Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms

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

1. Department of Medical Physics & Engineering, Strahlenklinik, Klinikum Offenbach, 63069 Offenbach, Germany.
2. Medical Physics Department, Medical School, Patras University, 26500 Rio, Greece
3. Institute of Communication & Computer Systems, National Technical University of Athens, 15773 Zografou, Athens, Greece.

Corresponding author:

Michael Lahanas
Dept. of Medical Physics & Engineering
Strahlenklinik, Klinikum Offenbach
Starkenburgring 66
63069 Offenbach am Main, Germany
Tel.: +49 – 69 – 8405 –3235
Fax. : +49 – 69 – 8405 –4481
E-mail: mlahanas@gmx.de

Abstract

We consider the behavior of the limited memory *L-BFGS* algorithm as a representative constraint-free gradient-based algorithm which is used for multiobjective (MO) dose optimization for intensity modulated radiotherapy (IMRT). Using a parameter transformation, the positivity constraint problem of negative beam fluences is entirely eliminated: a feature which to date has not been fully understood by all investigators.

We analyze the global convergence properties of *L-BFGS* by searching for the existence and the influence of possible local minima. With a fast simulated annealing algorithm *FSA* we examine whether the *L-BFGS* solutions are globally Pareto optimal. The three examples used in our analysis are a brain tumor, a prostate tumor and a test case with a C-shaped PTV. In one percent of the optimizations global convergence is violated. A simple mechanism practically eliminates the influence this failure and the obtained solutions are globally optimal. A single-objective dose optimization requires less than 4 seconds for 5400 parameters and 40000 sampling points.

The elimination of the problem of negative beam fluences and the high computational speed permits constraint-free gradient-based optimization algorithms to be used for MO dose optimization. In this situation, a representative spectrum of possible solutions is obtained which contains information such as the trade-off between the objectives and range of dose values. Using simple decision making tools the *best* of all the possible solutions can be chosen.

We perform MO dose optimization for the three examples and compare the spectra of solutions, firstly using recommended critical dose values for the organs at risk and secondly, setting these dose values to zero.

1. Introduction

The desired dose distribution in radiotherapy cannot always be obtained, due to physical limitations and to the existence of trade-offs between the various conflicting optimization objectives. We therefore have a MO optimization problem to solve. It is, though, important to realize that MO provides a spectrum of possible solutions and not just a single solution.

The implementation of MO dose optimization in brachytherapy was first described by Lahanas *et al* (1999) and in IMRT by Cotrutz *et al* (2001) is a new approach of inverse planning. Haas *et al* (1998) used for the first time to our knowledge a MO evolutionary algorithm in radiotherapy for optimization of the beam directions in two-dimensions without employing a dose optimization. In Haas (1999) the use of MO dose optimization for external beam radiotherapy is proposed.

The trial and error method used by treatment planners, which involves modifying the importance factors in order to obtain a satisfactory solution is in MO replaced by the determination of a representative set of so-called *efficient* solutions out of which the solution with the smallest compromise on all objectives for the treatment can be obtained. We have a set of solutions in which each solution is characterized by a different set of importance factors used to generate the solution.

This off-line or *a posteriori* approach provides information for all possible dose distributions which can be obtained for a given set of objective functions. This information is used for the selection of the best solution. The treatment planner is not assumed to have any knowledge of what the limitations of dose distributions are that can be obtained or what are the ideal importance factors. An *a priori* MO optimization method was proposed by Yu (1997) which assumes that the planner has such knowledge. The treatment planner is assumed to provide satisfaction constraints, goals, priorities, importance factors etc. Alternatively, there is the *a priori* approach of Xing *et al* (1999) where a solution is obtained using *ideal* dose-volume histograms (DVH). A search engine was proposed which determines the set of importance factors for which a solution is obtained with DVHs that are as close as possible to the ideal

DVHs in terms of a defined metric. Even if the algorithm described by Yu uses some artificial intelligence finally the *a priori* methods provide a single solution for which we do not know its *optimality*, i.e. its relation to all other possible solutions.

For both methods it is necessary to know the limitations of the single objective optimization algorithm which has to be applied many times with different set of importance factors.

We use as decision variables the square root of the beam fluences (weights) Cotrutz *et al* (2001) which eliminate the problem of solutions with negative beam fluences of previous algorithms. Previous methods, Mohan *et al* (1994), have considered modifications of the line search algorithms and/or correction mechanisms for the negative weights which cannot be avoided. This has had the effect of either reducing the quality and/or increasing the optimization time.

Important for deterministic gradient-based algorithms is the problem of global convergence as described for brachytherapy by Lahanas *et al* 2003 and in radiotherapy by Rowbottom *et al (*2002). It was reported that the results approximately depended on the initial value of the beam fluences (Llacer *et al* 2001). This was attributed either to local minima in which the algorithms were trapped, or as the result of corrections applied for the elimination or reduction of the number of non-physical solutions with negative beam fluences. One such correction method is the truncation to zero of all negative weights found at the end of each optimization step. The number of negative weights in some cases can be larger than the number of positive weights (Mohan *et al.* 1994). This approach alters the result which consequently may differ significantly from the actual global optimal solution.

Rowbottom *et al (*2002) observed *local minima* using a downhill simplex optimization algorithm. For gradient-based optimization algorithms such an analysis has never been presented. Llacer *et al* (2003), although the title of their paper is “Absence of multiple local minima…” find that the final score function values sometimes depend on the initial beam weights and the case studied. An iterative method was used and correction methods for negative fluences were applied.

The parameter transformation does not require any correction mechanism, therefore the influence of negative weights is eliminated and a constrained free true gradient-based optimization algorithm is used for the first time in IMRT dose optimization.

In comparison to the conjugate gradient algorithm used by Cotrutz *et al* (2001) we use the limited memory algorithm *L-BFGS* which is a few times faster and requires also less memory. We extend the study to three dimensional clinical cases with more than 5000 bixel intensities. We include a study of the dependence of the results on the initial value which defines the so-called global convergence properties and also the nature of the obtained minima. We obtain a representative efficient set using a set of importance factors. We compare the results using simulated annealing *SA* to determine whether the solutions obtained by *L-BFGS* are Pareto global optimal solutions. We study individual objective values to look for the presence of possible degenerate states. We use standard dose variances as objectives and not normalized scale invariant objectives used by Cotrutz *et al*.

Inverse planning in radiotherapy considers the problem of obtaining a solution that satisfies as best as possible clinical defined criteria given physical constraints and limitations. In principle we have the set all possible fluence distributions that are physical possible, the so-called *fluence space*. The set of all possible dose distribution defines the *dose space*. We want to obtain a dose distribution from this set and the corresponding fluence distribution that is as close as possible to a desired dose distribution. What part of the dose space is available depends also on which optimization algorithm we use and what objectives we consider. MO provides a representative set of the dose space available for a given set of objectives.

We apply such a comparison of the spectra of solutions obtained by MO dose optimization for the three examples, firstly using recommended critical dose values for the organs at risk (OAR) and secondly, setting these dose values to zero.

2. Methods

2.1 Constraint-free gradient-based optimization algorithms

We use the limited memory *BFGS* algorithm *L-BFGS* by Liu and Nocedal (1989) described by the following algorithm. Let $f(\mathbf{x})$ be a function twice differentiable and convex and $\mathbf{x}_k$ the vector of the *N* optimization parameters, $\mathbf{g}_k = \mathbf{g}(\mathbf{x}_k)$ the gradient of f at $\mathbf{x}_k$ and $\mathbf{H_k}$ the approximation of the inverse Hessian of f at the k^{th} iteration.

Giving a starting point $\boldsymbol{x}_0$ and $\boldsymbol{H}_0$ a positive definitive approximation of the inverse Hessian at $\boldsymbol{x}_0$

For iteration k = 0 until stopping criterion is satisfied

1. *Determine a descent direction* $\mathbf{p}_k = -\mathbf{H_k}\nabla f(\mathbf{x_k})$
2. *Line search: Choose a step size* $\alpha_k = \arg\min_{\alpha>0} f(\mathbf{x_k} + \alpha\mathbf{p_k})$
3. *Update* $\mathbf{x_{k+1}} = \mathbf{x_k} + \alpha\mathbf{p_k}$
4. *Update* $g_{\mathbf{k+1}}$
5. *Compute* $\mathbf{H}_{k+1}$ *by updating* $\mathbf{H}_k$
4. *k = k+1*

End.

At each iteration step a line search procedure is used to determine the step size for the optimizer of the objective function in a chosen direction. The line search algorithm (which should really be called a ray search algorithm) finds the exact step size ("exact line") to the minimum of f along the ray $\mathbf{x_k} + \alpha\mathbf{p_k}, \alpha \geq 0$ or an approximation ("soft line"). *L-BFGS* uses a backtracking line search. For convergence the step size has to be chosen such that a sufficient decrease criterion is satisfied, which depends on the local gradient and function value and is specified in *L-BFGS* by the Wolfe conditions, see Liu and Nocedal (1989).

The difference between the standard *BFGS* algorithm (Press *et al*) and *L-BFGS* is the method for the inverse Hessian update in step 5. *BFGS* requires $O(N^2)$ operations for the

update whereas *L-BFGS* requires only 4*mN* operations to calculate the descent direction indirectly from the *m* last values of $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ and $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$.

L-BFGS uses a set of *m* values $\{\mathbf{s}_i, \mathbf{y}_i\}, i = k-m,...,k-1$ and a basic diagonal matrix $\mathbf{H}_k^{(0)}$ is used for the update using the *m* pairs. A value of *m* = 5 is recommended. We say that $\mathbf{H}_{k+1}$ is obtained by updating $\mathbf{H}_k$ using the pairs $\{\mathbf{s}_i, \mathbf{y}_i\}$. Liu and Nocedal (1989) have found that the update can be done very effective using symmetry principles. After the new iterate the oldest of the *m* pairs $\{\mathbf{s}_i, \mathbf{y}_i\}$ is replaced by the newest pair.

The required memory for the m pairs $\{\mathbf{s}_i, \mathbf{y}_i\}$ is 2*mN*+O(*N*) whereas *BFGS* requires $N^2/2$ for the matrix $\mathbf{H}_k$. *L-BFGS* is very effective for problems with a very large number of parameters such as required for IMRT dose optimization.

We allow *L-BFGS* to run either until a maximum number of iterations is reached or the following criterion using the parameter ε is fulfilled.

$$\frac{\|\nabla f(\mathbf{x}_k)\|_2}{\max(1, \|\mathbf{x}_k\|_2)} < \varepsilon$$

where $\mathbf{x}_k$ is the optimal solution at the k^{th} iteration.

2.2 Multiobjective optimization

MO or multicriteria optimization or vector optimization is the problem of determining “A vector of decision variables which satisfies constraints and optimizes a vector function whose elements represent *M* objective functions Miettinen (1999).

We call *decision variables* $\mathbf{x}_j$, j=1,2,...,*N* for which values are to be chosen in an optimization problem. In order to know how "good" a certain solution is we need to have some criteria for evaluation. These criteria are expressed as computable functions $f_1(\mathbf{x}), f_2(\mathbf{x}),..., f_M(\mathbf{x})$ of the decision variables, which are called *objective functions*. These form a vector function **f**. In general, some of these will be in conflict with others, and some will have to be minimized while others are maximized. The multiobjective optimization problem can be

now defined as the problem to find the vector $\mathbf{x}=(x_1,x_2,...,x_N)$, i.e. solution which optimize the vector function **f**

The vector function **f**(**x**) is a function that maps the set **X** in the set **F** that represents all possible values of the objective functions. Normally we never have a situation in which all the $f_i(\mathbf{x})$ values have a optimum in **X** at a common point **x**. We therefore have to establish certain criteria to determine what would be considered an ''optimal'' solution. One interpretation of the term optimum in multiobjective optimization is the Pareto optimum.

A solution $\mathbf{x}_1$ dominates a solution x_2 if the two following conditions are true:

1) $\mathbf{x}_1$ is no worse than $\mathbf{x}_2$ in all objectives, i.e. $f_j(\mathbf{x}_1) \le f_j(\mathbf{x}_2) \ \forall \mathrm{j}=1,...,\mathrm{M}$
2) $\mathbf{x}_1$ is strictly better than $\mathbf{x}_2$ in at least one objective, i.e. $f_j(\mathbf{x}_1) < f_j(\mathbf{x}_2)$ for at least one $\mathrm{j} \in \{1,...,M\}$

We assume, without loss of generality, that this is a minimization problem. $\mathbf{x}_1$ is said to be non-dominated by $\mathbf{x}_2$ or $\mathbf{x}_1$ is *non-inferior* to $\mathbf{x}_2$ and $\mathbf{x}_2$ is dominated by $\mathbf{x}_1$. Among a set of solutions P, the non-dominated set of solutions P' are those that are not dominated by any other member of the set P. When the set P is the entire feasible search space then the set P' is called the *global Pareto optimal set*. If for every member **x** of a set P there exists no solution in the neighborhood of **x** then the solutions of P form a *local Pareto optimal set*. The image of the Pareto optimal set is called the *Pareto front*.

We produce a representative set of solutions by repeating the optimization with *L-BFGS* using a weighted sum $f_{Tot}(\mathbf{x})$ of the single objective functions $f_j(\mathbf{x})$, j=1,…,M. Each time a different set of importance factors is used from the set W of normalized and uniformly distributed vectors of weights.

$$f_{Tot}(\mathbf{x}) = \sum_{j=1}^{M} w_j f_j(\mathbf{x}),$$

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

We call k the sampling parameter. For *M* objectives and a sampling parameter k we have $\binom{M+k-1}{M-1}$ such combinations.

2.3 Decision making

The two tasks of multiobjective optimization are:

- Obtaining a representative set of non-dominated solutions.
- Selecting a solution from this set, i.e. the decision making (DCM) process.

For single objective optimization methods there is only one optimization result and the only decision necessary is whether or not to accept the solution. For MO decision-making tools are necessary to filter a single solution from the non-dominated set that matches at best the goals of the treatment planner.

DCM tools are used in this study similar to that developed for the Real-Time HDR brachytherapy prostate planning system SWIFT™ (Nucletron B.V., Veenendaal, The Netherlands). A display table of a list of values for all solutions of the objectives, DVHs for all OARs, the NT and the PTV of each solution and various parameters are of clinical interest. Additionally, the extreme dose values are also provided. The entire table for every such quantity can then be sorted and solutions can be selected and highlighted by the treatment planner. Various DVH constraints can be applied and solutions that do not satisfy the constraints are removed from the list. This reduces the number of solutions and simplifies the selection of an optimal solution. The DVHs of all selected solutions can be displayed and compared.

Other DCM tools used are projections of the Pareto front onto a pair of selected objectives. For *M* objectives the number of such projections is $\binom{M}{2}$.

The position of selected solutions can be seen in these projections. This helps to identify their position in the multidimensional Pareto front and to quantify the degree of correlation between the objectives and of the possibilities provided by the non-dominated set.

2.4 Variance based optimization objectives

As objective functions we use the PTV 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_{crit}^{OAR}.

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

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

$$f_{OAR} = \frac{1}{N_{OAR}} \sum_{j=1}^{N_{OAR}} \Theta\left(d_j^{OAR} - D_{crit}^{OAR} \right) \left(d_j^{OAR} - D_{crit}^{OAR} \right)^2$$

$\Theta(x)$ 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.

A Clarkson dose computation model was used. The dose value at each sampling point was obtained from the interpolation of tissue phantom ratio (TPR) values and off-center ratio (OCR) values. Such values are presented in tables and are derived from percentage depth dose and beam profile measurements in water phantom.

Quasi-randomly distributed sampling points in the target and in the OARs were used.

2.5 Global convergence analysis

Global convergence for an optimization algorithm describes the property of the algorithm which states that the result is independent on the initial value. The individual single objective functions were determined by running *L-BFGS* with different starting fluence profiles. The results of the optimization runs were stored and the statistical properties such as the average, minimum and maximum and one standard deviation (SD) values were calculated for each objective. This calculation was repeated depending on the number of bixel (also called beamlet) weights and number of objectives for different sets of importance factors. A representative set of the Pareto front was sampled in order to determine whether local minima Deasy (1997) or degenerate cases exist for some combinations of importance factors.

We have a degenerate state if the total objective function f_{Tot} for a fixed set of importance values can be obtained by different values of the individual objective functions, *i.e.* if two solutions *a*, *b* exist for which

$$f_{Tot} = \sum_{i=1}^{M} w^i f_a{}^i = \sum_{i=1}^{M} w^i f_b{}^i \quad \text{and for some } \mathrm{i} \in [1, M] \;\Rightarrow f_a{}^i \neq f_b{}^i$$

where $f_a{}^i$, $f_b{}^i$ are the optimal values for the i^{th} objective found in solution *a* and *b* respectively.

2.6 Optimization with simulated annealing

We compare the optimization results of *L-BFGS* with results obtained by *SA* as global convergence does not guarantee that the solution is globally optimal, i.e. a global minimum is obtained. *SA* in principle can escape from local minima and it is statistically proven that it asymptotical converges to the global minimum if a defined annealing schema and a visiting probability distribution is used. We use the fast *SA* algorithm (*FSA*) Szu and Hartley (1987) with a Cauchy visiting probability distribution.

The optimization tests were undertaken with a 933 MHz Intel Pentium III PC with 512 MB RAM memory.

2.7 Example cases

Two clinical cases are used: a patient with a brain tumor and a patient with prostate cancer. The third example is a phantom case which has been chosen to test the algorithm for a difficult case for IMRT. We term this the phantom patient test case. For all examples, a total of nine beams were used at angles 40°n, where n=0,…,8. We use the notation recommended by IEC 1993 for the specification of angles.

2.7.1 Phantom patient with C-shaped PTV

The phantom geometry consists of a C-shaped PTV with a spherical OAR adjacent to the smaller concave periphery of the PTV but not actually within the PTV, see Fig. 1. In IMRT such PTV and OAR geometry is difficult to plan. For this case only a two-dimensional case is considered with 567 bixels for the 9 beams. The number of sampling points used in the optimization is shown in Table I.

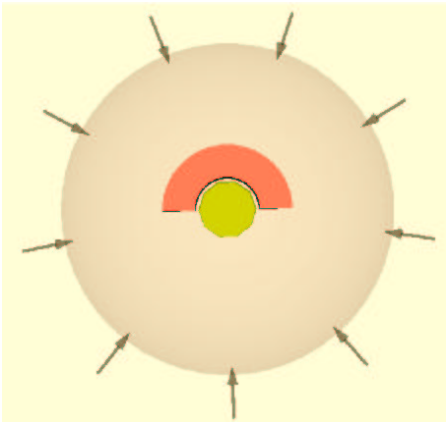

Figure 1 A phantom patient with a C-shaped PTV and spherical OAR with the nine beam directions shown.

Object	Number of Sampling Points
PTV	2426
NT	10029
OAR	1505

Table I. Statistics for the 13965 sampling points used for the C-shaped case.

2.7.2 Brain tumor patient

The brain tumor case consists of four structures, namely the PTV, the NT and both eyes, see Fig. 2. For this case 856 bixels for the 9 beams are considered. The number of sampling points used in the optimization is shown in Table II.

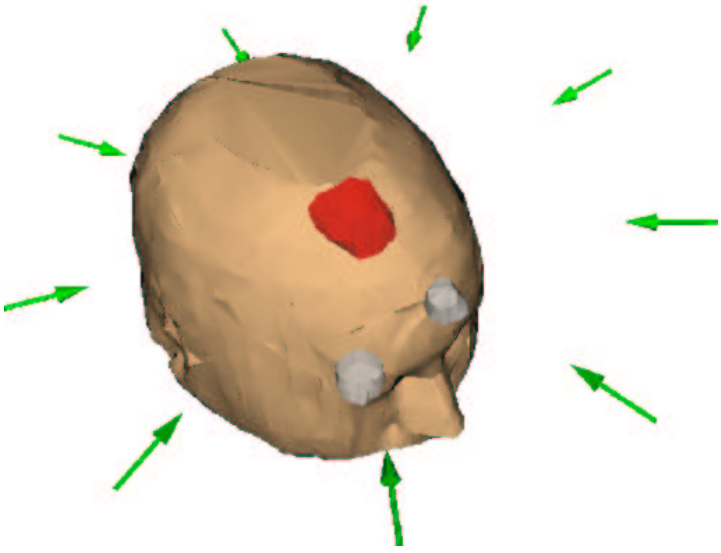

Figure 2 The brain tumor patient with the nine beam directions shown.

Object	Number of Sampling Points
PTV	3162
NT	20052
Left eye	3091
Right eye	3084

Table II. Statistics for the 29389 sampling points used for the brain tumor patient.

2.7.3 Prostate cancer patient

The prostate tumor case consists of six structures, the PTV, NT, bladder, rectum and the two femoral heads, see Fig. 3. For the optimization 5464 bixels are used. The number of sampling points in each structure is shown in Table III.

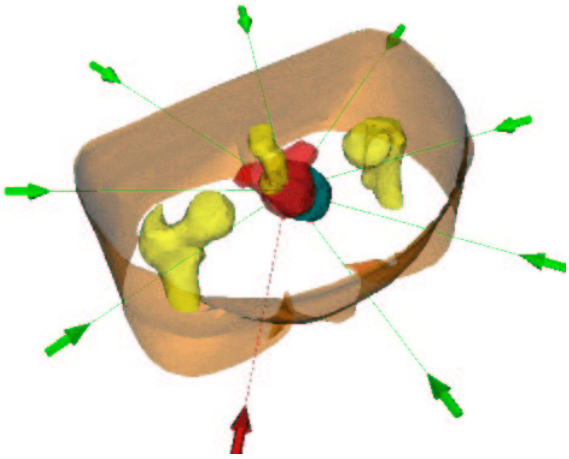

Figure 3 The prostate cancer patient with the nine beam directions shown.

Object	Number of Sampling Points
PTV	4852
NT	20515
Bladder	3275
Left femoral head	3360
Right femoral head	3350
Rectum	3638

Table III Statistics for the 38990 sampling points used for the prostate cancer patient.

3. Results

3.1 Convergence

Convergence of the optimization result with *L-BFGS* is observed after 300-500 iterations. The optimization results are virtually identical to the high precision optimization results obtained after 5000 iterations.

3.2 Global Convergence Analysis.

We investigate the existence of possible local minima and degenerate cases and the global convergence properties of *L-BFGS* for variance based objectives for the three cases. The critical dose for the OARs is set to 0. That is, we investigate the strictest requirements for the OARs and the NT which cannot be satisfied completely except for the trivial case for zero beam fluences, *i.e*. no beams at all. For each case we consider the convergence of a representative set of solutions specified by a set of importance factors. *L-BFGS* for each such set of importance factors is applied 100 times using always different random initial beam fluences.

3.2.1 Phantom patient with C-shaped PTV

The result of the configuration space analysis using *L-BFGS* for the C-shaped case is shown in Fig. 4 for a sampling parameter $k = 8$ which corresponds to 45 sets of importance factors. For each such set a distinct point on the Pareto front is expected, *i.e.* a Pareto global optimal solution.

The 45 solutions were obtained 100 times with different initial staring bixel intensities. The values of the individual objective functions, not including the importance factors, have been saved. The maximum number of iterations for each optimization run was 1000. Each of the 100 repetitions produces a distribution of 45 values for f_{PTV}, f_{NT}, f_{OAR} and f_{Tot}.

The results of these 100 distributions which represent 4500 optimization runs are shown in Fig. 4.

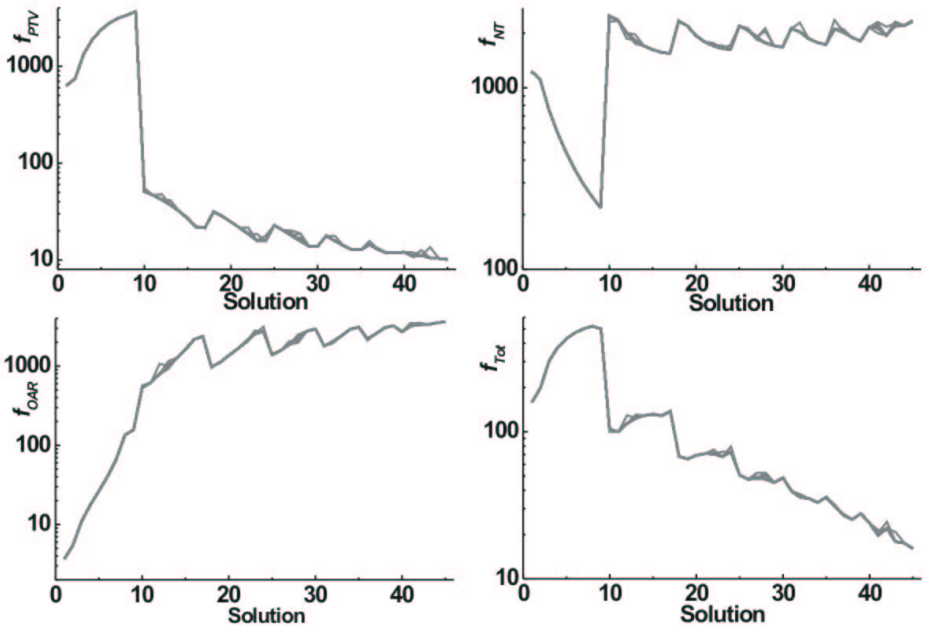

Figure 4 Distribution of the individual and the total objective function values for 45 solutions with different importance factors for the phantom patient with a C-shaped PTV obtained with *L-BFGS*. The optimization results for each set of importance factors repeated 100 times with different initial beam profiles is shown.

The results are so reproducible, *i.e.* independent on the initial beam fluences such that the 100 distributions included in Fig. 4 appear as a single distribution. Only in some optimization runs small deviations are observed which are due to a 1-2 percent convergence failure, see section 3.2.4.

3.2.2 Brain tumor patient

The result of the configuration space analysis using *L-BFGS* for the brain tumor case is shown in Fig. 5 for $k = 5$ which corresponds to 56 sets of importance factors. The optimization for each such set is repeated 100 times each time with different initial random beam fluences. The results for the 5600 optimization runs are presented in Fig. 5. For this 3D case the maximum number of iterations for each optimization run was set to 3000.

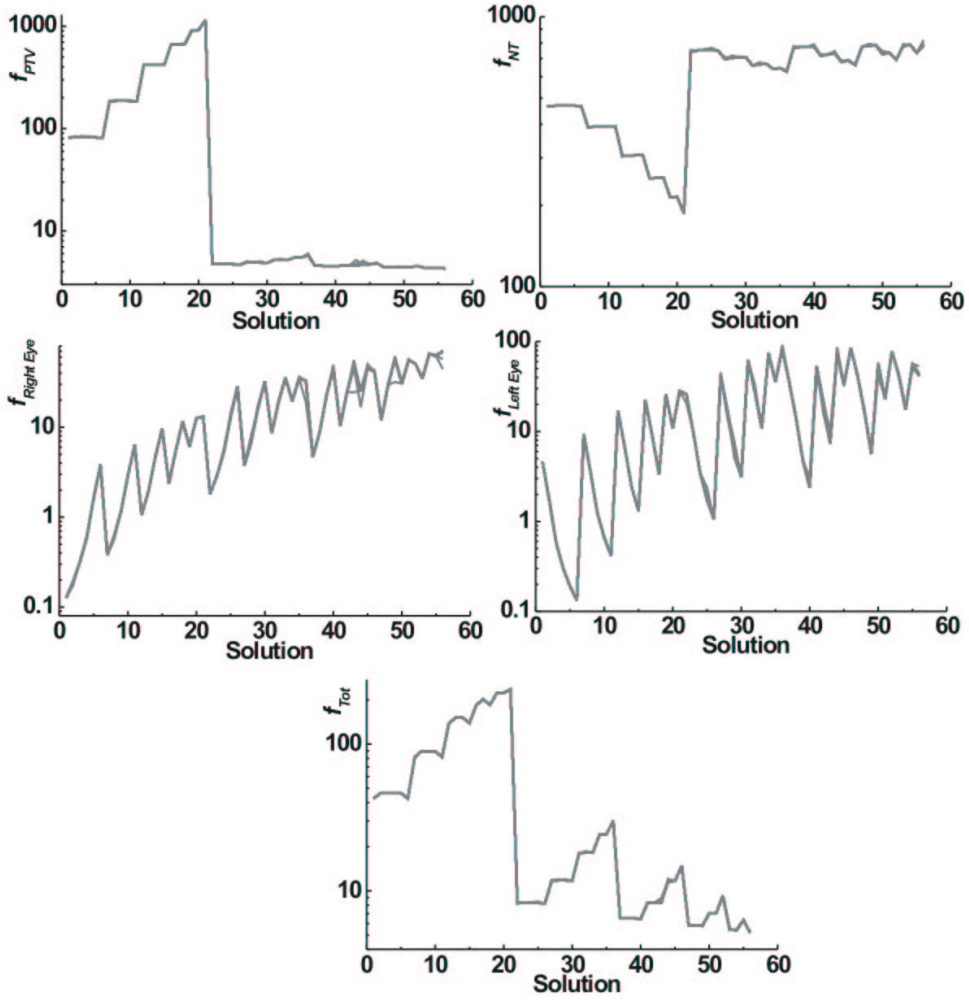

Figure 5 Distribution of the individual and the total objective function values for 56 solutions with different importance factors for the brain tumor patient obtained with *L-BFGS*. The optimization results for each set of importance factors repeated 100 times with different initial beam profiles is shown.

As for the C-shape case the 100 distributions appear as a single distribution and only in 1% of the cases we see some small differences.

3.2.3 Prostate cancer patient

The result of the configuration space analysis using *L-BFGS* with a maximum of 5000 iterations for the prostate cancer is shown in Fig. 6 for k = 3 which corresponds to 46 sets of importance factors. The values of f_{Tot} and the individual objective values for the 4600 solutions are shown. The results are independent on the initial value of the beamlet fluences.

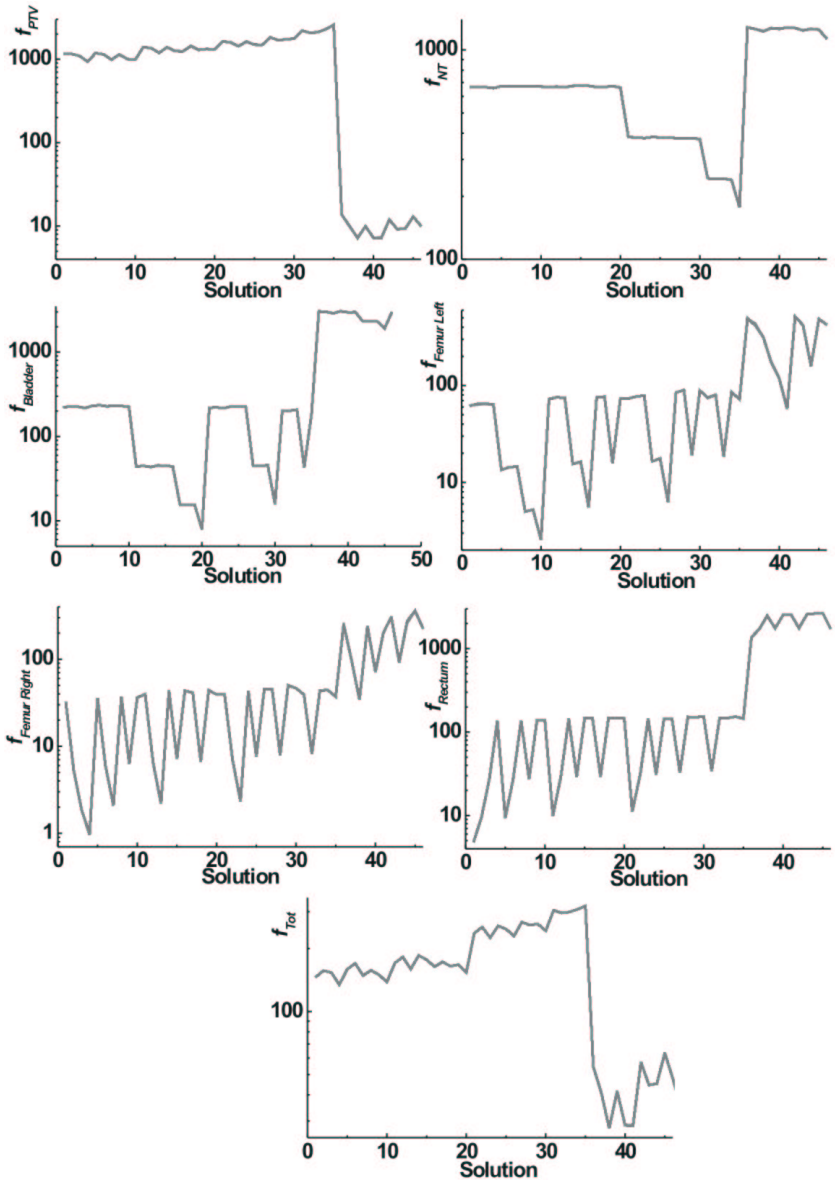

Figure 6 Distribution of the individual and the total objective function values for 46 solutions with different importance factors for the prostate cancer patient obtained with *L-BFGS*. The optimization results for each set of importance factors repeated 100 times with different initial beam profiles is shown.

3.2.4 Failure of global convergence and correction

For the prostate case no differences between the results using different initial beam weights are observed. For the C-shaped case global convergence is not established in 1-2% of the optimization runs. For these solutions the objective function value differs by up to 20% from the objective value of the remaining solutions. For the brain tumor case the number of failures is less than 1%. Similar results are found by applying different critical dose values for the OARs.

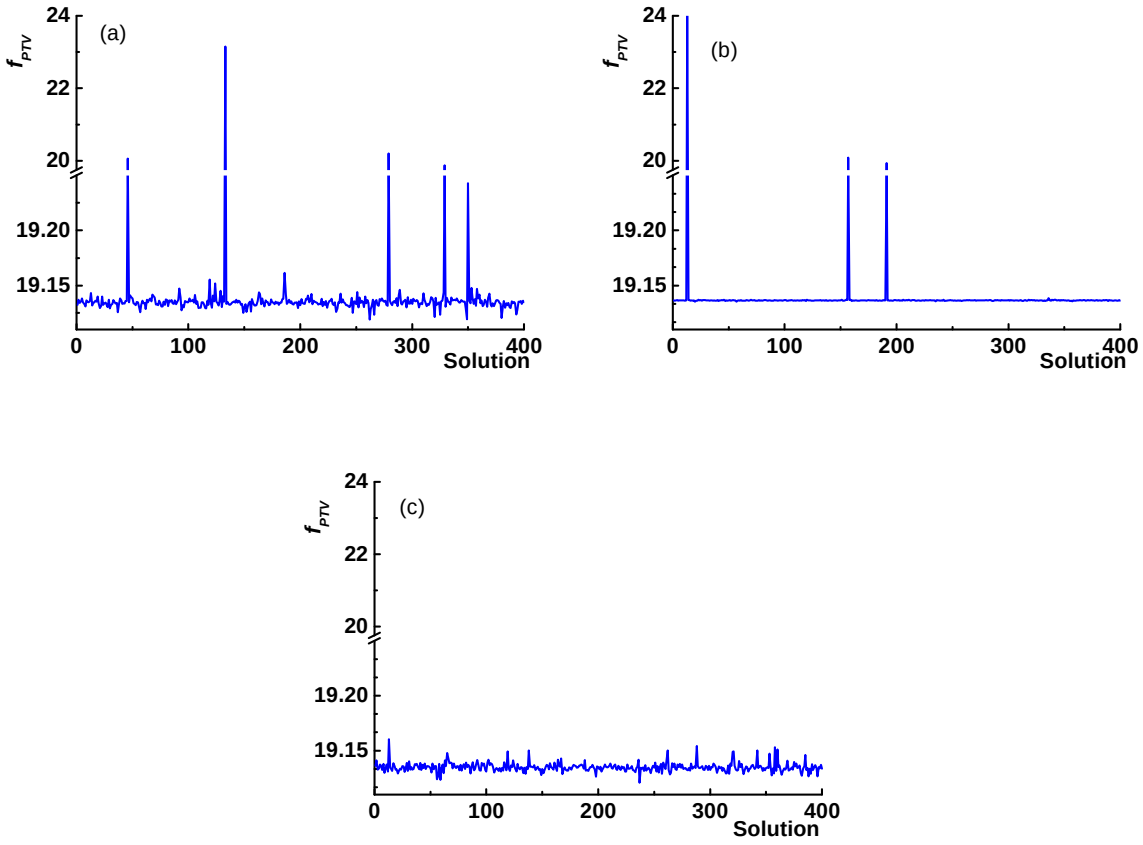

Figure 7 Distribution of the PTV dose variance f_{PTV} for 400 solutions obtained for the C-shaped PTV with *L-BFGS*. (a) Optimization using 500 iterations. For some solutions a violation of global convergence is observed. (b) Optimization using 2000 iterations; the accuracy improves but the algorithm is still trapped producing a few sub-optimal solutions. (c) Optimization using 500 iterations, including a correction mechanism as described in the text. Sub-optimal solutions are not more observed, global convergence is established.

The failure is shown for the C-shaped case using D_{crit} = 50% of D_{ref} for the spherical OAR. The optimization was repeated 400 times with different initial beam weights. The number of iterations was set to 500. The PTV dose variance of the obtained solutions is shown in Fig. 7.

A special scale is used to enhance the size of the fluctuations. For few solutions a violation of global convergence is observed, see Fig. 7(a). Repeating the optimization with another set using 2000 iterations improves the accuracy but still new sub-optimal solutions are produced, see Fig. 7(b).

The resulting DVHs for the PTV, NT and the OAR of all 400 solutions in Fig. 7(a) are shown in Fig. 8.

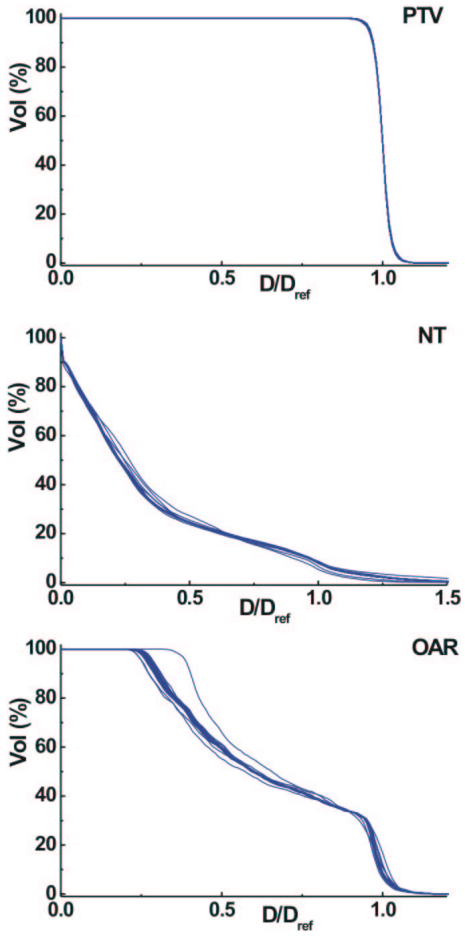

Figure 8 DVHs of the PTV, the NT and the OAR for the C-shaped case. The result of 400 solutions obtained with *L-BFGS* after 500 iterations is shown. The accuracy and the failure of global convergence is shown.

In brachytherapy (Lahanas et al 2003) our analysis showed that the line search method can be a reason for the failure. Whereas global convergence has been found for *L-BFGS* this

was not the case for *BFGS* which in a few percent of the cases failed to converge to the global optimum value.

We include a small disturbance in order to help *L-BFGS* to escape from local minima or optimization paths with a very small convergence. After every 100 iteration steps we add a very small constant term, beginning with a value of 10^{-5}, to all weights. This term decreases each time by a factor of 10. The correction is applied only three times.

The optimization was repeated 400 times using the correction method with 500 iterations. The PTV dose variance is shown in Fig. 7(c). No failure of global convergence is observed anymore.

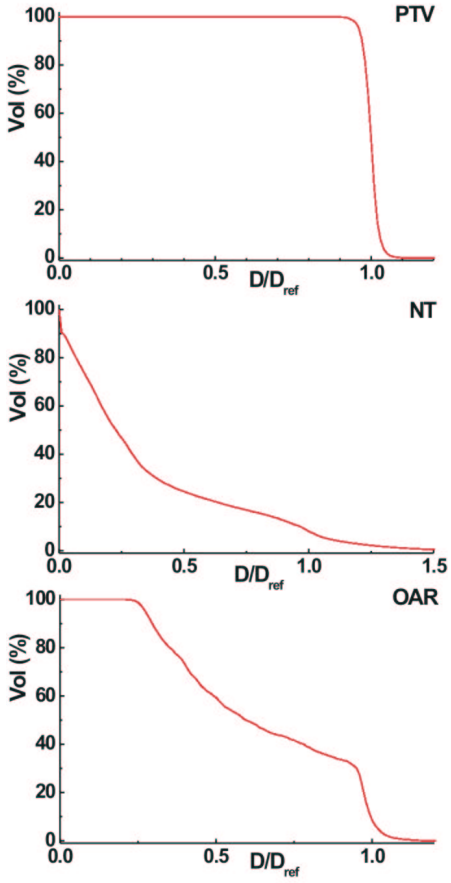

Figure 9 Dose-volume histograms for the C-shaped case for the PTV, the NT and the OAR. The result of 400 solutions obtained with *L-BFGS* after 2000 iterations is shown. A correction mechanism is applied during the optimization. The obtained DVHs are identical.

The DVHs of all 400 solutions obtained for the C-shaped case with the correction method is shown in Fig. 9. All 400 DVHs included in the plot are practical identical and appear as a single DVH.

It was reported in tomotherapy by Shepard *et al* (2000) that numerous combinations of beam weights can produce similar objective function values. We show in Fig. 10 for the C-shape case the distribution of the weights for all 400 solutions, i.e. all 400 intensity distributions are included in Fig. 10. The distributions are identical and appear as a single distribution; the resulting fluence profile is independent on the initial values used.

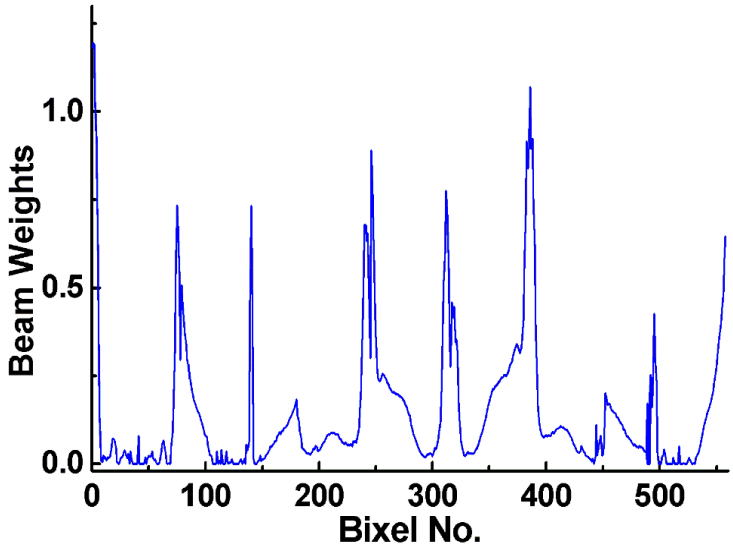

Figure 10 Distribution of the beam weights for the C-shaped case. The result of 400 solutions obtained by *L-BFGS* after 2000 iterations using different initial values for the beam weights is shown. A correction mechanism is applied during the optimization.

3.3 Comparison with fast simulated annealing

We compare the optimization results of *L-BFGS* with fast simulated annealing (FSA). We obtain a representative set of solutions obtained with uniformly distributed importance factor vectors and we compare the results using FSA. In this way we determine if the solutions obtained by *L-BFGS* are global optimal, i.e. if the representative Pareto front is a global Pareto front.

3.3.1 Phantom patient with C-shaped PTV

In Fig. 11 we compare the optimization results obtained with *L-BF*GS for the C-shaped PTV with results obtained with *FSA*. *L-BFGS* was running with a maximum of 5000 iterations and $\varepsilon=10^{-6}$. For *FSA* 1,000,000 iterations were used. We use D_{crit} = 0 for the OARs.

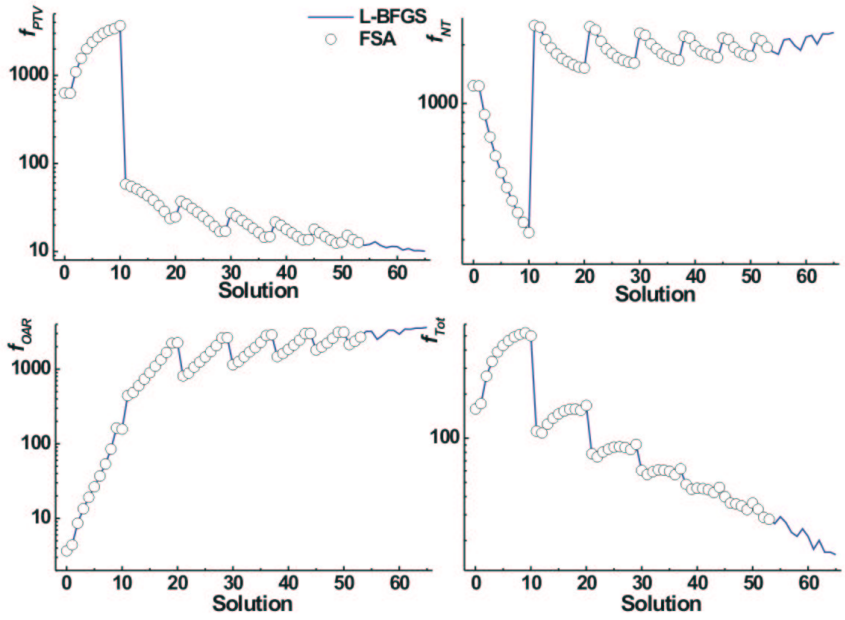

Figure 11 Comparison of the individual and total objective function values obtained for the C-shaped PTV when using *L-BFGS* and *FSA.*

The results show that *L-BFGS* provides Pareto global optimal solutions.

3.3.2 Brain tumor patient

The cost function f_{Tot} and the individual objective functions for the brain tumor patient obtained with *L-BFGS* and *FSA* are compared in Fig. 12. A non-dominated set with 56 solutions was produced by *L-BFGS*. *FSA* required more than 6 hours for each single optimization and used more than 2,000,000 iterations. Therefore because of this time factor we used a smaller set of solutions for *FSA*. We use D_{crit} = 0 for the OARs.

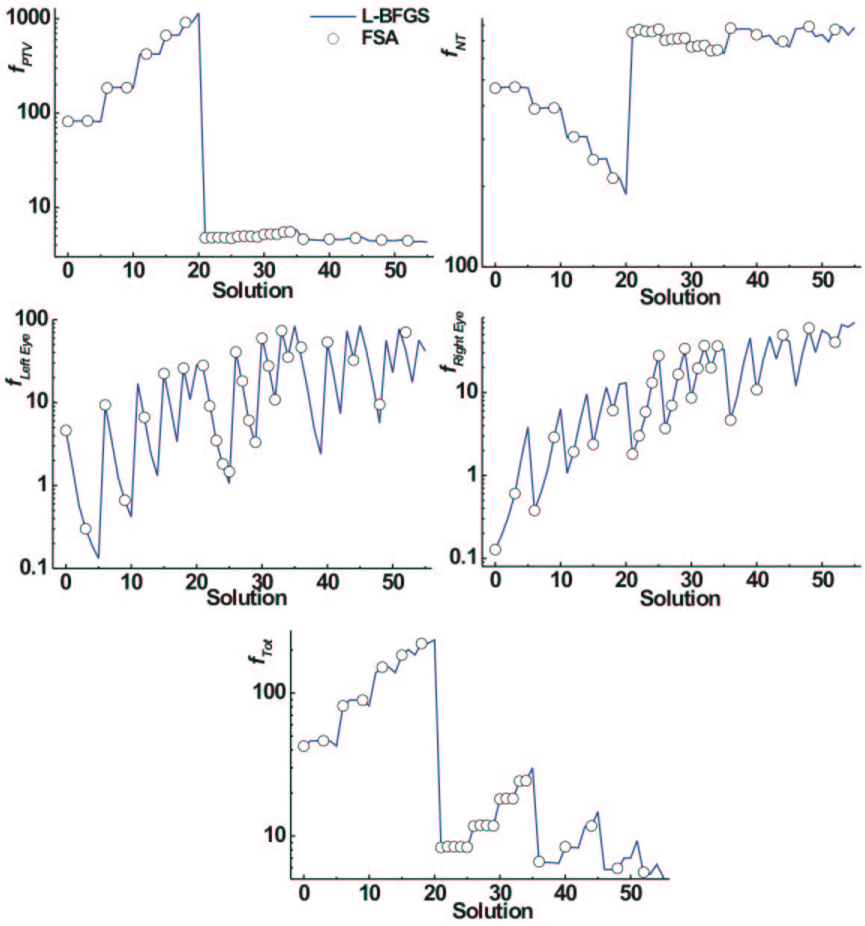

Figure 12 Comparison of the individual and total objective function values obtained for the brain tumor patient when using *L-BFGS* and *FSA.*

3.3.3 Prostate patient

The cost function f_{Tot} and the individual objective functions values for the prostate cancer obtained with *L-BFGS* and *FSA* are compared in Fig. 13. *FSA* required more than 10 hours for

each single optimization and used more than 2,000,000 iterations. Therefore because of this time factor we used a smaller set of solutions for *FSA*. We use D_{crit} = 0 for the OARs.

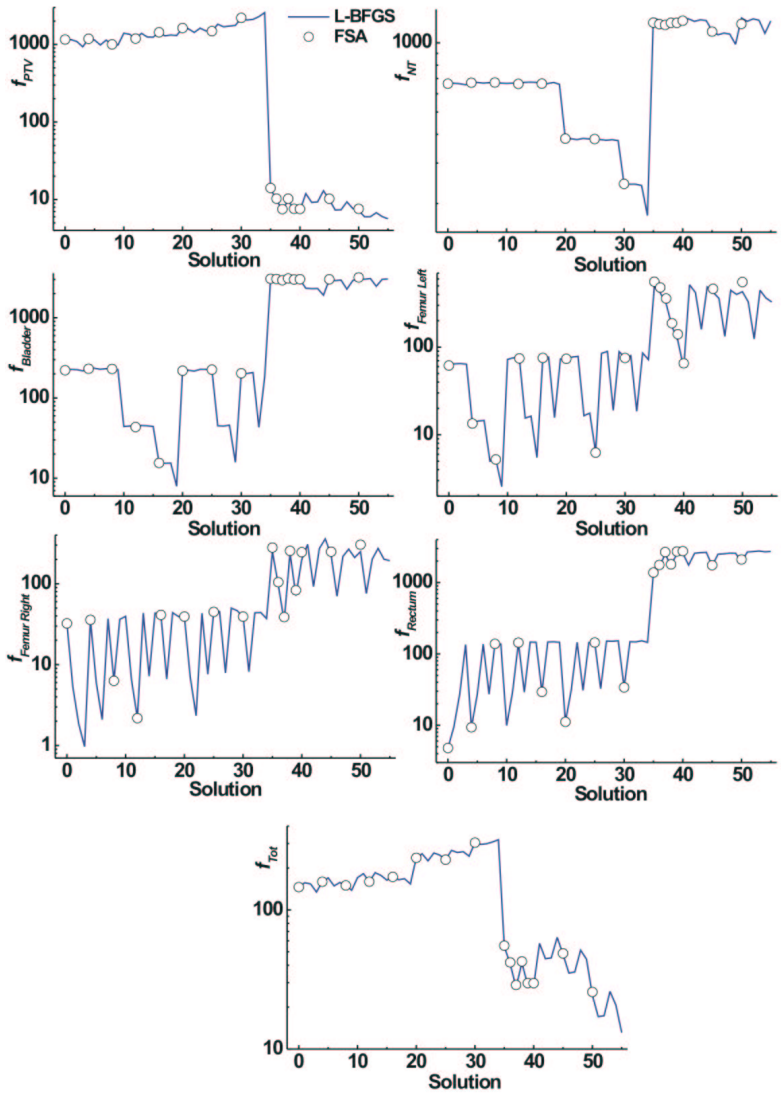

Figure 13 Comparison of the individual and total objective function values obtained for the prostate cancer patient when using *L-BFGS* and *FSA.*

Similar to the other cases also for the prostate tumor case *L-BFGS* provides Pareto global optimal solutions.

3.4 Multiobjective optimization

For the analysis of the spectrum of possible solutions obtained by MO we use as stopping criterion for *L-BFGS* a maximum of 500 iterations or a tolerance value $\varepsilon = 10^{-5}$.

Sufficient coverage of the PTV by the prescribed dose requires a very small dose variance. Tests show that a PTV importance factor $w_{PTV} > 0.9$ is required to obtain acceptable solutions. The corresponding dose variances for the OARs and NT are much larger. Rowbottom *et al* (2002) rescale for this reason the dose variances by the objective value found at the first optimization iteration. This requirement assumes that these values are not very different from the values at the end of the optimization. We have found that by multiplying the PTV objective by a factor of 100 we can use uniformly distributed importance factors in order to obtain a representative set of non-dominated solutions.

A similar result is obtained by multiplying the PTV importance factor by a parameter s = 100 and then normalizing the importance factors to obtain a sum equal to one. As an example we show the distribution of the importance factors or the C-shaped phantom case for 136 solutions with s = 1 and s = 100, see Fig. 14. Shown are the (w_{PTV}, w_{NT}) and (w_{NT}, w_{OAR}) values. A fraction of importance vectors still have a small w_{PTV} value but for the majority of solutions we have $w_{PTV} > 0.8$. We use all importance factors vectors produced in this way in this study even those which produce not clinical acceptable solutions. These are filtered in the decision making process and are used here only to indicate the position of these solutions on the Pareto front.

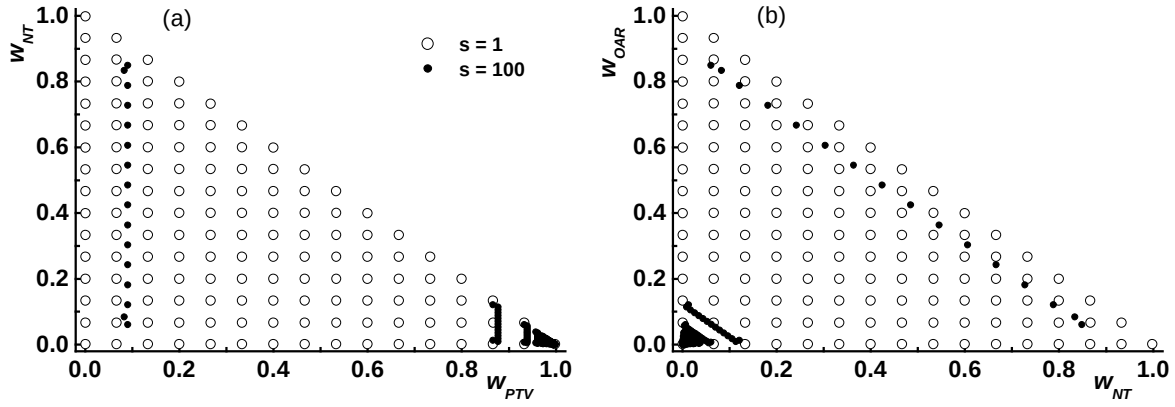

Figure 14 Distribution of importance factors for the C-shaped phantom case for 136 solutions with s = 1, and s = 100. (a) Distribution of (w_{PTV}, w_{NT}), (b) distribution of (w_{NT}, w_{OAR}). Note that most of the solutions have $w_{PTV} > 0.8$ when using a scaling parameter s = 100. A fraction of the solutions still have too small values which results in clinical not acceptable solutions.

We show as an example a high statistics case with $k = 60$ which corresponds to 1891 solutions using s = 1 for the C-shaped case, see Fig. 15. We show the PTV coverage fraction at the 95% of the prescription dose as a function of the importance factor w_{PTV} and as a function of the PTV dose variance f_{PTV}. Using uniformly distributed importance factors would produce only a few acceptable solutions. With a factor s = 100 we get a much larger fraction of acceptable solutions even with 136 solutions. Similar results are obtained for the other cases.

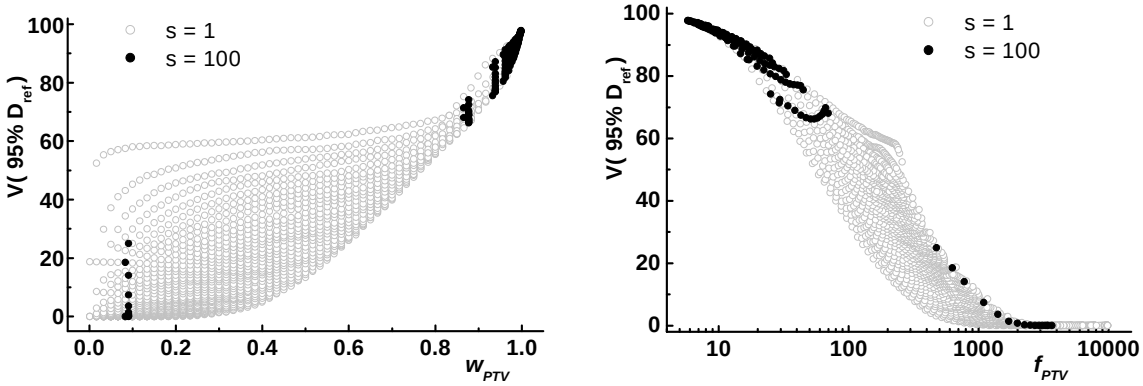

Figure 15 Dependence of the percentage of the PTV that is covered with the 95% of the prescription dose (a) as a function of the importance factor w_{PTV}, (b) as a function of the dose variance f_{PTV} for the C-shaped phantom case. The results are shown for 1891 solutions using s = 1 and 136 solution using s = 100.

In Fig. 16 we show the three two-dimensional projections of the Pareto front for the C-shaped case using 136 solutions obtained with s = 100 and 1891 solutions with s = 1.

We see the trade-off between f_{PTV} and f_{NT} and f_{OAR} respectively. This is a case where it is impossible to reduce the dose variance in the OARs and the NT below a specific value. The possible good candidates are concentrated at low f_{PTV} values to ensure a sufficient dose coverage for the PTV. This is true for most of the solutions obtained using s = 100.

A small fraction which has very small w_{PTV} values does not represent acceptable solutions, This fraction increases with the number of obejctives.

Most of the solutions with s = 1 are covering the entire extremely large Pareto front showing that this sampling method is very inefficient.

The f_{PTV} - f_{NT} projection, see Fig. 16(c), shows that f_{OAR} and f_{NT} have to be both relatively large. Note that as the beam directions are fixed the level of f_{NT} and the Pareto front depends on the beam orientation. This information is important in inverse planning where the trade-off between the objectives is analyzed for different beam orientations and number of beams. For the C-shaped case we have this Pareto front due to the concave shape of the PTV and the OAR position.

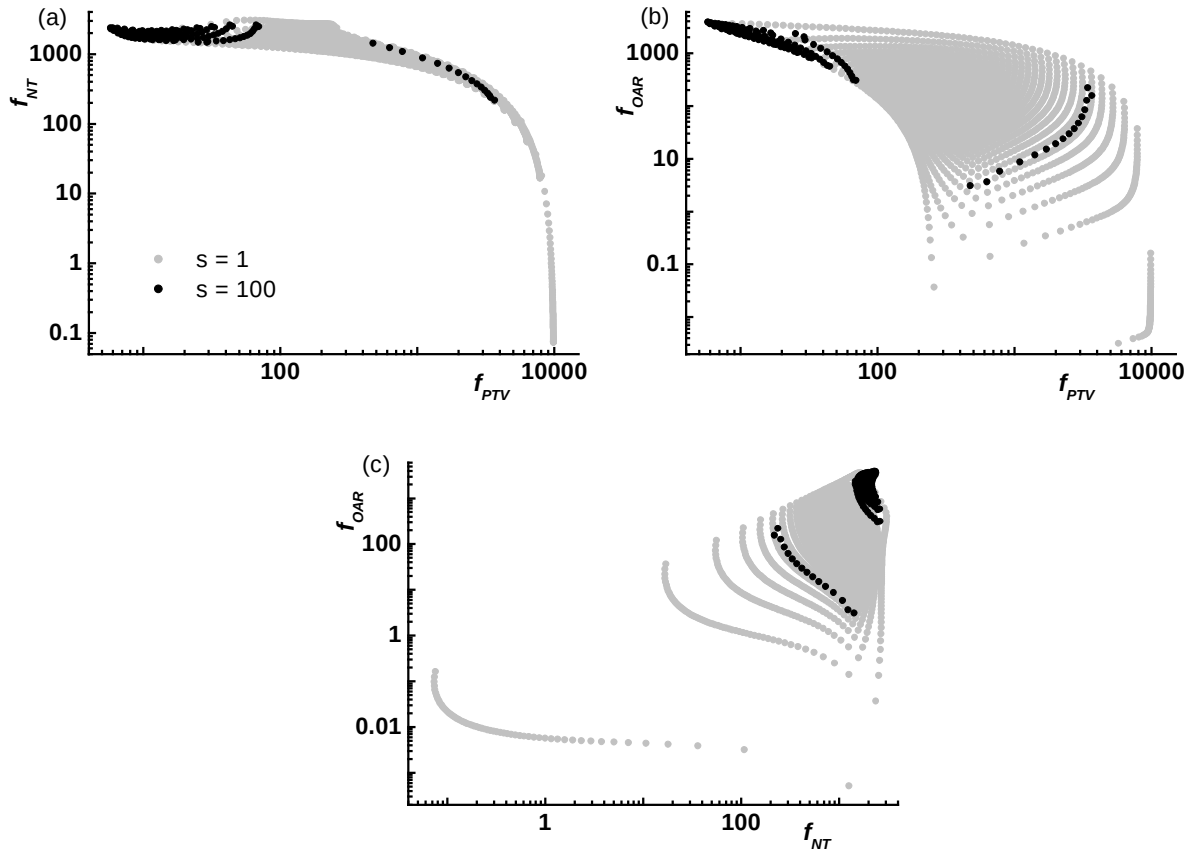

Figure 16 The three two-dimensional projections (a) (f_{PTV} , f_{NT}), (b) (f_{PTV} , f_{OAR}) and (c) (f_{NT}, f_{OAR}) of the three-dimensional Pareto front (f_{PTV} , f_{NT}, f_{OAR}). Included are the 1891 solutions obtained with k = 60 and s = 1 and the 136 solutions obtained by k = 15 and s = 100.

In Figure 17 we have a closer look at Figure 16(a) at small f_{PTV} values. Included are the 1891 solutions obtained with k = 60 and s = 1 and the 136 solutions obtained by k = 15 and s = 100. We obtain more clinical acceptable with the smaller set of solutions with s = 1 than the much larger set of solutions using s = 100. The number of more acceptable solutions for s = 100 is approximately 100. Thus more than 10000 solutions with s = 1 would provide the same number of non-dominated solutions in the region of interest. Note the rapid increase of f_{NT} as the dose variance in the PTV approaches its minimum value. At moderate f_{PTV} values it is possible to obtain solution with very different dose variances in the NT. Very rapidly as the dose variance approaches the minimum value f_{PTV} = 5.7 the variation of f_{NT} is reduced and we have to accept a variance value f_{NT} = 2370.

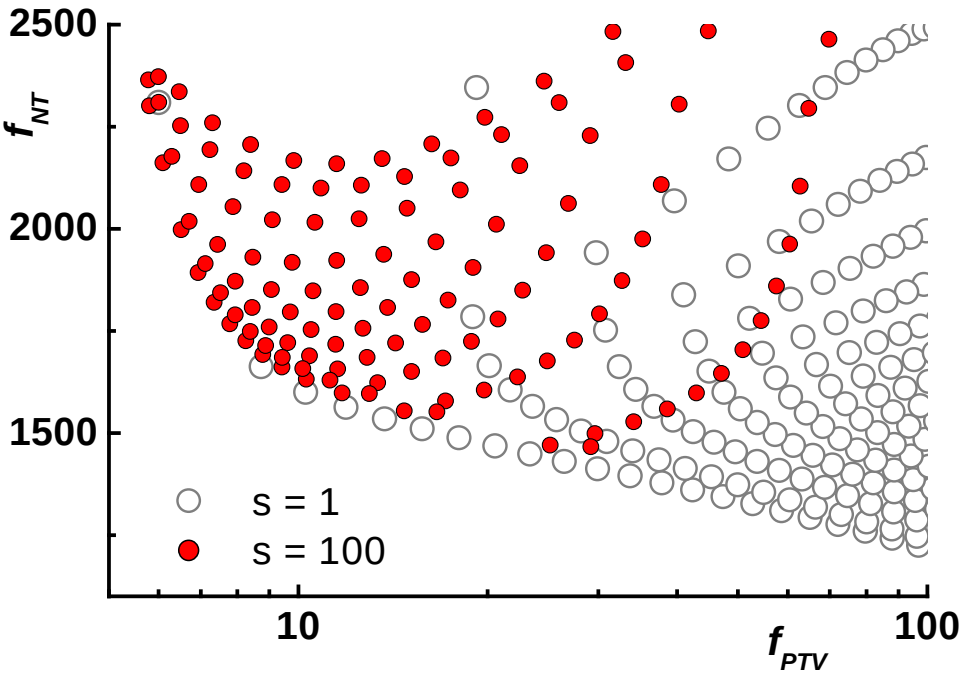

Figure 17 Projection of the Pareto front on the f_{NT} - f_{PTV} plane obtained with 1891 solutions using s = 1 and 136 solutions using s = 100. Note that the clinical acceptable solutions require a small variance and much more solutions are obtained with the smaller set of solutions using s = 1, this number is approximately as large as the parameter s = 100.

3.4.1 Results for the phantom patient with a C-shaped PTV

For the C-shaped PTV example, we performed an MO dose optimization using *L-BFGS* with 136 solutions. We compare the results with D_{crit} = 0 and D_{crit} = 50% of D_{ref} for the spherical OAR. The resulting DVHs for the PTV and OAR are shown in Fig. 18. The DVHs of solutions (filtered solutions) selected to have at least a 95% PTV coverage at 95% of D_{ref} are highlighted, where 97.8% was the largest value found.

For the optimization using D_{crit} = 0 a larger variety of solutions is observed for the OARs. Solutions can be found with a dose smaller in the OAR than any other solution obtained by MO where D_{crit} =0.5* D_{ref} was used.

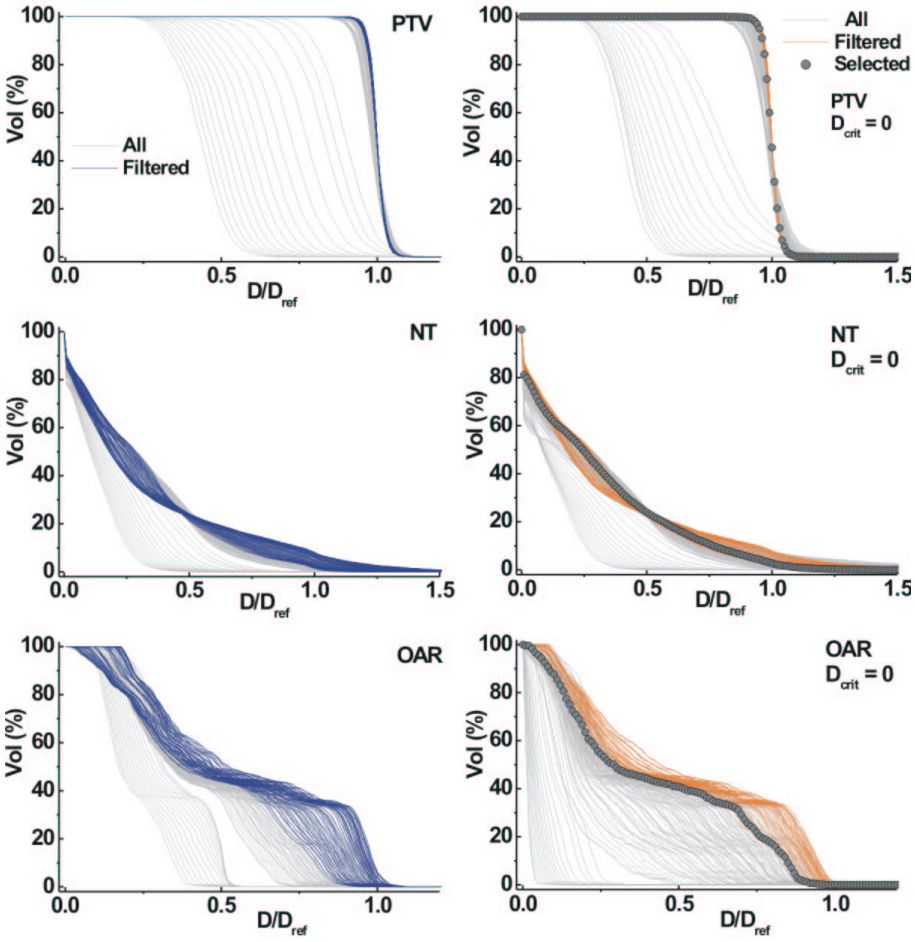

Figure 18 DVHs of 136 solutions (k = 15) generated by *L-BFGS* for the PTV, NT and the OAR for the C-shaped PTV example. Solutions filtered with at least 95% PTV coverage at 95% of D_{ref} are marked. The DVHs on the left side are obtained with $D_{crit} = 0.5\,D_{ref}$ and on the right side $D_{crit} = 0$ was used. The selected optimal solution from the set is shown.

The analysis of the two dimensional Pareto front projections for this example shows that there is a rapid increase of the dose variance in the OARs and the NT with increasing PTV coverage and with dose uniformity. The variance of the dose in the NT for acceptable solutions is larger than 1400. For this example the spectrum of solutions is restricted and if the OAR has to be considered then an increase of the dose variance in the PTV cannot be avoided. A solution has been selected which reduces the dose variance in the NT and in the spherical OAR simultaneously for the filtered solutions. The DVHs of this solution is shown in Fig. 18

3.4.2 Results for the brain tumor patient

A MO optimization was performed for the brain tumor example. The optimization is repeated with $D_{crit} = 0$ and $D_{crit} = 63.4\%$ of the prescription dose for the eyes. The resulting DVHs for 165 solutions ($k = 8$) are shown in Fig. 18. The largest PTV coverage at 95% of D_{ref} is 99.84%. The solutions with 99% PTV coverage are marked.

The observed spectrum of DVHs is large, showing the different results which can be obtained by modifying the importance factors. The DVH for the PTV for all cases is almost identical. The best results for the OARs, see Fig. 19, are obtained by using a optimization with zero critical dose for all OARs where solutions can be obtained that are not available if the optimization uses clinical acceptable dose limits for the eyes.

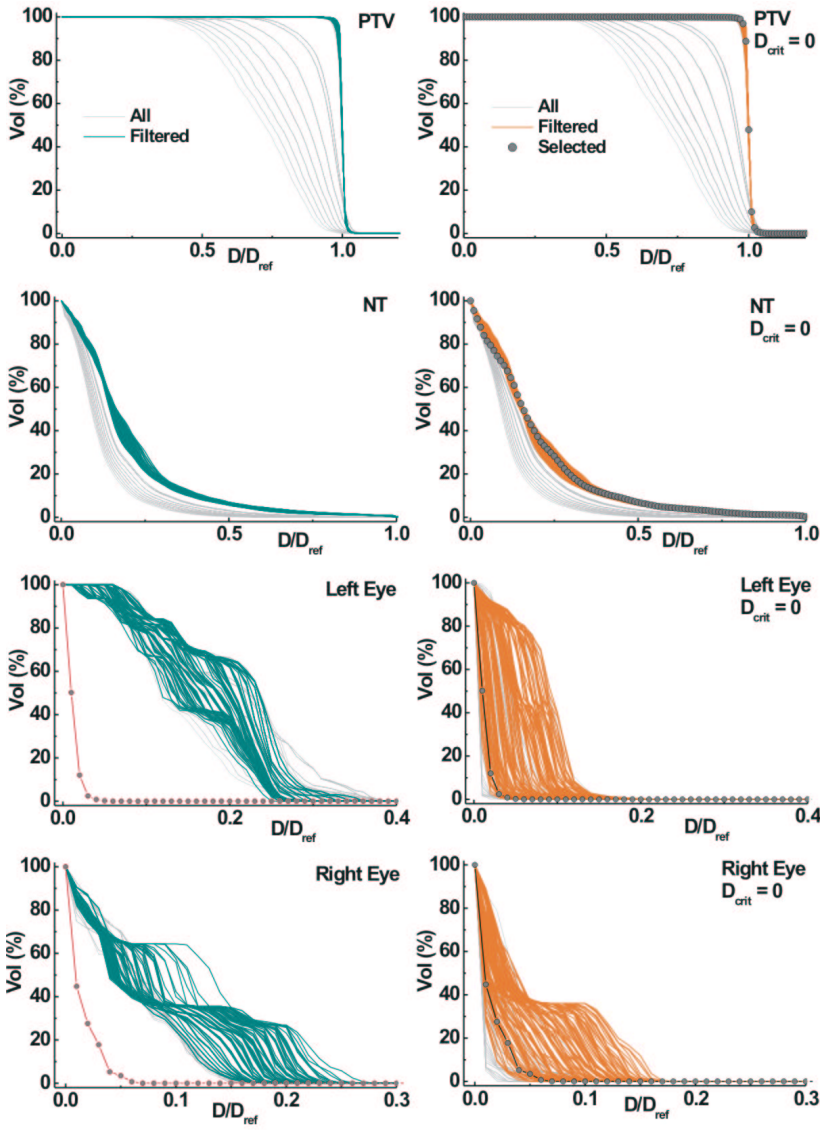

Figure 19 DVHs of 165 solutions (k = 8) generated by *L-BFGS* for the PTV, the NT and the OARs for the brain tumor case. Solutions filtered with at least 99% PTV coverage at 95% of D_{ref} are marked. The DVHs on the right side are obtained with D_{crit} = 0, whereas on the left side D_{crit} = 63.4% of D_{ref} was used for both eyes. The resulting DVHs for the NT and the OARs show the possibilities when using different importance factors. The selected optimal solution is shown.

Projections of the two-dimensional Pareto fronts show that the dose variance in the NT and much more in the left and right eye can be reduced significantly without any significant modification of the dose uniformity in the PTV. In contrast to the C-shaped PTV example the treatment planner has a larger spectrum of possible solutions.

A solution has been selected which for both eyes and the NT has simultaneously a value close to the corresponding smallest possible dose variance of all solutions. We obtained this by taking the solution with the smallest product of objective values for the eyes and the NT. The DVHs for this solution are shown in Fig. 19.

We performed two MO with optimizations using s = 1 and k = 16 with D_{crit} = 0 and D_{crit} = 63.4% of D_{ref} for the eyes. This corresponds to 969 solutions. A filter was applied for the PTV coverage to be 99% at 95% of D_{ref} with 116 and 135 solutions selected for the cases D_{crit} = 0 and D_{crit} = 63.4% D_{ref} respectively. The DVHs for the left eye are shown in Fig. 20. This higher statistics case confirms that we obtain better solutions using D_{crit} = 0. In comparison to Fig. 19 we see that using s = 100 even with only 165 solutions we obtain a broader spectrum of solutions, i.e. the Pareto front of clinical acceptable solutions is more efficiently sampled.

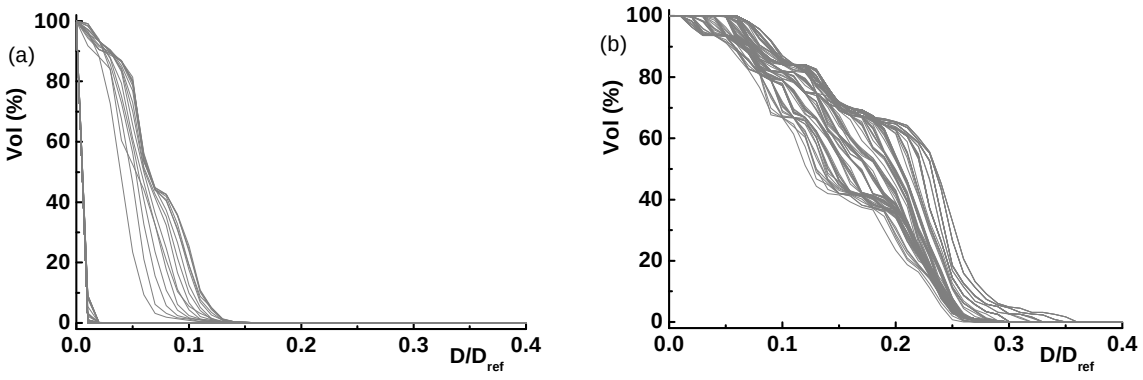

Figure 20 DVHs of the left eye obtained by 969 solution with s = 1 (k = 16) by *L-BFGS*. A filter was applied for the PTV coverage to be 99% at 95% of D_{ref} the result are (a) 116 solutions obtained with D_{crit} = 0 and (b) 135 solutions obtained with D_{crit} = 63.4% of D_{ref}.

3.4.3 Results for the prostate cancer patient

Two MO optimizations were performed for the prostate cancer example. For the first MO optimization we used D_{crit} = 0 for all OARs and for the second MO optimization the critical values for the bladder, left and right femur and the rectum were set to 87.8%, 70.3%, 70.3%, and 81% of the prescription dose respectively.

The two-dimensional projections of the Pareto front reveal that the bladder and rectum dose variances show a strong trade-off with the PTV coverage, see Fig. 21. The results are obtained using $D_{crit} = 0$ for all OARs.

For the femoral heads the dose variance is smaller than in the rectum and bladder, see Fig. 21. We also see that the range of solutions is large, *i.e.* if we do not use an optimal set of importance factors the dose variance can be very large. Note that a double logarithmic scale is used to reveal the fine structures of the Pareto front. Although solutions can be obtained which reduce strongly the dose variance in these structures without a significant change for the dose in the PTV, this cannot be accomplished simultaneously for the rectum and bladder. We have to accept a large dose variance for the bladder and rectum and we see that it is possible to reduce the dose variance in the PTV without a significant modification of the dose in the bladder and rectum.

For the left and right femoral head it is possible to obtain solutions with a small dose variance even at low f_{PTV} values. The large set of 2002 solutions obtained with s = 1 show that some of these solutions are not accessible by the smaller set of 126 solutions. For this case in order to reduce the further the dose in the femoral heads would require to generate more solutions. This is because the sampling parameter is small necessary to keep the number of solutions small for the six objectives.

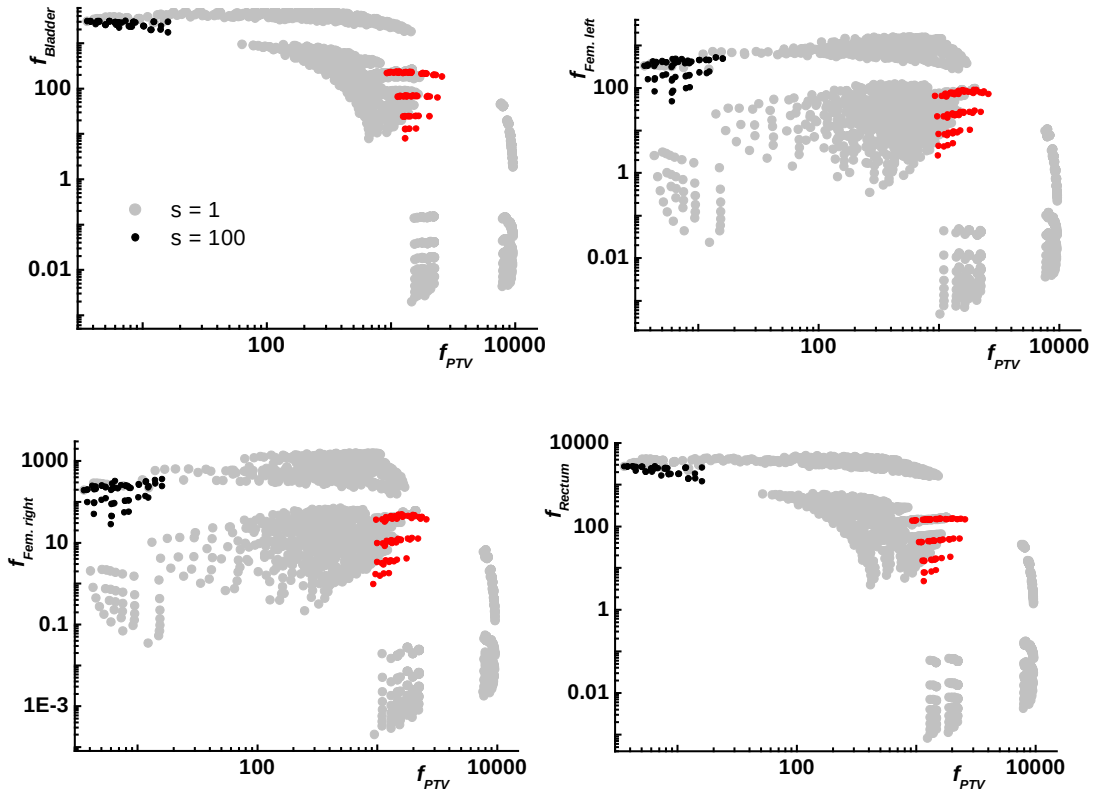

Figure 21 Projections of the Pareto front for the prostate case showing the correlation between the dose variance in the PTV and the dose variance in the OARs. The non-dominated solutions of a large set of 2002 solutions with s = 1 are shown. Included are the solutions obtained from a smaller set of 126 solutions using s = 100. The solutions which are clinical not acceptable are shown in red color.

The resulting DVHs for 126 solutions (k = 4) are shown in Fig. 22. The maximum PTV coverage at 95% of D_{ref} was found to be 98.8%. The DVHs of solutions selected have at least a PTV coverage of 97% are marked. Similarly to the first two examples studied, a larger spectrum of solutions is obtained for the case where the OAR critical dose value is set to 0.

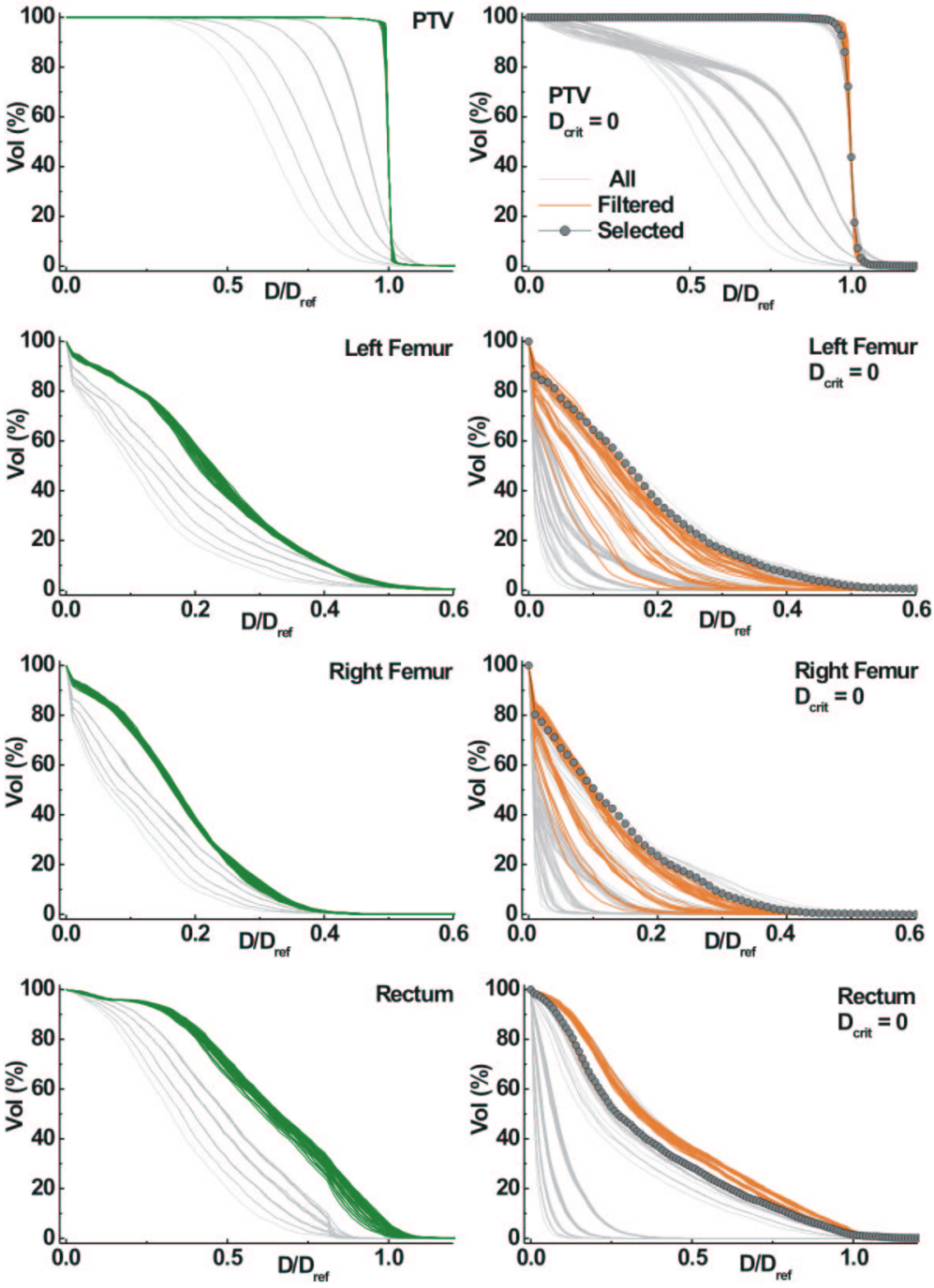

Figure 22 DVHs of 126 solutions (k = 4) generated by *L-BFGS* the PTV and the OARs for the prostate tumor example. Solutions with have at least 97% PTV coverage at 95% of D_{ref} are marked. We compare the results of a MO optimization with D_{crit} = 0 (DVHs on the right side) and clinical acceptable critical dose values for the OARs (DHVs on the right side). The selected optimal solution is shown.

A solution has been selected for which simultaneously, the dose variance of the rectum and bladder is close to the smallest possible value for a PTV coverage of at least 97% at 95% of D_{ref}. This solution is obtained by considering the minimum value of $f_{Bladder}*f_{Rectum}$ of the corresponding variances of the filtered solutions and the DVHs are shown in Fig. 22.

4. Discussion and conclusions

We studied the use of constraint-free gradient-based optimization algorithms for MO dose optimization in IMRT using as a representative the *L-BFGS* algorithm. The global convergence properties using variance based objectives have been analyzed. Results using *BFGS* and *FRPR,* Lahanas *et al* (2003) are not presented here but we found that they reproduce the *L-BFGS* results.

L-BFGS as with Newton based algorithms does not require the exact inverse Hessian matrix. Even if a smaller number of iterations were required for convergence, the use of an exact Hessian requires the initial values of the fluences to be not very different from their optimal values. *L-BFGS* is especially suitable for *N*-dimensional problems when *N* is very large.

The required memory is proportional only to the number of optimization parameters *N* and not proportional to N^2 as required by *BFGS which* requires much larger time for the optimization than *L-BFGS*.

Gradient-based optimization methods have been modified in the past by changing the line minimization routines, or by setting artificially negative weights to 0 in order to avoid negative beam fluences. It is, however, not clear how the theoretical established convergence properties of the algorithms are modified by this approach. Our mapping which considers as optimization parameters the square root of weights eliminates this problem.

The use of the square root of the bixel fluences as optimization parameters can be applied to all other non-linear optimization algorithms. This eliminates completely the problem of solutions with negative fluences. No artificial modifications of the line search routines is required which are important for gradient-based algorithms as some criteria have to be fulfilled by the line search for global convergence.

Llacer *et al* (2001) say that “it is not clear whether setting negative intermediate results to zero during the iterations of the gradient methods is important or not, since they can yield excellent results, but the statement can be made that, with approximately 5 to 10% of the beam weights becoming negative in the CG and NG methods in the later part of the iterative process,

the results cannot be precisely minimum least square solutions, as the theories for those methods require the existence of negative beam weights".

The *L-BFGS* algorithm used in this study does not require a quadratic function such as by NG methods, but has successfully been applied for various nonlinear optimization problems. The mapping of the decision variables avoids completely negative beam fluences.

A convergence analysis of *L-BFGS* shows that in some case the algorithm gets trapped in a local minimum. A very simple method can be used to practically eliminate this problem. Similar to simulated annealing the optimization path is disturbed slightly a few times and helps to escape from closed orbits or local minima.

While a similar result was observed by Llacer *et al* 2003, this analysis considers the global convergence of a true constraint free gradient-based optimization algorithm without any artificial line modifications. The exact origin of the failure in some cases is unknown. The local optimal solutions are such that the resulting DVHs do not differ much from the DVHs of the global optimal solutions. The proposed method which is applied only at three iterations removes the probability of failure so that we can say that global convergence is practically established.

As presented in Lahanas *et al* (2003) in a brachytherapy study, *L-BFGS* should not be used with a warm start option in which the results of a previous optimization are used to initialize the algorithm for a new optimization with only slightly modified importance factors. This could be used to increase the speed of the process of the generation of a representative efficient set. With this approach *L-BFGS* can prematurely converge as the starting point is required to be not on the border of the feasible objective space. For similar reasons found in brachytherapy by Lahanas *et al* (2003), statistical randomly selecting such a point is practical very unlikely.

A comparison of the optimization results with *FSA*, shows that the solutions are global optimal solutions. *FSA* required approximately 1000 times more time to approach the result of *L-BFGS*. This shows that for variance based objectives the use of *SA* algorithms is not necessary and for MO dose optimization *SA* is too slow.

The solutions of *L-BFGS* are global optimal for a representative set of importance factors. In the time where some other algorithms provide just one solution it is possible to obtain

a representative spectrum of solutions with valuable information for the treatment planner for the selection of the best solution.

L-BFGS produces a solution after 500 iterations which for practical purposes is almost identical to the global optimal solution. Eventually the algorithm could be stopped after 100 iterations if only the value of the total objective function is considered. Even so, there will be some improvement in the result for the NT and the OARs possible for the remaining 400 iterations. The reason is that the optimal solution requires a significant fraction of the importance factors to be distributed to w_{PTV} so that the contribution of the OARs and the NT to the score function is very small.

With current 3 GHz PCs the optimization time of the non-optimized code is 4 s for the prostate implant with 5464 bixels and 40,000 sampling points considering the NT and four OARs.

In the past results comparing different algorithms or sets of objectives have been presented using only single objective optimization algorithms with only one solution selected by trial and error. We think that a better approach is to obtain a representative set of non-dominated solutions. Based on the spectrum of solutions we are able to understand what dose distributions are available for a given set of objectives and a particular optimization method. The analysis which we performed showed that the solutions obtained are global optimal. It is important to know this for the comparison.

As an example we applied a MO optimization to obtain the spectrum of possible solutions for one test case and two clinical cases. The optimization was applied by using the recommended critical values for the OARs and the optimization was repeated with the critical values set to 0. For the last case the spectrum of solutions is larger and the dose in the OARs can be reduced more than any solution which can be obtained by the former case. A trivial explanation is that by using the step function parts of the dose space are not “visible” by the optimization algorithm. With $D_{crit} = 0$ critical dose the algorithm tries to reduce the dose variance including all dose values even those which could be considered as “acceptable”.

For the selection of the best solution we remove first clinical not acceptable solutions by applying constraints. The analysis of the tradeoff between the objectives and range of values using simple decision tools shows what we can obtain and for which objectives what compromise we have to make. For the remaining objectives that satisfy the clinical constraints we seek to obtain simultaneously a solution where the resulting objective value is close as possible to the minimum value.

An *a priori* automatic selection method as described by Xing *et al* (1999) based on the Euclidean metric between the DVHs of a solution and ideal DVHs does not give always a satisfactory result as the ideal DVHs can be different than any solution possible. DVHs of OARs that are better than the ideal DVHs or have a different form can be the reason that a solution is selected where the other objectives could be satisfied better by other solutions. As the treatment planner does not have the information, it cannot be decided if this is actually the case or how good is the solution.

A fuzzy logic approach has been presented by Li and Yin (2001) to consider the normal tissue versus PTV tradeoff and to obtain an optimal solution based on fuzzy logic. A fixed prescription dose is used for the OARs and Li and Yin try to obtain an optimal prescription for the NT. Our analysis shows that the best solutions can be obtained by setting the most natural choice $D_{crit} = 0$ for the NT and the OARs. Li and Yin express their concern that line search algorithms may not provide global optimal solutions due to the line step selection problem. Our analysis shows that the results of L-BFGS are global optimal.

The analysis of the spectrum of solutions shows that even if all solutions for the PTV are almost identical they are different especially for the OARs. A MO optimization is therefore important for the selection of a solution of high quality which reduces to the lowest possible level the dose in the OARs and the patient's body reducing thus the level of unnecessary radiation exposed to the patient, even beyond that what could be considered as clinical acceptable.

We used variance-based objectives and it remains to be seen what solutions can be obtained by other algorithms and other objectives which use DVH constraints, Cho *et al* (1998). The MO-approach considers the constraints in the decision process. We think that a

comparison of algorithms and/or different objective functions in radiotherapy should be based on comparing the spectrum of solutions of a representative set of non-dominated solutions and not on two solutions which we do not know which part of the Pareto front they represent. Such an analysis we have shown for two set of objective functions.

The analysis shows that for variance based objectives and using *L-BFGS* and the mapping which avoids cutoffs each solution is a simple point on the Pareto front, i.e. a unique solution is obtained. For DVH-based objectives we expect this not to be the case. For a fixed set of importance factors each solution will produce with different starting intensities a different distribution of intensities. Repeating the optimization with fixed importance factors but different staring points will produce also different points on the Pareto front the solutions covering some, probably small, part of the Pareto front. This is because DVH-based objectives allow more freedom to the dose distributions which can be obtained, whereas variance based objectives produce a unique solution.

We have for the first time observed multi-dimensional Pareto fronts for IMRT dose optimization with variance-based objectives and its increasing complexity as the number of OARs and the objectives increases. This provides the information of possibilities which we have. It is not an automatic procedure but it requires the treatment planner to understand the results and to be able to draw conclusions. This is more important for the next step of MO inverse planning where not only the intensities but also the beam orientation and the optimal number of beams has to be found. The results of the MO inverse planning will be presented in a further work.

We used a very simple sampling method with uniform and importance factors with a specific scaling for w_{PTV}. Using a uniformly distributed importance factors with s = 1 requires approximately 100 times more solutions to be processed than using s = 100 to obtain the same number of clinical acceptable solutions. Other better methods may exist such as using a few different scaling parameters s and smaller k in order to sample the Pareto front with different resolutions.

As the number of objectives increases a very large number of solutions could be required to obtain solutions more uniformly distributed on the Pareto front. The mapping from importance to objective space is highly complex. For this purpose MO evolutionary algorithms supported by *L-BFGS* seem to be a solution for this problem. Their strategy is to sample the Pareto front uniformly. For this *L-BFGS* is important as standard genetic algorithms perform very poor as an analysis showed which we did. With a support of *L-BFGS* hybrid genetic algorithms are able to obtain more than 1000 solutions in a few minutes.

V. Acknowledgements

This research was supported through a European Community Marie-Curie Fellowship (HPTM-2000-00011)

References

Cho P S, Lee S, Marks II R J, Ohm S, Sutlief S G and Phillips M H 1998 Optimization of intensity modulated beams with volume constraints using two methods: Cost function minimization and projections onto convex sets *Med. Phys.* **25** 435-43

Cotrutz C, Lahanas M, Kappas C and Baltas D 2001 A multiobjective gradient based dose optimization algorithm for conformal radiotherapy *Phys. Med. Biol.* **46** 2161-75

Deasy J O 1997 Multiple local minima in radiotherapy optimization problems with dose-volume constraints *Med. Phys.* **24** 1157-61

Haas O C L, Burnham K J and Mills J A 1998 Optimization of beam orientation in radiotherapy using planar geometry *Phys. Med. Biol.* **43** 2179-93

Haas O C L Radiotherapy Treatment Planning: New System Approaches. Springer Verlag London, Advances in Industrial Control Monograph, ISBN 1-85233-063-5, 1999.

IEC, International Electrotechnical Commission Subcommittee 62C, IEC Draft 1217: Radiotherapy Equipment – Coordinates, Movements and Scales, 26 March 1993

Lahanas M, Baltas D and Giannouli S 2003 Global convergence analysis of fast multiobjective gradient based dose optimization algorithms for high dose rate brachytherapy *Phys. Med. Biol.* **48** 599-617

Liu D C and Nocedal J 1989 On the limited memory BFGS method for large scale optimization *Mathematical Programming* **45** 503-28

Llacer J, Solberg T D, Promberger C 2001 Comparative behaviour of the Dynamically Penalized Likelihood algorithm in inverse radiation therapy planning *Phys. Med. Biol.* **46** 2637-63

Llacer J, Deasy J O, Bortfeld T R, Solberg T D and Promberger C 2003 Absence of multiple local minima effects in intensity modulated optimization with dose-volume constraints *Phys. Med. Biol.* **48** 183-210

Miettinen K M Nonlinear Multiobjective Optimization 1999 Kluwer Academic Publisher Boston

Mohan R, Wang X, Jackson A, Bortfeld T, Boyer A L, Kutcher G J, Leibel S A, Fuks Z and Ling C C 1994 The potential and limitations of the inverse radiotherapy technique *Radiother. Oncol.* **32** 232-48

Press W H, Teukolsky S A, Vetterling W T and Flannery B. P. 1992 Numerical Recipes in C,2nd ed. Cambridge University Press Cambridge England.

Rowbottom G R and Webb S 2002 Configuration space analysis of common cost functions in radiotherapy beam-weight optimization algorithms, *Phys. Med. Biol* **47** 65-77

Shepard D M, Olivera G H, Rechwerdt P J and Mackie T R 2000 Iterative approaches to dose optimization in tomotherapy *Phys. Med. Biol*. **45** 69-90

Szu H and Hartley R 1987 Fast Simulated Annealing *Phys. Lett. A* **122** 157-62

Xing L, Li J G, Donaldson S, Le Q T and Boyer A L 1999 Optimization of importance factors in inverse planning *Phys Med. Biol.* 44 2525-36

Yu Y 1997 Multiobjective decision theory for computational optimization in radiation therapy *Med. Phys.* **25** 445-54

Li R P and Yin F F Optimization of inverse treatment planning using a fuzzy weight function *Med. Phys.* **27** 691-700