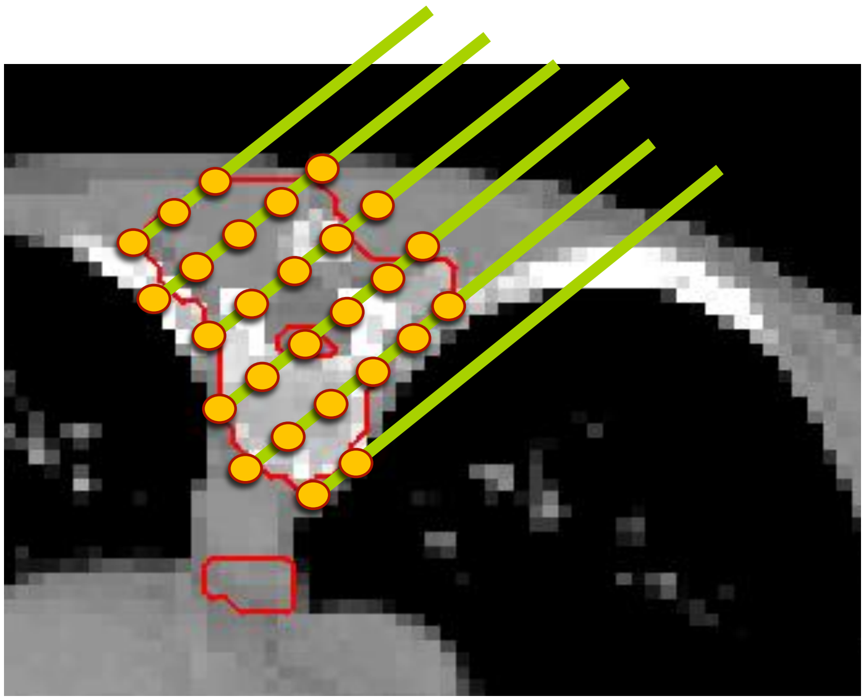
Treatment plan optimization

Intensity-modulated proton therapy (IMPT)



$$\underset{x}{\text{minimize}} \quad w_T \sum_{i \in T} (d_i - 70)^2 + w_H \sum_{i \in H} d_i$$

$$\text{subject to} \quad x_j \geq 0 \quad \forall j$$

$$d_i \leq 50 \quad \forall i \in S$$

Overview

1. CT scan is performed (possibly MRI, PET in addition)
2. Images are registered
3. Tumor volume and radiosensitive organs are delineated
4. Treatment plan optimization

Dose calculation (physics) + Optimization (mathematics)

A. Fluence map optimization

- Intensity-modulated proton therapy (IMRT)

B. IMRT planning

- leaf sequencing
- direct aperture optimization
- VMAT optimization

Treatment planning

High level goal

- achieve cancer cure
- avoid serious side effects

clinical outcome data

(dose-response relation)

Intermediate goal

- prescribed dose to the tumor
- maximum tolerance doses to the organs at risks

Treatment planning

Intermediate goal

- prescribed dose to the tumor
- maximum tolerance doses to the organs at risks

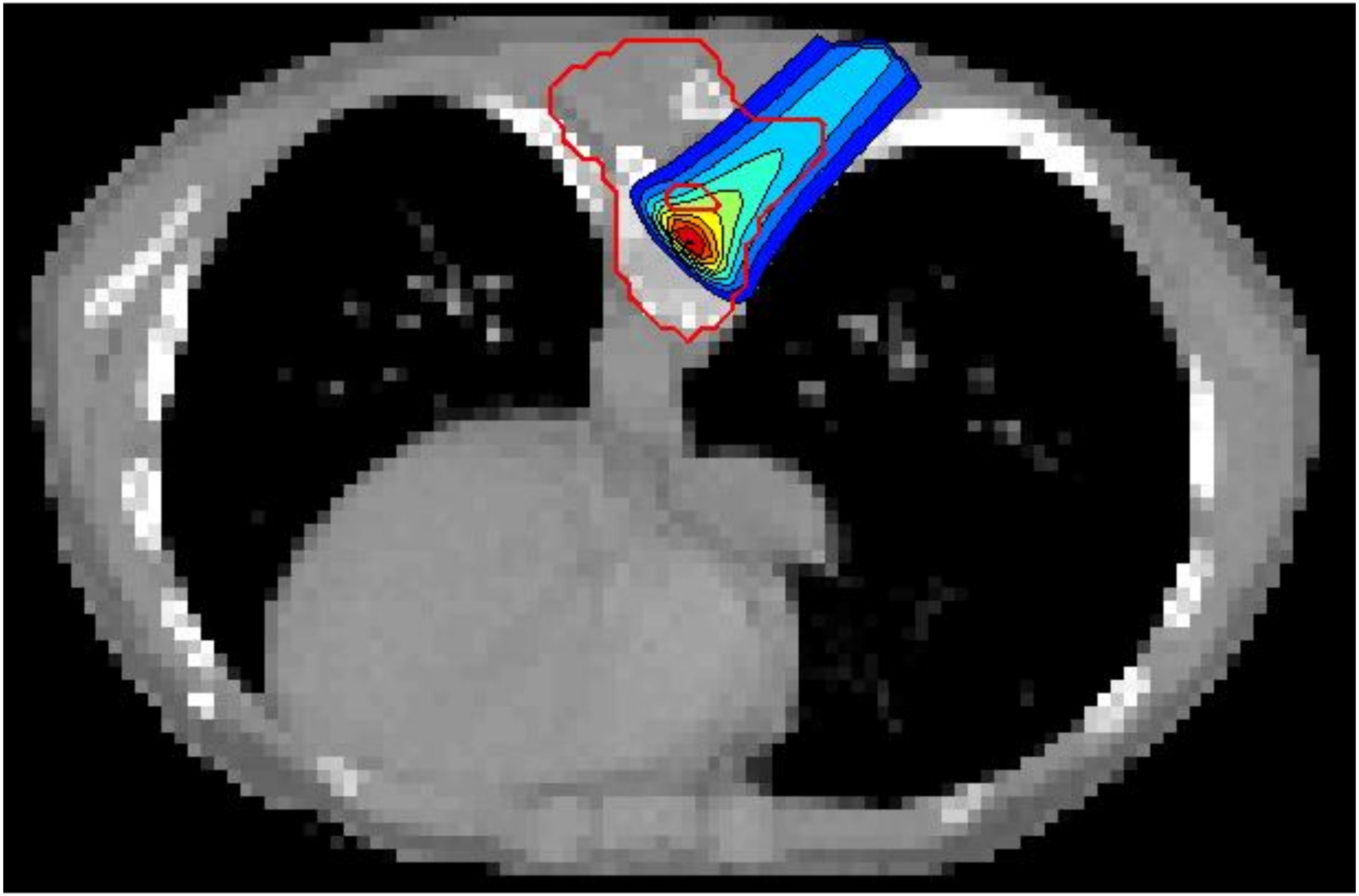
Treatment planning

(how can we realize the desired dose distribution)

Treatment plan parameters

- beam parameters (direction, energy, position, spot size)
- beam intensities

Treatment parameters



Treatment parameters

What are treatment parameters?

- incident direction (beam angle)
- beam size (sigma)
- initial energy
- lateral position
- intensity (fluence)

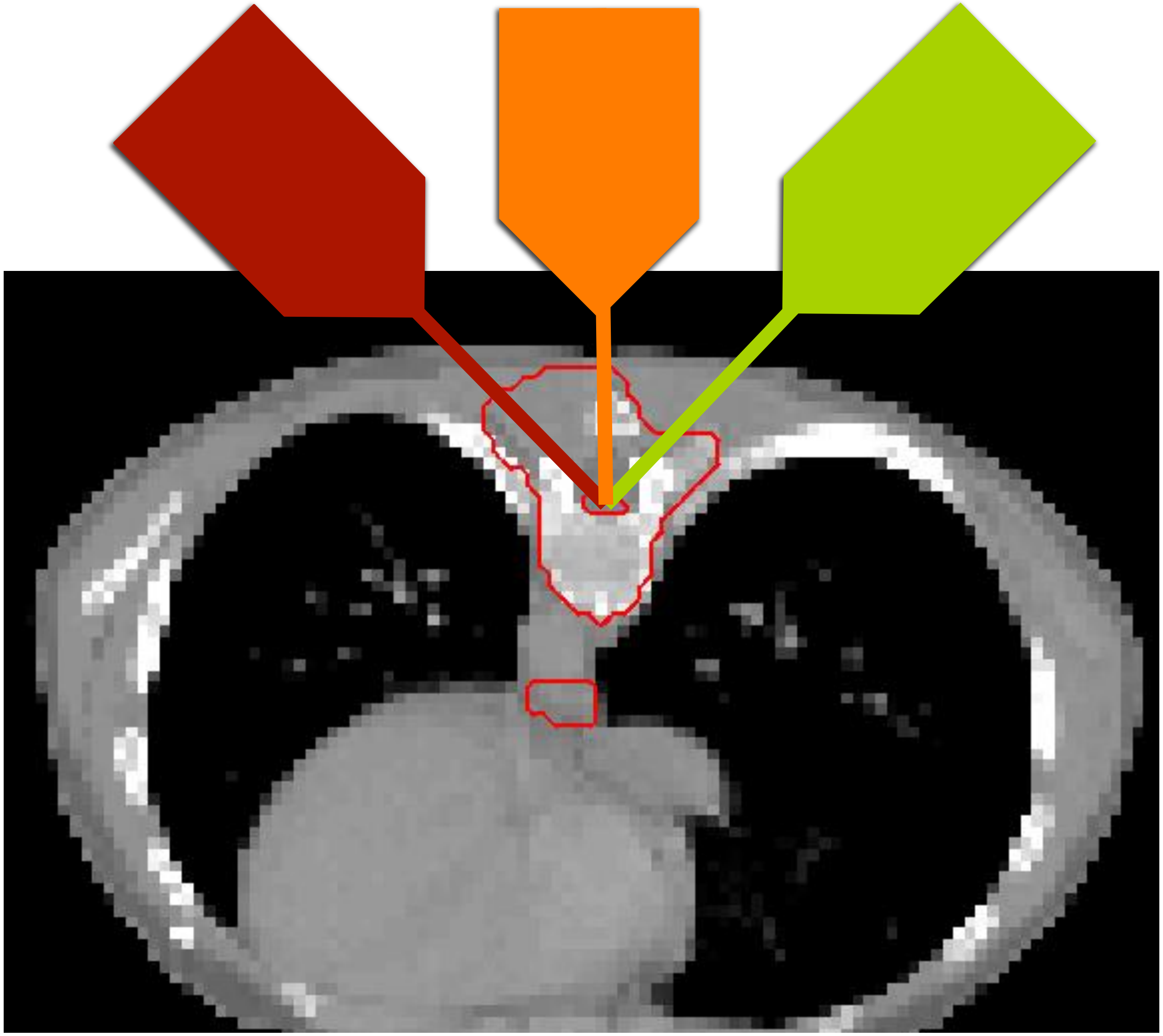
Treatment parameters

How are they determined?

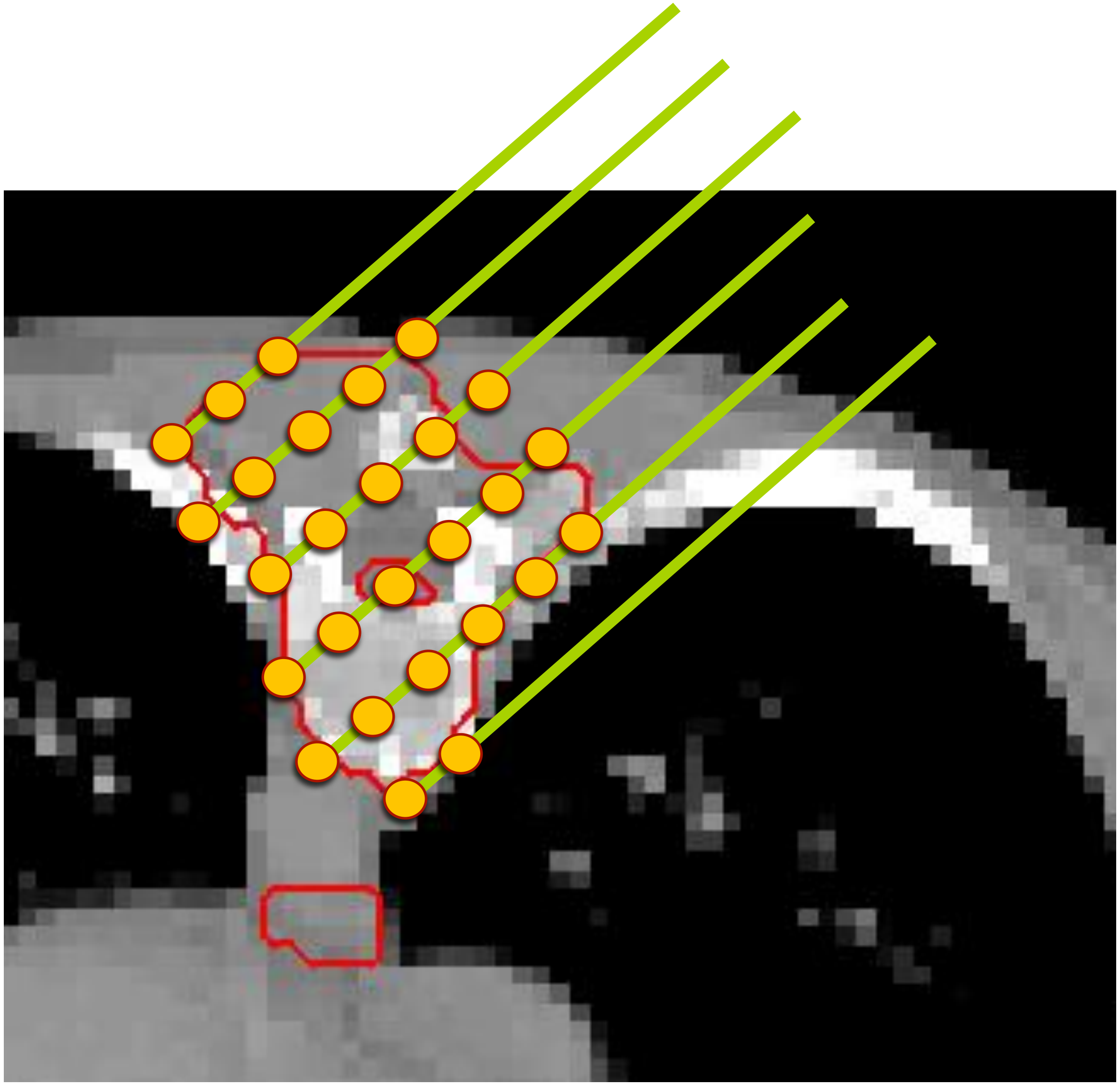
- incident direction (beam angle) — chosen manually (based on geometry)
- beam size (sigma) — fixed (determined by machine)
- initial energy
- lateral position

  — fixed pencil beam grid (to cover the entire tumor)
- intensity (fluence) — Determined by mathematical optimization methods

Beam directions



Pencil beam grid



Pencil beam grid

Lateral spacing: about one sigma (beam width)

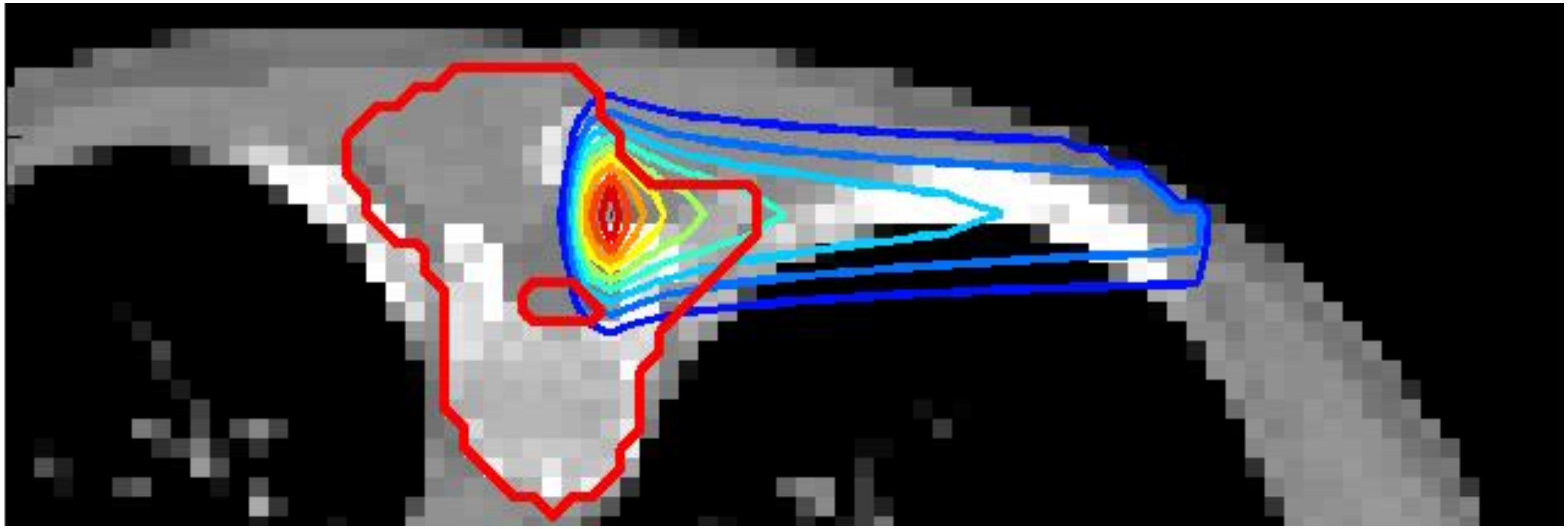
Energies: typically there is a fixed set of energies (e.g. at 5 mm distance in range)

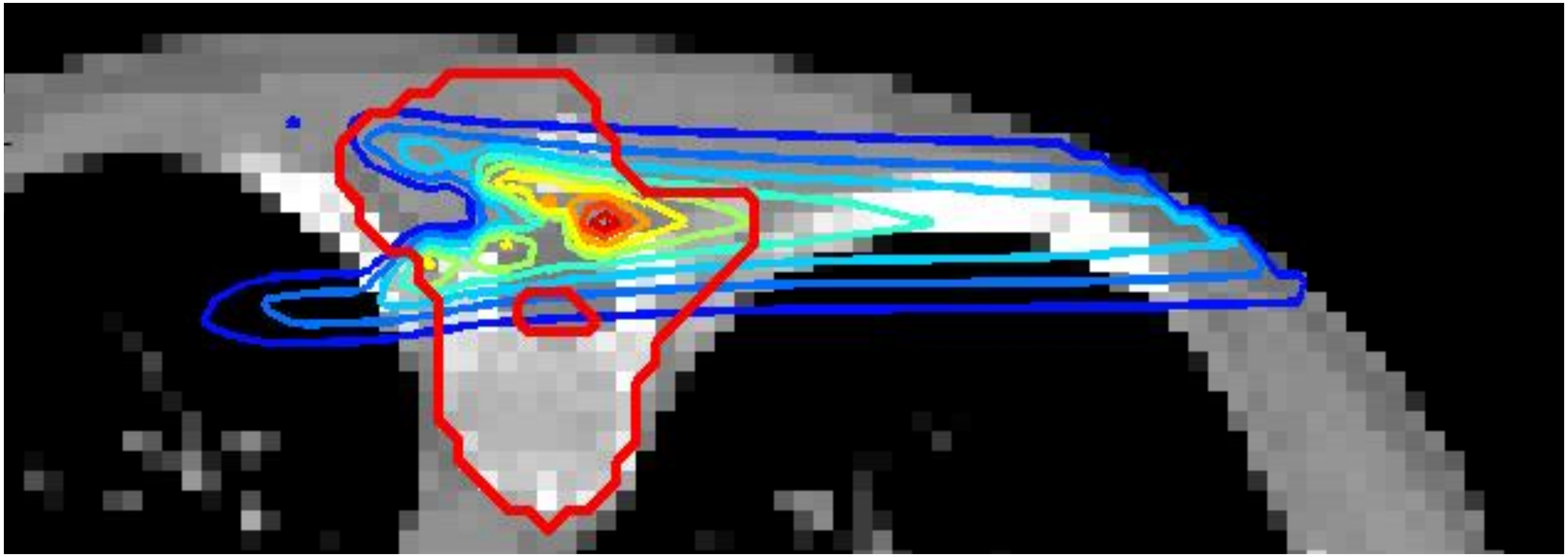
- need ray tracing in order to find the corresponding bragg peak locations
- pick the energies that fall into the target volume

➔ set of pencil beams characterized by (angle, energy, lateral position)

➔ calculate dose distributions of all pencil beams

Dose calculation





Dose-influence matrix

Pencil beam dose distribution:

$$\underbrace{D_{(\theta, x', y', E_0)}}_{\text{pencil beam attributes}} \underbrace{(x, y, z)}_{\text{position in the patient}}$$

Dose contribution of beam j to voxel i

$$D_{ij}$$

pencil beam j

voxel index i

Dose distribution

Total dose in voxel i is given by the sum over the contributions of all beams j

$$d_i = \sum_j x_j D_{ij}$$

Intensity of beam j

[Giga protons]

Dose contribution of beam j per unit intensity

[Gray per Giga proton]

Treatment planning problem

We want to determine the beam intensities x_j so that we closely approximate a desired dose distribution

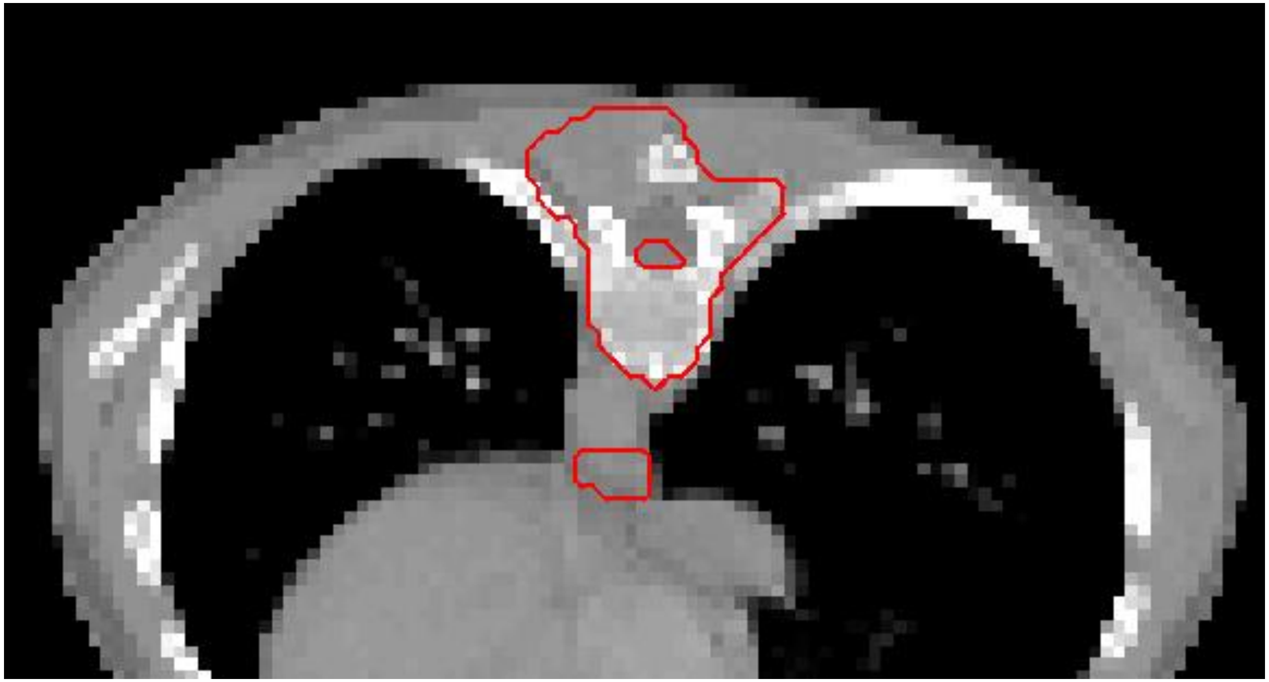
- The number of pencil beams is very large:

$$N \approx 10^3 \cdots 10^5$$

- cannot be done manually
- need for mathematical optimization methods

Mathematical optimization

Goals:
- deliver dose of 70 Gray to the tumor
- limit dose to spinal cord to 50 Gray
- limit dose to esophagus to 60 Gray
- minimize dose to the adjacent healthy tissues (lungs, heart, …)



Mathematical optimization

Mathematical optimization problems are defined through:

- Objectives (wishes)
- Constraints (definitely has to be fulfilled)

General formulation:

$$\underset{x}{\text{minimize}} \quad f(x) \qquad \text{(objective function)}$$

$$\text{subject to} \quad g_k(x) \le 0 \quad \forall k \qquad \text{(set of constraints)}$$

Mathematical optimization

Application to radiotherapy

$$\underset{x}{\text{minimize}} \quad \underbrace{w_T \sum_{i\in T} (d_i - 70)^2}_{\text{minimize deviation from 70 Gray in the tumor}} + \underbrace{w_H \sum_{i\in H} d_i}_{\text{minimize dose in healthy tissues}}$$

$$\text{subject to} \quad x_j \geq 0 \quad \forall j$$

fluence cannot be negative

$$d_i \leq 40 \quad \forall i \in S$$

limit dose to spinal cord to less than 40 Gray

Mathematical optimization

<u>as a function of beam weights:</u>

$$\underset{x}{\text{minimize}} \quad \underbrace{w_T \sum_{i \in T} \left(\sum_j x_j D_{ij} - 70 \right)^2 + w_H \sum_{i \in H} \sum_j x_j D_{ij}}$$

quadratic function of x_j

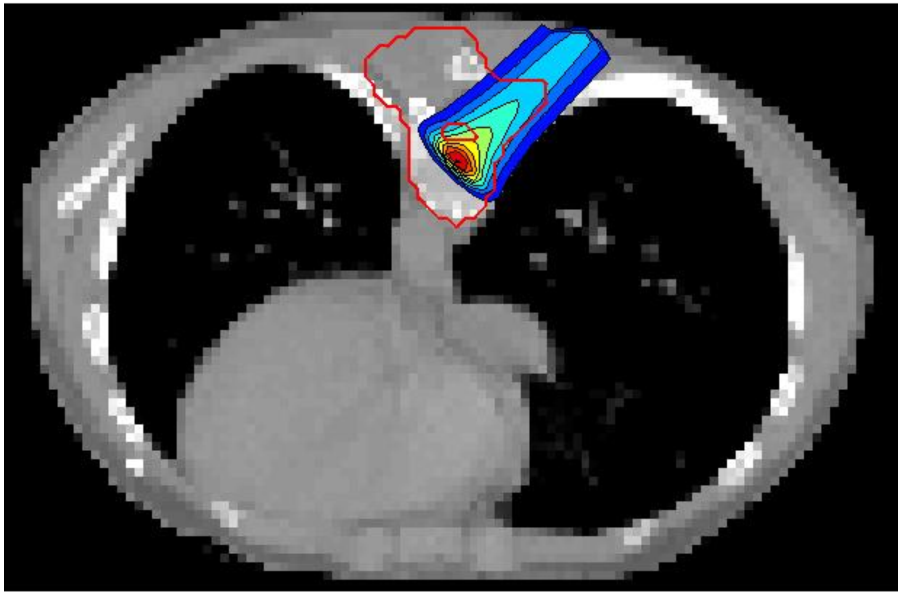
$$\text{subject to} \quad x_j \geq 0 \quad \forall j$$

$$\underbrace{\sum_j x_j D_{ij} \leq 40 \quad \forall i \in S}$$

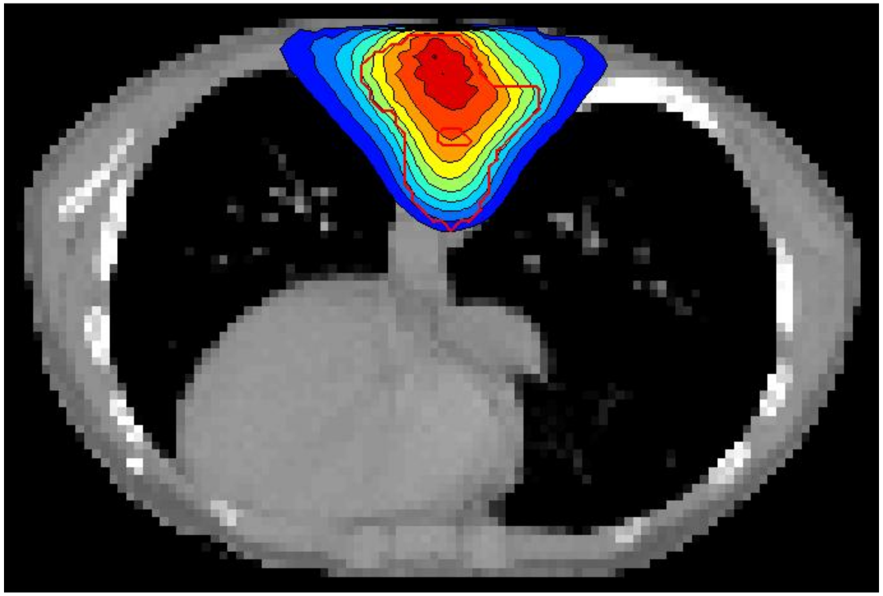
linear functions of x_j

Starting point

Single pencil beam



Equal intensity for all pencil beams



Goal:

determine beam intensities that yield a homogeneous dose in the tumor and spare healthy tissues

Optimization problem

<u>Spot intensity optimization problem:</u>

$$\underset{x}{\text{minimize}} \qquad \underbrace{\sum_{i \in T} \left(\sum_{j} x_j D_{ij} - 70 \right)^2}_{\text{quadratic function of } x_j}$$

$$\text{subject to} \qquad x_j \geq 0 \quad \forall j$$

$$\underbrace{\sum_{j} x_j D_{ij} \leq 40 \quad \forall i \in S}_{\text{linear functions of } x_j}$$

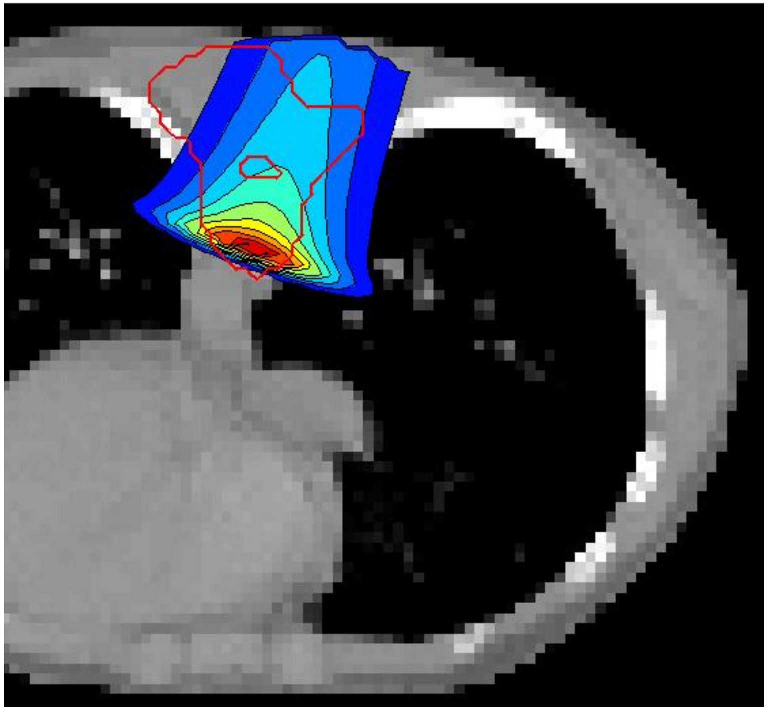
Geometry

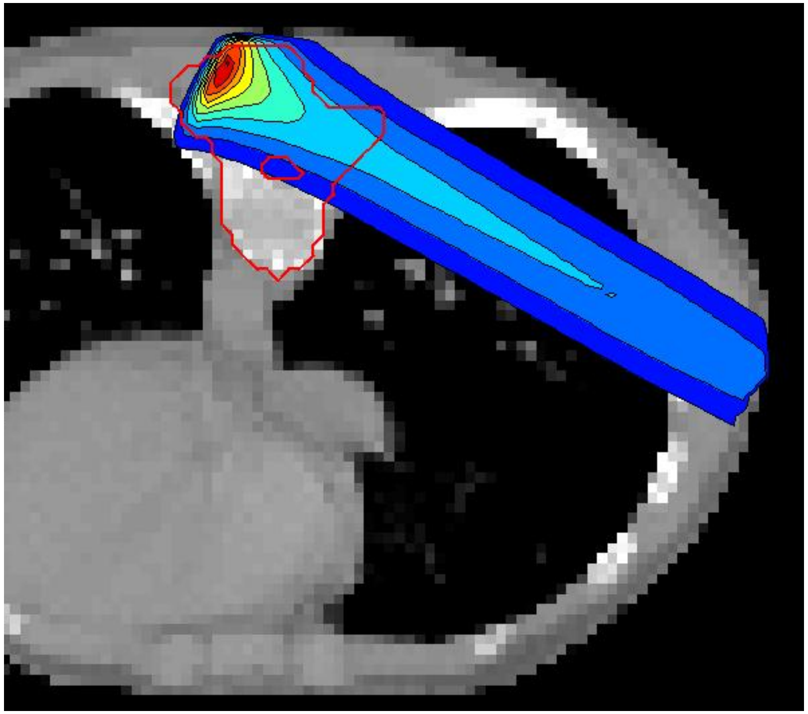
Visualize the optimization problem

- consider only 2 pencil beams

➔ 2 optimization variables (intensities)

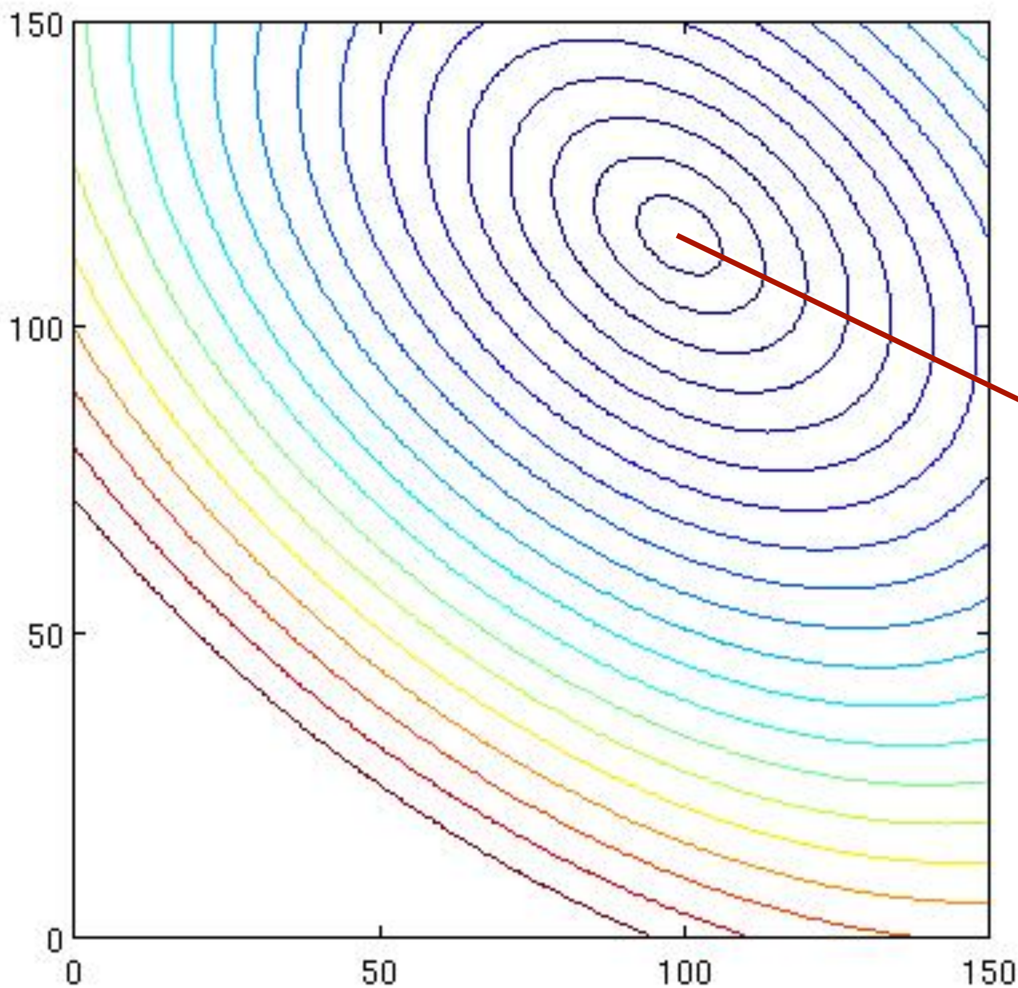
➔ we can plot the objective function and the constraints





Objective function

Objective function for this problem:

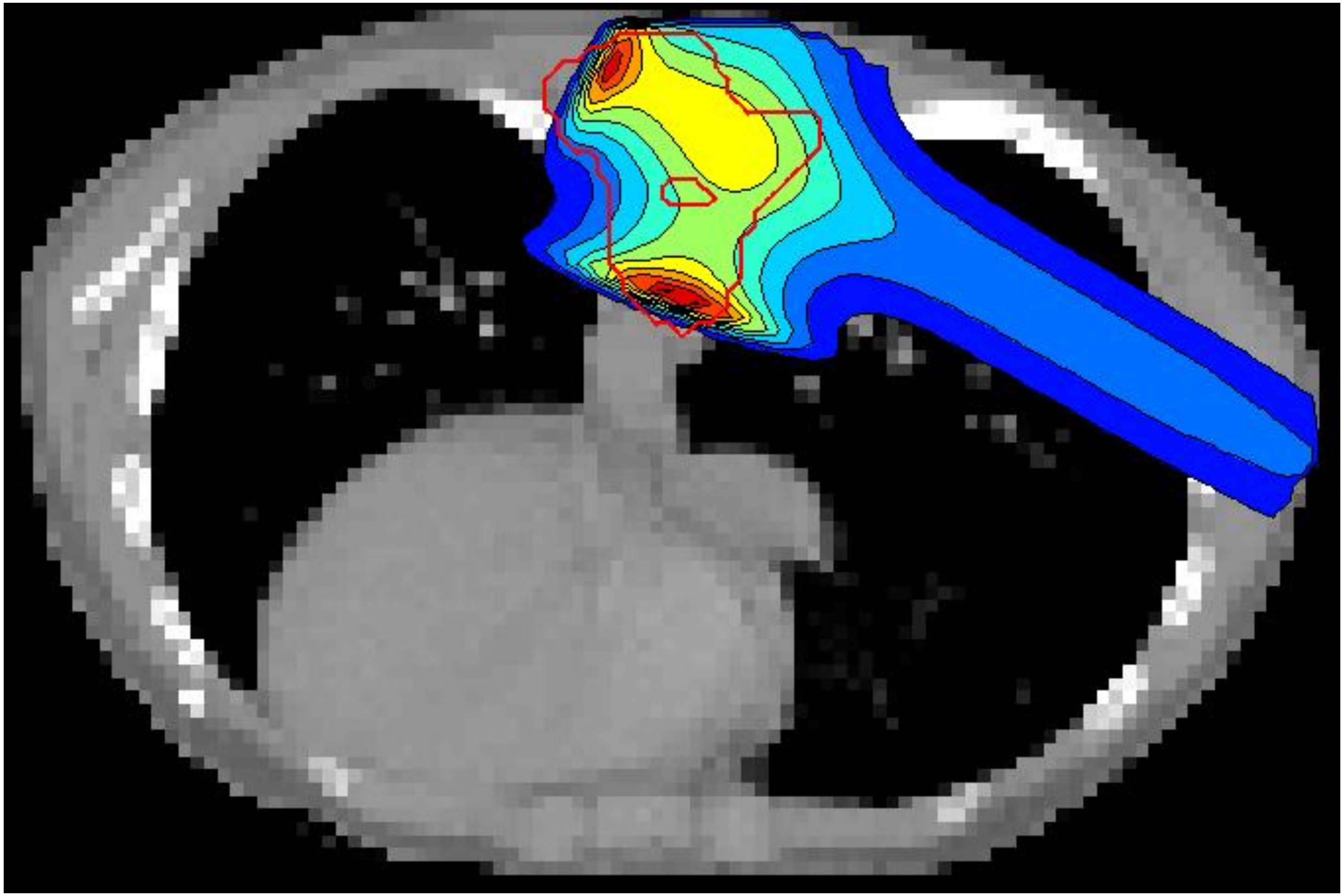


$$f(d) = \sum_{i \in T} \left(\sum_j x_j D_{ij} - 70 \right)^2$$

optimal solution

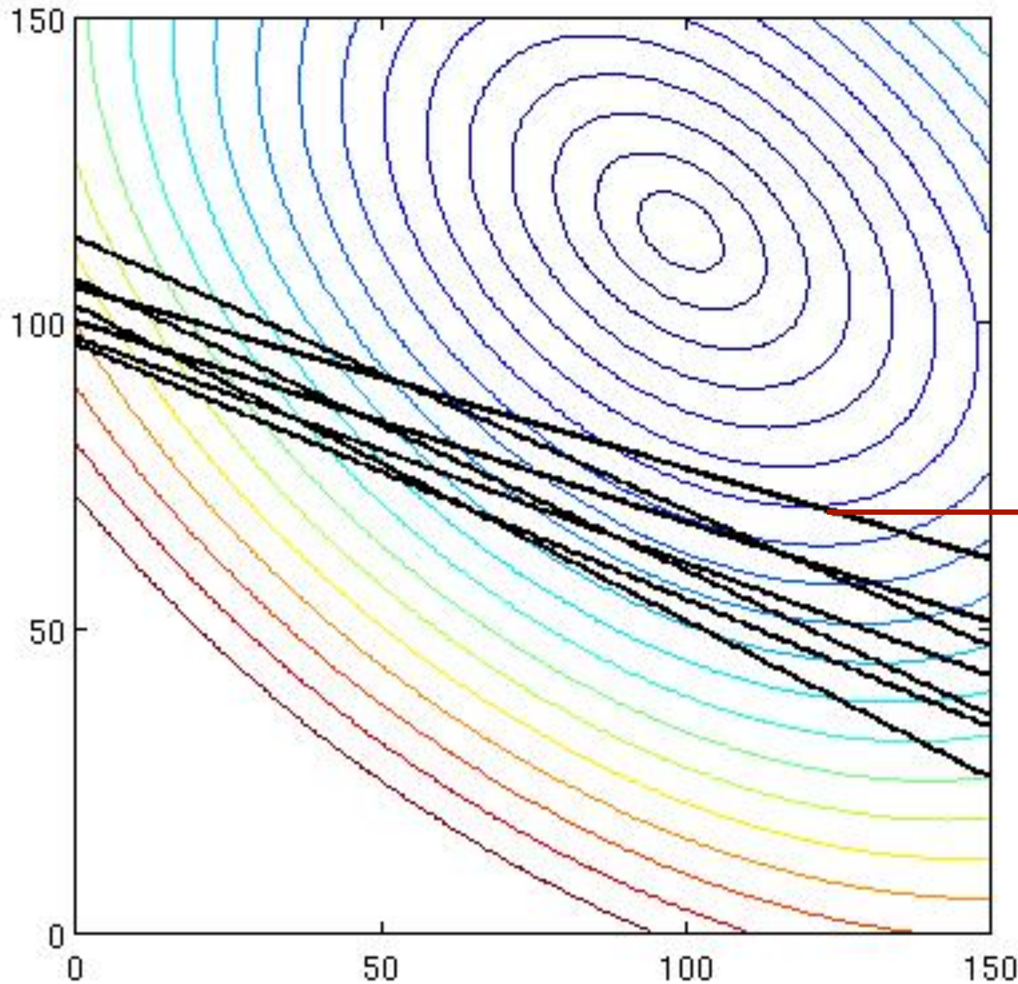
Optimal solution

Corresponding dose distribution:



Constraints

Constraints on the spinal cord are linear functions of the intensities:



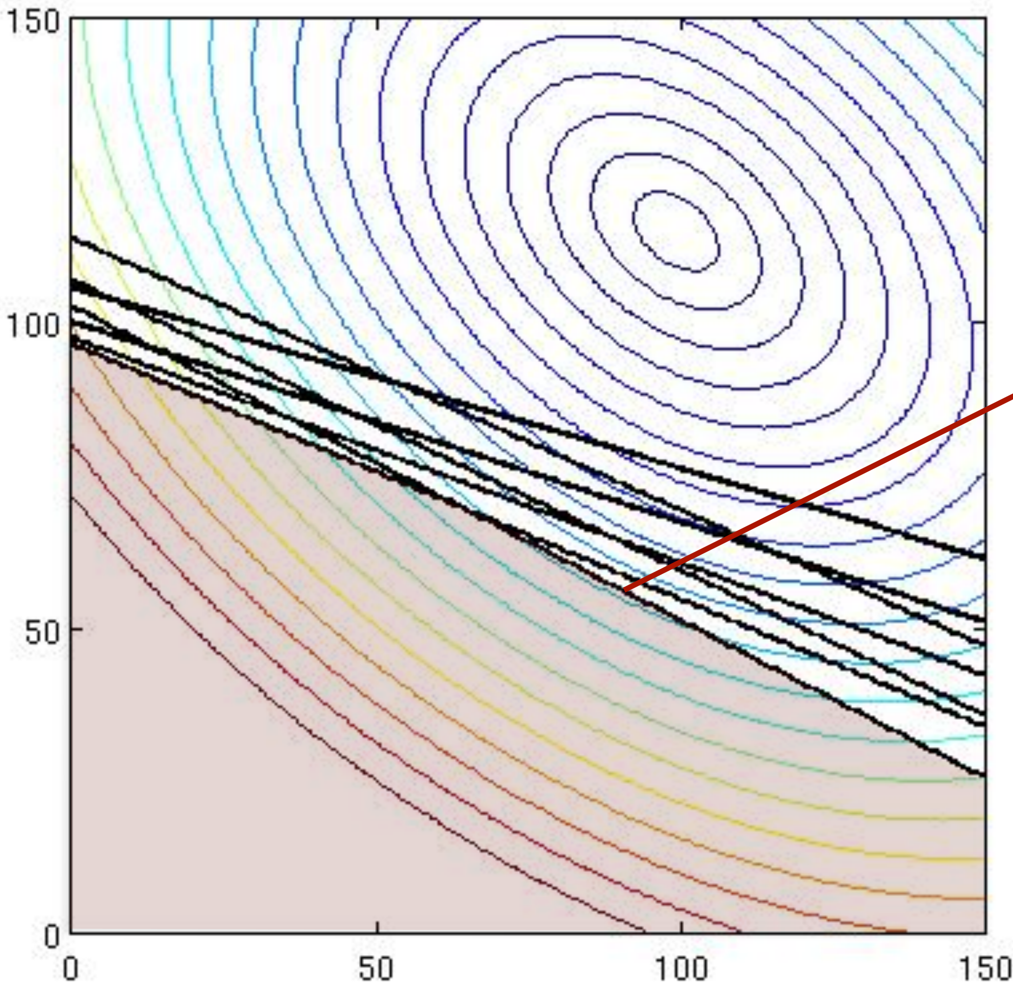
$$x_1 D_{i1} + x_2 D_{i2} \leq 40 \quad \forall i \in S$$

lines where constraints hold with equality

$$x_1 D_{i1} + x_2 D_{i2} = 40$$

(7 spinal cord voxels)

Constraints



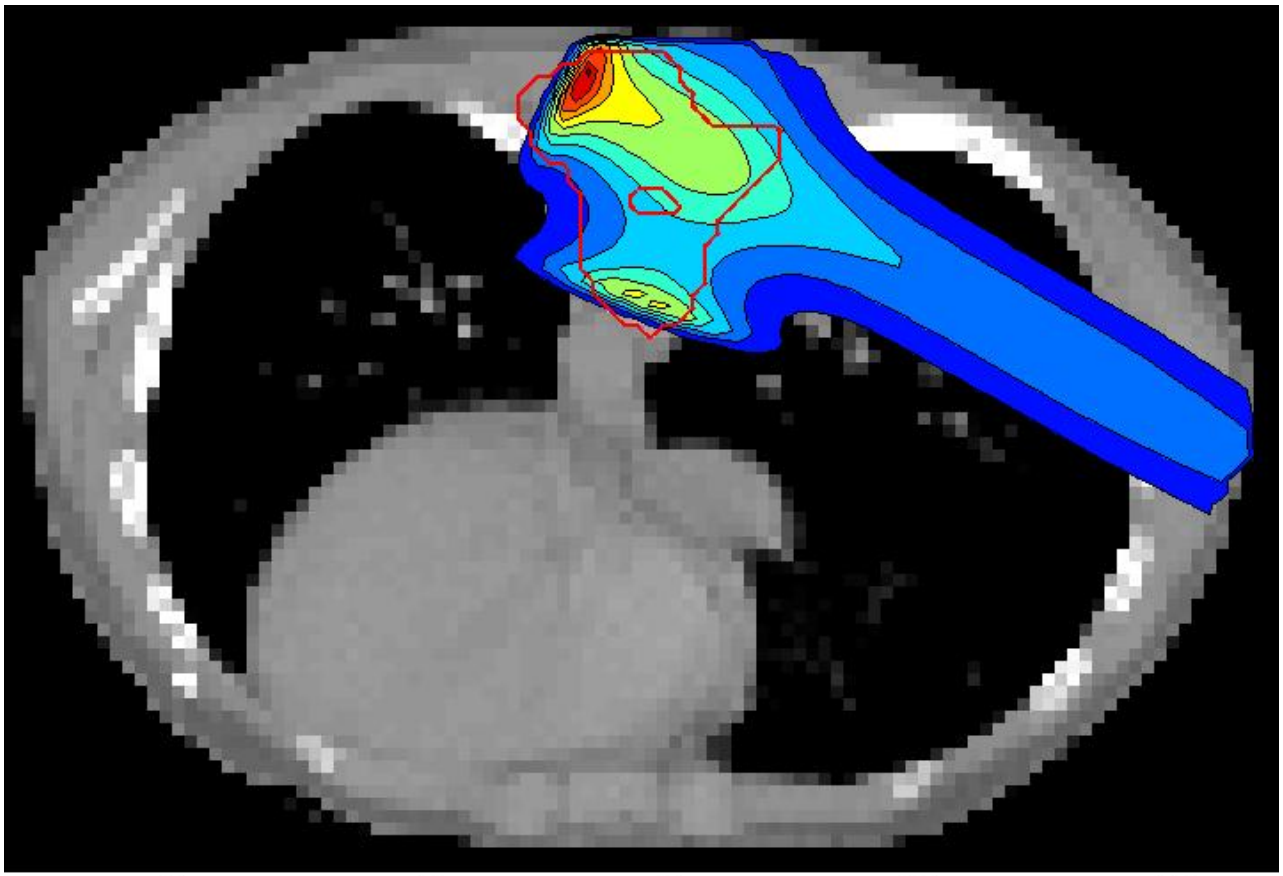
optimal solution
within the feasible set

Feasible set

(region where all
constraints are satisfied)

Constraints

Optimal dose distribution (satisfying the constraints)



(lower weight for the pencil beam that contributes more dose to the spinal cord)

Optimization algorithms

Here: two variables

➔ can find the optimum by calculating the objective function for all (x_1,x_2)

Realistically: $10^3 \ldots 10^5$ variables

➔ requires mathematical optimization methods

Gradient descent

First step:

Unconstrained optimization (consider only f(d))

Basic method:

Gradient descent

$$\nabla f = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{pmatrix}$$

Recall:

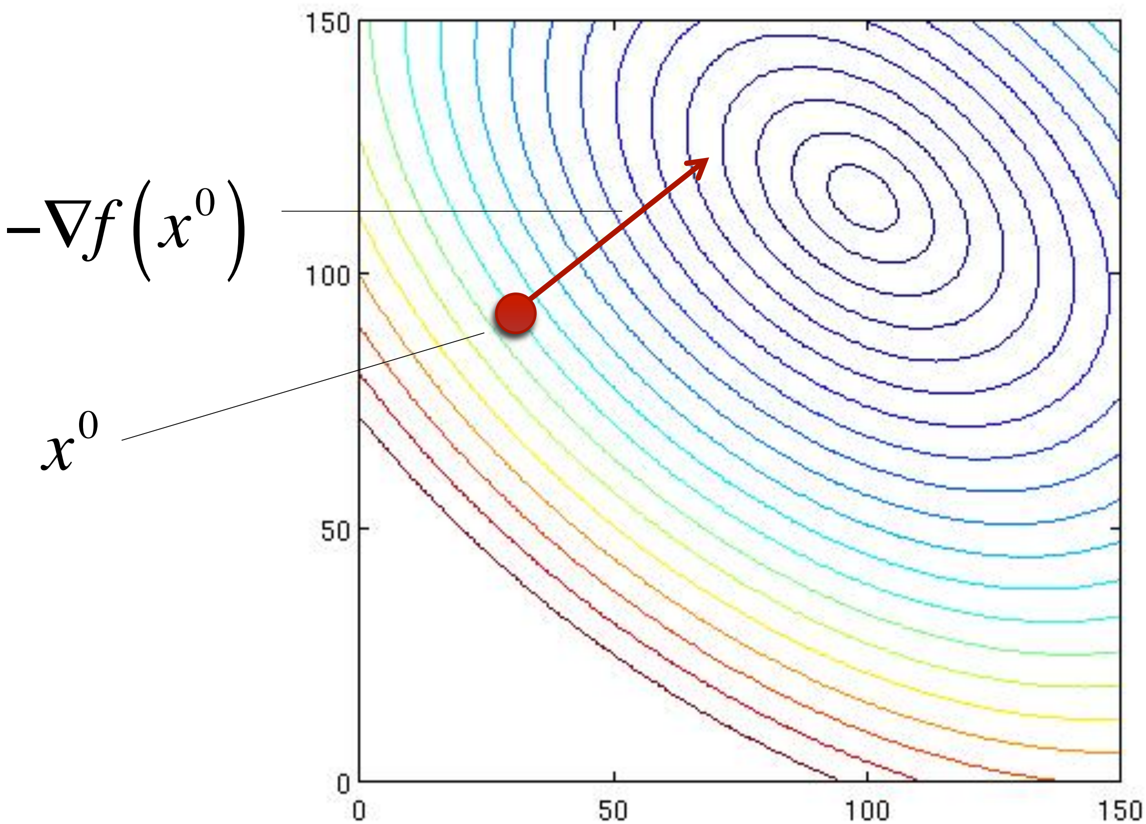
- Gradient vector is perpendicular to the isoline of the objective function f
- points in the direction of steepest increase of the function f

Gradient descent

Let x^0 be the initial guess of beam weights

We can get to a lower objective function value by talking a small step along the negative gradient direction

$$x^1 = x^0 - \alpha \nabla f\left(x^0\right)$$



Gradient descent

Iterative algorithm:

- initialize ($x_1^0 = 0$, $x_2^0 = 0$)
- choose small enough step size α > 0

while (stopping criterion not fulfilled)

$$x^{k+1} = x^k - \alpha \nabla f\left(x^k\right)$$

$$k \leftarrow k + 1$$

end

Gradient descent

Gradient calculation:

Objective:

$$f(x) = \sum_{i \in T} \left(d_i - 70\right)^2 = \sum_{i \in T} \left(\sum_j x_j D_{ij} - 70 \right)^2$$

Gradient: $$\frac{\partial f}{\partial x_j} = \sum_i \frac{\partial f}{\partial d_i} \frac{\partial d_i}{\partial x_j}$$ (Chain rule)

$$\frac{\partial f}{\partial x_j} = \sum_{i \in T} 2\left(d_i - 70\right) \frac{\partial d_i}{\partial x_j} = \sum_{i \in T} 2\left(d_i - 70\right) D_{ij}$$

Dose constraints

Handling dose constraints:

Penalty method:

convert constraint into a penalty term in the objective function

Constraint: $d_i \leq 40 \quad \forall i \in S$

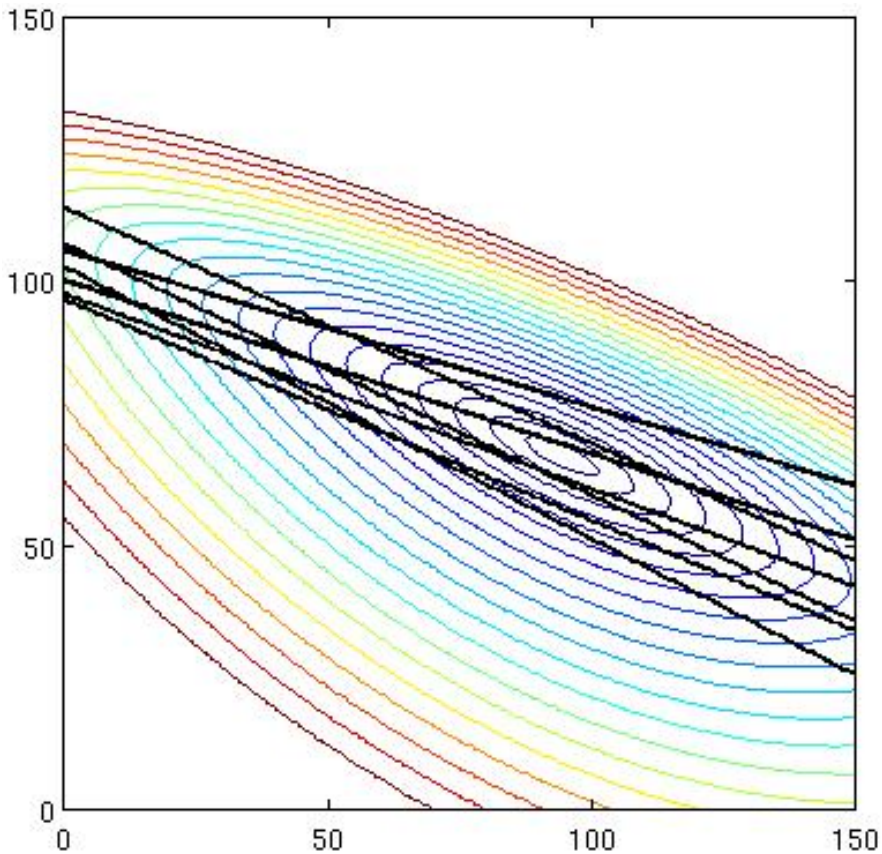
Penalty term: $f + \mu\,(d_i - 40)_+^2$ $\quad (d_i - 40)_+^2 = \begin{cases} (d_i - 40)^2 & d_i > 40 \\ 0 & \text{otherwise} \end{cases}$

➔ does not change the objective function within the feasible region

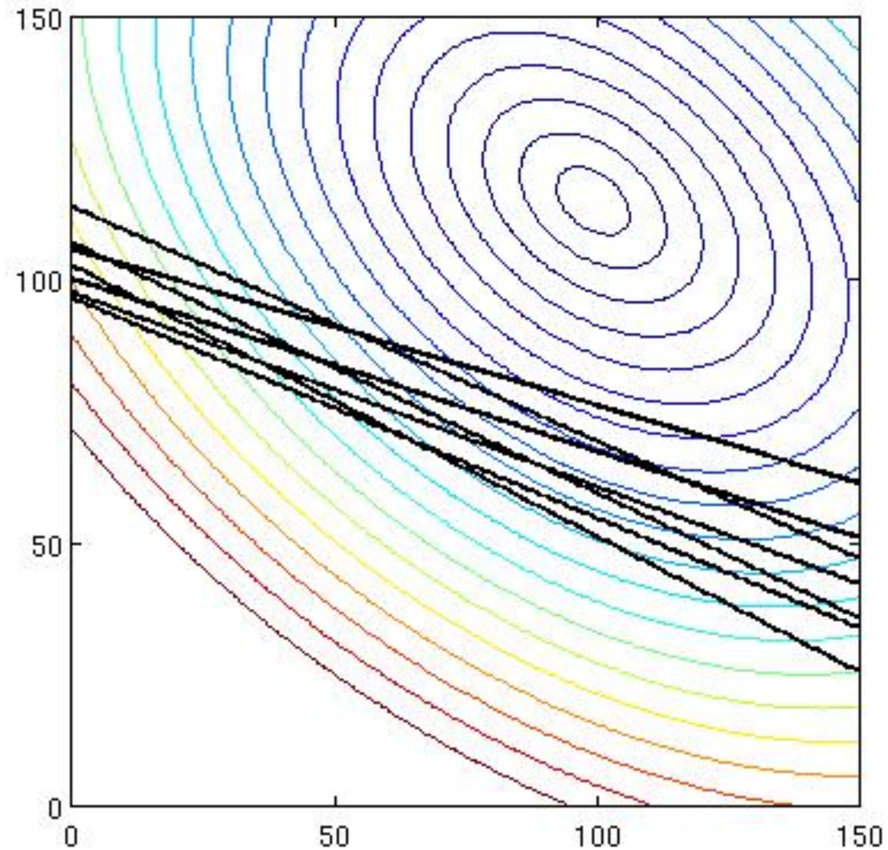
➔ creates an unconstrained optimization problem

Dose constraints

Visualization:



$$f + 10 \cdot \sum_{i \in S} (d_i - 40)_+^2$$

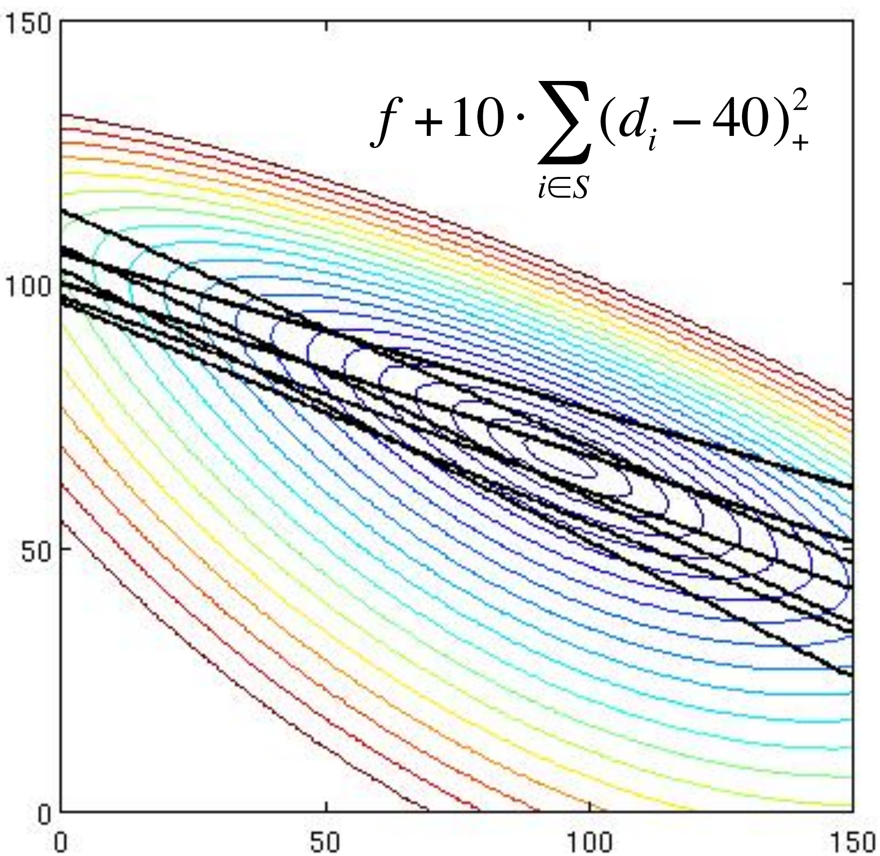


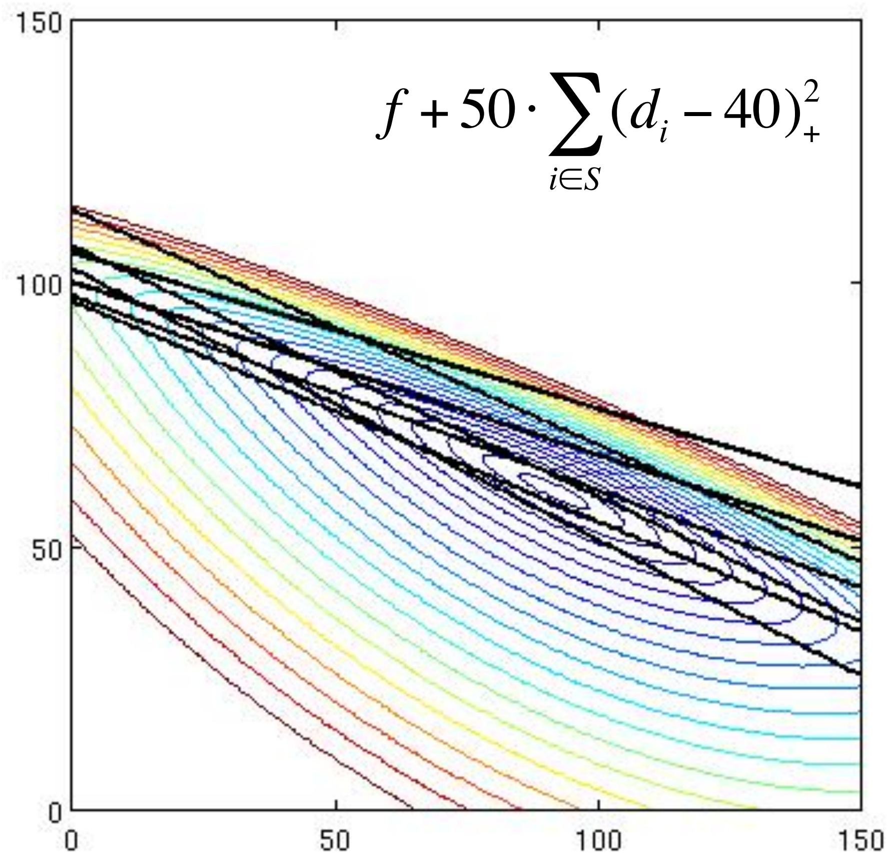
$$f$$

Dose constraints

Problem:

- constraints only fulfilled when μ -> ∞
- may lead to numerical problems, slow convergence





Projection method

Handling positivity constraints: $x_j \geq 0 \quad \forall j$

Projection method:

at every iteration, project current solution x^k onto the positivity constraint (i.e. find the closest x^k that fulfills the constraint)

while (stopping criterion not fulfilled)

$$x^{k+1} = x^k - \alpha \nabla f\left(x^k\right)$$

$$x_j^{k+1} \leftarrow \max\left(0, x_j^{k+1}\right)$$

$$k \leftarrow k+1$$

end

Projection method

Remarks:

Projection method suited for bound constraints:

- constraint hyper planes are perpendicular
- projecting on every constraint once fulfills all constraints

General linear constraints:

- projection can be done analytically (efficiently)
- but has to be done iteratively (projection on one constraint may violate one that was satisfied before)

Nonlinear constraints:

- have to approximate projection

IMPT case

3 beams, 238 pencil beams, 5 mm sigma, 5 mm spacing

Objective: $$f(d) = \sum_{i \in T} \left(d_i - 70\right)^2 + 50 \cdot \sum_{i \in S} \left(d_i - 40\right)_+^2$$

while (k<1000)

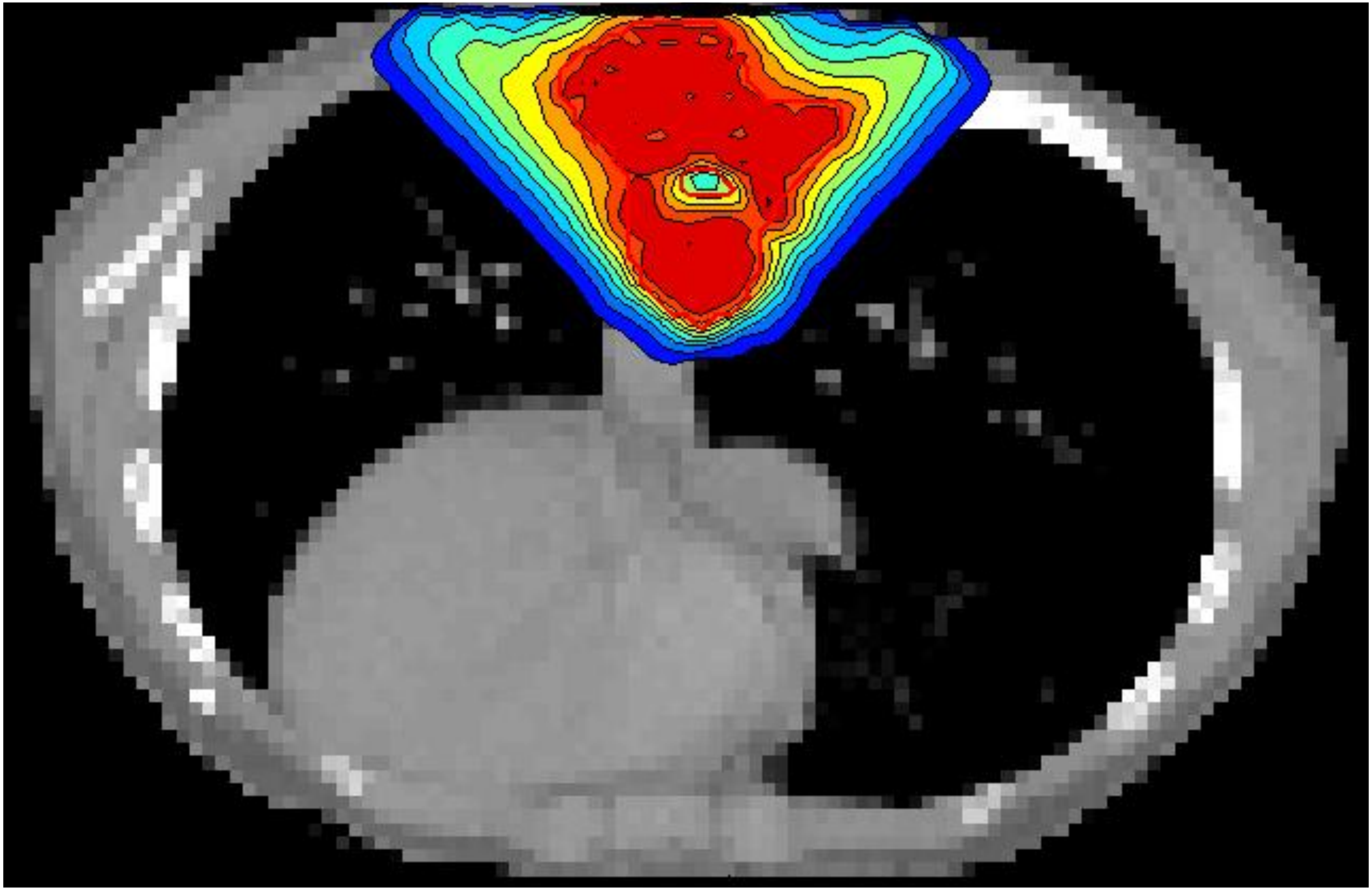
$$x^{k+1} = x^k - \alpha \nabla f\left(x^k\right)$$

$$x^{k+1} \leftarrow \max\left(0, x^{k+1}\right)$$

$$k \leftarrow k + 1$$

end

IMPT case



Improvements

Improvements:

Line search method:

- determining the best step size

Using curvature information

- improving the step direction using a quadratic approximation of the objective function

Constraint handling

- pure penalty method fulfills constraints only for $\mu \to \infty$
- better methods based on Lagrange multiplier theory

Newton method

Using curvature information

Newton method

The objective function is approximated as a quadratic function at the current beam intensities x

$$f(x+\Delta x) \approx f(x) + \Delta x \cdot \nabla f(x) + \tfrac{1}{2}\Delta x H(x) \Delta x$$

Hessian $$H(x) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f}{\partial x_2^2} \end{pmatrix}$$

(Taylor expansion at the current solution x)

Newton method

Idea:

go to the minimum of the quadratic approximation
in the next iteration

Quadratic approximation

$$f(x+\Delta x) \approx f(x) + \Delta x \cdot \nabla f(x) + \tfrac{1}{2}\Delta x H(x) \Delta x$$

find Δx that minimizes f(x+Δx)

$$\nabla_{\Delta x} f(x+\Delta x) = \nabla f(x) + H(x)\Delta x \overset{!}{=} 0$$

Solution: $\Delta x^* = -H^{-1}(x)\nabla f(x)$

Newton method

Newton method:

(unconstrained optimization)

while (stopping criterion not fulfilled)

$$x^{k+1} = x^k - \alpha^k H^{-1}\left(x^k\right)\nabla f\left(x^k\right)$$

$$k \leftarrow k+1$$

end

Gradient direction is modified by multiplication with the inverse Hessian

The Newton method has a natural step size α = 1

Newton method

Remarks:

The pure Newton method is not often used in practice

- Problems if second derivatives are close to zero
- Explicit calculation of the Hessian (plus its inverse!) is too time consuming if number of variables is large

But, many general purpose methods for nonlinear optimization are based on a quadratic approximation of the cost function

- Quasi-Newton methods (BFGS, L-BFGS) iteratively approximate the inverse Hessian
- Sequential quadratic programming (SQP)

Lagrange multiplier theory

Quadratic penalty functions:

$$\tilde{f}(d) = f(d) + \mu \cdot \sum_{i \in S} \left(d_i - 40\right)_+^2$$

Often used in radiotherapy planning, but

- constraints strictly fulfilled only for μ -> ∞
- ill-conditioning, slow convergence for large μ

Easiest extension: Augmented Lagrange function

$$\tilde{L}(d) = \underbrace{f(d) + \sum_{i \in S} \lambda_i \left(d_i - 40\right)}_{\text{Lagrange function}} + \sum_{i \in S} \mu_i \left(d_i - 40\right)_+^2$$

Lagrange multiplier theory

$$\tilde{L}(d) = f(d) + \sum_{i \in S} \lambda_i \left(d_i - 40\right) + \sum_{i \in S} \mu_i \left(d_i - 40\right)_+^2$$

Lagrange multipliers

Properties:

There exists a set of Lagrange multipliers λ ≥ 0 such that the unconstraint minimum of L(d) is identical to the constrained minimum of f(d) subject to d_i-40≤0

For large enough (but finite) μ, the augmented Lagrange function L(d) is convex around the minimum.

Remark: Lagrange multipliers not know a priori, have to be estimated using an algorithm