

Optimal Wideband Precoder For MIMO

October 27, 2021

For a MIMO wireless communication, we have sampled transmit signal $x[n] \in \mathbb{C}^M$, receive signal $y[n] \in \mathbb{C}^N$, and a channel $h[n] \in \text{Hom}(\mathbb{C}^M, \mathbb{C}^N)$ all with lengths P

$$y[n] = \sum_{n=0}^{P-1} h[n]x[n]$$

And in the frequency domain we have

$$Y[k] = H[k]X[k]$$

where $k = 0, \dots, P-1$.

We want to find an approximation of $h[n]$ denoted by \hat{h} such that y and $\hat{h}x$ are as close as possible, or, Y and $\hat{h}X$ as close as possible. Later we will show that these two requirements are equivalent.

Note that y and Y are elements of a $\text{Hom}(\mathbb{C}^N)$ -module, h and H are elements of another $\text{Hom}(\mathbb{C}^N)$ -module, and x as well as X are elements of a $\mathbb{C}^{M \times M}$ -module. To measure how close to elements of a module are, we can define an "inner product" analogy to the usual inner product defined on Hilbert spaces.

Definition 1 (Inner product on $\text{Hom}(\mathbb{C}^N)$ -modules). An auto-correlation on a P -dimensional free $\text{Hom}(\mathbb{C}^N)$ -module \mathcal{M} is a operation $\langle \cdot, \cdot \rangle: \mathcal{M} \times \mathcal{M} \rightarrow \text{Hom}(\mathbb{C}^N)$

$$\langle x, y \rangle = \sum_{n=0}^{P-1} x[n]y[n]^*$$

We also refer C_{xy} to $\langle x, y \rangle$ for convenience under some context.

Definition 2 (Orthogonality on modules). $x, y \in \mathcal{M}$ are orthogonal if

$$\langle x, y \rangle = 0$$

It is easy to verify that $\langle x, x \rangle$ is positive semi-definite. Therefore we can define the norm induced by the inner product

Definition 3 (Euclidean norm on modules). The Euclidean norm of $x \in \mathcal{M}$ is

$$||x|| = \langle x, x \rangle^{1/2}$$

It is obvious that the norm of any element is positive semi-definite. Hence, there is a partial order \preceq induced by the positive semi-definite cone. And we are able to measure the closeness of two elements in a module by this partial order.

The theorem follows tells us that we can study the closeness in either time domain or frequency domain.

Theorem 1 (Extended Parseval's Theorem).

$$||X|| = \sqrt{P}||x||$$

with the DFT defined by

$$X[k] = \sum_{n=0}^{P-1} x[n] \omega_P^{nk}$$

where

$$\omega_P^{nk} = e^{-j2\pi nk/P}$$

Proof.

$$\begin{aligned} ||X||^2 &= \sum_{k=0}^{P-1} X[k] X[k]^* \\ &= \sum_{k=0}^{P-1} \left(\sum_{m=0}^{P-1} x[m] \omega_P^{mk} \right) \left(\sum_{n=0}^{P-1} x[n] \omega_P^{nk} \right)^* \\ &= \sum_{k=0}^{P-1} \sum_{m=0}^{P-1} \sum_{n=0}^{P-1} x[m] x[n]^* \omega_P^{(m-n)k} \end{aligned}$$

Note that the summands of the outmost summation are non-zero only if $m = n$, thus

$$||X||^2 = \sum_{k=0}^{P-1} Px[k]x[k]^* = P||x||^2$$

completes the proof. \square

To find the optimal \hat{h} in time domain is equivalent to find it in the frequency domain. This amounts to say that we are trying to approximate the frequency-selective $H[k]$ with a constant \hat{h} over all frequencies.

We need to extend the orthogonality principle to modules. Before that we need another kind of product of elements from different modules.

Definition 4 (Cross-correlation). The cross-correlation is a binary operator defined on the product of two modules $\langle \cdot, \cdot \rangle: \mathcal{N} \times \mathcal{M} \rightarrow \text{Hom}(\mathbb{C}^M, \mathbb{C}^N)$

$$\langle x, y \rangle = \sum_{n=0}^{P-1} x[n]y[n]^*$$

And we sometimes refer C_{xy} to $\langle x, y \rangle$ for convenience under certain contexts.

When $N = M$, the cross-correlation coincides with auto-correlation.

Lemma 1. *Let $z \in \mathcal{N}$, $x \in \mathcal{M}$ s.t. $C_{zx} = 0$. Then $\forall \mu \in \text{Hom}(\mathbb{C}^M, \mathbb{C}^N)$ we have*

$$||z + \mu x|| \succeq ||z||$$

Proof. It is easy to verify that

$$||z + \mu x||^2 = ||z||^2 + ||\mu x||^2$$

and we are done. \square

Back to the main goal of this article, we want to approximate y with μx by an optimal μ so that

$$||y - \mu x||$$

is minimum w.r.t. the partial order induced by positive semi-definite cone.

It follows from lemma 1 that every $y \in \mathcal{N}$ can be decomposed into two parts

$$y = z + \mu x$$

where $C_{zx} = 0$ and $\mu \in \text{Hom}(\mathbb{C}^M, \mathbb{C}^N)$ and the μ is what we are looking for and the minimum norm is $\|z\|$.

The next question is how do we find μ and z given y and x .

We see that

$$\langle y, x \rangle = \langle z, x \rangle + \langle \mu x, x \rangle = \mu \langle x, x \rangle$$

or simply

$$C_{yx} = \mu C_{xx} \tag{1}$$

The unique μ can be found if C_{xx} is invertible. But what if it is not?

The existence of μ is assured by the nature of $\text{Hom}(\mathbb{C}^M, \mathbb{C}^N)$. The uniqueness is not assured, in fact it should not be unique. C_{xx} not being invertible means $\ker C_{xx}$ is non-trivial. If μ is a solution to equation 1, then $\forall \beta \in \text{Hom}(\mathbb{C}^M, \mathbb{C}^N)$ such that the restriction of β on $(\ker C_{xx})^\perp$ is trivial, $\mu + \beta$ is also a solution to equation 1.

We could use pseudo inverse of C_{xx} to find the solution such that the standard matrix norm $\text{tr}(\mu^* \mu)$ is minimum. And the pseudo inverse of a positive semi-definite matrix can be calculated by first applying spectral decomposition and then replacing the non-zero singular values by their reciprocals and leaving the zero ones unchanged.

But we are not content with the pseudo inverse as we want to know why it works at all.

The reasoning is not hard. We restrict C_{yx} and C_{xx} on $\mathbb{C}^M / \ker C_{xx}$ to get \tilde{C}_{yx} and \tilde{C}_{xx} respectively. Then \tilde{C}_{xx} is invertible by isomorphism theorem even if it is trivial (maps from a singleton to another singleton). Hence we have

$$\tilde{\mu} = \tilde{C}_{yx} \tilde{C}_{xx}^{-1} \in \text{Hom}(\mathbb{C}^M / \ker C_{xx}, \mathbb{C}^N)$$

where $\tilde{\mu}$ can be easily extended on $\text{Hom}(\mathbb{C}^M, \mathbb{C}^N)$ by assigning any value to μ on $\ker C_{xx}$. To make the standard matrix norm minimum, we can simply assign zero on $\ker C_{xx}$. The reasoning is similar to the proof of lemma 1. Note that \tilde{C}_{xx} is from discarding the zero singular values and the corresponding eigenvectors of C_{xx} . Thus the mechanism of pseudo inverse makes sense.

Remember our goal is to find \hat{h} . We do this in the frequency domain. Let $(\cdot)^\dagger$ denote taking pseudo inverse, then

$$\hat{h} = C_{YX} C_{XX}^\dagger = \left(\sum_{k=0}^{P-1} H[k] X[k] X[k]^* \right) (\|X\|^2)^\dagger$$

In SISO case, i.e., $M = N = 1$, this becomes

$$\hat{h} = \left(\sum_{k=0}^{P-1} H[k] |X[k]|^2 \right) \|X\|^{-2}$$

We see that the optimal wideband precoder is closely related to the transmit signal.

Another question: What is the practical meaning of the optimality in \hat{h} w.r.t. the partial order on S_+^N (positive semi-definite matrices of order N)?

Since the trace of a positive semi-definite matrix is non-negative, it preserves order. And $\text{tr}(yy^*)$ is but the energy of y . Hence \hat{h} is also optimal in the sense that it minimizes the error energy.

In MATLAB 2021a 5G Toolbox, however, it simply calculates the wideband precoder by linear average of $H[k]$'s over frequency indices. This unfortunately does not make good sense. For SISO case, it is OK since the linear average of $H[k]$ corresponds to $X[k]$ being constant in the frequency domain, or equivalently, impulse signal in the time domain. However for MIMO case, linear average of $H[k]$ amounts to say that

$$X[k]X[k]^* (\|X\|^2)^\dagger = \frac{1}{P} I_M, \quad \forall k$$

which implies $X[k]$ is constant over k . This is impossible because the rank of LHS is one whereas M for RHS.