Partially adaptive filtering using randomized projections

Olivier Besson, ISAE-SUPAERO

submitted to 5^{th} SONDRA workshop

We consider a scenario where an adaptive filter is used to retrieve a signal of interest among disturbance which consists of a strong low-rank interference and white noise. Additionally we address the situation where the number of training samples available K is much lower than the size of the observation space N and of the order $J - 2J$ where J is the size of the interference subspace. This is typically the case in a radar scenario with a possibly rapidly evolving interference environment or with a clutter that is non homogeneous, and thus only a few cells around the cell under test can be used. In order to cope with scarce training samples, a solution lies in a partially adaptive filter which relies on reducing the size of the observations -by means of a linear transformation- and then operating in this lower dimensional space. This approach is illustrated in Figure 1 with a generalized sidelobe canceler structure. Here $\mathbf{v}$ is the signature of the SoI (assumed to be unit-norm) and $\mathbf{V}_\perp$ is a semi-unitary matrix

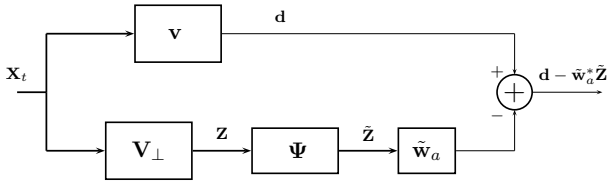

Figure 1: Structure of a partially adaptive filter with $\mathbf{\Psi}$ a $(N-1) \times R$ matrix.

orthogonal to $\mathbf{v}$. $\mathbf{X}_t$ stands for the $N \times K$ matrix of the training samples and $\mathbf{Z} = \mathbf{V}_\perp^* \mathbf{X}_t$ corresponds to the auxiliary channels. $\mathbf{\Psi}$ denotes the $(N-1) \times R$ matrix operating on $\mathbf{Z}$ which achieves dimension reduction. Its goal is to capture the interference subspace so that the reduced-dimension adaptive filter $\tilde{\mathbf{w}}_a$ can estimate from $\tilde{\mathbf{Z}} = \mathbf{\Psi}^* \mathbf{Z}$ the part of interference present in the main channel $\mathbf{d} = \mathbf{v}^* \mathbf{X}_t$.

This structure is especially efficient when the disturbance covariance matrix $\mathbf{\Sigma}$ is the sum of a low rank term plus a scaled identity matrix i.e., $\mathbf{\Sigma} = \mathbf{G}\mathbf{G}^* + \sigma^2 \mathbf{I}_N$ where $\mathbf{G}$ is a $N \times J$ matrix. Actually if $\mathbf{\Sigma}$ is known the optimal filter $(\mathbf{v}^* \mathbf{\Sigma}^{-1} \mathbf{v})^{-1} \mathbf{\Sigma}^{-1} \mathbf{v}$ coincides with a partially adaptive filter where $\mathbf{\Psi} = \mathbf{V}_\perp^* \mathbf{G}$ [1]. In practical situations where $\mathbf{\Sigma}$ is unknown, this partially adaptive filter can result in nearly optimal performance, provided of course that $\mathbf{\Psi}$ captures most of the range space of $\mathbf{V}_\perp^* \mathbf{G}$. Actually, the SNR loss at the output of a partially adaptive filter with a fixed $\mathbf{\Psi}$ is distributed according to a scaled beta distributed random variable where the scaling factor measures how much of the main interference subspace is retained through $\mathbf{\Psi}$.

Since the goal of $\mathbf{\Psi}$ is that most of the main subspace of $\mathbf{Z}$ carries on to $\tilde{\mathbf{Z}}$, a logical choice is to select the R principal left singular vectors of $\mathbf{Z}$ as the columns of $\mathbf{\Psi}$. This choice leads to the so-called principal component method. An alternative choice is to choose the left singular vectors which result in highest SNR, as in the cross spectral metric method. Now these methods require singular value decomposition (SVD) which can be computationally demanding. **Here we investigate an approach based on randomized methods which does not require SVD but provides a sufficiently accurate estimate of $\mathcal{R}\{\mathbf{V}_\perp^* \mathbf{G}\}$ for the purpose of partially adaptive filtering**.

This approach borrows ideas from recent methods for low-rank approximation of matrices based on random projections [2]. More precisely, for a given $m \times n$ matrix, one wishes to find an approximation $\mathbf{A} \approx \underset{m \times k}{\mathbf{B}} \; \underset{k \times n}{\mathbf{C}}$. The approximation is usually computed in two stages. A first stage consists of a *rangefinder*, whose goal is to construct a low-dimensional subspace that captures most of $\mathcal{R}\{\mathbf{A}\}$. In other words one wants to identify a $m \times (k+p)$ matrix $\mathbf{Q}$ with orthogonal columns such that $\mathbf{A} \approx \mathbf{Q}\mathbf{Q}^* \mathbf{A}$. Towards this end $\mathbf{Q}$ is generally obtained from a QR decomposition of $\mathbf{A}\mathbf{\Omega}$ where $\mathbf{\Omega}$ is a $n \times (k+p)$ random matrix. The second stage consists of reduced SVD computation of $\mathbf{Q}^* \mathbf{A}$ [2].

For the purpose of partially adaptive filtering, only the first step is necessary since we want $\mathbf{\Psi}$ to retain the principal subspace of $\mathbf{Z}$. Moreover, in contrast to the randomized low-rank approximation, we do not need an

approximate orthogonal basis for the subspace to be approximated. This suggests to use

$$\mathbf{\Psi} = \mathbf{Z}\mathbf{\Omega}$$

where $\mathbf{\Omega}$ is a $K \times R$ random matrix. If the covariance matrix of $\mathbf{X}_t$ is $\mathbf{\Sigma} = \mathbf{G}\mathbf{G}^* + \sigma^2\mathbf{I}_N$ with $\mathrm{rank}\,(\mathbf{G}) = J$ then it can be surmised that most of the subspace where $\mathbf{Z}$ lies will be retained provided that $R \geq J$. With this choice, one has $\tilde{\mathbf{Z}} = \mathbf{\Psi}^*\mathbf{Z} = \mathbf{\Omega}^*\mathbf{Z}^*\mathbf{Z}$. The vector $\tilde{\mathbf{w}}_a$, which is obtained by minimizing $\|\tilde{\mathbf{Z}}^*\tilde{\mathbf{w}}_a - \mathbf{d}^*\| = \|\mathbf{Z}^*\mathbf{\Psi}\tilde{\mathbf{w}}_a - \mathbf{d}^*\|^2$ can thus be written as $\tilde{\mathbf{w}}_a = (\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^*)^{-1}\tilde{\mathbf{Z}}\mathbf{d}^* = (\mathbf{\Omega}^*\mathbf{Z}^*\mathbf{Z}\mathbf{Z}^*\mathbf{Z}\mathbf{\Omega})^{-1}\mathbf{\Omega}^*\mathbf{Z}^*\mathbf{Z}\mathbf{d}^*$ and the equivalent length-N filter is then given by

$$\mathbf{w} = \mathbf{v} - \mathbf{V}_\perp\mathbf{\Psi}\tilde{\mathbf{w}}_a = \mathbf{v} - \mathbf{V}_\perp\mathbf{Z}(\mathbf{Z}^*\mathbf{Z})^{-1}\mathbf{P}_{\mathbf{Z}^*\mathbf{Z}\mathbf{\Omega}}\mathbf{d}^*$$

As for the matrix $\mathbf{\Omega}$ it can be possibly drawn from a Gaussian distribution, i.e., $\mathbf{\Omega}$ has independent entries drawn from a Gaussian distribution with zero mean and unit variance. Another possibility we explore is to set $\mathbf{\Omega}(i_r, r) = 1$ and 0 otherwise, where $(i_1, \ldots, i_R)$ is a random partition of $(1, \ldots, K)$. This idea amounts to select only R columns of $\mathbf{Z}$, i.e., $\mathbf{\Psi} = \begin{bmatrix} z_{i_1} & z_{i_2} & \cdots & z_{i_R} \end{bmatrix}$.

For illustration purposes Figure 2 compares the distribution of the SNR loss obtained with $\mathbf{\Psi} = \mathbf{Z}\mathbf{\Omega}$ to that of the PC method. It is seen that there is almost no difference between the new methods and the PC method, while the latter requires eigenvalue decomposition. In other scenarios, we will show that it can be beneficial to choose $R > J$.

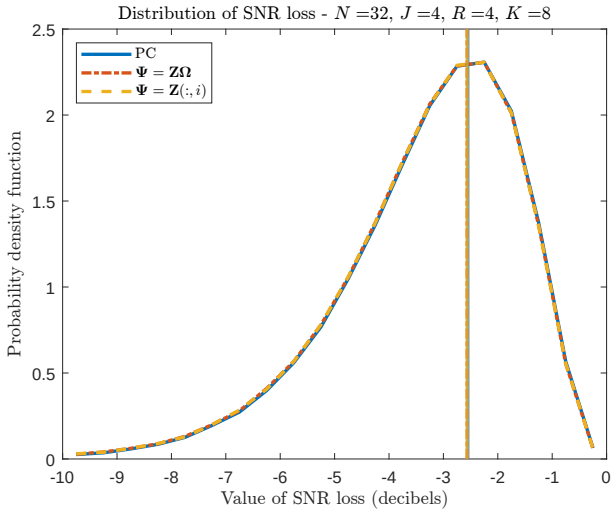

Figure 2: Distribution of the SNR loss. Scenario: uniform linear array with $N = 32$ elements at $\lambda/2$. Source at $0°$, interference located at $-13°$, $-5°$, $6°$ and $9°$ with respective interference to noise ratio 30, 30, 20 and 30 dB.

References

[1] J. Ward. Space-time adaptive processing for airborne radar. Technical Report 1015, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA, December 1994.

[2] N. Halko, P. G. Martinsson, and J. A. Tropp. Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.