Antenna array optimization for channel model using spherical harmonics decomposition

Leandro R. Ximenes, André L. F. de Almeida

***Resumo*— Este artigo aborda a otimização de arranjos de antenas. Para este fim, usamos um modelo de canal recente, baseado em harmônicos esféricos, que permite a dissociação dos padrões das antenas do meio de propagação. Foram derivadas duas técnicas de otimização baseadas nos métodos de Gauss-Newton (GN) e do Gradiente (SD), bem como seus custos computacionais em função do número de antenas e da ordem esférica. Nossos resultados fornecem uma comparação desses métodos, variando basicamente estes dois parâmetros, mostrando que ambos podem dar resultados ótimos, sendo o método do GN mais rápido e mais estável para os casos de dinâmica lenta.**

***Palavras-Chave*— MIMO, Otimizaçã de antenas, Harmônicos esféricos, Gauss-Newton, Algoritmo do Gradiente**

***Abstract*— This paper addresses the optimization antenna arrays. For this purpose, we use a recent channel model based on the use of spherical harmonics, which allows decoupling the patterns of the antennas from the propagation medium. We also derived two techniques of optimization based on methods of Gauss-Newton (GN) and of Steepest-Descent (SD), as well as their computational costs as functions of number of antennas and of the spherical order. Our results provide a comparison of these methods, varying basically these two parameters, showing that both can give optimal results, being GN method faster and stabler for low dynamic cases.**

***Keywords*— MIMO, Antenna Optimization, Radiation pattern, nsaSpherical Harmonics, Gauss-Newton, Steepest-Descent**

I. INTRODUCTION

With the rapid growth of wireless data traffic, new standards have been struggling towards antenna and propagation characteristics matching. This concept is part of we call exploiting the spatial diversity. The benefits of multiple inputs Multiple outputs (MIMO) systems falls under the multiple signals arriving at the receiving antennas through the (so-called) multipaths ([1], [2]). In a matter of fact, spatial diversity is only one of the diversities related to antennas, as well as the pattern diversity and polarization diversity, but except as otherwise stated, it is treated here as the one that encompasses all of them.

Hence, antennas have ascended from minor supporting actors to the main actors in the wireless communications. Until recently designing antennas (and arrays of them) was based basically only on power specifications, choosing specific radiation patterns characteristics and frequency bands ([3], [4], [5], and others). In other words, the use of arrays of antennas corresponded to alter the directivity, reduce side lobes, among other points of smaller importance for signal correlation and symbol interference. The research field in antennas was reduced then to the beamforming technique, until the promising use of MIMO systems raised the possibility of designs specified antennas to exploit the richness of scattered signals ([6], [7], etc).

This work is part of antennas optimization techniques for MIMO systems. Since the construction of the channel matrix does not take into account only the power gain of each path, but also the phases of the same, these design techniques present innovation upon conventional ones. Another point of innovation is that our channel model is built here by the use of Spherical Harmonics. This model allows to separate the matrix channel into receiving antennas, propagation medium and transmitting antennas ([8]). This separation of the antennas in this way is crucial to optimize the radiating elements. Without this, one must know exactly the total correlation matrix of all the antennas, without even know the notion of optimal radiation pattern, besides being computationally very expensive. This is contrary to our method, where the coefficients of spherical harmonics to generate the pattern can be easily found. It is also important to note that the techniques identify the optimal radiation pattern (including amplitude and phase), but still does not take into account the practicalities of physical construction of the antennas.

This paper begins with a basic description of the channel model using spherical harmonics in the section II. In section III, the mathematical formulation about two optimization methods are presented: Gauss-Newton (GN) and Steepest-Descent (SD), followed by the presentation of their computational costs. Inside the Simulations section (section IV), these methods are compared in terms of flops according to some parameters, as the spherical orders. Finally, a section V brings the conclusions and perspectives about this work.

II. CHANNEL DECOMPOSITION MODEL

Let $E_a(\theta, \phi)$ and $E_b(\theta, \phi)$ be the electric fields for a receiving and a transmitting antenna, respectively, whose representations by spherical harmonics expansions ([9], [10]) may be given according to

$$E_a(\theta, \phi) = \sqrt{4\pi} \sum_{lr=0}^{Lr} \sum_{m=-lr}^{lr} A_l^m Y_l^m(\theta, \phi) \tag{1}$$

$$E_b(\theta, \phi) = \sqrt{4\pi} \sum_{lt=0}^{Lt} \sum_{m=-lt}^{lt} B_l^m Y_l^m(\theta, \phi) \tag{2}$$

where Lt and Lr are the greatest used orders for representing the functions $E_a(\theta, \phi)$ and $E_b(\theta, \phi)$. Likewise, A_l^m and B_l^m are the spherical harmonics coefficients for each electric field.

Leandro R. Ximenes, André L. F. de Almeida GTEL-Wireless Telecom Research Group, Federal University of Ceará, Fortaleza, Brazil E-mails: {leandro,andre}@gtel.ufc.br.

Organizing the coefficients of the spherical harmonics in vectors, we get:

$$\mathbf{a} = [A_0^0 \;\; A_1^{-1} \;\; A_1^0 \;\; A_1^1 \;\; A_2^{-2} \;\; ... \;\; A_{Lr}^{Lr}] \quad (3)$$

$$\mathbf{b} = [B_0^0 \;\; B_1^{-1} \;\; B_1^0 \;\; B_1^1 \;\; B_2^{-2} \;\; ... \;\; B_{Lt}^{Lt}] \quad (4)$$

Assuming also that the receiving array of a wireless link has P antennas, and the transmitting one has Q antennas, then in a MIMO system the coefficients can be grouped as

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}(1) \\ \mathbf{a}(2) \\ \vdots \\ \mathbf{a}(P) \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{b}(1) \\ \mathbf{b}(2) \\ \vdots \\ \mathbf{b}(Q) \end{bmatrix}, \quad (5)$$

These matrix configurations allow to rewrite the channel matrix as the multiplication of three independent matrices: $\mathbf{A}$ (receiving array), $\mathbf{M}$ (propagation medium), and $\mathbf{B}$ (Transmitting array) ([8]). Considering that both arrays have two polarizations (vertical and horizontal), and therefore assuming that the channel matrix has four subchannels (i.e. the number of combinations for co-and cross polarizations), we can finally express them as

$$\mathbf{H_{vv}} = \mathbf{A_v}\mathbf{M_{vv}}\mathbf{B_v}^T. \quad (6)$$

$$\mathbf{H_{vh}} = \mathbf{A_h}\mathbf{M_{vh}}\mathbf{B_v}^T. \quad (7)$$

$$\mathbf{H_{hv}} = \mathbf{A_v}\mathbf{M_{hv}}\mathbf{B_h}^T. \quad (8)$$

$$\mathbf{H_{hh}} = \mathbf{A_h}\mathbf{M_{hh}}\mathbf{B_h}^T. \quad (9)$$

where the subscripts for $\mathbf{H}$ and $\mathbf{M}$ show the polarization from transmitter side (left subscript) to the receiver side (right subscript). Among the usefulness of this model, it is possible, for instance, designing a receiver that favors the V-V polarization for one antenna pair, and the H-V polarization for some other, as well as to analyse the isolated impact of the propagation scenario on the overall Cross-Polarization Ratio (XPR). Of course, the channel knowledge envolving the polarization has to be known *a priori* to optimize in this fashion.

Separating the channel in this way requires that another step has to be taken so that our channel is coherent with most current bidirectional channel models (e.g. [11]), among others). This step is to return the matrices (Eqs. 6-9) into the classical single MIMO channel $\mathbf{H}$ form

$$\mathbf{H} = \begin{bmatrix} \mathbf{H_{vv}} + \mathbf{H_{hv}} \\ \mathbf{H_{vh}} + \mathbf{H_{hh}} \end{bmatrix} \quad (10)$$

III. OPTIMIZING ANTENNAS' EXPANSION COEFFICIENTS

For sake of antenna optimization let us define

$$f = \mathbf{H}_o - \mathbf{AMB}^T \quad (11)$$

as the function that relates the error between any desired channel $\mathbf{H}_o$ and channel composed by specific matrices $\mathbf{A}$, $\mathbf{M}$ and $\mathbf{B}$. In the matter of fact, our goal is to match this set of three matrices such that minimizes the norm

$$||f|| = ||\mathbf{H}_o - \mathbf{AMB}^T|| \quad (12)$$

Next section will approach two methods for antenna optimization: Gauss-Newton and Steepest Descent.

A. Gauss-Newton method

The Gauss-Newton (GN) method let us to find the set of solutions for a number of linear or nonlinear equations using an iterative process. Being $\mathbf{B}^T$ the variable to be optimized (so the function (11) is minimized), then:

$$\mathbf{B}_{k+1}^T = \mathbf{B}_k^T - (\mathbf{J}_f(\mathbf{B}_k^T)^H \mathbf{J}_f(\mathbf{B}_k^T))^\dagger \mathbf{J}_f(\mathbf{B}_k^T)^H f(\mathbf{B}_k^T) \quad (13)$$

where k is the iterative index, and † is the Moore-Penrose pseudoinverse. The maximum value achieved for k is usually given by a stopping criterion. In fact, using the Gauss-Newton method for only one linear function (of one unknown variable) does not need an iterative process. Therefore, in our case, if one variable to be found ($\mathbf{B}$), the jacobian term $\mathbf{J}_f$ become a simple differencial operation,

$$\mathbf{J}_f(\mathbf{B}_k^T) = \frac{\partial(\mathbf{H} - \mathbf{AMB}_k^T)}{\partial \mathbf{B}_k^T} = -\mathbf{AM} \quad (14)$$

and the method of Gauss-Newton is reduced to a simply use of pseudoinverse in a closed-form solution. Substituing (11) and (14) in (13), we finally obtain

$$\mathbf{B}^T = (\mathbf{AM})^\dagger \mathbf{H} \quad (15)$$

Even though the GN method was reduced to a single equation, it could be applied to optimize not only one variable, but two or three. In other words, the Gauss-Newton method could find simultaneously unknown matrices $\mathbf{A}$ and/or $\mathbf{M}$ and/or $\mathbf{B}$. For this reason, we will keep the terminology Gauss-Newton for this non-iterative method.

B. Steepest-descent method

Minimizing the quadratic norm (12), the cost function J can be given by

$$J[\mathbf{B}^T] = ||f||^2 = E[||\mathbf{H} - \mathbf{AMB}^T||^2] \quad (16)$$

The optimization using the steepest-descent method tell us to find the minimum point of the cost function through using its gradient. In other words, assuming that there is only one point of minimum and no point of maximum, then simply

$$\nabla J[\mathbf{B}^T] = \frac{\partial J[\mathbf{B}^T]}{\partial \mathbf{B}^T} = 0 \quad (17)$$

$$\nabla J[\mathbf{B}^T] = -2(\mathbf{AM})^H [E[\mathbf{H}] - \mathbf{AMB}T] \quad (18)$$

and the equations for the steepest-descent algorithm is obtained

$$\mathbf{B}_{k+1}^T = \mathbf{B}_k^T - \frac{\mu}{2} \nabla J[\mathbf{B}^T] \quad (19)$$

$$\mathbf{B}_{k+1}^T = \mathbf{B}_k^T + \mu(\mathbf{AM})^H (E[\mathbf{H}] - \mathbf{AMB}_k^T) \quad (20)$$

C. Convergence properties and numerical complexity

The computational cost comparison between both methods is done based on the flops required by each one for convergence. The flops are standard forms of comparison and measurement, where 1 flop corresponds to a float-point operation (sum or multiplication).The comparisons are done basically by looking at the ergodic capacities and the number of flops required until the steady-state.

1) Gauss-Newton computational cost: The equation (15) can be implemented by several forms, each one having its own computational cost. On the other hand, the evaluation of the pseudoinverse of a matrix may be treated equivalently, in term of flops, to the computation of the Singular Value Decomposition (SVD) of the same matrix. In fact many mathematical libraries (e.g. IMSL, LAPACK) compute the pseudoinverse by means of the SVD algorithm ([12]). The computational cost in flops for equation (15) using the application of the Golub-Reisch SVD algorithm is

$$O = 9M_t^3 + 8M_t^2 N_r + M_t[4N_r^2 + N_r(2M_r - 1) + N_t(2N_r - 1)] \quad (21)$$

If it is assumed a spherical harmonics truncating order $L = L_r = L_t$, such that $M = M_r = M_t = (L+1)^2$, then the equation above can be rewritten as

$$O = 9(L+1)^6 + 10(L+1)^4 N_r + (L+1)^2(4N_r^2 + 2N_t N_r - N_t - N_r) \quad (22)$$

For paired systems ($N_r = N_t = N$), simpler equation can be derived

$$O = 9(L+1)^6 + 10(L+1)^4 + N(L+1)^2(6N-2) \quad (23)$$

2) Steespest Descent computational cost: Starting from equation (20), the number of multiplications (*Om*) and summations (*Os*) for one iteration are given by

$$Om = M_t N_t(M_r + N_r + 1) + M_r N_r(N_t + M_t) \quad (24)$$

$$Os = N_r N_t(M_r + M_t + 1) + M_t M_r(N_t + N_r) - M_r N_t - M_t N_r \quad (25)$$

Once again, if $M_r = M_t = (L+1)^2$,

$$Om = (L+1)^4(N_r + N_t) + (L+1)^2 N_t(2N_r + 1) \quad (26)$$

$$Os = (L+1)^4(N_t + N_r) + (L+1)^2(2N_r N_t - N_t - Nr) + N_r N_t \quad (27)$$

and if $N_r = N_t = N$,

$$Om = 2(L+1)^4 N + (L+1)^2(2N^2 + N) \quad (28)$$

$$Os = 2(L+1)^4 N + (L+1)^2(2N^2 - 2N) + N^2 \quad (29)$$

Then, the overall number of flops required for one iteration is

$$O = 4(L+1)^4 N + (L+1)^2(4N^2 - N) + N^2 \quad (30)$$

IV. Simulation Results

The upper bound for ergodic capacity during optimization is set by the proper choice of $\mathbf{H}_o$ for equation (11). In fact, $\mathbf{H}_o$ could be chosen such that the optimization would maximize/minimize other metrics. Our optimal channel is chosen as a full rank, full orthogonal matrix. Then we assume, for convenience, the identity matrix. The identity matrix is the best matrix to show the independent subchannels one MIMO system can have, as well as we can easily see its eigenvalues.

In the simulation example we use $\mathbf{A}$ composed by four antennas, each one with random patterns. The number of spherical harmonics modes is set to $M = 16$ (truncated at spherical order $L = 3$). The matrix $\mathbf{B}$ has arbitrarily the same number of antennas and spherical modes. In this first simulation, the propagation scenario is also random, where $\mathbf{M}$ here is not built from spatial information of clusters and theirs gains like in the last simulation of this paper, but simply as a gaussian distributed complex random matrix with zero mean and unitary variance. The power of both $\mathbf{A}$ and $\mathbf{M}$ are unitary, as well as the channel matrix, and therefore the capacity increase is due to augment of the spatial diversity.

A. Simulation I - Rich-scattering scenario

1) Steepest Descent (SD): Other steepest descent parameters are the following:

- adaptive step $\mu = 0.001$
- Number of statistical realizations for mean: $R = 100$

The ergodic capacity for 1000 iterations of the steepest-descent algorithm as function of the spherical order L are seen at figure 1, and as function of the transmitting antennas N at figure 3.

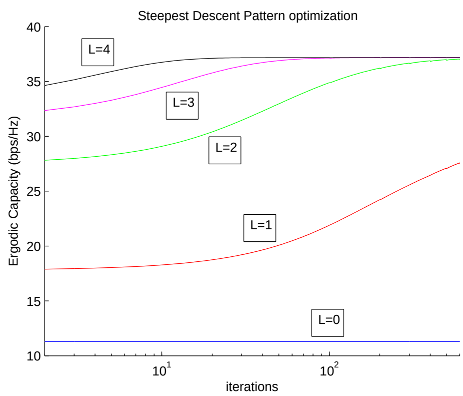
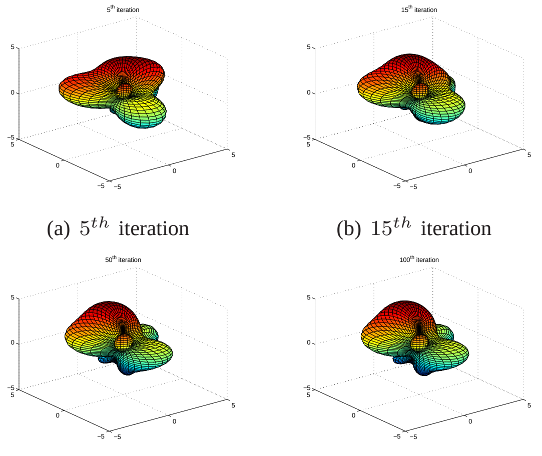


Fig. 1. Steepest-Descent antennas optimization. Ergodic capacity vs Spherical order L

The radiation pattern of the transmitting array is progressively changed as seen by the sequence (Fig.2), where it is shown the patterns for 5^{th}, 15^{th}, 50^{th} and 100^{th} iterations.

2) Gauss-Newton versus Steepest-Descent: Admitted that both methods achieve convergence, and the optimization errors are small enough so the capacity keeps itself within the steady-state, the comparison between them are mainly done by taking into account the computational costs.



(a) 5^{th} iteration (b) 15^{th} iteration

(c) 50^{th} iteration (d) 100^{th} iteration

Fig. 2. Steepest-Descent pattern optimization

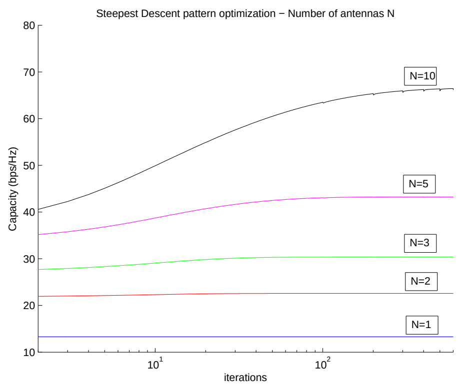


Fig. 3. Steepest-Descent antennas optimization. Ergodic capacity vs Number of antennas N

Datas about convergence of the two methods are inserted in the tables I and II. For the Gauss-Newton (GN) method, it is presented the optimized capacity and the number of flops to achieve it. In addition, for the steepest-descent is shown the number of flops for one iteration (*flops/k*), the number of iterative steps (*k*) and then the total number of flops.

In spite of the fact that pseudoinverse has a high computational cost, which makes the number of flops by iteration greater for the GN method, the several steps required by the steepest descent make it more expensive at the end. The figure 4 shows the cost-behavior of the SD algorithm for different steps' sizes. Note that the value of the step does not change the number of flops by iteration, but the number of required

	Gauss-Newton (GN)	
Order	flops(total)	Capacity (bps/Hz)
$L=0$	107	11.28
$L=1$	1,088	37.16
$L=2$	8,163	37.16
$L=3$	39,424	37.16
$L=4$	146,875	37.16

TABELA I

GN COMPUTATIONAL COST

	Steepest Descent (SD)			
Order	*k*	flops/*k*	flops(total)	Capacity (bps/Hz)
$L=0$	†	92	†	11.29
$L=1$	†	512	†	33.26
$L=2$	960	1,852	1,777,920	37.16
$L=3$	77	5,072	390,544	37.16
$L=4$	24	11,516	276,384	37.16

TABELA II

SD COMPUTATIONAL COST - † NOT ACHIEVED CONVERGENCE

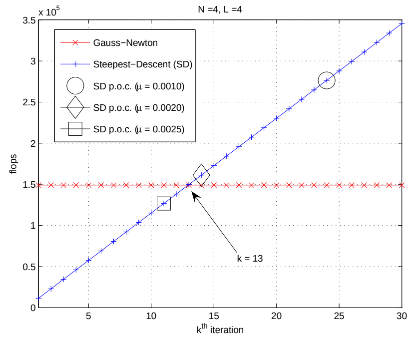


Fig. 4. Computational cost comparison between GN and SD with different adaptation steps

steps to reach the point of convergence (p.o.c.). Moreover, even being clear that increasing this value turns the steepest descent faster, its stability is not guaranteed.

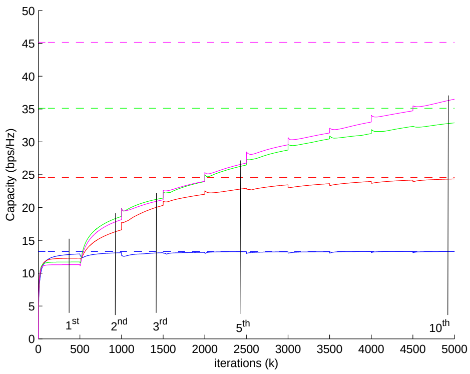
It is also noteworthy that hardly the SD algorithm will be faster considering that the GN takes only one iteration through time. In fact, the advantage of steepest descent over the Gauss-newton will come for highly dynamic channels, making several matrix inversions very cumbersome.

B. *Simulation II - bidirectional, single-scattering scenario*

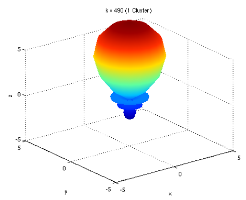
Once the computational cost of the Gauss-Newton does not change based on the number of clusters, but on the number of antennas and spherical modes, the impact of using bidirectional scenarios will change only for the SD technique.

This change is due to the fact that realistic scenarios may not provide enough richness to achieve the theoretical MIMO performance. To illustrate this effect, the simulation here verifies the capacity gains by adding up single-scattering clusters in the propagation medium. The design parameters for the steepest-descent were kept, but after every 500 iterations a new cluster was allocated to the scenario. In other words, a cluster with unity power and a specific pair of DOA/DOD was added to the calculation of the matrix $\mathbb{M}$ ([8]).

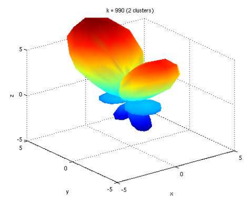
Ten clusters were added in total. All directions of arrival and departure were chosen with zero mean and unit variance Gaussian distributions with respect to the azimuthal coordinate. Regarding the elevation angle, DOAs were assumed in an uniform distribution between $\theta = 0$ and $\theta = \pi/2$ (horizontal plane), while DODs of each added cluster followed a fixed progression of 10 degrees steps from $\theta = \pi/2$. The choice of the distributions cited above was based on the natural position of higher elevation for the transmission antennas in relation



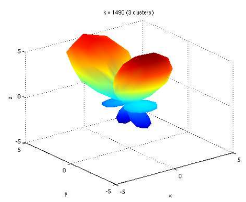
(a) Capacity (bps/Hz) for different number of antennas ($N = 1, 2, 3, 4$). Dashed lines are theoretical ones



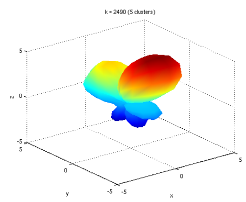
(b) 1^{st}



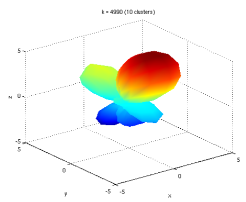
(c) 2^{nd}



(d) 3^{rd}



(e) 5^{th}



(f) 10^{th}

Fig. 5. SD optimization for bidirectional channels with progressive addition of single-scattering clusters.5(b)-5(f) illustrate some optimized radiation patterns

to the reception ones (mobile terminals), and the existence of rays coming from all directions in the horizontal plane.

The figure 5 shows this simulation. The figure 5(a) brings the optimization behavior when the number N = 1, 2, 3 and 4. In the same plot it is seen the optimal performance (dashed lines). The ladder-form of the curves (mainly seen for $N = 4$) happens due to the addition of new clusters. The vertical lines cut the curves where are plotted the radiation patterns shown from fig.5(b) to fig.5(f). The number at the bottom corresponds to the number of cluster added.

It is interesting to note that smaller arrays exploit less the richness of scenarios, and consequently the gains after optimization are smaller. This could be seen also in the figure for i.i.d.channel (fig.1 and fig.3). Moreover, the promising performance of MIMO systems ([1], [2]) is chiefly ruled by the richness of scenarios (where the number and positions of scattering clusters are crucial), once the adaptive process tended slower to the optimal curve for a low number of clusters.

V. Conclusion

We have presented an updated MIMO channel model, and also a comparative analysis envolving two methods of antenna optimization based on it. These are the Gauss-Newton (GN) and the Steepest-Descent (SD). Meanwhile the GN is reduced a simple equation for one unknown matrix, and requires some extra computational work to calculate the pseudoinverse , the SD algorithm needed several iterations, surpassing the overall cost of GN for most of the cases, being the worst between both. The exception comes when dealing with high-dynamical systems, where the fixed and non-adaptive solution of the GN may lead to poorer behaviors, even though no problems with stability are presented.

The methods of antenna optimizations here have shown very interesting results and applicability on increasing the capacity, although it is required knowledge of the channel matrix (in this case, knowledge of one end of the wireless link and the propagation scenario). This is also a drawback of some other optimization methods like the water-filling, but however, the approach of antenna optimization using spherical harmonics modelling presents a new and promising field of study, since some future considerations and discoveries may greately reduce such problem.

Referências

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications - Springer Netherlands*, vol. 6, 1998.

[2] E. Telatar, "Capacity of multi antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, 1999.

[3] O. M. Bucci, G. D'Elia, G. Mazzarela, and G. Panariello, "Antenna pattern synthesis: A new general approach," *Procedings of the IEEE*, vol. 82, 1994.

[4] A. C. Ludwig, "Radiation pattern synthesis for circular aperture horns," *IEEE Trans. Antennas and Propagation*, vol. 14, 1996.

[5] S. Choi and D. Yun, "Design of an adaptive antenna array for tracking the source of maximum power and its application to cdma mobile communications," *IEEE Trans. Antennas and Propagation*, vol. 45, 1997.

[6] H. Arai, "Adaptive pattern controlled handset antenna by analog phase shifters," in *Eucap 2009 - 3rd European Conference on Antennas and Propagation*, Berlin, Germany, 2009.

[7] V. Zuniga, A. T. Erdogan, and T. Arslan, "Adaptive pattern optimization for antenna arrays by phase perturbations using particle swarm optimization," in *2010 NASA/ESA Conference on Adaptive Hardware and Systems*, Anaheim, California, 2010.

[8] L. R. Ximenes and A. L. F. de Almeida, "Capacity of mimo antenna systems using spherical harmonics expansion," in *IEEE 72th Vehicular Technology Conference, VTC 2010-Fall*, Ottawa, 2010, pp. 1–5.

[9] J. Rahola, F. Belloni, and A. Richter, "Modelling of radiation patterns using scalar spherical harmonics with vector coefficients," in *Eucap 2009 - 3rd European Conference on Antennas and Propagation*, Berlin, Germany, 2009.

[10] T. Abhayapala, T. Pollock, and R. Kennedy, "Characterization of 3D spatial wireless channels," in *IEEE 58th Vehicular Technology Conference, VTC 2003-Fall*, Ottawa, 2003, pp. 123–127.

[11] *WINNER II Channel Models - 2007 deliverable*, D1.1.2 v1.2 ed., Information Society Technologies, 2008.

[12] G. H. Golub and C. Reinsch, "Singular value decomposition and least square solutions," *Numer. Math.*, vol. 14, 1970.