{\text{PAST}}/(1 - p_{\text{imp}})$, where r_{PAST} and p_{imp} are respectively the convergence rate of the PAST measures in *Gaussian noise*, and the occurrence probability of the impulses. However, it is shown in the next section that the impulses will further slow down the adaptation of the algorithm.

V. SIMULATION RESULTS

A. DOA Tracking

The performance of the proposed robust subspace-tracking algorithm is evaluated in a DOA tracking application. In general, the DOA can be estimated by the ESPRIT [26], [33] and the MUSIC [28], [31] algorithms. The problem of estimating DOA under Gaussian mixtures was recently studied using the EM algorithm [20]. The example considered here mainly focused on efficient and robust subspace tracking using the PAST-based algorithms in impulsive noise. The TLS-ESPRIT [26] is employed in our simulation to compute the DOA from the signal subspace estimate. In order to compare the proposed robust subspace tracking algorithm with PAST, the simulation settings are similar to those adopted in [42]. We investigate a uniform linear array with $N = 9$ sensors impinged by three plane sinusoidal waves. Data vectors are generated according to the signal model in (1). Both PAST and robust PAST are employed to track the signal subspace of the same set of data vectors. Then, the DOAs $\theta_k(k = 1, 2, 3)$ of these three plane sinusoidal waves are estimated by TLS-ESPRIT, based on the signal subspace estimates of the PAST and robust PAST algorithms. θ_1 is set to vary linearly from $20°$ to $40°$, while θ_2 varies linearly from $40°$ to $20°$. θ_3 is set to be a constant of $10°$ when there is no subspace system change. For the system change case, θ_3 changes from $10°$ to $0°$ at the time instant of 200th snapshot. Background noise is assumed to be an additive white Gaussian noise (AWGN) with a variance of 1, i.e., 0 dB. Both individual and consecutive impulsive noises are modeled as Gaussian noise with a power of 20 dB and a probability of occurrence of $1 \cdot 10^{-1}$. They intrude
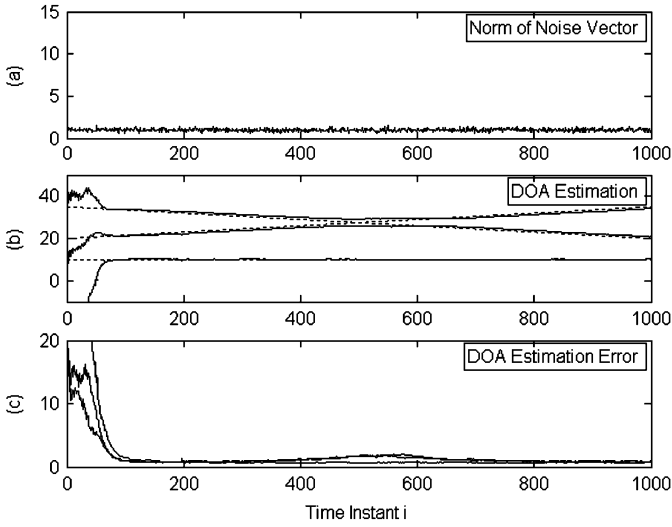
Fig. 1. Estimated DOA of the PAST subspace tracking algorithm in Gaussian noise. (a) One realization of normalized Frobenious-norm of error vector $\|\underline{e}(i)\|_F/\sqrt{N}$. (b) Estimation of DOA (degree). (c) Estimation error of DOA (degree) with impulsive noise power of 20 dB.
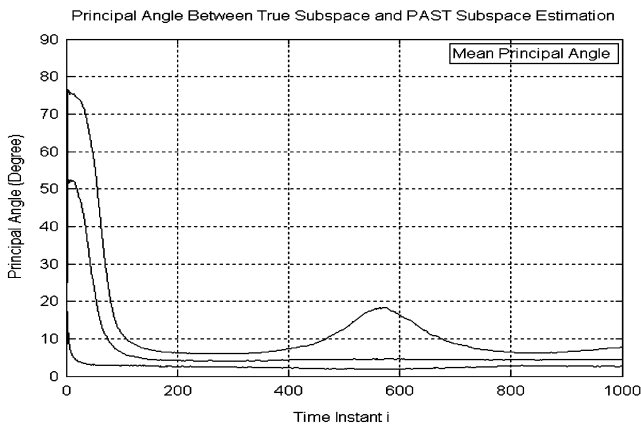

Fig. 2. Mean principle angle between the true and the PAST estimated subspace in Gaussian noise.

the background noise at the 200th and 600th symbols, and last for 50 and 100 symbols, respectively. The first two sinusoidal waves have relative power of 3 dB, and the third one has a relative power of 0 dB, all with respect to the background noise. The forgetting factors β, λ_μ, and λ_σ are all set to be 0.98 while N_w is set to 11. The number of Monte Carlo simulation is 100.

Fig. 1 shows the estimated DOA of the PAST algorithm in Gaussian noise, which is identical to those reported in [42]. Fig. 1(a) shows one realization of the normalized error vector norm $\|\underline{e}(i)\|_F/\sqrt{N}$. Fig. 2 is the corresponding mean principle angle between the true and the PAST estimated subspaces. Fig. 3 shows the DOA estimate of the PAST algorithm in impulsive noise. It is evident from Fig. 3 that the PAST algorithm is vulnerable to the presence of individual or consecutive impulsive noise. The tracking of the signal subspace is substantially interfered and the estimation error of the DOA is very large. The discrepancy between the true subspace and the PAST estimate can be seen more clearly from the mean principle angle plot in Fig. 4. The PAST subspace estimate deviates substantially from the true subspace when impulsive noise is present. Similar
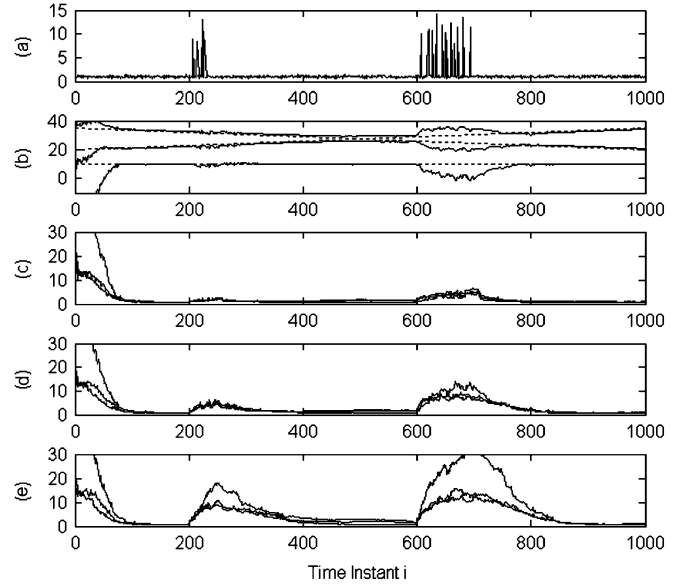

Fig. 3. Estimated DOA of the PAST subspace tracking algorithm in impulsive noise. (a) One realization of normalized Frobenious-norm of error vector $\|\underline{e}(i)\|_F/\sqrt{N}$. (b) Estimation of DOA (degree) (true—dotted line, estimated—solid line). Estimation error of DOA (degree) with different power of impulsive noise: (c) 15 dB, (d) 20 dB, and (e) 25 dB.
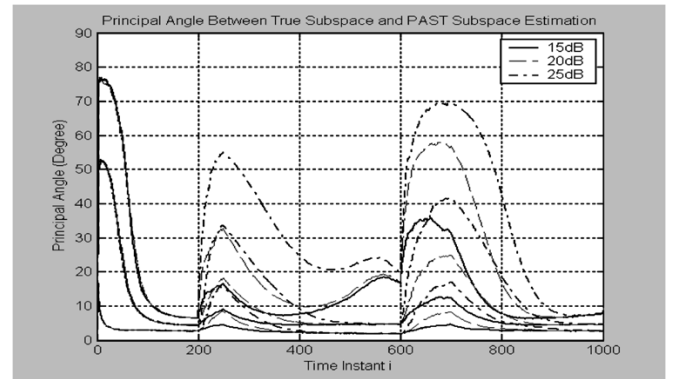

Fig. 4. Mean principle angle between the true and the PAST estimated subspace in impulsive noise.

results also occur when the power of the impulsive noise is changed to 15 dB and 25 dB. Fig. 5 shows the DOA estimate of the proposed robust PAST algorithm in impulsive noise. It can be seen that the robust PAST algorithm is much more robust to the impulsive noise than the conventional PAST algorithm. This demonstrates its robustness over its conventional counterpart for the contaminated Gaussian impulsive noise model. From Fig. 6, we can also see that the robust PAST subspace estimate is very close to the true subspace with a small principle angle even when excessive impulsive noise is experienced. Its performance is also less sensitive to the variation of the power of the impulsive noise. Due to page limitation, a power of 20dB is chosen for the impulsive noise in the following simulations for evaluating the performance of various algorithms during system change. The DOA estimation and mean principle angle errors of the proposed robust PAST algorithm in Gaussian noise are shown in Fig. 7 and Fig. 8, respectively. They are approximately equal to those of the PAST in Gaussian noise, as shown in Figs. 1 and 2. This demonstrates the robustness of the proposed algorithm in both
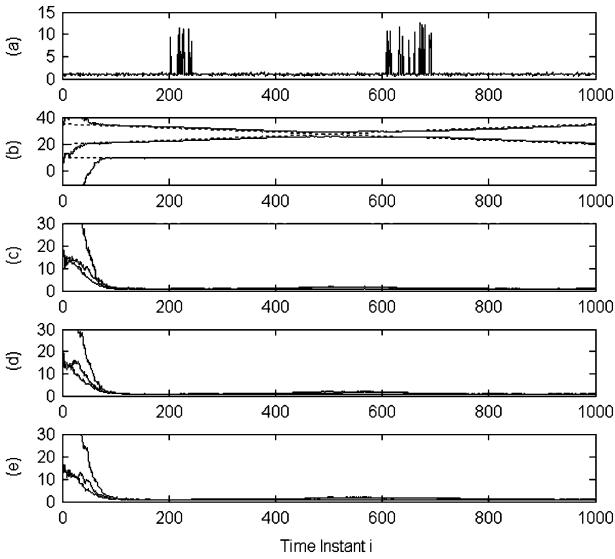
Fig. 5. Estimated DOA of the robust PAST subspace tracking algorithm in impulsive noise. (a) One realization of normalized Frobenious-norm of error vector $\|\underline{e}(i)\|_F / \sqrt{N}$. (b) Estimation of DOA (degree) (true—dotted line, estimated—solid line). Estimation error of DOA (degree) with different power of impulsive noise: (c) 15 dB, (d) 20 dB, and (e) 25 dB.
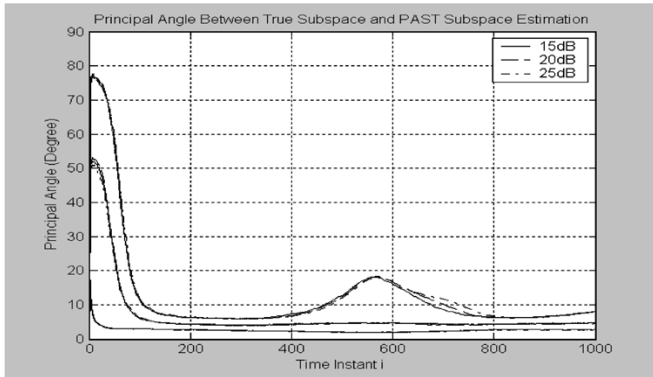

Fig. 6. Mean principle angle between the true and the robust PAST estimated subspace in impulsive noise.
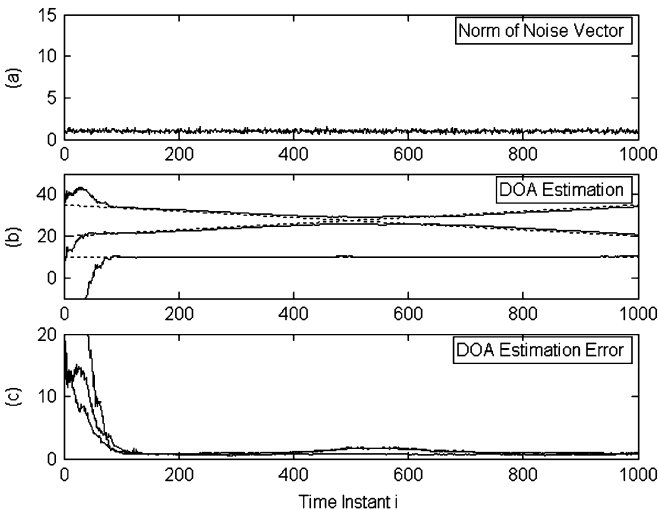

Fig. 7. Estimated DOA of the robust PAST subspace tracking algorithm in Gaussian noise. (a) One realization of normalized Frobenious-norm of error vector $\|\underline{e}(i)\|_F / \sqrt{N}$. (b) Estimation of DOA (degree) (true—dotted line, estimated—solid line). (c) Estimation error of DOA (degree) with impulsive noise power of 20 dB.
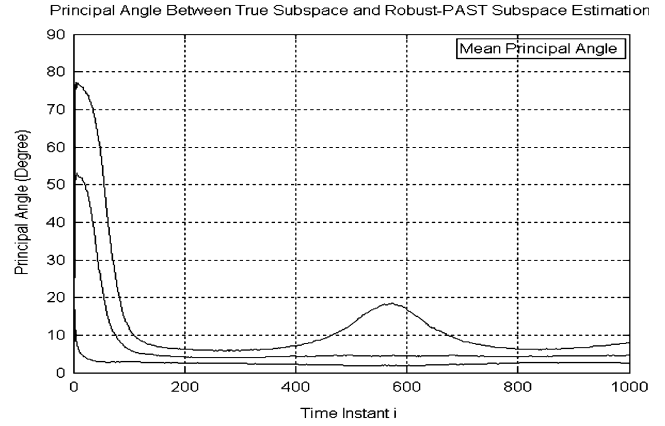

Fig. 8. Mean principle angle between the true subspace and the robust PAST estimated subspace in Gaussian noise.
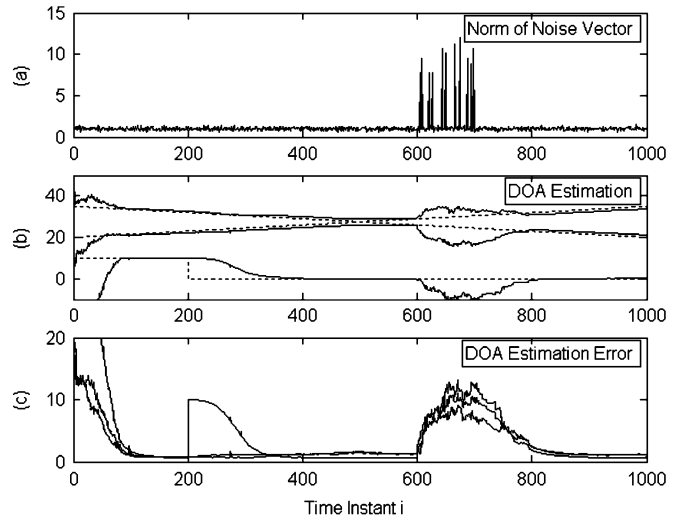

Fig. 9. Estimated DOA of PAST subspace tracking algorithm in impulsive noise with system change. (a) One realization of normalized Frobenious-norm of error vector $\|\underline{e}(i)\|_F / \sqrt{N}$. (b) Estimation of DOA (degree) (true—dotted line, estimated—solid line). (c) Estimation error of DOA (degree) with impulsive noise power of 20 dB.

Gaussian and impulsive noise environment. Fig. 9 shows the performance of the PAST algorithm under both system change (at 200 snapshot) and impulsive noise (from 600 to 700 snapshot). It can be seen that the PAST algorithm is able to track the system change at the 200th snapshot. The 1st and 2nd DOA estimates are less affected. Its behavior under impulsive noise, however, is quite different. All DOA estimates are significantly interfered. Also, the estimation error does not seem to converge in the presence of impulsive noise. Such behavior can be seen more clearly from the principle angles in Fig. 10. Fig. 11 shows the performance of the proposed algorithm under both system change and impulsive noise. It suggests that the proposed algorithm is also capable of tracking the system change with approximately the same speed while providing improved robustness to impulsive noise over the conventional PAST algorithm. This is also supported by the principle angles as shown in Fig. 12. Due to page limitation, simulation results of using different values of ξ are omitted. The performances are very similar if ξ are not chosen at the tail part of the distribution (i.e., too large or too small).
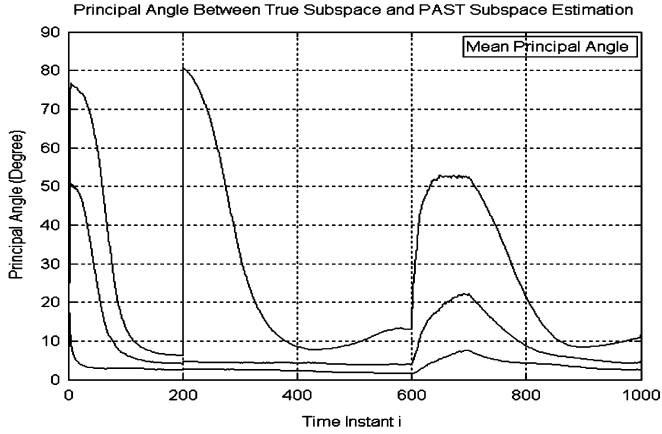
Fig. 10.   Mean principle angle between the true and the PAST estimated subspace in impulsive noise with system change.
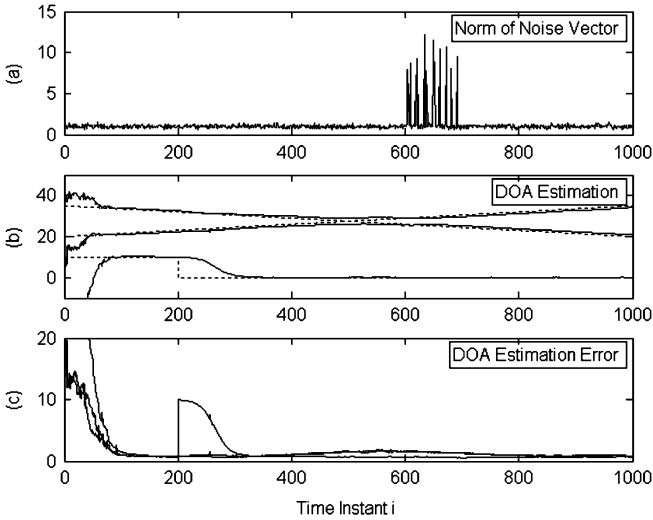

Fig. 11.   Estimated DOA of the robust PAST subspace tracking algorithm in impulsive noise with system change. (a) One realization of normalized Frobenious-norm of error vector $\|\underline{e}(i)\|_F/\sqrt{N}$. (b) Estimation of DOA (degree) (true—dotted line, estimated—solid line). (c) Estimation error of DOA (degree) with impulsive noise power of 20 dB.
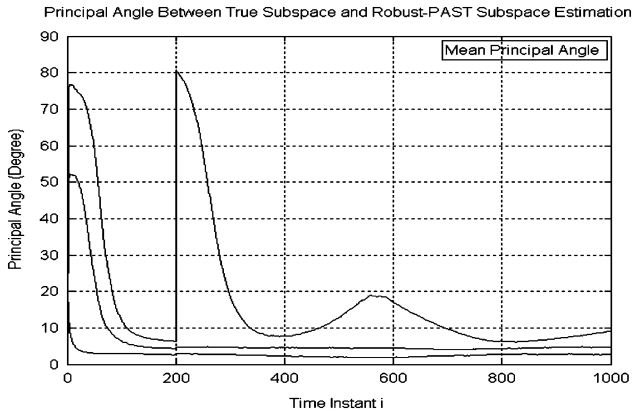

Fig. 12.   Mean principle angle between the true and the robust PAST estimated subspace in impulsive noise with system change.

B. Convergence Performance

The convergence rate of the proposed robust PAST algorithm is evaluated by comparing with the measures $f_{\mathrm{ortho}}(i)$ and $f_{\mathrm{proj}}(i)$ for the PAST under Gaussian noise described in
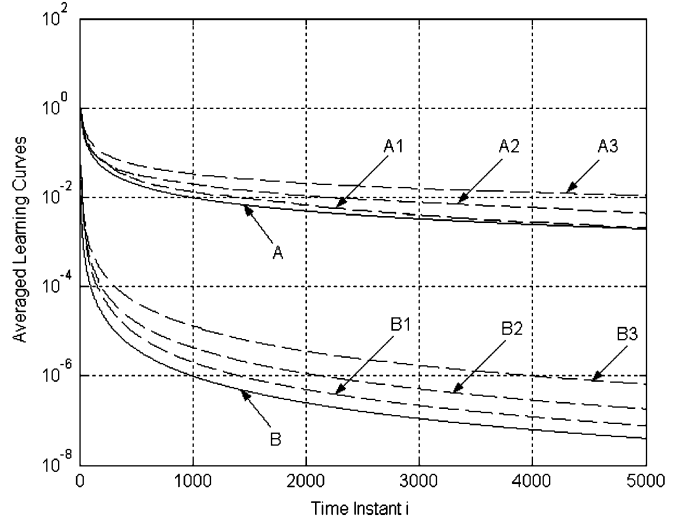

Fig. 13.   Averaged Learning Curve of robust PAST. Curve A: Theoretical curve of $f_{\mathrm{proj}}(i)$ of PAST under Gaussian noise. Curve A1, A2, A3: Experimental learning curves $f_{\mathrm{proj}}(i)$ of robust PAST under impulsive noise with occurrence probabilities of 3%, 5%, and 10%, respectively. Curve B: Theoretical curve of $f_{\mathrm{ortho}}(i)$ of PAST under Gaussian noise. Curve B1, B2, B3: Experimental curves of $f_{\mathrm{ortho}}(i)$ of robust PAST under impulsive noise with occurrence probabilities of 3%, 5%, and 10%, respectively.

Section IV. Fig. 13 shows the curves for $f_{\mathrm{ortho}}(i)$ and $f_{\mathrm{proj}}(i)$ (the corresponding measures obtained from $r_{\mathrm{PAST}}/(1 - p_{\mathrm{imp}})$ mentioned in Section IV-A for the occurrence probabilities considered are similar because the plot is in log scale) and those for the robust PAST algorithm under impulsive noise with 3%, 5%, and 10% occurrence probabilities. The simulation setting is similar to [43], where the input data vector $\underline{x}(i)$ is a stationary Gaussian stochastic process with correlation matrix $\boldsymbol{C}_{xx} = \mathrm{diag}(10, 9, 8, 3, 2, 1)$. The first three signals are the signals of interest, and the corresponding signal subspace is tracked by the robust PAST algorithm. The forgetting factor β is set to be one. The initial value of $\boldsymbol{W}(0)$ is chosen to be the leading submatrix of the identity matrix. The result is averaged over 100 Monte Carlo trials. Though the convergence rates of the robust PAST algorithm are slowed down by the sporadic impulsive noise, the trend of convergence does not seems to be disturbed. It is because, accordingly to the convergence analysis in Section IV, the robust PAST algorithm will converge to $\boldsymbol{C}_{\rho\_xx}$. This substantiates the convergence analysis presented in Section IV. However, because of the impulsive noise, the rate is now much lower than $f_{\mathrm{ortho}}(i)$ and $f_{\mathrm{proj}}(i)$. In fact, the higher the occurrence probability of impulsive noise, the lower will be the convergence rate.

VI. CONCLUSION

A new robust PAST algorithm for robust subspace tracking in impulsive noise environment is presented. It is based on a new robust autocorrelation matrix estimate, called the M-estimator, which is derived from the maximum likelihood estimation of a multivariate Gaussian process under CG noise. A systematic method for incorporating this new estimator into the PAST algorithm is developed. Moreover, a new restoring mechanism is proposed to combat the hostile effect of long burst of impulses. The convergence of the robust PAST algorithm is

analyzed using the ODE method. Both theoretical and simulation results show that the robust PAST algorithm offers improved robustness over the conventional PAST algorithm. On the other hand, the performance of the new algorithm in nominal Gaussian noise is very close to that of the PAST algorithm.

APPENDIX A
ML-ESTIMATION IN CONTAMINATED GAUSSIAN NOISE

The probability density function (pdf) of a multivariable normal distribution is

$$f_{(\boldsymbol{m},\boldsymbol{R})}(\underline{x}) = \frac{1}{((2\pi)^m |\boldsymbol{R}|)^{1/2}}$$
$$\times \exp\left[-\frac{1}{2}(\underline{x}-\underline{m})^T \boldsymbol{R}^{-1}(\underline{x}-\underline{m})\right] \quad \text{(A-1)}$$

where $\underline{x} = [x_1, x_2, \ldots, x_m]^T$, $\underline{m}$ is the mean vector, and $\boldsymbol{R}$ is the covariance matrix. The pdf of the contaminated Gaussian model is given by the weighted sum of two (or in general more) Gaussian distributions with means and covariances given by $(\underline{m}_0, \boldsymbol{R}_0)$, and $(\underline{m}_1, \boldsymbol{R}_1)$ as follows:

$$f_{CG}(\underline{x}) = p_0 f_{(\underline{m}_0, \boldsymbol{R}_0)}(\underline{x}) + (1-p_0) f_{(\underline{m}_1, \boldsymbol{R}_1)}(\underline{x})$$
$$= \alpha_0 \exp\left[-\frac{1}{2}(\underline{x}-\underline{m}_0)^T \boldsymbol{R}_0^{-1}(\underline{x}-\underline{m}_0)\right]$$
$$+ \alpha_1 \exp\left[-\frac{1}{2}(\underline{x}-\underline{m}_1)^T \boldsymbol{R}_1^{-1}(\underline{x}-\underline{m}_1)\right]$$
$$\text{(A-2)}$$

where $\alpha_0 = (p_0)/(((2\pi)^m |\boldsymbol{R}_0|)^{1/2})$ and $\alpha_1 = (1-p_0)/(((2\pi)^m |\boldsymbol{R}_1|)^{1/2}) = (p_1)/(((2\pi)^m |\boldsymbol{R}_1|)^{1/2})$. In other words, the random variable x is generated from the Gaussian distributions $f_{(\underline{m}_0, \boldsymbol{R}_0)}(\underline{x})$ and $f_{(\underline{m}_1, \boldsymbol{R}_1)}(\underline{x})$ with a probability of p_0 and $p_1 = 1 - p_0$, respectively. In the CG model, the additive impulsive noise is modeled by $f_{(\underline{m}_1, \boldsymbol{R}_1)}(\underline{x})$. $\boldsymbol{R}_1$ is usually much larger than $\boldsymbol{R}_0$ to emulate the impulsive nature of the impulsive noise, while p_1 represents the probability of occurrence of the impulsive noise. Given a set of observations $\{\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n\}$, our goal is to estimate the parameters of the process $f_{(\underline{m}_0, \boldsymbol{R}_0)}(\underline{x})$, i.e., $(\underline{m}_0, \boldsymbol{R}_0)$. First of all, note that the probability of observing these n observations is

$$l = \prod_{i=1}^n f_{CG}(\underline{x}_i)$$
$$= \prod_{i=1}^n \left[\alpha_0 \exp\left[-\frac{1}{2}(\underline{x}_i-\underline{m}_0)^T \boldsymbol{R}_0^{-1}(\underline{x}_i-\underline{m}_0)\right]\right.$$
$$\left. + \alpha_1 \exp\left[-\frac{1}{2}(\underline{x}_i-\underline{m}_1)^T \boldsymbol{R}_1^{-1}(\underline{x}_i-\underline{m}_1)\right]\right] \quad \text{(A-3)}$$

which is called the likelihood function of the CG model. The principle of ML estimation is to choose the unknown parameter θ (i.e., $(\underline{m}_0, \boldsymbol{R}_0)$ in our case) for which l is maximized. If l is a differentiable function of θ, a necessary condition for l to have a maximum (not at the boundary) is $(\partial l/\partial \theta) = 0$. Note l also depends on $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$. The solution of (A-3), which depends on $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$, is called the *maximum likelihood (ML) estimate*. We may replace the condition $(\partial l/\partial \theta) = 0$ by $(\partial \ln(l)/\partial \theta) = 0$, because $\ln(l)$, the log-likelihood function, is a monotonic increasing function of l. To estimate $\underline{m}_0$, we take the logarithm of (A-3) and obtain

$$\ln l = \sum_{i=1}^n \ln \left[\alpha_0 \exp\left[-\frac{1}{2}(\underline{x}_i-\underline{m}_0)^T \boldsymbol{R}_0^{-1}(\underline{x}_i-\underline{m}_0)\right]\right.$$
$$\left. + \alpha_1 \exp\left[-\frac{1}{2}(\underline{x}_i-\underline{m}_1)^T \boldsymbol{R}_1^{-1}(\underline{x}_i-\underline{m}_1)\right]\right]. \quad \text{(A-4)}$$

Taking the partial derivatives of $\ln(l)$ with respect to $\underline{m}_0$ yields the equation shown at the bottom of the page. Here, the derivative of a scalar function f with respect to an $(N \times M)$ matrix $\boldsymbol{X}$ of independent variable x_{ij} is an $(N \times M)$ matrix with the (i, j) entry given by $\partial f/\partial x_{ij}$. Setting $(\partial \ln l/\partial \underline{m}_0) = \underline{0}$ and noting that $\boldsymbol{R}_0$ is nonsingular, one gets

$$\hat{\underline{m}}_0 = \frac{\sum_{i=1}^n w_m(\underline{x}_i, \hat{\underline{m}}_0, \underline{m}_1, \boldsymbol{R}_0, \boldsymbol{R}_1) \cdot \underline{x}_i}{\sum_{i=1}^n w_m(\underline{x}_i, \hat{\underline{m}}_0, \underline{m}_1, \boldsymbol{R}_0, \boldsymbol{R}_1)}$$
$$= \sum_{i=1}^n \rho_m(\underline{x}_i, \hat{\underline{m}}_0, \underline{m}_1, \boldsymbol{R}_0, \boldsymbol{R}_1) \cdot \underline{x}_i \quad \text{(A-5)}$$

where

$$w_m(\underline{x}_i, \underline{m}_0, \underline{m}_1, \boldsymbol{R}_0, \boldsymbol{R}_1)$$
$$= \left\{1 + (\alpha_1/\alpha_0) \exp\left[\frac{1}{2}(\underline{x}_i-\underline{m}_0)^T \boldsymbol{R}_0^{-1}(\underline{x}_i-\underline{m}_0)\right.\right.$$
$$\left.\left. - \frac{1}{2}(\underline{x}_i-\underline{m}_1)^T \boldsymbol{R}_1^{-1}(\underline{x}_i-\underline{m}_1)\right]\right\}^{-1},$$

and

$$\rho_m(\underline{x}_i, \underline{m}_0, \underline{m}_1, \boldsymbol{R}_0, \boldsymbol{R}_1)$$
$$= \frac{w_m(\underline{x}_i, \underline{m}_0, \underline{m}_1, \boldsymbol{R}_0, \boldsymbol{R}_1)}{\sum_{i=1}^n w_m(\underline{x}_i, \underline{m}_0, \underline{m}_1, \boldsymbol{R}_0, \boldsymbol{R}_1)}$$
.

This is a nonlinear equation in $\hat{\underline{m}}_0$. Also, note that the ML estimate of $\hat{\underline{m}}_0$ is a weighted sum of the observed samples. For estimating $\boldsymbol{R}_0$, we again take the partial derivatives of $\ln(l)$ with respect to $\boldsymbol{R}_0^{-1}$ and get (A-6) shown at the bottom of the page.

$$\frac{\partial \ln l}{\partial \underline{m}_0} = -\sum_{i=1}^n \frac{\alpha_0 \boldsymbol{R}_0^{-1}(\underline{x}_i-\underline{m}_0) \exp\left[-\frac{1}{2}(\underline{x}_i-\underline{m}_0)^T \boldsymbol{R}_0^{-1}(\underline{x}_i-\underline{m}_0)\right]}{\alpha_0 \exp\left[-\frac{1}{2}(\underline{x}_i-\underline{m}_0)^T \boldsymbol{R}_0^{-1}(\underline{x}_i-\underline{m}_0)\right] + \alpha_1 \exp\left[-\frac{1}{2}(\underline{x}_i-\underline{m}_1)^T \boldsymbol{R}_1^{-1}(\underline{x}_i-\underline{m}_1)\right]}.$$

$$\frac{\partial \ln l}{\partial \boldsymbol{R}_0^{-1}} = \sum_{i=1}^n \frac{(\alpha_0/2)\left[-(\underline{x}_i-\underline{m}_0)(\underline{x}_i-\underline{m}_0)^T + \boldsymbol{R}_0^T\right] \exp\left[-\frac{1}{2}(\underline{x}_i-\underline{m}_0)^T \boldsymbol{R}_0^{-1}(\underline{x}_i-\underline{m}_0)\right]}{\alpha_0 \exp\left[-\frac{1}{2}(\underline{x}_i-\underline{m}_0)^T \boldsymbol{R}_0^{-1}(\underline{x}_i-\underline{m}_0)\right] + \alpha_1 \exp\left[-\frac{1}{2}(\underline{x}_i-\underline{m}_1)^T \boldsymbol{R}_1^{-1}(\underline{x}_i-\underline{m}_1)\right]}$$
$$= (1/2)\sum_{i=1}^n w_m(\underline{x}_i, \underline{m}_0, \underline{m}_1, \boldsymbol{R}_0, \boldsymbol{R}_1)\left[-(\underline{x}_i-\underline{m}_0)(\underline{x}_i-\underline{m}_0)^T + \boldsymbol{R}_0^T\right]. \quad \text{(A-6)}$$

Setting the derivative to zero and noting $\boldsymbol{R}_0$ is symmetric, we have

$$\hat{\boldsymbol{R}}_0^T = \hat{\boldsymbol{R}}_0$$
$$= \sum_{i=1}^{n} \rho_m(\underline{x}_i, \underline{m}_0, \underline{m}_1, \hat{\boldsymbol{R}}_0, \boldsymbol{R}_1) \cdot (\underline{x}_i - \underline{m}_0)(\underline{x}_i - \underline{m}_0)^T.$$
(A-7)

Again it can be seen that this is a nonlinear equation in $\hat{\boldsymbol{R}}_0$ and the ML estimate $\hat{\boldsymbol{R}}_0$ is a weighted sum of the estimates $(\underline{x}_i - \underline{m}_0)(\underline{x}_i - \underline{m}_0)^T$. Careful examination review that the weight $w_m(\underline{x}_i, \underline{m}_0, \underline{m}_1, \hat{\boldsymbol{R}}_0, \boldsymbol{R}_1)$, which depends on $\underline{x}_i, \underline{m}_0, \underline{m}_1, \hat{\boldsymbol{R}}_0$ and $\boldsymbol{R}_1$, gets smaller and smaller as the magnitude of $\|\underline{x}_i - \underline{m}_0\|$ increases. This is reasonable as extra-ordinary large value of $\|\underline{x}_i - \underline{m}_0\|$ indicates that the observation become more and more unreliable and the weight should decrease accordingly to de-emphasis their effects on the estimates. For m = 1, (A-5) and (A-7) reduces to the ML estimates of the mean $\hat{\mu}_0$ and $\hat{\sigma}_0^2$ variance of a scalar process as follows:

$$\hat{\mu}_0 = \frac{\sum_{i=1}^{n} w(x_i, \mu_0, \mu_1, \sigma_0, \sigma_1) \cdot x_i}{\sum_{i=1}^{n} w(x_i, \mu_0, \mu_1, \sigma_0, \sigma_1)} \quad \text{and}$$
$$\hat{\sigma}_0^2 = \frac{\sum_{i=1}^{n} w(x_i, \mu_0, \mu_1, \sigma_0, \sigma_1)(x_i - \mu_0)^2}{\sum_{i=1}^{n} w(x_i, \mu_0, \mu_1, \sigma_0, \sigma_1)} \quad \text{(A-8)}$$

where $w(x_i, \mu_0, \mu_1, \sigma_0, \sigma_1) = \{1 + (\alpha_1/\alpha_0)\exp[((x_i - \mu_0/\sqrt{2}\sigma_0))^2 - ((x_i - \mu_1/\sqrt{2}\sigma_1)^2)]\}^{-1}$. Further, if the distribution is Gaussian, i.e., $w_m(\cdot) = 1$, then (A-5) and (A-7) reduce to the familiar estimates:

$$\underline{m}_0 = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i \quad \text{and} \quad \hat{\boldsymbol{R}}_0 = \frac{1}{n}\sum_{i=1}^{n} (\underline{x}_i - \underline{m}_0)(\underline{x}_i - \underline{m}_0)^T$$
(A-9)

for multivariate Gaussian distributions. In robust statistics, $\rho(\cdot)$ is chosen as a fixed function so that the sensitivity of the estimate to variation of the nominal pdf is minimized. This is the basic motivation of our robust correlation matrix introduced in Section II. The details of choosing $\rho(\cdot)$ is explained in Appendix B. For complex Gaussian process with pdf: $f_{(\underline{m},\boldsymbol{R})}(\underline{x}) = (1/\pi^m)|\boldsymbol{R}| \exp[-(\underline{x} - \underline{m})^H \boldsymbol{R}^{-1}(\underline{x} - \underline{m})]$, $w_m(\underline{x}_i, \underline{m}_0, \underline{m}_1, \boldsymbol{R}_0, \boldsymbol{R}_1)$ in (A-5) and (A-6) can be shown to be $\{1 + (p_0/p_1)(|\boldsymbol{R}_1|/|\boldsymbol{R}_0|) \exp[(x_i - \underline{m}_0)^T \boldsymbol{R}_0^{-1}(\underline{x}_i - \underline{m}_0) - (x_i - \underline{m}_1)^T \boldsymbol{R}_1^{-1}(\underline{x}_i - \underline{m}_1)]\}^{-1}$ and the matrix transpose ^{T} in (A-7) will be replaced by the Hermitian transpose ^{H}.

APPENDIX B
MULTIVARIATE RECURSIVE LEAST M-ESTIMATE (RLM) ALGORITHM

Let $\underline{e}_n$ be the estimation error in fitting the observations $\underline{y}_n, n = 1, \ldots, i$, by a model $\underline{y}(\underline{x}_n, \underline{a})$ with input $\underline{x}_n$ and parameter vector $\underline{a}$, which is to be determined. The log-likelihood function is then

$$\ln l = \sum_{n=1}^{i} \ln p_e(\underline{y}_n - \underline{y}(\underline{x}_n, \underline{a})).$$
(B-1)

For simplicity, we have assumed that $\underline{e}_n$ are independent. Maximizing the log-likelihood function is equivalent to the minimization of $-\ln l = \sum_{n=1}^{i} -\ln p_e(\underline{y}_n - \underline{y}(\underline{x}_n, \underline{a}))$. Denote $\rho_{p_e}(\underline{e}_n) = -\ln p_e(\underline{y}_n - \underline{y}(\underline{x}_n, \underline{a}))$, the ML estimate of $\underline{a}$ is

$$\underline{a}_{ML} = \arg\min_{\underline{a}} \sum_{n=1}^{i} \rho_{p_e}(\underline{y}_n - \underline{y}(\underline{x}_n, \underline{a})).$$
(B-2)

If $\underline{e}$ is Gaussian distributed, (B-2) reduces to the conventional least squares estimation

$$\underline{a}_{LS} = \arg\min_{\underline{a}} \sum_{n=1}^{i} \|\underline{y}_n - \underline{y}(\underline{x}_n, \underline{a})\|_2^2.$$
(B-3)

On the other hand, if the residual error is modeled as a contaminated Gaussian distribution, because of the presence of additive impulsive noise, then one should minimize (B-2). Since impulsive noise is usually of short time duration and time varying, its statistics are rather difficult to estimate accurately. Instead of estimating these quantities in real-time, the basic idea of the robust statistic-based estimator is to choose $\rho(\cdot)$ as a fixed function such as the Cauchy or Lorentzian distribution: $\rho(e) = \log(1 + e^2/2)$ (see [40, pp. 700–702]). The solution to (B-2) is then referred to as the M-estimator, or Maximum likelihood-like estimator, of $\underline{a}$. Since $\underline{e}_n = \underline{y}_n - \underline{y}(\underline{x}_n, \underline{a})$ is a vector, let's assume that $\rho_{p_e}(\underline{e}_n)$ is equal to $\rho(\|\underline{e}_n\|_2^2)$, where $\rho(\sqrt{e})$ is an M-estimate function such as the Cauchy, Huber, or modified Huber function. For linear estimation, we have $\underline{y}(\underline{x}_n, \boldsymbol{W}) = \boldsymbol{W} \cdot \underline{x}_n$. Differentiating (B-2) with respect to $\underline{a}$ yields the following necessary condition for $\boldsymbol{W}$ as follows:

$$2\sum_{n=1}^{i} \lambda^{i-n} \cdot \rho'\left(\left\|\underline{y}_n - \boldsymbol{W}\underline{x}_n\right\|_2^2\right) \cdot (\boldsymbol{W}\underline{x}_n\underline{x}_n^T - \underline{y}_n\underline{x}_n^T) = \boldsymbol{0}$$
(B-4)

where $\rho'(\cdot)$ is the derivative of $\rho(\cdot)$, and λ is a forgetting factor introduced to enable tracking of time varying systems. In what follows, we shall assume that the observations and inputs are derived from a time series and we shall replace $\underline{y}_n$ and $\underline{x}_n$ by $\underline{y}(n)$ and $\underline{x}(n)$, and vice versa. This yields

$$\boldsymbol{W}(i)\boldsymbol{C}_{\rho\_xx}(i) = \boldsymbol{C}_{\rho\_yx}(i)$$
(B-5)

where $\boldsymbol{C}_{\rho\_xx}(i) = \sum_{n=1}^{i} \lambda^{i-n}\rho'(\|\underline{e}(n)\|_2^2) \cdot \underline{x}(n)\underline{x}^T(n)$, and $\boldsymbol{C}_{\rho\_yx}(i) = \sum_{n=1}^{i} \lambda^{i-n}\rho'(\|\underline{e}(n)\|_2^2) \cdot \underline{y}(n)\underline{x}^T(n)$. Note, (B-5) is a nonlinear system of equation and, in principle, an iterative algorithm like some kind of gradient or Newton method is required to solve for the optimal M-estimator. Fortunately in recursive estimation, as mentioned in the introduction, a rough prior knowledge of $\boldsymbol{W}$ (say the subspace estimate in the current paper) is usually available from previous iterations. For the modified Huber function, $\rho'(e)$ is equal to one when $|e|$ is less than Γ, and zero otherwise. In other words, when the estimation error is abnormally large, the current observation will be discarded (similar to a hard decision). Other M-estimate functions might lead to slightly different weighting of the observation. Using the recurrent relations

$$\boldsymbol{C}_{\rho\_xx}(i) = \lambda\boldsymbol{C}_{\rho\_xx}(i-1) + \rho'\left(\|\underline{e}(i)\|_2^2\right) \cdot \underline{x}(i)\underline{x}^T(i),$$
$$\boldsymbol{C}_{\rho\_yx}(i) = \lambda\boldsymbol{C}_{\rho\_yx}(i-1) + \rho'\left(\|\underline{e}(i)\|_2^2\right) \cdot \underline{y}(i)\underline{x}^T(i)$$
(B-6)

and the matrix inversion lemma, $(\mathbf{A} + \mu \underline{x}\underline{y}^T)^{-1} = \mathbf{A}^{-1}(\mathbf{I} - (\mu\underline{x}\underline{y}^T\mathbf{A}^{-1})/(1 + \mu\underline{y}^T\mathbf{A}^{-1}\underline{x}))$, one gets the following robust multivariate RLM algorithm:

$$\mathbf{V}(i) = \lambda^{-1}\mathbf{V}(i-1)[\mathbf{I} - \rho' \cdot \underline{x}(i)\mathbf{K}(i)],$$
$$\mathbf{K}(i) = \frac{\underline{x}^T(i)\mathbf{V}(i-1)}{\lambda + \rho' \cdot \underline{x}^T(i)\mathbf{V}(i-1)\underline{x}(i)},$$
$$\boldsymbol{W}(i) = \boldsymbol{W}(i-1) + \rho'\underline{e}(i)\mathbf{K}(i) \qquad \text{(B-7)}$$

where we have write $\rho'(\|\underline{e}(i)\|_2^2)$ as ρ' to save notation. (B-7) is a generalization of the RLM algorithm in [46], [47], [49] for matrix parameters. Therefore, by assuming that $\rho'(\|\underline{e}(i)\|_2^2)$ depends weakly on $\boldsymbol{W}(i)$, we can solve (B-5) using (B-7), by treating $\rho'(\|\underline{e}(i)\|_2^2)$ as a constant. This amounts to the relaxation of (B-5). $\rho'(\|\underline{e}(i)\|_2^2)$ helps to determine whether the incoming signal vector is potentially corrupted by impulsive noise or not. In the proposed robust PAST algorithm, the resulting robust algorithm, after removing these corrupted measurements, is able to converge eventually to the subspace of the robust covariance matrix, though with a slower convergence speed. This is supported theoretically by the convergence analysis and simulation results in Sections IV and V, respectively. For the proposed robust PAST algorithm, $\rho'(\|\underline{e}\|_2^2)$ is chosen as $\rho'(\|\underline{e}\|_F - \mu_e)$. It is because the error vector might not be zero mean when the algorithm is still converging. For notation convenience, we also denote $\rho'(\cdot)$ by $\rho_r(\cdot)$. The matrix transpose ^{T} in (B-7) can be replaced by the Hermitian transpose ^{H} if the input is complex.
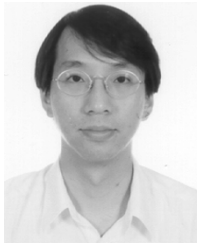
REFERENCES

[1] C. H. Bischof and G. M. Shroff, "On updating signal subspaces," *IEEE Trans. Signal Process.*, vol. 40, pp. 96–105, 1992.

[2] K. L. Blackard, "Measurements and models of radio frequency impulsive noise for indoor wireless communication," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 7, pp. 991–1001, Sep. 1993.

[3] J. R. Bunch, C. P. Nielsen, and D. Sorenson, "Rank-one modification of the symmetric eigenproblem," *Numerische Mathematik*, vol. 31, pp. 31–48, 1978.

[4] B. Champagne and Q. G. Liu, "Plane rotation-based EVD updating schemes for efficient subspace tracking," *IEEE Trans. Signal Process.*, vol. 46, no. 7, pp. 1886–1900, Jul. 1998.

[5] A. Chkeif *et al.*, "Spatial-temporal blind adaptive multiuser detection," *IEEE Trans. Commun.*, vol. 48, no. 5, pp. 729–732, May 2000.

[6] S. C. Chan and Y. X. Zou, "Convergence analysis of the recursive least M-estimate adaptive filtering algorithm for impulse noise suppression," in *Proc. 14th International Conf. on Digital Signal Processing (DSP2002)*, vol. 2, Santorini, Greece, Jul. 1–3, 2002, pp. 663–666.

[7] P. Common and G. H. Golub, "Tracking a few extreme singular values and vectors in signal processing," in *Proc. IEEE*, Aug. 1990, pp. 1327–1343.

[8] C. E. Davila, "Efficient, high performance, subspace tracking for time-domain data," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3307–3315, Dec. 2000.

[9] R. D. DeGroat, "Noniterative subspace tracking," *IEEE Trans. Signal Process.*, vol. 40, pp. 571–577, 1992.

[10] G. H. Golub and C. F. V. Loan, *Matrix Computation*, 3rd ed. Baltimore, MD: The John Hopkins University Press, 1996.

[11] F. Gustafsson, *Adaptive Filtering and Change Detection*. New York: Wiley, 2000.

[12] T. Gustafsson, "Instrumental variable subspace tracking using projection approximation," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 669–681, Mar. 1998.

[13] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley, 1986.

[14] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall, 1991.

[15] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.

[16] I. Karasalo, "Estimating the covariance matrix by signal subspace averaging," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, pp. 8–12, 1986.

[17] J. Karhunen, "Adaptive algorithms for estimating eigenvectors of correlation type matrices," in *Proc. IEEE ICASSP'84*, San Diego, CA, Mar. 1984, pp. 14.6.1–14.6.4.

[18] S. R. Kim and A. Efron, "Adaptive robust impulsive noise filtering," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 1855–1866, Aug. 1995.

[19] S. Kiole, "Adaptive threshold nonlinear algorithm for adaptive filters with robustness against impulsive noise," in *Proc. IEEE ICASSP'96*, Atlanta, GA, Nov. 1992, pp. 1644–1647.

[20] R. J. Kozick and B. M. Sadler, "Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3520–3535, Dec. 2000.

[21] H. J. Kushner and D. S. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Cambridge, MA: MIT Press, 1978.

[22] X. Li and H. H. Fan, "Blind channel identification: Subspace tracking method without rank estimation," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2372–2382, Oct. 2001.

[23] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.

[24] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, 1995.

[25] W. K. Pratt, *Digital Image Processing*, 2nd ed. New York: Wiley, 1991.

[26] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameter via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[27] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison Wesley, 1991.

[28] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectrum Estimation Workshop*, 1979, pp. 243–258.

[29] P. Spasojevic and X. Wang, "Improved robust multiuser detection in non-Gaussian channels," *IEEE Signal Process. Lett.*, vol. 8, no. 3, pp. 83–86, Mar. 2000.

[30] G. W. Stewart, "An updating algorithm for subspace tracking," *IEEE Trans. Signal Process.*, vol. 40, pp. 1535–1541, 1991.

[31] P. Stoica and T. Soderstrom, "Statistical analysis of MUSIC and subspace rotation estimate of sinusoidal frequencies," *IEEE Trans. Signal Process.*, vol. 39, no. 8, pp. 1836–1847, Aug. 1991.

[32] P. Strobach, "Bi-iteration SVD subspace tracking algorithms," *IEEE Trans. Signal Process.*, vol. 45, no. 5, pp. 1222–1240, May 1997.

[33] A. L. Swindlehurst, B. Ottersten, R. Roy, and T. Kailath, "Multiple invariance ESPRIT," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 867–881, Apr. 1992.

[34] S. Visuri, H. Oja, and V. Koivunen, "Blind channel identification using robust subspace estimation," in *Proc. IEEE 11th IEEE Workshop on Statistical Signal Processing, Singapore*, Aug. 6–8, 2001, pp. 281–284.

[35] X. Wang and H. V. Poor, "Robust multi-user detection in non-Gaussian channels," *IEEE Trans. Signal Process.*, vol. 47, no. 2, pp. 289–305, Feb. 1999.

[36] Y. Wen, S. C. Chan, and K. L. Ho, "Robust subspace tracking in impulsive noise," in *Proc. IEEE Int. Conf. Communications (ICC2001)*, vol. 3, Helsinki, Finland, June 11–14, 2001, pp. 892–896.

[37] ——, "A robust subspace tracking algorithm for subspace-based blind multiuser detection in impulsive noise," in *Proc. 14th Int. Conf. Digital Signal Processing (DSP2002)*, vol. 2, Santorini, Greece, 2002, pp. 1289–1292.

[38] ——, "Robust subspace tracking based blind channel identification in impulsive noise environment," in *Proc. European Signal Processing Conf. (EUSIPCO2002)*, vol. II, Toulouse, France, Sep. 3–6, 2002, pp. 641–644.

[39] ——, "Robust recursive bi-iteration singular value decomposition (svd) for subspace tracking and adaptive filtering," in *Proc. ISCAS'2003*, Bangkok, Thailand, May 2003.

[40] W. H. William *et al.*, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge University Press, 1997.

[41] X. Yang, T. K. Sarkar, and E. Arvas, "A survey of conjugate gradient matrix," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 1550–1556, 1989.

[42] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[43] ——, "Convergence analysis of the subspace tracking algorithms PAST and PASTd," in *Proc. IEEE ICASSP'96*, vol. 3, May 1996, pp. 1759–1762.

[44] B. Yang and F. Gersemsky, "Asymptotic distribution of recursive subspace estimators," in *Proc. IEEE ICASSP'96*, vol. 5, Atlanta, GA, May 1996, pp. 2856–2859.

[45] B. Yang, "Asymptotic convergence analysis of projection approximation subspace tracking algorithms," *Signal Process.* (Special Issue on Subspace Methods for Detection and Estimation, Part I: Array Signal Processing and Subspace Computations), vol. 50, pp. 123–136, Apr. 1996.

[46] Y. X. Zou, S. C. Chan, and T. S. Ng, "A robust M-estimate adaptive filter for impulsive noise suppression," in *Proc. IEEE ICASSP'99*, vol. 4, Phoenix, AZ, Mar. 1999, pp. 1765–1768.

[47] ——, "A recursive M-estimate algorithm for robust adaptive filtering in impulse noise," *IEEE Signal Process. Lett.*, vol. 7, no. 11, pp. 324–326, Nov. 2000.

[48] ——, "Fast least mean M-estimate algorithms for robust adaptive filtering in impulse noise," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 47, no. 12, pp. 1564–1569, Dec. 2000.

[49] ——, "Robust M-estimate adaptive filtering," in *Inst. Elec. Eng. Signal Process.*, vol. 148, Aug. 2001, pp. 289–294.

Yu Wen received the B.S. degree in electrical engineering from Beijing Institute of Technology, Beijing, China, in 1994 and the M.S. degree in electrical engineering from The University of Electronic Science and Technology of China, Chengdu, China, in 1997. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong.

His research interests are in the area of statistical signal processing, wireless communications, MIMO, OFDM, and 4G communications systems.

Shing-Chow Chan (S'87–M'92) received the B.Sc. (Eng.) and Ph.D. degrees from the University of Hong Kong in 1986 and 1992, respectively.

In 1990, he joined City Polytechnic of Hong Kong as an Assistant Lecturer and later as a University Lecturer. Since 1994, he has been with the Department of Electrical and Electronic Engineering, the University of Hong Kong, where he is currently an Associate Professor. He was a Visiting Researcher at Microsoft Corporation, Redmond, WA, in 1998 and at Microsoft China in 1999. His research interests include fast transform algorithms, filter design and realization, multirate signal processing, communications signal processing, and image-based rendering.

Dr. Chan is currently a member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society. He was Chairman of the IEEE Hong Kong Chapter of Signal Processing from 2000 to 2002.

Ka-Leung Ho (SM'04) received the B.Sc.(Eng.) and the M.Phil. degrees in electrical engineering from the University of Hong Kong in 1971 and 1973 respectively, and the Ph.D. degree from the University of London, U.K., in 1977.

In 1984, he joined the Department of Electrical and Electronic Engineering at the University of Hong Kong. His current research interests include signal processing and communications systems.

Dr. Ho is a Chartered Engineer of the Engineering Council, U.K., a Fellow of the Institution of Electrical Engineers (IEE), and Member of HKIE.