# ON MULTICAST BEAMFORMING AND ADMISSION CONTROL FOR UMTS-LTE

*E. Matskani, N.D. Sidiropoulos*[*] *L. Tassiulas*[†]

Dept. of ECE, Tech. Univ. of Crete
73100 Chania - Crete, Greece

Dept. of CE & T, Univ. of Thessaly
38221 Volos, Greece

**ABSTRACT**

Transmit beamforming for physical layer multicasting is emerging as an appealing transmission modality for next-generation cellular wireless systems, notably UMTS-LTE. Optimal design of the transmit beamformer(s) from a quality of service perspective is a hard computational problem; however, convex approximation tools have been shown to yield high-quality approximate solutions. Recently, Lozano proposed a particularly simple adaptive multicast beamforming algorithm that aims to serve a certain percentage of users. In parallel, convex approximation algorithms were developed for the more general problem of joint co-channel multicast beamforming and admission control. In this paper, we focus on the important (in view of recent standardization activity) special case of a single multicast group, and put the two approaches to the test. Through simple examples, we pinpoint issues regarding convergence of Lozano's algorithm. In numerical experiments with measured channel data, we show that a convex approximation approach is preferable performance-wise. At the same time, we find merits in the simplicity of Lozano's approach, and suggest a way to improve its performance.

Keywords: Multicasting, beamforming, admission control, UMTS-LTE

## 1. INTRODUCTION

Multicasting has recently gained renewed momentum as an important transmission modality for wireless networks. Multicasting bridges the gap between two widely used information dissemination paradigms: broadcasting, where common information is delivered to all nodes in a network; and independent unicast transmissions, consisting of many simultaneous point-to-point links. The middle ground between the two is important for existing and emerging applications, such as Internet TV, pay-per-view, streaming audio programming, and software updates.

Multicasting over wired networks has been much studied, and there are effective multicast routing solutions for wired networks. Wireless networks are different, in a number of ways. Along with fading and interference comes the wireless "broadcast advantage": it is possible to reach multiple destinations with a single physical-layer transmission. With the growing availability of transmit antenna arrays, this opens the door for multicasting at the physical layer: it is possible to beamform in a way that steers energy towards a *group* of receivers, while minimizing interference to all others. This can be viewed as a generalization of traditional transmit beamforming, where a single beam (lobe) is designed to steer energy in the direction of a single receiver. In *multicast beamforming*, the transmit beamformer generally creates multiple lobes, to serve a group of receivers.

To the best of our knowledge, Lopez [7] was the first to ponder about the potential of multicast beamforming. Lopez suggested a simple beamformer design approach: maximize the group-average received Signal to Noise Ratio (SNR). This is intuitive, and it yields a particularly simple design problem: the beamformer weight vector is the dominant singular vector of the channel matrix. Unfortunately, this does not guarantee a certain SNR to each receiver in the group. This is a serious drawback, because streaming media traffic requires Quality of Service (QoS), and the multicast rate is determined by the weakest link, not the average link quality.

Multicast beamforming under SNR constraints was first treated in [13] (see also references therein), where it was shown that the problem is NP-hard, yet also amenable to convex approximation tools (see also [9]). The case of multiple interfering multicast groups has been treated in [5] and it includes transmit beamforming for the multiuser downlink (multiple interfering unicast transmissions, one per user) and hybrid co-channel multicast and unicast scenarios as special cases. The multiuser downlink case is convex, as shown in [1] (see also [4]). There are other special cases that are convex - see [6]. The joint beamforming and admission control problem for the multiuser downlink has been treated in [11], and its extension to the case of multiple interfering multicast groups is reported in [12].

In this paper, we focus on a special case of [12], namely, that of a single multicast group, which is important in view of recent standardization activity in the context of UMTS-LTE; see Lozano [8], who proposed a particularly simple adaptive multicast beamforming algorithm that aims to serve a certain percentage of users. Lozano's algorithm operates under a fixed power budget, and attempts to maximize the minimum received SNR. Introducing the option to reject difficult-to-serve users for the benefit of the remaining ones, Lozano's algorithm works well in certain cases, considering its simplicity on one hand and NP-hardness of the problem on the other. Rejecting a certain percentage of users is a form of admission control, however, and therefore the proper baseline for [8] would be one that jointly accounts for beamforming and admission control.

Through simple examples, we pinpoint issues regarding convergence of Lozano's algorithm. In numerical experiments with measured channel data, we show that the convex approximation approach is preferable performance-wise. At the same time, we find merits in the simplicity of Lozano's approach, and suggest a way to improve its performance.

[*]Tel: +302821037227, Fax: +302821037542, E-mail: (matskani,nikos)@telecom.tuc.gr. Supported in part by ARL/ERO contract N62558-06-0340, and E.U./FP6 COOPCOM, WIP.

[†]E-mail: leandros@uth.gr. Supported in part by E.U./FP6 WIP.

## 2. FORMULATION AND CONVEX APPROXIMATION

Consider a base station with $N$ transmit antenna elements, and $K$ single-antenna receivers that wish to subscribe to the same multicast. Let $\mathbf{h}_i$ denote the $N \times 1$ complex baseband-equivalent channel from the transmit antenna array to receiver $i$, $i \in \mathcal{U} := \{1, \cdots, K\}$, and $\mathbf{w}^H$ denote the $1 \times N$ beamforming weight vector applied to the $N$ transmitting elements. The problem of interest can be stated as follows

$$\min_{\mathbf{w} \in \mathbb{C}^N} \|\mathbf{w}\|_2^2 \tag{1}$$

$$\text{subject to} : \|\mathbf{w}\|_2^2 \leq P, \tag{2}$$

$$\frac{|\mathbf{w}^H \mathbf{h}_i|^2}{\sigma_i^2} \geq c_i, \ \forall i \in \mathcal{U}, \tag{3}$$

where $\sigma_i^2$ is the additive noise power at receiver $i$ and $c_i$ stands for the associated minimum SNR requirement. When (2) is not enforced, the resulting problem is always feasible, provided that $\mathbf{h}_i \neq \mathbf{0}_{N \times 1}$, $\forall i$; still, finding an optimum solution is generally NP-hard, as shown in [13]. An upper bound on transmission power is important in practice for a number of reasons - including regulatory / co-channel interference considerations, and power amplifier limitations. In this case, an important concern is that (1)-(3) can be infeasible, which brings up the issue of admission control.

When admission control is necessary, a natural objective is to maximize the number of users that can be served at pre-specified SNR levels for a given $P$. An alternative would be to fix the number (or percentage) of users to be admitted and maximize the minimum SNR among those only. The second approach cannot guarantee pre-specified minimum SNR levels, but both can be used to trade-off between coverage of the subscriber population and minimum SNR that can be offered to the admitted part of the population for a given $P$. Clearly, the smaller the coverage the higher minimum SNR (and thus multicast rate) can be offered to the admitted users. Similar to the concept of a receiver operating characteristic, multicast performance in this context is characterized by the minimum SNR / multicast rate - coverage curve, parameterized by $P$.

We begin with the first approach. Mathematically, we aim for

$$S_o = \text{argmax}_{S \subseteq \mathcal{U}, \mathbf{w} \in \mathbb{C}^N} |S| \tag{4}$$

$$\text{subject to} : \|\mathbf{w}\|_2^2 \leq P, \tag{5}$$

$$\frac{|\mathbf{w}^H \mathbf{h}_i|^2}{\sigma_i^2} \geq c_i, \forall i \in S, \tag{6}$$

where $|S|$ denotes the cardinality of $S$. Given $S_o$, we then

$$\min_{\mathbf{w} \in \mathbb{C}^N} \|\mathbf{w}\|_2^2 \tag{7}$$

$$\text{subject to} : \frac{|\mathbf{w}^H \mathbf{h}_i|^2}{\sigma_i^2} \geq c_i, \ \forall i \in S_o. \tag{8}$$

The above is a special case of the setup considered in [12], which accounts for multiple interfering co-channel multicasts. It is shown in [12] (see also [11]) that it is possible to recast the two-stage problem in (4) - (8) in the following convenient equivalent form:

$$\min_{\mathbf{w} \in \mathbb{C}^N, \{s_i \in \{-1,+1\}\}_{i \in \mathcal{U}}} \epsilon \|\mathbf{w}\|_2^2 + (1-\epsilon) \sum_{i \in \mathcal{U}} (s_i + 1)^2 \tag{9}$$

$$\text{subject to} : \|\mathbf{w}\|_2^2 \leq P, \tag{10}$$

$$\frac{|\mathbf{w}^H \mathbf{h}_i|^2 + \delta^{-1}(s_i + 1)^2}{\sigma_i^2} \geq c_i, \ \forall i \in \mathcal{U}, \tag{11}$$

where $\delta \leq \min_{i \in \mathcal{U}} \frac{4 c_i^{-1}}{\sigma_i^2}$, $\epsilon < \frac{1}{P/4+1}$, and the $s_i$'s are admission control variables: $s_i = -1$ ($s_i = +1$) means that receiver $i$ is admitted (rejected). The above reformulation is convenient because it yields an effective convex approximation of the original problem formulation, which is non-convex and NP-hard. This is so because (4)–(8) (and its reformulation in (9)–(11)) contains the version considered in [13] (without the explicit power constraint, or, equivalently, $P = \infty$), which is already NP-hard. Define $\mathbf{W} := \mathbf{w}\mathbf{w}^H$, $\mathbf{H}_i := \mathbf{h}_i \mathbf{h}_i^H$, $\mathbf{s}_i := [s_i \ \ 1]^T$, and $\mathbf{S}_i := \mathbf{s}_i \mathbf{s}_i^T$. It can be shown [12] that (9)-(11) is equivalent to

$$\min_{\mathbf{W} \in \mathbb{C}^{N \times N}, \{\mathbf{S}_i \in \mathbb{R}^{2 \times 2}\}_{i \in \mathcal{U}}} \epsilon \text{Tr}(\mathbf{W}) + (1-\epsilon) \sum_{i \in \mathcal{U}} \text{Tr}(\mathbf{1}_{2 \times 2} \mathbf{S}_i) \tag{12}$$

$$\text{subject to} : \text{Tr}(\mathbf{W}) \leq P, \tag{13}$$

$$\frac{\text{Tr}(\mathbf{H}_i \mathbf{W}) + \delta^{-1} \text{Tr}(\mathbf{1}_{2 \times 2} \mathbf{S}_i)}{\sigma_i^2} \geq c_i, \ \forall i \in \mathcal{U}, \tag{14}$$

$$\mathbf{W} \geq 0, \text{rank}(\mathbf{W}) = 1, \tag{15}$$

$$\mathbf{S}_i \geq 0, \text{rank}(\mathbf{S}_i) = 1, \mathbf{S}_i(1,1) = \mathbf{S}_i(2,2) = 1, \ \forall i \in \mathcal{U}. \tag{16}$$

Dropping the rank-one constraints, we obtain a semidefinite program that is a convex relaxation[1] of the original problem and can be efficiently solved using modern interior point solvers [2]. Alas, a solution of the relaxed problem generally only provides a lower bound on the objective of the original problem; but there is considerable optimization literature exploring ways to generate a high-quality approximate solution of the original problem from a solution of the relaxed problem. In our context, the following trimmed-down (isolated multicast) version of the algorithm in [12] seems to work best among the several options we tried.

**Algorithm 1** *Multicast Membership Deflation by Relaxation (MDR):*

1. $\mathcal{U} \leftarrow \{1, ..., K\}$
2. *Solve the relaxed problem, and let* $\check{\mathbf{W}}$ *denote the resulting transmit covariance matrix*
3. $\check{\mathbf{w}}$ = *principal component of* $\check{\mathbf{W}}$*, scaled to power* $Tr(\check{\mathbf{W}})$.
4. *For each* $i \in \mathcal{U}$*, check whether* $|\check{\mathbf{w}}^H \mathbf{h}_i|^2 / \sigma_i^2 \geq c_i$*. If true* $\forall$ $i \in \mathcal{U}$*, stop (feasible solution has been found); else pick user with largest gap to its target SNR, remove from* $\mathcal{U}$ *and go to step 2.*

## 3. LOZANO'S ALGORITHM

Lozano's algorithm [8] is very simple and well-suited for adaptive on-line implementation. Its performance is remarkably good in certain cases, given that the problem it attempts to solve is non-convex and NP-hard.

Lozano's algorithm starts with an initial weight vector $\mathbf{w}_0$. At iteration $t$, $t \in \{1, 2, \cdots\}$ it first computes the previously attained SNR values for all users $P \mathbf{w}_{t-1}^H \mathbf{H}_i \mathbf{w}_{t-1}$, $i \in \mathcal{U}$, where $\mathbf{H}_i$ is defined as either $\mathbf{h}_i \mathbf{h}_i^H / \sigma_i^2$ or its expectation, depending on context; and $\mathbf{w}_t$ denotes the weight vector at iteration $t$. It then sorts the resulting SNRs and leaves out the users with the smallest SNRs. The cut-off SNR threshold may be fixed to enforce a strict SNR constraint for a variable number of users; or adjusted in each iteration to ensure that a fixed number of users will be considered, albeit without a strict SNR guarantee. The second option works better

[1] Note that (12)-(16) is a non-convex quadratically constrained quadratic program, and rank relaxation can be interpreted as its bi-dual problem.

in practice. Either way, the algorithm selects the weakest user, $i^*$, from the list of non-excluded users, makes a gradient step in its direction, i.e., $\mathbf{w}_t = \mathbf{w}_{t-1} + \mu \mathbf{H}_{i^*}\mathbf{w}_{t-1}$ followed by normalization $\mathbf{w}_t = \mathbf{w}_t/||\mathbf{w}_t||_2$, and continues to the next iteration.

Despite its conceptual simplicity, Lozano's algorithm exhibits intricate convergence behavior. In order to appreciate related issues, consider the following very simple scenario: there are $N = 2$ transmit antennas and $K = 2$ users, with associated channels $\mathbf{h}_1 = [1\ 0]^T$ and $\mathbf{h}_1 = [0\ 1]^T$ (each user only listens to a single transmit antenna). Let $\sigma_1 = \sigma_2 = P = 1$. If both users should be served, the optimal solution is $\mathbf{w} = \frac{1}{\sqrt{2}}[1\ 1]^T$, attaining an SNR of $\frac{1}{2}$ for each user. Lozano's algorithm initialized with $\mathbf{w}_0 = \mathbf{h}_1$ (say, because it was previously serving only user 1, and now user 2 comes into play) has a fixed point at $\mathbf{h}_1$, which is in the null space of $\mathbf{H}_2$ - thus user 2 is simply shut off from the system for all $\mu$. This shows that the algorithm can converge to a very suboptimal point. A small perturbation of either $\mathbf{h}_2$ or $\mathbf{w}_0$ takes the algorithm away from this undesirable fixed point; for small enough $\mu$, the iterates typically approach the optimum solution, albeit slowly. Beyond the usual speed - misadjustment trade-off, however, for this toy problem Lozano's algorithm typically exhibits limit cycle behavior when randomly initialized. Figure 1 illustrates this behavior for $\mu = 0.1$. Choosing a smaller $\mu$ helps reduce the magnitude of the oscillation, but the problem persists even for $\mu = 10^{-4}$.

Summarizing, Lozano's algorithm may fail to converge, or converge to a suboptimal solution, and is sensitive with respect to initialization and problem instantiation. These issues do crop up in realistic problem setups, however the algorithm performs better, on average, than what the above toy example suggests. Given its simplicity, the algorithm should be seriously considered as a candidate for multicast beamforming.

## 4. RESULTS

MDR fixes a certain minimum SNR and seeks to optimize coverage (number of users served) for a given $P$. Lozano's algorithm, on the other hand, fixes coverage (percentage of users served) and attempts to maximize minimum SNR for a given $P$. As such, the best way to compare the two is by means of the respective minimum SNR - coverage curves, parameterized by $P$.

In our experiments, we used measured channel data downloaded from the University of Alberta at http://www.ece.ualberta.ca/∼mimo/ (see also [3]). Details about data selection and pre-processing can be found in [10].

We used instantaneous channel vectors (rank-one channel covariance matrices) and the reported results are averages over 30 temporal channel snapshots, spanning 30 seconds. We also tested the case of long-term Channel State Information (CSI) as in [8], where only average received SNRs can be guaranteed. We estimated the associated channel covariance matrices by averaging each channel vector over the 30 temporal snapshots; i.e., with $\mathbf{h}_{i,n}$ denoting the channel from the transmit antenna array to receiver $i$ at time $n \in \{1, 2, \cdots, 30\}$, we used $\hat{\mathbf{H}}_i = \frac{1}{30}\sum_{n=1}^{30} \mathbf{h}_{i,n}\mathbf{h}_{i,n}^H$ in place of $\mathbf{H}_i$ for both algorithms.

In addition to MDR and Lozano's algorithm, we include two additional algorithms in the comparison. One is a baseline algorithm that starts with a given (common) SNR target $c_i = c$ and finds a maximal subset of users that can be served at or above the desired SNR for a given $P$ using enumeration (ENUM). Recall that the multicast beamformer design problem is NP-hard even for a fixed subset of users; but, if enumeration over all subsets (using the potentially higher-rank relaxation solution to test each subset) ends up returning a rank-one solution, then this solution is overall optimal - because one cannot possibly serve more users, even with a higher-rank covariance. Surprisingly, in all cases considered except for $K = 10$ and no admission control (full coverage), ENUM indeed returns the optimal solution - hence it can be used as the ultimate performance baseline. Of course, its complexity is exponential in $K$ and thus prohibitive for large $K$.

Given the sensitivity of Lozano's algorithm with respect to initialization, we propose using the average SNR beamformer of Lopez [7] to initialize Lozano's algorithm, instead of the $[1\ 0\ \cdots 0]$ initialization suggested in [8]. This starts the iterations at a reasonable point and, as we will see, consistently improves the results. We call this variation the LLI algorithm (Lozano with Lopez Initialization).

In all experiments, the parameters were set as follows: $N = 4$; $K = 10$ or 30; $P = 30$; $c_i = c$ and $\sigma_i^2 = \sigma^2 = 1$, $\forall i$; for MDR, $\epsilon = 10^{-10}$, $\delta < \min_{i \in \mathcal{U}} 4c^{-1}/\sigma_i^2$. For Lozano's algorithm and LLI, $\mu = 10^{-2}$ for $K = 10$; $\mu = 10^{-3}$ for $K = 30$; and convergence is declared when the difference in minimum SNR drops below $10^{-3}$. These values were empirically optimized for the fastest possible convergence in under $10^3$ iterations (note that associated analytical guidelines were not provided in [8], and recall the potential for limit cycles).

For $K = 10$, we selected measurements 1, 3, 4, 6, 7, 9, 13, 15, 12, 17 in Fig. 1 of [10], distributed in six locations (denoted L1-L3, L5-L7 in Fig. 1 of [10]). For $K = 30$, we selected six users around each of L1, L2, L3, and 4 users around L5, L6, L7.

The results are summarized in Fig. 2, 3, for the case of instantaneous CSI; and Fig. 4 for long-term CSI. For $K = 10$, we can afford the ENUM baseline, and notice from Fig. 2 that MDR performs very close to the optimum in terms of the minimum SNR - coverage curves. Lozano's algorithm is far behind - the average coverage gap is up to 5 users (50%) for a given average minimum SNR, while the average minimum SNR gap is up to 5 dB for a given average coverage. The proposed LLI variant falls between Lozano and MDR. The situation is similar for $K = 30$ in Fig. 3, except that we cannot afford ENUM in this case, and, interestingly, LLI approaches the performance of MDR. For the long-term CSI (single set of covariance matrices) $K = 10$ results in Fig. 4, note that MDR is somewhat further away from optimum in this case; Lozano's algorithm is still away from MDR, and LLI is between the two.

## 5. DISCUSSION AND CONCLUSIONS

MDR performs close to the optimum in those cases where it is possible to use enumeration as a baseline. In all cases considered, MDR outperforms Lozano's algorithm by a significant margin. Interestingly, our simple modification of Lozano's algorithm significantly improves performance, and in certain cases brings it close to that of MDR (cf. Fig. 3).

In so far as complexity is concerned, we note that, for the above setup, Lozano's algorithm requires between $10^{-3}$ and $10^{-1}$ seconds per problem instance, whereas MDR between $10^{-2}$ and 1 second - so Lozano is about two orders of magnitude faster in our implementation. ENUM takes about 2 minutes for $K = 10$. $LLI$ is generally faster than Lozano's algorithm, due to the better initialization, despite the up-front complexity of computing the dominant singular vector that maximizes the group-average SNR. The latter can be computed in an iterative fashion using the power method, which makes the complete solution very appealing from a practical perspective.

We conclude that MDR is the best option performance-wise, and its complexity is within reach of today's base stations; however, further improvements of LLI hold the promise of bridging the performance gap at a much lower complexity. We are currently investigating such improvements.

## 6. REFERENCES


[1] M. Bengtsson and B. Ottersten, "Optimal and Suboptimal Transmit Beamforming," ch. 18 in *Handbook of Antennas in Wireless Communications*, L. C. Godara, Ed., CRC Press, Aug. 2001.

[2] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004; see also http://www.stanford.edu/∼boyd/cvxbook.html.

[3] P. Goud Jr., R. Hang, D. Truhachev, and C. Schlegel, "A Portable MIMO Testbed and Selected Channel Measurements," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 51490, 10 pages, 2006. doi:10.1155/ASP/2006/51490

[4] F.-R. Farrokhi, K.R. Liu, and L. Tassiulas, "Transmit Beamforming and Power Control for Cellular Wireless Systems," *IEEE J. Select. Areas Commun.*, vol. 16, p. 1437, October 1998.

[5] E. Karipidis, N.D. Sidiropoulos, and Z.-Q. Luo, "Quality of Service and Max-min-fair Transmit Beamforming to Multiple Co-channel Multicast Groups," *IEEE Trans. on Signal Processing*, to appear (conf. version in *Proc. IEEE CAMSAP 2005*, Dec. 12-14, Puerto Vallarta, Mexico).

[6] E. Karipidis, N.D. Sidiropoulos, Z.-Q. Luo, "Far-field Multicast Beamforming for Uniform Linear Antenna Arrays," *IEEE Trans. on Signal Processing*, vol. 55, no. 10, pp. 4916-4927, Oct. 2007.

[7] M. J. Lopez, "Multiplexing, scheduling, and multicasting strategies for antenna arrays in wireless networks," Ph.D. thesis, Dept. of Elect. Eng. and Comp. Sci., MIT, Cambridge, MA, 2002.

[8] A. Lozano, "Long-Term Transmit Beamforming for Wireless Multicasting," in *Proc. ICASSP 2007*, Apr. 15-20, Honolulu, Hawaii.

[9] Z.-Q. Luo, N.D. Sidiropoulos, P. Tseng, P., and S. Zhang, "Approximation Bounds for Quadratic Optimization with Homogeneous Quadratic Constraints," *SIAM Journal on Optimization*, vol. 18, no. 1, pp. 1-28, 2007.

[10] E. Matskani, N.D. Sidiropoulos, Z.-Q. Luo, and L. Tassiulas, "A Second-order Cone Deflation Approach to Joint Multiuser Downlink Beamforming and Admission Control," in *Proc. SPAWC 2007*, June 17-20, Helsinki, Finland.

[11] E. Matskani, N.D. Sidiropoulos, Z.-Q. Luo, and L. Tassiulas, "Convex Approximation Techniques for Joint Multiuser Downlink Beamforming and Admission Control," *IEEE Trans. on Wireless Communications*, to appear.

[12] E. Matskani, N.D. Sidiropoulos, Z.-Q. Luo, and L. Tassiulas, "Joint Multicast Beamforming and Admission Control," in *Proc. CAMSAP 2007*, December 12-14, St. Thomas, U.S. Virgin Islands.

[13] N.D. Sidiropoulos, T.N. Davidson, and Z-Q. Luo, "Transmit Beamforming for Physical Layer Multicasting," *IEEE Trans. on Signal Processing*, vol. 54, no. 6, part 1, pp. 2239-2251, June 2006.


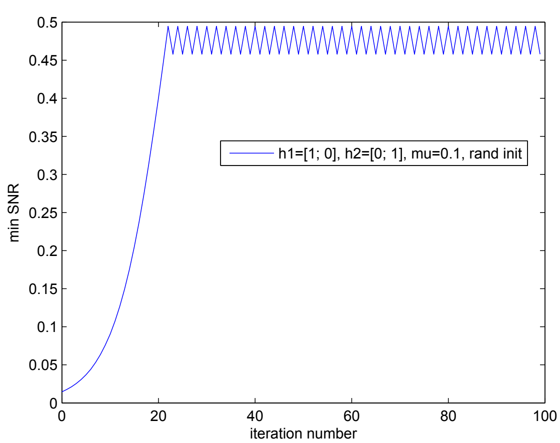


**Fig. 1**. Example of limit cycle behavior of Lozano's algorithm.

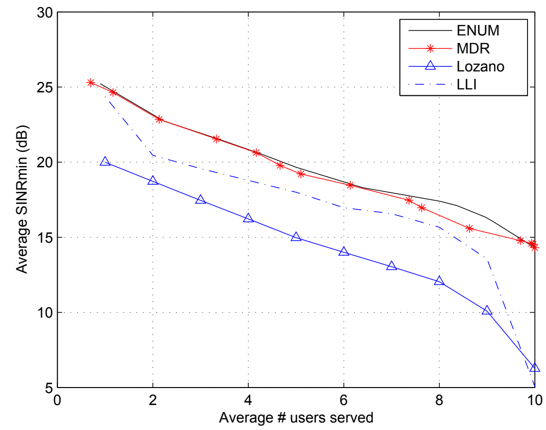


**Fig. 2**. Average SINRmin versus average number of users served: 30 measured channel snapshots. $K = 10$, $P = 30$.

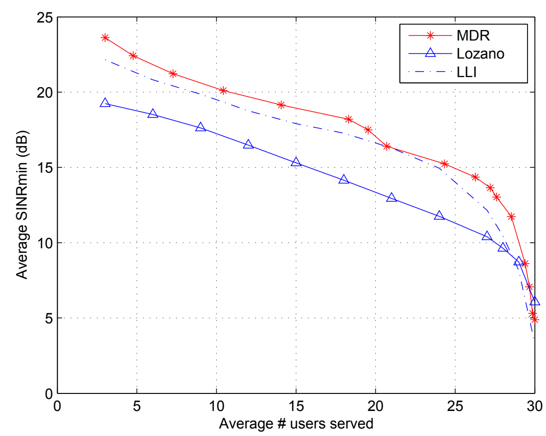


**Fig. 3**. Average SINRmin versus average number of users served: 30 measured channel snapshots. $K = 30$, $P = 30$.

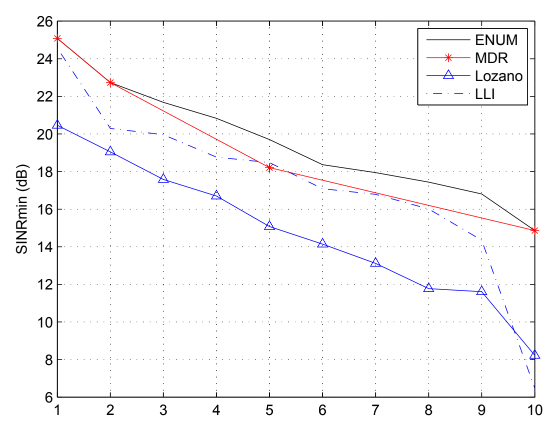


**Fig. 4**. SINRmin versus number of users served: long-term CSI. $K = 10$, $P = 30$.