Statistical Signal Processing for Radionuclide Tomography

Alfred O. Hero

Depts. EECS and BioMedEng

University of Michigan - Ann Arbor

`hero@eecs.umich.edu`

`http://www.eecs.umich.edu/ hero/index.html`

Outline

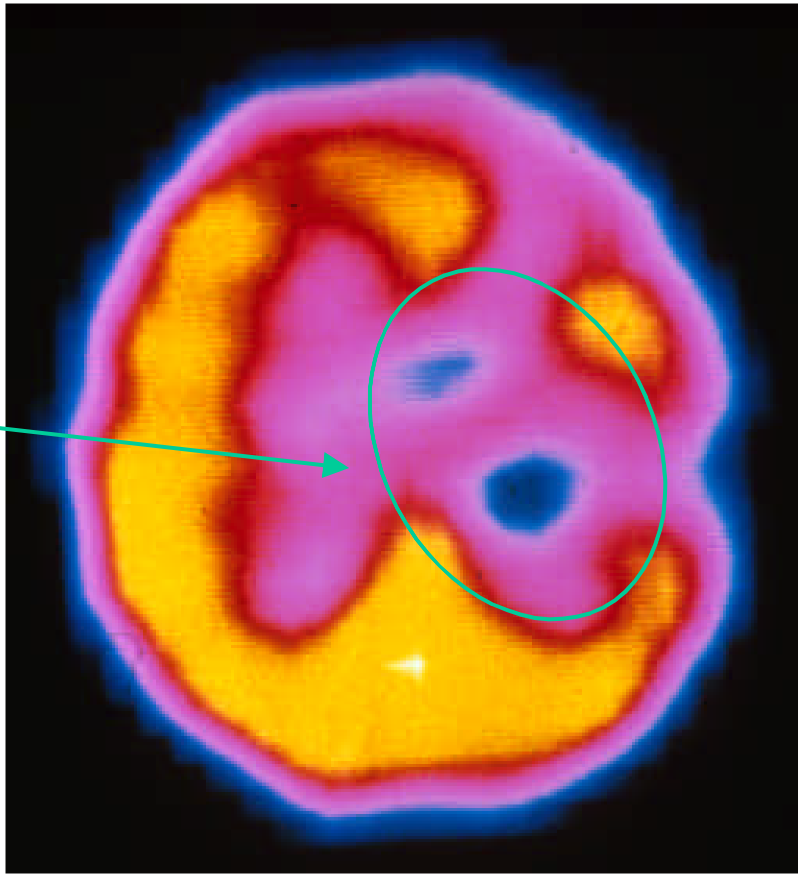
1. Radionuclide imaging background: ECT
2. System model
3. Radionuclide imaging algorithms
4. Radionuclide imaging with MRI/CT side information
5. Bounds and feasibility studies

Single Photon Emission Computed Tomography (SPECT)

- 1958 - Anger camera
- 1963 - first ECT device
- 1964 - parallel-hole collimators
- 1972 - statistical image reconstruction
- 1973 - first CT scanner
- Late 70's - first commercial SPECT (Tomomatic)
- 1979 - dual head SPECT & fan-beam collimators
- 1980 - triple head SPECT
- 1984 - ring geometry SPECT
- 90's - combined CT/SPECT and PET/SPECT

Transverse Section

- Brain Transverse Section
- HMPAO blood flow study
- ***Diagnosis: Evidence of stroke in left cortex***



R Transverse Section L

Tomographic Image

Tomographic Brain Imaging

- Brain Transverse Section
- X-Ray CT
- ***Diagnosis: Lesions in left cortex***

Tomographic Image

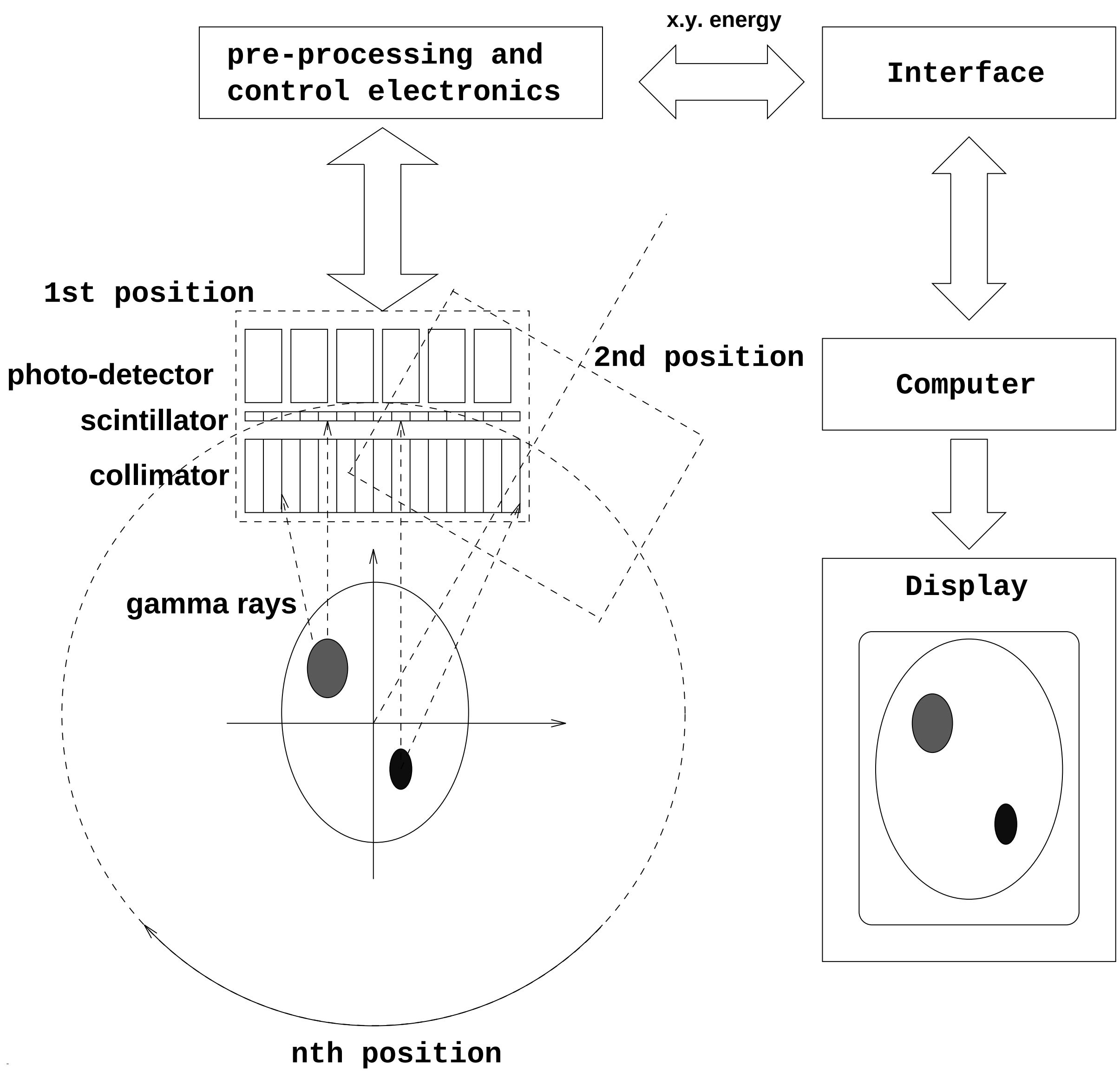

pre-processing and
control electronics
x.y. energy
Interface
1st position
photo-detector
scintillator
collimator
2nd position
Computer
gamma rays
Display
nth position

Figure 1.

Directions in radionuclide tomography

- 3D emission computed tomography (ECT)
- Imaging spatio-temporal processes
- Fusion of anatomical side information
- New detector materials, collimators, projection geometries

1. System Model

- Object intensity distribution: $\underline{\lambda} \in \mathbf{R}^N$
- Detector intensity (fluence) distribution: $\underline{\mu} = \mathbf{A}\underline{\lambda} + \underline{e}$
  - $\mathbf{A}$ = $M \times N$ system matrix
  - $\underline{e}$ = Background intensity (assumed known)
- Projection Data: $\{Y_i\}_{i=1}^{M}$ independent Poisson
- Pseudo-linear system model

$$\underline{Y} = \mathbf{A}\,\underline{\lambda} + \underline{e} + \underline{n}$$

- $\underline{n}$ is vector of independent shifted Poisson random variables
  - $E[\underline{n}] = \underline{0}$
  - $\text{cov}(\underline{n}) = \text{diag}(\underline{\mu}) = \text{diag}(\mathbf{A}\underline{\lambda} + \underline{e})$ = **Signal dependent!**

What makes this a hard problem?

- **Poisson likelihood** is difficult to maximize over $\underline{\lambda}$

- **A** is very large

  2D $\Rightarrow MN = (512 * 128)(128^2) \approx$ 1 Gigabyte (10^9)

  3D $\Rightarrow MN = (512 * 128)(256)(64)(128^3) \approx$ 2 Petabytes (10^{15})

- **A** is typically poorly conditioned

- estimates of $\underline{\lambda}$ must be positive

- estimates of $\underline{\lambda}$ must be spatially smooth

2. Image Reconstruction Algorithms

I. Algebraic Reconstruction (AR): Solve $\underline{Y} = \mathbf{A}\underline{\lambda} + \underline{e}$ for $\underline{\lambda}$

II. Statistical Reconstruction: Iteratively maximize log-likelihood (ML)

$$L_Y(\underline{\lambda}) = \ln f(Y;\, \underline{\lambda}) = \sum_{m=1}^{M} Y_m \ln(\mu_m) - \mu_m - Y_m!$$

or maximize penalized likelihood (PML):

$$\Phi_Y(\underline{\lambda}) = L_Y(\underline{\lambda}) - \beta \underline{\lambda}^T \mathbf{P} \underline{\lambda}$$

- $\underline{\mu} = \mathbf{A}\underline{\lambda} + \underline{e}$
- $\mathbf{P}$ is n.n.d. smoothing matrix
- $\beta > 0$ is smoothing parameter

Expectation-Maximization Algorithm

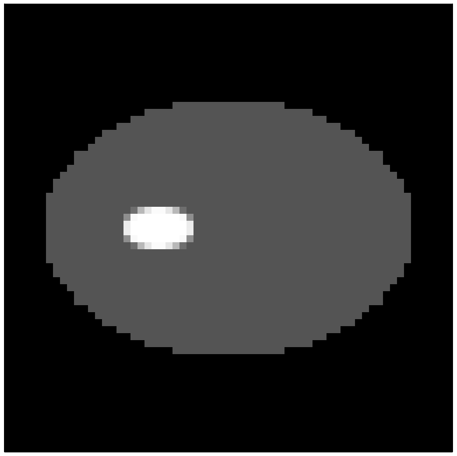
- Hypothesize $\underline{X}$= "Complete data"

- **"Q-Term":** Optimal estimate of unknown distribution

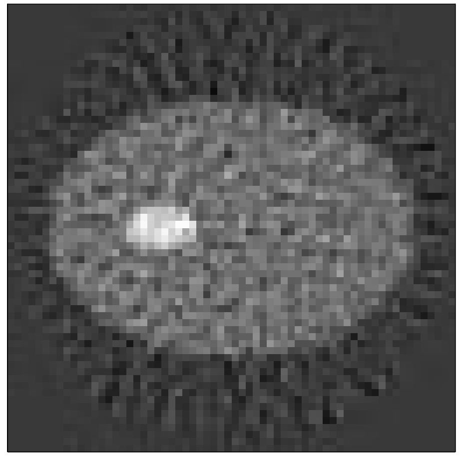
$$Q(\underline{\lambda},\underline{\lambda}^k) := E\left[\ln f(X;\underline{\lambda})|\underline{Y};\ \underline{\lambda}^k\right].$$

- **The EM Algorithm:**
  - **E–Step:** Evaluate Q-term.
  - **M–Step:** Solve $$\underline{\lambda}^{k+1} = \arg\max_{\underline{\lambda}\geq 0} Q(\underline{\lambda},\underline{\lambda}^k).$$

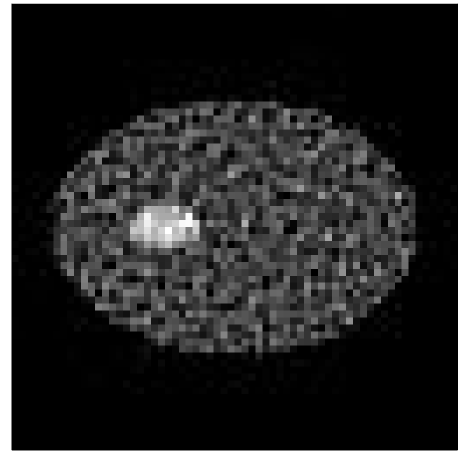
Emission Phantom



FBP Reconstruction of ECT Data



ML−EM Reconstruction of ECT Data

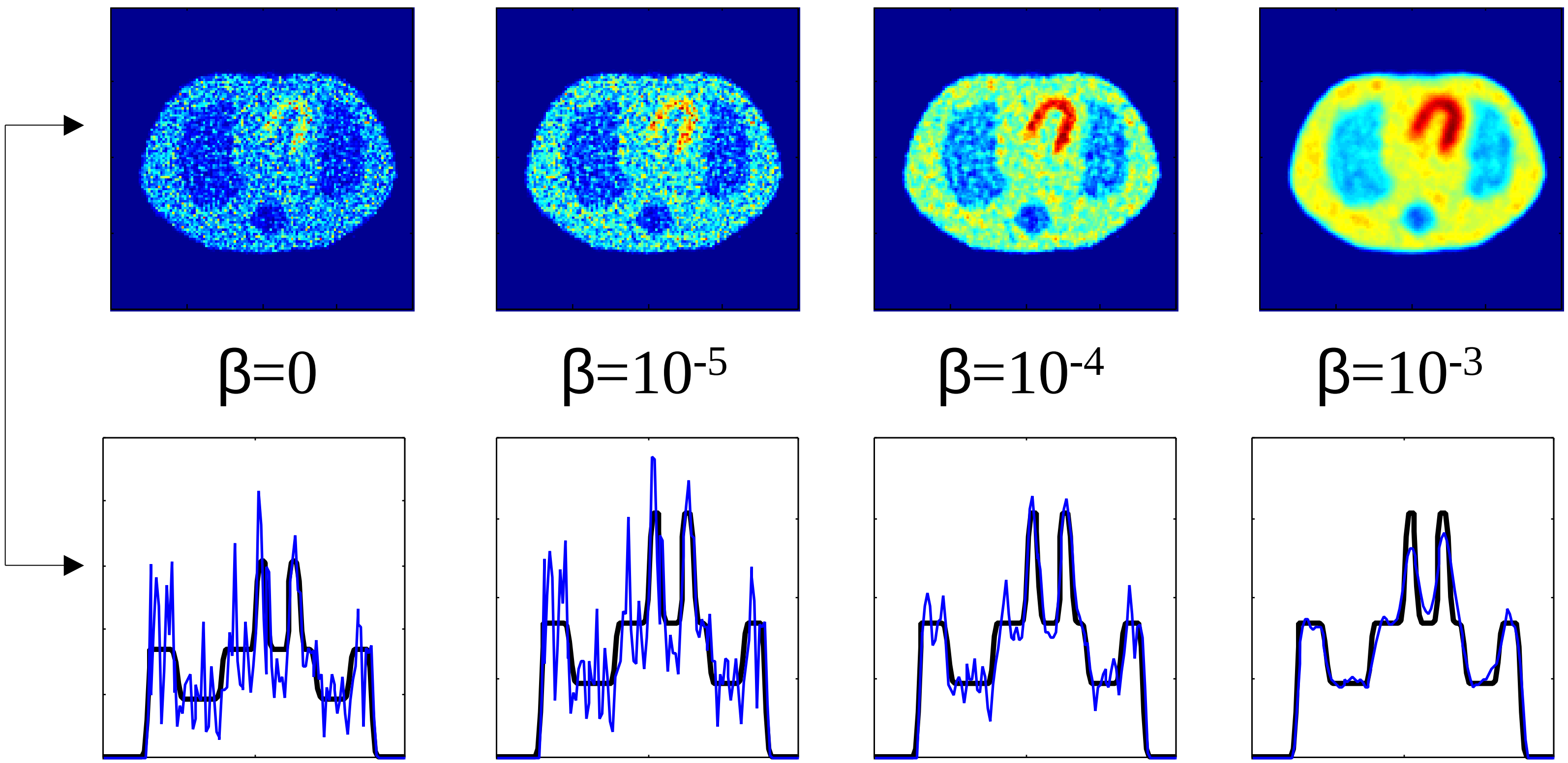


Uptake estimate: $\hat{\alpha} = \underline{1}_{\mathrm{ROI}}^{T} \underline{\hat{\lambda}}$

- FBP oversmooths and ignores Poisson statistics or *prior* information
- ML undersmooths and ignores *smoothness* information

Example: EM Algorithm with Roughness Penalty

10^6 counts, 100 iterations



ML-EM Aceleration Methods

Define new Q function

$$Q_i(\lambda_i, \underline{\lambda}^k) = E\left[\ln f(X_i;\ \lambda_i, \underline{\lambda}^K_{-i}) | \underline{Y};\ \underline{\lambda}^k\right].$$

- SAGE ML estimator [Fessler&Hero SP94, IP95]

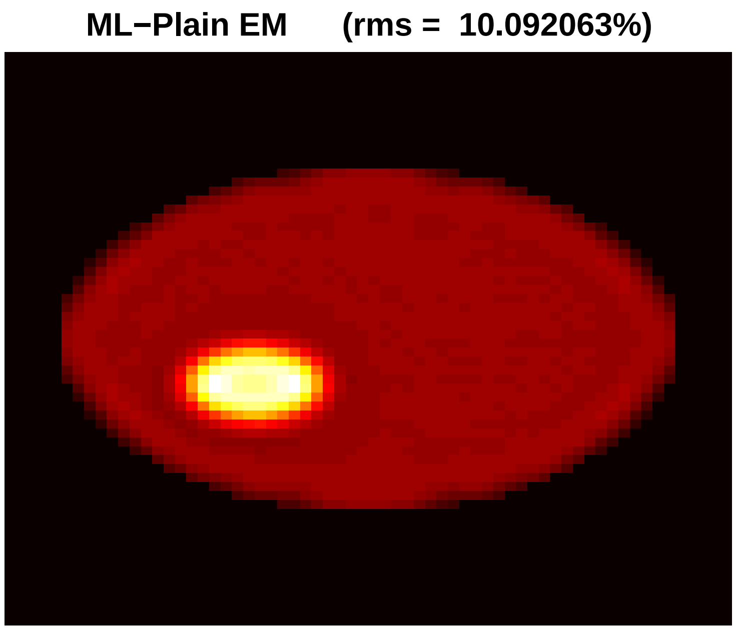
$$\lambda_i^{k+1} = \arg\max_{\lambda_i \geq 0}\ Q_i(\lambda_i, \underline{\lambda}^k)$$

- Kullback Proximal Point Acceleration [Chretien & Hero IT00]

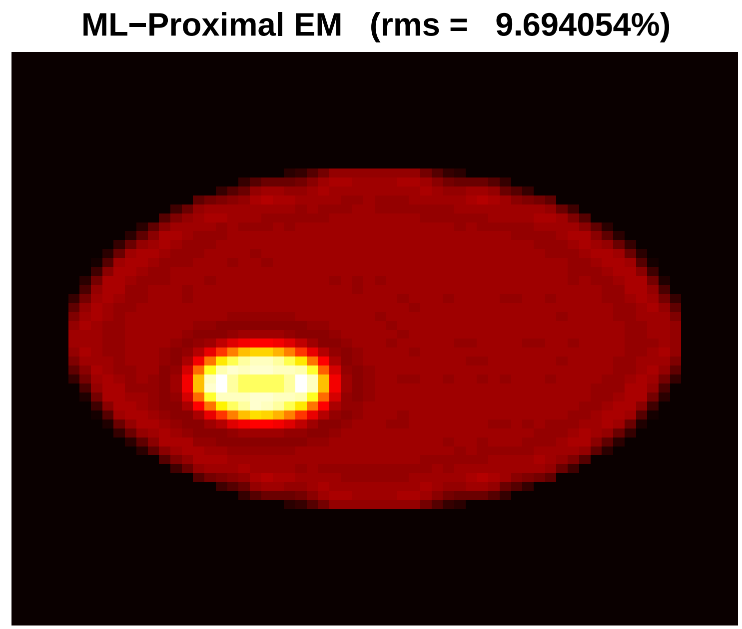
$$\lambda_i^{k+1} = \arg\max_{\lambda_i \geq 0}\ \left\{(1-\rho_k)\ln f(\underline{Y};\ \lambda_i, \underline{\lambda}^k_{-i}) + \rho_k Q_i(\lambda_i, \underline{\lambda}^k)\right\}$$

where $\rho_k > 0$ is a relaxation sequence.

<u>Plain EM and KPP–EM</u>



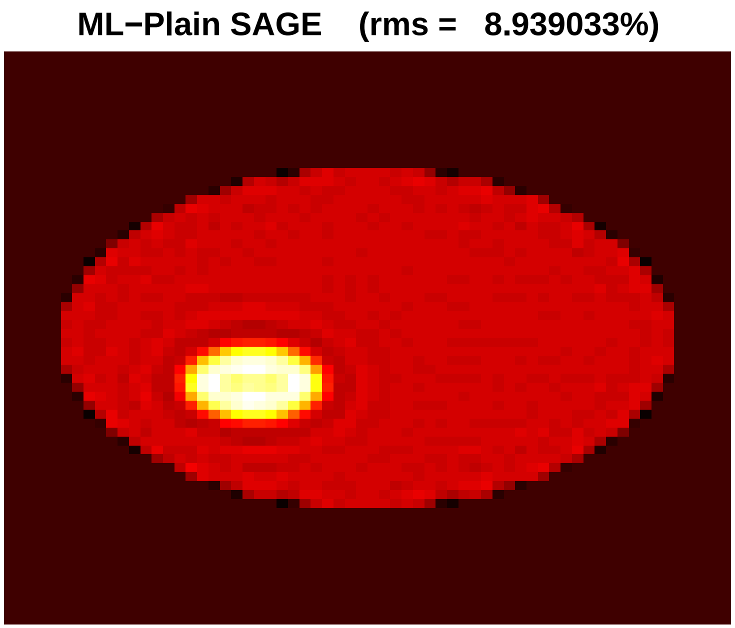
(a) Plain EM



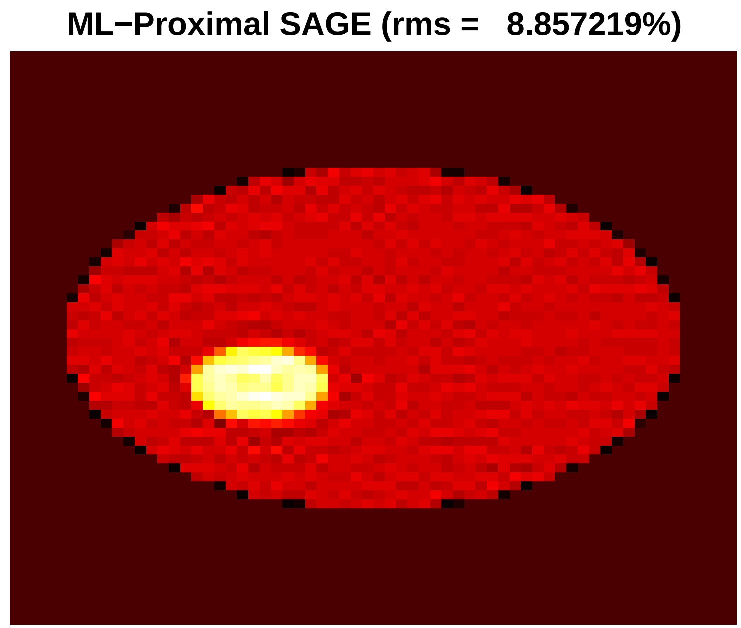
(b) KPP–EM

- Reconstructed image after 50 iterations. (% RMS are included in the title of each figure.)

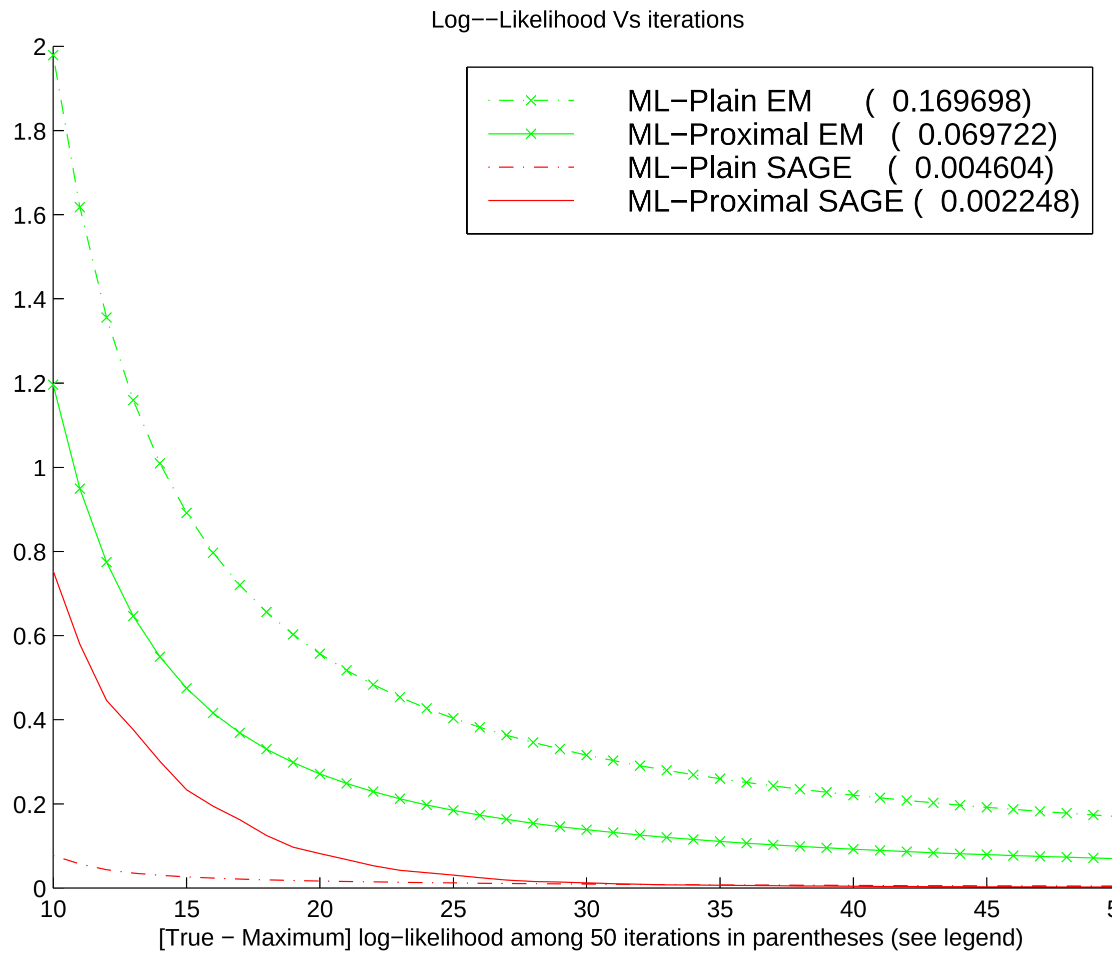
Plain SAGE and KPP–SAGE



(a) Plain SAGE



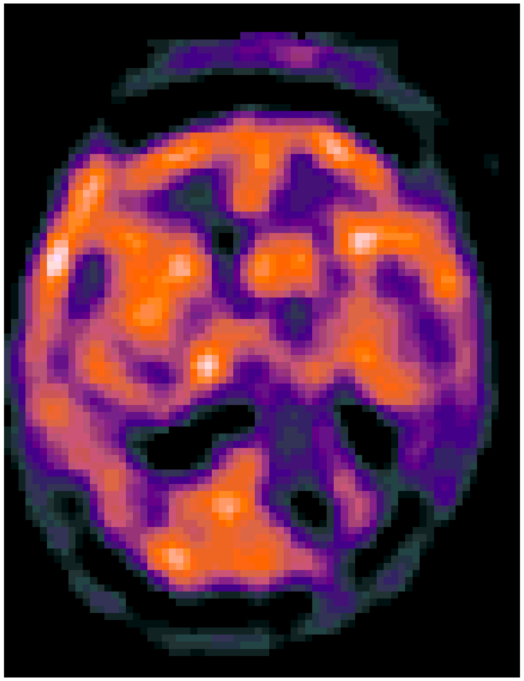
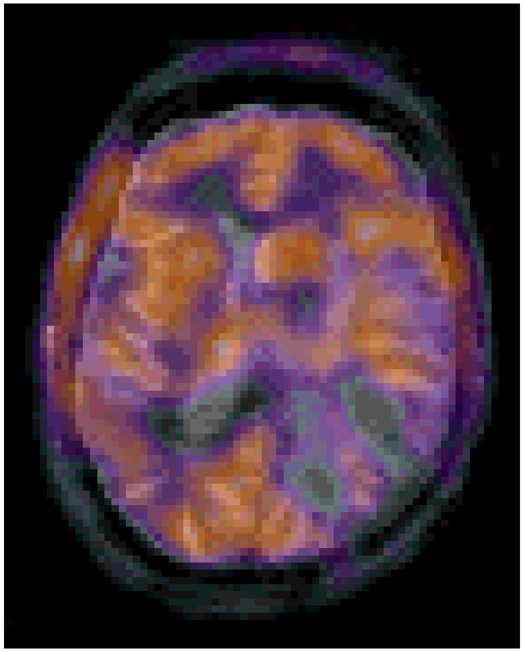
(b) KPP–SAGE



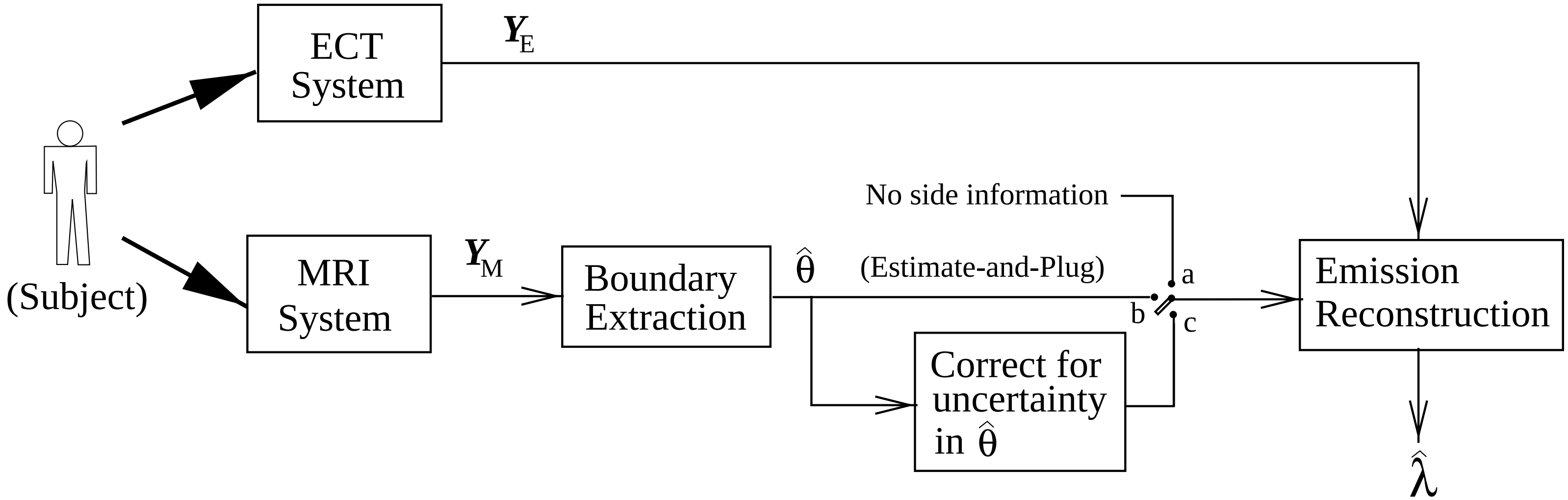
- True log–likelihood minus log–likelihood *Vs* number of iterations ($k = 11, \ldots 50$) for various methods (2–D).

Combining MRI and SPECT For Functional Imaging

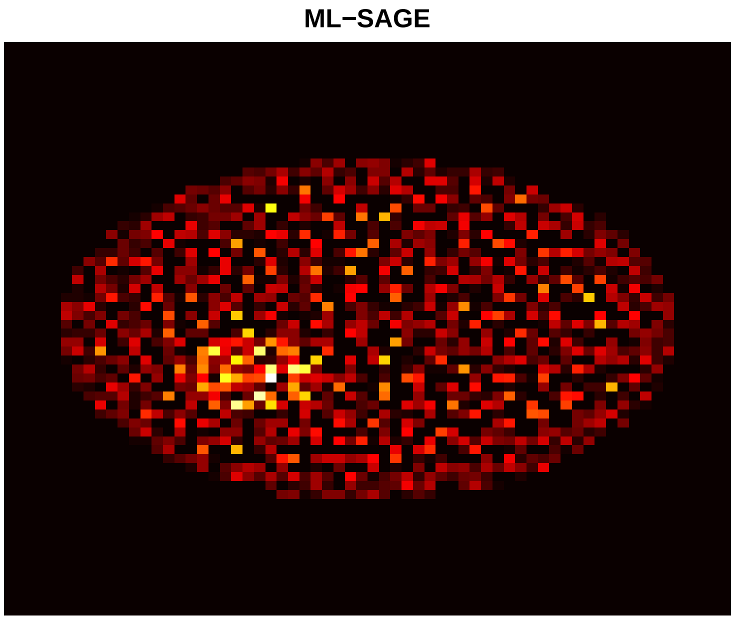
- Tracers: oxygen, glucose, iodine, antibodies, etc.
- "Functional" information about physiological processes
- Uptake estimation in a region of interest

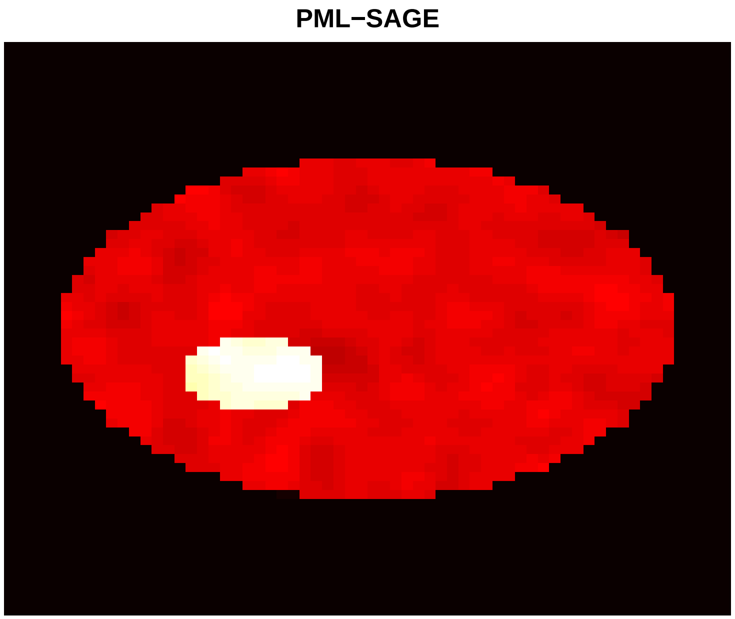
Use of MRI-derived Organ Boundaries



ML vs PML Reconstruction with Perfect Side Info



(c) Maximum Likelihood (ML)



(d) Penalized ML (PML)

- Note: ML image is obtained with $\beta = 0$.

Gibbs Weight Mapping $\omega_{jk}(\underline{\theta})$

- **Non-Negativity**

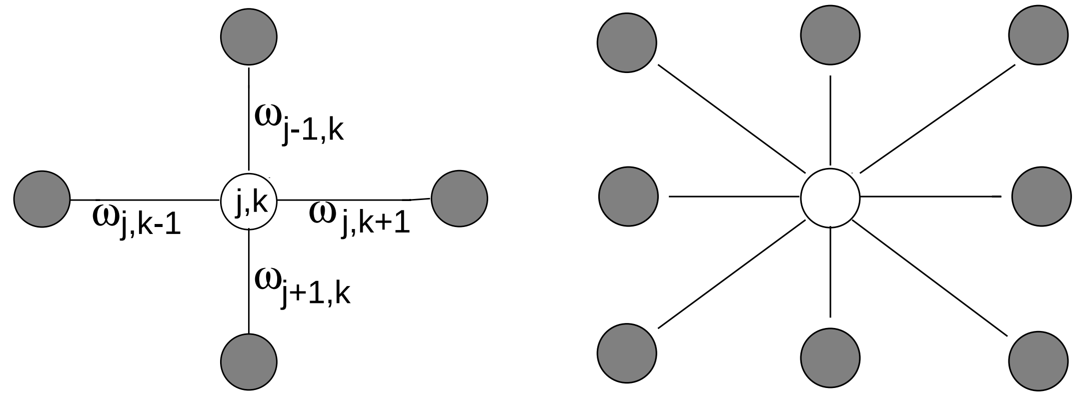
  $\omega_{jk} \geq 0, \ \forall j, k$

- **Symmetry**

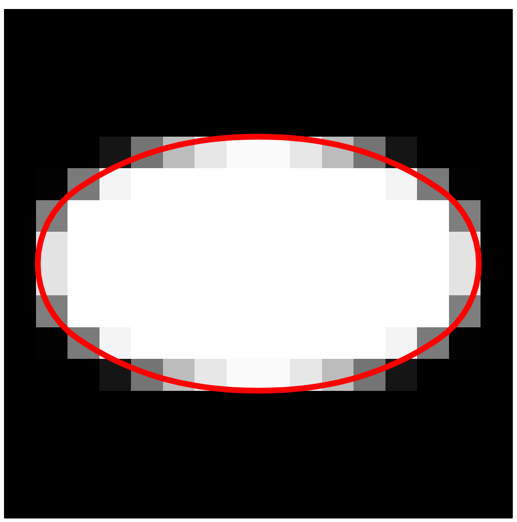
  $\omega_{jk} = \omega_{kj}, \ \forall j, k$

- **Locality**

  $\omega_{jk} = 0$ (non-neighbors)



Example Normalized Image and Boundary



- **Boundary relaxed weight mapping**

  Normalized "set membership" image $J(\underline{\theta})$ used to prevent smoothing across boundary.

ECT Reconstruction via Spatially Variant Gibbs Model

- Select $\underline{\lambda}$ to maximize penalized likelihood

$$\Phi_{Y,\underline{\omega}}(\underline{\lambda}) = \ln f(Y_E;\, \underline{\lambda}) - \beta \underline{\lambda}^T \mathbf{P} \underline{\lambda}$$

- Penalty

$$\underline{\lambda}^T \mathbf{P} \underline{\lambda} = \sum_{j,k \in N} \omega_{jk} |\lambda_j - \lambda_k|^2$$

- N is clique (neighborhood)

- $\omega_{jk} = \omega_{jk}(\underline{\theta})$

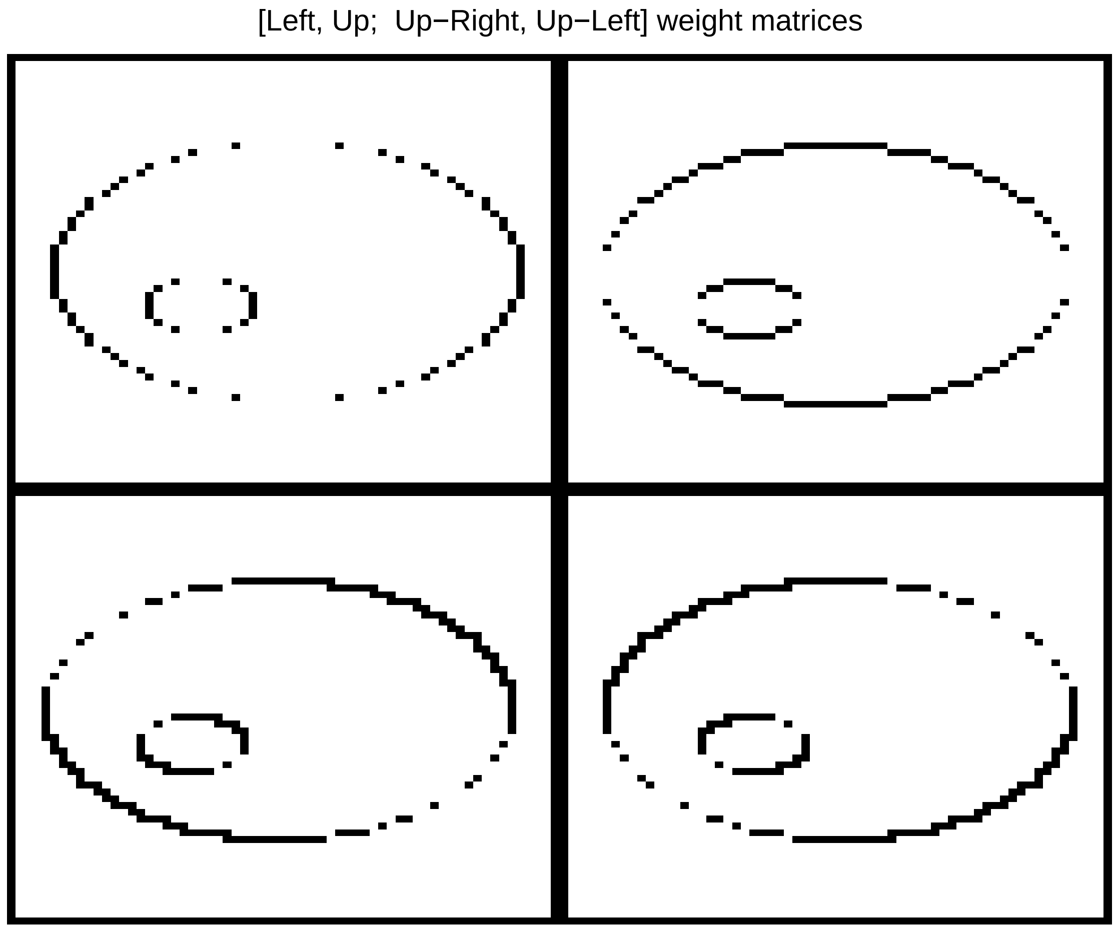


Figure 2. Ideal weights: a) left, b) up, c) up-right, d) up-left.

Incorporation of Imperfect Side Info

For unknown $\underline{\theta}$ we could either

- Use "estimate and plug weights": $\omega_{ij} = \omega_{ij}(\hat{\underline{\theta}})$
- Use minmax averaged weights [Hero & etal, IT99]

$$\begin{aligned}
\tilde{w}_{ij}(\hat{\underline{\theta}}) &= \frac{|\hat{F}_{\hat{\underline{\theta}}}|^{\frac{1}{2}}}{(\sqrt{2\pi})^p} \cdot \int_{\Theta} w_{ij}(\underline{\theta}) \exp\left\{-\tfrac{1}{2}(\underline{\theta} - \hat{\underline{\theta}})^T \hat{F}^{\dagger}_{\hat{\underline{\theta}}} (\underline{\theta} - \hat{\underline{\theta}})\right\} d\underline{\theta} \\
&= w_{ij}(\hat{\underline{\theta}}) * h(\hat{\underline{\theta}})
\end{aligned}$$

- $\hat{F}_{\hat{\underline{\theta}}}$ is "empirical Fisher information" matrix
- h is a Gaussian convolution kernel

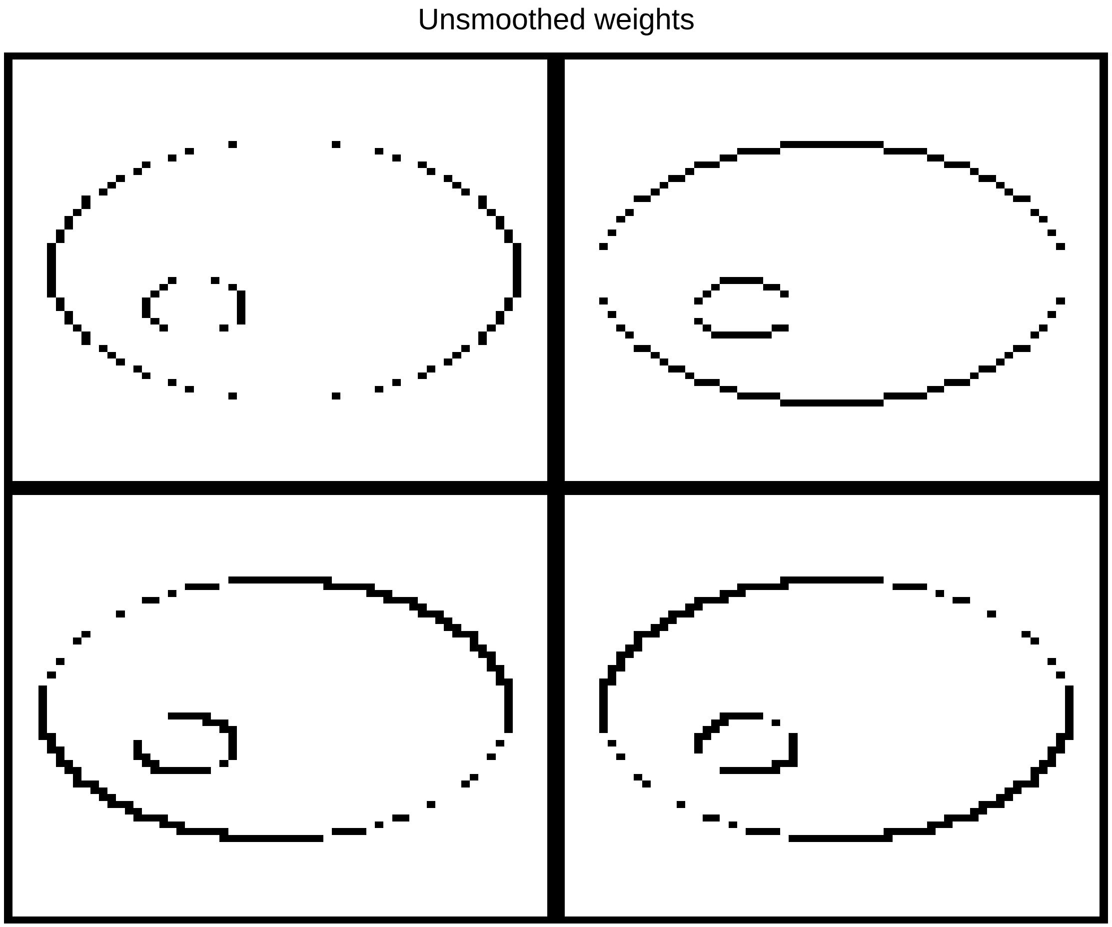


Figure 3. Extracted weights: a) left, b) up, c) up-right, d) up-left.

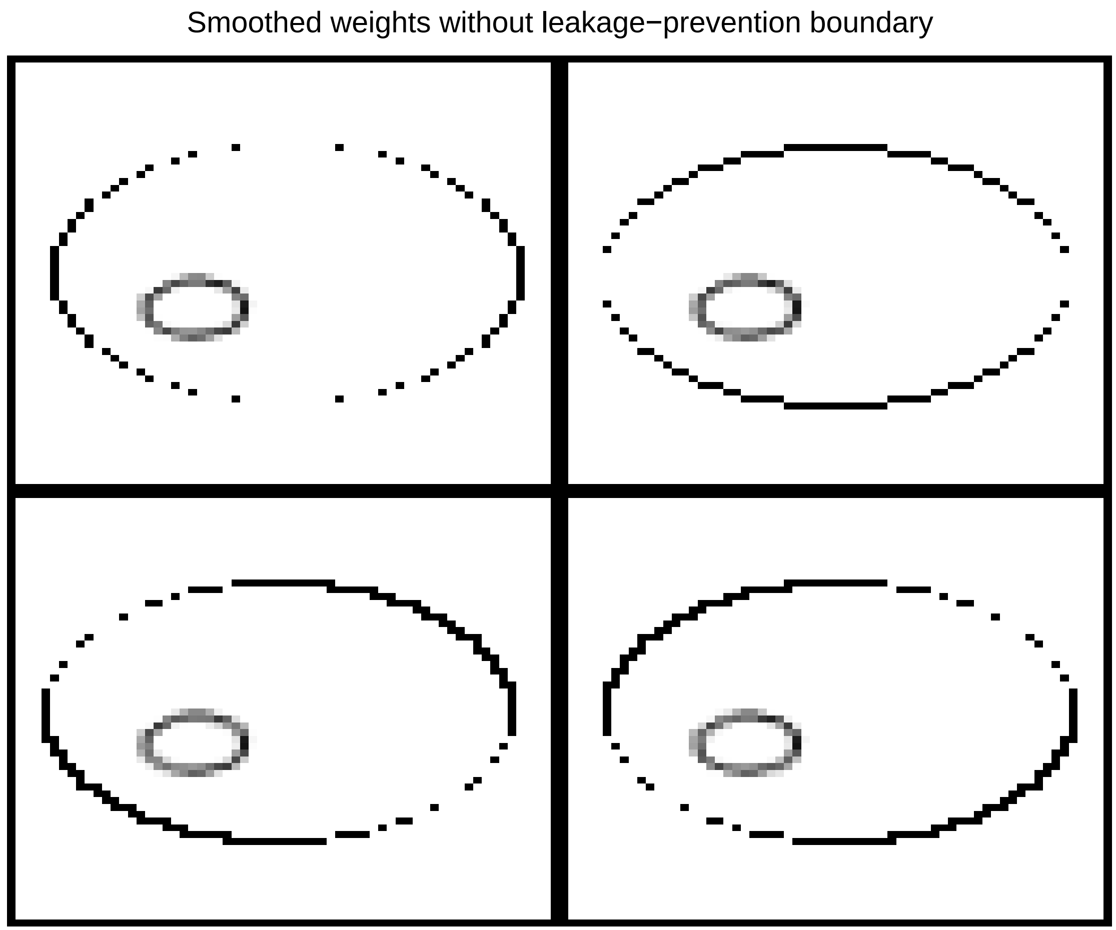
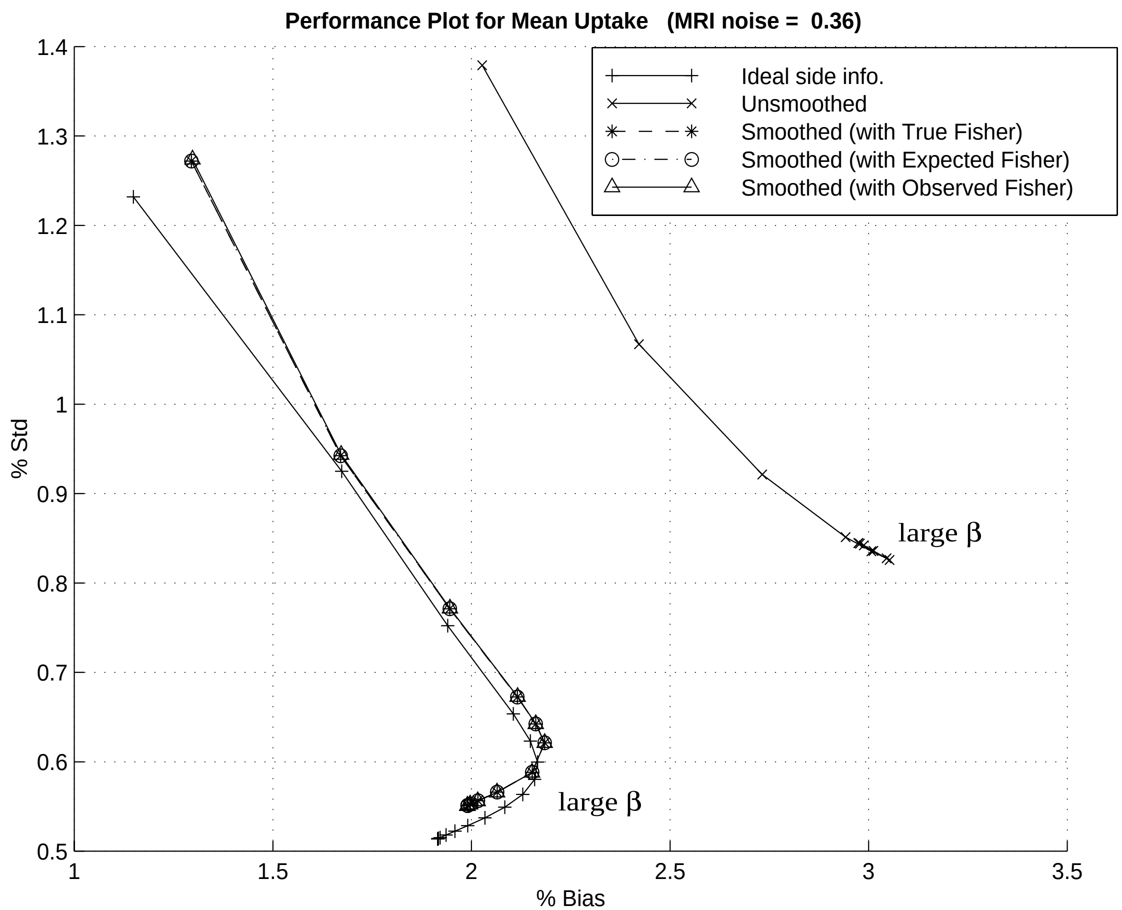


Figure 4. Smoothed weights: a) left, b) up, c) up-right, d) up-left.



(a) Mean Uptake Bias Variance Tradeoff

3. Bounds and Feasibility Studies

General statistical measures of performance

Let $\hat{\underline{\lambda}}$ be an estimator of source intensity $\underline{\lambda} \in \mathbf{R}^N$.

- Estimator bias: $\text{bias}_{\underline{\lambda}}(\hat{\lambda}_p) = E_{\underline{\lambda}}[\hat{\lambda}_p] - \underline{\lambda}$

- Estimator variance: $\text{var}_{\underline{\lambda}}(\hat{\lambda}_p)$

- Estimator MSE: $\text{var}_{\underline{\lambda}}(\hat{\lambda}_p) + \text{bias}^2_{\underline{\lambda}}(\hat{\lambda}_p)$

Standard performance measures for imaging systems

Let $\underline{\lambda} = \underline{e}_p$ denote a point source intensity at pixel location p

$$\underline{e}_p = [0, \ldots, 0, 1, 0, ..., 0]^T$$

Point spread function (PSF): $\underline{h}_p = E_{\underline{\lambda}=\underline{e}_p}[\hat{\lambda}_p]$

- Point source sensitivity (volume of PSF): $\eta_p = \sum_{j=1}^{N} h_p^2(j)$

- Recoverable resolution (width of PSF):

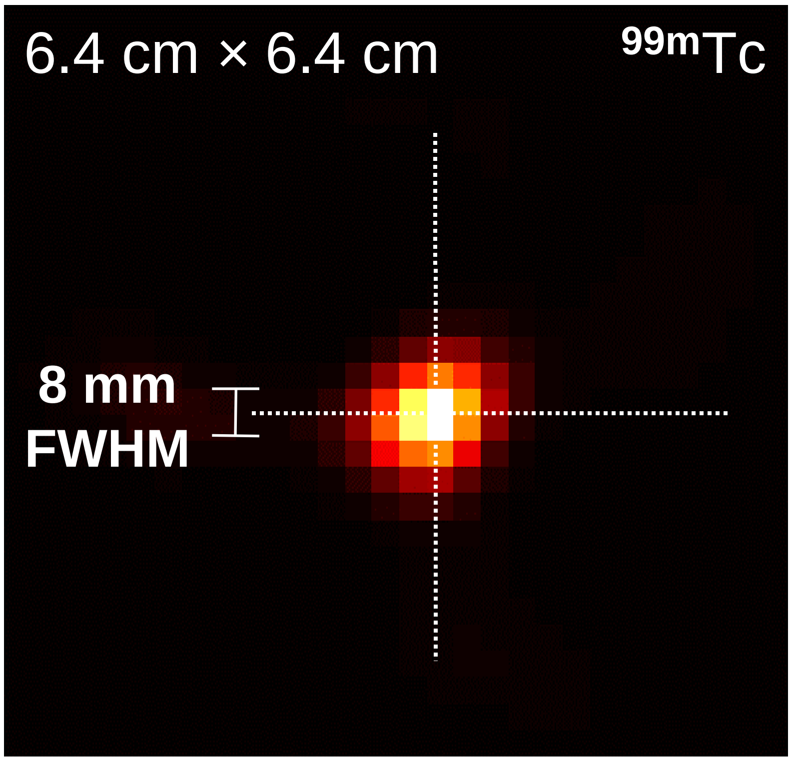
$$\text{FWHM}\{h\} = \sqrt{\frac{1}{\eta_p} \sum_{j=1}^{N} (j-p)^2 \, h_p^2(j)} = \|\underline{h}_p - \underline{e}_p\|$$

$\|\underline{z}\|$: 2nd-moment-of-inertia norm on $\mathbf{R}^N$.

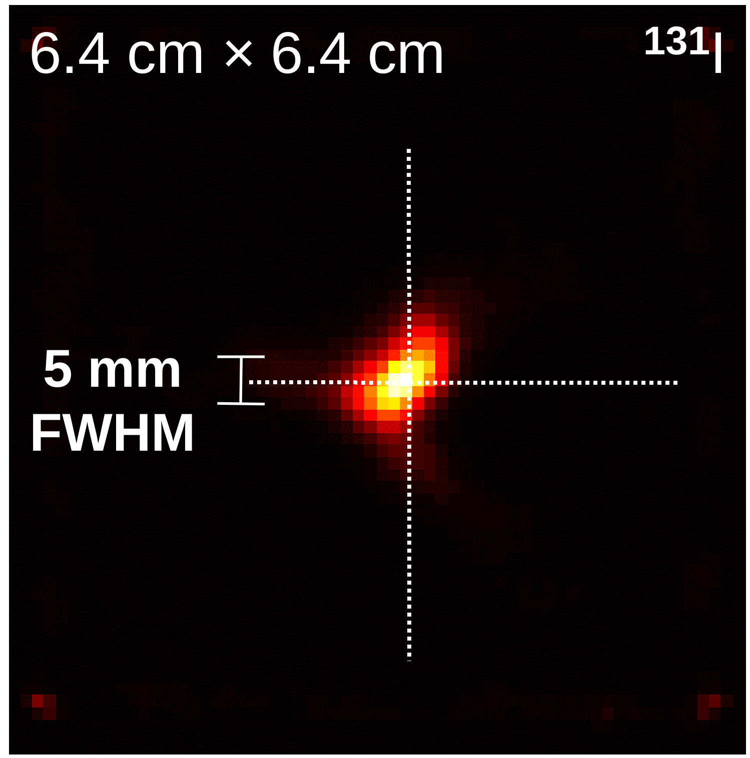
^{99m}Tc and ^{131}I Point Source Images (2D)

Single on-axis point source at 10 cm

100K coincident events



100K coincident events



Fisher information matrix

$$F_Y(\underline{\lambda}) = E_{\underline{\lambda}}[\nabla_{\underline{\lambda}} \ln f(\underline{Y};\underline{\lambda}) \, \nabla^T_{\underline{\lambda}} \ln f(\underline{Y};\underline{\lambda})]$$

Unbiased CR Bound

Assume $\hat{\underline{\lambda}}$ is any unbiased estimator, $\mathrm{bias}_{\underline{\lambda}}(\hat{\lambda}_p) = 0$, $\underline{\lambda} \in \mathbf{R}^N$. Then

$$\mathrm{var}_{\underline{\lambda}}(\hat{\lambda}_p) \geq \underline{e}_p^T F_Y^{-1} \underline{e}_p$$

Biased CR Bound

Assume $\hat{\underline{\lambda}}$ is any estimator such that bias gradient $\nabla_{\underline{\lambda}} \mathrm{bias}_{\underline{\lambda}}(\hat{\lambda}_p)$ equals $\underline{\beta}$. Then

$$\mathrm{var}_{\underline{\lambda}}(\hat{\lambda}_p) \geq [\underline{e}_p + \underline{\beta}]^T F_Y^{-1} [\underline{e}_p + \underline{\beta}]$$

$\Rightarrow$ **applicability of these CR bounds is very limited**

<u>Uniform CR Bound</u> [Hero& etal SP96]

Assume $\hat{\underline{\lambda}}$ is any estimator such that bias gradient norm $\|\nabla_{\underline{\lambda}} \text{bias}_{\underline{\lambda}}(\hat{\lambda}_p)\|$ is less than δ, $0 \leq \delta \leq 1$. Then:

$$\text{var}_{\underline{\lambda}}(\hat{\lambda}_p) \geq B(\underline{\theta}, \delta)$$

where

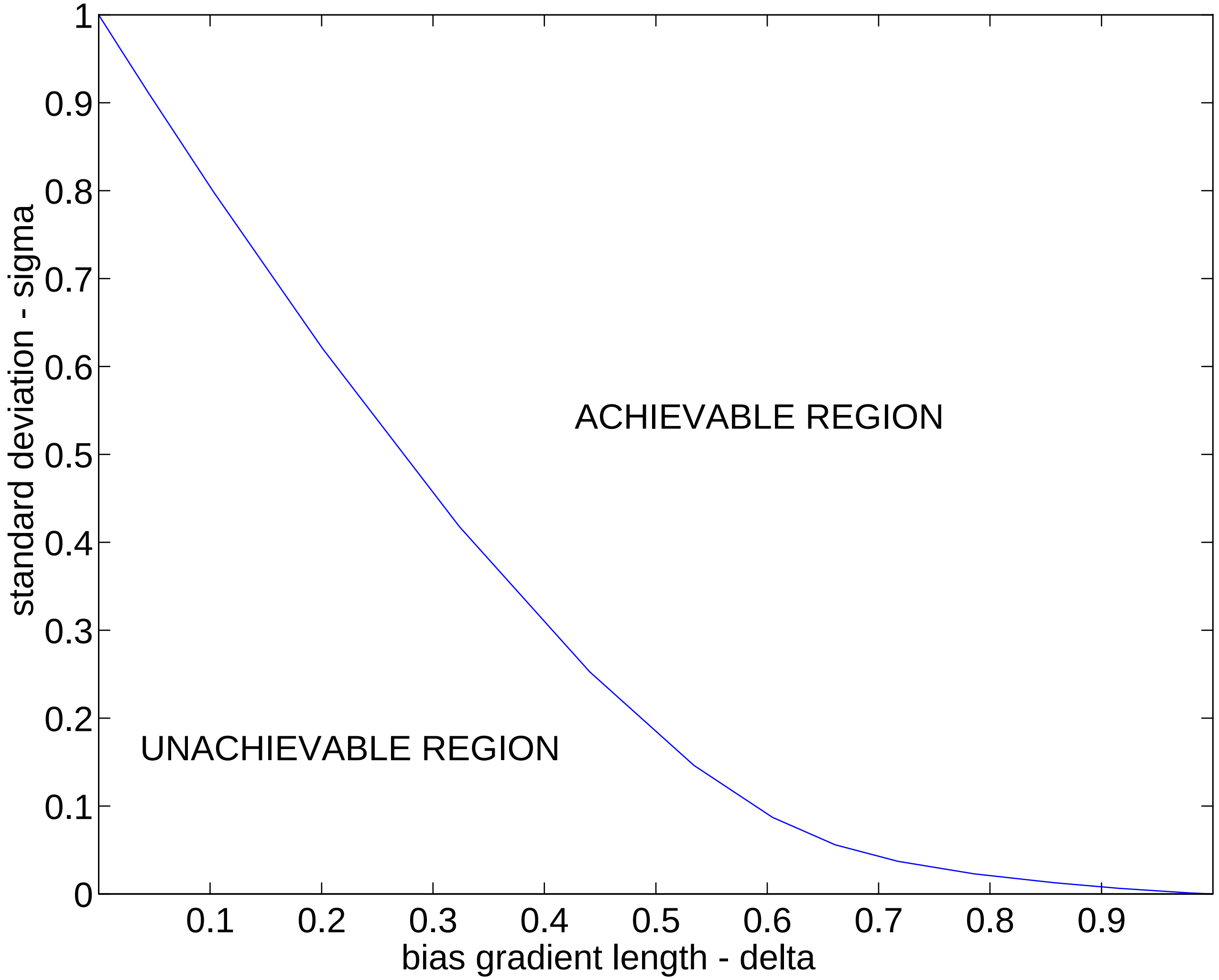
$$\begin{aligned} B(\underline{\theta}, \delta) &= [\underline{e}_p + \underline{d}_{min}]^T F_Y^{-1} [\underline{e}_p + \underline{d}_{min}] \\ &= \rho^2 \underline{e}_p^T [I + \rho F_Y]^{-1} F_Y [I + \rho F_Y]^{-1} \underline{e}_p, \end{aligned}$$

$$\underline{d}_{min} = -[I + \rho F_Y]^{-1} \underline{e}_p$$

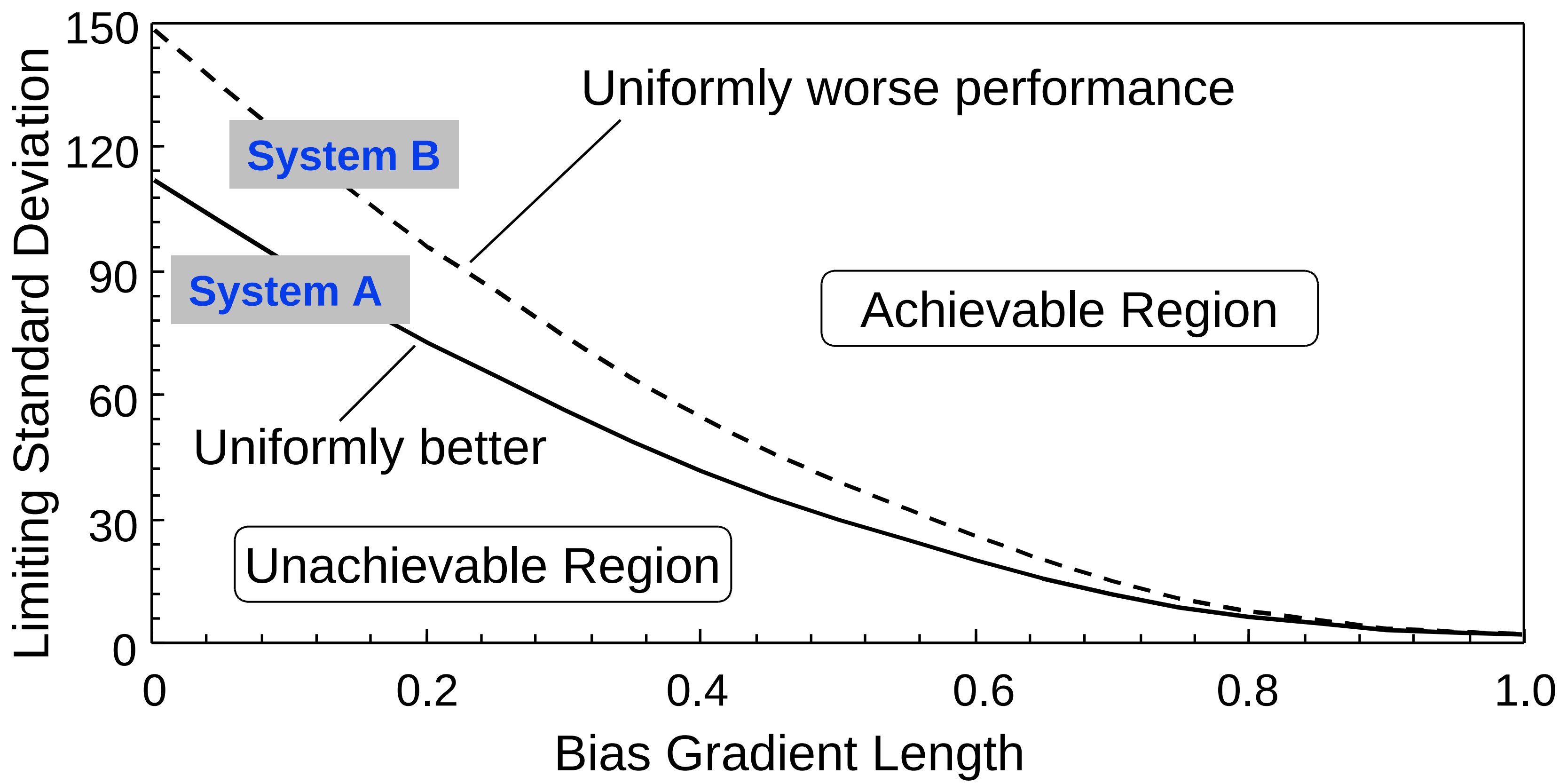
and ρ is given by solution to

$$g(\rho) = \|\underline{d}_{min}\|^2 = \delta^2 \qquad \rho \geq 0$$

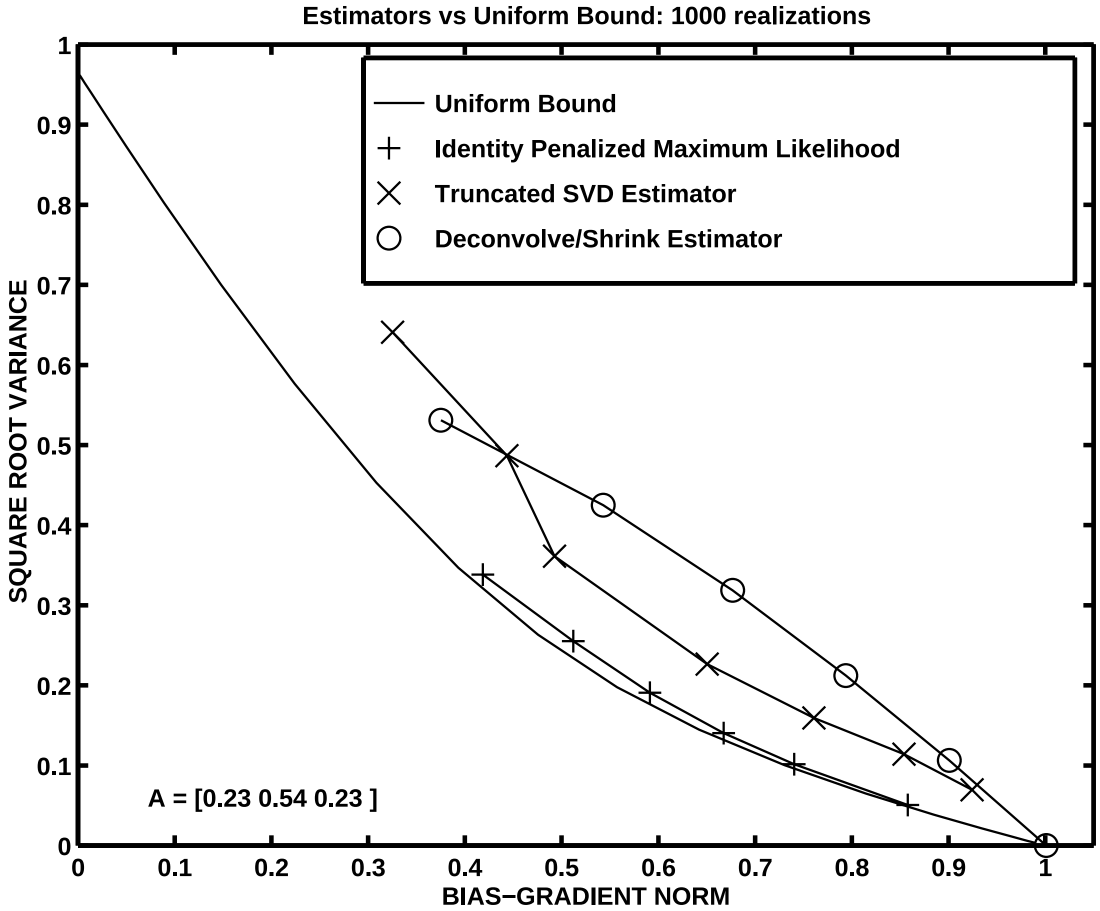
Bias-Variance Plane



Example Uniform Cramer-Rao Bound Curve



Achievability of Uniform CRB?



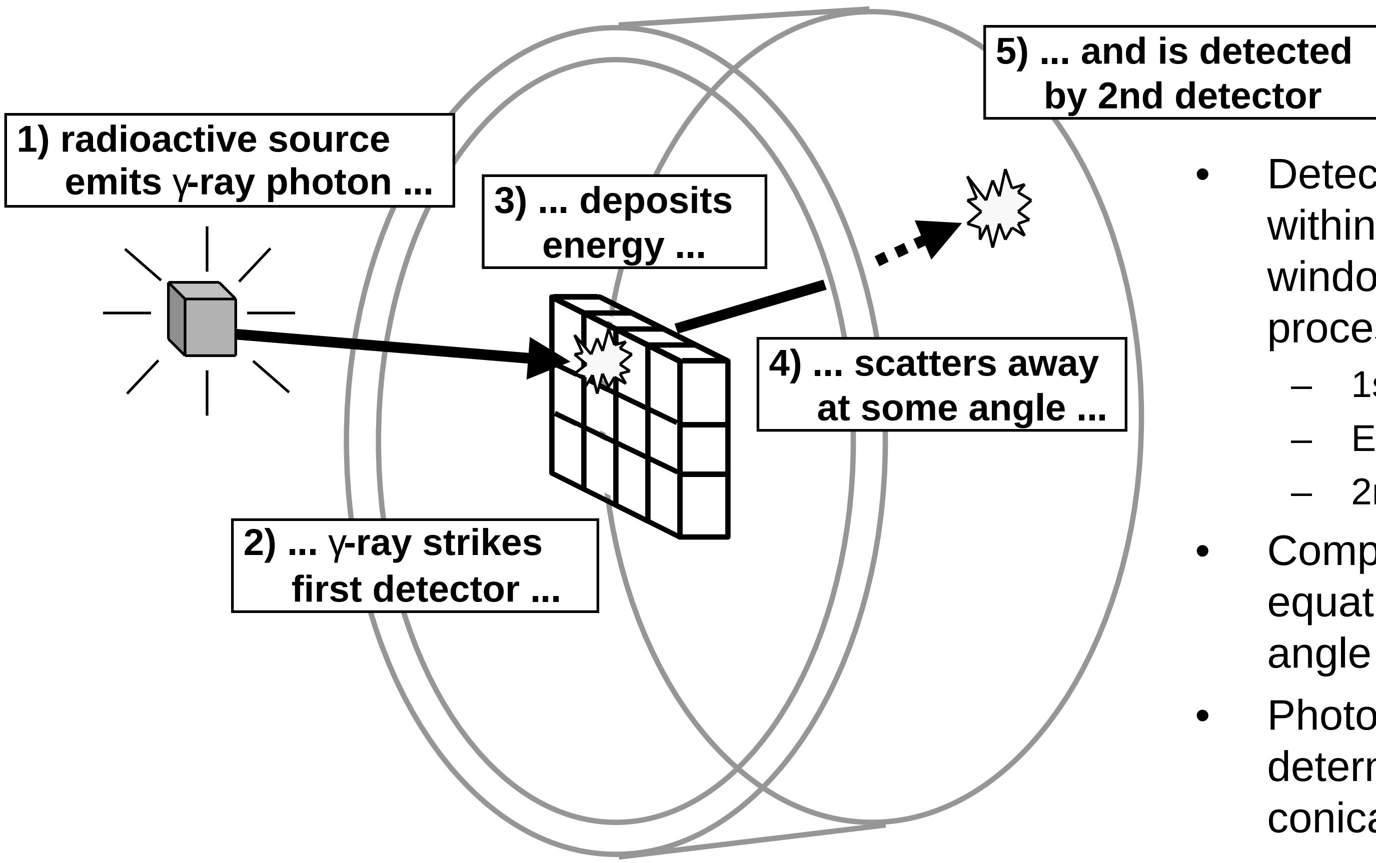
Relation between bias gradient and recoverable resolution

For PML estimator $\hat{\underline{\lambda}} = \text{argmax}_{\underline{\lambda}} \ln f(Y;\underline{\lambda}) - \beta\underline{\lambda}^T \mathbf{P}\underline{\lambda}$

$$\|\nabla_{\underline{\lambda}} \text{bias}_{\underline{\lambda}}(\hat{\lambda}_p)\| = \underbrace{\| \left[F_Y(\underline{\lambda})[F_Y(\underline{\lambda}) + \beta\mathbf{P}]^{-1} - I_N\right] \underline{e}_p\|^2}_{\|E_{\underline{\lambda}=\underline{e}_p}[\hat{\lambda}_p] - \underline{e}_p\|} + O(1/\beta)$$

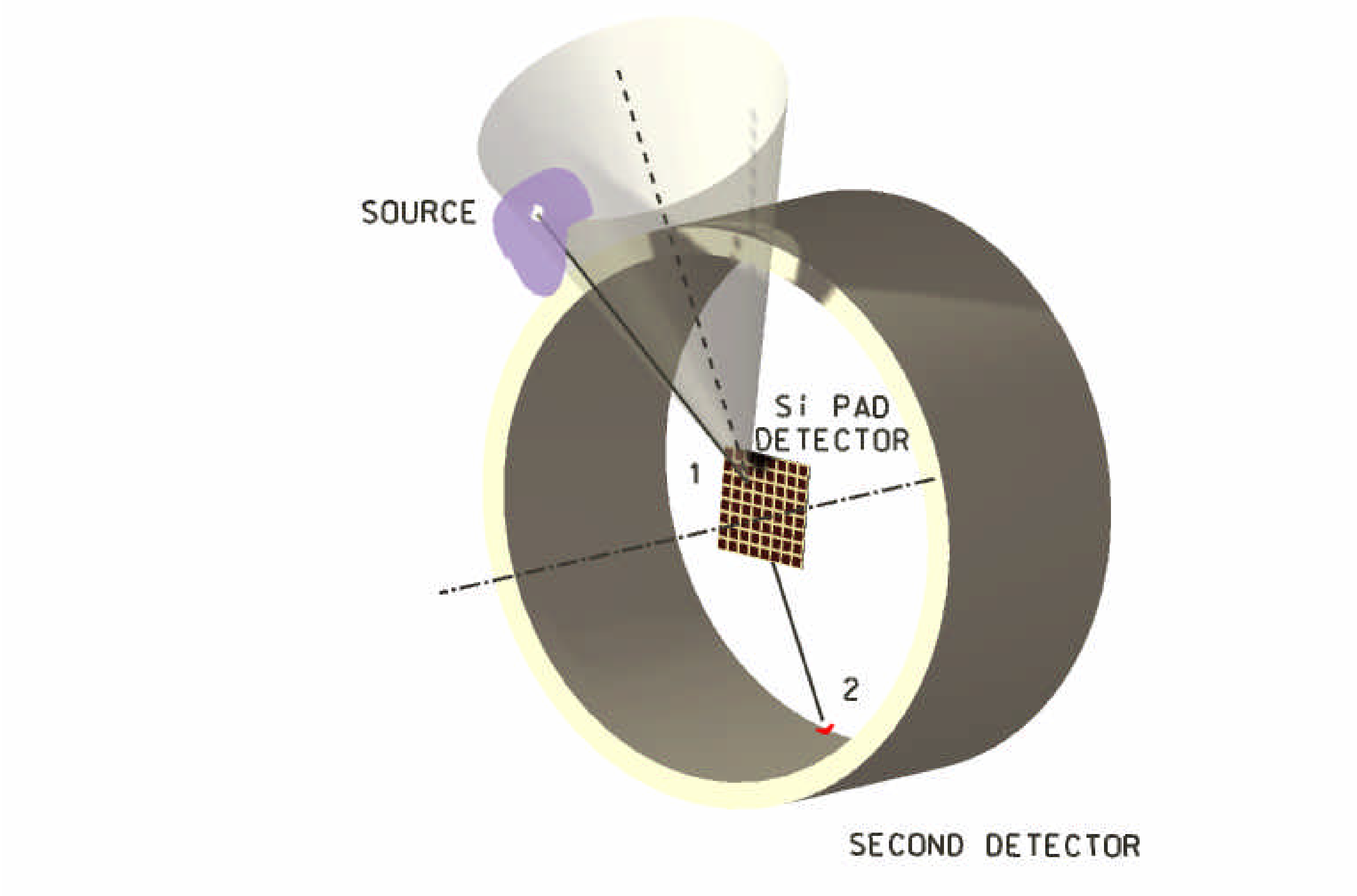
$$= \eta_p \, \text{FWHM}\{\underline{h}_p\} + O(1/\beta)$$

Compton-SPECT Camera Operating Principles

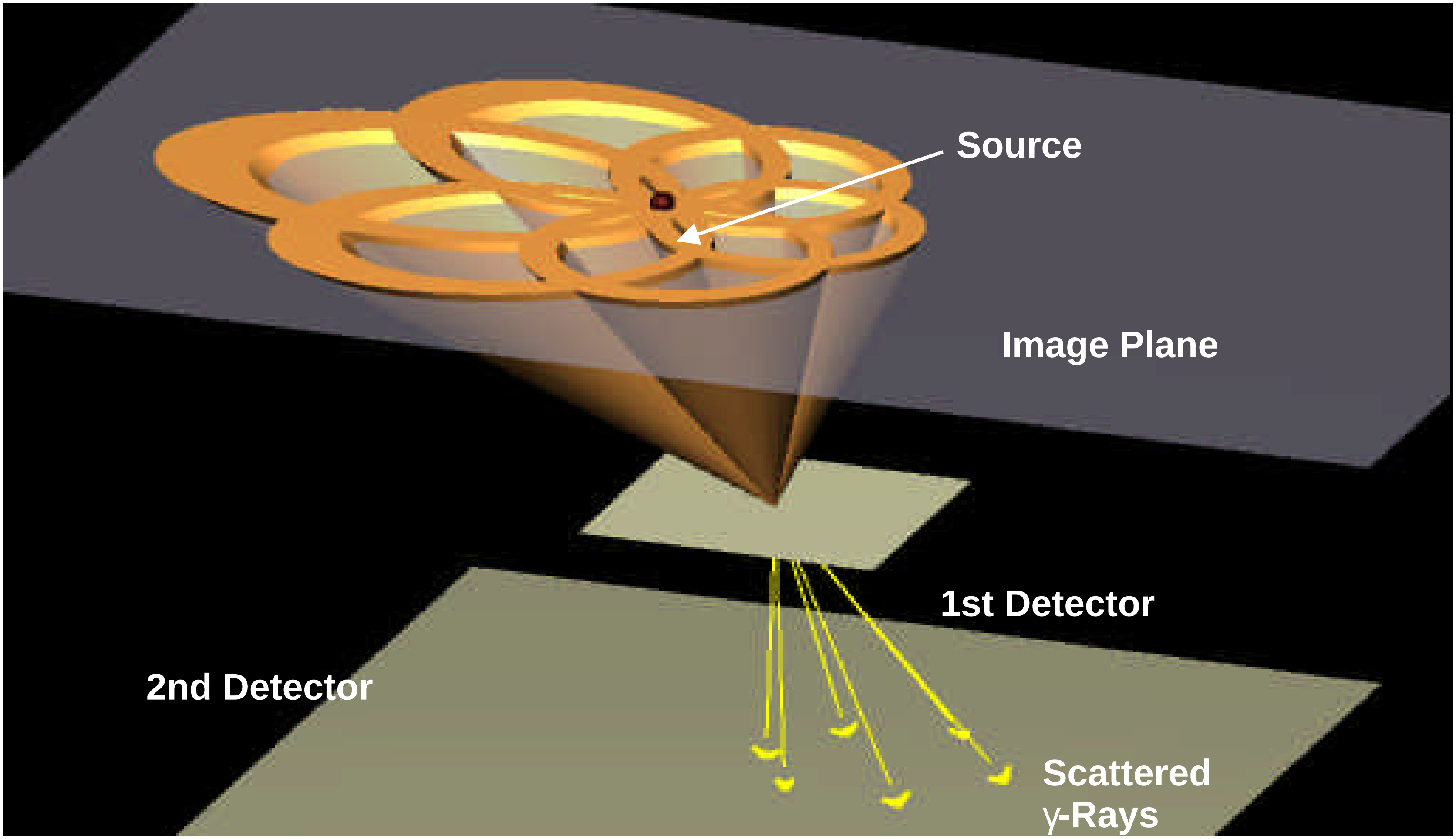


- Detections occurring within a small time window are recorded for processing
  - 1st Detector Position
  - Energy Deposited
  - 2nd Detector Position
- Compton scatter equation relates scatter angle / energy
- Photon direction determined to within a conical ambiguity

Single Measurement Backprojection Cone

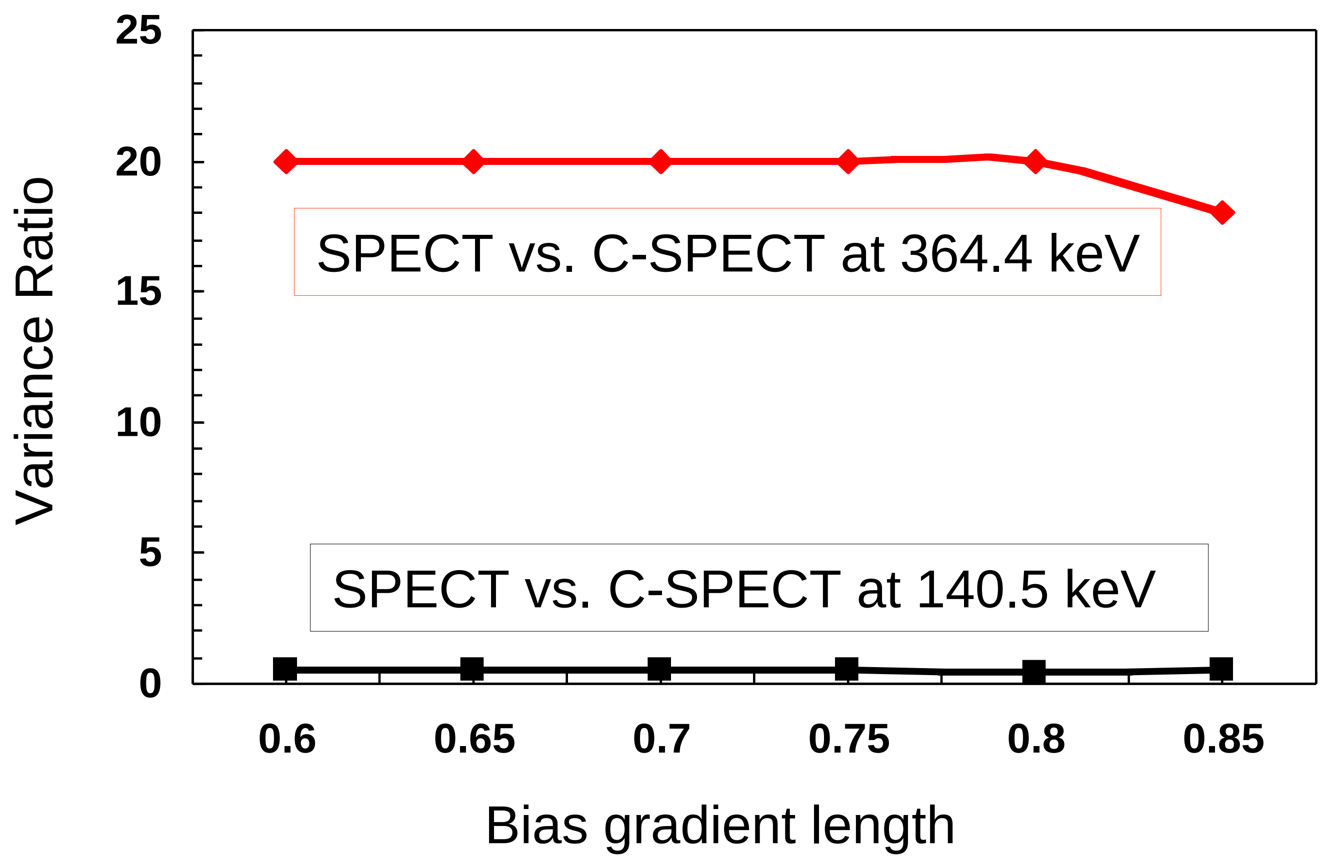


Multiple Measurements Intersecting at Source Location



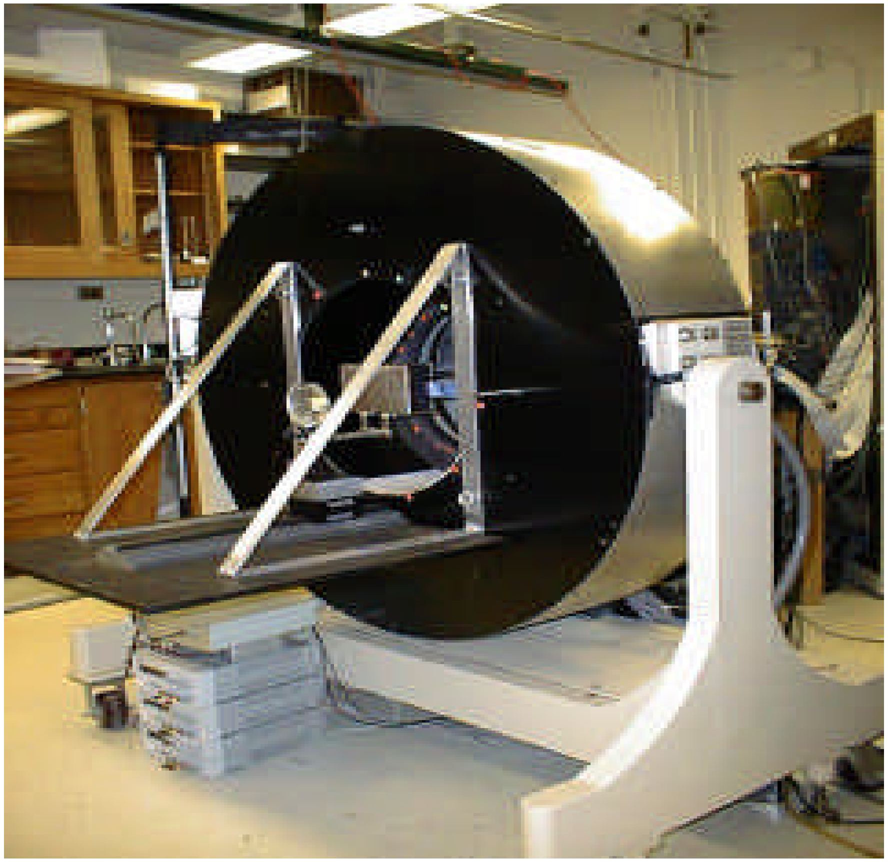
Compton Advantage

- Same imaging time (take efficiency into account)
- Assume a 9x9x0.5 cm^3 silicon 1st-detector (20x efficiency advantage)



University of Michigan Compton-SPECT System

- Silicon 1st-Detector
  - 4.5cm x 1.4cm x 0.03cm
- NaI 2nd Detector
  - 50cm diameter
  - 10cm deep
  - 11 detector modules, arranged around circumference



4. Conclusions

Tools of statistical SP have played an important role in

- design/acceleration of iterative reconstruction algorithms [Fessler&etal]

- optimal fusion of information across imaging modalities [Robinson&etal]

- benchmark studies for new imaging modalities [Clinthorne&etal]

Other areas of impact

- shape estimation (CRB, active ballons, spherical harmonics) [Robinson&etal]

- optimization of imaging subsystems (collimation, detector trajectories, etc) [Sauve&etal]

- multi-modality multi-scan image registration [Hero&etal]

- detection and confidence regions in tomographic imaging [Hero&Zhang]