Self-Interference Cancellation For Full-Duplex MIMO Transceivers

Ahmed Masmoudi, Tho Le-Ngoc
Department of Electrical and Computer Engineering, McGill University, Canada
Email: ahmed.masmoudi@mcgill.ca; tho.le-ngoc@mcgill.ca

Abstract—Full-duplex operation requires effective self-interference (SI) cancellation that in turn needs reliable SI channel estimation. In this paper, we develop two estimation algorithms suitable for a 2-stage SI cancellation structure. By exploiting the sparsity of the SI channel, we first derive a compressed sensing-based SI channel estimation algorithm to be used in the first SI cancellation stage at radio-frequency (RF) to reduce the SI. We then develop a subspace-based algorithm to jointly estimate the residual SI channel, the intended channel and the transmitter nonlinearities for the second SI cancellation stage at baseband. Including the intended received signal in the estimation process is the main advantage of the proposed algorithm as compared to previous works that assume it as additive noise. Simulation results show that the proposed algorithms outperform the least-square (LS) algorithm and offer higher signal-to-residual-interference-and-noise ratio (SINR) over a large received signal-to-noise ratio (SNR) range.

I. INTRODUCTION

Current wireless communication systems operate in a half-duplex fashion by transmitting and receiving in different time/frequency slots. One alternative approach to increase the transmission rate of wireless network is full-duplex transmission by allowing simultaneous transmission/reception on the same channel. Full-duplex systems can potentially double the capacity compared to half-duplex systems if the additional self-interference (SI) signal can be perfectly suppressed. In full-duplex systems, the SI is several orders of magnitude higher than the signal of interest because the transceivers use closely-spaced transmit and receive antennas. Therefore, successive cancellation stages have been implemented to sufficiently reduce the SI in order to detect the intended signal [1], [2]. The passive cancellation level is performed by Tx-Rx isolation provided in the multi-antenna sub-system. In the radio-frequency (RF) cancellation level, a replica of the SI signal is created and subtracted from the received signal before the low noise amplifier (LNA) and analog-to-digital converter (ADC) to avoid saturation and additional SI suppression can be done after ADC at the baseband. Even if a complete knowledge of the SI signal is available, the cancellation performance is limited by system imperfections, especially transmitter nonlinearities. Several works [3] [4] [5] have pointed out the impact of transceiver impairments on the system performance.

The work presented in this paper is partly supported by the R&D Contract Advanced Transmission Multiple-Access Schemes for Future Radio Access from Huawei Technologies Canada.

In this work, the SI cancellation is performed in two steps. During a short half-duplex period, noting that the channel exhibits a sparse structure dominated by relatively small number of clusters of significant paths, we adapt the recent advances in the theory of compressed sensing [6], [7] to estimate the SI channel. The unknown channel coefficients are obtained using a mixed-norm optimization criteria which returns the non-zero coefficients and sets the other to zero. Note that the compressed sensing-based channel estimation has been considered in the angle domain, delay-Doppler domain and angle delay-Doppler domain [8]. We prove the restricted isometry property (RIP) [6] of the sensing matrix, a key property to apply compressed sensing, and we provide the regularization parameter to sufficiently reduce the SI. Then, during the full-duplex operation, the received signal at the baseband includes the intended signal from the other transceiver and the residual SI after the RF cancellation stage. A subspace-based algorithm to jointly estimate the residual SI channel, the intended channel and the power amplifier (PA) and imbalanced IQ mixer nonlinearity coefficients is developed. The proposed algorithm considers the intended signal in the estimation process instead of ignoring it as done in previous work [3] [9]. The SI channel can be obtained without any knowledge of the intended transmitted signal while a small training sequence is needed to solve the ambiguity matrix of the intended channel.

II. SYSTEM MODEL AND CANCELLATION SCHEME

We consider two transceivers operating in a full-duplex fashion, i.e., simultaneously transmitting and receiving in the same frequency and time slot. A simplified block diagram of a MIMO full-duplex transceiver with $N_t$ transmit (Tx) streams and $N_r$ receive (Rx) streams is presented in Fig. 1. The simultaneous transmission and reception creates self-interference (SI) to be cancelled before reliable detection. SI can be gradually suppressed through different cancellation stages. The RF SI cancellation stage is done prior to the LNA/ADC to avoid overloading/saturation, followed by the baseband cancellation stage located after the ADC to reduce the residual SI. Without loss of generality, we focus on the received signal at one transceiver. After digital-to-analog converter (DAC), the transmitted OFDM signal of the Tx stream $q$, $x_q(t)$, is passed through an imbalanced IQ mixer to obtain:

$$x_q^{IQ}(t) = g_{1,q}x_q(t) + g_{2,q}x_q^*(t), \quad (1)$$
where $g_{1,q}$ and $g_{2,q}$ are the response of the IQ mixer to the direct signal and its image, respectively. Before transmitting, $x_{t,q}^{IQ}(t)$ is amplified by a nonlinear PA, modeled with a Hammerstein nonlinearity, whose output is:\[ \tag{1} x_{q}^{PA}(t) = (\alpha_{1,q}x_{t,q}^{IQ}(t) + \alpha_{3,q}x_{t,q}^{IQ}(t)x_{t,q}^{IQ}(t))^2 \ast f(t), \] where $f(t)$ models the memory of the PA and $\alpha_{1,q}$ and $\alpha_{3,q}$ are the linear gain and the third-order gain, respectively. $\ast$ denotes the convolution operation. For multipath channels, the received signal of the Rx stream $r$ is: \[ \tag{2} y_{r}^{RX}(t) = \sum_{q=1}^{N_{t}} h_{r,q}^{c}(t) \ast x_{q}^{PA}(t) + \sum_{q=1}^{N_{t}} h_{r,q}^{s}(t) \ast s_{q}(t) + w_{LNA}^{(r)}(t), \] where $s_{q}(t)$ is the intended signal from the Tx stream $q$ of the other intended transceiver. $h_{r,q}^{c}(t)$ is the multipath coupling channel from the Tx stream $q$ to Rx stream $r$ of the same transceiver and $h_{r,q}^{s}(t)$ is the intended channel from Tx stream $q$ of the other intended transceiver to Rx stream $r$. $w_{LNA}^{(r)}(t)$ is the additive thermal noise. Then, the received signal passes through the LNA: \[ \tag{3} y_{r}^{LNA}(t) = k_{LNA}y_{r}^{RX}(t) + w_{LNA}(t), \] where $w_{LNA}(t)$ is the additive noise caused by the LNA and $k_{LNA}$ is the gain of the LNA. Finally, variable gain amplifier (VGA) adjusts the amplitude of the received signal to match the operation range of the ADC. Injecting (2) and (3) into (4) and assuming unity linear gain for the PA and IQ mixer, the output samples are given by: \[ \tag{4} y_{r}(n) = \sum_{q=1}^{N_{t}} \sum_{l=0}^{L} h_{r,q}(l)x_{t,q}^{IQ}(n-l) + \alpha_{3,q}h_{r,q}(l)x_{t,q}^{IQ}(n-l) + h_{r,q}^{s}(n-l) + w_{r}(n), \] where $x_{t,q}^{IQ}(n) = x_{t,q}^{IQ}(n)|x_{t,q}^{IQ}(n)|^2$, the equivalent channel responses are: \[ h_{r,q}^{c}(l) = K_{LNA}h_{r,q}(l) \ast f(l), \]
\[ h_{r,q}^{s}(l) = K_{LNA}h_{r,q}(l), \] and $w_{r}(n)$ collects the quantization noise, the LNA noise and the thermal noise. From (5), the received vector $y(n) = [y_{1}(n), \ldots, y_{N_r}(n)]^{T}$ at the $N_r$ antennas can be written as: \[ \tag{5} y(n) = \sum_{l=0}^{L} X_{iq}(n-l)h_{iq}^{c}(l) + X_{ip3}(n-l)(\alpha_{3} \otimes I_{N_r})h_{ip3}^{c}(l) + S(n-l)h_{us}^{c}(l) + w(n), \] where $X_{iq}(n)$ is a $N_r \times N_t$ Toeplitz matrix with the first column given by the $N_r \times 1$ vector $[x_{1,q}^{IQ}(n), 0, \ldots, 0]$ and the first row given by $[x_{1,q}^{IQ}(n), x_{2,q}^{IQ}(n), \ldots, x_{N_t,q}^{IQ}(n)] \otimes e_1$ with $e_1$ being the $1 \times N_r$ vector having one in the first element and zeroes elsewhere and $\otimes$ being the Kronecker product. The matrices $S(n)$ and $X_{ip3}(n)$ are constructed in the same way as $X_{iq}(n)$ using the sequences $\{q_{s}(n)\}$ and $\{q_{t,q,i}(n)\}$, respectively. In (7), the channel vectors are given by: \[ \tag{6} h_{iq}^{c}(l) = \begin{bmatrix} h_{1,q}^{cT}(l), \ldots, h_{N_t,q}^{cT}(l) \end{bmatrix}^{T}, \]
\[ \tag{7} h_{us}^{c}(l) = \begin{bmatrix} h_{1,q}^{usT}(l), \ldots, h_{N_t,q}^{usT}(l) \end{bmatrix}^{T}, \] and $\alpha_3$ is a $N_t \times N_t$ diagonal matrix whose diagonal elements are $\{\alpha_{3,q}\}$. Now let $h_{iq}^{s} = [h_{iq}^{sT}(0), \ldots, h_{iq}^{sT}(L)]^{T}$ and $h_{us}^{s} = [h_{us}^{sT}(0), \ldots, h_{us}^{sT}(L)]^{T}$ be the $N_tN_r(L+1) \times 1$ vectors gathering all the SI channel and the intended channel coefficients, respectively. And define: \[ \tag{8} X_{iq} = \begin{bmatrix} X_{iq}(0) & X_{iq}(N-1) & \cdots & X_{iq}(N-L) \\ X_{iq}(1) & X_{iq}(0) & \cdots & \vdots \\ \vdots & \vdots & \ddots & X_{iq}(N-1) \\ X_{iq}(N-1) & X_{iq}(N-2) & \cdots & X_{iq}(N-L-1) \end{bmatrix}, \]
\[ \tag{9} S = \begin{bmatrix} S(0) & S(N-1) & \cdots & S(N-L) \\ S(1) & S(0) & \cdots & \vdots \\ \vdots & \vdots & \ddots & S(N-1) \\ S(N-1) & S(N-2) & \cdots & S(N-L-1) \end{bmatrix}. \] The $N_rN \times N_tN_r(L+1)$ self-signal matrix $X_{iq}$ includes samples transmitted from the same transceiver affected by the IQ mixer and the $N_rN \times N_tN_r(L+1)$ intended signal matrix $S$ contains samples transmitted from the intended transmitter. Then we can express the total received vector $y = [yT(0), \ldots, yT(N-1)]^{T}$, after removing the cyclic prefix, as: \[ \tag{10} y = X_{iq}h_{iq} + X_{ip3}I_{L+1} \otimes \alpha_3 \otimes I_{N_r}h_{ip3}^{c} + Sh_{us}^{c} + w, \] where $w$ is the $N_rN \times 1$ thermal noise vector and $X_{ip3}$ is defined in the same way as $X_{iq}$. The main detriment of full-duplex systems is the large SI (shown by the $1st$ term in (11)) compared to the intended
signal (the 3rd term in (11)) and the presence of nonlinear distorted SI (the 2nd term in (11)). This requires multiple cancellation stages to reduce the SI such that the intended signal can be detected [2]. In the following sections, we propose two estimation algorithms to be implemented in the two cancellation stages.

III. COMPRESSED SENSING ESTIMATOR FOR RF CANCELLATION STAGE

The major task in the RF cancellation stage is to reduce the SI before the ADC to avoid overloading/saturation. That is, to recreate the SI and subtract it, we need an initial estimate of the SI channel vector $\mathbf{h}_{si}$. This estimate is obtained during the initial half-duplex period. During this period, the automatic gain control located before the ADC adjusts the dynamic range of the strong SI signal to not saturate the ADC and the following processing can be done in digital domain. The received signal in (11) reduces to:

$$y = \mathbf{X}\mathbf{h}_{si} + \mathbf{w},$$

(12)

where the matrix $\mathbf{X}$ is built in the same way as $\mathbf{X}_{ip}$ from the linear part of the SI transmitted signal and the transmitter impairments are ignored and will be reconsidered in the digital cancellation stage. Since the self-signal matrix $\mathbf{X}$ is known, the classical algorithms to solve this problem rely on linear estimators [10], [11]. However, these methods do not exploit the sparsity of the SI channel. As confirmed by measurements [2], the SI channel between close-by antennas exhibits a strong path component compared to the reflected paths. Therefore, the problem turns out to estimating a sparse channel from the observation $y$ or, mathematically, finding $\mathbf{h}$ which minimizes $\|\mathbf{h}\|_1$ such that $y = \mathbf{Xh}$. This is, however, an intractable combinatorial optimization problem. Recently, it has been shown that, for sparse vector $\mathbf{h}$, minimizing $\|\mathbf{h}\|_1$ instead of $\|\mathbf{h}\|_0$ returns the exact same solution as the original problem [12]. The new optimization problem:

$$\arg\min_{\mathbf{h}} \|\mathbf{h}\|_1 \text{ such that } y = \mathbf{Xh},$$

(13)

is convex and can be solved by linear programming. In practice, the observation vector $y$ is corrupted by noise. Therefore, the constraint in (13) is replaced by $\|y - \mathbf{Xh}\|_2^2 \leq \lambda$, for a given $\lambda$, to include the effects of the additive noise. This optimization problem can be recast as a second-order cone programming [7].

To include the transmitter impairments when cancelling the SI in the RF, the output of the transmitter PA is taken as a reference signal and convoluted with the estimated channel [1]. Therefore, if $\mathbf{h}_{si}$ is known, we will have $\|y - \mathbf{X}_{ip}\hat{\mathbf{h}}_{si} - \mathbf{X}_{ip}(I_{L+1} \odot \alpha_{3} \odot I_{N_r})\hat{\mathbf{h}}\|_2^2 = \|\mathbf{w}\|_2^2 \approx \sigma^2 N_r N$ for sufficiently large noise vector $\mathbf{w}$, where $\sigma$ is the noise variance. In practice, the exact channel $\mathbf{h}$ cannot be obtained. Let $\mathbf{h}_{r}^{si}$ denote the residual channel, i.e., $\mathbf{h}_{r}^{si} = \mathbf{h}_{si} - \hat{\mathbf{h}}$ for an estimate $\hat{\mathbf{h}}$ of the SI channel. Then we have:

$$y = \mathbf{X}_{ip}\hat{\mathbf{h}}_{si} - \mathbf{X}_{ip}(I_{L+1} \odot \alpha_{3} \odot I_{N_r})\hat{\mathbf{h}} = \mathbf{X}_{ip}\mathbf{h}_{r}^{si} + \mathbf{X}_{ip}(I_{L+1} \odot \alpha_{3} \odot I_{N_r})\mathbf{h}_{r}^{si} + \mathbf{w},$$

(14)

In order to detect the intended signal and estimate the residual channel $\mathbf{h}_{r}^{si}$ in the baseband cancellation stage, the residual interference should be reduced to the same power level of the intended signal so that the different components (LNA, ADC) will not be overloaded/saturated. Accordingly, the regularization parameter $\lambda$ should satisfy $(P_s + \sigma^2)N_r N \leq \lambda$ where $P_s$ is the power of the intended signal.

The reconstruction capability of $\mathbf{h}_{si}$ from the samples $\mathbf{y}$ using compressed sensing requires the matrix $\mathbf{X}$ to satisfy the restricted isometry property (RIP) of order $2S$ [6], with $S$ being the number of non-zero elements in $\mathbf{h}_{si}$, i.e., $\mathbf{X}$ satisfies the RIP of order $2S$ with parameter $\delta_S \in [0, 1]$ if, for every vector $\mathbf{\theta}$ such that $\|\mathbf{\theta}\|_0 \leq 2S$, we have:

$$(1 - \delta_S)\|\mathbf{\theta}\|^2_2 \leq \|\mathbf{X}\mathbf{\theta}\|^2_2 \leq (1 + \delta_S)\|\mathbf{\theta}\|^2_2.$$  

(15)

To prove the matrix $\mathbf{X}$ satisfying the RIP, we need to show that the eigenvalues of the Gramian matrix $\mathbf{G}_T = \mathbf{X}_T^H \mathbf{X}_T$ are in the interval $[1 - \delta_S, 1 + \delta_S]$, for all subsets $T \subset \{1, \ldots, N_r(N+1)\}$ with cardinality $|T|$ no larger than $2S$, where $\mathbf{X}_T$ is formed from the columns of $\mathbf{X}$ indexed by the entries of $T$. According to the Geršgorin’s Disc theorem [13], the eigenvalues of $\mathbf{G}_T$ lie in the union of the $|T|$ discs $d_i$ centered at $\mathbf{G}_T(i, i)$ and with radius $\sum_{j\neq i}|\mathbf{G}_T(i, j)|$, for $i = 1, \ldots, |T|$. That is, for given real $\delta_S$ and $\delta_0$ in $[0, 1]$ satisfying $\delta_S = \delta_d + \delta_0$, if we have $|\mathbf{G}_T(i, i) - 1| < \delta_d$ and $|\mathbf{G}_T(i, i)| < \delta_0/2S$, then all the eigenvalues of $\mathbf{G}_T$ are in the union of the discs $d_i$ for $i = 1, \ldots, |T|$ and thus in the range $[1 - \delta_S, 1 + \delta_S]$. It is shown in [14] that the eigenvalues of $\mathbf{G}_T$ are in the range $[1 - \delta_S, 1 + \delta_S]$ with probability approaching 1 and, by the same occasion, the matrix $\mathbf{X}$ satisfies the RIP with probability approaching 1.

IV. BASEBAND CANCELLATION STAGE USING SUBSPACE

When switching to the full-duplex communication mode, the SI channel estimated during the training period is used to reduce the SI in the RF cancellation stage. Using similar vector structures as in Section II, the received signal at the $N_r$ antennas is:

$$y_q(n) = \sum_{q=1}^{N_r} \sum_{l=0}^{L} h_{q}^{l}(l) s_{q}^{l}(n-l) + \alpha_{3,q} h_{q}^{l}(l) x_{q,ip3}(n-l) + h_{q}^{us}(l) \delta_{q}(n-l) + w(n),$$

(16)

where $h_{q}^{l}(l)$ is the residual SI channel from transmit antenna $q$ to the $N_r$ receive antennas. Next, we will estimate the residual SI channel and the nonlinearity coefficients from $y_q(n)$ to be used in the baseband cancellation stage. The input signal in (16) also contains the intended signal. Therefore, the unknown intended signal should be incorporated in the estimation process. In this section, we develop a subspace-based method for jointly estimating the SI channel, intended channel and the nonlinearity coefficients. Before presenting the algorithm, we rewrite (16) in a more compact form:

$$y_q(n) = \sum_{l=0}^{L} H_{r}^{l}(l) x(n-l) + H_{u}^{l}(l) s(n-l) + w(n),$$

(17)
where:
\[
\begin{align*}
x(n) &= [x_1^{(n)} + \alpha_3 x_1, x_2^{(n)} + \alpha_2 x_2, \ldots, x_{2N}^{(n)} + \alpha_1 x_{2N}]^T, \\
s(n) &= [s_1(n), s_2(n), \ldots, s_{2N}(n)]^T, \\
H^{1*}(l) &= [h_1^{1*}(l), h_2^{1*}(l), \ldots, h_{2N}^{1*}(l)], \\
H^{2*}(l) &= [h_1^{2*}(l), h_2^{2*}(l), \ldots, h_{2N}^{2*}(l)].
\end{align*}
\]

Then, we gather the two channel matrices \(H^{1*}(l)\) and \(H^{2*}(l)\) in one matrix \(H(l) = [H^{1*}(l) \ H^{2*}(l)]\) and define the \(N_r \times M \times 2N_r N_t\) block Toeplitz matrix:

\[
\begin{pmatrix}
H(0) & 0 & \ldots & 0 & H(L) & \ldots & H(1) \\
H(1) & H(0) & \ddots & \vdots & \vdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\
H(L) & \vdots & \ddots & 0 & H(0) & \ddots & \vdots \\
0 & H(L) & \ddots & \vdots & \vdots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & H(L)
\end{pmatrix}
\]

(19)

where \(M = N + L\). Gathering the transmitted signals in the \(2N_r N_t \times 1\) vector \(u = [x_1^T(0), s_1^T(0), \ldots, x_1^T(N - 1), s_1^T(N - 1)]^T\), the received samples is written as:

\[
y_c = H u + w.
\]

(20)

**A. Subspace-Based Estimator**

Under the assumption of uncorrelated noise and signal samples, the covariance matrix \(R_{y_c}\) of \(y_c\) is given by:

\[
R_{y_c} = H R_u H^H + \sigma^2 I_{N_r M},
\]

(21)

where \(I_{N_r M}\) is the identity matrix and \(R_u\) is the \(2N_r N_t \times 2N_r N_t\) covariance matrix of \(u\).

The signal subspace is defined as the span of the columns of the matrix \(H\) and the noise subspace is its orthogonal complement. Noting that the dimension of the signal subspace is \(2N_r N_t\) [15], the dimension of the noise subspace is \(d_n = N_r M - 2N_r N_t\). The signal subspace is spanned by the \(2N_r N_t\) eigenvectors of \(R_{y_c}\), corresponding to the largest eigenvalues and the other \(d_n\) eigenvectors, corresponding to \(\sigma^2\), span the noise subspace [16]. Therefore, the columns of \(H\) satisfy:

\[
\nu_i^H H = 0, \quad i = 1, 2, \ldots, d_n.
\]

(22)

where \(\nu_i\), for \(i = 1, \ldots, d_n\), are the \(d_n\) vectors spanning the noise subspace. From (19), the matrix \(H\) is entirely defined by the matrices \(H(0), \ldots, H(L)\). Therefore, instead of looking for the whole \(N_r M \times 2N_r N_t\) matrix \(H\), we restrict our search only on the matrices \(H(l)\). Then, the set of equations in (22) are rearranged as:

\[
\begin{align*}
\sum_{l=0}^{L} \nu_i^H (n + L - l) H(l) &= 0, \quad n = L + 1, \ldots, N \\
\sum_{l=0}^{L} \nu_i^H (n + L - l) H(l) + \sum_{l=n}^{L} \nu_i^H (M - l + n) H(l) &= 0, \\
& \quad n = 1, \ldots, L
\end{align*}
\]

(23)

where each eigenvector \(\nu_i\) is written as:

\[
\nu_i = [\nu_i^T(M), \nu_i^T(M - 1), \ldots, \nu_i^T(1)]^T.
\]

(24)

By defining:

\[
\Theta_i = \begin{pmatrix}
\nu_i^H(L + 1) & \nu_i^H(L) & \ldots & \nu_i^H(1) \\
\nu_i^H(L + 2) & \nu_i^H(L + 1) & \ldots & \nu_i^H(2) \\
\vdots & \vdots & \ddots & \vdots \\
\nu_i^H(N + L) & \nu_i^H(N + L - 1) & \ldots & \nu_i^H(N) \\
0 & \nu_i^H(N + L) & \ldots & \nu_i^H(N + 1) \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \nu_i^H(N + L)
\end{pmatrix} +
\]

(25)

and \(\tilde{H} = [H^T(0), \ldots, H^T(L)]^T\), (23) is equivalent to:

\[
\Theta_i \tilde{H} = 0, \quad i = 1, \ldots, d_n.
\]

(26)

Collecting all the \(\Theta_i\) matrices in a \(Nd_n \times (L + 1)\) matrix \(\Theta = [\Theta_1^T, \Theta_2^T, \ldots, \Theta_{d_n}^T]^T\), we rewrite (26) in a more compact form as:

\[
\Theta \tilde{H} = 0.
\]

(27)

From (27), the columns of \(\tilde{H}\) can be obtained from the basis of the null space of \(\Theta\). To do so, we perform the singular value decomposition of \(\Theta\) and choose the \(2N_r\) right singular vectors as the columns of \(\tilde{H}\). Let \(\tilde{H}_0\) be a solution to (27) obtained as discussed above. For any \(2N_r \times 2N_r\) invertible matrix \(C\), the matrix \(\tilde{H} = \tilde{H}_0 C\) is also a solution to (27). Therefore, the channel matrix is obtained up to a matrix ambiguity. We give in the next section a simple method to solve this ambiguity.

In practice, the channel order \(L\) is unavailable. Therefore, an over estimated value of \(L\) is considered without affecting the estimation process. This is a common property to subspace-based estimators [15].

**B. Solving the Ambiguity Matrix**

If \(H_0\) represents the block Toeplitz matrix constructed as in (19) from the estimated matrix \(\tilde{H}_0\), then the received vector in (20) is written as:

\[
y_c = H_0 (I_N \otimes C) u + w.
\]

(28)

Multiplying \(y_c\) by the pseudo-inverse \(H_0^{-\dagger}\) of \(H_0\), we obtain:

\[
\tilde{y}_c = (I_N \otimes C) u + w.
\]

(29)

Then, we divide \(C\) in two \(N_t \times N_t\) matrices \(C^L\) and \(C^u\) (i.e., \(C = [C^L \ C^u]\)) where the two matrices are associated with the residual SI and intended channels, respectively. By dividing the vector \(\tilde{y}_c\) into \(N\) vectors of size \(2N_t \times 1\), i.e.,
\( \tilde{y}_c = [\tilde{y}_c^T(0), \tilde{y}_c^T(1), \ldots, \tilde{y}_c^T(N-1)]^T \), we rewrite (29) as follows:

\[
\tilde{y}_c(n) = C' x(n) + C'' u s(n) + w(n), \quad n = 0, \ldots, N-1
\]

where \( x(n) = x_1(n) + g x_1^*(n) + \alpha_3 x_{ip3}(n) \):

\[
\tilde{y}_c(n) = C' x_1(n) + C'_{iq} x_1^*(n) + C'_{ip3} x_{ip3}(n) + C'' u s(n) + w(n),
\]

(31)

where \( x_1(n) = [x_1(n), \ldots, x_{N_1}(n)]^T \) and \( x_{ip3}(n) = [x_{1,ip3}(n), \ldots, x_{N_1,ip3}(n)]^T \). \( g \) is the diagonal matrix whose diagonal elements are \( \{g_{2,q}\} \) and \( C'' u q = C' q u \) and \( C'_{ip3} = C'_{i} \alpha_3 \) are two matrices capturing the effects of the IQ mixer and the PA, respectively. Rearranging the term \( C' x_1(n) \) as \( \tilde{x}_i(n) C' \) where \( C' \) is the \( 2N_1 \times 1 \) vector obtained by stacking all the columns of \( C'^T \) on the top of each other and \( \tilde{x}_i(n) = I_{2N_1} \otimes \tilde{x}_1^T(n) \). \( \tilde{C}'_{iq}', \tilde{C}'_{ip3} ' \) and \( \tilde{x}_{ip3}(n) \) are also defined in the same way. Then, from (30), a linear estimator of \( C', \tilde{C}'_{iq} \) and \( \tilde{C}'_{ip3} \) is given by [17]:

\[
\begin{bmatrix}
\tilde{x}_{i q} \\
\tilde{C}
\end{bmatrix} = \left( \sum_{n=0}^{N-1} \tilde{x}(n) \tilde{x}(n)^H \right)^{-1} \sum_{n=0}^{N-1} \tilde{x}(n) \tilde{y}_c(n).
\]

(32)

where \( \tilde{x}(n) = [\tilde{x}_i(n) \tilde{x}_i^*(n) \tilde{x}_{ip3}(n)] \). Note that the elements of \( \tilde{x}_{ip3}(n) \) are approximated by \( x_q(n) |x_q(n)|^2 \). Using the previous estimates in (32), we obtain a cleaner version of \( \tilde{y}_c(n) \) as \( \tilde{y}_c(n) = \tilde{x}_i(n) \tilde{C}' \tilde{x}_i^*(n) \tilde{C}' - \tilde{x}_{ip3}(n) \tilde{C}'_{ip3} \). Assuming that a sequence of pilot symbols are inserted in the subcarriers indexed by \( P = \{p_1, \ldots, p_P\} \), then the intended transmitted signal at antenna \( q \) is the sum of:

\[
s_q^0(n) = \sum_{i=1}^{P} S_q(p_i) e^{j2\pi kp_N}, \quad s_q^d(n) = \sum_{k \notin P} S_q(k) e^{j2\pi kn/N},
\]

(33)

where the first sequence \( s_q^0(n) \) contains the pilot symbols and the second sequence \( s_q^d(n) \) contains the unknown data symbols. By separating the pilot and data sequences in the expression of \( \tilde{C}_q \), \( \tilde{C}'_{iq} \) can be obtained as:

\[
\tilde{C}_q' = \left( \sum_{n=0}^{N-1} \tilde{s}_p^H(n) \tilde{\varphi}(n) \right)^{-1} \sum_{n=0}^{N-1} \tilde{s}_p^H(n) \tilde{\varphi}(n),
\]

(34)

where \( \tilde{\varphi}(n) \) is defined in the same way as \( \tilde{\varphi}(n) \) using the pilot sequence \( s_q^0(n) \) instead of the self signal \( x_q(n) \).

V. SIMULATION RESULTS

In the simulations, we consider a MIMO full-duplex using OFDM-4-QAM with an input intended-signal-to-SI power ratio (SIR\textsubscript{input}) before cancellation sets to \(-50\) dB, i.e., SI power is assumed to be \(50\) dB higher than the intended received signal power. The wireless channels are represented by multipath fading models with 9 paths (i.e., \( L = 8 \)). The SI channels are measured while the intended channel taps are generated as complex zero-mean i.i.d. Gaussian random variables. A complete transmission chain is implemented to model the transmitter PA and IQ mixer, and the receiver LNA and ADC. To properly assess the performance of the proposed approach, we compare it to the LS estimator by considering two scenarios. In the first scenario, the estimate of the residual SI is \( X^\# y \) where the intended received signal is unknown and considered as noise (referred as noisy LS). The second scenario assumes perfectly known intended signal and both the residual SI and intended channels are estimated as follows (referred as joint LS):

\[
\left( \hat{R}_{y y}^{\text{joint}} \right) = [X S]^\# y.
\]

(35)

In practice, the samples estimate \( \hat{R}_{y y} \) of the covariance matrix \( R_{y y} \) is obtained by a time-average of the received samples over \( T \) transmit OFDM symbols.

Fig. 2 depicts the SINR after the different cancellation stages versus SNR for \((N_I = 1, N_r = 2)\) and \((N_I = 2, N_r = 4)\). The sample covariance matrix is obtained using \( T = 50, 70 \) and 100 OFDM blocks. It can be seen that the proposed algorithm greatly outperforms the LS-based when the intended signal is considered as noise. In this scenario, the estimation error of the residual SI channel from the LS estimator is very high, and thus, instead of reducing the SI, it adds additional error, which degrades the SINR after the baseband cancellation stage lower than the SINR after the RF cancellation stage. Therefore, a joint estimation strategy is preferred in order to obtain good cancellation performance. The cancellation performance of the joint LS-based using known intended symbols is greatly improved and comparable with that of the proposed algorithm at low to medium SNR. However, the joint LS algorithm needs known training symbols from the intended transceiver while the proposed algorithm does not, and hence is more bandwidth-efficient.

The cancellation performances are closely related to the accuracy of the estimated parameters. In Figs. 3 and 4, we show the mean square error (MSE) of the estimated SI channel versus the SNR for \((N_I = 1, N_r = 2)\) and \((N_I = 2, N_r = 4)\), respectively. Since the subspace-based algorithm uses the covariance matrix \( R_{y y} \) of the received signal, the performance of the proposed algorithm is closely related to the number of blocks used to estimate \( R_{y y} \). For 70
blocks, the corresponding MSE approaches that with perfectly known matrix $R_p$. The MSE of the LS-based estimator in the first scenario is very poor and thus is not included in the figures. On the other hand, the LS-based joint estimation of the two channels presents relatively good performance at the expense of additional training sequence.

For completeness, performance curves for the MSE of the intended channel estimate are drawn in Fig. 5. To solve the ambiguity of the intended channel, the subspace algorithm needs some training symbols. To illustrate the impact of the training length, we vary the amount of known symbols in an OFDM block. Compared to the LS estimator, our algorithm offers better performance using less known training symbols.

VI. CONCLUSION

In this work, we developed two estimators for RF and baseband cancellation stages in full-duplex MIMO systems. A compressed sensing-based estimator is proposed for the RF cancellation stage to reduce the SI power to the same level of the intended signal. Then, in the baseband cancellation stage, a joint estimation of the residual SI and intended channels, as well as the transmitter impairments is performed using a subspace procedure. To the best of our knowledge, it is the first time that an estimation technique exploits the information bearing in the unknown data to cancel the SI in the digital domain.

REFERENCES

[1] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, “Achieving single channel, full duplex wireless communication,” in Proc. ACM MobiCom, New York, NY, USA, 2010, pp. 1–12.
[2] M. Duarte, C. Dick, and A. Sabharwal, “Experiment-driven characterization of full-duplex wireless systems,” IEEE Trans. Wireless Comm., vol. 11, no. 12, pp. 4296–4307, 2012.
[3] E. Ahmed, A. M. Eltawil, and A. Sabharwal, “Self-interference cancellation with nonlinear distortion suppression for full-duplex systems,” in Proc. ASILOMAR Signals, Syst., Comput., 2013, pp. 1199–1203.
[4] ——, “Rate gain region and design tradeoffs for full-duplex wireless communications,” IEEE Trans. Wireless Comm., vol. 12, no. 7, pp. 3556–3565, 2013.
[5] D. Korpi, T. Riihonen, V. Syrjala, L. Anttila, M. Valkama, and R. Wichman, “Full-duplex transceiver system calculations: Analysis of ADC and linearity challenges,” IEEE Trans. Wireless Comm., vol. 13, no. 7, pp. 3821–3836, July 2014.
[6] E. Candes and T. Tao, “Decoding by linear programming,” IEEE Trans. Inf. Theory, vol. 51, no. 12, pp. 4203–4215, 2005.
[7] E. Candes, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” Comm. Pure appl. math., vol. 59, no. 8, pp. 1207–1223, 2006.
[8] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, “Compressed channel sensing: A new approach to estimating sparse multipath channels,” Proceedings of the IEEE, vol. 98, no. 6, pp. 1058–1076, 2010.
[9] D. Korpi, L. Anttila, V. Syrjala, and M. Valkama, “Wideband linear digital self-interference cancellation in direct-conversion full-duplex transceiver,” IEEE Journal on Selected Areas in Communications, vol. 32, no. 9, pp. 1674–1687, Sept 2014.
[10] X. Ma, G. B. Giannakis, and S. Ohno, “Optimal training for block transmissions over doubly selective wireless fading channels,” IEEE Trans. Signal Process., vol. 51, no. 5, pp. 1351–1366, 2003.
[11] H. Minn and N. Al-Dhahir, “Optimal training signals for MIMO OFDM channel estimation,” IEEE Trans. Wireless Comm., vol. 5, no. 5, pp. 1158–1168, 2006.
[12] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” SIAM journal on scientific computing, vol. 20, no. 1, pp. 33–61, 1998.
[13] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, “Toeplitz compressed sensing matrices with applications to sparse channel estimation,” IEEE Trans. Inf. Theory, vol. 56, no. 11, pp. 5862–5875, 2010.
[14] A. Masmoudi and T. Le-Ngoc, “2-stage self-interference cancellation in full-duplex communication systems,” submitted to IEEE Trans. Vehicular Technology, 2014.
[15] E. Moulard, 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.
[16] R. Schmidt, “Multiple emitter location and signal parameter estimation,” IEEE Trans. Antennas and Propagation, vol. 34, no. 3, pp. 276–280, 1986.
[17] S. M. Kay, Fundamentals of statistical signal processing, Volume 1: Estimation theory. Prentice Hall, 1993.