Optimization in Medical Research

Shuo Yang, Chuang Miao, Jie Liu,
Timothy Thomas

4/11/2013

Steepest Descent method

- For the optimization problem: $\min_{x \in R^n} f(x)$ solve for the local minimizer x_*.

1. initial guess $x_0 \approx x_*$;

For k=0,1,2,…until stop criteria satisfied:

2. calculate the search direction $d_k = -g_k = -\nabla f(x_k)$;

3. perform line search $\min_{\alpha_k > 0} f(x_k + \alpha_k d_k)$;

4. update $x_{k+1} = x_k + \alpha_k p_k$.

Steepest Descent method

Two aspects concerned:

1. How to choose the step size?
2. What conditions should be satisfied for iteration convergence?

How to determine the step size?

Option 1: Solve for the global minimizer of the 1D function $\min_{\alpha>0} f(x+\alpha p)$ (usually computationally expensive)

Option 2: Inexact line search: find an interval containing suitable step size, then refine the step size within the interval.

Wolfe conditions: α should satisfy

(1) $f(x+\alpha p) \leq f(x) + c_1 \alpha f'(x)$

(2) $f'(x+\alpha p) \geq c_2 f'(x)$ $\qquad (0 < c_1 < c_2 < 1)$

selecting c_1 and c_2 produces an interval which brackets the optimal step size.

“Backtracking” Algorithm for Line Search

Algorithm (Backtracking Line Search)

begin
  Choose $\bar{\alpha} > 0$ and $\rho, c \in (0, 1)$
  $\alpha := \bar{\alpha}$
  while $\phi(\alpha) \geq \phi(0) + c\alpha\phi'(0)$ do
    $\alpha := \rho\alpha$
  end
  $\alpha_k := \alpha$
end

Option 3: choose α empirically (often a small constant), as shown in the following applications.

Convergence of the iteration

- Objective function f should satisfy:

f is bounded;

the gradient $\nabla f(x)$ *is Lipschitz continuous (exist a constant L, the gradient is bounded between –L and L);*

- The line search α_k satisfies the Wolfe conditions.

The "zig-zag" Behavior

Solve for the exact line length:

$$\frac{d}{d\alpha} f(x_k - \alpha g_k) = 0 \qquad\qquad (g = \nabla f)$$

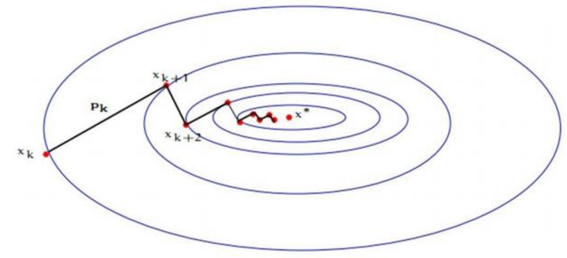


Figure1. Typical steepest descent steps

- $g_k^T \mathrm{g}(x_k - \alpha g_k) = 0$
- $g_k^T g_{k+1} = 0$ - the successive search directions are perpendicular.

When the line length α is close to the exact one, usually we have $g_k^T g_{k+1} \approx 0$

Applications of the Gradient Descent Method

1. Computational health informatics

(nonlinear approach for Bayesian network parameter learning)

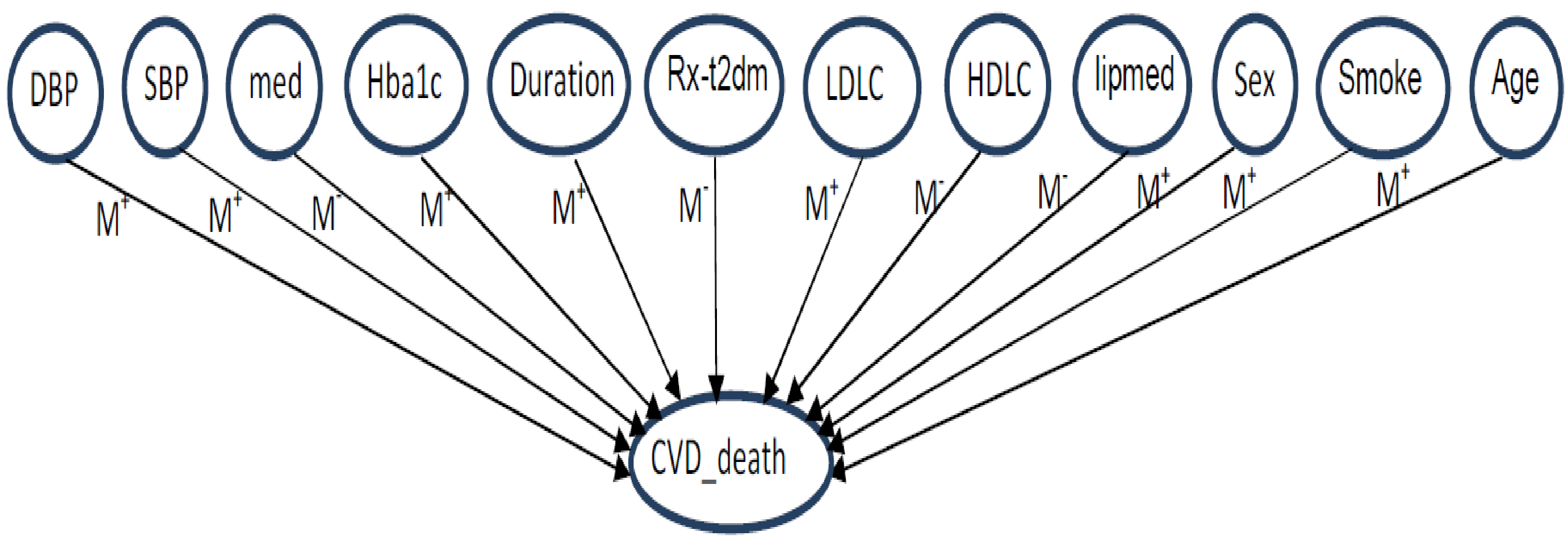
2. CT imaging

(iterative method for image recon for a linear system)

Uncertainty in Medical Knowledge

- 1. Observational: How certain am I that I saw what I thought I saw? (P(observation))
- 2. Occurrence: How likely is it that the event happened? (P(event))
- 3. Causal: How influential are different preconditions in making the event happen? (P(event|cond1) vs P(event|cond2))
- 4. Temporal: How does the chance of true occurrence change over time?
- 5. Model: Have I modeled everything properly?

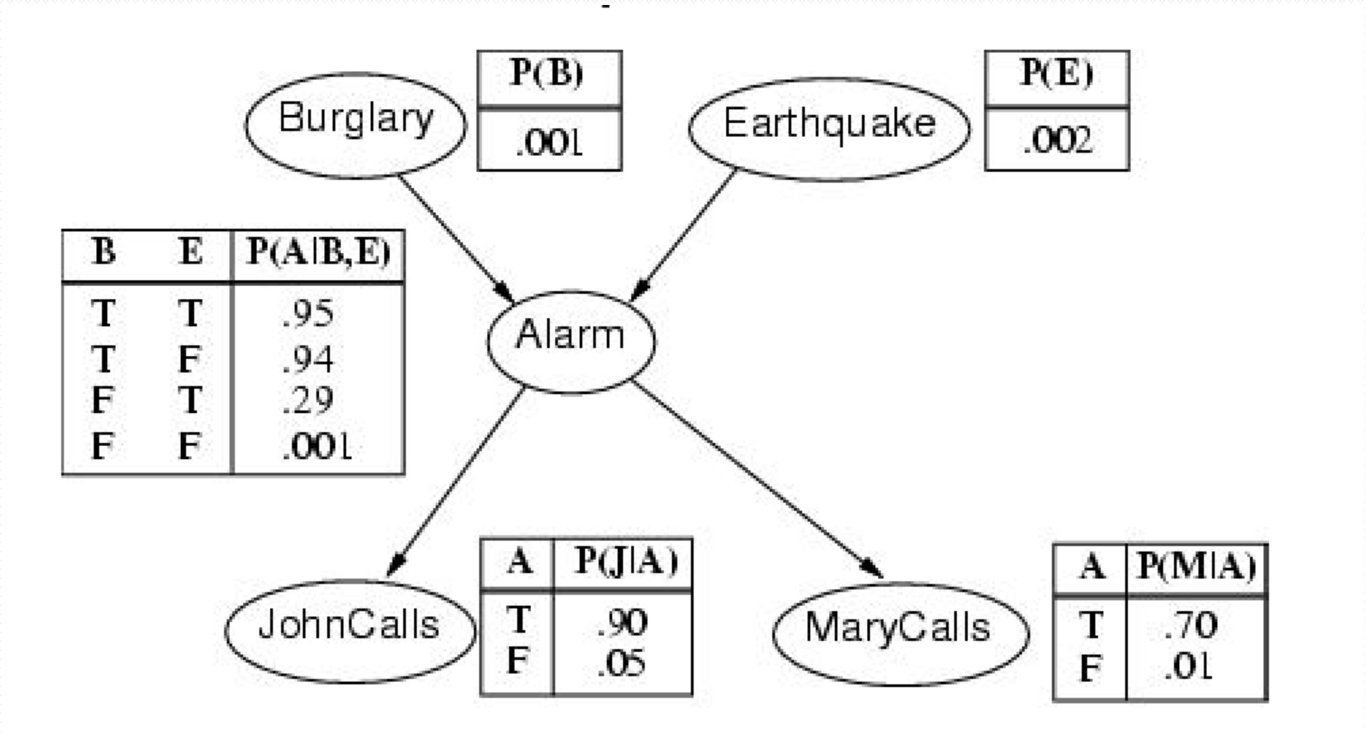
Example



What is the probability of deaths from cardiovascular disease based on these patient's physiological indexes?

Bayesian Net can answer such queries and capture the probability of these uncertainties based on the observations.

Bayesian Network

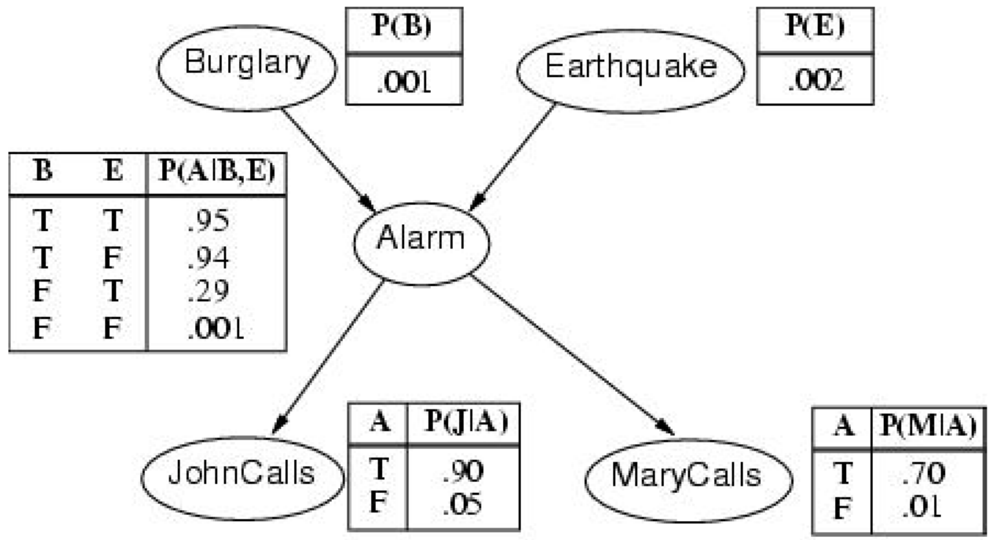


- ➢ Each node corresponds to a random variable
- ➢ Directed arrows connects pairs of nodes
- ➢ Each node has a conditional probability distribution

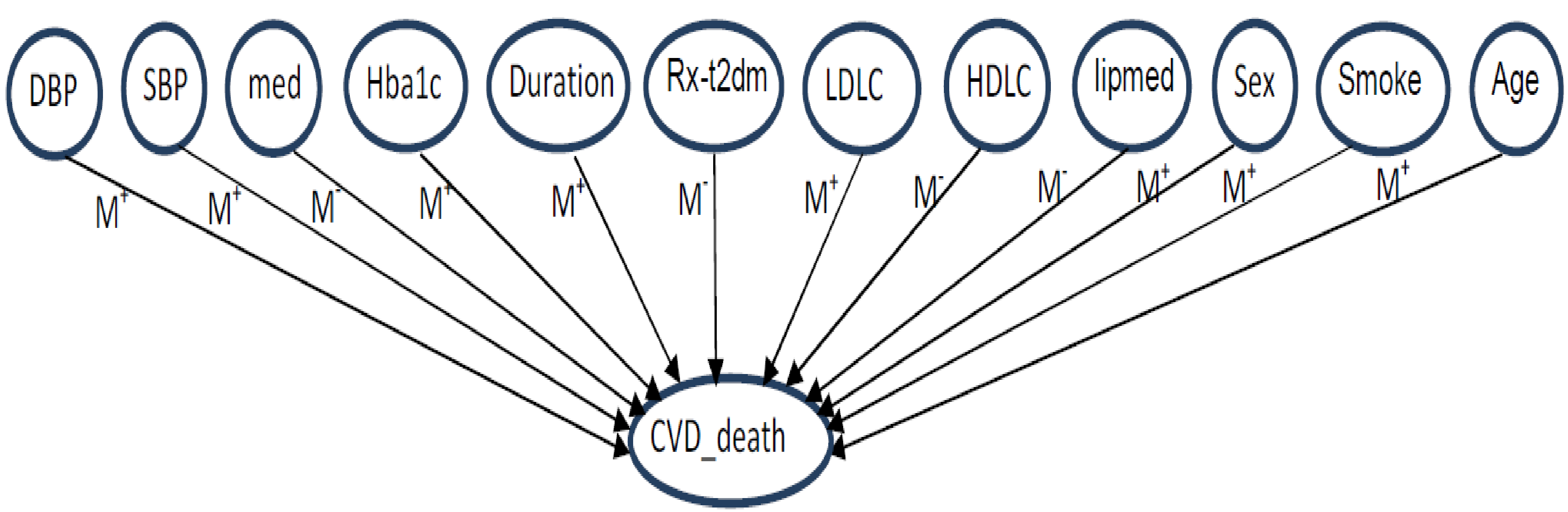
Objective Function

- Given the structure of a Bayesian network
- The optimized CPTs are learned by maximize the likelihood of the training data given the proposed model:

$$\hat{\boldsymbol{\theta}} = \boldsymbol{argmax}_{\theta \in \Theta^Q} \prod_{i=1}^{n} \prod_{j=1}^{v_i} \prod_{k=1}^{r_i} \theta_{ijk}^{N_{ijk}}$$



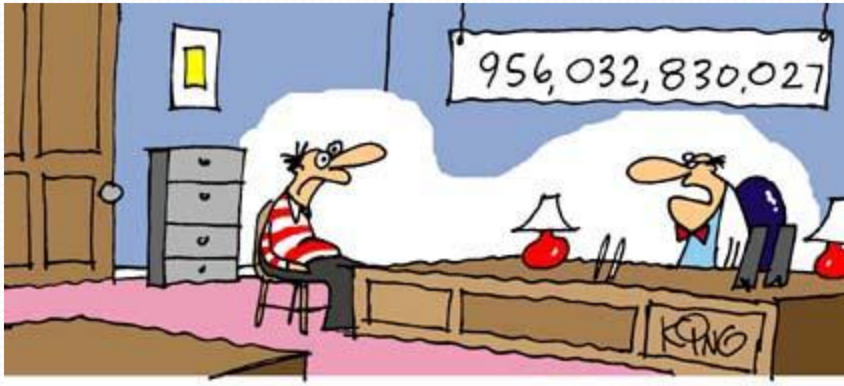
Why Need Qualitative Constraints?



- Number of different configurations of the features: $2^{12} = 4096$
- Only 100 records of CVD death

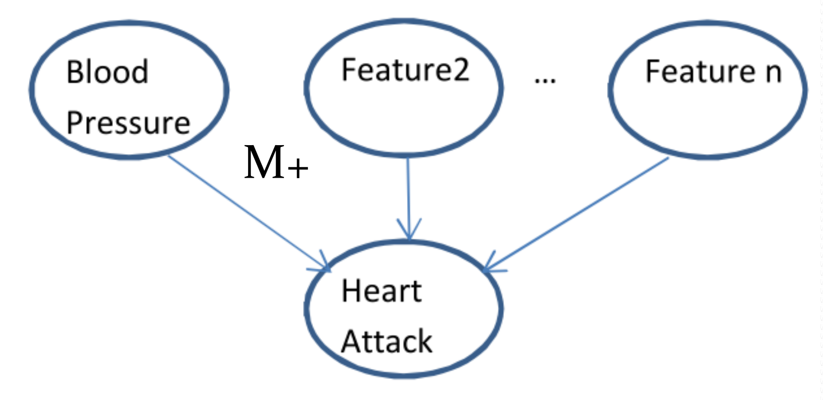
Monotonicity

- Increase in Cholesterol level increases the risk of heart attack
- Increase in global temperature increases the risk of sea level rise
- Hypothesis: Such domain knowledge can help improve the prediction accuracy beyond merely learning from data



"That number has nothing to do with the lottery or the stock market. That's your cholesterol level."

Monotonicity – Probabilistic Influence



- $P(HA = 1 \mid BP = high, C) \geq P(HA = 1 \mid BP = low, C)$

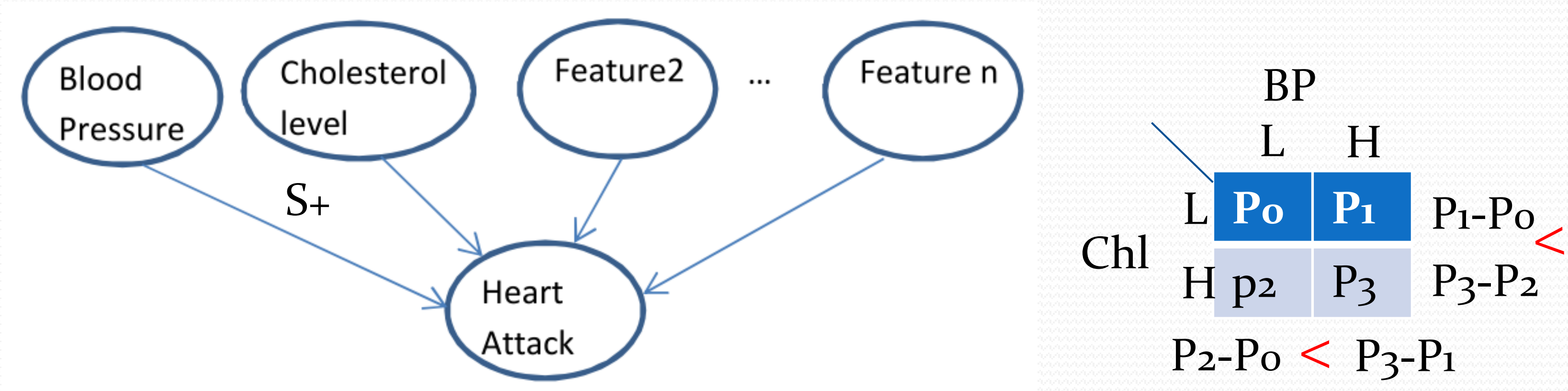
Where $C \in \times_{j \neq BP} X_j$ means all contexts (configurations of other parents)

Synergistic Interactions

- Increasing the cholesterol level increases the risk of heart attack more effectively at higher level of blood pressure than at lower levels

- Increasing the size of breast tumor increases the recurrence rate more efficiently at old age than at young age

Synergistic Interactions – Probabilistic Influence



- *Effect of Blood Pressure on Heart Attack*$= P(HA = 1 \mid BP = high, C) - P(HA = 1 \mid BP = low, C)$

- $P(HA = 1 \mid BP = high, Chl = high, \mathrm{C}) + P(HA = 1 \mid BP = low, Chl = low, C) \geq P(HA = 1 \mid BP = low, Chl = high, C) P(HA = 1 \mid BP = high, Chl = low, C)$
- Where $C \in \times_{j \neq BP, Chl} X_j$ means all contexts (configurations of other parents)

Penalty and Objective Functions

- The function of monotonic constraint is defined as:

$$\boldsymbol{\delta = P\left(X_i \le k_c \middle| Pa_i^{j_2}\right) - P\left(X_i \le k_c \middle| Pa_i^{j_1}\right) + \epsilon}$$

- The function of synergy constraint is defined as:

$$\boldsymbol{\delta = P\left(Y \le k_c \middle| X_1^i, X_2^j\right) + P\left(Y \le k_c \middle| X_1^{i+1}, X_2^{j+1}\right)}$$
$$\boldsymbol{-P\left(Y \le k_c \middle| X_1^{i+1}, X_2^j\right) - P\left(Y \le k_c \middle| X_1^i, X_2^{j+1}\right) + \epsilon}$$

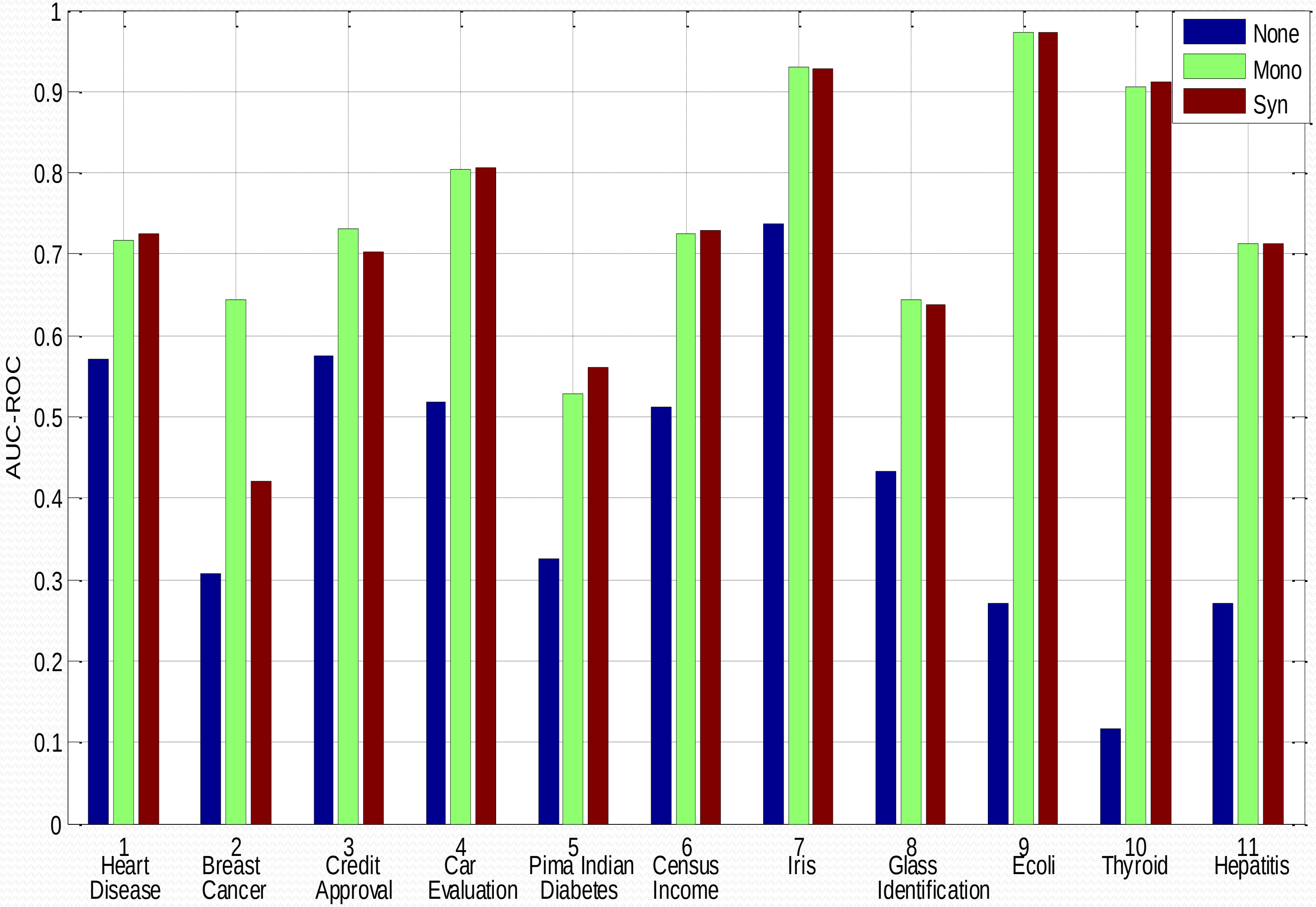
- The corresponding penalty function is $\boldsymbol{P_{j_1,j_2}^{i,k_c} = I_{(\delta>0)}\delta^2}$

  where I=1 when $\delta > 0$ and I=0 when $\delta \le 0$

- Objective function:

$$\boldsymbol{J(\mu_{ijk}) = J_L(\mu_{ijk}) - \omega \sum\nolimits_{j'} P_{j'}^{i,k_c}}$$

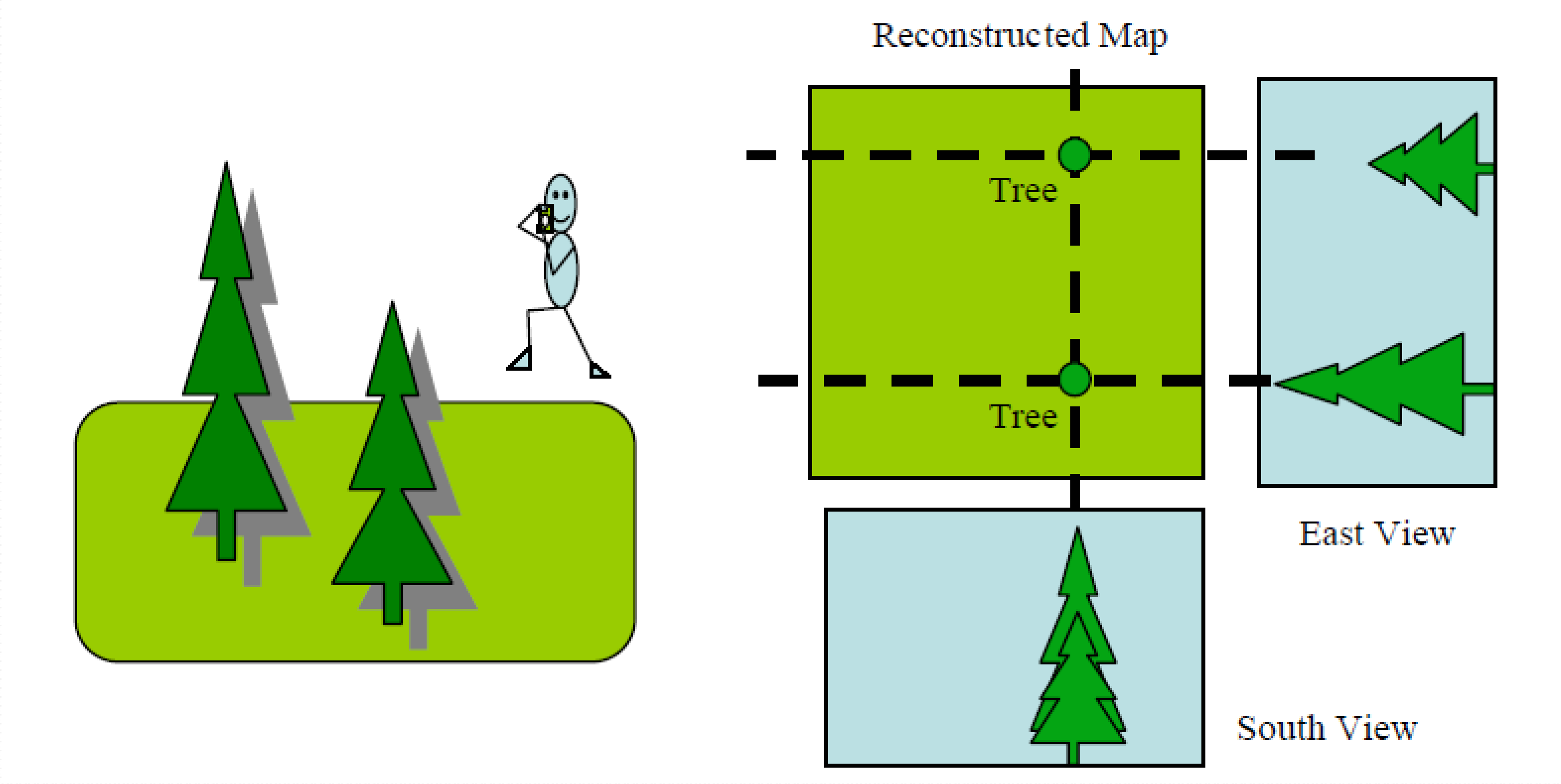
Experiments



Tomography

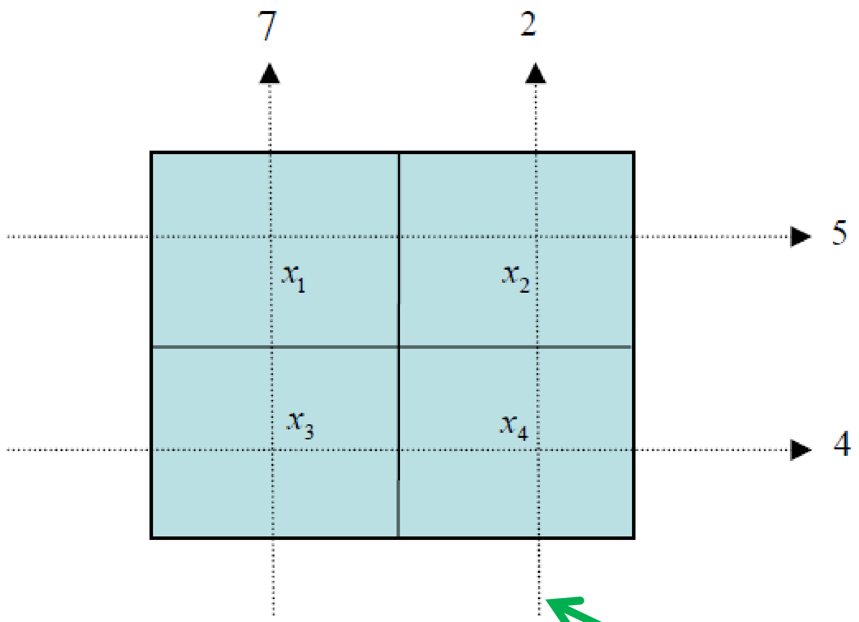
"***tomos***": Greek word, means a section, a slice or a cut.

"***Tomography***": is the process of imaging a cross section.



G. Zeng, "Medical Image Reconstruction: A Conceptual Tutorial", China Higher Education Press, Beijing, China 2009.

Computed Tomography (CT): a mathematical problem



$$\begin{cases} x_1 + x_2 = 5 \\ x_3 + x_4 = 4 \\ x_1 + x_3 = 7 \\ x_2 + x_4 = 2 \end{cases}$$

Ray sum, Line Integral or a Projection.

What if the problem gets more complicated, more trees, larger matrix? **More views!**

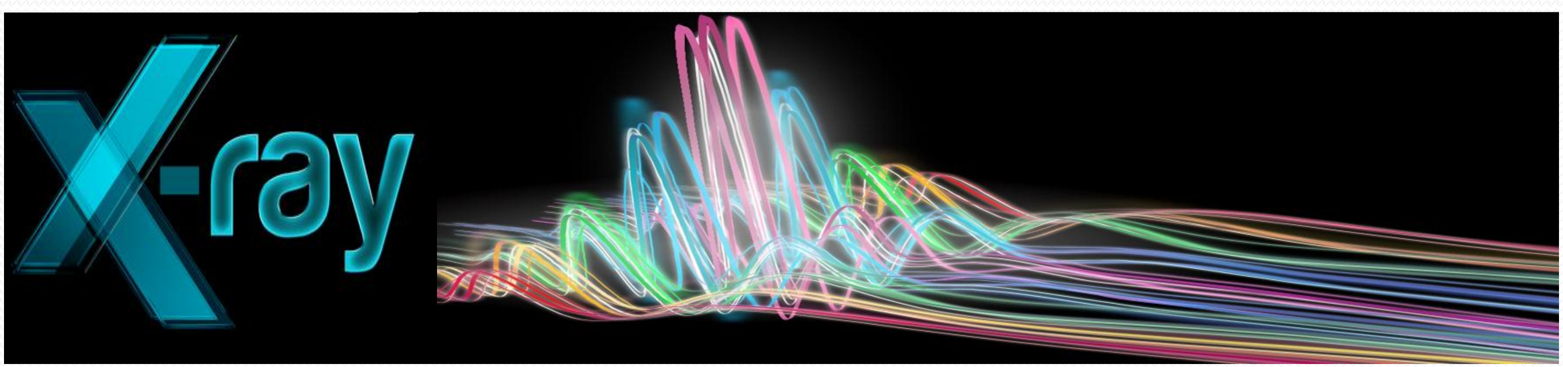
G. Zeng, "Medical Image Reconstruction: A Conceptual Tutorial", China Higher Education Press, Beijing, China 2009.

CT Basics: X-ray

X-ray is a special light with certain **wavelength**. Meanwhile, it can also be viewed as a sets of **photons**.

Under **ideal** imaging conditions

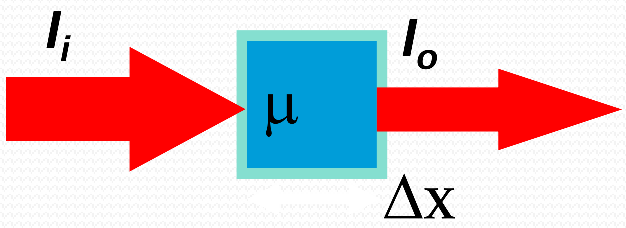
- X-ray energy is **monochromatic**
- There is **no scattering** in imaging process
- The imaging chains are **linear**



http://arstechnica.com/science/2012/06/turning-an-infrared-laser-into-an-x-ray-source/

Beer Law--Exponential Attenuation Model

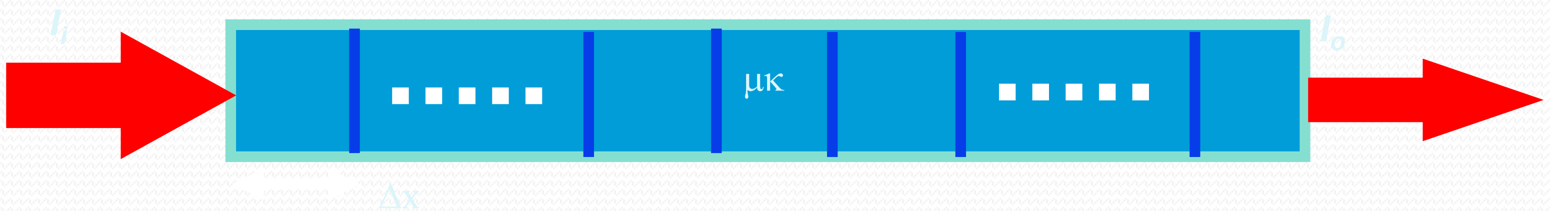
- The relation between x-ray and the object can be expressed as Beer's Law.



$$I_o = I_i e^{-\mu \Delta x}$$

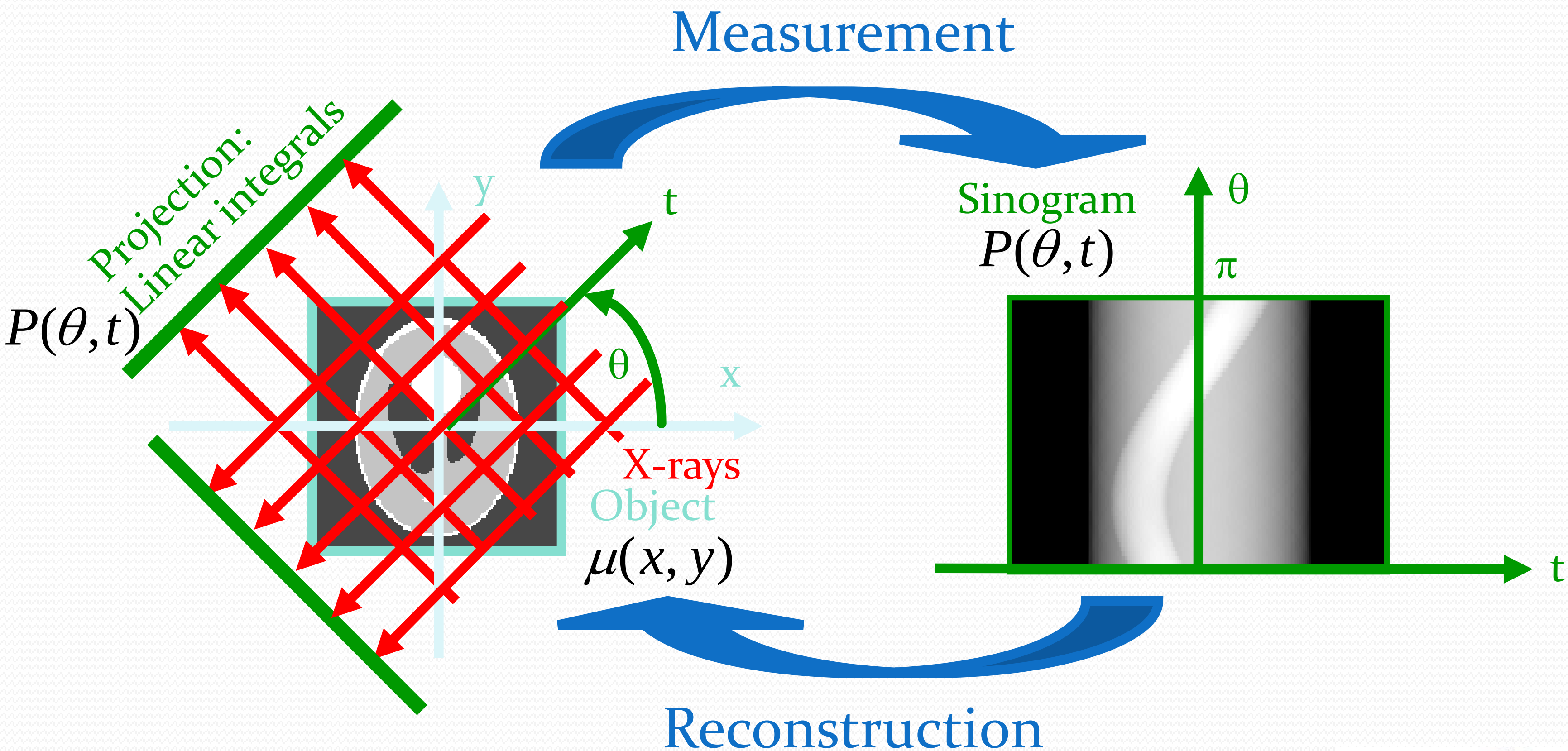
I_i: input intensity of X-ray
I_o: output intensity of X-ray
μ: linear X-ray attenuation

$$I_o = I_i e^{-\sum_k \mu_k \Delta x}$$



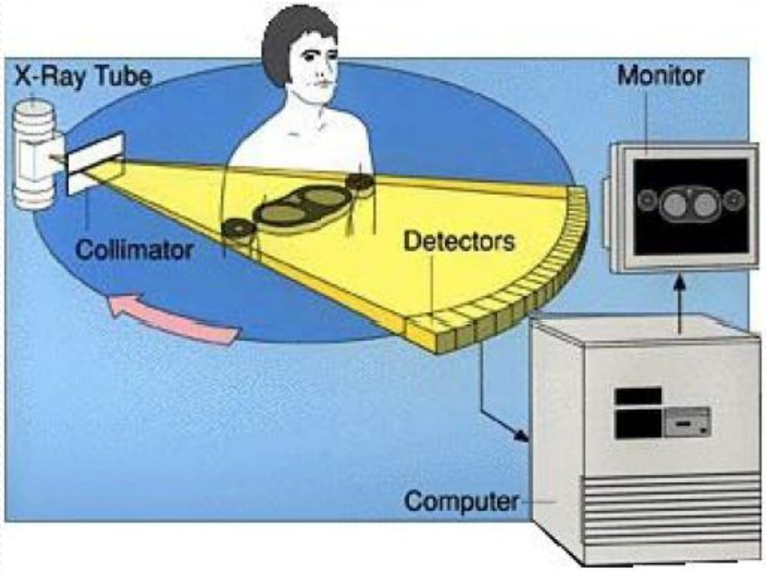
$$\boxed{\text{Ray sum: } \ln\left(\frac{I_i}{I_o}\right) = \lim_{\Delta x \to 0} \sum_k \mu_k \Delta x = \int_{-\infty}^{+\infty} \mu(x)dx}$$

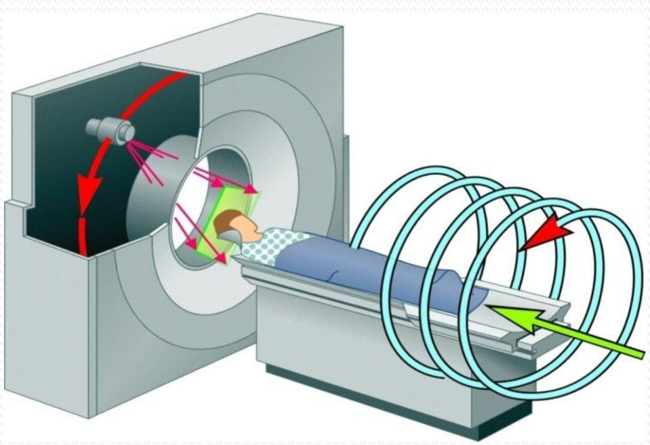
Randon Transform & Sinogram

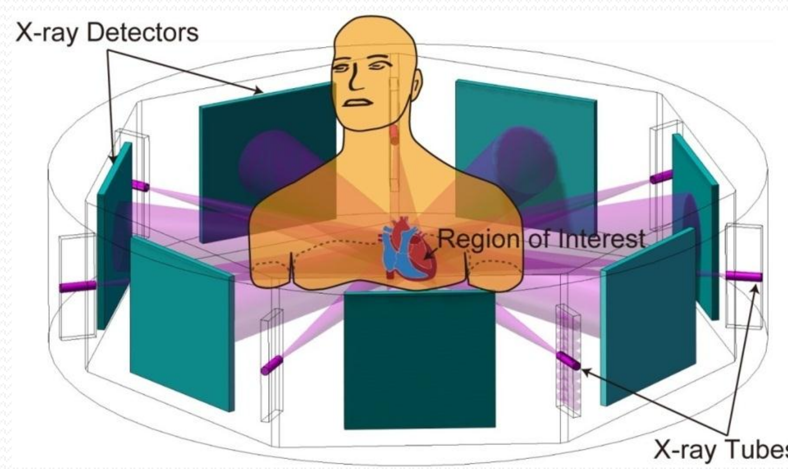


$$P(\theta,t)=\int_{l(\theta,t)}\mu(x,y)dl$$

How does CT works?







http://www.imaginis.com/ct-scan/how-does-ct-work

http://www.medimagingsales.com/medical-imaging-equipment/ct/siemens/siemens-somatom-definition-dual-source-ct

Wang G, Lin TH, Cheng PC, Shinozaki DM, Kim HG: Proc. SPIE Vol. 1556, p. 99-112, July 1991 (Scanning Microscopy Instrumentation, Gordon S. Kino;

Algorithms

$$\hat{X}_{LS} = argmin_X \|AX - b\|^2$$

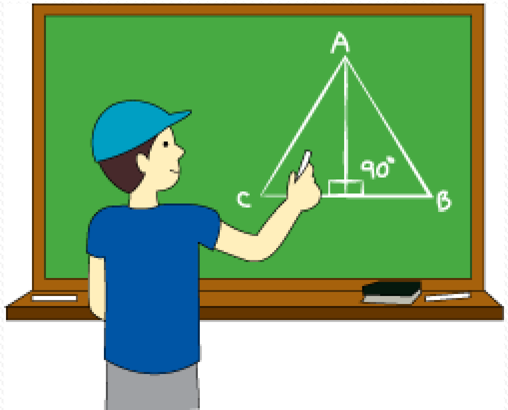
$$\nabla = 2A^T(AX - b)$$

$$X^{k+1} = X^k + \alpha_k \cdot A^T(AX - b)$$

α_k : step length, small number.
Preset the iteration number (200) as the stopping criteria.

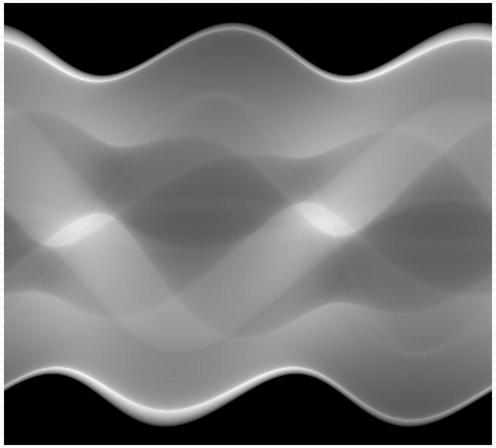
Experiment: Main Geometry Parameters

Geometry	Parameters
# of Detector Cells	888
# of Views	984
FOV in Radius (mm)	249.2
Radius of Scanning (mm)	538.5
Scanning Range (degrees)	360

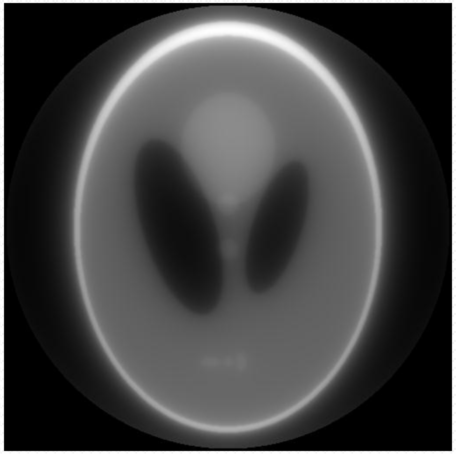


Experiment: Simulated Phantom

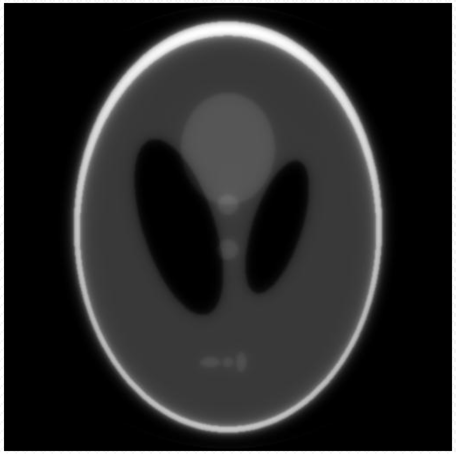
(a) Simulated phantom



(b) Projection data



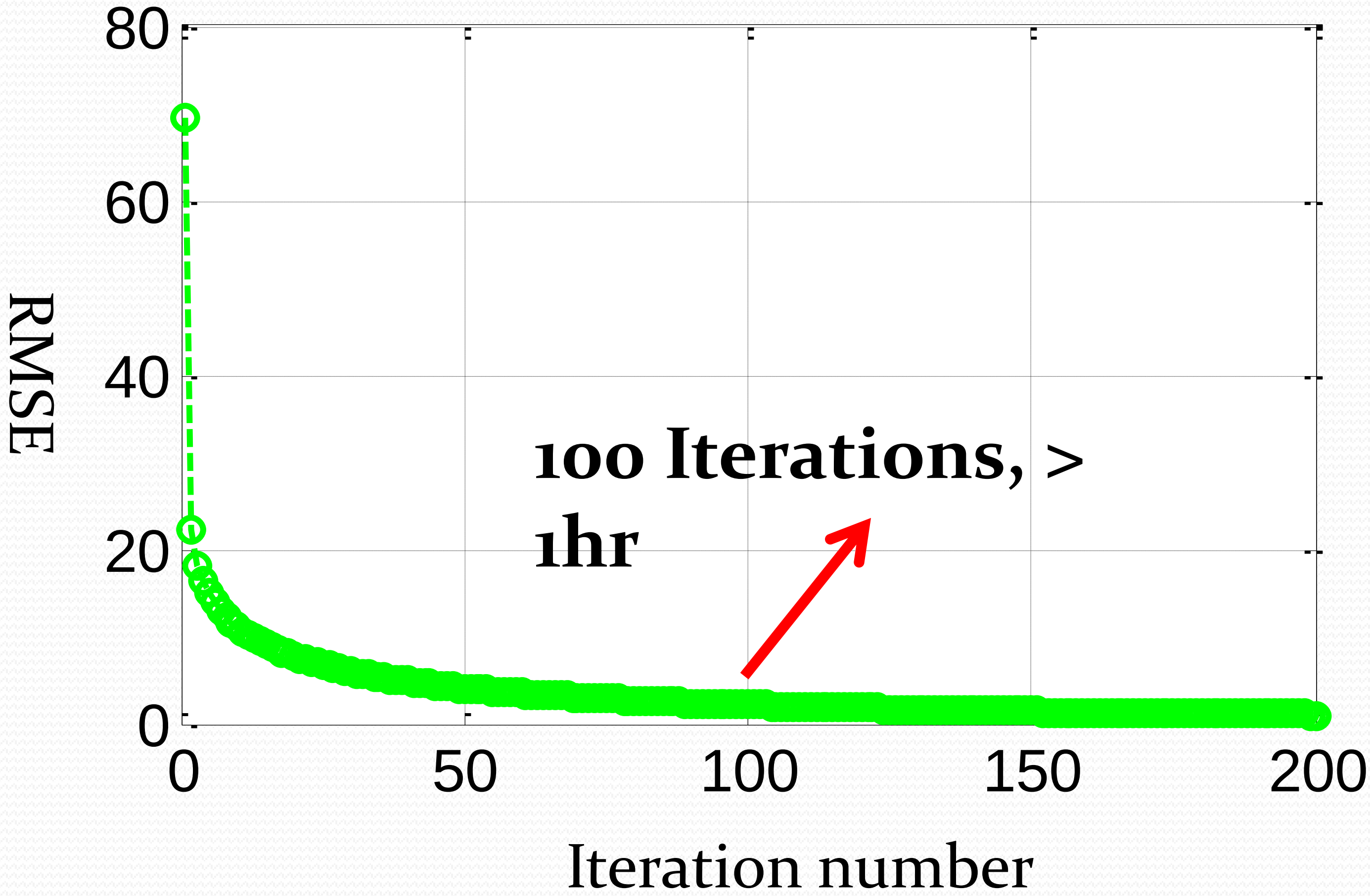
(c) 10 Iterations



(d) 50 Iterations

(e) 200 Iterations

Convergence Property



Convergence speed is **slow!**

Reference

- Prof. Plemmons's class note: "An overview of unconstraint optimization"
- http://en.wikipedia.org/wiki/Gradient_descent
- J. Kinsella: "Course notes for MS4327 Optimization"
- http://en.wikipedia.org/wiki/Health_informatics#cite_note-48
- Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall, 1995, ISBN.
- E. Altendorf, A. Resticar, and T. Dietterich, *Learning from sparse data by exploiting monotonicity constraints*. In UAI, pages 18-26, 2005.
- Yorke ED, Keall P, Verhaegen F. Anniversary, *Role of medical physicists and the AAPM in improving geometric aspects of treatment accuracy and precision*. Medical Physics. 2008; 35(3):828-839
- Burnham, C.A. and G.L. Brownell, *A Multi-Crystal Positron Camera.* Nuclear Science, IEEE Transactions on, 1972. **19**(3): p. 201-205.
- Ter-Pogossian, M.M., et al., *A positron-emission transaxial tomograph for nuclear imaging (PETT).* Radiology, 1975. **114**(1): p. 89-98.
- Phelps, M.E., et al., *Application of annihilation coincidence detection to transaxial reconstruction tomography.* J Nucl Med, 1975. **16**(3): p. 210-24.
- Andersen, A.H., *Algebraic reconstruction in CT from limited views.* Medical Imaging, IEEE Transactions on, 1989. **8**(1): p. 50-55.
- Kuhl, D.E. and R.Q. Edwards, *Image Separation Radioisotope Scanning.* Radiology, 1963. **80.**
- Jasczak, R.J., *Tomographic radiopharmaceutical imaging.* Proceedings of the IEEE, 1988. **76**(9)
- Neumann, D.R., N.A. Obuchowski, and F.P. Difilippo, *Preoperative 123I/99mTc-sestamibi subtraction SPECT and SPECT/CT in primary hyperparathyroidism.* J Nucl Med, 2008. **49**(12): p. 2012-7.
- Sheil, W.C., *Magnetic Resonance Imaging (MRI Scan).* MedicineNet.com. Retrieved 08 Feb. 2013.
- Herman, G.T. and A. Lent, *Iterative reconstruction algorithms.* Computers in Biology and Medicine, 1976. **6**(4): p. 273-294.
- Hendee, W.R. and C.J. Morgan, *Magnetic Resonance Imaging Part I -- Physical Principles.* Western Journal of Medicine, 1984. **141**(4): p. 491-500.


Thank You