Multiuser MIMO Detection With Composite NUV Priors

Gian Marti, Raphael Keusch, and Hans-Andrea Loeliger
ETH Zurich, Dept. of Information Technology & Electrical Engineering
{marti, keusch, loeliger}@isi.ee.ethz.ch

***Abstract*—Normals with unknown variances (NUV) representations encompass variational representations of sparsifying norms and priors for sparse Bayesian learning. Recently, a binarizing NUV prior has been proposed and shown to work very well on certain approximation problems. We elaborate on this new prior and begin to explore its use for recovery problems. Concretely, we apply the method to the multiuser multiple-input multiple-output (MIMO) detection problem. Empirically, the method outperforms existing approaches based on convex relaxations and is more robust than a method based on approximate message-passing.**

I. Introduction

Many problems in signal processing amount to finding a signal $\mathbf{x} \in \mathbb{R}^n$ that corresponds to some given signal $\mathbf{y} \in \mathbb{R}^m$ through the linear relationship $\mathbf{y} \approx \Phi\mathbf{x}$ for a given $\Phi \in \mathbb{R}^{m\times n}$. A typical way of posing such problems in a formal way is

$$\operatorname*{argmin}_{\mathbf{x}\in\mathcal{O}^n} \|\mathbf{y} - \Phi\mathbf{x}\|^2, \tag{1}$$

where $\mathcal{O} \subset \mathbb{R}$ is a constraint set. If $\mathcal{O}$ is convex, then (1) is convex and readily solvable with convex optimizers. In many practical applications, however, $\mathcal{O} = \{a, b\}$ is binary. In that case, (1) is NP-hard [1]. Finding the optimal solution is therefore often impractical, and one resorts to approximations.

This problem characterization encompasses two fundamentally different problem classes: Problems of *recovery*, where $\mathbf{y}$ is known to have been generated by some specific $\mathbf{x}$ which we aim to recover (e.g., in channel coding or detection), and problems of *approximation*, where we aim to construct an $\mathbf{x}$ to approximate $\mathbf{y}$ under the linear operator Φ (e.g., in control or compression). It is a common experience that methods which work well for recovery need not work well for approximation.

A new method for estimating binary input signals using a new composite NUV (Normal with unknown variance) prior[1] has been proposed in [4] and shown to work well on certain approximation problems in a control setting.

In this paper, we assess the performance of this method for recovery, specifically, for multiuser (MU) multiple-input multiple-output (MIMO) detection. We find that the method shows excellent performance and is thus suitable for problems both of approximation and recovery. A second contribution of the paper lies in complementing [4] with an analysis of *why* the proposed method yields binary estimates.[2]

[1]NUV priors are well known to encompass variational representations of sparsifying norms and priors for sparse Bayesian learning [2], [3].

[2][4] proposes the method in two variations: "joint MAP estimation" and "type-II estimation". This paper only considers the latter, superior variant.

The paper is outlined as follows: In Section II, we introduce the method from [4] in a scalar setting and analyze its ability to yield binary estimates. Section III extends the method to general linear systems and provides an explicit algorithm. We also show how to adapt the method for larger discrete constellations. In Section IV, we apply the method to the MU-MIMO detection problem and analyze the choice of its only hyperparameter. Section V provides an empirical comparison with existing methods. Section VI concludes the paper.

II. The Binarizing Composite NUV Prior

We introduce the composite NUV prior from [4] in a simple but illuminating scalar setting: We want to estimate $x \in \{a, b\}$ based on a scalar observation y such that $y \approx x$.[3] Specifically, we assume a probabilistic setting where the likelihood function $p(y|x)$ is Gaussian with mean $\mu = y$ and variance s^2, $p(y|x) = \mathcal{N}\big(x|\mu, s^2\big)$.[4] In this scalar setting, our prior is

$$\rho(x, \theta) \triangleq \mathcal{N}\big(x|a, \sigma_a^2\big)\mathcal{N}\big(x|b, \sigma_b^2\big), \tag{2}$$

where $\theta = (\sigma_a^2, \sigma_b^2)$. We thereby interrelate the variable of interest, x, with the two newly introduced (virtual) variances (σ_a^2, σ_b^2) through the improper joint prior $\rho(x, \theta)$ by means of the composition of two Normals with unknown variances (NUVs). A factor graph [5] of the resulting statistical model

$$p(y|x)\rho(x, \theta) = \mathcal{N}\big(x|\mu, s^2\big)\mathcal{N}\big(x|a, \sigma_a^2\big)\mathcal{N}\big(x|b, \sigma_b^2\big) \tag{3}$$

is shown in Fig. 1. The unknown variances θ are then obtained by Maximum-A-Posteriori (MAP) estimation:

$$\hat{\theta} = \theta_{\text{MAP}} \triangleq \operatorname*{argmax}_{\theta} \tilde{p}(\theta|y), \tag{4}$$

where

$$\tilde{p}(\theta|y) = \int_{-\infty}^{\infty} p(y|x)\rho(x, \theta)\, dx, \tag{5}$$

is the (improper) posterior of θ given y. Fixing this choice of variances $\hat{\theta}$, we estimate $\hat{x}$ with the MAP estimate

$$\hat{x} = x_{\text{MAP}}(\hat{\theta}) \triangleq \operatorname*{argmax}_{x} p(y|x)\rho(x, \hat{\theta}). \tag{6}$$

Ideally, $\hat{x}$ would be the member of the set $\mathcal{O} = \{a, b\}$ closer to μ (i.e., to y). Is there any reason to think that this will be the case? Indeed there is. To understand this, we start by defining

$$\mu_\theta \triangleq \frac{b\sigma_a^2 + a\sigma_b^2}{\sigma_a^2 + \sigma_b^2} \quad \text{and} \quad \sigma_\theta^2 \triangleq \frac{\sigma_a^2\sigma_b^2}{\sigma_a^2 + \sigma_b^2}, \tag{7}$$

[3]Throughout this paper, we assume without loss of generality that $a < b$.

[4]We use this notation for reasons that should become clear in Section III.

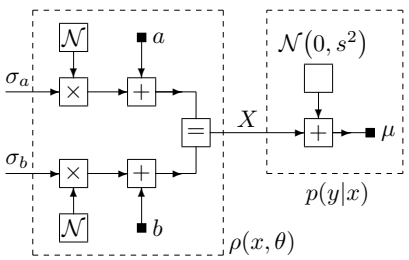


Fig. 1: Factor graph of the statistical model (3) for fixed observation $y = \mu$.

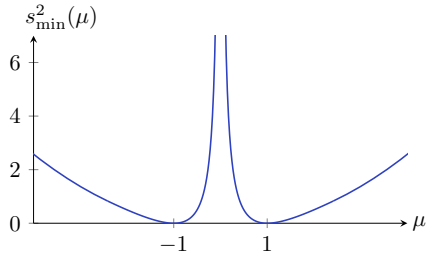


Fig. 2: Minimal s^2 required for binarization as a function of μ.

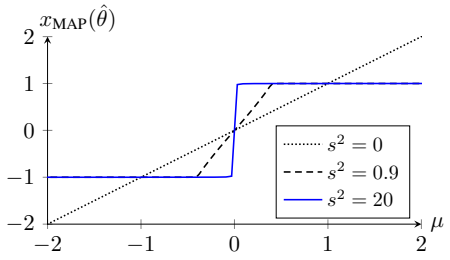


Fig. 3: The estimate $x_{\text{MAP}}(\hat{\theta})$ as a function of μ.

which allows us to rewrite $\rho(x, \theta)$ as

$$\rho(x,\theta) = p(x|\theta) \cdot \rho(\theta) = \mathcal{N}\big(x|\mu_\theta, \sigma_\theta^2\big) \cdot \frac{e^{-\frac{(b-a)^2}{2(\sigma_a^2+\sigma_b^2)}}}{\sqrt{2\pi(\sigma_a^2+\sigma_b^2)}}. \quad (8)$$

Note that $\rho(x, \theta)$ for fixed θ is Gaussian (up to the scale factor $\rho(\theta)$). Consequently, the MAP estimate in (6) becomes

$$x_{\text{MAP}}(\hat{\theta}) = \frac{s^2\mu_{\hat{\theta}} + \sigma_{\hat{\theta}}^2\mu}{s^2 + \sigma_{\hat{\theta}}^2}. \quad (9)$$

Moreover, if exactly one of the two variances $\hat{\theta} = (\sigma_a^2, \sigma_b^2)$ is zero, then $\rho(x, \hat{\theta}) \propto \delta(x-a)$ or $\rho(x, \hat{\theta}) \propto \delta(x-b)$ (depending on which of the variances is zero), and $x_{\text{MAP}}(\hat{\theta}) \in \{a, b\}$. Motivated by this observation, we call such a $\hat{\theta}$ *binarizing*.

To find out whether the maximization of the posterior in (5) yields a binarizing $\hat{\theta}$, we rewrite the posterior $\tilde{p}(\theta|y)$ as

$$\tilde{p}(\theta|y) = \rho(\theta)\,\mathcal{N}\big(\mu - \mu_\theta \,|\, 0, \sigma_\theta^2 + s^2\big), \quad (10)$$

where we used (8) and the fact that $p(y|x) = \mathcal{N}\big(x|\mu, s^2\big)$. Plugging in $\rho(\theta)$, taking logarithms, changing the sign, and dropping irrelevant constants yields the negative-log-posterior (NLP) function, whose minima are the maxima of (10)

$$\mathcal{L}(\theta) \triangleq \log(\sigma_a^2 + \sigma_b^2) + \frac{(a-b)^2}{\sigma_a^2 + \sigma_b^2} + \log(\sigma_\theta^2 + s^2) + \frac{(\mu - \mu_\theta)^2}{\sigma_\theta^2 + s^2}. \quad (11)$$

We then have the following theorem:

Theorem 1: For $\mu < (a+b)/2$, the NLP $\mathcal{L}(\theta)$ is minimized at the binarizing point $\sigma_a^2 = 0$ and $\sigma_b^2 = (a-b)^2$ and has no other critical points, resulting in $x_{MAP} = a$, if and only if

$$s^2 > \begin{cases} (3-\sqrt{8})(a-\mu)(b-\mu), & \text{if } \mu < a - \frac{|a-b|}{\sqrt{2}} \\ \frac{(a-\mu)^2|a-b|}{(a+b)-2\mu}, & \text{if } a - \frac{|a-b|}{\sqrt{2}} \le \mu < \frac{a+b}{2}. \end{cases}$$

Likewise, for $\mu > (a+b)/2$, the NLP $\mathcal{L}(\theta)$ is minimized at the binarizing point $\sigma_a^2 = (b-a)^2$ and $\sigma_b^2 = 0$ and has no other critical points, resulting in $x_{MAP} = b$, if and only if

$$s^2 > \begin{cases} (3-\sqrt{8})(a-\mu)(b-\mu), & \text{if } \mu > b + \frac{|a-b|}{\sqrt{2}} \\ \frac{(b-\mu)^2|a-b|}{2\mu-(a+b)}, & \text{if } \frac{a+b}{2} < \mu \le b + \frac{|a-b|}{\sqrt{2}}. \end{cases}$$

Moreover, a binarizing θ can only be a minimum of $\mathcal{L}(\theta)$ if its positive variance equals $(a-b)^2$, i.e., if $\sigma_a^2 + \sigma_b^2 = (a-b)^2$.

A proof is given in [6]. Some remarks are in order:

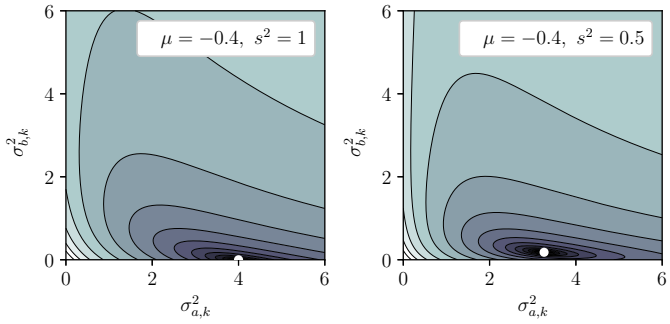
1) For any $\mu \neq (a+b)/2$, for sufficiently large s^2 (i.e., for a sufficiently noisy statistical model), the procedure naturally yields a binary estimate $x_{\text{MAP}}(\hat{\theta})$. In this respect, the method differs from methods based on convex relaxation, which typically require some sort of projection onto the constraint set at the end.
2) For any given $\mu \neq (a+b)/2$ (and sufficiently large s^2), only the constellation member closer to μ is a (global or local) minimum of $\mathcal{L}(\theta)$. In this respect, the method differs from simple concave regularization as in [7], which may induce a local optimum also at the farther constellation member.
3) The optimization problem (4) is not convex. However, thanks to the guarantees of Theorem 1, even a local optimization procedure (such as EM, see Section III) is guaranteed to converge to the desired binary solution.

The behavior is further illustrated with Figs. 2–4, all of which use $\{a, b\} = \{-1, +1\}$: Fig. 2 depicts the minimal s^2 satisfying the conditions of Theorem 1 as a function of μ. Fig. 3 depicts the estimate $x_{\text{MAP}}(\hat{\theta})$ as a function of μ for different values of s^2, also showing the method's behavior when s^2 is too small to enforce binarization. Fig. 4 shows contour plots of $\mathcal{L}(\theta)$ for a fixed observation μ and two choices of s^2, one of which is large enough for binarization while the other is not. The figure highlights the smooth profile of $\mathcal{L}(\theta)$.

III. From the Scalar Case to Linear Systems

The extension from the scalar case to general linear systems is straightforward and enables tackling the problem (1) for binary constellations $\mathcal{O} = \{a, b\}$. In this case, the prior is $\rho(\mathbf{x}, \boldsymbol{\theta}) = \prod_k \rho(x_k, \theta_k)$ with parameters (variances) $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n), \theta_k = (\sigma_{a,k}^2, \sigma_{b,k}^2)$, and with $\rho(x_k, \theta_k)$ as in (2). The likelihood changes from $p(y|x) = \mathcal{N}\big(x|\mu, s^2\big)$ to $p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\Phi\mathbf{x} \,|\, \mathbf{y}, \sigma^2\mathbb{I})$, where σ^2 is a hyperparameter.

We remark that, for fixed $\boldsymbol{\theta}$, the whole model is essentially Gaussian (up to a scale factor). This means that from any "input terminal" x_k of $\mathbf{x}$, the local observation corresponds to the model from Section II, where the local mean μ_k and variance s_k^2 are functions of $\mathbf{y}, \sigma^2$, and the remaining $x_{k'}, k' \neq k$. If $\mu_k \neq (a+b)/2$ and if the local variance s_k^2 at the input terminal (which is a monotonous function of σ^2) is large enough, we may therefore expect to observe binarization.



(a) s^2 sufficiently large. (b) s^2 too small.

Fig. 4: Contour plot of $\mathcal{L}(\theta)$ for $\mu = -0.4$ and two different choices of s^2. The white dot indicates the minimum.

The MAP estimate (4) becomes

$$\boldsymbol{\theta}_{\text{MAP}} = \operatorname*{argmax}_{\boldsymbol{\theta}} \tilde{p}(\boldsymbol{\theta}|\mathbf{y}) = \operatorname*{argmax}_{\boldsymbol{\theta}} \int_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})\rho(\mathbf{x},\boldsymbol{\theta})\, d\mathbf{x}. \quad (12)$$

Due to the Gaussian form of $\rho(\mathbf{x},\boldsymbol{\theta})$, the integral in (12) is tractable, but solving the resulting non-convex optimization problem is hard. An obvious candidate for obtaining a surrogate estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_{\text{MAP}}$ in iterative fashion is the Expectation Maximization (EM) algorithm [4], [8]. The updates are

$$\boldsymbol{\theta}^{(i+1)} = \operatorname*{argmax}_{\boldsymbol{\theta}} \mathsf{E}_{p(\mathbf{x}|\mathbf{y},\boldsymbol{\theta}^{(i)})}[\log p(\mathbf{X},\mathbf{y}|\boldsymbol{\theta})] \quad (13)$$

$$= \operatorname*{argmax}_{\boldsymbol{\theta}} \mathsf{E}_{p(\mathbf{x}|\mathbf{y},\boldsymbol{\theta}^{(i)})}[\log p(\mathbf{X}|\boldsymbol{\theta})]. \quad (14)$$

This optimization problem separates nicely into the individual $\sigma^2_{a,k}$ and $\sigma^2_{b,k}$, yielding the updates

$$\big(\sigma^2_{a,k}\big)^{(i+1)} = \mathsf{E}_{p(\mathbf{x}|\mathbf{y},\theta^{(i)})}\big[(X_k - a)^2\big] \quad (15a)$$

$$\big(\sigma^2_{b,k}\big)^{(i+1)} = \mathsf{E}_{p(\mathbf{x}|\mathbf{y},\theta^{(i)})}\big[(X_k - b)^2\big]. \quad (15b)$$

The quantities on the Right-Hand-Side (RHS) of (15) can be conveniently calculated with Gaussian message passing [9].

For a given estimate $\hat{\boldsymbol{\theta}}$, the MAP estimate of $\mathbf{x}$ is

$$\mathbf{x}_{\text{MAP}}(\hat{\boldsymbol{\theta}}) = \operatorname*{argmax}_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})\, \rho(\mathbf{x},\hat{\boldsymbol{\theta}}) \quad (16)$$

$$= \big(\Sigma^{-1}_{\hat{\boldsymbol{\theta}}} + \sigma^{-2}\Phi^{\mathsf{T}}\Phi\big)^{-1}\big(\Sigma^{-1}_{\hat{\boldsymbol{\theta}}}\boldsymbol{\mu}_{\hat{\boldsymbol{\theta}}} + \sigma^{-2}\Phi^{\mathsf{T}}\mathbf{y}\big), \quad (17)$$

where $\boldsymbol{\mu}_{\hat{\boldsymbol{\theta}}} \triangleq (\mu_{\hat{\boldsymbol{\theta}},1},\dots,\mu_{\hat{\boldsymbol{\theta}},n})^{\mathsf{T}}, \sigma^2_{\hat{\boldsymbol{\theta}}} \triangleq (\sigma_{\hat{\boldsymbol{\theta}},1},\dots,\sigma_{\hat{\boldsymbol{\theta}},n})^{\mathsf{T}}, \Sigma_{\hat{\boldsymbol{\theta}}} \triangleq \operatorname{diag}\big(\sigma^2_{\hat{\boldsymbol{\theta}}}\big)$. This MAP estimate is obtained as a byproduct ($\mathbf{m}_{\mathbf{X}}$ in Algorithm 1) of the EM algorithm. The resulting algorithm is called BiNUV-EM and outlined in Algorithm 1 ("$\odot$" and "$./$" denote pointwise multiplication and division, respectively). It may not be entirely obvious from the equations, but if the k-th component of $\boldsymbol{\theta}$ is binarizing, $\theta_k = (0,(b-a)^2)$ or $\theta_k = ((b-a)^2,0)$, then the corresponding entry of the MAP estimate $\mathbf{m}_{\mathbf{X}}[k]$ will be a or b as a result, respectively.[5]

As argued above, if σ^2 is large, then the final $\mathbf{m}_{\mathbf{X}}$ should contain (almost) binary entries. But to guarantee the syntactic correctness of our result, we ultimately apply a rounding step.

[5]Strictly speaking, the quantities (17), $\overrightarrow{\mathbf{W}}_{\mathbf{X}}, \vec{\boldsymbol{\xi}}_{\mathbf{X}}, \mathbf{V}_{\mathbf{X}}, \mathbf{m}_{\mathbf{X}}$ are undefined if any θ_k is binarizing. A mathematically rigorous treatment would consider these expressions under limits to zero. This is irrelevant in practice, as the variances obtained with Algorithm 1 will not converge to *exactly* zero in a finite number of iterations.

Algorithm 1 BiNUV-EM

1: **Input:** $\mathbf{y}, \Phi, \sigma^2, \boldsymbol{\theta}^{(0)}, T$
2: $\overleftarrow{\mathbf{W}}_{\mathbf{X}} = \sigma^{-2}\,\Phi^{\mathsf{T}}\Phi,\ \overleftarrow{\boldsymbol{\xi}}_{\mathbf{X}} = \sigma^{-2}\,\Phi^{\mathsf{T}}\mathbf{y}$
3: $(\sigma^2_{a,k})^{(0)}, (\sigma^2_{b,k})^{(0)} \leftarrow \boldsymbol{\theta}^{(0)},\ k = 1,\dots,n$
4: **for** $t = 0,\dots,T-1$ **do**
5: $\quad \boldsymbol{\gamma} = \boldsymbol{\sigma}^2_a \odot \boldsymbol{\sigma}^2_b./(\boldsymbol{\sigma}^2_a + \boldsymbol{\sigma}^2_b)$
6: $\quad \boldsymbol{\mu} = (b\boldsymbol{\sigma}^2_a + a\boldsymbol{\sigma}^2_b)./(\boldsymbol{\sigma}^2_a + \boldsymbol{\sigma}^2_a)$
7: $\quad \overrightarrow{\mathbf{W}}_{\mathbf{X}} = \operatorname{diag}(1./\boldsymbol{\gamma}),\ \vec{\boldsymbol{\xi}}_{\mathbf{X}} = \overrightarrow{\mathbf{W}}_{\mathbf{X}}\boldsymbol{\mu}$
8: $\quad \mathbf{V}_{\mathbf{X}} = (\overrightarrow{\mathbf{W}}_{\mathbf{X}} + \overleftarrow{\mathbf{W}}_{\mathbf{X}})^{-1},\ \mathbf{m}_{\mathbf{X}} = \mathbf{V}_{\mathbf{X}}(\vec{\boldsymbol{\xi}}_{\mathbf{X}} + \overleftarrow{\boldsymbol{\xi}}_{\mathbf{X}})$
9: $\quad (\sigma^2_{a,k})^{(t+1)} = (\mathbf{m}_{\mathbf{X}}[k] - a)^2 + \mathbf{V}_{\mathbf{X}}[k,k],\ k = 1,\dots,n$
10: $\quad (\sigma^2_{b,k})^{(t+1)} = (\mathbf{m}_{\mathbf{X}}[k] - b)^2 + \mathbf{V}_{\mathbf{X}}[k,k],\ k = 1,\dots,n$
11: **Output:** $\hat{\mathbf{x}} = \text{round}(\mathbf{m}_{\mathbf{X}})$

A drawback of BiNUV-EM as outlined in Algorithm 1 is the high complexity of the matrix inversion in line 8, which may be unattractive for certain practical applications. In many applications, however, the linear operator Φ exhibits a structure which makes it possible to avoid the calculation of this inverse, as for instance in linear state-space models [4], [5].

A. Towards larger constellations

So far, the entire discussion pertains to binary constellations. We now discuss how the method can also be used for larger constellations. While it is tempting to simply extend the binary prior (2) to $\rho(x_k,\theta_k) = \prod_{q\in\mathcal{O}} \mathcal{N}(x_k|q,\sigma^2_q)$, this does not work well: It leads to uneven preference of the constellation points, with edge points discouraged and central points preferred. Instead, we propose to restrict ourselves to constellations of the form $\mathcal{O} = \{c + \ell\cdot d\ :\ \ell = 0,\dots,L\}$. Any $x \in \mathcal{O}$ can be represented as $x = c + \sum_{\ell=1}^{L} x^{(\ell)}$, where the $x^{(\ell)}$ take value in $\{0,d\}$. This representation is not unique, which is unproblematic as the tuple $(c, x^{(1)},\dots,x^{(L)})$ is of interest only via its sum x. In fact, the redundant representation makes inference more robust. The original problem (1) can then be restated as

$$\operatorname*{argmin}_{\tilde{\mathbf{x}}\in\{0,d\}^{nL}} \|\tilde{\mathbf{y}} - \tilde{\Phi}\tilde{\mathbf{x}}\|^2, \quad (18)$$

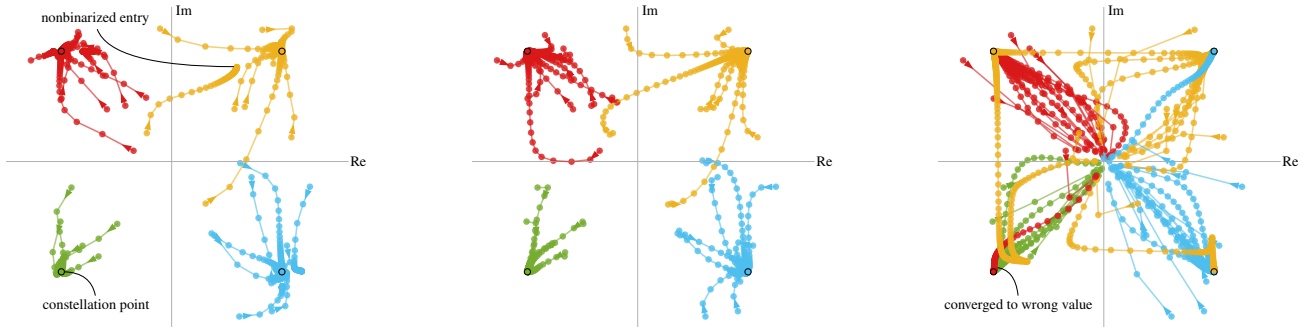
where $\tilde{\mathbf{y}} \triangleq \mathbf{y} - c\sum_{k=1}^{n}\boldsymbol{\phi}_k$, and $\tilde{\Phi} \triangleq \Phi \otimes \mathbf{1}_{1\times L}$, yielding an nL-dimensional binary input estimation problem. The peculiar structure of $\tilde{\Phi}$ (L identical copies of every column) should be exploited for simplifying the computations of the EM algorithm, but poses no problem, as the measurement matrix is not required to have distinct columns. The remaining question is how to break the symmetry in the iterative algorithm. We propose to do this by initializing the variances $(\sigma^{2,(\ell)}_{0,k}, \sigma^{2,(\ell)}_{d,k})$ with different values for the different ℓ, which works well.

IV. Application to MU-MIMO Detection

We now discuss the application of our method to the MU-MIMO detection problem. In this problem, an m-antenna base station observes the vector

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{z}, \quad (19)$$

where $\Phi \in \mathbb{C}^{m\times n}$ is the channel matrix, $\mathbf{x}$ is the $\mathcal{O}_{\mathbb{C}}$-valued vector of complex symbols transmitted by n users, and $\mathbf{z} \overset{\text{IID}}{\sim} \mathcal{N}_{\mathbb{C}}(0,\mathsf{N}_0)$ is noise. Φ is assumed to be known at the base



(a) $\sigma^2 = \mathsf{N}_0$ and $T = 80$ iterations. (b) $\sigma^2 = 5\mathsf{N}_0$ and $T = 80$ iterations. (c) $\sigma^2 = 100\mathsf{N}_0$ and $T = 40$'000 iterations.

Fig. 5: Trajectory of the posterior means $\mathbf{x}_{\text{MAP}}(\hat{\boldsymbol{\theta}})$ ($= \mathbf{m}_{\mathbf{X}}$ in Algorithm 1) for the iterative EM estimates $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(i)}$, $i = 1, 2, \ldots$.

station. For applying our method, we require that $\mathcal{O}_{\mathbb{C}}$ is a QAM constellation and transform the problem to the real domain via

$$\mathbf{y}_{\mathbb{R}} = \begin{bmatrix} \Re(\mathbf{y}) \\ \Im(\mathbf{y}) \end{bmatrix}, \quad \Phi_{\mathbb{R}} = \begin{bmatrix} \Re(\Phi) & -\Im(\Phi) \\ \Im(\Phi) & \Re(\Phi) \end{bmatrix}, \quad \mathbf{x}_{\mathbb{R}} = \begin{bmatrix} \Re(\mathbf{x}) \\ \Im(\mathbf{x}) \end{bmatrix}. \tag{20}$$

We then consider the problem $\min_{\mathbf{x}_{\mathbb{R}} \in \mathcal{O}^{2n}} \|\mathbf{y}_{\mathbb{R}} - \Phi_{\mathbb{R}}\mathbf{x}_{\mathbb{R}}\|_2^2$, where $\mathcal{O} = \Re(\mathcal{O}_{\mathbb{C}})$. If the constellation $\mathcal{O}_{\mathbb{C}}$ is QPSK, the entries of $\mathbf{x}_{\mathbb{R}}$ are representable by independent binary variables. If $\mathcal{O}_{\mathbb{C}}$ is 16QAM, they can be represented independently as sums of three binary variables. After making a choice for σ^2, setting the initialization point[6] $\boldsymbol{\theta}^{(0)}$ and the number of iterations T, one can execute BiNUV-EM on $(\mathbf{y}_{\mathbb{R}}, \Phi_{\mathbb{R}}, \sigma^2, \boldsymbol{\theta}^{(0)}, T)$.

We note that the computational complexity of BiNUV-EM as outlined in Algorithm 1 is not feasible for MU-MIMO detection in practice. However, the focus of our current study lies primarily in evaluating the performance of the composite NUV prior on a practical recovery problem. We note that, for the application of MU-MIMO detection, experiments suggest that an approximation of Algorithm 1 based on iterative scalar message-passing can be used for reducing the complexity without significantly affecting the performance.

A. How to choose σ^2

Theorem 1 shows that the proposed method enforces binarization as long as the variances at the local terminals are large enough. Since these variances are monotonously increasing in σ^2, it may seem obvious that one should choose σ^2 very large. But this would be a premature conclusion—there are two good reasons for not choosing σ^2 too large:

The first reason is that EM needs longer to converge for large σ^2, so a large σ^2 increases the computational burden.

The second reason is subtler: It can be shown that a binarizing $\boldsymbol{\theta}$ with $\{0, (a-b)^2\}$-valued components θ_k is a local minimum of $\mathcal{L}(\boldsymbol{\theta})$ if and only if for every $k \in [1:n]$,

$$\boldsymbol{\phi}_k^{\mathsf{T}}\left[\boldsymbol{\phi}_k - \frac{2}{c_{+,k} - c_{0,k}}(\mathbf{y} - \Phi\boldsymbol{\mu}_{\boldsymbol{\theta}})\right] \geq \frac{[\boldsymbol{\phi}_k^{\mathsf{T}}(\mathbf{y} - \Phi\boldsymbol{\mu}_{\boldsymbol{\theta}})]^2}{\sigma^2}, \tag{21}$$

where $c_{+,k} - c_{0,k}$ should be understood as either $b - a$ (if $(\sigma_{a,k}^2, \sigma_{b,k}^2) = (0, (b-a)^2)$), or (if $(\sigma_{a,k}^2, \sigma_{b,k}^2) = ((b-a)^2, 0)$) as $a - b$. It is evident that this criterion is the stricter, the smaller σ^2 is. As a consequence, while a too small σ^2 may not enforce the desired binarization, it seems that an overly large σ^2 may lead to an increased number of (spurious) local minima at binarizing points $\boldsymbol{\theta}$.

This behaviour is illustrated in Fig. 5 at the example of a MIMO detection problem with $n = m = 32$. The constellation is QPSK, the entries of Φ are drawn IID from $\mathcal{N}_{\mathbb{C}}\big(0, \frac{1}{m}\big)$. The signal-to-noise-ratio (SNR) is 12dB. All subfigures of Fig. 5 represent the same problem instance. They show the trajectory of the MAP estimate (17) for $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(i)}$, $i = 1, 2, \ldots$ obtained in the EM iterations (14), starting from the LMMSE configuration. The colors correspond to the different true input values. In Fig. 5(c), only the iterations $1, 4, 9, 16, \ldots$ are displayed.

In Fig. 5(a), σ^2 equals the physical noise energy N_0. After 80 iterations the algorithm has converged and most—but not all—entries of $\mathbf{x}$ have binarized (i.e., have converged to constellation points). But after rounding, the estimate is error-free.

In Fig. 5(b), σ^2 is five times larger than N_0. After 80 iterations the algorithm has converged, and here all entries of $\mathbf{x}$ have binarized. The estimate contains no errors.

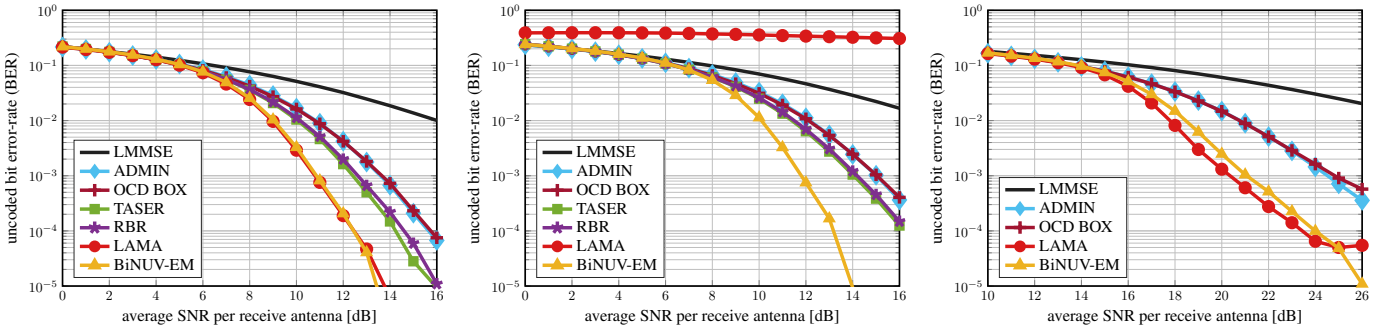
In Fig. 5(c), σ^2 is much larger than N_0. EM now needs roughly 40'000 iterations to converge. And while all entries of $\mathbf{x}$ have binarized, some have converged to the wrong value—the algorithm has converged to a spurious local minimum.

The optimal value of σ^2 for a given setting can be determined experimentally. The choice of a good σ^2 appears to be quite robust to changes of the SNR (e.g., $\sigma^2 = 2\mathsf{N}_0$ works well for a wide range of SNR) or the problem dimension.

V. Empirical Evaluation

We evaluate BiNUV-EM empirically in comparison with existing approaches for MU-MIMO detection. The channel model is as in (19), and the matrix Φ is either drawn IID from $\mathcal{N}_{\mathbb{C}}\big(0, \frac{1}{m}\big)$, or it is constructed according to the Jakes model [10] (with antennas spaced at half the wavelength) for correlated channels: In that case, $\Phi = R_m^{\frac{1}{2}}\Phi_{\text{IID}}R_n^{\frac{1}{2}}$, where $\Phi_{\text{IID}} \overset{\text{IID}}{\sim} \mathcal{N}_{\mathbb{C}}\big(0, \frac{1}{m}\big)$ and where R_m, R_n are given by $[R_m]_{i,j} = [R_n]_{i,j} = J_0(|i-j| \cdot \pi)$, with J_0 being the zero-order first-kind Bessel function. The number of users and of base station antennas are both set to 32 (but the method works also for under- and overdetermined problems). For QPSK, we set $\sigma^2 = 2\mathsf{N}_0$, and for 16QAM, we set $\sigma^2 = \mathsf{N}_0$. We use $T = 50$ iterations.

[6] The initializtion point $\boldsymbol{\theta}^{(0)}$ can be chosen such that the first iteration of Algorithm 1 corresponds to the LMMSE estimate.



(a) QPSK, 32x32 system, IID matrices Φ. (b) QPSK, 32x32 system, Jakes model. (c) 16QAM, 32x32 system, IID matrices Φ.

Fig. 6: Uncoded bit error-rate (BER) vs. average signal-to-noise ratio (SNR) for different MIMO detection methods.

We compare BiNUV-EM with an LMMSE baseline and methods that rely on convex relaxation: the semidefinite relaxation (SDR) based methods TASER [11] and RBR [12], and the box-constrained methods ADMIN [13] and OCD BOX [14]. We also compare with the approximate message passing (AMP) based algorithm LAMA [15], which is provably optimal for large IID Gaussian matrices.

Fig. 6 shows the uncoded bit error-rate (BER) as a function of the average signal-to-noise ratio (SNR) for three different setups: Figs. 6(a) and 6(b) use a QPSK constellation, where Fig. 6(a) uses IID channel matrices and Fig. 6(b) uses the correlated channel matrices of the Jakes model. Fig. 6(c) uses a 16QAM constellation and IID channel matrices.

BiNUV-EM outperforms the convex-relaxation-based methods in all the settings. (TASER and RBR do not support larger constellations than QPSK.) For QPSK in an IID Gaussian setting (Fig. 6(a)), it performs on par with LAMA. For the correlated channel matrices (Fig. 6(b)) of the Jakes model, LAMA breaks down completely (even with damping), while BiNUV-EM continues to perform well. When using the approach of Section III-A for larger constellations such as 16QAM (Fig. 6(c)), BiNUV-EM loses 0.5dB at BER=10^{-3} compared to LAMA, but clearly outperforms the box-constrained methods. Note that BiNUV-EM—unlike LAMA—shows no signs of an error floor. Furthermore, switching to correlated channel matrices would lead to a breakdown of LAMA as in Fig. 6(b).

We conclude that BiNUV-EM can show excellent empirical performance in binary recovery problems, and that our proposed approach to larger constellations also works quite well, at least for constellations of moderate size.

VI. Conclusion

We have provided an analysis of the method proposed in [4] for the estimation of binary input signals that explains *why* the method produces binary estimates. We have then adopted the method to the MU-MIMO detection problem and compared it with existing methods. We found that it outperforms methods based on convex relaxations and that it works better for correlated matrices than an AMP-based method. Future work will be aimed at exploring and improving the complexity/performance tradeoff of composite-NUV based methods for MU-MIMO detection to obtain a practical algorithm.

Acknowledgement

We thank Oscar Castañeda, Charles Jeon, and Christoph Studer for helpful discussions and for providing their simulators to us.

References

[1] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*. Springer, 2012, vol. 2.

[2] H.-A. Loeliger, B. Ma, H. Malmberg, and F. Wadehn, "Factor graphs with NUV priors and iteratively reweighted descent for sparse least squares and more," in *Int. Symposium on Turbo Codes & Iterative Information Processing (ISTC) 2018, Hongkong, China*, Dec. 3–7, 2018.

[3] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.

[4] R. Keusch, H. Malmberg, and H.-A. Loeliger, "Binary control and digital-to-analog conversion using composite NUV priors and iterative Gaussian message passing," in *Procs. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 5330–5334.

[5] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295–1322, 2007.

[6] R. Keusch and H.-A. Loeliger, "A binarizing NUV prior and its use for M-level control and digital-to-analog conversion," arXiv:2105.02599.

[7] Y. Meslem, A. Aïssa-El-Bey, and M. Djeddou, "Large-scale MIMO receiver based on finite-alphabet sparse detection and concave-convex optimization," in *Procs. IEEE 21st Int. Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2020, pp. 1–5.

[8] P. Stoica and Y. Selen, "Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: a refresher," *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 112–114, 2004.

[9] H.-A. Loeliger, L. Bruderer, H. Malmberg, F. Wadehn, and N. Zalmai, "On sparsity by NUV-EM, Gaussian message passing, and Kalman smoothing," in *2016 Information Theory & Applications Workshop (ITA), La Jolla, CA*, Jan. 31 Feb. 5, 2016.

[10] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2636–2647, 2003.

[11] O. Castaneda, T. Goldstein, and C. Studer, "Data detection in large multi-antenna wireless systems via approximate semidefinite relaxation," *IEEE Trans. on Circuits and Systems I: Regular Papers*, vol. 63, no. 12, pp. 2334–2346, 2016.

[12] H.-T. Wai, W.-K. Ma, and A. M.-C. So, "Cheap semidefinite relaxation MIMO detection using row-by-row block coordinate descent," in *Procs. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 3256–3259.

[13] S. Shahabuddin, M. Juntti, and C. Studer, "ADMM-based infinity norm detection for large MU-MIMO: Algorithm and VLSI architecture," in *Procs. IEEE Int. Symp. on Circuits and Systems (ISCAS)*, 2017, pp. 1–4.

[14] M. Wu, C. Dick, J. R. Cavallaro, and C. Studer, "High-throughput data detection for massive MU-MIMO-OFDM using coordinate descent," *IEEE Trans. on Circuits and Systems I: Regular Papers*, vol. 63, no. 12, pp. 2357–2367, 2016.

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