ADAPTATION WITH REDUCED-SIZE MESSAGE PASS TO PRECODER SELECTION IN MULTI-CELL MIMO SYSTEMS

Igor M. Guerreiro, Charles C. Cavalcante

Wireless Telecommunications Research Group
Universidade Federal do Ceará
Fortaleza, 60455-760 Brazil
e-mail: {igor,charles}@gtel.ufc.br.

Dennis Hui

Ericsson Research
San José, CA 95134, USA
e-mail: dennis.hui@ericsson.com

ABSTRACT

This paper addresses the optimization of a multicell network where each cell needs to select precoding matrices for beamforming in a distributed way. A modified version of the adaptive min-sum algorithm that considers reduced-size messages in the message-passing procedure is proposed. The message-passing technique, where each message is reduced to convey only part of usual information, is applied to maximize the system throughput leading to a faster adaptation procedure. Simulation results show this reduction leads to a small average performance loss in system throughput of approximately 6%, while generally reduces the message size in about 73%. Also, the proposed method still outperforms the baseline greedy technique.

1. INTRODUCTION

The deployment of dense networks in new generations of wireless communication systems is still an open problem to be solved. Hundreds of cells in a cellular network make the notion of optimality in terms of data rate a hard challenge. In many cases, each cell needs to dynamically coordinate a set of parameters. An example of parameter coordination is the selection of precoders to overcome channel impairments. A joint optimization in such scenario would demand a huge signaling load and high computational burden in order to reach the globally optimal solution. On the other hand, each cell may optimize its own objective in a selfish manner to decrease both the computational complexity and the signaling load. However, the network in a global sense would experience a poor performance. Intuitively, any approach lying in between would possibly satisfy the demand for dense network deployment and still obtaining high data rate.

The work in [1] proposed an iterative method to precoder selection for beamforming purposes based on a message-passing algorithm, namely min-sum algorithm, in factor graphs [2] considering ideal message pass. However, the message size can be large and may cause an undesired overhead over the channel, as we have discussed in [3], where we analyze the signaling load and the computational burden of the message-passing algorithm with full message computation. Compared to a centralized (exhaustive search) solution, the message-passing algorithm with full message computation has much lower signaling load, but still higher when compared with a greedy approach. To decrease even more the overhead for the adaptation of messages, this paper addresses a distributed technique to the problem of precoder selection in a multi-cell network, where a modified message-passing technique that considers the exchange of reduced-size messages is proposed. Different degrees of message size reduction are assessed, which affects the overall network performance but significantly decreases the signaling load among access nodes.

2. SYSTEM MODEL

Let N be the number of (communication) nodes in a network with N cells. A node, each one in a cell, represents a base-station (BS) that has N_t transmit antennas and is associated with a user equipment (UE) that has N_r receive antennas. Each node i has a local discrete parameter p_i, whose value is drawn from a finite set $\mathcal{P}$. Each parameter p_i represents a precoding matrix index (PMI) for BS i indicating which precoder BS i should select and use at a certain radio resource block to transmit signals.

Each node i has a list $\mathcal{N}_i$ of neighboring nodes whose choices of parameter values can affect its local performance. For convenience, also let $\mathcal{A}_i \equiv \mathcal{N}_i \cup \{i\}$ denote the "inclusive" neighbor list of node i. Let $\mathbf{p}_{\mathcal{A}_i}$ denote the vector of those parameters of nodes in $\mathcal{A}_i$, with its ordering of parameters determined by the sorted indices in $\mathcal{A}_i$. Besides, each BS i transmits precoded and spatially multiplexed vector $\mathbf{x}_i$ to its associated UE i. The $N_t \times 1$ vector $\mathbf{x}_i$ is defined as $\mathbf{x}_i = \mathbf{W}_i \mathbf{s}_i$, where $\mathbf{s}_i$ is the $N_s \times 1$ symbol vector, N_s is the number of data streams, and $\mathbf{W}_i \in \mathcal{W}$ is the $N_t \times N_s$ precoding matrix indexed by the parameter p_i. Symbols are assumed to be uncorrelated and signals to have unit average magnitute, i.e. $\mathbb{E}\{\mathbf{x}_i^H \mathbf{x}_i\} = 1$. The codebook $\mathcal{W}$ is the finite set of all precoding matrices available for every communication node for beamforming purposes.

The sampled incoming signal vector $\mathbf{y}_i$ at UE i is given by

$$\mathbf{y}_i = \sqrt{g_{ii}}\mathbf{H}_{ii}\mathbf{x}_i + \sum_{j \in \mathcal{N}_i} \sqrt{g_{ji}}\mathbf{H}_{ji}\mathbf{x}_j + \mathbf{v}_i \,, \qquad (1)$$

where $\mathbf{H}_{ji}$ denotes the $N_r \times N_t$ multiple-input-multiple-output (MIMO) channel matrix between the BS j and UE served by BS i in downlink, and $\mathbf{v}_i$ is a $N_r \times 1$ zero-mean circularly symmetric complex Gaussian noise vector. The parameter g_{ji} is the path gain of each signal, here modeled in a simplified way as being $g_{ji} = \left(\frac{1}{d_{ji}}\right)^{\alpha}$, where the constant α refers to the path loss exponent and d_{ji} is the distance between BS j and UE i. Each MIMO channel matrix is obtained from a data set of measured channel matrices [4, 5], whose elements are further transformed into zero-mean random variables of unit variance, which yields each matrix $\mathbf{H}_{ji}$ (see more details in [1]). In turn, the path gain g_{ji} is inserted to the resulting matrix $\mathbf{H}_{ji}$ according to (1). This way, such a channel model provides a suitable scenario for precoder selection, which usually benefits from the characteristics of spatially-correlated channels. The second term on the right-hand side of (1) refers to the interfer-

Authorized licensed use limited to: UNIVERSIDADE FEDERAL DO CEARA. Downloaded on December 05,2022 at 13:42:49 UTC from IEEE Xplore. Restrictions apply.

ence caused by neighboring nodes.

Associated with each node i is a local performance metric or cost $M_i(\mathbf{p}_{\mathcal{A}_i})$, which is a function of those parameters in $\mathcal{A}_i$. It represents the negative of the data throughput [6, 7] of the cell corresponding to BS i measured by

$$M_i(\mathbf{p}_{\mathcal{A}_i}) = -\log\det\left(\mathbf{I} + |g_{ii}|\mathbf{R}_i^{-1}\mathbf{H}_{ii}\mathbf{W}_i\mathbf{W}_i^H\mathbf{H}_{ii}^H\right), \quad (2)$$

where $\mathbf{R}_i$ denotes the covariance matrix of the noise-plus-interference at the UE served by BS i. Each node i is assumed to be capable of communicating with all nodes in $\mathcal{A}_i$.

2.1. Ideal Message Pass to Precoder Selection

Let $\mathbf{p} \equiv \begin{bmatrix} p_1 & p_2 & \cdots & p_N \end{bmatrix}^T$ be a vector collecting all the parameters in the network, where $p_i \in \mathcal{P},\ i = 1, 2, \ldots, N$. Then, each node i aims at finding, in a distributed fashion, its own optimal parameter p_i^*, which is the corresponding component of the optimal global parameter vector $\mathbf{p}^*$ that minimizes the global performance metric M given by

$$M(\mathbf{p}) \equiv \sum_{i=1}^{N} M_i(\mathbf{p}_{\mathcal{A}_i}). \quad (3)$$

The min-sum algorithm [1] is executed on a loopy factor graph to marginalize out the global performance metric $M(\mathbf{p})$ in an iterative and efficient fashion. Such a factor graph is comprised by N factor nodes and N variable nodes. Factor node i is associated with local performance metric $M_i(\mathbf{p}_{\mathcal{A}_i})$ and variable node i is associated with local parameter p_i. Besides, each factor node i connects to the variable nodes in its neighbor list $\mathcal{A}_i$. That is, varible nodes pass messages only to neighboring factor nodes. The (aggregated) message from variable node k to factor node i is defined as

$$\mu_{p_k \to M_i}(p_k) = \sum_{j \in \mathcal{A}_k \setminus \{i\}} \mu_{M_j \to p_k}(p_k). \quad (4)$$

Similarly, factor nodes pass messages only to neighboring variable nodes. The (summary) message from factor node i to variable node k is defined as

$$\mu_{M_i \to p_k}(p_k) = \min_{\mathbf{p}_{\mathcal{A}_i \setminus \{k\}}} \left\{ M_i(\mathbf{p}_{\mathcal{A}_i}) + \sum_{j \in \mathcal{A}_i \setminus \{k\}} \mu_{p_j \to M_i}(p_j) \right\}, \quad (5)$$

which is normalized to have zero mean to not increase endlessly. The notation $\setminus\{k\}$ means the underlying operator is performed over all associated indexes except to index k.

Then, the algorithm iterates by computing and passing messages following a simultaneous message-passing scheduling, which is based on the flooding schedule [8]. Upon receipt of the message $\mu_{M_i \to p_k}(p_k)$, each variable node k then compute message $\mu_{p_k \to M_i}(p_k)$ to factor node i for each $i \in \mathcal{A}_k$. The parameter for node i is determined at its variable node i by

$$p_i^* = \arg\min_{p_i} \left\{ \sum_{j \in \mathcal{A}_i} \mu_{M_j \to p_i}(p_i) \right\}. \quad (6)$$

The algorithm runs until a stopping criterion is reached, either a predetermined maximum number of iteration or when the set of parameters computed in (6) converges to a fixed state, i.e.,

$$p_i^{(n+1)} = p_i^{(n)}, \quad \forall i = 1, 2, \ldots, N, \quad (7)$$

for sufficiently large n, where n is an iteration index.

3. REDUCED-SIZE MESSAGE PASS

Assume each message can be represented by a table of values with each entry corresponding to one of the possible values of its associated single variable. Since each parameter p_k has $L = |\mathcal{P}|$ possible values, then each message of variable p_k is simply a table with L entries. To be passed on an edge, each of the L entries may be converted into a sequence of bits. Consider that each entry is represented by N_b bits. Then, each table of L entries roughly has $L_b = N_b L$ bits. The amount of L_b bits may cause an undesired overhead over the channel represented by an edge. To overcome this, a reduction in the size of messages formulated in (4) and (5) is proposed.

The main idea here is to allow only the J smallest values of each message to be passed on, where $1 \leq J \leq L$. Table 1 shows an example of message as a table of values for different values of J, and $\mathcal{P} = \{1, 2, 3\}$. A message with no reduction in size is obtained by setting $J = 3$, as it would convey its 3 possible values. For $J < 3$, the message conveys only its J smallest values and a size reduction is then obtained.

Table 1. Example of Message as a Table of Values

	Message Value		
Parameter value	$J = 3$	$J = 2$	$J = 1$
1	-2.9	-	-
2	-5.3	-5.3	-5.3
3	-3.1	-3.1	-

Basically, an additional step is considered after the computation of each message to be passed on. Let $\mathbf{m}_{M_i \to p_k}^{(l)}$ be a message vector that conveys the lth smallest value of each message $\mu_{M_i \to p_k}$, computed in (5), and its associated parameter value, given by

$$\mathbf{m}_{M_i \to p_k}^{(l)} = \left[\min_{p_k}^{(l)}\{\mu_{M_i \to p_k}\} \quad \arg\min_{p_k}^{(l)}\{\mu_{M_i \to p_k}\} \right]^T, \quad (8)$$

where $\min^{(l)}\{\cdot\}$ and $\arg\min^{(l)}\{\cdot\}$ denote the lth smallest quantity in the bracket and the argument associated with the smallest quantity in the bracket, respectively. To comprise the J smallest values, message vectors are stacked to generate the message matrix $\mathbf{M}_{M_i \to p_k}^{(J)}$, which is given by

$$\mathbf{M}_{M_i \to p_k}^{(J)} = \left[\mathbf{m}_{M_i \to p_k}^{(1)} \quad \cdots \quad \mathbf{m}_{M_i \to p_k}^{(J)} \right]. \quad (9)$$

Note that the same formulation in (8) and (9) can be applied to generate message vectors $\mathbf{m}_{p_k \to M_i}^{(l)}$ and message matrices $\mathbf{M}_{p_k \to M_i}^{(J)}$, respectively, for each message $\mu_{p_k \to M_i}$ computed in (4).

To keep the same framework of the min-sum algorithm, incoming message matrices must be mapped back onto full-size messages. Then, the update rules defined in (4) and (5) can be properly used. Let $m_{j,l}$ be an auxiliary variable which stands for the (j, l)th entry of message matrix $\mathbf{M}_{M_i \to p_k}^{(J)}$. Then, each incoming message from factor node i to variable node k is redefined as $\tilde{\mu}_{M_i \to p_k}(p_k)$, which is given by

$$\tilde{\mu}_{M_i \to p_k}(p_k)\Big|_{p_k = m_{2,l}} = \begin{cases} m_{1,l}, & \text{for } l = 1, \ldots, J, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

Analogously, let $m'_{j,l}$ be an auxiliary variable which stands for the

Authorized licensed use limited to: UNIVERSIDADE FEDERAL DO CEARA. Downloaded on December 05,2022 at 13:42:49 UTC from IEEE Xplore. Restrictions apply.

(j,l)th entry of message matrix $\mathbf{M}^{(J)}_{p_k \to M_i}$. Then,

$$\tilde{\mu}_{p_k \to M_i}(p_k)\Big|_{p_k = m'_{2,l}} = \begin{cases} m'_{1,l}\,, & \text{for } l = 1, \ldots, J\,, \\ 0\,, & \text{otherwise.} \end{cases} \tag{11}$$

It is worth mentioning that message values not conveyed by message matrices $\mathbf{M}^{(J)}_{M_i \to p_k}$ and $\mathbf{M}^{(J)}_{p_k \to M_i}$ are set to zero in both messages $\tilde{\mu}_{M_i \to p_k}(p_k)$ and $\tilde{\mu}_{p_k \to M_i}(p_k)$. However, they could be set to any nonnegative value.

3.1. Reduction in Size

From the formulation above, the size (in bits) of each message matrix can then be given by

$$L_{\text{reduced}} = J\left(N_b + \lceil \log_2 L \rceil\right)\,, \tag{12}$$

where $\lceil \cdot \rceil$ stands for the ceiling function. The second term on the right-hand side of (12) comes from the fact that any value of variable p_k may be assumed to be an integer index. To measure the reduction in number of bits, let η be the percentage ratio of L_{reduced} to L_b, defined as

$$\eta = 1 - \frac{L_{\text{reduced}}}{L_b}. \tag{13}$$

In fact, effective size reduction is obtained whether J is upper bounded by

$$J < L/\left(1 + K\right)\,,$$

where $K = \lceil \log_2 L \rceil / N_b$ and assuming that N_b and L may be fixed in the long run. Table 2 shows the behavior of η for $N_b = 32$ and different values of L and J. For instance, if $L = 4$ and $J = 1$, the reduction is about 73%.

Table 2. Message Size Reduction for $N_b = 32$

	Percentage ratio η			
L	$J=1$	$J=2$	$J=3$	$J=4$
3	0.646	0.291	-	-
4	0.734	0.468	0.203	-
5	0.781	0.562	0.343	0.125

The motivation for characterizing each message by its J smallest values comes from the intuition that they convey the most significant part of the message from one node to the other. Clearly, the decrease in the amount of information needed to be passed on is considerable.

4. SIMULATION RESULTS

The global performance metric defined in (3) for the precoder selection problem is investigated in order to evaluate how it behaves statistically in terms of cumulative distribution functions (CDFs). The reduced-size message-passing technique, for J ranging from 1 to 3, is compared with the greedy solution [9], which is expected to provide a sub-optimal result, and with the coordinate descent technique [10], which is expected to reach a near-optimal solution. For the coordinate descent technique, a total of ten iterations was considered as its stopping criterion. The graph-based technique [1] with full-size message pass is also assessed. Moreover, the 50th CDF percentile of the system throughput is evaluated to realize how much gain each distributed technique obtains over the iterations. The convergence speed, inversely proportional to the average number of iterations until convergence per simulation run, of both distributed approaches is qualitatively assessed in terms of CDF curves for only the cases in which the algorithms converge. Additionally, the convergence rate, defined as the ratio of the number of runs in which the algorithms converge to the total number of simulation runs, is shown for both distributed techniques. A total of 850 runs (independent simulations) were conducted for statistical purposes considering a simultaneous message-passing scheduler for both algorithms.

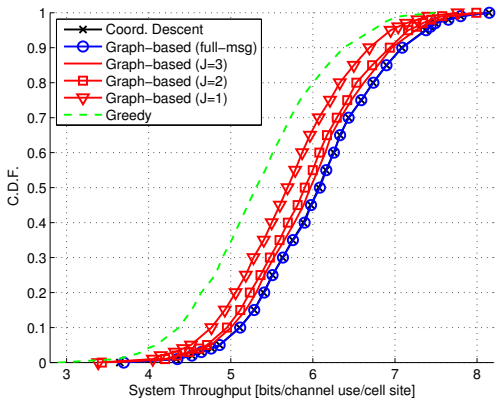

Fig. 1. Performance analysis of graph-based technique for the precoder selection problem in terms of system throughput in 19-node network.

A hexagon layout with $N = 19$ cells and a single communication node in each cell was adopted. The position of each communication node is at random following a uniform distribution. The MIMO setup is such that each transmitter has $N_t = 2$ available transmit antennas and $N_s = 1$ data streams to be transmitted, and each receiver has $N_r = 1$ receive antennas. Then, the codebook $\mathcal{W}$ is defined as

$$\mathcal{W} = \left\{ \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\j\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-j\end{bmatrix} \right\}. \tag{14}$$

Consequently, the parameters to be coordinated are four PMIs for all the cells. Each MIMO channel matrix was drawn from a set containing 324,000 samples of measured channel. The channel responses were kept constant over iterations. The average signal-to-noise ratio (SNR) value was set to 20dB, the pathloss exponent equals 3.76 and the cell radius equals 500 meters (and consequently the intercell distance equals $500\sqrt{3}$). In the greedy technique, the parameter initialization is at random, i.e. nodes pick one of the PMIs randomly at the beginning of each simulation run. In the graph-based technique, the initial messages defined in (4) are equal to zero and an ideal error-free message pass is considered. The maximum number of iterations allowed in each simulation run is 100.

The graph-based technique outperforms the greedy technique for any value of parameter J, while for full-size message pass it obtains the same performance as the coordinate descent approach. Those results can be realized in the CDF curves shown in Figure 1. Also, the larger the parameter J, the closer the performance compared to the full-size message case, but the smaller the reduction in size. For example, the ratio η equals 73.4% for $J = 1$, 46.9% for $J = 2$ and 20.3% for $J = 3$. The maximum achievable system throughput is about 8.14 bits per channel use and per cell site, reached by both the coordinate descent and the graph-based techniques. The graph-based technique for $J = 1$ reaches about 7.76 and the greedy technique reaches about 7.48 bits per channel use and per cell site at the most.

Authorized licensed use limited to: UNIVERSIDADE FEDERAL DO CEARA. Downloaded on December 05,2022 at 13:42:49 UTC from IEEE Xplore. Restrictions apply.

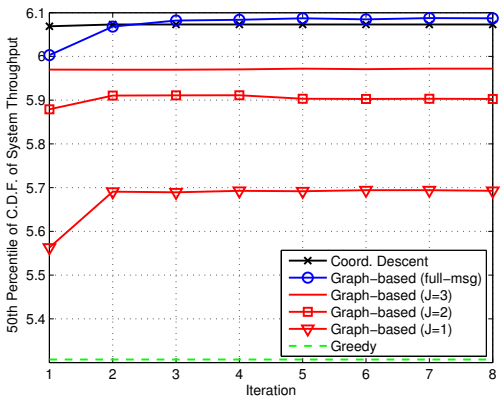

Fig. 2. CDF percentile of system throughput per iteration in 19-node network. It shows how much gain the graph-based technique obtains over the greedy.

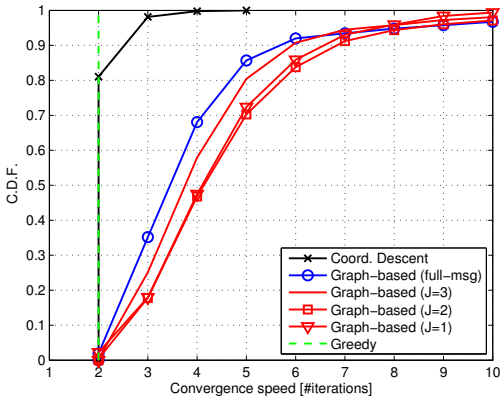

Fig. 3. Performance analysis of graph-based technique for the precoder selection problem in terms of convergence speed in 19-node network.

In general, the graph-based technique obtains most of the gain in system throughput during the first three iterations. Figure 2 shows the 50th percentile of CDF of system throughput obtained per iteration. Particularly, at the first iteration the graph-based technique for $J = 1$ outperforms the greedy solution with a percentage gain of 4.9%, while for full-size message pass it does with a gain of 13.2%. Further, at the third iteration it outperforms the greedy approach with a percentage gain of 7.2% for $J = 1$ and 14.7% for full-size message pass. On the other hand, a loss in system throughput of 6.5% can be observed by setting $J = 1$ compared to the full-size message pass at the third iteration.

In terms of convergence, Figure 3 shows that the graph-based technique demands more number of iterations to converge than the greedy solution. On average, its convergence speed is about 4.8 iterations. It is worth mentioning that it does not converge only in a small number of simulation runs. Specifically, it converges with 95% probability for $J = 1$, and with 97% probability for the full-size message case. Conversely, the greedy solution converges with 100% probability satisfying the stopping criterion always at the second iteration.

5. CONCLUSIONS

The graph-based method for distributed parameter coordination considers the impact of nodes decisions on their neighboring nodes. The message pass is only among neighbors. Without message-size reduction, such a technique reaches the (near) optimal solution. To reduce the message size, a modified version of the graph-based methodology was proposed. In this new approach, each income message conveys only part of the values along with their associated indices. After reception, each message is padded with zeros in order to be computed to generate new messages. Numerical results show that the graph-based technique with message size reduction still provides good gains in the global cost over the greedy solution. A loss in system throughput is observed in comparison with the full-size message pass case, whereas the resulting message size reduction proposed in this work is significantly large. It is worth noting that in general the graph-based approach is totally adaptable to any discrete problem of parameter coordination and it is scalable to any network size.

6. REFERENCES

[1] I. M. Guerreiro, D. Hui, and C. C. Cavalcante, "A distributed approach to precoder selection using factor graphs for wireless communication networks," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, p. 83, 2013.

[2] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498 –519, feb 2001.

[3] I. M. Guerreiro, C. C. Cavalcante, and D. Hui, "Precoder selection scheme based on message-passing approach: a practical perspective," *IOP Conference Series: Materials Science and Engineering*, vol. 67, no. 1, p. 012023, 2014.

[4] Y. Selen and H. Asplund, "3G LTE simulations using measured MIMO channels," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, 30 2008-dec. 4 2008, pp. 1 –5.

[5] M. Binelo, A. L. F. De Almeida, J. Medbo, H. Asplund, and F. R. P. Cavalcanti, "MIMO channel characterization and capacity evaluation in an outdoor environment," in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, Sept., pp. 1–5.

[6] Telatar, Emre, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, 1999.

[7] Scutari, G., Palomar, D. P., and Barbarossa, S., "Optimal linear precoding strategies for wideband noncooperative systems based on game theory - part I: Nash equilibria," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1230–1249, March 2008.

[8] F. Kschischang and B. Frey, "Iterative decoding of compound codes by probability propagation in graphical models," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 2, pp. 219 –230, feb 1998.

[9] I. Guerreiro and C. Cavalcante, "A distributed approach for antenna subset selection in mimo systems," in *2010 7th International Symposium on Wireless Communication Systems (ISWCS 2010)*, sept. 2010, pp. 199 –203.

[10] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.

Authorized licensed use limited to: UNIVERSIDADE FEDERAL DO CEARA. Downloaded on December 05,2022 at 13:42:49 UTC from IEEE Xplore. Restrictions apply.