Robust Space-Time Adaptive Processing Using Projection Statistics

André P. des Rosiers[1], Gregory N. Schoenig[2], Lamine Mili[3]

1: Adaptive Processing Section, Radar Division
United States Naval Research Laboratory
4555 Overlook Ave. SW, Washington, DC 20375 USA
E-mail: apd@nrl.navy.mil

2: Advanced Research and Engineering Solutions
Science Applications International Corporation
4501 Daly, Dr., Suite 400, Chantilly, VA 20151 USA
E-mail: gregory.n.schoenig@saic.com

3: Alexandria Research Institute
Virginia Polytechnic Institute and State University
206 N. Washington St., Suite 400, Alexandria, VA 22314 USA
E-mail: lmili@vt.edu

***Abstract*—In this paper, *projection statistics* (PS) are applied to detect and mitigate outliers in adaptive processing algorithms. Outliers present in training data result in slow adaptive processor convergence. In radar, these outliers may arise from clutter discretes, desired targets, or jammers. PS provide a computationally tractable technique for identifying and mitigating outlier data samples prior to adaptive processing. We demonstrate that well known processing methods, such as sample matrix inversion and its variants can be made robust to these outlier impairments by incorporating PS into the algorithm formulation.**

***Index Terms*—Adaptive processing, array signal processing, space-time signal processing, space-time adaptive processing, robust statistics.**

I. Introduction

Linear adaptive processing methodologies use the input data stream to estimate the statistics of the system noise and external interference. In the radar signal processing context, accurately estimating parameters of the background noise is necessary for developing the signal-to-interference-plus-noise ratio (SINR) used in target detection. During the adaptation process, a window of data samples surrounding the sample under test, is used to estimate the appropriate weight vector for optimizing the detection of a target that may be present and canceling the jamming or clutter interference. The number of statistically independent training samples required for the performance of the adaptive processor to be sufficiently close to the optimum (nominally within 3 dB) is used as a measure of algorithm convergence [1]. For example, the number of samples needed for the sample matrix inversion (SMI) algorithm is roughly twice the number of system degrees of freedom (DOF). Fast convergence in training samples is a desirable property of adaptive processors, since the noise in many cases is non-stationary due to terrain inhomogeneity, resulting in noise coherence times on the order of tens of samples long. Another form of interference is impulsive noise; characteristic of clutter discretes, smart jammers, and other in-band radar and communication systems. These forms of interference degrade severely the convergence performance of SMI and other forms of adaptive processing. Thus, the presence of impulsive noise often requires adaptation intervals longer than the interference noise coherence time. If samples due to impulsive noise are modeled as statistical outliers, the powerful mathematical tools of robust statistics can be utilized [2]. Previous authors have used robust methods in a Gram Schmidt form of the generalized sidelobe canceler and obtained an adaptive processor with a convergence rate superior to that of SMI [3]. This method, although elegant in its formulation, is difficult to analyze and requires significant computational load.

This work was supported by the Office of Naval Research

In this paper, we introduce to this radar problem a robust statistical method, known as *projection statistics* (PS), a body of work currently applied to outlier detection in power system state estimation [4]. We show that PS provide a basis for analytical development, and offer a computational complexity that scales linearly with the dimensionality of the adaptive system.

II. Projection Statistics

In this section we denote $A_x = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_P\}$ as a set of P data points in $N-$dimensional space. The region of N-dimensional space dominated by the bulk of the P data points is referred to as the *point cloud*. To obtain the PS, this point cloud is projected onto specific one-dimensional unit vectors and a metric which is a function of the maximum of these projections is assigned to each point. The set of these metrics constitute the PS and measure the distance from a given point to the worst case one-dimensional projection of the bulk of the data set. These metrics enable outlier detection by providing a measure of the degree of association of each point to the bulk. Not only do the PS offer good indication of outliers, they are rapid to compute even for high dimensional problems.

The conventional measure for determining the distance from point, $\mathbf{x}_i$, to the bulk of a point cloud is the Mahalanobis distance (MD)

$$\mathrm{MD}_i = \sqrt{(\mathbf{x}_i - \mathrm{M}(A_x))^T \widehat{\mathbf{R}}_x^{-1} (\mathbf{x}_i - \mathrm{M}(A_x))} \qquad (1)$$

where the location estimate, $\mathrm{M}()$, is the sample average of its argument and $\widehat{\mathbf{R}}_x$ is the sample covariance matrix estimate of the set A_x. This measure can be used to identify outliers, however it is non robust when outliers appear in clusters; the so-called *masking effect* [4].

It was observed in [5] that the MD has an alternate formulation[1], that is,

$$\mathrm{MD}_i = \max_{\|\mathbf{v}\|=1} \frac{\left|\mathbf{x}_i^T\mathbf{v} - \mathrm{M}\left(\mathbf{x}_1^T\mathbf{v}, \cdots, \mathbf{x}_P^T\mathbf{v}\right)\right|}{\sigma\left(\mathbf{x}_1^T\mathbf{v}, \cdots, \mathbf{x}_P^T\mathbf{v}\right)} \qquad (2)$$

where the scale estimate, $\sigma()$, is the sample standard deviation of the projection of the data points on the direction of $\mathbf{v}$. Equality in (2) is met when all directions $\mathbf{v}$ are searched. A robust alternative to MD is made by using the sample median, $\mathrm{med}()$, for $\mathrm{M}()$ and the median-absolute-deviation (MAD) for $\sigma()$ denoted

$$\mathrm{MAD}(\mathbf{v}) = 1.4826 \cdot \mathrm{med}_i \left|\mathbf{x}_i^T\mathbf{v} - \mathrm{med}_j\left(\mathbf{x}_j^T\mathbf{v}\right)\right|. \qquad (3)$$

Since it is not possible to investigate all directions $\mathbf{v}$, it was suggested that one should use only those directions that originate from the coordinate-wise median and pass through each data point. The resulting distances computed with the robust version of MD are called the projection statistics. Denoted PS_i, $i = 1, \cdots, P$, Algorithm 1 details the steps for computing the PS [6], [7].

Determining which samples are outliers proceeds by comparing the PS_i with a threshold derived from assuming A_x is distributed as a multivariate Gaussian. When the sample points follow a N-variate Gaussian density the quadratic form, MD_i^2, follows approximately[2] a chi-squared distribution with N-DOF with probability $1-\alpha$ and is denoted $\mathrm{MD}_i^2 \sim \chi^2_{N,\alpha}$. Therefore, an outlier could be identified when

$$PS_i > \sqrt{\chi^2_{N,\alpha}} \qquad (4)$$

with α set to an appropriate level [4]. One should note, that real radar data often does not follow a Gaussian rule rendering this particular threshold inadequate for detecting outliers and will need to be suitably modified.

[1] In this paper superscript T and H denote the transposition and conjugate-transposition operator, respectively.

[2] Approximately since the actual density is a function of the eigenvalues of $\widehat{\mathbf{R}}_x$ which for correlated cases are not all equal.

Algorithm 1 Method for computing the projection statistics.

We let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_P]$ be a $N \times P$ matrix of P, N-dimensional data points. The n-th element of the p-th data point of matrix $\mathbf{X}$ is denoted x_{np}.

1. Calculate the coordinate-wise median of $\mathbf{X}$ denoted as vector, $\mathbf{m} = \{\mathrm{med}_p(x_{1p}), \cdots, \mathrm{med}_p(x_{Np})\}$, $p = 1, \cdots, P$.
2. Calculate the unit norm directions $\mathbf{v}_p = \mathbf{u}_p / \|\mathbf{u}_p\|$ where $\mathbf{u}_p = \mathbf{x}_p - \mathbf{m}$ and $\|\mathbf{u}_p\| = \sqrt{u_{1p}^2 + \cdots + u_{Np}^2}$ for all $p = 1, \cdots, P$.
3. For each $\mathbf{v}_i$, $i = 1, \cdots, P$, compute the standardized projections of $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_P\}$ onto $\mathbf{v}_i$ as follows:
   (a) Compute projection $\mathbf{z}_{\mathrm{i}} = \mathbf{X}^T\mathbf{v}_i$.
   (b) Compute the location estimate of the projection using the median over the elements of $\mathbf{z}_{\mathrm{i}}$: $m_i = \mathrm{med}(z_{1i}, \cdots, z_{Pi})$.
   (c) Compute the scale estimate of the projection: $\mathrm{MAD}_i = 1.4826 \cdot \mathrm{med}|\mathbf{z}_i - m_i|$.
   (d) Find the standardized projection vector: $\mathbf{s}_i = \frac{|\mathbf{z}_i - m_i|}{\mathrm{MAD}_i}$. The p-th element of $\mathbf{s}_i$ is denoted s_{ip}.
4. For each $p = 1, \cdots, P$ find the projection statistics using $PS_p = \max_i(s_{ip})$.

III. Adaptive Processing Methods

In this section we incorporate PS into two adaptive processing methods in order to improve their immunity to outlier impairments. PS will be used as a diagnostic tool to identify outlier samples thereby providing a mechanism for down-weighting their influence on the covariance matrix estimation of the data samples.

We let $\mathbf{X}$ be a $N \times P$ data matrix with independent and circularly symmetric[3] complex Gaussian distributed N-length random vectors with covariance matrix $\mathbf{R}_x$[4]. We partition $\mathbf{X} = \begin{bmatrix} \widetilde{\mathbf{x}} \\ \widetilde{\mathbf{X}} \end{bmatrix}$ where $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{X}}$ are the first row and $N-1$ remaining rows of $\mathbf{X}$, respectively. Using the generalized sidelobe canceler (GSC) formulation [8], we wish to detect a signal by estimating parameters of the interference and noise. This is formalized by the expression $\mathbf{y} = \widetilde{\mathbf{x}} - \mathbf{w}^H\widetilde{\mathbf{X}}$ where $\mathbf{y}$ is the GSC output and $\mathbf{w}$ are the linear weights used to estimate the noise on the main channel, $\widetilde{\mathbf{x}}$, from the data on the auxiliary channels, $\widetilde{\mathbf{X}}$.

We use a system with N DOF (denoted as channels) which corresponds to any combination of sensors, pulses, time slots, frequency bins, etc. For example, in space-time adaptive processing (STAP) N is equal to the number of antenna elements times the number of radar pulses being processed.

In STAP, the detection space is divided into angles and perceived Doppler shifts of the returned radar signals. Op-

[3] Independent real and imaginary parts.

[4] This is the true covariance defined as $\mathbf{R}_x = \mathrm{E}\left\{\mathbf{X}\mathbf{X}^H\right\}$ where $\mathrm{E}\{\}$ is the expectation operator.

erating in a two dimensional (2-D) detection space enables clutter cancellation in a 2-D angle/Doppler plane providing sensitive regions over which detection of low-power targets is possible. Such detection would not be possible in a classical radar system.

The developed weights, $\mathbf{w}$, are used to detect signal energy by matching to the transmitted waveform in the receiver and generating a maximum output SINR value. Optimum weights, $\mathbf{w}_{opt}$, are computed based on the well known minimum variance distortion-less response (MVDR) processor,

$$\mathbf{w}_{opt} = \frac{\mathbf{R}_x^{-1}\mathbf{s}}{\mathbf{s}^H\mathbf{R}_x^{-1}\mathbf{s}} \tag{5}$$

where the $\mathbf{s}$ vector is a space-time steering vector for which the STAP processor is tuned, and includes the influences from both the space and time steering vectors through the Kronecker product, $\mathbf{s} = \mathbf{s}_{time} \otimes \mathbf{s}_{space}$.

The classical linear approach to estimating $\mathbf{R}_x$ uses K data samples from N sources, where $N \leq K \leq P$. The maximum likelihood (ML) estimate for this $N \times N$ covariance matrix is the sample covariance matrix (SCM) denoted, $\widehat{\mathbf{R}}_x = \frac{1}{K}\mathbf{X}\mathbf{X}^H$. Adaptive processing using SMI uses the SCM for estimating $\mathbf{R}_x$ and requires roughly $K = 2N$ data samples per channel for performance within 3 dB of optimal SINR [9]. Performance degrades substantially as fewer samples are used due to ill conditioning of $\widehat{\mathbf{R}}_x$.

To obtain lower sample support, it was proven in [1] that a ML estimate $\mathbf{R}_x$ exists for $K < N$ provided the ambient noise floor is known. This method, know as fast maximum likelihood (FML), amounts to limiting the smallest eigenvalues of $\widehat{\mathbf{R}}_x$ to the noise floor, thereby properly conditioning the SCM matrix prior to inversion and use in (5).

IV. Results

We present in this section simulation results of two adaptive algorithms incorporating PS in the presence of impulsive noise. The intent of this controlled simulation study is to highlight the salient features PS provide in convergence rate improvements and offer topics for future research directions.

A. Preliminaries

The simulation configuration follows the approach in [1] and [10]. Using a GSC formulation, the desired steering vector, $\mathbf{s} = [1, 0, \cdots, 0]^T$, is an length N vector. We separate the N channels into one main channel and $N-1$ auxiliary channels. For our simulations the desired signal is present only in the main channel and therefore is not adaptively weighted and is given a unit weight factor. Auxiliary channels, however, are adaptively weighted and the weights are computed from training data that is statistically independent of the data sample under test.

The figure of merit for comparing algorithm performance is the convergence rate of the average output SINR given as,

$$\rho = \frac{1}{L \cdot \mathrm{SINR}_{\mathrm{opt}}} \sum_{l=1}^{L} \frac{\left|\mathbf{w}_l^H \mathbf{s}\right|^2}{\mathbf{w}_l^H \widehat{\mathbf{R}}_x \mathbf{w}_l} \tag{6}$$

and normalized by the optimal SINR obtained when the optimal linear weighting is used, $\mathrm{SINR}_{\mathrm{opt}} = \mathbf{s}^H\mathbf{R}_x^{-1}\mathbf{s}$. In (6), $\mathbf{w}_k$ is the random weight computed using a specific algorithm such as SMI or FML at the l-th realization of L algorithm runs.

Of the K data samples used for training, K_{out} of the samples contain outliers. The resulting covariance matrix for the overall training data set can be written as, $\mathbf{R}_x = \mathbf{R} + \sum_{k=1}^{K_{out}} \sigma_{k,out}^2 \mathbf{s}_{k,out}\mathbf{s}_{k,out}^H$ where $\mathbf{R}$ is the true covariance matrix containing no outliers and $\mathbf{s}_{k,out}$ and $\sigma_{k,out}^2$ are the k-th outlier steering vector and power, respectively. For our simulations we use $\mathbf{s}_{k,out} = [1, 0, \cdots, 0]^T$, that is, the outlier has the same steering vector as the desired signal. In addition, outliers are positioned in the first K_{out} samples of the training set, and model as zero mean complex Gaussian random variables with variance $\sigma_{k,out}^2$.

PS provides outlier diagnostics and must be unified with the adaptive algorithms. A simple approach would be to prune the flagged samples from the matrix $\mathbf{X}$ prior to developing the SCM. Alternatively, one could weight the samples providing a graceful attenuation of the flagged samples. We choose to use the latter. Using PS we develop a weighting matrix that is applied to either SMI or FML and we denote these adaptive algorithms as PS-SMI and PS-FML[5], respectively. We form a diagonal weighting matrix $\Omega = \mathrm{diag}(\omega_1, \omega_2, \cdots, \omega_K)$ by comparing each PS_i to the threshold from (4) and computing

$$\omega_i = \min\left(1, \frac{b}{PS_i^2}\right) \tag{7}$$

where $b = \chi^2_{N,\alpha}$ and $\alpha = 0.975$. With these PS derived weights in (7) we define

$$\mathrm{PS-SMI}: \quad \widehat{\mathbf{R}}_x = \frac{1}{S_\omega}\mathbf{X}\Omega\mathbf{X}^H \tag{8}$$

$$\mathrm{PS-FML}: \quad \widehat{\mathbf{R}}_x = \frac{1}{S_\omega}\mathbf{X}\Omega\mathbf{X}^H + \sigma_n\mathbf{I}_N \tag{9}$$

where σ_n is the background noise and is assumed to be the same for each channel, $\mathbf{I}_N$ is an $N \times N$ identity matrix, and $S_\omega = \sum_{k=1}^{K}\omega_k$.

Threshold b was defined in Section II assuming that samples $\mathbf{x}_i$ are distributed as *real* Gaussian distributed N-variate random variables. For a complex data matrix $\mathbf{X}_c$ which is $N \times P$, we form a real matrix $\mathbf{X}$ of size $2N \times P$ by partitioning the real and imaginary components as $\mathbf{X} = \begin{bmatrix} \mathrm{real}(\mathbf{X}_c) \\ \mathrm{imag}(\mathbf{X}_c) \end{bmatrix}$. Note that this has the effect

[5] We use a variant of FML called *diagonal loading* [11]. True FML as introduced in [1] adds the channel ambient noise power term to the low eigenvalues of $\widehat{\mathbf{R}}_x$.

of doubling the dimension of the space and DOF used in computing the PS, but does not change the DOF of the adaptive processing algorithms. One could also take the modulus of the complex data, however this reduces the outlier information content, potentially reducing the effectiveness of the PS algorithm. This issue needs further investigation. Lastly, variations in choice of b may be needed to accommodate non Gaussian data.

B. Performance Evaluation

We now demonstrate some representative simulations showing the advantage of using PS in outlier detection. In all the examples the number of channels is $N = 20$ and the number of simulation runs averaged for each convergence performance data point is $L = 100$. No jammers are included in these simulations in order to isolate the convergence improvement using PS. We note that since outlier detection is performed independently of spatial filtering, high jammer powers mask the outliers when the outlier powers are comparable to the jammer power. Further work is needed to incorporate PS directly into the adaptive filter algorithms in order to spatially filter these jammers. We are currently researching an method which iterates between PS and MVDR processing.

As a baseline, we plot in Figure 1 the normalized average SINR, ρ, versus K for SMI and FML without outlier mitigation when 2 and 10 outliers with 30 dB power are included in the training set. In Figure 1(a), until $K = 20$ is reached, SMI has flat performance, after which we observe that in 40 samples SMI achieves -14 dB SINR for $K_{out} = 2$, and -23 dB SINR for $K_{out} = 10$. Not surprisingly, due to the presence of the outliers, SMI fails to converge to the -3 dB SINR in $K = 40$ samples as predicted in [9]. FML also has slow performance in the presence of outliers although better than SMI. Typical FML convergence performance in the absence of outliers is twice the number of narrrowband barrage jammers [1]. In these simulations we do not include jammers. Thus, the slow convergence is only due to the influence of the outliers. In Figure 1(b), performance is in fact negative until $K = 10$. This indicates that positive FML convergence is exhibited after $K = K_{out}$.

Re-simulating the same scenarios using PS-SMI and PS-FML shows much improved performance. In Figure 1(a), for 2 outliers PS-SMI is close to -3 dB SINR after 40 training samples, and PS-FML converges to -3 dB SINR in $K = 7$ samples. However, in Figure 1(b), for 10 outliers PS-SMI converges to -7 dB SINR in $K = 40$ samples requiring $K = 60$ samples to converge to -3 dB. On the other hand, in Figure 1(b) PS-FML shows a convergence to -3 dB SINR in 22 samples demonstrating its superior performance to PS-SMI.

In Figure 2, we plot ρ versus K of SMI and PS-SMI as the outlier power is increased from -10 dB to 50 dB in 10 dB increments. We observe for SMI that as the outlier power is increased, the number of training samples needed for convergence is also increased. Using PS-SMI, convergence is rapid for all outlier power levels beyond $K = 20$. Increasing the number of outliers to 10 in Figure 3, reveals that the convergence improvement is a function of the outlier power level. PS-SMI does not provide an improvement in convergence until the outlier power level is above 10 dB. We also observe a SINR cross-over point at -20 dB SINR and $K = 28$ samples. At higher SINR, a larger outlier power provides for a more rapid convergence. This behavior is indicative of larger probability tails due to outliers. Large probability distribution tails in the training data increase the significance of the higher order moments which are important for convergence at higher SINR.

Figure 4 plots the convergence performance of ρ versus K for FML and PS-FML for 10 outliers. We clearly see the dramatic improvement in performance after $K = 20$ is reached. At low values of K, performance is negative until $K = K_{out} = 10$. A cross-over point is reached for the same reason as described for Figure 3.

V. Conclusions

In this paper, we investigate the use of projection statistics in order to mitigate impulsive interference in adaptive processors. These statistical methods hold great promise for robustness and computational efficiency of space-time adaptive algorithms.

In this paper, we demonstrate methods of incorporating the PS into the SMI and FML adaptive processing algorithms. By appropriate weighting of the data samples in the processor formulations, different degrees of outlier immunity can be obtained. We display simulation comparisons of these modified processors denoted PS-SMI and PS-FML and show analysis of convergence performance for these modified processing topologies.

References

[1] M. Steiner and K. Gerlach, "Fast converging adaptive processor for a structured covariance matrix," *IEEE Trans. Aerospace and Electronic Systems*, vol. 36, no. 4, pp. 1115–1126, Oct. 2000.

[2] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley, 1987.

[3] M. Picciolo and K. Gerlach, "Median cascaded canceller using reiterative processing," IEEE Radar Conference, May 2003, pp. 71–78.

[4] L. Mili, M. Cheniae, N. Vichare, and P. Rousseeuw, "Robust state estimation based on projection statistics," *IEEE Trans. on Power Systems*, vol. 11, no. 2, pp. 1118–1127, May 1996.

[5] D. Donoho, *Breakdown Properties of Multivariate Location Estimators*, Ph.D. thesis, Harvard University, 1982.

[6] P. J. Rousseeuw and B. C. Van Zomeren, *Directions in Robust Statistics, Part II*, chapter Robust Distances: Simulations and Cutoff Values, pp. 195–203, Springer-Verlag, 1991.

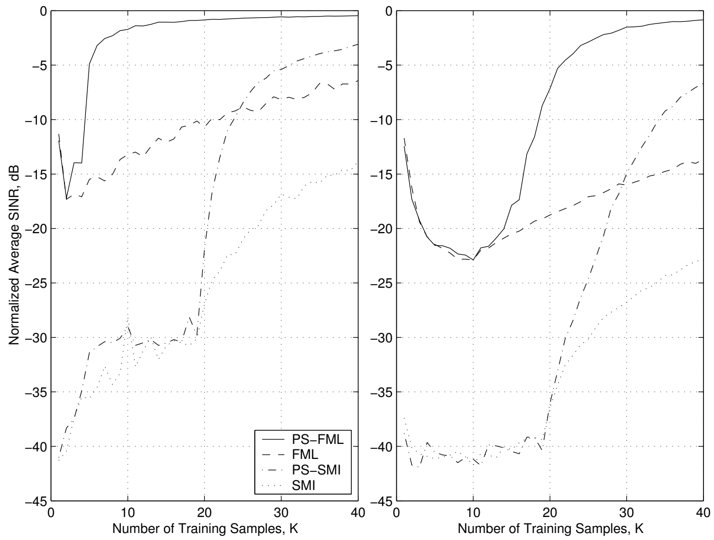
[7] Lamine Mili, *lecture notes on Robust Estimation and Filtering*, Virginia Polytechnic Institute and State University, 2003.

[8] Simon Haykin, *Adaptive Radar Detection and Estimation*, Wiley, 1992.

[9] I. Reed, D. Mallett, and L. Brennan, "Rapid convergence in adaptive arrays," *IEEE Trans. Aerospace and Electronic Systems*, vol. 10, no. 4, pp. 853–863, June 1974.

[10] K. Gerlach, "Outlier resistant adaptive matched filtering," *IEEE Trans. Aerospace and Electronic Systems*, vol. 38, no. 3, pp. 885–900, July 2002.

[11] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 397–401, April 1988.



(a) (b)

Figure 1. Normalized average SINR convergence for SMI, PS-SMI, FML, and PS-FML, outlier power = 30 dB, $N = 20$, (a) $K_{out} = 2$, (b) $K_{out} = 10$.

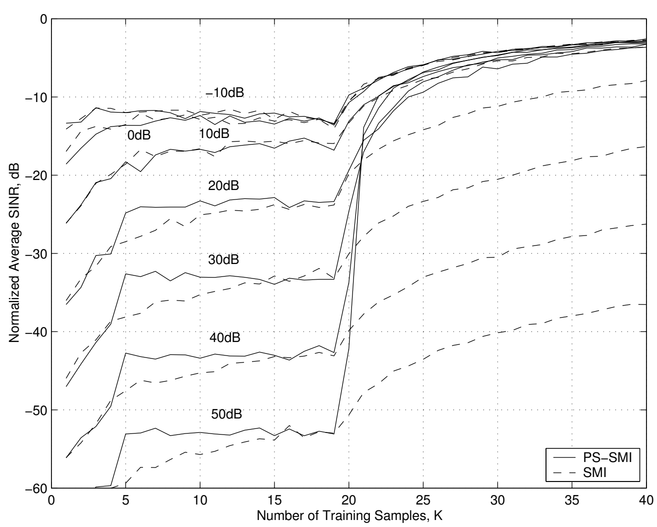


Figure 2. Normalized average SINR convergence for SMI and PS-SMI at different outlier powers, $N = 20$, $K_{out} = 2$.

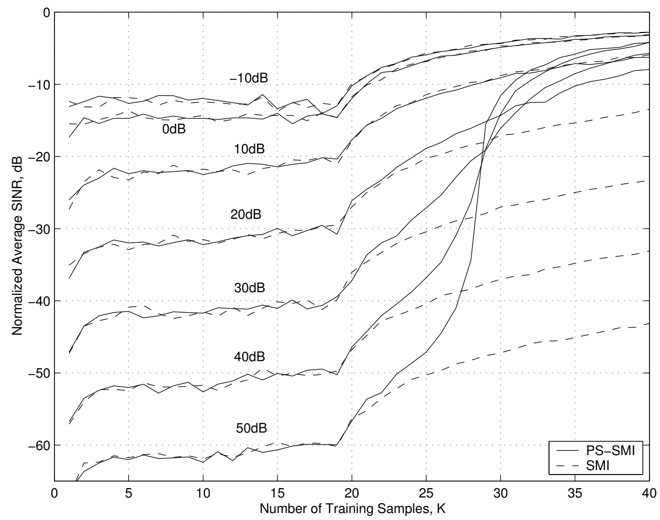


Figure 3. Normalized average SINR convergence for SMI and PS-SMI at different outlier powers, $N = 20$, $K_{out} = 10$.

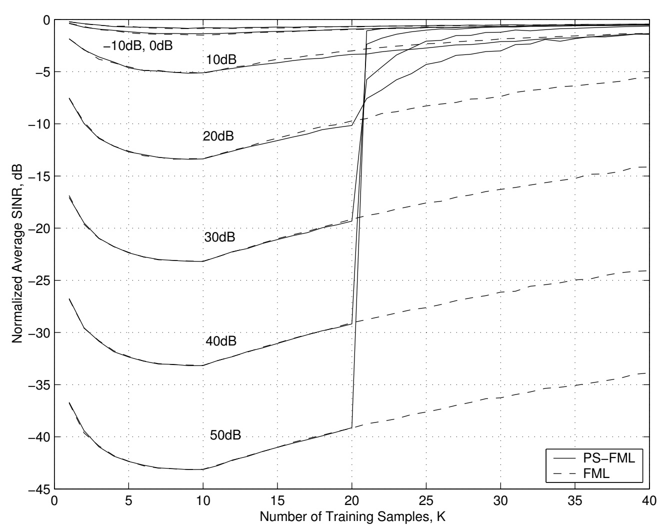


Figure 4. Normalized average SINR convergence for FML and PS-FML at different outlier powers, $N = 20$, $K_{out} = 10$.