Intensity Modulated Beam Radiation Therapy Dose Optimization with Multiobjective Evolutionary Algorithms

Michael Lahanas[1], Eduard Schreibmann[1,2], Natasa Milickovic[1] and Dimos Baltas[1,3]

[1] Department of Medical Physics and Engineering, Strahlenklinik, Klinikum Offenbach, 63069 Offenbach, Germany.
[2] Medical Physics Department, Medical School, Patras University,
[3] Institute of Communication and Computer Systems, National Technical University of Athens, 15773 Zografou, Athens, Greece.

Abstract. We apply the NSGA-II algorithm and its controlled elitist version NSGA-IIc for the intensity modulated beam radiotherapy dose optimization problem. We compare the performance of the algorithms with objectives for which deterministic optimization methods provide global optimal solutions. The number of parameters to be optimized can be up to a few thousands and the number of objectives varies from 3 to 6. We compare the results with and without supporting solutions. Optimization with constraints for the target dose variance value provides clinical acceptable solutions.

1 Introduction

Every year more than one million patients only in the United States will be diagnosed with cancer. Half of these will be treated with radiation therapy [1]. In teletherapy or external radiotherapy beams of penetrating radiation are directed at the tumor. Along their path through the patient body the beams deposit energy. Cancer cells have a smaller probability than healthy normal cells to survive the radiation damage. The dose is the amount of energy deposited per unit of mass. The physical and biological characteristics of the patient anatomy and of the source, such as intensity and geometry are used for the calculation of the dose function, i.e. the absorbed dose as a function of the location in the body. A physician depending on the patient, the size etc. prescribes the so called desired dose function. The objectives of dose optimization are to deliver a sufficient high dose in the cancerous tissue and to protect the surrounding normal tissue (NT) and sensitive structures from excessive radiation. The problem is to determine a intensity distribution for the radiation sources so that the resulting dose function is equal to the desired dose function. The calculation of the dose function for a given intensity distribution is possible with a high accuracy, whereas the inverse problem, i.e. the determination of the intensity distribution for a given dose function is with some exceptions not possible as the inverse dose operator produces non-physical solutions with negative intensities. Optimization algorithms

are used to minimize the difference between the desired and the obtained dose function.

In the past the multiobjective (MO) dose optimization problem has been transformed into a single objective problem using as a score function the weighted sum of the individual objective functions. The weights, called also importance factors, have been determined by trial and error methods. The treatment planner was required to often repeat the optimization with other importance factors until some satisfactory solution was obtained. We have recognized that a better method is to produce a representative set of solutions and to select out of these the best possible. Even if methods have been proposed to select automatically optimal weights [2], these require additional importance factors which are *a priori* not known. The MO optimization method is important as it provides the treatment planner information about the trade-off between the objectives and the limitations of the available solutions. Gradient based optimization algorithms can be used only for variance based objectives [3]. The most used optimization method is simulated annealing (SA)[4] which can be applied for all types of objectives but requires a very large number of iterations and it is practical not possible in clinical relevant time to produce a sufficient large number of solutions. We used in the past MO evolutionary algorithms for dose optimization in brachytherapy [5], which is another radiation based cancer treatment method. The resulting dose distribution must satisfy similar objectives. In radiotherapy the sources are outside the patients body and the problem is to find the beam directions and intensity of the beams so that the resulting dose distribution satisfies various criteria. The number of parameters is much larger than in brachytherapy and it can be as large as 2000 - 10000. Previous methods used in radiotherapy include SA, iterative approaches and filtered back-projection [6]. We applied successfully gradient based optimization algorithms which are fast enough to produce a large number of solutions [3]. Problems such as the selection of beams in two dimensions have been considered with MO evolutionary algorithms by Haas *et al* [7],[8]. Various single objective genetic algorithms have also been used. Knowles *et al* [9],[10]used EA for training neural networks which bypass the optimization problem by learning from previous optimization results.

From our experience with deterministic MO algorithms we know that high quality solutions can only be obtained by analyzing the trade-off information provided by a representative non-dominated set. We use realistic three-dimensional cases with a large number of optimization parameters. We study the possibility of the use of MO evolutionary algorithms for the intensity modulated beam radiotherapy (IMRT) dose optimization problem.

2 Methods

2.1 Intensity modulated beam radiotherapy

In IMRT each beam is divided in a number of small beamlets (bixels), see Fig. 1. The intensity of each beamlet can individually be adjusted. The geometry of the planning target volume (PTV) which includes the tumor and a margin

and organs at risk (OAR) is specified by contours. Tools can be used to select the number of beams and their directions based on geometric criteria. The dose at each sampling point in the PTV and other structures is calculated from the contributions of each beamlet of each beam and its corresponding intensity. A sparse dose matrix is precalculated and contains the dose value at each sampling point from each bixel with a unit radiation intensity. The intensity (weights) of the beamlets have to be determined such that the produced dose distribution is "optimal".

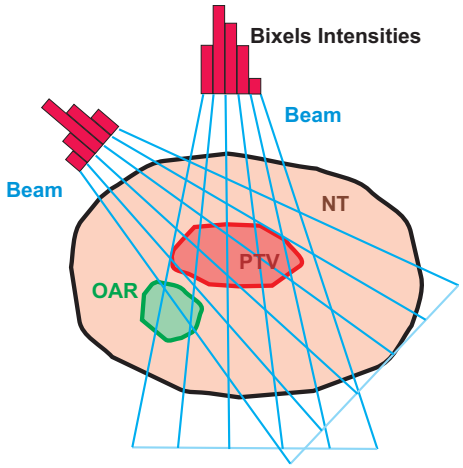

Fig. 1. Principle of IMRT dose optimization. The contours of the body, the PTV and one OAR are shown. The problem is to determine the intensities of the tiny subdivisions (bixels) of each beam, so that the resulting dose distribution is optimal.

Radiation oncologist use for the evaluation of the dose distribution quality a cumulative dose volume histogram (DVH) for each structure (PTV, NT or OARs), which displays the fraction of the structure that receives at least a specified dose level. If the objectives are expressed in terms of DVHs related values, then the objectives are called DVH based objectives.

2.2 Variance based objectives

The optimization goals can be expressed also with variance based objectives which are only indirect related to the DVHs values but they are used because deterministic gradient based optimization algorithms can be applied. The objective functions are: for the PTV the dose variance f_{PTV} around the prescription dose D_{ref}, for NT the sum of the squared dose values f_{NT} and for each OAR the variance f_{OAR} for dose values above a specific critical dose value D_{cr}^{OAR}.

$$f_{PTV} = \frac{1}{N_{PTV}} \sum_{j=1}^{N_{PTV}} (d_j^{PTV} - D_{ref}^{PTV})^2 \tag{1}$$

$$f_{NT} = \frac{1}{N_{NT}} \sum_{j=1}^{N_{NT}} (d_j^{NT})^2 \tag{2}$$

$$f_{OAR} = \frac{1}{N_{OAR}} \sum_{j=1}^{N_{OAR}} \Theta(d_j^{OAR} - D_{cr}^{OAR})(d_j^{OAR} - D_{cr}^{OAR})^2 \tag{3}$$

$\Theta(x)$ is the Heaviside step function. d_j^{PTV}, d_j^{NT} and d_j^{OAR} are the calculated dose values at the j-th sampling point for the PTV, the NT and each OAR respectively. N_{PTV}, N_{NT} and N_{OAR} are the corresponding number of sampling points.

2.3 L-BFGS

We generate a representative set of non-dominated global optimal solutions with the L-BFGS algorithm [11]. This deterministic gradient based optimization algorithm requires derivatives of the objective function. We produce a representative set of solutions using a weighted sum $f(x)$ of the M single-objective functions $f_j(x)$, $j = 1, \ldots, M$. Normalized and uniformly distributed weights w_j are taken from the set of importance vectors W.

$$f(x) = \sum_{j=1}^{M} w_j f_j(x) \tag{4}$$

$$W = \left\{ [w_1, ..., w_M]; \sum_{j=1}^{M} w_j = 1; w_i \in [\frac{0}{k}, \frac{1}{k}, \ldots, \frac{k-1}{k}, 1] \right\} \tag{5}$$

We call k the sampling parameter. L-BFGS is especially suited for high-dimensional problems as for N parameters it uses indirectly an approximation of the Hessian matrix using only an order of $5N$ instead N^2 operations. A representative set of solutions is generated by repeating the optimization with L-BFGS each time with a different vector of importance factors from the set W . The stopping criteria are 300 iterations or a tolerance value of 10^{-6}.

2.4 NSGA-II and NSGA-IIc

For the MO evolutionary optimization we use the non-dominated sorting genetic algorithm NSGA-II [12] and the controlled elitist version NSGA-IIc [13]. The population of these algorithms can be initialized by a specific number of solutions generated with L-BFGS. We call these solutions *supporting solutions* [14],[15]. This is necessary as the number of objectives depending on the number

of OARs to be considered is in the range 3-6. The calculation of the objective values requires a significant fraction of the optimization time due to the large number of sampling points where the dose has to be calculated. It is practical not possible to allow the algorithm to evolve for thousands of generations. In clinical practice time is important and a representative set of the Pareto front has to be found in a few minutes. In order to obtain a sufficient large set of solutions we archive all non-dominated solutions found during the optimization. Dominated solutions are removed from this archive. This external archived population is not used in the optimization directly like in the PAES [16] or SPEA [17] algorithm. We include a comparison of NSGA-II with its controlled elitist version NSGA-IIc where a lateral diversity is kept by allowing a sufficient number of individuals to survive in various non-dominated fronts. The distribution is specified by the geometric parameter G. For the crossover operator we use simulated binary crossover SBX [18], [19]. This operator produces near parent solutions with increasing probability as the population evolves. The probability of generating near parent solutions increases with a parameter, the distribution index η_c. For mutation we use a similar operator specified by the corresponding index η_m. The mutation and crossover probability used was 0.01 and 0.9 respectively.

3 Results

3.1 Comparison of unconstrained optimization with NSGA-II and NSGA-IIc

We compare the results from L-BFGS with NSGA-II and NSGA-IIc for two clinical cases. The first is a brain tumor case, see Fig. 2. We consider only a two-dimensional dose optimization, i.e. only the dose distribution in a slice is optimized. Four beams are used and each beam is divided in 22 bixels. The number of parameters to be optimized is 88. We have five objectives by considering the PTV, the left and right lobus temporalis, the left eye and the NT, i.e. the brain as OARs. We use this clinical case to study the dependence of the evolutionary algorithms results on the population size, number of generations and other genetic parameters. The second case is a prostate tumor case, see Fig. 3. Again four beams are used. Each beam is divided in $22 \cdot 22$ bixels. The number of parameters for a full three-dimensional optimization is 1936. The dose is calculated at 15000 sampling points. For this case we have with the PTV the NT and two OARs four objectives. We use this case in order to look at the performance of NSGA-II, if the number of parameters is very large. The calculations were performed using a 933 MHz Intel III Windows NT computer with 512 MB RAM.

The population size was 200. For the crossover and mutation we set $\eta_c = \eta_m = 10$. The number of accumulated non-dominated solutions in the archive obtained by NSGA-IIc as a function of the geometric factor G after 200 generations is shown in Fig. 4. A maximum at 0.6 is observed which is close to G=0.65 found in [20] for a completely different problem. We use therefore G=0.65 for

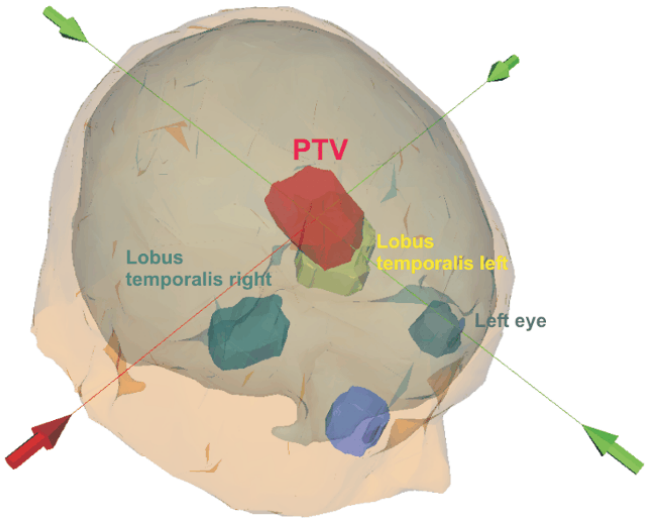

Fig. 2. 3D anatomy of the head tumor case. The PTV and the OARs are shown. The orientation of the four beams is shown.

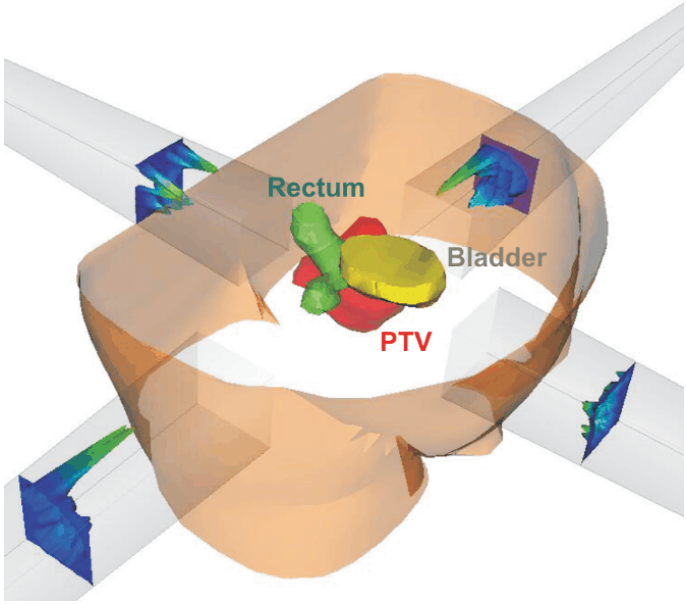

Fig. 3. 3D anatomy of the prostate tumor case. The PTV and the two OARs are shown. The orientation of the four beams is shown together with the intensity profile of the beams for one selected solution.

NSGA-IIc. The archived population from NSGA-II and NSGA-IIc without supporting solutions after 1000 generations is shown in Fig. 5. NSGA-IIc covers a larger part of the Pareto front found by L-BFGS, whereas for NSGA-II stagnation or premature convergence is observed.

The distribution of the archived non-dominated solutions of NSGA-II and NSGA-IIc with and without support is shown after 200 generations in Fig. 6 for the two out of ten two-dimensional projections of the Pareto front. The number of supported solutions were 35 (k=3) generated by L-BFGS. Without the supporting solutions parts of the Pareto front are not accessible even after a few thousand generations. Even if the supporting solutions improve the performance of NSGA-II its controlled elitist version is superior in the coverage and number of accumulated non-dominated solutions.

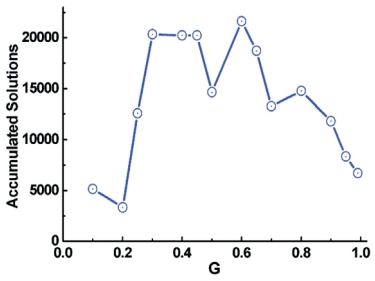

Fig. 4. Number of archived non-dominated solutions of NSGA-IIc as a function of the geometric factor G.

The six two-dimensional projections for the prostate tumor case are shown in Fig. 7. The results of NSGA-II and NSGA-IIc with supporting solutions are compared with the corresponding 2380 generated L-BFGS solutions. The population size was increased to 1000 for this high dimensional problem. The sampling parameter $k = 4$ was used with corresponds to 35 supported solutions.

3.2 Constraint optimization with NSGA-II

In IMRT only a part of the Pareto front is of practical interest. The dose variance in the PTV for clinical acceptable solutions is very small whereas for the OARs and the NT it can be much larger. It is important to apply constraints for the value of f_{PTV}.

3.3 Comparison of deterministic and evolutionary optimization results

We compare the spectrum of solutions (non-dominated set) obtained by a sequential application of L-BFGS and constrained optimization with the supported

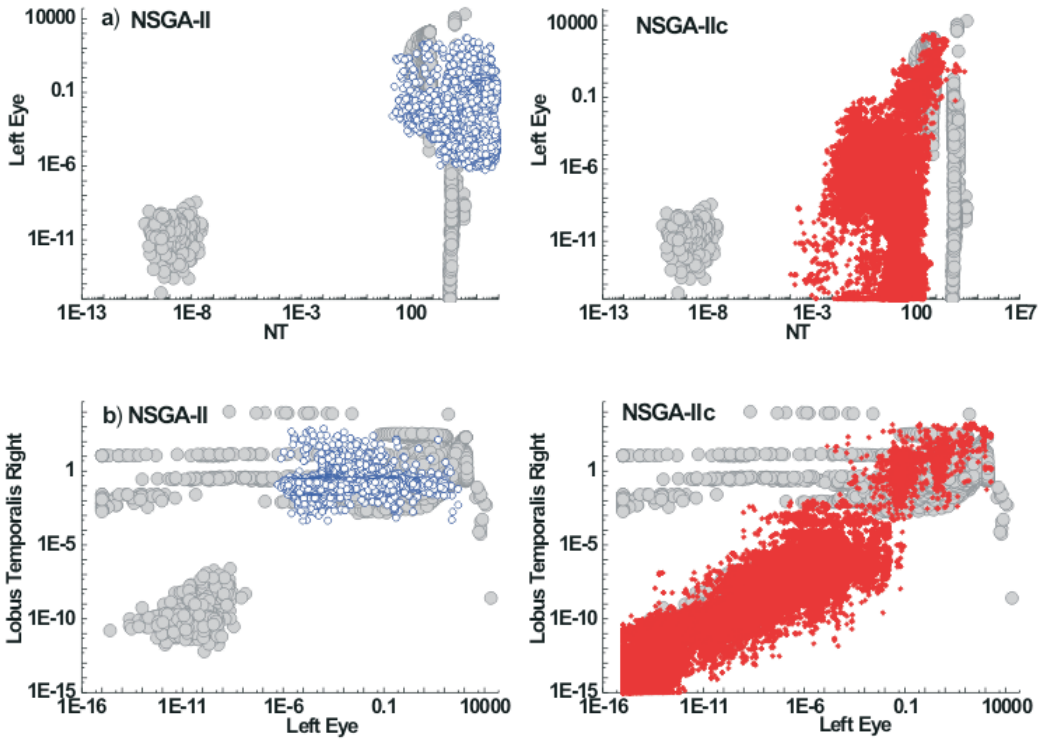

Fig. 5. Two examples of two-dimensional projections of the Pareto front of the archived non-dominated solutions for the five-dimensional dose optimization problem for the brain case. The solutions of NSGA-II and NSGA-IIc after 1000 generations are shown. The result of L-BFGS is included. The objectives values shown are the dose variances in organs at risk.

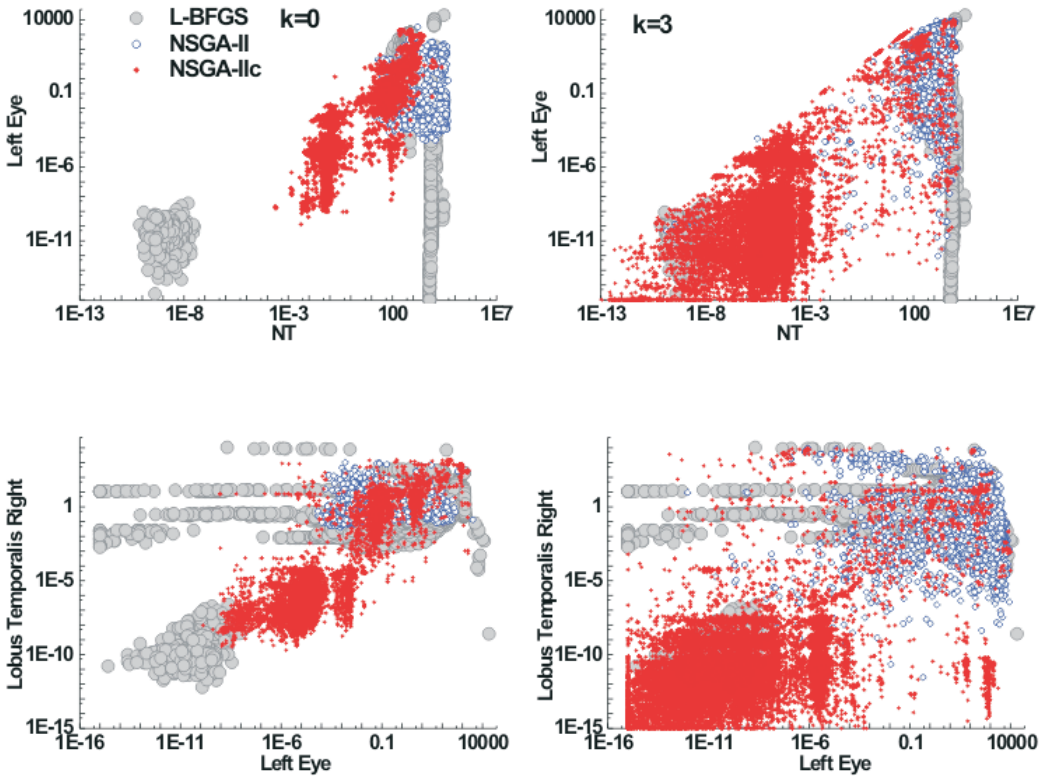

Fig. 6. The result as in Fig. 5 but only after 200 generations. The result from NSGA-II and NSGA-IIc with and without supporting solutions is shown.

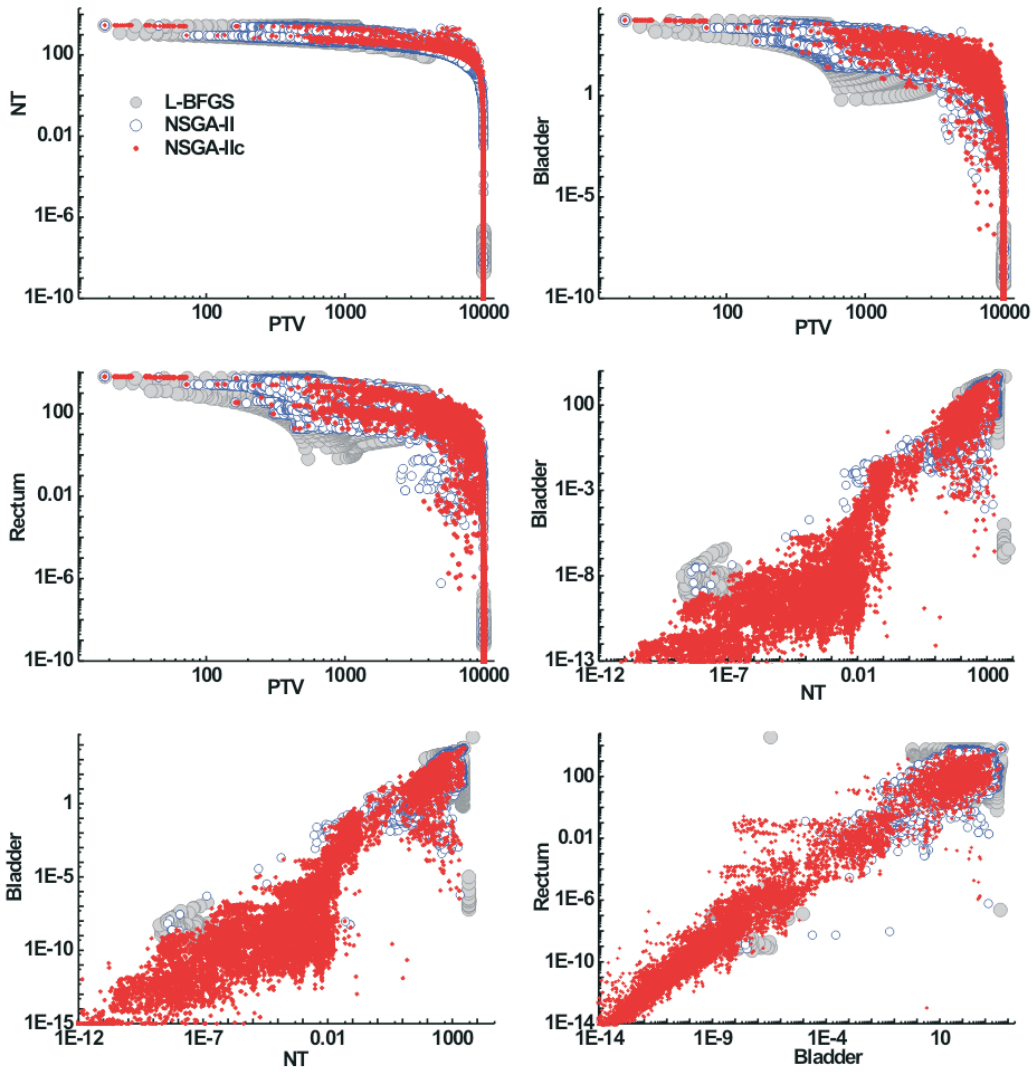

Fig. 7. Example of a the six two-dimensional projections of the four-dimensional Pareto front for the prostate tumor case. The result from L-BFGS, NSGA-II and NSGA-IIc with supported solutions.

NSGA-II algorithm. Our current implementation of NSGA-IIc does not support constraint optimization. The optimization with NSGA-II is performed with a population size of 500 and 100 generations. The resulting 2D projections of the Pareto front using 30 supported L-BFGS solutions is shown in Fig. 8. The result is compared with the Pareto front obtained by L-BFGS with 815 solutions. The calculation time for L-BFGS was 3860 s and for NSGA-II 271 s. We perform a 3D-dimensional dose optimization for the brain tumor with 856 bixels and nine beams. A display table with a list of objectives and dosimetric values of all solutions is used for the selection of the best solution. Constraints on these values can be applied. Solutions that satisfy these constraints are marked and their corresponding DVHs highlighted.

The spectrum of DVHs of solutions from NSGA-II is shown in Fig. 9 for the PTV, the left and right eye. The best solution has been selected as the solution with the smallest product of dose variances for the NT and the eyes. This solution obtained from the L-BFGS algorithm and NSGA-II is marked in Fig. 9.

4 Discussion and conclusions

In the past many single-objective optimization methods have been proposed for the dose optimization problem in radiotherapy. These methods were single-objective whereas the dose optimization problem is a MO problem. Gradient based optimization algorithms can be used with variance based objectives but if other objectives such as radiobiological or DVH based are used then stochastic algorithms such as SA are not efficient to produce a representative set of solutions. With a support of a small fraction of solutions by deterministic algorithms or SA the MO algorithm NSGA-IIc is able to produce efficiently a representative set of non-dominated solutions. The supporting solutions are necessary to guide a fraction of the population in parts of the objective space which are accessible only after a very large number of generations with Pareto ranking algorithms. A similar idea uses the genetic local search algorithm MOGLS [21] which using a scalarization of the individual objective functions performs a repeated local search in randomly specified directions. Also the local search hybrid method used with NSGA-II in [22] increased its performance. The supported solution approach is applied only at the start of the algorithm and requires a small number of solutions to be initialized, whereas the local search approach would require for the high dimensional and MO problem a very large number of function evaluations. The number of supported solutions depends on the number of objectives. A sampling parameter at least $k = 3$ is required.

The archiving of non-dominated solutions found is necessary in order to have a sufficient large set of representative solutions for the large number of objectives. It additionally allows the population size to be reasonable small and reduces the optimization time in order to obtain an approximate similar number of solutions as the archive contains. The number of accumulated solutions in 200 generations was 15000-20000 for both cases studied so that finally a filter has to be applied

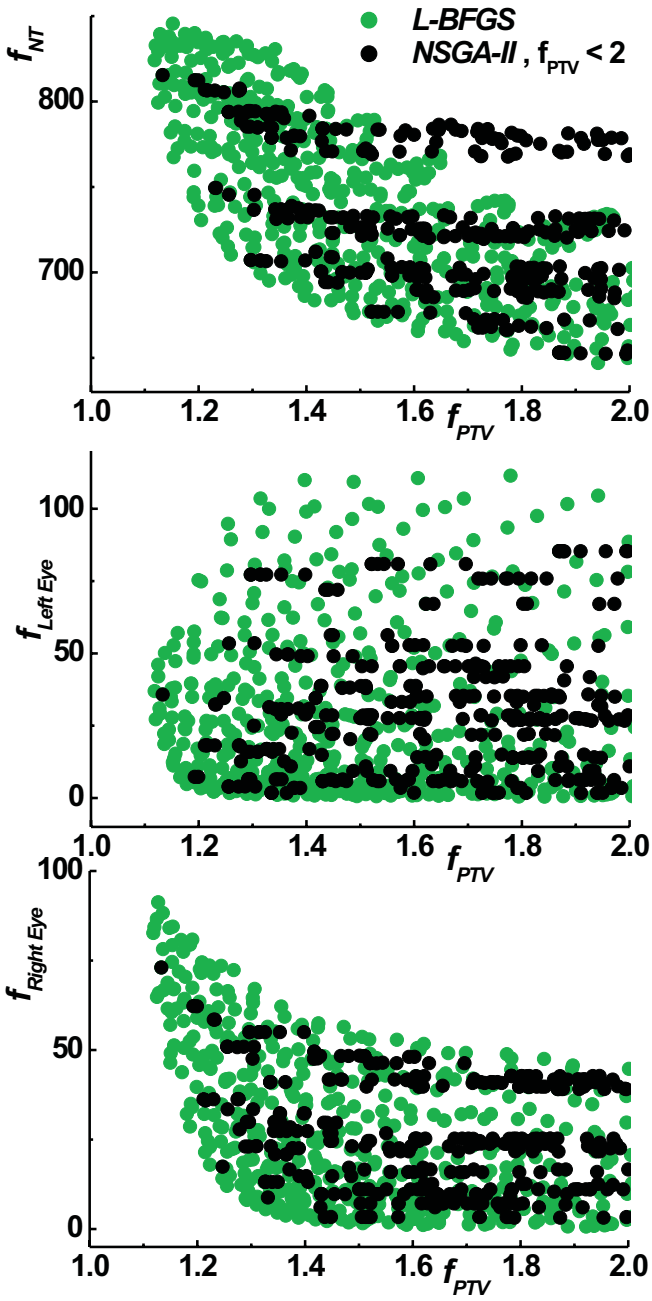

Fig. 8. Two-dimensional projections of the four-dimensional Pareto front for a brain tumor case. The result using L-BFGS and NSGA-II with supported solutions and the constraint $f_{PTV} < 2$ is shown.

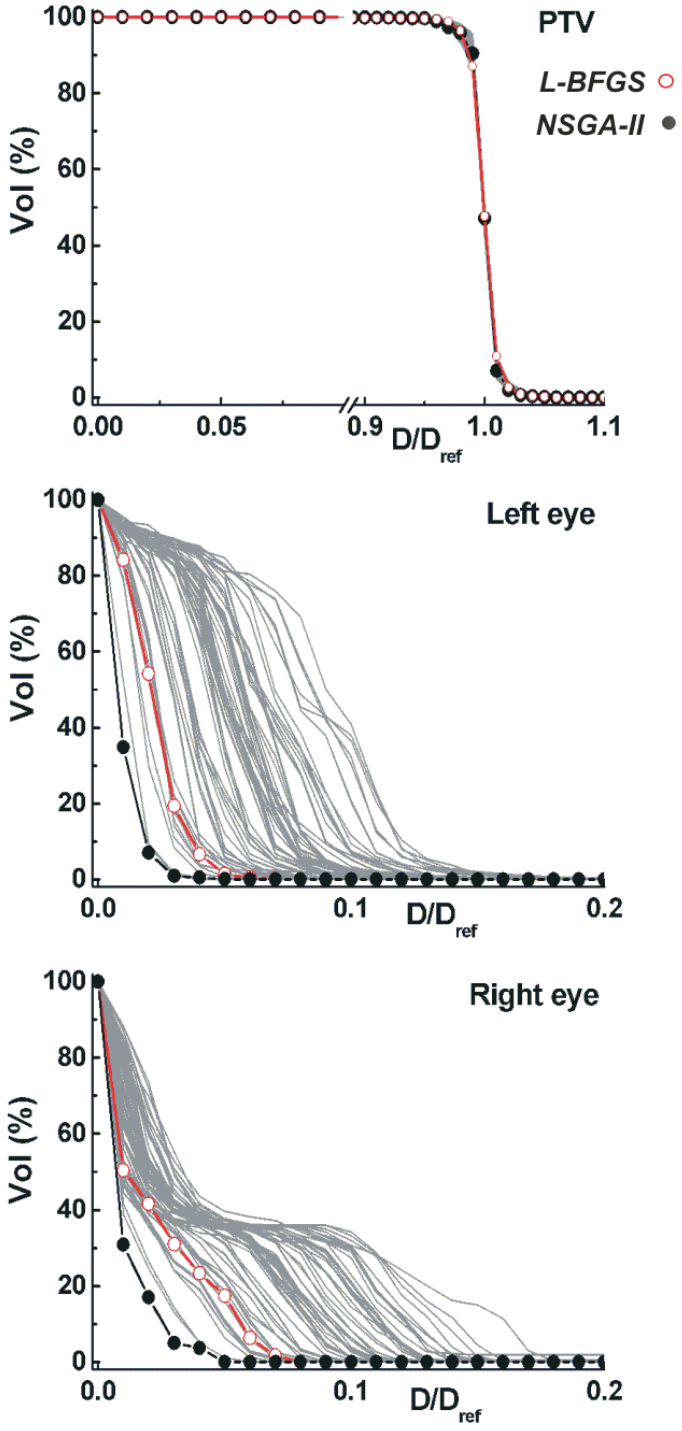

Fig. 9. Spectrum of DVHs of solutions obtained by NSGA-II for the PTV, the left and right eye for a brain tumor case. The best selected solution obtained by L-BFGS and NSGA-II is shown.

to reduce the number to an acceptable level. The results show that a lateral diversity is important to avoid a premature convergence and the controlled elitist algorithm benefits more from the support. Even without support NSGA-IIc is able to approach the global Pareto front better than NSGA-II which prematurely converges. For the problem with 1936 parameters a sufficiently large population with at least 500-1000 members is required. IMRT requires the use of constraint optimization with a PTV dose variance value less than 10. The spectrum of the obtained DVHs shows that we have a true MO optimization problem. NSGA-II with constraints for the f_{PTV} value produces clinical acceptable solutions even if the best result from L-BFGS is better. MO dose optimization for IMRT requires only a few minutes, whereas L-BFGS needs more than 10 times more time to produce a comparable number of solutions. The Pareto front obtained from NSGA-II is close to the global optimal Pareto front and this distance can possibly be reduced if constrained optimization is applied using NSGA-IIc. We hope that MO optimization with evolutionary algorithms will be applied in IMRT where currently only single-objective optimization algorithms are used. MO optimization provides not only a satisfactory solution but the best possible. This reduces the dose in the OARs and in the NT to a minimum possible level.

We want to use MO evolutionary optimization algorithms for inverse planning in IMRT where the optimal number of beams and their orientation additional must be found. For readers interested in experimenting with the IMRT dose optimization problem an extension of the MOMHLIB library [23] used in this study, together with dosimetric data sets and additional information, is available at the website: www.mlahanas.de/IMRTOpt.html.

Acknowledgments. The Research was Supported through a European Community Marie Curie Training Fellowship (HPTM -2000-00011).

References

1. Peres, C. A., and Brady, L. W.: Principles and practice of radiotherapy. Lippincott-Raven, Philadelphia, 3rd edition, 1998.
2. Xing, L., Li, J.G., Donaldson, S., Le, Q.T. and Boyer, A. L.: Optimization of importance factors in inverse planning. Phys Med. Biol., **44** 2525 – 2536, 1999.
3. Cotrutz, C., Lahanas, M., Kappas, C. and Baltas, D.: A multiobjective gradient based dose optimization algorithm for conformal radiotherapy. Phys. Med. Biol. **46** 2161 – 2175, 2001.
4. Webb, S.: Optimization of conformal radiotherapy dose distributions by simulated annealing. Phys. Med. Biol., **34** 1349 – 1370, 1990.
5. Lahanas, M., Baltas, D. and Zamboglou, N.: A hybrid evolutionary algorithm for multiobjective anatomy based dose optimization in HDR brachytherapy, to be published in Phys. Med. Biol. 2003.
6. Bortfeld, T., Ürkelbach, J., Boesecke, R. and Schlegel, W.: Methods of image reconstruction from projections applied to conformation therapy, Phys. Med. Biol. **35** 1423 – 1434, 1990.
7. Haas, O. C. L., Burnham K. J. and Mills J. A.: On Improving the selectivity in the treatment of cancer: a systems modelling and optimisation approach . J. Control Engineering Practice, **5** 1739 – 45, 1997

8. Haas, O. C. L.: Radiotherapy Treatment Planning: New System Approaches. Springer Verlag London, Advances in Industrial Control Monograph, ISBN 1-85233-063-5, 1999.
9. Knowles, J. D., Corne, D. and Bishop J. M.: Evolutionary Training of Artificial Neural Networks for Radiotherapy Treatment of Cancers in Proceedings of the 1998 IEEE International Conference on Evolutionary Computation IEEE Neural Networks Council, 0-7803-4871-0, pp 398–403, pp. 398–403
10. Knowles, J. D. and Corne, D.: Evolving Neural Networks for Cancer Radiotherapy, in Chambers, L.(ed.), Practical Handbook of Genetic Algorithms: Application 2nd Edition, Chapman Hall/CRC Press, pp. 443–448. ISBN L-58488-240-9, 2000.
11. Liu, D.C., and Nocedal, J.: On the limited memory BFGS method for large scale optimization. Mathematical Programming **45** 503 – 528, 1989.
12. Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. Technical Report 20001, Indian Institute of Technology, Kanpur, Kanpur Genetic Algorithms Laboratory (KanGAL), 2000.
13. Deb, K. and Goel, T.: Controlled elitist non-dominated sorting genetic algorithms for better convergence, in Proceedings of the first international conference. EMO 2001, Zurich, Switzerland, edited by E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello, D. Corne, Lecture Notes in Computer Science Vol. 1993, Springer 67 – 81, 2001
14. Gandibleaux, X., Morita, H. and Katoh, N.: The Supported Solutions Used as a Genetic Information in Population Heuristic, in Proceedings of the first international conference. EMO 2001, Zurich, Switzerland, edited by E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello, D. Corne, Lecture Notes in Computer Science Vol. 1993, Springer 429 – 42. 2001.
15. Milickovic, N., Lahanas, M., Baltas, D. and Zamboglou, N.: Comparison of Evolutionary and Deterministic Multiobjective Algorithms for Dose Optimization in Brachytherapy, in Proceedings of the first international conference. EMO 2001, Zurich, Switzerland, edited by E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello, D. Corne, Lecture Notes in Computer Science Vol. 1993, Springer 167 – 180. 2001.
16. Knowles, J.D., Corne, D.W.: Approximating the nondominated front using the Pareto Archived Evolution Strategy. Evolutionary Computation **8** 149 – 172, 2000.
17. Zitzler, E., Thiele, L.: Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach. IEEE Transactions on Evolutionary Computation **37** 257 – 271, 1999.
18. Deb, K. and Agrawal, R.B.: Simulated binary crossover for continuous search space. Complex Systems **9** 115 – 148, 1995.
19. Deb, K. and Beyer, H.G.: Self-Adaptive genetic Algorithms with Simulated Binary Crossover. Evolutionary Computation **9** 197 – 221, 2001.
20. Deb, K., Multi-Objective Optimization using Evolutionary Algorithms, Chichester, Wiley, UK, 200l.
21. Jaszkiewicz, A. Genetic local search for multiple objective combinatorial optimization, Technical Report RA-014/98, Institute of Computing Science, Poznan University of Technology, 1998.
22. Goel, T. and Deb, K.: Hybrid Methods for Multi-Objective Evolutionary Algorithms. KanGAL Report Number 2001004, 2001.
23. http://www-idss.cs.put.poznan.pl/˜jaskiewicz/MOMHLIB/