State-of-the-art of optimization methods and software
and a Radiation Therapy Treatment Application

Tamás Terlaky

Industrial and Systems Engineering

Lehigh University

The RTT application part is joint work with

Mohammad R. Oskoorouchi (California State University San Marcos)

Dionne Aleman, Hamid R. Ghaffari (University of Toronto)

Outline

- Optimization: models, algorithms, software
  - Roots of linear optimization (LO)
  - Nonlinear Optimization -- Optimization classes
  - Black-Box / Derivative Free Optimization
  - Conic Linear Optimization: LO, SOCO, SDO
  - Robust Linear Optimization
  - Software
- Gamma knife surgery
  - Leksel Gamma Knife Perfexion
  - duration optimization
  - Computational experience
- Comparing SILO with MOSEK and a projected gradient alg.

Roots of Optimization

Farkas Gyula (1847–1930)

- Let $g, g_1, \ldots, g_m \in R^n$ given and $x \in R^n$,
- then $g_i^T x \leq 0, \;\; i = 1, ..., m$ imply $g^T x \leq 0$,
- **if and only if**
- there exists $0 \leq \lambda \in R^n$ such that $g = \lambda_1 g_1 + \ldots + \lambda_n g_n$.
- **Generalizations:** Convex Farkas Lemma (Farkas, 1906, no regularity)
  Infinite linear systems (Haar 1918)
  Discrete version (Murota, Tamura)

Motivation coming from theoretical mechanics,
Farkas studied conditions for mechanical equilibrium.
He built on the principles about virtual work (Fourier, Bernoulli).

Minimize functions / solve linear and nonlinear equations

Newton, Lagrange, Gauss, Euler, ……

What we really mean by

Optimization/Optimal

Misinterpretations, misunderstandings

Keywords	Should be	Frequently
optimal	best solution/design; no better solution or design exists that satisfies the requirements	best I got; in the given time, with my limited resources...
optimization	a discipline: theory, algorithms, analysis, software ... domain expertise	"programming" a process try to do better

Optimization, design optimization as a rigorous, mathematically formulated discipline has very limited exposure in engineering science, physics, undergraduate curriculum

OPTIMIZATION – as "WE" see it

$$\min \quad f(x)$$

$$g_j(x) \leq 0, \quad j = 1, \ldots, m,$$

$$h_i(x) = 0, \quad i = 1, \ldots, k,$$

where $x \in \mathcal{C} \subseteq R^n$, $f(x), g_j(x), h_i(x)$ are functions.

Linear Optimization
Quadratic Optimization
Convex Optimization
Nonlinear Optimization
Derivative Free Optimization
Bi-level Optimization
Global Optimization
Optimization with Equilibrium constraints
Discrete Optimization
Combinatorial Optimization
Network flows
Mixed-Integer Optimization
Stochastic Optimization
Robust Optimization
Multi-objective Optimization
Heuristic Optimization

- **analyze mathematical properties**
- **design and analyze algorithm**
- **implement software and solve problems**

Q: Where the functions, problem data come from?

BlackBox: Derivative Free Optimization

- Coordinate search (Hooke, Jeevs) $\min f(x)$
- Simplex methods (Nelder, Mead)
- Pattern search (Torczon, Dennis, Audet)
- Filtering (Kelley)

DFO: (Powell, Conn, Scheinberg, Toint, Vicente)

1. Choose a set of linearly independent model functions;
   (usually linear and quadratic monomials)
2. Choose a set of sample well poised points and a TR radius;
   (the same number as the number of functions)
3. Evaluate the function at the sample pints;
   (call the BlackBox)

DFO (part 2)

4. Find the interpolation function in the model space;

   (solve a linear system of equations)

   (independent model functions, poised point set)

5. Optimize the model around the best point in the trust region;

   (trust region subproblem)

6. Update the point set and the interpolating model function;

   (keep the best points, with good geometry, not far from best point)

   (recalculate the interpolation function, possibly adding new model function(s) and adjust, reduce/increase the TR radius)

7. Eventually improve the geometry of the model;

   (this is another trust region subproblem)

8. Repeat steps 5–7;

Cone-Linear Optimization (CLO)

Cone linear optimization problems play a crucial role in the theory, algorithms and applications of modern optimization.
The Primal-dual pair of CLO problems is given as

$$
\begin{array}{llll}
(P) & \min & c^T x & \\
& \text{s.t.} & Ax - b & \in \mathcal{C}_1 \\
& & x & \in \mathcal{C}_2
\end{array}
\qquad
\begin{array}{llll}
(D) & \max & b^T y & \\
& \text{s.t.} & c - A^T y & \in \mathcal{C}_2^* \\
& & y & \in \mathcal{C}_1^*,
\end{array}
$$

where $b, y \in \mathbb{R}^m$, $c, x \in \mathbb{R}^n$, $A : m \times n$ matrix, $\mathcal{C}_1, \mathcal{C}_2$ are convex cones and $\mathcal{C}_i^* = \{s \in \mathbb{R}^n : x^T s \geq 0, \ \forall x \in \mathcal{C}_i\}$ are the dual cones for $i = 1, 2$.

Important classes of CLO (not all) are solvable efficiently

(in polynomial time)

by using interior point methods.

Poyhedral cones: Linear Optimization

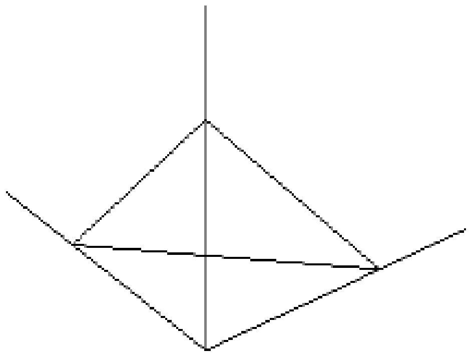
Polyhedral cones are either of the following:

(i) the set $\{0\}$; (ii) the whole space $\mathbb{R}^n$; or

(iii) positive orthant $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x \geq 0\}$.

Optimization problems where the cones $\mathcal{C}_1$ and $\mathcal{C}_2$ are either of these polyhedral cones are *linear optimization (LO) problems.*

Their duals are LO problems as well



The nonnegative orthant R^3

Significance

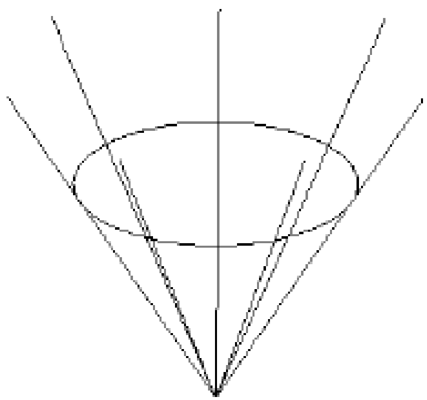
Huge number of applications, incl. trust design, transportation, planning.

Huge problems can be solved efficiently, even on PC, by using modern IPM software.

The Second order or Ice cream cone

$$\mathcal{S}_2^n := \left\{ x \in \mathbb{R}^n \, : \, \sqrt{\sum_{i=1}^{n-1} x_i^2} \le x_n \right\}.$$

"Ice cream cone" is coming from the 3D shape of the cone.



The ice-cream cone L^3

The second order cone is self-dual:

$$(\mathcal{S}_2^n)^* = \mathcal{S}_2^n.$$

Optimization problems where the cones $\mathcal{C}_1$ and $\mathcal{C}_2$ are either polyhedral or second order cones are *second order cone optimization(SOCO)* problems.

Significance

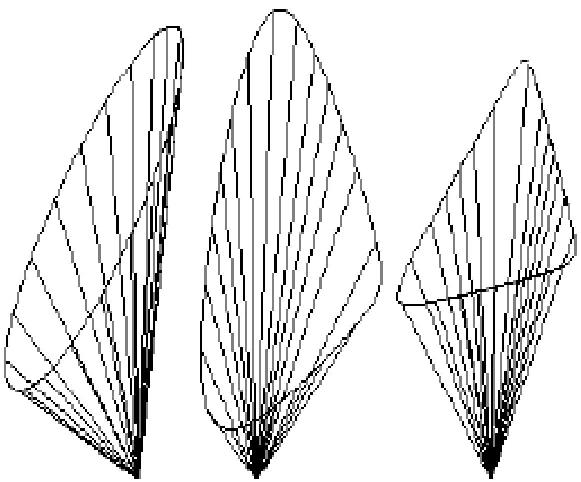
Norm and distance minimization, Tschebyshev approximation, robust optimization.

The Semidefinite cone

The semidefinite cone in $\mathbb{R}^{n\times n}$ is defined as

$$\mathcal{S}^n := \left\{X \in \mathbb{R}^{n\times n} : X = X^T, z^T X z \geq 0 \, \forall z \in \mathbb{R}^n\right\}$$

i.e. the matrices X are symmetric and positive semidefinite, denoted as $X \succeq 0$.



3 random 3D cross-sections of $\mathbb{S}^4_+$

Optimization problems where the cones $\mathcal{C}_1$ and $\mathcal{C}_2$ are either polyhedral, second order or semidefinite cones are called *semidefinite optimization (SDO) problems*.

The semidefinite cone is self-dual: $(\mathcal{S}^n)^* = \mathcal{S}^n$.

The SDO optimization problem

Let A_i, $i = 1, \cdots, n$ and C, X be $n \times n$ symmetric matrices, $b, y \in \mathbb{R}^m$ and let $\mathrm{Tr}(\cdot)$ denote the trace of a matrix.

The primal-dual SDO problem is defined as

$$
\begin{array}{llll}
(SP) & \min & \mathrm{Tr}(CX) & \\
 & \text{s.t.} & \mathrm{Tr}(A_i X) - b_i & \geq 0, \ \forall i \\
 & & X & \succeq 0
\end{array}
\qquad
\begin{array}{llll}
(SD) & \max & b^T y & \\
 & \text{s.t.} & C - \sum_{i=1}^{m} A_i y_i & \succeq 0 \\
 & & y & \geq 0.
\end{array}
$$

Significance

Robust optimization, trust design, Linear Matrix Inequalities (LMI),
Eigenvalue/singular-value optimization, sensor networks,
Convex relaxation of nonconvex/integer problems,...
(and many more)

Robust Linear Optimization

Classic – Polyhedral (scenario) approach

$$(P) \quad \min \quad c^T x \qquad \text{s.t.} \quad a_j^T x - b_j \geq 0 \quad \forall j$$

Let (a_j, b_j) be uncertain, it is coming from a polyhedral set (e.g. convex combination of "scenario" data points):

$$\left\{ \begin{pmatrix} a_j \\ -b_j \end{pmatrix} = \sum_{i=1}^{n_j} \begin{pmatrix} a_j^i \\ -b_j^i \end{pmatrix} \lambda_j^i \;\middle|\; \sum_{i=1}^{n_j} \lambda_j^i = 1,\ \lambda_j^i \geq 0 \right\}$$

The inequality $a_j^T x \geq b_j$ must be true for all possible values of $(a_j^T, -b_j)$:

$$\left[\sum_{i=1}^{n_j} \begin{pmatrix} a_j^i \\ -b_j^i \end{pmatrix} \lambda_j^i \right]^T \begin{pmatrix} x \\ 1 \end{pmatrix} \geq 0 \quad \text{for all} \quad \sum_{i=1}^{n_j} \lambda_j^i = 1,\ \lambda_j^i \geq 0$$

Infinitely many constraints!

iff $[a_j^i]^T x - b_j^i \geq 0$ for $i = 1, \ldots, n_j$

Finally the problem stays linear as:

$$(RP) \quad \min \quad c^T x \qquad \text{s.t.} \quad [a_j^i]^T x - b_j^i \geq 0 \ \text{for} \ i = 1, \ldots, n_j \quad \forall j$$

Disadvantages: – Huge number of linear inequalities

– Polyhedral uncertainty set not realistic.

Robust Linear Optimization

$$(P) \quad \min \quad c^T x$$
$$\text{s.t.} \quad a_j^T x - b_j \geq 0 \qquad \forall\, j$$

Let (a_j, b_j) be uncertain, it is coming from an ellipsoid (e.g. level set of a distribution):

$$\left\{ \begin{pmatrix} a_j \\ -b_j \end{pmatrix} = \begin{pmatrix} a_j^0 \\ -b_j^0 \end{pmatrix} + Pu \;\middle|\; u \in \mathbb{R}^k,\ u^T u \leq 1 \right\}$$

The inequality $a_j^T x \geq b_j$ must be true for all possible values of $(a_j^T, -b_j)$:

$$\left[\begin{pmatrix} a_j^0 \\ -b_j^0 \end{pmatrix} + Pu \right]^T \begin{pmatrix} x \\ 1 \end{pmatrix} \geq 0 \quad \forall u : u^T u \leq 1 \quad \textbf{iff} \quad [a_j^0]^T x - b_j^0 + \min_{u^T u \leq 1} \left\{ (Pu)^T \begin{pmatrix} x \\ 1 \end{pmatrix} \right\} \geq 0$$

$$[a_j^0]^T x - b_j^0 - \left\| P^T \begin{pmatrix} x \\ 1 \end{pmatrix} \right\|_2 \geq 0$$

This is a nondifferentiable norm constraint: **(See second order cones)**

$$\left\| P^T \begin{pmatrix} x \\ 1 \end{pmatrix} \right\|_2 \leq [a_j^0]^T x - b_j^0.$$

Single nonlinear, norm-constraint!

Eigenvalue Optimization

Given $n \times n$ symmetric matrices $A_1, \ldots, A_m$.

Problem: Find a nonnegative combination of the matrices that has the maximal smallest eigenvalue.

Solution: $$\max \left\{ \lambda \;\middle|\; \begin{array}{r} \sum_{i=1}^{m} A_i y_i - \lambda I \ \text{ is positive semidefinite} \\ y_i \geq 0 \ \ i = 1, \ldots, m \end{array} \right\}$$

Problem: Find a nonnegative combination of the matrices that has the smallest maximal eigenvalue.

Solution: $$\min \left\{ \lambda \;\middle|\; \begin{array}{r} \lambda I - \sum_{i=1}^{m} A_i y_i \ \text{ is positive semidefinite} \\ y_i \geq 0 \ \ i = 1, \ldots, m \end{array} \right\}$$

The semidefiniteness constraint is not differentiable, not easy to calculate when formulated by explicit functions, e.g., min-eigenvalue, determinant (of minors). **See Semidefinite Optimization.**

Solvability of CLO problems – Use IPMs

Classic Linear Optimization

Large scale LO problems are solved efficiently.
High performance packages (CPLEX, XPRESS-MP, MOSEK, GuRoBi, PcX, LIPSOL, CLP) offer simplex and interior point solvers.
Problems solved with 10^8 variables.
All commercials solvers offer powerful mixed integer engines.

SOCO and SDO

Polynomial solvability established.
Traditional software is unable to handle conic constraints.
Specialized software is developed. (SeDuMi, SDPT3, CSDP, DSDP, SDPpack, SDPA, SDPHA, MOSEK etc.)
SOCO: Problems solved with 10^6 variables.
SDO: solved with 10^4 dimensional matrices.

Solvability of CLO problems – Use IPMs

Developments on Modeling Languages

For conic and convex optimization problems.

Convex problems can be solved more efficiently than NLOs.

Novel Convex Modeling Environments. (YALMIP, CVX CERR)

SDO, SOCO, eigenvalue, determinant, all known convex functions – convex calculus.

IPMs for General Nonlinear Problems

Polynomial solvability of convex problems, including

Geometric, entropy and ℓ_p-norm programming.

Implementations for non-convex problems as well.

Specialized software is developed. (IPOPT, IPTOPT-C, KNITRO, MOSEK, LOQO, PENNON, etc.)

Problems solved with 10^4 dimensional matrices.

Mixed Integer Nonlinear Optimization Problems

BARON, COUENNE, BonMin

Disjunctive Conic Cuts for Mixed Integer Second Order Cone Optimization

Tamás Terlaky, Dept. ISE, Lehigh University, Bethlehem, PA

Joint work with Pietro Belotti (Clemson U.), Julio Goez (Lehigh U.), Ted Ralphs (Lehigh U.)

Mixed Integer Second order cone optimization

$$\begin{aligned} \text{minimize}: \quad & c^T x \\ \text{subject to}: \quad & Ax = b \\ & x \in \mathrm{K} \\ & x \in \mathrm{Z}^d \times \mathrm{R}^{n-d} \end{aligned}$$

- $A \in \mathrm{R}^{m\times n}, c \in \mathrm{R}^n, b \in \mathrm{R}^m$
- $x = [(x^1)^T, (x^1)^T, \ldots, (x^1)^T]^T$
- $\mathrm{L}^{n_i} = \{x^i \mid x_1^i \geq \| x_{2:n}^i \|\}$, Lorentz cone
- $\mathrm{K} = \{\mathrm{L}_1^{n_1} \times \mathrm{L}_2^{n_2} \times \ldots \times \mathrm{L}_k^{n_k}\}$
- Rows of A are linearly independent

Proposition (Convex hull of the intersection of a disjunction and a convex set)

Consider a closed convex set E and two halfspaces $\mathrm{A} = \{x \in \mathrm{R}^n : a^T x \leq \alpha\}$, $\mathrm{B} = \{x \in \mathrm{R}^n : b^T x \leq \beta\}$ such that they do not intersect inside E, i.e., $\mathrm{E} \cap \mathrm{A} \cap \mathrm{B} = \phi$. Denote $\mathrm{A}^= = \{x \in \mathrm{R}^n : a^T x = \alpha\}$, $\mathrm{B}^= = \{x \in \mathrm{R}^n : b^T x \leq \beta\}$. If $A^=$ and $\mathrm{B}^=$ are bounded, and there exists a convex cone K such that $\mathrm{K} \cap \mathrm{A}^= = \mathrm{E} \cap \mathrm{A}^=$ and $\mathrm{K} \cap \mathrm{B}^= = \mathrm{E} \cap \mathrm{B}^=$, then $\mathrm{conv}(\mathrm{E} \cap (\mathrm{A} \cup \mathrm{B})) = \mathrm{E} \cap \mathrm{K}$.

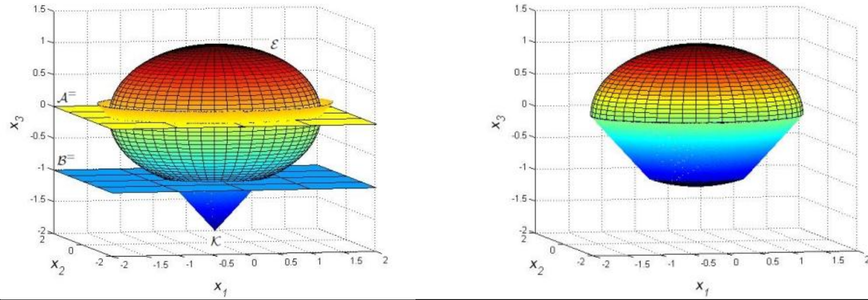
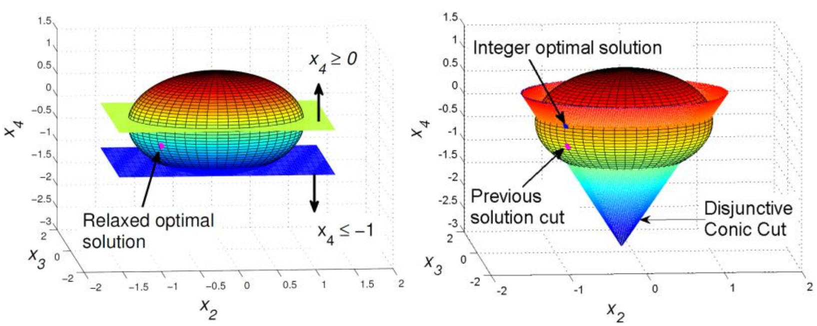


Illustration of the disjunctive conic cut procedure



Theorem (Uniparametric family of quadrics)

Let $\mathcal{Q} = (Q, q, \rho)$, where Q is a positive definite, $q \in \mathrm{R}^n$, $\rho \in \mathrm{R}$. The quadric given by $\mathcal{Q}$ is the set $\{x : x^T Q x + q^T x + \rho = 0\}$. Given two hyperplanes $\mathcal{A}_1 = \{x \in \mathrm{R}^n : a_1^T x \leq \alpha_1\}$, $\mathcal{A}_2 = \{x \in \mathrm{R}^n : a_2^T x \leq \alpha_2\}$, the family of quadrics having the same intersection with the two hyperplanes as the quadric given by $\mathcal{Q}$ is parametrized by $\tau \in \mathrm{R}$ as $\mathcal{Q}(\tau) = (Q(\tau); q(\tau); \rho(\tau))$, where

$$Q(t) = Q + \tau c\left(a_1 a_2^T + a_2 a_1^T\right), \quad q(\tau) = q - \tau c\left(\alpha_1 a_2^T + \alpha_2 a_1^T\right), \quad \rho(\tau) = \rho + 2\tau c \alpha_1 \alpha_2,$$

with $c = 1/\left(2a_1^T a_2\right)$ if $a_1^T a_2 \neq 0$, and $c = 1$ if $a_1^T a_2 = 0$.

Let $\hat{\tau}$ be the larger root of equation $q(\tau) > q(\tau)^T Q(\tau) q(\tau) - \rho(\tau) = 0$. The quadric generated by $\mathrm{Q} = (Q(\tau), q(\tau), \rho(\tau))$ gives the disjunctive conic cut.

Conclusions

- We developed new disjunctive conic cuts for MISOCO.
- It is algebraically simple to find the disjunctive conic cut for MISOCO problems.

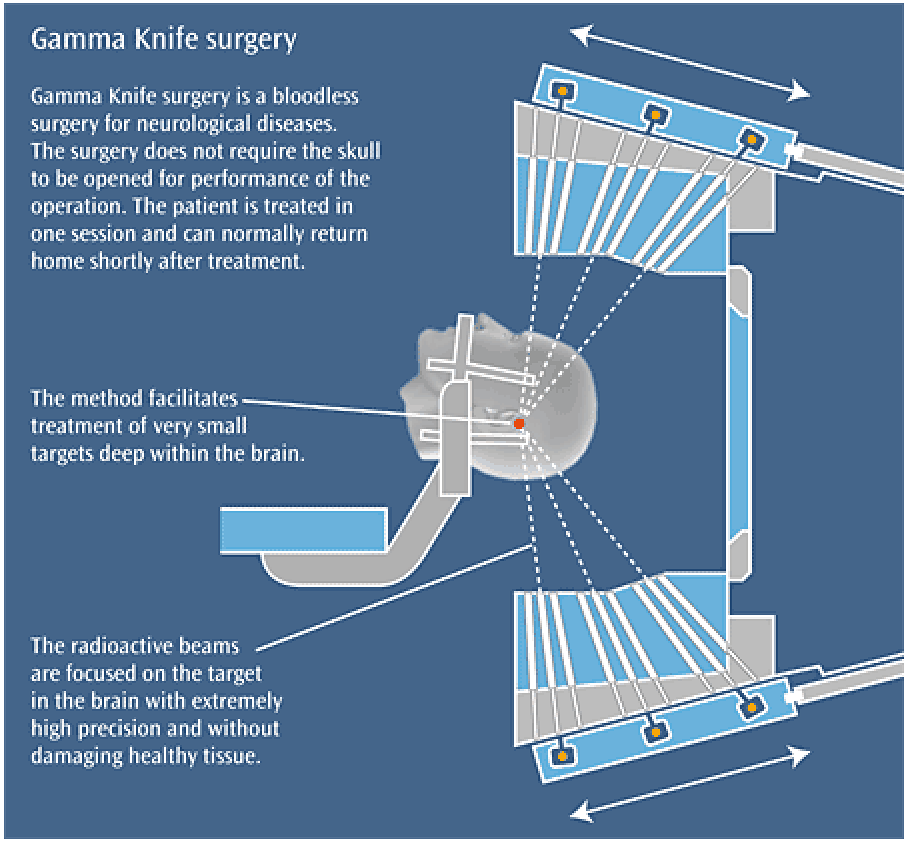
A Radiation Therapy Treatment Application

Joint work with

Mohammad R. Oskoorouchi (California State University San Marcos)

Dionne Aleman, Hamid R. Ghaffari (University of Toronto)

Gamma knife surgery



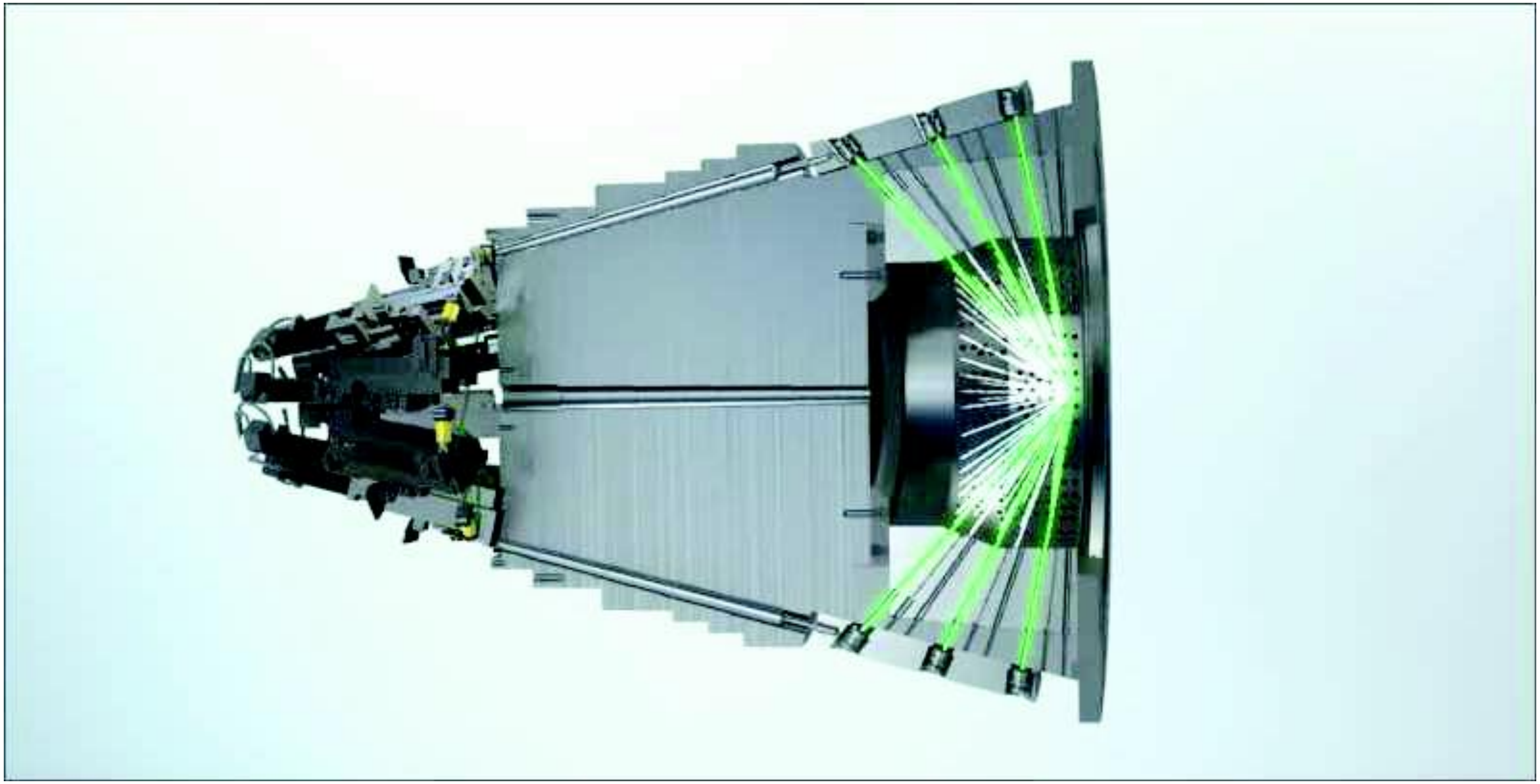
Gamma knife Perfexion unit

Images courtesy of http://www.elekta.com

Gamma knife treatment planning

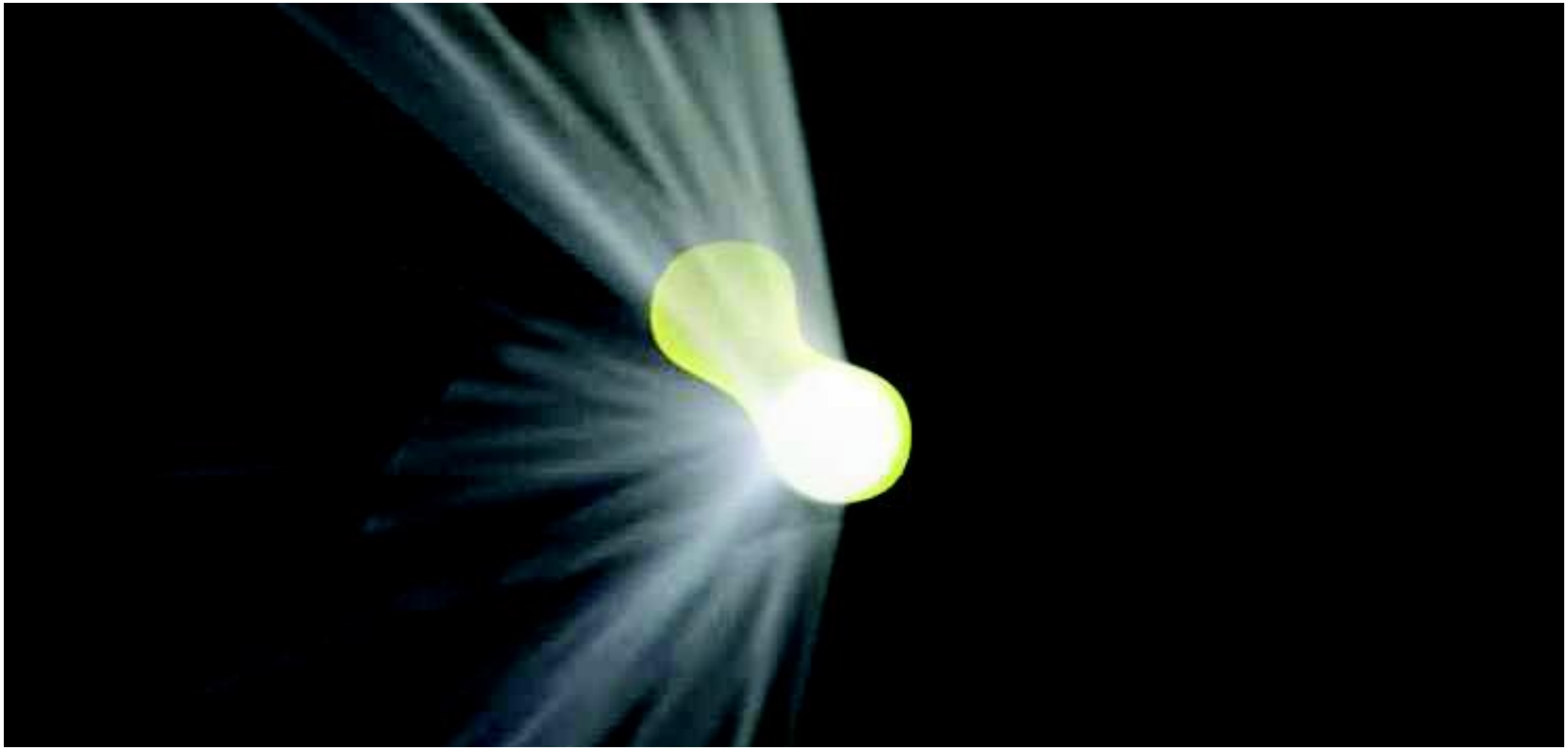
- Each target structure should receive a minimum prescription dose
- Critical structures should not be over dosed
- Target structure should not be over dosed either

Dose delivery



Images courtesy of http://www.elekta.com

Treatment



Images courtesy of http://www.elekta.com

Sector Duration Optimization

(Ghaffari, Aleman, Ruscin, Jaffry 2009)

$$\min \sum_{s\in S}\sum_{j=1}^{v_s} F_s(z_{js})$$

s.t.

$$z_{js} = \sum_{I\in\Theta}\sum_{b\in B}\sum_{c\in C} D_{Ibcjs} t_{Ibc} \qquad s\in S,\ j=1,\ldots,v_s$$

$$t_{Ibc} \geq 0, \qquad I\in\Theta, b\in B, c\in C$$

where,

$$F_s(z_{js}) = \begin{cases} \dfrac{\overline{w}_s}{v_s}\left(z_{js}-T_s\right)^2 & z_{js}\geq T_s \\ \dfrac{\underline{w}_s}{v_s}\left(T_s-z_{js}\right)^2 & z_{js}< T_s \end{cases}$$

Semi-infinite Linear Formulation

$$\min \sum_{j=1}^{v_s} \sum_{s \in S} \delta_{js}$$

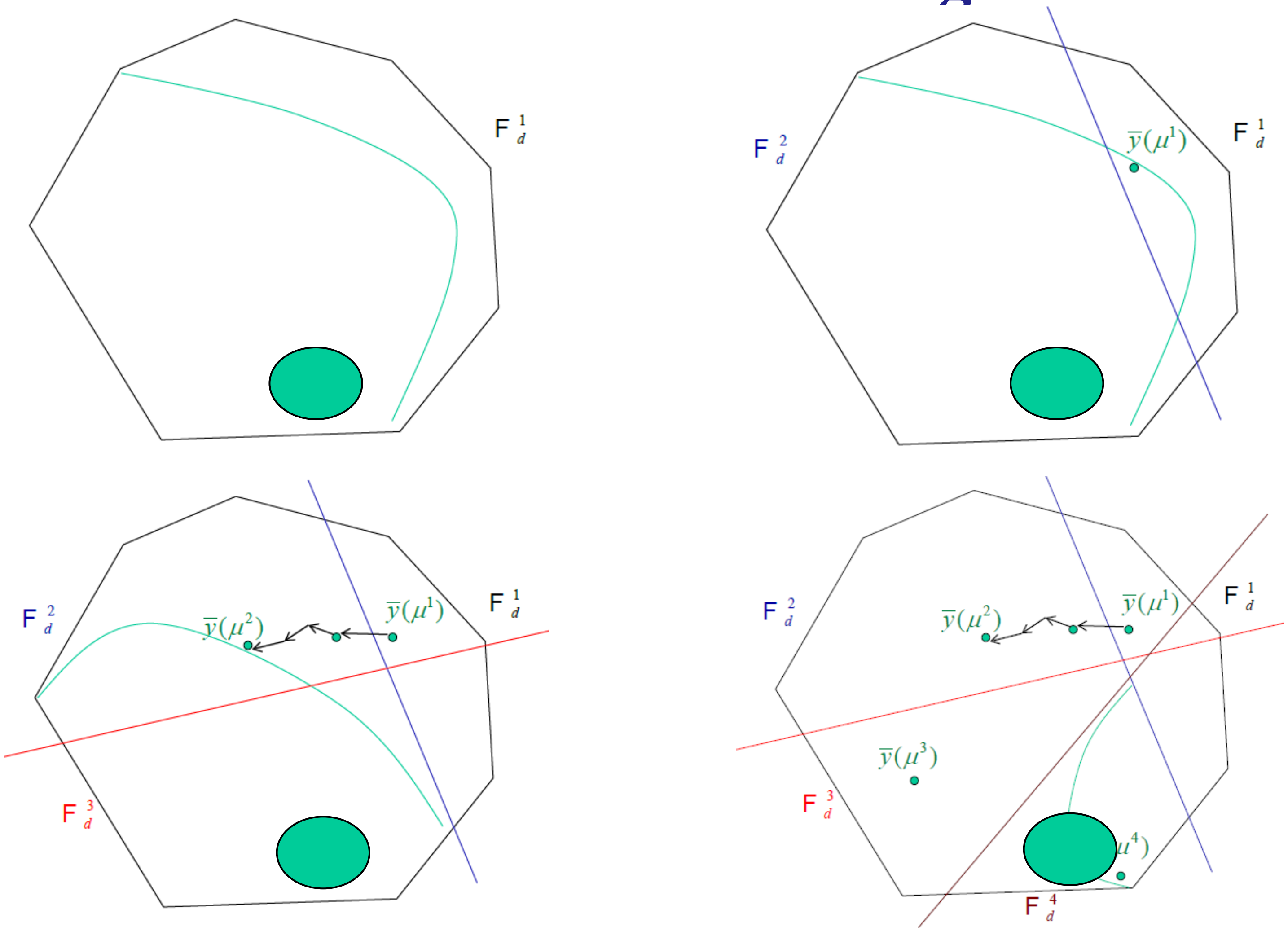
s.t.

$$\begin{cases} \dfrac{\overline{w}_s}{v_s}\left(z_{js} - T_s\right)^2 \le \delta_{js}, & z_{js} \ge T_s \\ \dfrac{\underline{w}_s}{v_s}\left(T_s - z_{js}\right)^2 \le \delta_{js}, & z_{js} < T_s \end{cases} \quad s \in S, \forall j \text{ convex set}$$

$$z_{js} = \sum_{I \in \Theta} \sum_{b \in B} \sum_{c \in C} D_{Ibcjs} t_{Ibc} \qquad s \in S,\ j = 1, \ldots, v_s$$

$$t_{Ibc} \ge 0, \qquad I \in \Theta, b \in B, c \in C$$

Semi-infinite Optimization:
Interior Point Column Generation Algorithm



SOCO Formulation

$$\min \sum_{j=1}^{v_s} \sum_{s\in S} \delta_{js}$$

s.t.

$$-\bar{y}_{js} \le z_{js} - T_s \le \underline{y}_{js} \qquad s\in S,\ j=1,\dots,v_s$$

$$\frac{1}{v_s}\left(\bar{w}_s \bar{y}_{js}^2 + \underline{w}_s \underline{y}_{js}^2\right) \le \delta_{js}, \qquad s\in S,\ j=1,\dots,v_s$$

$$\bar{y}_{js} \ge 0,\ \ \underline{y}_{js} \ge 0 \qquad s\in S,\ j=1,\dots,v_s$$

$$z_{js} = \sum_{I\in\Theta}\sum_{b\in B}\sum_{c\in C} D_{Ibcjs} t_{Ibc} \qquad s\in S,\ j=1,\dots,v_s$$

$$t_{Ibc} \ge 0, \qquad I\in\Theta, b\in B, c\in C$$

SOCO Formulation

$$\min \sum_{s\in S} \delta_s$$

s.t.

$$-\bar{y}_{js} \le z_{js} - T_s \le \underline{y}_{js} \qquad s\in S,\ j=1,\dots,v_s$$

$$\frac{1}{v_s}\sum_{j=1}^{v_s}\left(\bar{w}_s \bar{y}_{js}^2 + \underline{w}_s \underline{y}_{js}^2\right) \le \delta_s, \qquad s\in S$$

$$\bar{y}_{js} \ge 0,\ \ \underline{y}_{js} \ge 0 \qquad s\in S,\ j=1,\dots,v_s$$

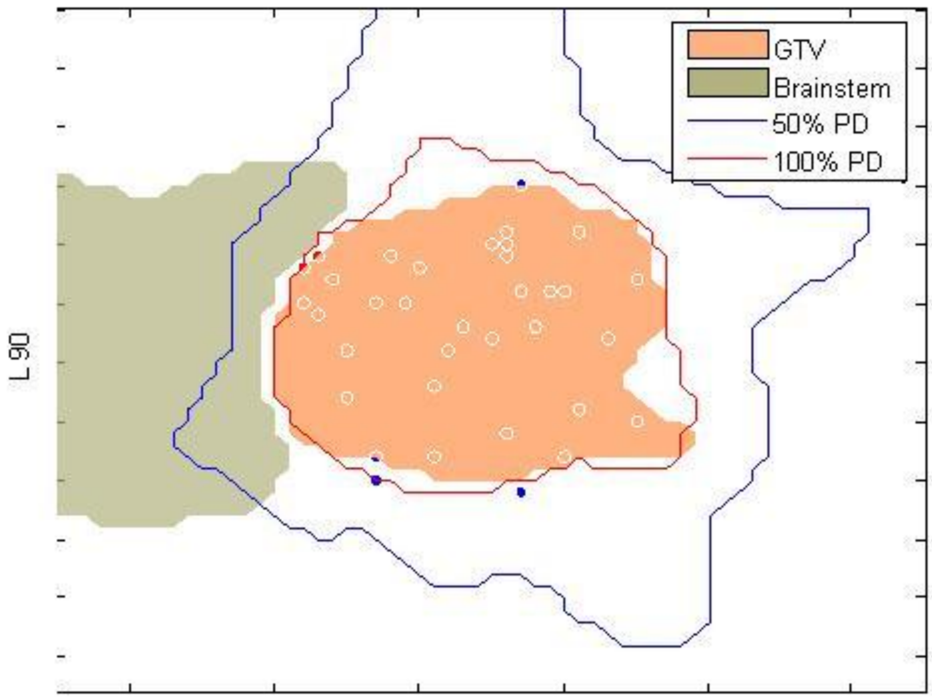
$$z_{js} = \sum_{I\in\Theta}\sum_{b\in B}\sum_{c\in C} D_{Ibcjs} t_{Ibc} \qquad s\in S,\ j=1,\dots,v_s$$

$$t_{Ibc} \ge 0, \qquad I\in\Theta, b\in B, c\in C$$

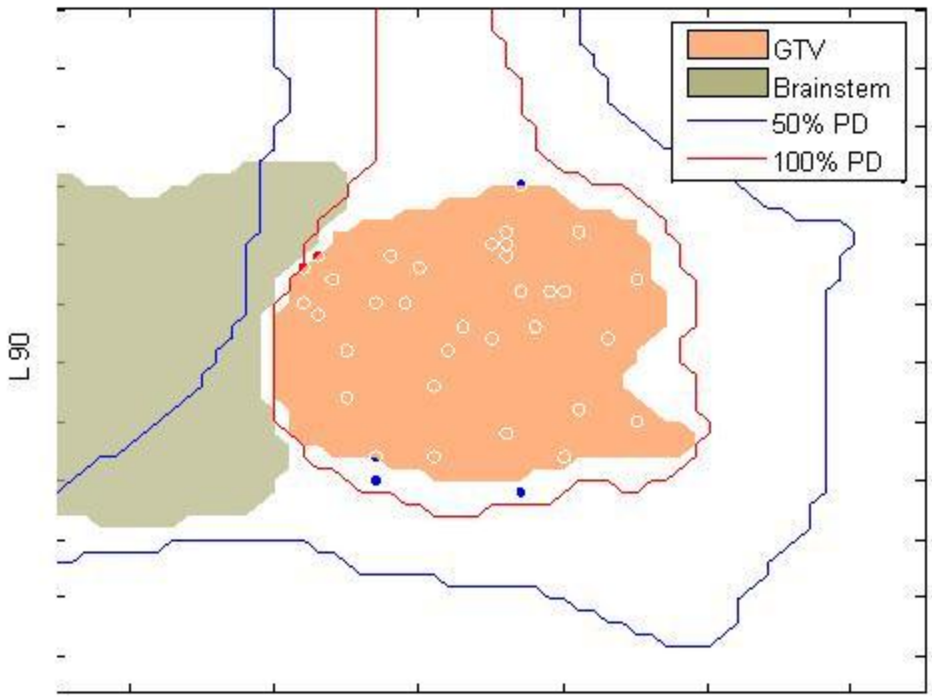
Computational Results

	Problem Dimension			Objective Value			CPU Time (min)		
nIso	**PRG**	**IPCG**	**MOSEK**	**PRG**	**IPCG**	**MOSEK**	**PRG**	**IPCG**	**MOSEK**
10	240	480+141	99336	57.06	56.84	56.88	60.36	1.19	2.85
15	360	720+160	99456	58.08	57.92	59.96	79.28	2.02	4.74
20	480	960+189	99576	53.77	53.70	56.74	148.41	4.22	9.18
25	600	1200+181	99696	43.57	43.45	48.52	349.13	5.22	10.01
30	720	1440+188	99816	41.89	41.44	45.54	466.71	7.56	12.36
35	840	1680+187	99936	40.48	39.35	39.52	877.01	11.14	22.57
45	1080	2160+199	100176	40.71	39.78	39.96	1520.25	21.15	29.47
55	1320	2640+200	100416	35.61	35.22	35.53	3053.94	31.22	50.89
65	1560	3120+206	100656	37.78	36.22	--	4051.90	46.37	--
105	2520	5040+209	101616	--	33.43	--	9999.99	129.22	--

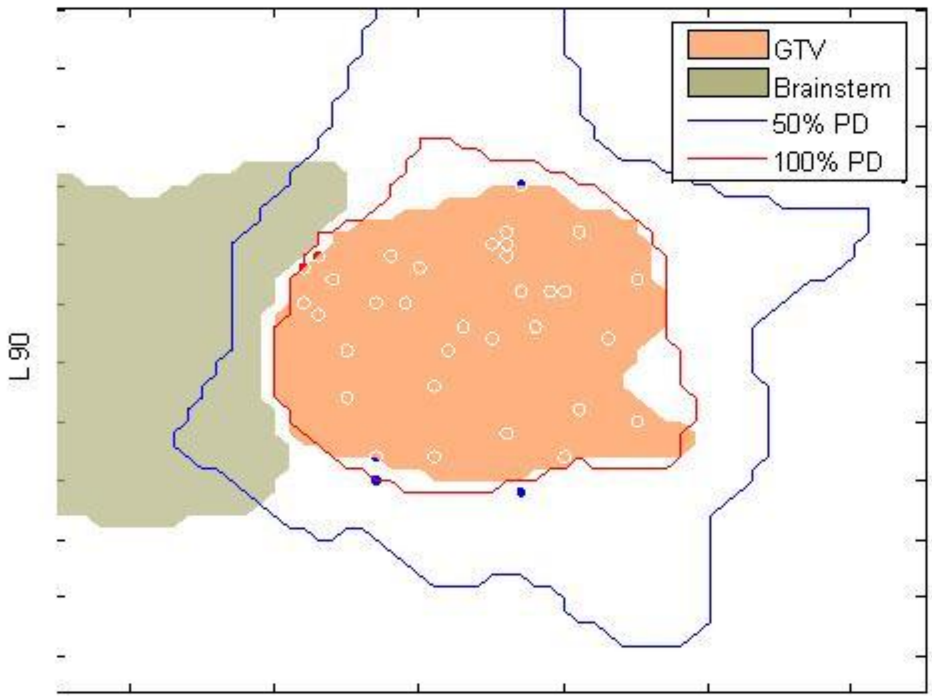
Brain image (projected gradient)



Brain image (MOSEK)



Brain image (IPCG)



Summary

- Radiation Therapy Treatment (IMRT, IGRT, PTT...) is a rich area of novel optimization problems, new technologies bring new problems
- Many modeling, algorithmic and software options, so choose the best model, algorithm, software

Case Study

- Analyzed models for an optimal Perfexion treatment plan having fixed isocenter locations
- Compared algorithms for the models: IPCG outperforms PCG classical IPMs (MOSEK)

Thanks

Questions?

Acknowledgments

- Lehigh University
- Elekta (Stockholm, Sweden)
- NSERC Discovery Grants of Canada
- The Canadian Foundation for Innovation
- Princess Margaret Hospital (Toronto, Canada)