

DETECTION OF PILOT-HOPPING SEQUENCES FOR GRANT-FREE RANDOM ACCESS IN MASSIVE MIMO SYSTEMS

Ema Becirovic, Emil Björnson and Erik G Larsson

Book Chapter

Cite this chapter as:

Becirovic, E., Björnson, E., G, E. DETECTION OF PILOT-HOPPING SEQUENCES FOR GRANT-FREE RANDOM ACCESS IN MASSIVE MIMO SYSTEMS, In (eds), 2019 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), : IEEE; 2019, pp. 8380-8384. ISBN: 978-1-4799-8131-1

International Conference on Acoustics Speech and Signal Processing ICASSP, 1520-6149, No.

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-160640>

Copyright: IEEE

<http://www.ieee.org/>

Copyright: Institute of Electrical and Electronics Engineers (IEEE) <http://www.ieee.org/index.html>
©2018 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. <http://www.ieee.org/index.html>



DETECTION OF PILOT-HOPPING SEQUENCES FOR GRANT-FREE RANDOM ACCESS IN MASSIVE MIMO SYSTEMS

Ema Becirovic, Emil Björnson and Erik G. Larsson

Dept. of Electrical Engineering, Linköping University, Linköping, Sweden
Email: {ema.becirovic, emil.bjornson, erik.g.larsson}@liu.se

ABSTRACT

In this paper, we study an active user detection problem for massive machine type communications (mMTC). The users transmit pilot-hopping sequences and detection of active users is performed based on the received energy. We utilize the channel hardening and favorable propagation properties of massive multiple-input multiple-output (MIMO) to simplify the user detection. We propose and compare a number of different user detection methods and find that using non-negative least squares (NNLS) is well suited for the task at hand as it achieves good results as well as having the benefit of not having to specify further parameters.

Index Terms— massive MIMO, machine type communication, compressed sensing.

1. INTRODUCTION

In massive machine type communications (mMTC) there is a huge number of users that sporadically transmit small amounts of data to the base station [1]. This means that the number of active users at a given time instant is typically much smaller than the total number of users in the system. Moreover, the users that we consider are not concerned about latency as the data they are transmitting is not critical. An example of these types of devices are simple internet-of-things (IoT) sensors.

The mMTC users do not want to spend many resources on overhead signaling such as random access or requesting to be scheduled as they only transmit small amounts of data. Therefore, these users will benefit from using a grant-free access scheme. Grant-free random access for mMTC with massive MIMO has been studied in many papers, for instance in [2–4], and different techniques are summarized in [5].

To completely avoid pilot contamination in massive MIMO, all the users need to be assigned mutually orthogonal pilots during training. However, in the mMTC case, the number of users is far too large to be assigned to orthogonal pilots in advance and temporary assignment of orthogonal pilots is not possible in grant-free access. To resolve this issue we can let the users have unique non-orthogonal pilots or let the users transmit the pilots in unique pilot-hopping sequences during many coherence intervals.

The notion that the users transmit pilot-hopping sequences for activity detection is introduced in [6], however the activity detection is, in that paper, assumed to be perfect. In this paper, we introduce an efficient way to solve this problem as well as analyze the performance of the method. We assume that the data transmission is only in the uplink and that the latency requirements are lenient

User 1	ϕ_2	$D_1(1)$	ϕ_1	$D_1(2)$	ϕ_1	$D_1(3)$	ϕ_2	$D_1(4)$
User 2	ϕ_1	$D_2(1)$	ϕ_1	$D_2(2)$	ϕ_2	$D_2(3)$	ϕ_2	$D_2(4)$
User 3	ϕ_1	$D_3(1)$	ϕ_2	$D_3(2)$	ϕ_1	$D_3(3)$	ϕ_1	$D_3(4)$

Fig. 1: Three users transmitting pilot-hopping sequences of length four using two pilots, ϕ_1 and ϕ_2 . Additionally, in coherence interval t the user k also transmits uplink data $D_k(t)$.

which means that we can postpone the channel estimation to the active users until the end of the pilot-hopping sequences. Each coherence interval (τ_c samples) in the proposed scheme is split into two parts: the uplink pilot phase (τ_p samples) and the uplink data phase ($\tau_c - \tau_p$ samples). This is unlike canonical massive MIMO systems where each coherence interval also has a downlink data phase [7]. In Fig. 1, an example of four coherence intervals of three users using two orthogonal pilots, ϕ_1 and ϕ_2 , is pictured in the proposed scheme. The figure also depicts how the users transmit data; user k transmits data $D_k(t)$ in coherence interval t .

1.1. Related Works

Related works on detecting active users in grant-free random access with massive MIMO study one coherence interval at a time. The AMP algorithm from [8] is demonstrated in [2] to be a promising technique for active user detection in mMTC. In [2] pilots are generated randomly and active user detection is performed by the AMP algorithm. Further, [3] studies how a few number of bits can be embedded in the pilot sequences for non-coherent data transmission. In [4], the authors consider other compressed sensing techniques such as non-negative least squares (NNLS) and derive a scaling law on parameters required for reliable detection of active users.

1.2. Contributions

In this work, we consider a setup with multiple coherence intervals where the users transmit pilot-hopping sequences as introduced in [6]. We observe that with the channel hardening and favorable propagation properties of massive MIMO, the problem can be simplified and solved with compressed sensing techniques. We propose a novel user detection algorithm that performs very well for the task at hand.

2. SYSTEM MODEL

We consider a single-cell system where the base station is equipped with M antennas and serves K users. However, not all of these

This work was supported in part by ELLIIT and in part by the Swedish Research Council (VR).

users are active. We assume a block fading model where in each coherence interval the channel is assumed to be time invariant and frequency flat. There are τ_p orthogonal pilots to choose from and the pilot-hopping sequences are T coherence intervals long. At the pilot phase of coherence interval t , the base station receives

$$\mathbf{Y}^t = \sum_{k=1}^K \sum_{j=1}^{\tau_p} \alpha_k S_{j,k}^t \sqrt{\tau_p p_k} \mathbf{g}_k^t \phi_j^H + \mathbf{N}^t, \quad t = 1, \dots, T \quad (1)$$

where

$$\alpha_k = \begin{cases} 1 & \text{if user } k \text{ is active,} \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

$$S_{j,k}^t = \begin{cases} 1 & \text{if user } k \text{ sends pilot } j \text{ at pilot phase } t, \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

p_k is the transmit power of user k , $\mathbf{g}_k^t \in \mathbb{C}^M$ is the channel between user k and the base station in coherence interval t , and $\phi_j \in \mathbb{C}^{\tau_p}$ is the j :th pilot consisting of τ_p symbols. The pilots are mutually orthogonal, $\phi_i^H \phi_j = 0, i \neq j$, and have unit norm, $\|\phi_i\| = 1$. Finally, $\mathbf{N}^t \in \mathbb{C}^{M \times \tau_p}$ is noise with i.i.d. $\mathcal{CN}(0, \sigma^2)$ elements. The model assumes that the users are synchronized in the sense that all the active users start their pilot-hopping sequence at the same coherence interval.

In each coherence interval, the base station computes an estimate of the received signal energy over each pilot as

$$E_{i,t} = \frac{(\mathbf{Y}^t \phi_i)^H (\mathbf{Y}^t \phi_i)}{M} - \sigma^2 = \frac{\|\mathbf{Y}^t \phi_i\|^2}{M} - \sigma^2. \quad (4)$$

Assuming a block independent Rayleigh fading channel with large-scale fading β_k , $\mathbf{g}_k^t \sim \mathcal{CN}(\mathbf{0}, \beta_k \mathbf{I})$ we can use the properties of channel hardening, $\frac{\|\mathbf{g}_k^t\|^2}{M} \rightarrow \beta_k$, as $M \rightarrow \infty$, and favorable propagation, $\frac{(\mathbf{g}_k^t)^H (\mathbf{g}_{k'}^t)}{M} \rightarrow 0$, as $M \rightarrow \infty, k \neq k'$, to see that

$$E_{i,t} = \frac{\|\mathbf{Y}^t \phi_i\|^2}{M} - \sigma^2 \rightarrow \sum_{k=1}^K \alpha_k S_{i,k}^t \tau_p p_k \beta_k \quad \text{as } M \rightarrow \infty. \quad (5)$$

In practice, the number of antennas M will be finite but these limits can be used as to obtain a simpler design. The limits are normally tight at around 50 antennas [9, p. 192].

Our single-cell model also gives accurate results for multi-cell systems. When considering a multi-cell scenario, using orthogonal pilots is not feasible as the coherence interval is limited. Considering a reuse factor of one, the effect of pilot contamination from neighboring cells will aggregate in the target cell which might impair the user detection (and channel estimation quality). However, considering a large enough pilot reuse factor the effects of pilot contamination from other cells will be relatively small and therefore not affect the target cell much. In this case, the product of the reuse factor and the number of orthogonal pilots would still have to be smaller than the coherence interval.

3. DETECTING ACTIVE USERS USING THE ASYMPTOTIC ENERGIES

In this section, we solve the user detection problem by studying the asymptotic energies in (5). We place all the received energies over the T pilot phases and τ_p pilots in a vector

$$\mathbf{y} = (E_{1,1} \dots E_{\tau_p,1} \dots E_{1,t} \dots E_{1,T} \dots E_{\tau_p,T})^T. \quad (6)$$

Using terminology from compressed sensing, we also create a $\tau_p T \times K$ measurement matrix of the asymptotic behavior of the energies:

$$\mathbf{A} = \begin{pmatrix} S_{1,1}^1 \tau_p p_1 \beta_1 & \dots & S_{1,K}^1 \tau_p p_K \beta_K \\ \vdots & \ddots & \vdots \\ S_{\tau_p,1}^1 \tau_p p_1 \beta_1 & \dots & S_{\tau_p,K}^1 \tau_p p_K \beta_K \\ \vdots & \ddots & \vdots \\ \vdots & S_{i,k}^t \tau_p p_k \beta_k & \vdots \\ \vdots & \ddots & \vdots \\ S_{1,1}^T \tau_p p_1 \beta_1 & \dots & S_{1,K}^T \tau_p p_K \beta_K \\ \vdots & \ddots & \vdots \\ S_{\tau_p,1}^T \tau_p p_1 \beta_1 & \dots & S_{\tau_p,K}^T \tau_p p_K \beta_K \end{pmatrix}. \quad (7)$$

There are τ_p^T unique pilot-hopping sequences, therefore, the number of users must be $K \leq \tau_p^T$. However, the product of the number of pilots and the sequence length is most likely smaller than the number of users: $\tau_p T \leq K$. Therefore, the measurement matrix \mathbf{A} is wide which means that we have more unknowns than measurements.

The measurement matrix is sparse when orthogonal pilots are used and only T elements per column will be non-zero. This stems from the fact that one user can only transmit one pilot at a time.

To detect the active users, we use the fact that in the asymptotic case when M is large \mathbf{y} is a linear combination of the user activities:

$$\mathbf{y} \rightarrow \mathbf{A} \boldsymbol{\alpha} \quad \text{as } M \rightarrow \infty, \quad (8)$$

where $\boldsymbol{\alpha} = (\alpha_1 \dots \alpha_k \dots \alpha_K)^T$. This comes from (5), i.e., the channel hardening and favorable propagation properties of the channel. In the non-asymptotic case this will only hold approximately hence our goal is to minimize the error between the asymptotic energies and the observed energies.

The fact that we have a sparse input signal $\boldsymbol{\alpha}$ and a wide measurement matrix \mathbf{A} indicates that we should consider using compressed sensing techniques. We start by posing the following problem:

$$\underset{\boldsymbol{\alpha} \in \{0,1\}^K}{\operatorname{argmin}} \lambda \|\boldsymbol{\alpha}\|_0 + \|\mathbf{A} \boldsymbol{\alpha} - \mathbf{y}\|_2^2 \quad (9)$$

where the first term is imposing sparsity in the solution with a weight of $\lambda \geq 0$ and the second term aims to minimize the error. This problem is combinatorial and hard to solve when there are many users, i.e., when K is large. Therefore, we relax the problem to the non-negative LASSO (NLASSO)

$$\underset{\boldsymbol{\alpha} \geq 0}{\operatorname{argmin}} \lambda \|\boldsymbol{\alpha}\|_1 + \|\mathbf{A} \boldsymbol{\alpha} - \mathbf{y}\|_2^2 \quad (10)$$

where instead the sparsity is induced by the ℓ_1 -norm. To detect the users, the elements of the resulting $\boldsymbol{\alpha}$ are compared to a threshold.

It is known that for measurement matrices, \mathbf{A} , that fulfil the self-regularizing property [10, Condition 1] the ℓ_1 -regularization term in the NLASSO problem is not needed to induce sparsity. Hence, our problem can also be posed as the NNLS problem

$$\underset{\boldsymbol{\alpha} \geq 0}{\operatorname{argmin}} \|\mathbf{A} \boldsymbol{\alpha} - \mathbf{y}\|_2^2 \quad (11)$$

where we benefit from not having to choose a regularization parameter λ . The relation between the NLASSO and the NNLS is discussed in [10–13]. In essence, both the methods can be related to the ℓ_1 -squared non-negative regularization.

4. MAXIMUM LIKELIHOOD DETECTION OF ACTIVE USERS

To not have to rely on asymptotic approximations, in this section, we derive the maximum likelihood (ML) detector of the user detection problem. If we despread the received signal in (1) at time t onto the pilot i we get

$$\mathbf{Y}^t \boldsymbol{\phi}_i = \sum_{k=1}^K \alpha_k S_{i,k}^t \sqrt{\tau_p p_k} \mathbf{g}_k^t + \mathbf{N}^t \boldsymbol{\phi}_i \quad (12)$$

which conditioned on the activity of the users $\boldsymbol{\alpha}$ is distributed as

$$\mathbf{Y}^t \boldsymbol{\phi}_i | \boldsymbol{\alpha} \sim \mathcal{CN} \left(\mathbf{0}, \left(\sum_{k=1}^K \alpha_k S_{i,k}^t \tau_p p_k \beta_k + \sigma^2 \right) \mathbf{I} \right). \quad (13)$$

The norm squared of the despread signal is distributed as follows:

$$\|\mathbf{Y}^t \boldsymbol{\phi}_i\|^2 | \boldsymbol{\alpha} \sim \left(\sum_{k=1}^K \alpha_k S_{i,k}^t \tau_p p_k \beta_k + \sigma^2 \right) \frac{1}{2} \chi_{2M}^2. \quad (14)$$

The squared norms are parts of the vector \mathbf{y} defined in (6). We define a new vector

$$\tilde{\mathbf{y}} = \left(\|\mathbf{Y}^1 \boldsymbol{\phi}_1\|^2 \dots \|\mathbf{Y}^1 \boldsymbol{\phi}_{\tau_p}\|^2 \dots \|\mathbf{Y}^t \boldsymbol{\phi}_i\|^2 \dots \|\mathbf{Y}^T \boldsymbol{\phi}_1\|^2 \dots \|\mathbf{Y}^T \boldsymbol{\phi}_{\tau_p}\|^2 \right)^T \quad (15)$$

$$= M(\mathbf{y} + \sigma^2 \mathbf{1}). \quad (16)$$

The elements of $\tilde{\mathbf{y}}$ are conditionally independent given $\boldsymbol{\alpha}$ as all the sources of randomness are independent for different coherence intervals t and pilots i .¹ Because of the independence, the joint probability density function is the product of the probability density functions of each element. We let \mathbf{a}_r^T denote the r :th row of the matrix \mathbf{A} defined in (7). The probability density function of the vector $\tilde{\mathbf{y}}$ is then

$$p(\tilde{\mathbf{y}} | \boldsymbol{\alpha}) = \prod_{r=1}^{\tau_p T} \frac{1}{\Gamma(M)(\mathbf{a}_r^T \boldsymbol{\alpha} + \sigma^2)^M} \tilde{y}_r^{M-1} \times \exp \left(-\frac{\tilde{y}_r}{\mathbf{a}_r^T \boldsymbol{\alpha} + \sigma^2} \right). \quad (17)$$

Hence, the ML detector is

$$\begin{aligned} & \operatorname{argmax}_{\boldsymbol{\alpha} \in \{0,1\}^K} p(\tilde{\mathbf{y}} | \boldsymbol{\alpha}) \\ &= \operatorname{argmin}_{\boldsymbol{\alpha} \in \{0,1\}^K} \sum_{r=1}^{\tau_p T} M \ln(\mathbf{a}_r^T \boldsymbol{\alpha} + \sigma^2) + \frac{\tilde{y}_r}{\mathbf{a}_r^T \boldsymbol{\alpha} + \sigma^2}. \end{aligned} \quad (18)$$

We note that finding the ML detector is a difficult problem without doing an exhaustive search.

5. MAXIMUM A POSTERIORI DETECTION OF ACTIVE USERS

In this section, we assume a prior distribution on the user activity and derive the maximum a posteriori (MAP) detector of the user detection problem. If we assume that the activity of each user is

¹However, if the pilots were not orthogonal the elements of $\tilde{\mathbf{y}}$ would only be independent over time but not over pilots.

Table 1: The simulation parameters used unless anything else is stated.

Parameter	Value	Parameter	Value
K	1000	SNR	0 dB
M	70	τ_p	10
P_α	0.01	T	10

independent and Bernoulli distributed with a P_α probability of being active, i.e.,

$$p(\boldsymbol{\alpha}) = \prod_{k=1}^K P_\alpha^{\alpha_k} (1 - P_\alpha)^{1 - \alpha_k}, \quad (19)$$

we can state the MAP detector

$$\begin{aligned} & \operatorname{argmax}_{\boldsymbol{\alpha} \in \{0,1\}^K} p(\tilde{\mathbf{y}} | \boldsymbol{\alpha}) p(\boldsymbol{\alpha}) \\ &= \operatorname{argmin}_{\boldsymbol{\alpha} \in \{0,1\}^K} \sum_{r=1}^{\tau_p T} M \ln(\mathbf{a}_r^T \boldsymbol{\alpha} + \sigma^2) + \frac{\tilde{y}_r}{\mathbf{a}_r^T \boldsymbol{\alpha} + \sigma^2} \\ & \quad + \ln \left(\frac{1 - P_\alpha}{P_\alpha} \right) \mathbf{1}^T \boldsymbol{\alpha}. \end{aligned} \quad (20)$$

Finding the MAP detector is, just as finding the ML detector, difficult without doing an exhaustive search. Note that the difference between (18) and (20) is a scaled summation of $\alpha_1, \dots, \alpha_K$ which can in the binary case be seen as the zero-norm of $\boldsymbol{\alpha}$ or in the non-negative case be seen as the ℓ_1 -norm of $\boldsymbol{\alpha}$. This implies that the MAP detector enforces sparser solutions than the ML detector.

6. SIMULATIONS

In this section, we present some numerical results to illustrate the performance of our proposed detection methods using asymptotic energies, (10) and (11). We compare to the performance of the statistical methods in (18) and (20). The simulation parameters, unless otherwise stated, can be found in Table 1. The users' pilot-hopping sequences are chosen uniformly at random from all the possible τ_p^T sequences. The users perform power control in the form of statistical channel inversion [9]

$$p_k = p \frac{\min_{k'} \beta_{k'}}{\beta_k} = p \frac{\beta_{\min}}{\beta_k}, \quad (21)$$

where $\beta_{\min} = \min_{k'} \beta_{k'}$ is the large-scale fading of the user with the worst channel and p is a system wide power parameter. The aim with this power control scheme is that the received power at each of the base station antennas from each user should be equal, $p_k \beta_k = p \frac{\beta_{\min}}{\beta_k} \beta_k = p \beta_{\min}$. The signal-to-noise ratio is defined as $\text{SNR} = \frac{p \beta_{\min}}{\sigma^2}$.

The probability of missed detection, p_m , and the probability of false alarm, p_{fa} , are defined as

$$p_m = \frac{\#\text{undetected active users}}{\#\text{active users}} \quad \text{and} \quad p_{fa} = \frac{\#\text{detected inactive users}}{\#\text{inactive users}},$$

respectively.

The NNLS is solved by an implementation of the Lawson-Hanson algorithm [14]. The curves denoted by $\overline{\text{ML}}$ and $\overline{\text{MAP}}$ are suboptimal solutions to problems (18) and (20) where the binary

Table 2: References to the detectors used in the simulations.

Detector	Equation	Detector	Equation
$\overline{\text{ML}}$	suboptimal (18)	NLASSO	(10)
$\overline{\text{MAP}}$	suboptimal (20)	NNLS	(11)
Greedy iterative	(22), (23)		

constraints have been relaxed to non-negative constraints. The non-convex optimization problems are solved with a gradient descent to find a local optimum. The gradient descent step size is found with a backtracking algorithm and the best of 50 solutions is used to increase the probability of finding the global optimum. The NLASSO is solved with an implementation in CVX [15] with the MOSEK solver [16]. The *greedy iterative* algorithm works by repetitively detecting the user with the highest accumulated energy,

$$\hat{k} = \underset{k}{\operatorname{argmax}} \sum_{t=1}^T \sum_{i=1}^{\tau_p} E_{i,t} S_{i,k}^t, \quad (22)$$

and removing the corresponding asymptotic energy,

$$E_{i,t} := E_{i,t} - S_{i,\hat{k}}^t \tau_p p_{\hat{k}} \beta_{\hat{k}}, \quad (23)$$

until the remaining energy falls below a threshold. The references to the respective equations for the detectors are summarized in Table 2.

From Fig. 2 we can see that when the number of antennas is small the NNLS method does not perform well compared to the sub-optimal ML and MAP. The reason for this is that the approximation made by taking M to the limit will not hold. When M is larger, the NNLS will perform better. This is because the ML and MAP are solved suboptimally due to the prohibitive complexity of an exhaustive search. We can also conclude that the greedy iterative algorithm is not well suited for this problem. The reason for this is that users with colliding pilots will have higher accumulated energies which can lead to false alarm.

In Fig. 3, we can compare NLASSO with NNLS. In this simulation the NNLS is solved like the NLASSO but with the regularization parameter λ set to 0. We see that the regularization parameter is not needed and that \mathbf{A} has self-regularizing properties. This is holds regardless of the number of antennas M . Although not shown in the figure, the same results can be seen with larger regularization parameters λ .

Fig. 4 shows that increasing the number of orthogonal pilots and the length of the sequence significantly improves the performance. Increasing the number of pilots or the length of the pilot-hopping sequence will make the matrix \mathbf{A} taller.

7. CONCLUSIONS

In this paper we studied an active user detection problem for mMTC in massive MIMO. The scheme is based on detecting pilot-hopping sequences by looking at the received energy. We saw that when there is a large number of base station antennas we can utilize the channel hardening and favorable propagation properties to simplify the problem which can then be solved with compressed sensing techniques. We observed that the non-negative least squares approach is suitable to solve the problem as it performs well and benefits from not having to set additional parameters, e.g., regularization parameters.

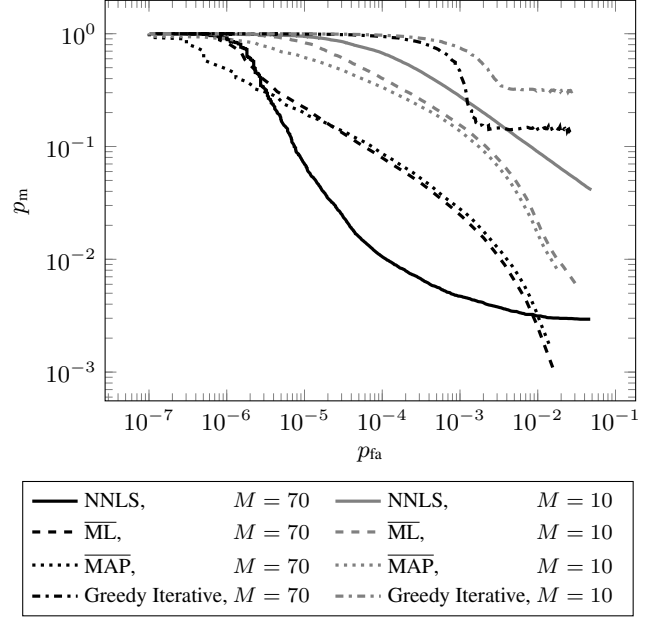


Fig. 2: ROC-curves of different methods using few antennas, $M = 10$, and many antennas, $M = 70$, at the base station.

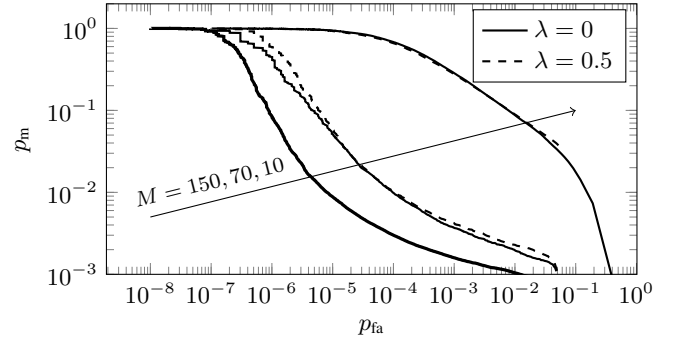
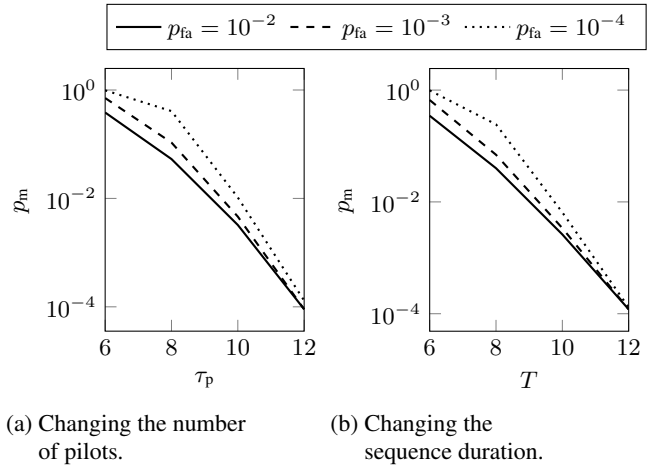


Fig. 3: ROC-curves comparing NNLS with NLASSO.



(a) Changing the number of pilots.

(b) Changing the sequence duration.

Fig. 4: Probability of missed detection given a certain probability of false alarm of the proposed scheme, solved with NNLS while changing a parameter.

8. REFERENCES

- [1] *IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond*, ITU-R Std. M.2083-0, 2015.
- [2] L. Liu and W. Yu, “Massive connectivity with massive MIMO—part I: Device activity detection and channel estimation,” *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2933–2946, June 2018.
- [3] K. Senel and E. G. Larsson, “Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach,” *IEEE Transactions on Communications*, 2018, doi: 10.1109/TCOMM.2018.2866559.
- [4] S. Haghighatshoar, P. Jung, and G. Caire, “A new scaling law for activity detection in massive MIMO systems,” *ArXiv e-prints*, Mar. 2018.
- [5] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, “Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the internet of things,” *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 88–99, Sept 2018.
- [6] E. de Carvalho, E. Björnson, J. H. Sørensen, E. G. Larsson, and P. Popovski, “Random pilot and data access in massive MIMO for machine-type communications,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 7703–7717, Dec 2017.
- [7] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.
- [8] D. L. Donoho, A. Maleki, and A. Montanari, “Message-passing algorithms for compressed sensing,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.
- [9] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO networks: Spectral, energy, and hardware efficiency,” *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.
- [10] M. Slawski and M. Hein, “Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization,” *Electron. J. Statist.*, vol. 7, pp. 3004–3056, 2013.
- [11] A. M. Bruckstein, M. Elad, and M. Zibulevsky, “On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations,” *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 4813–4820, Nov 2008.
- [12] R. Kueng and P. Jung, “Robust nonnegative sparse recovery and the nullspace property of 0/1 measurements,” *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 689–703, Feb 2018.
- [13] S. Foucart and D. Koslicki, “Sparse recovery by means of nonnegative least squares,” *IEEE Signal Processing Letters*, vol. 21, no. 4, pp. 498–502, April 2014.
- [14] C. Lawson and R. Hanson, *Solving Least Squares Problems*. Society for Industrial and Applied Mathematics, 1995.
- [15] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” <http://cvxr.com/cvx>, Mar. 2014.
- [16] M. ApS, *MOSEK Optimization Toolbox for MATLAB 8.1.0.63*, 2018. [Online]. Available: <https://docs.mosek.com/8.1/toolbox/index.html>