33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Cramér-Rao Bounds on the Performance of Simulated Annealing and Genetic Algorithms in EEG Source Localization

D. I. Escalona-Vargas, D. Gutiérrez, *Member, IEEE*, and I. Lopez-Arevalo

***Abstract*— In this paper, we evaluate the performance of simulated annealing (SA) and the genetic algorithm (GA) when used for electroencephalographic (EEG) source localization. The performance is evaluated on the variance of the estimated localizations as a function of the optimization's initialization parameters and the signal-to-noise ratio (SNR). We use the concentrated likelihood function (CLF) as objective function and the Cramér-Rao bound (CRB) as a reference on the performance. The CRB sets the lower limit on the variance of our estimated values. Then, our simulations on realistic EEG data show that both SA and GA are highly sensitive to noise, but adjustments on their parameters for a fixed SNR value do not improve performance significantly. Our results also confirm that SA is more sensitive to noise and its performance may be affected by correlated sources.**

I. Introduction

Solution to the inverse problem in electroencephalography (EEG) consists of finding the current distributions within the brain using EEG measurements. Many times solving the inverse problem involves an iterative solution of the forward problem (i.e., computing the electric potentials over the scalp given a current source), then it is important to have efficient analytical and numerical solutions of the forward problem in order to minimize the computational burden [1]. Furthermore, the optimization process may fall in the case where the cost function to be minimized contains several local minima around the neuroelectric sources which could make difficult the estimation process.

Many metaheuristics algorithms for global optimization have been used in the solution of the inverse problem, and most of them report high accuracy on the estimation of multiple dipoles [2] [3]. Some examples of such algorithms are the simulated annealing (SA) algorithm [4], genetic algorithms (GA) [5] [6], particle swarm optimization (PSO) [7], and tabu search (TS). Nevertheless, a strict statistical study on the variability of these results under realistic conditions has not yet been performed, and the establishment of realistic confidence intervals as a function of the parameter space of the metaheuristic algorithms remains an open task. Therefore, in this paper we propose to evaluate the performance of some of these metaheuristics algorithms using the Cramér-Rao bound (CRB).

This work was supported by the National Council of Science and Technology (CONACyT-Mexico) under Grant 101374.

D. I. Escalona-Vargas and I. Lopez-Arevalo are with the Center of Research and Advanced Studies (Cinvestav) at Tamaulipas, 87130, Cd. Victoria, Mexico descalona@cinvestav.mx, ilopez@tamps.cinvestav.mx

D. Gutiérrez is with Cinvestav at Monterrey, 66600, Apodaca, Mexico dgtz@ieee.org

The CRB establishes a lower bound on the variance of any unbiased estimator for a predefined set of parameters. The CRB is independent of the algorithm used for the estimation and it defines a universal performance limit among unbiased estimators [8]. Therefore, the CRB serves as a reference to analyze the variability of an estimation process as a function of the parameters of the optimization algorithm.

In our case, we are interested in evaluating how much the performance of the SA and GA is affected when changing their initialization parameters. Such evaluation is performed as a function of the signal-to-noise ratio (SNR), thus not only *ad-hoc* conditions but also realistic scenarios are considered. In Section II, we define the optimization problem in terms of the concentrated likelihood function (CLF) and define its CRB. In the same section, we briefly describe the SA and GA. In Section III, we present numerical examples using simulated EEG data to demonstrate the applicability of the proposed methods. Section V discusses the results and future work.

II. Methods

In this section, we define the measurement model of the EEG generated by multiple dipole sources and the CLF is established as cost function. Later, the CRB is introduced, as well as the SA and GA.

A. Concentrated Likelihood Function

Let Y_k be the matrix of measurements for $k = 1, 2, \ldots, K$ independent experiments obtained from an array of $m = 1, 2, \ldots, M$ EEG sensors at $t = 1, 2, \ldots, N$ time samples. These measurements are made in the presence of additive noise E_k distributed as $\mathcal{N}(0, \sigma^2)$ and uncorrelated in time and space. Under these conditions, the measurement model is given by

$$Y_k \quad = \quad A(\boldsymbol{\theta})\boldsymbol{q}(t) + E_k, \tag{1}$$

where $\boldsymbol{q}$ is the *dipole moment*, $\boldsymbol{\theta}$ in our case corresponds to the dipole's position (i.e., $\boldsymbol{\theta} = \boldsymbol{r}_q = [r_{q_x}, r_{q_y}, r_{q_z}]^T$), and $A(\boldsymbol{\theta})$ is the *array response* matrix of size $M \times 3$ where the mth row corresponds to the *kernel vector*. Such kernel vector relates the electric potential at an observation point with the current source, the position of the mth electrode, and the volume conductor's geometry and electrical properties (see [1], [9] for a more detailed description). In the case of p distinct dipoles, equation (1) holds with $\boldsymbol{q}(t)$ and $A(\boldsymbol{\theta})$ substituted with $\boldsymbol{q}(t) = [\boldsymbol{q}_1(t),\ \boldsymbol{q}_2(t),\ \ldots,\ \boldsymbol{q}_p(t)]^T$, and $A(\boldsymbol{\theta}) = [A_1(\boldsymbol{\theta}),\ A_2(\boldsymbol{\theta}),\ \ldots,\ A_p(\boldsymbol{\theta})]$.

Therefore, the CLF can be defined as [10]

$$F(\boldsymbol{\theta}) = \text{tr}\{(I - A(A^T A)^{-1} A)\hat{R}\}, \quad (2)$$

where $\text{tr}\{\cdot\}$ is the trace operator, $\hat{R}$ is a consistent estimate of the covariance matrix, and $A = A(\boldsymbol{\theta})$ for simplicity in the notation. It can be shown that an unbiased estimate of $\boldsymbol{\theta}$ (denoted as $\hat{\boldsymbol{\theta}}$) can be obtained by minimizing (2), which is known as the maximum likelihood (ML) estimate [8].

B. Cramér-Rao Bound

The CRB is a theoretical lower limit on the variance that provides a benchmark against which we can compare the performance of any unbiased estimator. In our case, the CRB is defined by [11]

$$CRB(\boldsymbol{\theta}) = \sigma^2 \left[\sum_{t=1}^{N} Q(t) D^T P_A^{\perp} D Q(t) \right]^{-1}, \quad (3)$$

where $Q(t)$ is a block diagonal matrix of size $9p \times 3p$ in which the moments of the p dipoles are arranged, D is a matrix $(M \times 9p)$ containing the partial derivatives of the desired parameters, $P_A^{\perp} = I - A(A^T A)^{-1} A^T$, and σ^2 is the signal's variance. For more details on the calculation of (3) see [12].

Therefore, our goal is to evaluate the performance in optimizing (2) when SA and GA are used. Hence, unbiased estimates $\hat{\boldsymbol{\theta}}$ will be obtained under different conditions, then their variance will be compared against the CRB. In the case of optimal performance, the variance of the SA and GA estimates is expected to be close to the CRB.

C. Simulated annealing (SA)

SA is a stochastic simulation method originally proposed in [13] which has proven useful in a wide range of complex combinatorial optimization problems. The process of annealing is analogous to the process of the optimization, in which the value of the cost function takes the role of the energy of the system, and the global optimum corresponds to the energy of the ground state of the system. In the SA algorithm, choosing a suitable *anneal schedule* is very important in order to ensure that the SA converges to a global optimum. This includes the choice of an initial temperature (T_o) and a temperature decrement rule. In this paper, we evaluate two traditional scheduling strategies:

(a) *Selection of initial temperature:* the value of T_o affects the trade off of the computational cost and the possibility of finding a global optimum. T_o should be high enough to *melt* the system at the beginning of the annealing. However, choosing an overly high value will consume too much computer time. In practice, T_o is often determined by [15]

$$T_o = \frac{\Delta \bar{E}}{\ln(\beta^{-1})}, \quad (4)$$

where $\Delta \bar{E}$ is the average increase in the cost function and β is a predetermined acceptance ratio.

(b) *Temperature decrement rule:* in order to avoid an impractically long time of computation, the temperature decrement should not be too slow. However, if the decrement is too fast, thermal equilibrium cannot be reached at each temperature. For this reason, a usually adopted decrement rule is the following [4]:

$$T_l = \alpha T_{l-1} \quad (5)$$

where T_l is the l-th temperature and $\alpha < 1$ controls the decrease in temperature.

For a more comprehensive description of SA, see [14].

D. Genetic Algorithm (GA)

GA is an iterative optimization technique inspired in the mechanics of genetics and natural selection. A group of candidate solutions, which are described as binary sequences (chromosomes), is viewed as a generation of certain population. Operations such as selection, crossover, and mutation are applied upon these chromosomes to produce new generations with better fitness. In this paper, we evaluate the GA taking into account the following criteria:

(a) *Roulette wheel selection mechanism:* in this case, GA assigns to each individual a selection probability which is directly proportional to its fitness. Then, the area of a sector in a roulette wheel (from where the individuals will be chosen) is set proportional to the fitness values. Finally, the selection of individuals is performed by independent spins of the wheel and, since better individuals have more space on the wheel, they will have more chance to be chosen.

(b) *Two point crossover mechanism:* here, GA creates crossover children of the given population using two available parents and choosing two random points A and B in its binary sequence. Then, the child will have the genes of the first parent at the locations after A and before B, and the genes of the second parent after B and before A.

(c) *Gaussian mutation mechanism:* genes in an individual are randomly interchanged using a Gaussian distribution.

(d) *Hybrid algorithm:* in many cases, combining GA with another method can significantly enhance the effectiveness of the GA [16]. In our case, we are interested in evaluating a hybrid algorithm which combines GA and constrained nonlinear optimization.

In addition to comparing the performance of all those variations of the GA, we evaluate the effect of different sizes of the population. For a more detailed description of GA see [17].

III. Numerical examples

We performed a series of numerical experiments for simulated EEG data corresponding to one and two dipole sources within a four-sphere head model [1]. In both cases, the data was generated using an array of $M = 37$ sensors. For the data produced by one dipole, the source was located (in spherical coordinates) at $[\vartheta = 0.5235,\ \varphi = -1.2,\ \varrho =$

$0.083]^T$ where ϑ and φ are the azimuth and the elevation angles given in radians, respectively, and ϱ is the eccentricity in meters. For the case of two dipoles, they were located at $[\vartheta = 0.5235,\ \varphi = 0.6,\ \varrho = 0.083]^T$ and $[\vartheta = 0.5235,\ \varphi = -0.6,\ \varrho = 0.083]^T$. The dipoles' magnitudes were allowed to change in time according to

$$q_\vartheta(t) = 15e^{-(\frac{t-60}{8})^2} - 5e^{-(\frac{t-40}{17})^2}[nA \cdot m], \quad (6a)$$

$$q_\varphi(t) = 13e^{-(\frac{t-60}{12})^2} - 3e^{-(\frac{t-40}{17})^2}[nA \cdot m], \quad (6b)$$

$$q_\varrho(t) = 0, \quad (6c)$$

i.e., $\boldsymbol{q}(t) = [q_\vartheta(t), q_\varphi(t), q_\varrho(t)]^T = \boldsymbol{q}_1(t) = \boldsymbol{q}_2(t)$. Note that, in the case of the experiments with two dipoles, this assumption would simulate correlated sources. Finally, we added uncorrelated random noise, distributed as $\mathcal{N}(0, \sigma^2)$ with different values of σ^2 in order to achieve mean values of $\mathrm{SNR} = -2, -1, 0.2, 0.6$ dB.

Under these conditions, the estimates of the source localizations were calculated by minimizing (2) with ϑ and φ being the parameters of interest, while ϱ was kept fixed to the surface of the sphere modeling the brain cortex (i.e., $\boldsymbol{\theta} = [\vartheta, \varphi, 0.083]^T$). Then, SA and GA were used in the optimization process and the influence of varying their operational parameters was evaluated, as well as the effect of different SNR values. The estimation process was repeated $K = 100$ times for each SNR value with independent noise realizations. Then, we computed the standard deviations of those estimates (denoted as $\sigma_{\hat{\theta}_x}, \sigma_{\hat{\theta}_y}, \sigma_{\hat{\theta}_z}$ in Cartesian coordinates), and compared them to the CRB square roots.

In the case of SA, the parameters we evaluated were T_o and the temperature decrement rule. The initial point of the search was considered as $\vartheta_o = 0,\ \varphi_o = 0$. We introduced lower and upper bounds constrains into the parameters of interest in all the cases. In two of the experiments, the SA initial temperatures were $T_o = 50, 100$ and the temperature decrement rule was computed as $T_{l+1} = T_o 0.95^l$ for $l = 0, 1, 2, \ldots$. In the third experiment, we used (4) with $\beta = 0.8$ and $\Delta\bar{E}$ was approximated from the mean value of 20 representative examples. Then, the optimization was performed using (5) with $\alpha = 0.85$.

For the GA optimization, we tested its performance for population sizes of $20, 50, 100,$ and 500 chromosomes in the experiments corresponding to one dipole, while for two dipoles we used lower and upper bounds constrains into the values of ϑ and φ and a population size of 100 chromosomes. Then, three experiments were performed: in the first one we used a traditional process, i.e. used the genetic operators described in Section II-D. In the second, we added a hybrid algorithm with a local constrained search based on a nonlinear multivariate algorithm. In the third experiment, we used the same selection and crossover operators as in the traditional process but changed the mutation operator by one which randomly generates directions that adapt as a function of the last successful generation [18]. In all the GA experiments, the individuals within the initial population were generated randomly without restrictions.

IV. Results and Discussion

The results of all our experimentation are shown in Figures (1) and (2) for the case of estimating one dipole, and Figures (3) and (4) for simultaneously estimating two dipoles.

In all cases we note that, for high SNR, the standard deviations of the estimates are close to the CRB square root without significant differences introduced by the variations in the operational parameters of the SA and GA algorithms. In the case of GA, we did not find significant differences between optimizing the CLF with a low population size (20 chromosomes) in comparison to larger populations (500 chromosomes). In addition, the standard deviations were very similar independently of the genetic operators used or if hybridizing was applied for all population sizes.

Similar results were obtained in the case of SA, where varying T_o or the decrement rule did not change the variances. However, in the example corresponding to estimating two dipoles, we found that SA was more sensitive to changes in T_o under all conditions of SNR. Nevertheless, our results

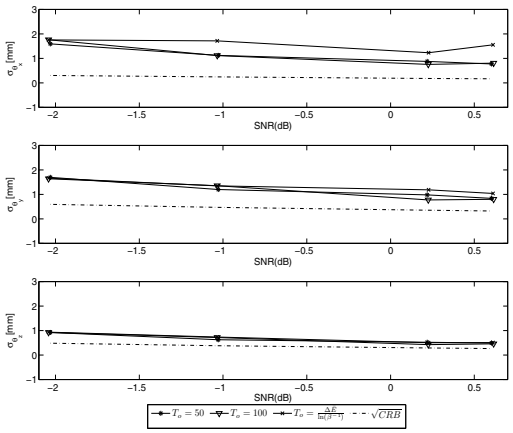

Fig. 1. Standard deviation of $\hat{\boldsymbol{\theta}}$ for one dipole using SA with different initialization parameters as a function of SNR.

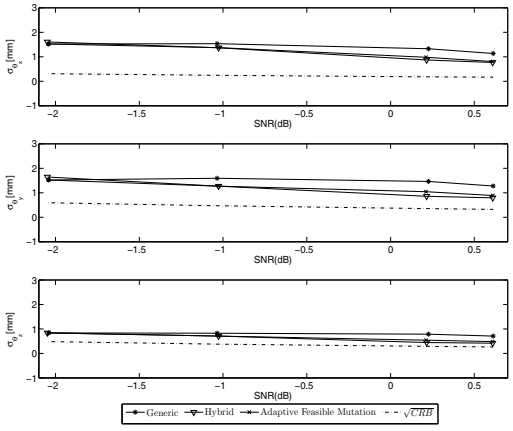

Fig. 2. Standard deviation of $\hat{\boldsymbol{\theta}}$ for one dipole as a function of SNR and for different implementations of the GA.

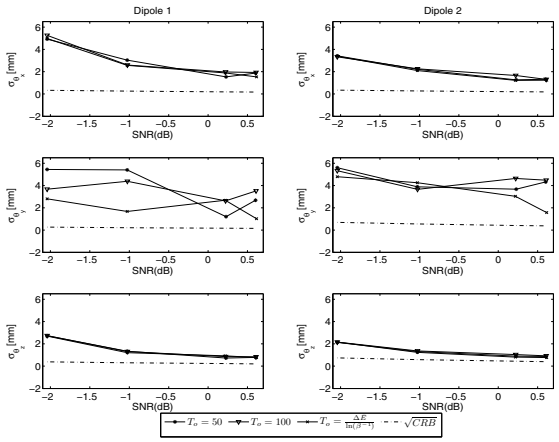

Fig. 3. Standard deviation of $\hat{\boldsymbol{\theta}}$ for two dipoles using SA with different initialization parameters as a function of SNR.

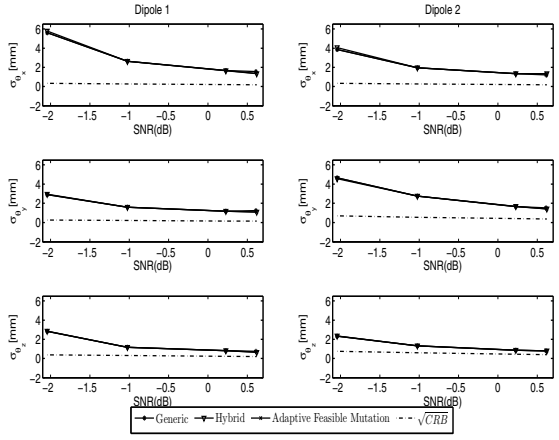

Fig. 4. Standard deviation of $\hat{\boldsymbol{\theta}}$ for two dipoles as a function of SNR and for different implementations of the GA.

showed that in a two-dipole estimation SA benefits from using a very high value of T_o like the one computed using (4). However, this strategy makes SA less attractive in comparison to GA as the computational cost increases.

In summary, GA and SA are effective in localizing one dipole. GA is particularly efficient given that the population size used in the experiments was rather small. Furthermore, our results also showed that the hybrid GA with local search was a suitable strategy. Finally, we noted that GA outperforms SA in the case of two dipoles as SA seems to be easily trapped in local minima, which is reflected by the standard deviations of SA being relatively larger than those obtained through GA and very different to those for the case of estimating one dipole.

V. Conclusion

We presented a scheme to evaluate the performance of SA-based and GA-based EEG source localization methods using the CRB as reference. Our study showed that no significant variations on the performance are introduced by changing the selection of the operational parameters of the SA and GA, while in general GA has a better performance. However, both methods seem to be very sensitive to the presence of noise, while SA also seems to be sensitive to the introduction of correlated sources as seen in beamforming methods [19]. Our preliminary tests also showed that GA has the advantage of being computationally less costly than SA. However, a more strict evaluation of the computational times is required. This evaluation will be part of the future work, as well as a more intensive experimentation using real EEG data, and the evaluation of other metaheuristic methods such as PSO and differential evolution (DE).

References

[1] J. Mosher, R. Leahy, and P. Lewis, "EEG and MEG: forward solutions for inverse methods, " *IEEE Transactions on Biomedical Engineering*," vol. 46, no. 3, pp. 245-259, 1999.

[2] D. McNay, E. Michielssen, R. L. Rogers, S. A. Taylor, M. Akhtari, and W. W. Sutherling, "Multiple source localization using genetic algorithms," *Journal of Neuroscience Methods*, vol.64, pp. 163-172, 1996.

[3] K. Uutela, M. Hämäläinen, and R. Salmelin, "Global optimization in the localization of neuromagnetic sources," *IEEE Transactions on Biomedical Engineering*, vol. 45, pp. 716-723, 1998.

[4] D. Khosla, M. Singh and M. Don, "Spatio-temporal EEG source localization using simulated annealing," *IEEE Transactions on Biomedical Engineering*, vol. 11, pp. 1075-190, 1997.

[5] C. Sequeira, F. Sanchez-Quesada, M. Sancho, I. Hidalgo, and T. Ortiz, "A genetic algorithm approach for localization of deep sources in MEG," *Physica Scripta*, vol. 118, pp. 140-142, 2005.

[6] C. Jiang, J. Ma, B. Wang, and L. Zhang, "Multiple signal classification based on genetic algorithm for MEG sources localization," in *Proceedings of the 4th International Symposium on Neural Networks*, Nanjing, China, 2007, pp. 1133-1139.

[7] K. E. Parsopoulos, F. Kariotou, G. Dassios, and M. N. Vrahatis, "Tackling magnetoencephalography with particle swarm optimization, " *International Journal of Bio-Inspired Computation*, vol. 1, no. 1/2, pp. 32-49, 2009.

[8] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley and Sons, New York, 1968

[9] D. Gutiérrez and A. Nehorai, "Array response kernels for EEG and MEG in multilayer ellipsoidal geometry," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1103-1111, 2008.

[10] P. Stoica and A. Nehorai, "On the concentrated stochastic likelihood function in array signal processing," *Circuits, Systems, and Signal Processing*, vol. 14, no. 5, pp. 669-674, 1995.

[11] D. Gutiérrez, A. Nehorai y C. Muravchik, "Estimating brain conductivities and dipole source signals with EEG arrays," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 12, pp. 2113-2122, 2004.

[12] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *Acoustics, IEEE Transactions on Speech and Signal Processing*, vol. 37, pp. 720-741, 1989.

[13] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671-680, 1983.

[14] E. H. L. Aarts, and J. H. M. Korst *Simulated Annealing and Boltzmann Machines*, John Wiley and Sons, 1989.

[15] H. Shakouri, K. Shojaee, and M. Behnam, "Investigation on the choice of the initial temperature in the simulated annealing: a mushy state SA for TSP, " in *17th Mediterranean Conference on Control & Automation*, Greece, 2009, pp. 1050-1055.

[16] X. Li, T. Jiang, and D. J. Evans, "Medical image reconstruction using a multi-objective genetic local search algorithm," *International Journal of Computer Mathematics*, vol.74, 301-314, 2000.

[17] C. R. Reeves and J. E. Rowe, *Genetic Algorithms – Principles and Perspectives: A Guide to GA Theory*, Kluwer Academic Publishers, Boston, MA, 2002.

[18] A. Neubauer, "Adaptive non-uniform mutation for genetic algorithms," in *Proceedings of the 5th Fuzzy Days International Conference*, Dortmund, Germany, 1997, pp. 24-34.

[19] B. D. Van Veen, W. van Drongelen, M. Yuchtman, and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering," *IEEE Transactions on Biomedical Engineering*, vol. 44, pp. 867-80, 1997.