A SET-THEORETIC APPROACH TO MIMO DETECTION

Jochen Fink, Renato L. G. Cavalcante, Zoran Utkovski, and Sławomir Stańczak

Technische Universität Berlin and *Fraunhofer Heinrich-Hertz-Institute*, Berlin, Germany

ABSTRACT

In this paper, we propose a set-theoretic framework for MIMO detection. Various low-complexity MIMO detection algorithms achieve excellent performance on i.i.d. Gaussian channels, but they typically incur high performance loss if realistic channel models are considered. Compared to existing low-complexity iterative detectors such as approximate message passing (AMP), the proposed algorithms do not impose any structure the channel matrix. Simulations with a realistic channel model show that the proposed methods are competitive with detectors based on orthogonal AMP (OAMP), which compute matrix inverses in each iteration. At the same time, the proposed methods do not require matrix inverses, and their complexity is similar to AMP.

***Index Terms*—** MIMO detection, nonconvex optimization, adaptive projected subgradient method, superiorization.

1. INTRODUCTION

Because of the growing interest in large-scale multi-antenna systems, low-complexity algorithms for MIMO detection have been increasingly gaining attention. The MIMO detection problem has been studied for decades (see [1], [2] for an overview of MIMO detection algorithms). While algorithms that achieve close to optimal performance (such as sphere decoding [3]) are typically too computationally demanding for large-scale systems, low-complexity techniques often perform poorly on realistic channels [4]. The authors of [5] propose a low-complextity MIMO detector based on approximate message passing (AMP). They show that this individually-optimal large-MIMO AMP (IO-LAMA) algorithm is optimal for MIMO detection over i.i.d. Gaussian channels in the large-system limit under some additional conditions. In [6], the authors relax the assumption of i.i.d. Gaussian channels by proposing an orthogonal AMP (OAMP) algorithm for MIMO detection over the more general class of unitarily invariant channel matrices. In contrast to the AMP detector proposed in [5], each iteration of OAMP involves a matrix inversion in order to compute the linear minimum mean square error (LMMSE) estimate, making OAMP more computationally complex than IO-LAMA. Many recent publications on MIMO detection [4, 7–12] propose deep-unfolded versions of iterative detectors. Despite their celebrated success, some of these techniques have been found to suffer considerable performance losses on realistic channels. The authors of [4] mitigate this problem by proposing an online training scheme, which in turn increases the computational cost compared to deep-unfolded algorithms that are trained offline. As the hyperparameters used during training can influence severely the performance of the resulting deep-unfolded detectors, we restrict our attention to untrained detectors in this paper. However, we note that, owing to their iterative structure, the proposed algorithms can be readily used as a basis for deep-unfolded detection algorithms.

In this paper, we approach MIMO detection from a set-theoretic perspective. We propose iterative MIMO detectors based on the adaptive projected subgradient method (APSM) [13] and the superiorization methodology [14]. The proposed detectors have a per-iteration complexity similar to IO-LAMA. In contrast to IO-LAMA, the proposed methods do not impose assumptions on the channel matrix, so their convergence guarantees are not restricted to i.i.d. Gaussian channels. Moreover, simulations show that, despite their low complexity, the proposed methods can outperform OAMP on realistic channels.

Preliminaries and Notation

Throughout this paper, $\mathbb{N}$, $\mathbb{R}_+$, and $\mathbb{R}$ denote the sets of nonnegative integers, nonnegative real numbers, and real numbers, respectively. Unless specified otherwise, lower case letters denote scalars, lower case bold letters denote column vectors, and upper case bold letters denote matrices. The ith entry of a column vector $\mathbf{x}$ is denoted by x_i, the $N \times N$-identity matrix is denoted by $\mathbf{I}_N$, and the spectral norm of a matrix $\mathbf{A}$ is denoted by $\|\mathbf{A}\|_2$.

Given a finite dimensional Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ with induced norm $\|\cdot\|$ and a proper, lower-semicontinuous convex function $f : \mathcal{H} \to (-\infty, +\infty]$, the proximal mapping prox_f associated with f is defined by $(\forall \mathbf{x} \in \mathcal{H})$

$$\mathrm{prox}_f(\mathbf{x}) \in \arg\min_{\mathbf{y}\in\mathcal{H}} \left(f(\mathbf{y}) + \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 \right). \qquad (1)$$

Since f is proper, lower-semicontinuous, and convex, the solution set in (1) is a singleton in general, and its unique element is referred to as $\mathrm{prox}_f(\mathbf{x})$. In this work, we extend the notion of proximal mappings to proper closed (possibly nonconvex) functions. In this case, we denote by $\mathrm{prox}_f(\mathbf{x})$ a unique point selected from the solution set of (1).[1] Similarly, if $\mathcal{C}$ is a nonempty closed nonconvex set, we denote by $P_{\mathcal{C}}(\mathbf{x})$ a unique point selected from the set of projections of $\mathbf{x}$ onto $\mathcal{C}$.

2. PROBLEM STATEMENT

We consider a MIMO system with K transmit- and N receive antennas. For square constellations (QPSK, QAM), which are commonly used in practice, we can describe the system using the real-valued signal model [16]

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w},$$

where $\mathbf{y} \in \mathbb{R}^{2N}$ is the received signal, $\mathbf{H} \in \mathbb{R}^{2N\times 2K}$ is the channel matrix, $\mathbf{x} \in \mathbb{R}^{2K}$ is the transmit signal with coefficients $(\forall i \in [2K] := \{1,\ldots,2K\})$ $x_i \in \mathcal{A} \subset \mathbb{R}$ drawn i.i.d. from a uniform distribution over the finite set $\mathcal{A}$ of real-valued constellation points, and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \frac{\sigma^2}{2}\mathbf{I}_{2N})$ is a vector of i.i.d. Gaussian noise samples.

The objective of MIMO detection is to estimate the transmit signal vector $\mathbf{x}$ based on knowledge of the channel $\mathbf{H}$ and the received signal vector $\mathbf{y}$. Since $\mathbf{x}$ is distributed uniformly over the constellation alphabet, and $\mathbf{w}$ is a vector of Gaussian noise, the optimal

[1] Note that this set is nonempty if $(\mathcal{H} = \mathbb{R}^J, \langle\cdot,\cdot\rangle)$ is a finite dimensional real Hilbert space and the function $\mathbf{y} \mapsto f(\mathbf{y}) + \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2$ is coercive for all $\mathbf{x} \in \mathcal{H}$ [15, Theorem 6.4].

detector uses the maximum likelihood (ML) criterion given by

$$\mathbf{s}^\star \in \arg\max_{\mathbf{x}\in\mathcal{S}} p(\mathbf{y}|\mathbf{x}) = \arg\min_{\mathbf{x}\in\mathcal{S}} \|\mathbf{H}\mathbf{x}-\mathbf{y}\|_2^2, \tag{2}$$

where $\mathcal{S} := \mathcal{A}^{2K} \subset \mathbb{R}^{2K}$ is the discrete set of feasible transmit signal vectors. The ML problem is known to be NP-hard [17] (and, in fact, NP-complete [18]). Therefore, various suboptimal approximations have been proposed. In this paper, we approach the problem from a set-theoretic perspective, which allows us to devise low-complexity approximation techniques with provable convergence properties without imposing any additional assumptions.

3. ALGORITHMIC SOLUTION

To apply results from set-theoretic estimation, we formulate Problem (2) in a real Hilbert space $\left(\mathcal{H} := \mathbb{R}^{2K}, \langle\cdot,\cdot\rangle\right)$ equipped with standard Euclidean inner product

$$(\forall \mathbf{x},\mathbf{y}\in\mathcal{H}) \quad \langle \mathbf{x},\mathbf{y}\rangle := \mathbf{y}^T\mathbf{x}$$

inducing the Euclidean norm $\|\mathbf{x}\| = \sqrt{\langle\mathbf{x},\mathbf{x}\rangle}$. In this Hilbert space, we can express the ML problem in (2) as

$$\underset{\mathbf{x}\in\mathcal{H}}{\text{minimize}} \; \|\mathbf{H}\mathbf{x}-\mathbf{y}\|^2 + \iota_{\mathcal{S}}(\mathbf{x}), \tag{3}$$

where and $\iota_{\mathcal{S}} : \mathcal{H} \to \mathbb{R}_+ \cup \{+\infty\}$ is the indicator function of $\mathcal{S}$ given by

$$(\forall \mathbf{x}\in\mathcal{H}) \quad \iota_{\mathcal{S}}(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x}\in\mathcal{S} \\ +\infty & \text{otherwise.}\end{cases}$$

In the following, we propose a superiorized APSM for MIMO detection. Similarly to IO-LAMA (AMP), which alternates between gradient steps and (Gaussian) denoising steps, the proposed algorithm interleaves subgradient projections onto sublevel sets of increasing cost with denoising steps defined by hard slicing or soft thresholding. In Section 3.1, we relax the discrete set $\mathcal{S}$ to its convex hull, and devise an APSM method with provable convergence guarantees to tackle this relaxed problem. Due to relaxation, we cannot ensure that this algorithm produces feasible constellation vectors. Therefore, in Section 3.2, we propose superiorized [14] version of this algorithm by adding bounded perturbations in each iteration with the intent to steer the iterate towards the nonconvex set $\mathcal{S}$. Although NP-completeness of Problem 2 prohibits proving convergence to an optimal point, we characterize useful features of the point to which the superiorized algorithm converges in Section 3.3.

3.1. An Adaptive Projected Subgradient Method for MIMO Detection

By defining $\rho^\star := \|\mathbf{H}\mathbf{s}^\star - \mathbf{y}\|^2$, we can express (3) as the nonconvex feasibility problem

$$\text{Find } \mathbf{x} \text{ such that } \mathbf{x}\in\mathcal{C}_{\rho^\star}\cap\mathcal{S}. \tag{4}$$

Here, $\mathcal{S}\subset\mathcal{H}$ is the discrete set of feasible transmit signals, and

$$(\forall\rho\geq 0) \quad \mathcal{C}_\rho := \left\{\mathbf{x}\in\mathcal{H} \mid \|\mathbf{H}\mathbf{x}-\mathbf{y}\|^2 \leq \rho\right\}$$

is a sub-level set of the objective function in (2), also known as a stochastic property set [19]. By replacing the discrete set $\mathcal{S}$ with its convex hull

$$\mathcal{B} := \left\{\mathbf{x}\in\mathcal{H} \mid \|\mathbf{x}\|_\infty \leq a_{\max}\right\},$$

where $a_{\max} = \max_{a\in\mathcal{A}} |a|$, we can relax Problem (4) into the convex feasibility problem

$$\text{Find } \mathbf{x} \text{ such that } \mathbf{x}\in\mathcal{C}_{\rho^\star}\cap\mathcal{B}. \tag{5}$$

If $\rho^\star$ were known, Problem (5) could be solved using alternating projection methods such as projections onto convex sets [20]. In the following, we build upon a technique shown in [21], where the objective is to solve the problem

$$\text{Find } \mathbf{x} \text{ such that } \mathbf{x}\in\left(\bigcap_{n\geq n_0}\mathcal{C}_{\rho_n}\right)\cap\mathcal{B},$$

for some $n_0\in\mathbb{N}$, given a sequence $(\mathcal{C}_{\rho_n})_{n\in\mathbb{N}}$ of stochastic property sets. As in [21], we approach this problem with the APSM [13]. To do so, we define a sequence of continuous convex functions $(\forall n\in\mathbb{N})$ $\Theta_n : \mathcal{H}\to\mathbb{R}_+$ by

$$(\forall n\in\mathbb{N})(\forall\mathbf{x}\in\mathcal{H}) \quad \Theta_n(\mathbf{x}) := \left(\|\mathbf{H}\mathbf{x}-\mathbf{y}\|^2 - \rho_n\right)_+,$$

Then we use the APSM to minimize asymptotically this sequence of functions (see the definition of asymptotic minimization in [13]) over the set $\mathcal{B}$ by iteratively applying the recursion

$$\mathbf{x}_0\in\mathcal{H}, \quad (\forall n\in\mathbb{N})\; \mathbf{x}_{n+1} := T_n(\mathbf{x}_n), \tag{6}$$

where [2]

$$T_n(\mathbf{x}) := \begin{cases} P_{\mathcal{B}}\left(\mathbf{x} - \mu_n \frac{\Theta_n(\mathbf{x})}{\|\Theta_n'(\mathbf{x})\|^2}\Theta_n'(\mathbf{x})\right) & \text{if } \Theta_n'(\mathbf{x})\neq\mathbf{0} \\ P_{\mathcal{B}}(\mathbf{x}) & \text{otherwise.}\end{cases} \tag{7}$$

Here, $(\forall n\in\mathbb{N})$ $\Theta_n' : \mathcal{H}\to\mathcal{H}$ defines a subgradient $(\forall\mathbf{x}\in\mathcal{H})$

$$\partial\Theta_n(\mathbf{x}) \ni \Theta_n'(\mathbf{x}) = \begin{cases} 2\mathbf{H}^T(\mathbf{H}\mathbf{x}-\mathbf{y}) & \text{if } \Theta_n(\mathbf{x}) > 0 \\ \mathbf{0} & \text{otherwise}\end{cases}$$

of Θ_n at $\mathbf{x}$, and $\mu_n\in[\varepsilon_1, 2-\varepsilon_2]\subset(0,2)$ is a relaxation parameter. If we choose the elements of $(\rho_n)_{n\in\mathbb{N}}$ to increase monotonically in such a way that $(\exists n_0\in\mathbb{N})$ $\rho_n > \rho^\star$, the recursion in (6) is guaranteed to converge (see Section 3.3). Moreover, if ρ_0 is sufficiently small and $(\rho_n)_{n\in\mathbb{N}}$ increases sufficiently slowly, the final objective value $\lim_{n\to\infty}\|\mathbf{H}\mathbf{x}_n-\mathbf{y}\|^2$ will be close to optimal.

3.2. Superiorization

Replacing the discrete constellation alphabet $\mathcal{S}$ by its convex hull $\mathcal{B}\supset\mathcal{S}$ can potentially limit the performance of the algorithm in (6), because it ignores available information on the prior distribution of $\mathbf{x}$. Therefore, in the following, we use the superiorization methodology [14], [22] to exploit this knowledge. The idea of superiorization is to add small perturbations to the iterates of a so-called *basic algorithm*, with the intent to reduce slightly the value of some superiorization objective. For the problem at hand, we use the recursion in (6) as a basic algorithm. By showing that this algorithm is *bounded perturbation resilient*, convergence of the perturbed algorithm

$$\mathbf{x}_0\in\mathcal{H}, \quad (\forall n\in\mathbb{N})\; \mathbf{x}_{n+1} := T_n(\mathbf{x}_n + \beta_n\mathbf{v}_n) \tag{8}$$

can still be guaranteed, given that $(\beta_n\mathbf{v}_n)_{n\in\mathbb{N}}$ are bounded perturbations, i.e., that $(\beta_n)_{n\in\mathbb{N}}$ is as summable sequence in $\mathbb{R}_+$ and that $(\exists r\in\mathbb{R}_+)(\forall n\in\mathbb{N})$ $\|\mathbf{v}_n\|\leq r$. This allows us to introduce perturbations designed in such a way that they steer the iterates closer to the nearest point in the nonconvex set $\mathcal{S}$.

Objective functions for superiorization are typically convex. Nevertheless, we will consider nonconvex objective functions in the following. Moreover, in a slight deviation from [14] and [22], we use proximal mappings instead of subgradients of the superiorization objective to define the perturbations. This idea was also used in [23]. It allows for a simple trade-off between the perturbations' magnitude and their contribution to reducing the objective value.

[2]The projection onto $\mathcal{B}$ for $\Theta_n'(\mathbf{x}) = \mathbf{0}$ is not present in [13]. It ensures that the APSM generates a sequence in $\mathcal{B}$ regardless of the perturbations added in (8).

In order to introduce prior information on the constellation alphabet, we are interested in superiorization objective functions $f : \mathcal{H} \to \mathbb{R}_+ \cup \{+\infty\}$ that satisfy $f(\mathbf{x}) = 0$ if and only if $\mathbf{x} \in \mathcal{S}$. One example of such a function is the indicator function $f_{\ell_2} := \iota_S$. The proximal mapping associated with f_{ℓ_2} is given by

$$\operatorname{prox}_{f_{\ell_2}}(\mathbf{x}) = P_{\mathcal{S}}(\mathbf{x}).$$

Here, $P_{\mathcal{S}}$ denotes a projection onto the set $\mathcal{S}$. Since $\mathcal{S}$ is not convex, this point is not unique for all $\mathbf{x} \in \mathcal{H}$. However, a projection onto $\mathcal{S}$ always exists because the set is closed. In this way, we can devise perturbations of the form $(\forall n \in \mathbb{N})$

$$\mathbf{v}_n^{(\ell_2)} := P_{\mathcal{S}}(\mathbf{x}_n) - \mathbf{x}_n. \tag{9}$$

As the primary objective of MIMO detection is to reduce the symbol error ratio (SER), one could instead use a superiorization objective that penalizes the number of coefficients of the estimate $\hat{\mathbf{x}}$ which lie outside the the set of valid constellation points, i.e.,

$$\sum_{k:\hat{x}_k \notin \mathcal{A}} 1 = \|\hat{\mathbf{x}} - P_{\mathcal{S}}(\hat{\mathbf{x}})\|_0, \tag{10}$$

where $\|\cdot\|_0$ denotes the ℓ_0 pseudo-norm. Borrowing a well-known result from compressed sensing [24], we replace the ℓ_0 pseudo-norm in (10) with the ℓ_1-norm to define an alternative superiorization objective

$$(\forall \mathbf{x} \in \mathcal{H}) \quad f_{\ell_1}(\mathbf{x}) := \|\mathbf{x} - P_{\mathcal{S}}(\mathbf{x})\|_1.$$

Note that f_{ℓ_1} is still nonconvex due to the projection onto the nonconvex set $\mathcal{S}$. Nevertheless, $(\forall \lambda \geq 0)$ we can define a proximal mapping associated with λf_{ℓ_1} by

$$\operatorname{prox}_{\lambda f_{\ell_1}}(\mathbf{x}) = \phi_\lambda\left(\mathbf{x} - P_{\mathcal{S}}(\mathbf{x})\right) + P_{\mathcal{S}}(\mathbf{x}),$$

where $(\forall \lambda \geq 0)\ \phi_\lambda : \mathcal{H} \to \mathcal{H}$ is the soft-thresholding operator

$$(\forall \mathbf{x} \in \mathcal{H})(\forall k \in [2K]) \quad \phi_\lambda(\mathbf{x})|_k := \operatorname{sign}(x_k)(|x_k| - \lambda)_+.$$

As a result, we obtain perturbations of the form $(\forall n \in \mathbb{N})$

$$\mathbf{v}_n^{(\ell_1)} := \operatorname{prox}_{\lambda_n f_{\ell_1}}(\mathbf{x}_n) - \mathbf{x}_n \tag{11}$$

3.3. Convergence of the Proposed Algorithms

In the following, we investigate the convergence of the proposed algorithms. We begin by stating the following new result. Owing to page limitations, the proof of this theorem will be published elsewhere.

Theorem 1. *Define* $(\forall n \in \mathbb{N})\ \Theta_n^\star := \inf_{\mathbf{x}\in\mathcal{B}} \Theta_n(\mathbf{x})$ *and* $\Omega_n := \{\mathbf{x} \in \mathcal{B} \mid \Theta_n(\mathbf{x}) = \Theta_n^\star\}$. *The APSM in* (6) *and* (7) *is bounded perturbation resilient in the following sense. Suppose that* $(\beta_n \mathbf{v}_n)_{n\in\mathbb{N}}$ *are bounded perturbations and let* $(\forall n \in \mathbb{N})\ \mu_n \in [\varepsilon_1, 2 - \varepsilon_2] \subset (0, 2)$. *Then the sequence produced by superiorized APSM in* (8) *converges to a point* $\hat{\mathbf{x}} \in \mathcal{B}$, *given that*

1. $(\Theta_n'(\mathbf{x}_n))_{n\in\mathbb{N}}$ *is bounded*
2. *there exists a bounded sequence* $(\Theta_n'(\hat{\mathbf{x}}))_{n\in\mathbb{N}}$, *where* $(\forall n \in \mathbb{N})\ \Theta_n'(\hat{\mathbf{x}}) \in \partial\Theta_n(\hat{\mathbf{x}})$
3. $(\exists n_0 \in \mathbb{N})(\forall n \geq n_0)\ \Theta_n^\star = 0$ *and* $\Omega := \bigcap_{n\geq n_0} \Omega_n$ *has nonempty interior, i.e.,* $(\exists \mathbf{z} \in \Omega)(\exists \delta > 0)$ *satisfying* $\{\mathbf{x} \in \mathcal{H} \mid \|\mathbf{x} - \mathbf{z}\| \leq \delta\} \subset \Omega$.

Moreover, the point $\hat{\mathbf{x}}$ *minimizes all but finitely many functions of the sequence* $(\Theta_n)_{n\in\mathbb{N}}$.

In light of Theorem 1, to ensure that the proposed algorithms converge, we only need to show that the proposed perturbations are bounded.

Proposition 1. *The sequences of perturbations proposed in* (9) *and* (11) *are bounded.*

Proof: *Since* $\mathcal{B}$ *is compact, we can define* $c := \max_{\mathbf{x}\in\mathcal{B}} \|\mathbf{x}\|$. *By* (7) *and the definition of a projection,* $(\forall n \in \mathbb{N})\ \mathbf{x}_n \in \mathcal{B}$ *and* $(\forall \mathbf{x} \in \mathcal{H}) P_{\mathcal{S}}(\mathbf{x}) \in \mathcal{S} \subset \mathcal{B}$. *Consequently, we have*

$$\|\mathbf{v}_n^{(\ell_2)}\| = \|P_{\mathcal{S}}(\mathbf{x}) - \mathbf{x}_n\| \leq \|P_{\mathcal{S}}(\mathbf{x})\| + \|\mathbf{x}_n\| \leq 2c$$

and

$$\begin{aligned}\|\mathbf{v}_n^{(\ell_1)}\| &= \|\phi_\lambda\left(\mathbf{x} - P_{\mathcal{S}}(\mathbf{x})\right) + P_{\mathcal{S}}(\mathbf{x}) - \mathbf{x}_n\| \\ &\leq \|\phi_\lambda\left(\mathbf{x} - P_{\mathcal{S}}(\mathbf{x})\right)\| + \|P_{\mathcal{S}}(\mathbf{x}) - \mathbf{x}_n\| \\ &\leq 2\|P_{\mathcal{S}}(\mathbf{x}) - \mathbf{x}_n\| \leq 4c,\end{aligned}$$

which concludes the proof. □

Finally, we apply Theorem 1 and Proposition 2 to prove the convergence of the proposed method.

Proposition 2. *Let* $(\rho_n)_{n\in\mathbb{N}}$ *be a sequence in* $\mathbb{R}_+$ *satisfying* $(\exists n_0 \in \mathbb{N})\ (\exists \eta > 0)\ (\forall n \geq n_0)\ \rho_n \geq \rho^\star + \eta$. *Then the proposed algorithm in* (8) *with perturbations according to* (9) *or* (11) *and summable* $(\beta_n)_{n\in\mathbb{N}}$ *in* $\mathbb{R}_+$ *is guaranteed to converge to a point* $\hat{\mathbf{x}} \in \mathcal{B}$ *minimizing all but finitely many functions of the sequence* $(\Theta_n)_{n\in\mathbb{N}}$.

Proof: *It remains to show that the three conditions in Theorem 1 are satisfied.*

1. *Since* $(\forall n \in \mathbb{N})\ \Theta_n(\mathbf{x}) = 0 \Rightarrow \Theta_n'(\mathbf{x}) = \mathbf{0}$, *it is sufficient to consider the case* $\Theta_n(\mathbf{x}) > 0$. *Here, we have that* $\Theta_n'(\mathbf{x}) = 2\mathbf{H}^T(\mathbf{H}\mathbf{x} - \mathbf{y})$, *so* $(\forall \mathbf{x} \in \mathcal{B})$
$$\begin{aligned}\|\Theta_n'(\mathbf{x}_n)\| &\leq 2\|\mathbf{H}^T\mathbf{H}\mathbf{x}_n\| + 2\|\mathbf{H}^T\mathbf{y}\| \\ &\leq 2\|\mathbf{H}^T\mathbf{H}\|_2 \cdot \|\mathbf{x}_n\| + 2\|\mathbf{H}^T\mathbf{y}\| \\ &\leq 2c\|\mathbf{H}^T\mathbf{H}\|_2 + 2\|\mathbf{H}^T\mathbf{y}\|,\end{aligned}$$
Consequently, the sequence $(\Theta_n'(\mathbf{x}_n))_{n\in\mathbb{N}}$ *is bounded.*
2. *Since* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *is a sequence in the closed set* $\mathcal{B}$, $\hat{\mathbf{x}} = \lim_{n\to\infty} \mathbf{x}_n \in \mathcal{B}$. *Therefore, we can apply the same argument as above.*
3. *Let* $\mathbf{s}^\star$ *denote the maximum likelihood estimate in* (2) *and define* $\mathcal{E} := \{\mathbf{x} \in \mathcal{H} \mid \|\mathbf{s}^\star - \mathbf{x}\| \leq \varepsilon\}$ *with some positive* $\varepsilon \leq \frac{\sqrt{\rho^\star+\eta}-\sqrt{\rho^\star}}{\|\mathbf{H}\|_2}$. *All* $u \in \mathcal{H}$ *with* $\|\mathbf{u}\| \leq 1$ *satisfy*
$$\begin{aligned}\|\mathbf{H}(\mathbf{s}^\star + \varepsilon\mathbf{u}) + \mathbf{y}\|^2 &= \rho^\star + 2\varepsilon\langle \mathbf{H}\mathbf{s}^\star - \mathbf{y}, \mathbf{H}\mathbf{u}\rangle + \varepsilon^2\|\mathbf{H}\mathbf{u}\|^2 \\ &\leq \rho^\star + 2\varepsilon\sqrt{\rho^\star}\|\mathbf{H}\|_2 + \varepsilon^2\|\mathbf{H}\|_2^2 \\ &\leq \rho^\star + \eta.\end{aligned}$$
Therefore, by the premise of this proposition, $(\forall n \geq n_0)$ $(\forall \mathbf{x} \in \mathcal{E})\ \Theta_n(\mathbf{x}) = 0$, *i.e.,* $\mathcal{E} \subseteq \mathcal{C}_{\rho_n}$. *Now, we define a set with nonempty interior by* $\mathcal{Q} := \{\mathbf{x} \in \mathcal{H} \mid (\forall i \in \{1, \ldots, 2K\})\ s_l \leq x_i \leq s_u\}$, *where* $(\forall i \in \{1, \ldots, 2K\})$
$$\tilde{s}_i := \operatorname{sign}(s_i^\star) \cdot \left(|s_i^\star| - \frac{\varepsilon}{\sqrt{2K}}\right),$$
$s_l := \min(\tilde{s}_i, s_i)$, *and* $s_u := \max(\tilde{s}_i, s_i)$. *Note that* $\mathcal{Q} \subset \mathcal{E}$. *Moreover,* $\mathcal{Q} \subset \mathcal{B}$ *for sufficiently small* $\varepsilon > 0$. *Consequently* $\mathcal{Q} \subset \Omega$, *which concludes the proof.* □

4. NUMERICAL RESULTS

In this section, we compare the performance of the following algorithms:

- The APSM basic algorithm in (6) (`APSM`)
- The superiorized APSM in (8) with perturbations according to (9) (`APSM-L2`)

- The superiorized APSM in (8) with perturbations according to (11) (`APSM-L1`)
- The AMP-based MIMO detector (IO-LAMA) proposed in [5] (`AMP`)
- The detector based on OAMP [6] (`OAMP`)
- The LMMSE estimate given by $\mathbf{x}_{\text{LMMSE}} = (\mathbf{H}^T\mathbf{H}+\sigma^2\mathbf{I})\mathbf{H}^T\mathbf{y}$ (`LMMSE`).

We consider a system with $K = 16$ single antenna transmitters and $N = 64$ receive antennas and 16-QAM constellation. As in [4], we assume perfect power allocation, i.e., we normalize the columns of the channel matrix $\mathbf{H}$ to unit 2-norm. For the APSM algorithms, we set $(\forall n \in \mathbb{N})\ \rho_n = 5 \cdot 10^{-5} \cdot 1.06^n$ and $\mu_n = 0.7$. The perturbations of `APSM-L2` are scaled using the sequence $(\beta_n = b^n)_{n\in\mathbb{N}}$ with $b = 0.9$. For `APSM-L1`, we set $(\forall n \in \mathbb{N})\ \lambda_n = 0.005$ and $\beta_n = 0.9999$. Figure 1 shows the SER throughout the iterations, averaged over 10000 i.i.d. Gaussian channel matrices with 9 dB SNR. It can be seen that both `AMP` and `OAMP` achieve ML performance within about 10 iterations. The proposed methods do not achieve ML performance. However, they still outperform `LMMSE`. Figure 2

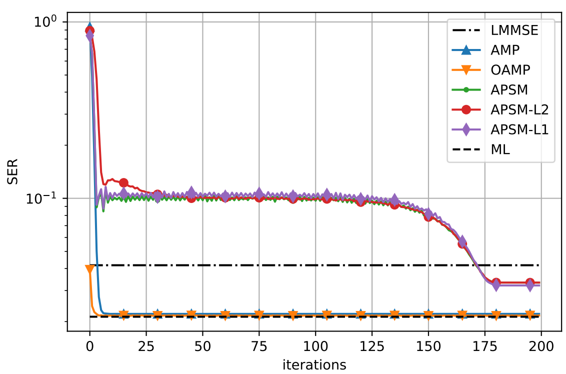

Fig. 1. SER as a function of the number of iterations averaged over 10000 realizations of i.i.d. Gaussian channels.

shows the he SER as a function of the number of iterations averaged over 10000 (single-subcarrier) 3GPP channels with 18 dB SNR. The single-subcarrier channels are drawn at random from a dataset that was generated using the code provided with [4]. While all APSM-type algorithms achieve a lower SER than `LMMSE`, `AMP` fails to re-

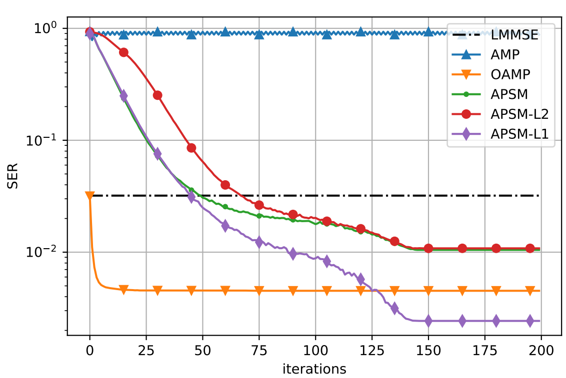

Fig. 2. SER as a function of the number of iterations averaged over 10000 3GPP channels.

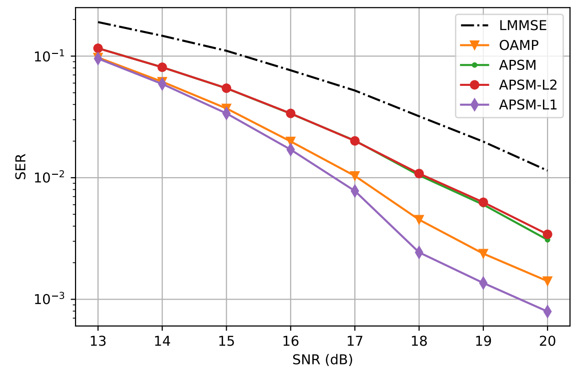

Fig. 3. SER as a function of the SNR averaged over 10000 3GPP channels.

duce the SER throughout the iterations. Superiorization based on the indicator function $f_{\ell_2} = \iota_{\mathcal{S}}$ (`APSM-L2`) does not improve the performance compared to the unperturbed basic algorithm (`APSM`). By contrast, the SER achieved by `APSM-L1` is about an order of magnitude below the unperturbed basic algorithm `APSM`, even outperforming the more complex `OAMP` detector. Figure 3 shows the average SER as a function of the SNR for 3GPP channels. Since `AMP` did not converge for 3GPP channels, it is excluded from this comparison. It can be seen that `APSM-L1` achieves a lower SER than `OAMP` for all SNR levels. The perturbations of `APSM-L2` did not achieve a significant improvement over the unperturbed `APSM`.

5. CONCLUSION

In this paper, we proposed iterative MIMO detectors with convergence guarantees based on a superiorized adaptive projected subgradient method. Unlike IO-LAMA, the proposed methods are not restricted to i.i.d. Gaussian channels. Simulations show that the proposed methods can outperform OAMP on realistic channels. Moreover, in contrast to OAMP, the proposed detectors do not require matrix inverses, so they have a per-iteration complexity similar to IO-LAMA.

6. REFERENCES

[1] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 1941–1988, 2015.

[2] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3109–3132, 2019.

[3] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Transactions on Information theory*, vol. 45, no. 5, pp. 1639–1642, 1999.

[4] M. Khani, M. Alizadeh, J. Hoydis, and P. Fleming, "Adaptive neural signal detection for massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5635–5648, 2020.

[5] C. Jeon, R. Ghods, A. Maleki, and C. Studer, "Optimality of large MIMO detection via approximate message passing," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 1227–1231, IEEE, 2015.

[6] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.

[7] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, IEEE, 2017.

[8] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2554–2564, 2019.

[9] G. Gao, C. Dong, and K. Niu, "Sparsely connected neural network for massive MIMO detection," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, pp. 397–402, IEEE, 2018.

[10] M.-W. Un, M. Shao, W.-K. Ma, and P. Ching, "Deep MIMO detection using ADMM unfolding," in *2019 IEEE Data Science Workshop (DSW)*, pp. 333–337, IEEE, 2019.

[11] N. T. Nguyen and K. Lee, "Deep learning-aided tabu search detection for large MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 4262–4275, 2020.

[12] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "A model-driven deep learning network for MIMO detection," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 584–588, IEEE, 2018.

[13] I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," *Numerical Functional Analysis and Optimization*, 2005.

[14] Y. Censor, "Weak and strong superiorization: between feasibility-seeking and minimization," *arXiv preprint arXiv:1410.0130*, 2014.

[15] A. Beck, *First-order methods in optimization*. SIAM, 2017.

[16] E. Telatar, "Capacity of multi-antenna gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[17] D. Micciancio, "The hardness of the closest vector problem with preprocessing," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 1212–1215, 2001.

[18] A. Del Pia, S. S. Dey, and M. Molinaro, "Mixed-integer quadratic programming is in NP," *Mathematical Programming*, vol. 162, no. 1, pp. 225–240, 2017.

[19] I. Yamada, K. Slavakis, and K. Yamada, "An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1091–1101, 2002.

[20] H. Stark, Y. Yang, and Y. Yang, *Vector space projections: a numerical approach to signal and image processing, neural nets, and optics*. John Wiley & Sons, Inc., 1998.

[21] R. L. G. Cavalcante, S. Stanczak, D. Schupke, and J. Klaue, "Low-complexity distributed set-theoretic decoders for analog fountain codes," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pp. 1480–1484, IEEE, 2018.

[22] Y. Censor, R. Davidi, and G. T. Herman, "Perturbation resilience and superiorization of iterative algorithms," *Inverse Problems*, vol. 26, no. 6, p. 065008, 2010.

[23] J. Fink, R. L. G. Cavalcante, and S. Stańczak, "Multi-group multicast beamforming by superiorized projections onto convex sets," *IEEE Transactions on Signal Processing*, vol. 69, pp. 5708–5722, 2021.

[24] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.