Optimization in Imaging and Therapy: Iterative CBCT

Adam Wang, PhD
Assistant Professor
Department of Radiology
Stanford University

Stanford MEDICINE

adamwang@stanford.edu

AAPM Annual Meeting | San Antonio, TX | July 15, 2019

Disclosures

Shareholder of Varian Medical Systems

Research support
GE Healthcare
Siemens Healthineers
Varex Imaging
NIH U01EB023822

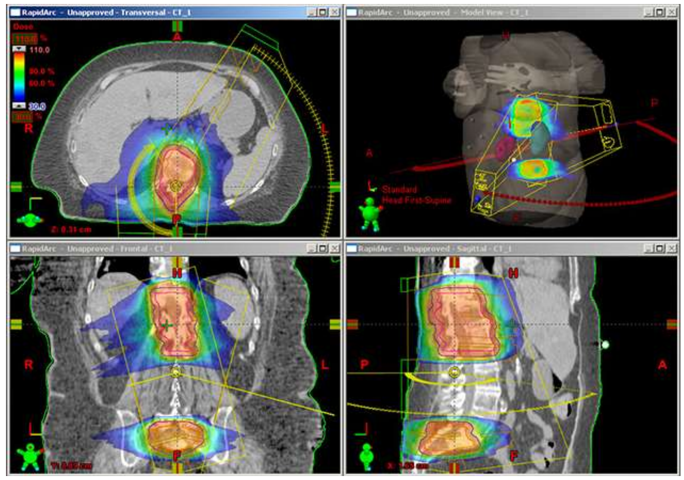
2

Radiotherapy demands optimized plans

Patient-specific treatment plan

Dose to target and organs-at-risk quantified

Very carefully planned, but is this what we deliver?



Eclipse (Varian)

3

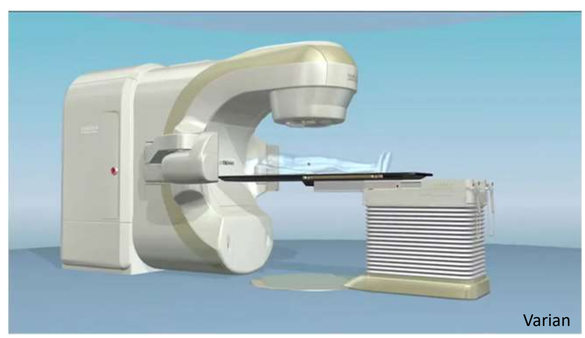
Imaging is the key to better radiotherapy

Cone-beam CT is used for patient setup, primarily based on bony anatomy

Truly optimized treatments should use:

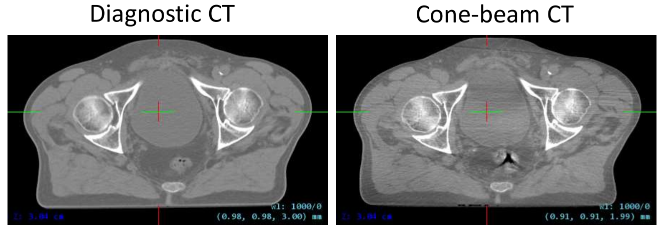
- Soft-tissue based patient positioning
- Adaptive radiotherapy
- Dose accumulation

Need high-quality CBCT images!

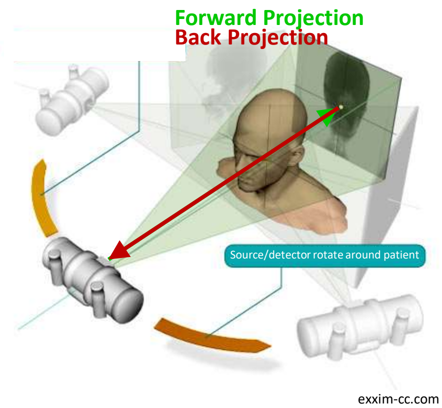


Varian

Diagnostic CT | Cone-beam CT



CBCT reconstruction basics

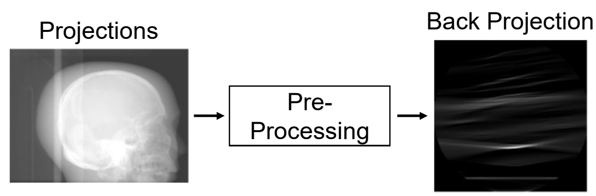


Pre-processing steps

- Scatter correction, normalization, beam hardening (spectral) correction

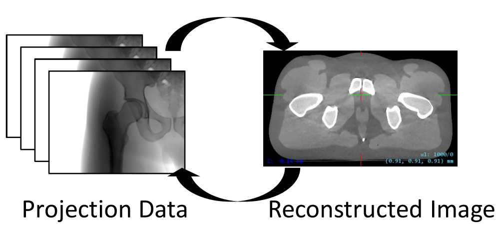
Analytic reconstruction

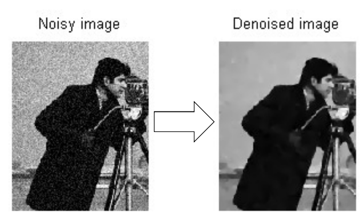
- Filtered back-projection
- Exact solution for noiseless, central axial slice
- Noise creates streaks, incomplete data causes cone-beam artifacts



Reconstruction as an optimization problem

Goal: Enforce data consistency, with image regularization





Wikipedia: Total Variation

$$\underset{\mu}{\operatorname{argmin}} \; |y - A\mu|_W^2 + R(\mu)$$

Corrected projections (scatter, beam hardening) y

Reconstructed image μ; Forward projector A

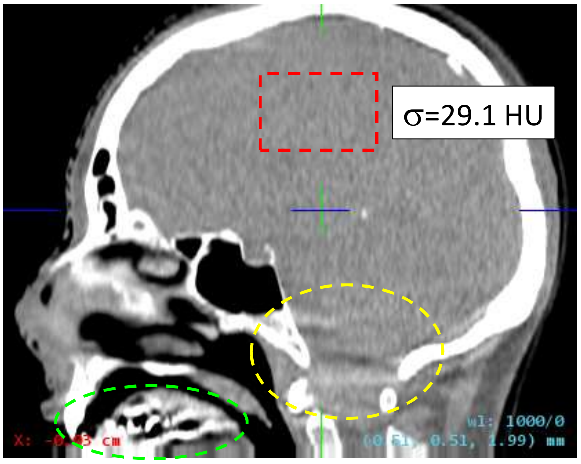
Statistical weighting W: Downweights low-fidelity, noisy rays

Image regularization $R(\cdot)$: Edge-preserving noise reduction

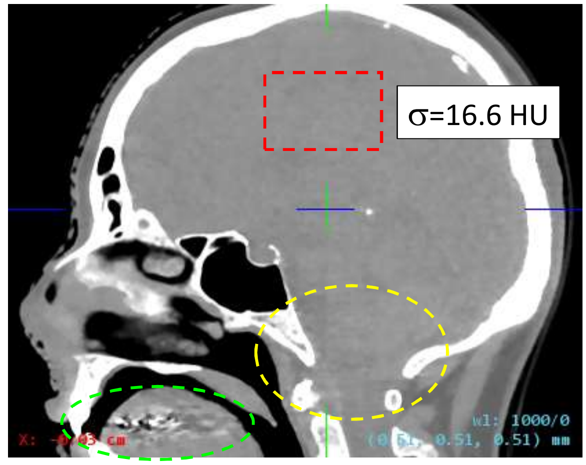
6

Head – Noise and Artifact Reduction

Conventional (FBP)

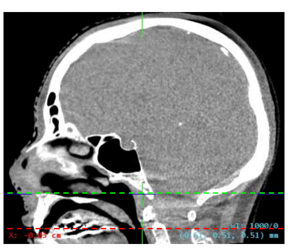


Iterative



Wang et al, AAPM 2016 [Best in Physics: Imaging]

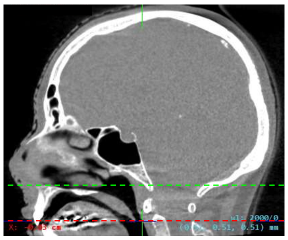
Head – Conventional Reconstruction



Cone-Beam Artifact

Metal Artifact

Head – Iterative Reconstruction



Cone-Beam Artifact

Metal Artifact

The optimization problem

Model-based image reconstruction

Basic Poisson statistical model of quantum noise

$$y_i \sim \text{Poisson}\{I_{0,i} \mathrm{e}^{-[\mathbf{A}\mu]_i}\}$$

Negative log-likelihood of all measurements (assuming independence)

$$-L(\mu; y) = \sum_i I_{0,i} \mathrm{e}^{-[\mathbf{A}\mu]_i} + y_i [\mathbf{A}\mu]_i$$

- Can get more sophisticated by adding energy dependence, scatter, electronic noise, source/detector blur, etc.

Combining likelihood and regularization = penalized likelihood (PL):

Objective Log-likelihood Regularization

$$\hat{\mu} = \underset{\mu}{\text{argmin}}\{\Phi(\mu; y) \triangleq -L(\mu; y) + \beta R(\mu)\}$$

10

Image regularization

Roughness penalty on difference between neighboring voxels

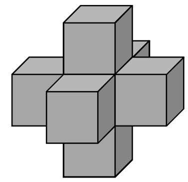
$$R(\mu) = \sum_j \sum_{k \in N} w_{jk} \Psi(\mu_j - \mu_k)$$

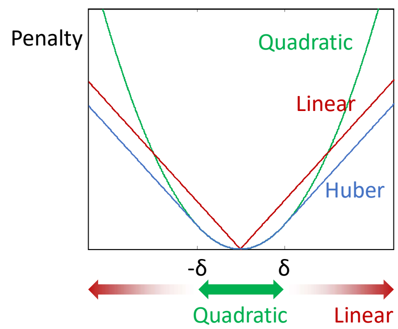
Quadratic penalty enforces smoothness throughout image

$$R(\mu) = \sum_j \sum_{k \in N} w_{jk} (\mu_j - \mu_k)^2$$

Huber penalty preserves edges with less penalty for larger differences

$$\Psi_H(x) = \begin{cases} \frac{1}{2\delta} x^2, & |x| \le \delta \\ |x| - \frac{\delta}{2}, & |x| > \delta \end{cases}$$





11

Noise-Resolution Tradeoff

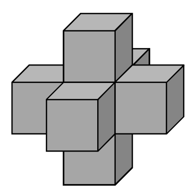
Roughness penalty on difference between neighboring voxels

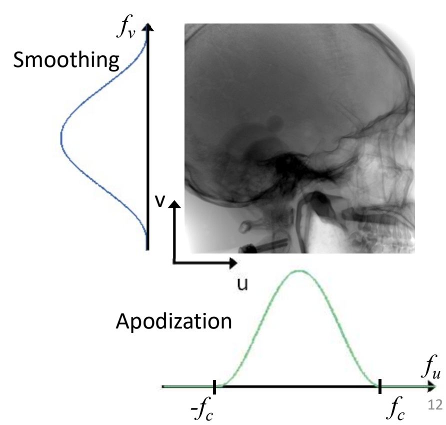
$$R(\mu) = \sum_j \sum_{k \in N} w_{jk} \Psi(\mu_j - \mu_k)$$

Filtered Backprojection

Apodization window with cutoff frequency f_c

Fair comparison should also smooth in z-direction

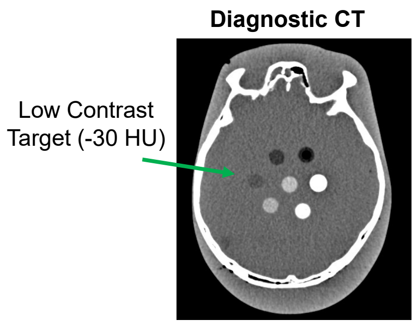


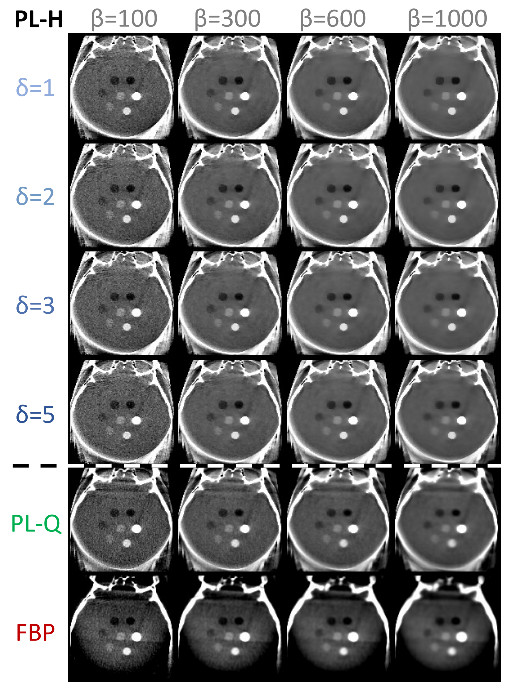


12

Noise-Resolution Tradeoff

Low contrast sphere in head phantom





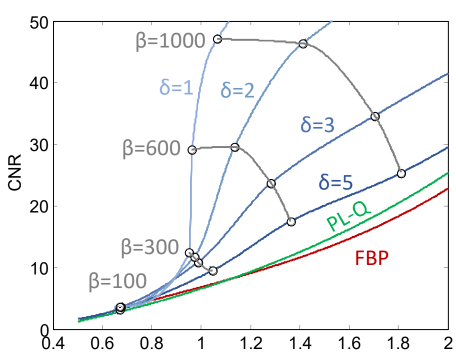
Noise-Resolution Tradeoff

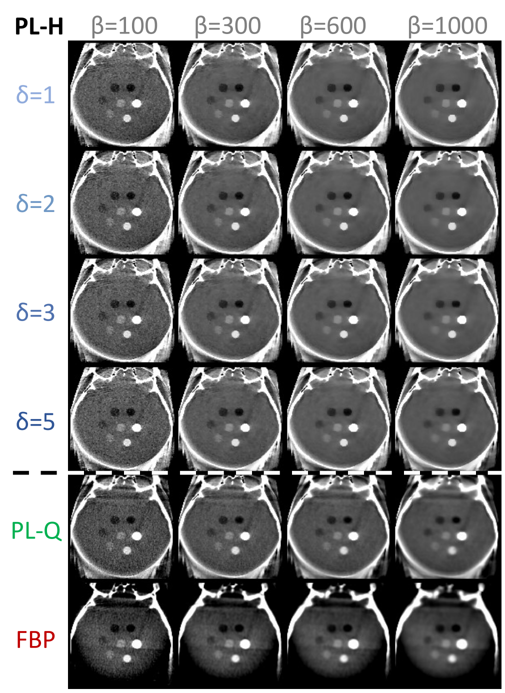
Low contrast sphere in head phantom

Increasing smoothness (L to R)

PL-Q and FBP offer similar tradeoffs

Smaller δ provides greater edge-preservation





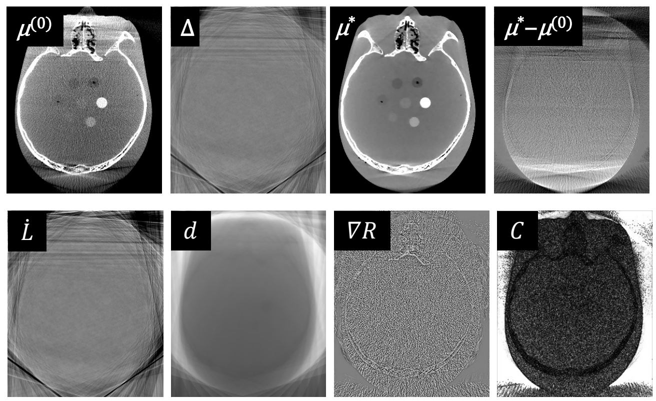
Solving the optimization problem

Popular method: Separable Quadratic Surrogates (SQS)[1] fits quadratic to objective function
Separable = update voxels simultaneously
Quadratic = minimization at each step
Surrogates = guarantees convergence

Conventional SQS

Initialize $\mu = \mu^{(0)}$

For $n = 1,2,3,\ldots,N$	iterations
1) $l = \boldsymbol{A}\mu$	forward project
2) $\dot{L} = \boldsymbol{A^T}(y - I_0 e^{-l})$	likelihood gradient
3) $d = \boldsymbol{A^T}(y \cdot c(l))$	likelihood curvature
4) $\nabla R = \sum_{k\in\mathcal{N}_j} \dot{\psi}(\mu_j - \mu_k)$	reg. gradient
$C = \sum_{k\in\mathcal{N}_j} 2\omega_\psi(\mu_j - \mu_k)$	reg. curvature
5) $\Delta = -\frac{\dot{L}+\beta\nabla R}{d+\beta C}$	compute update
$\mu := [\mu + \Delta]_+$	update image



[1]Erdogan and Fessler, Phys Med Biol 44(11) 1999.

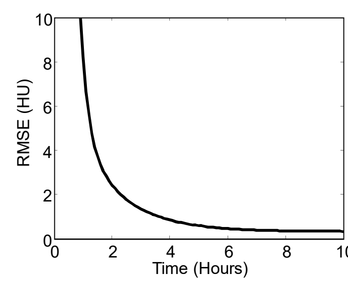
15

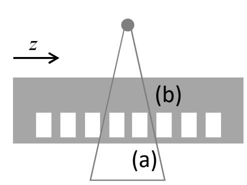
Reconstruction time

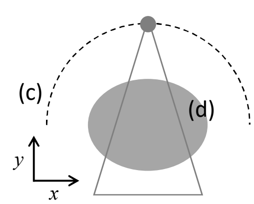
Iterative Reconstruction Tends to be Slow
- Default PL reconstruction: ~200 iterations, ~2 hrs
- FBP reconstruction: few sec

Incomplete Data Slows Convergence Further
a) Cone-beam artifacts (e.g., away from central slice)
b) Longitudinal truncation (e.g., 20 cm coverage)
c) Undersampling / incomplete orbit (e.g., ~180° orbit)
d) Lateral truncation (e.g., 26 cm FOV)

Acceleration Methods
- Ordered subsets
- Fast, GPU-based projectors
- Momentum-based methods to accelerate reconstruction time by an order of magnitude[1,2]





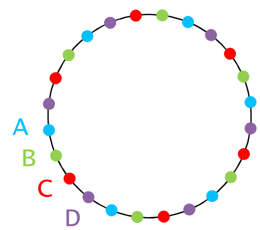


[1] Nesterov, Math. Program. Ser. A 103, 2005
[2] Kim, Ramani, and Fessler, IEEE TMI 2015

16

Ordered subsets

Conventional SQS-M	
Initialize $\mu = \mu^{(0)}$	
For $n = 1,2,3,\ldots,N$	iterations
For $m = 1,2,3,\ldots,M$	subsets
1) $l = \mathbf{A}_m \mu$	forward project
2) $\dot{L} = M[\mathbf{A}_m^T \dot{h}]$	likelihood gradient
3) $d = M[\mathbf{A}_m^T(\gamma \cdot c(l))]$	likelihood curvature
…	
Compute Δ	
$\mu := [\mu + \Delta]_+$	update image



A · C · B · D | A · C · B · D | A · C · B · D | A · C · B · D | · · ·

Each subset produces a near-equivalent update as the full set

…but with substantially reduced projection time (1/M)

Ordering helps maximize the separation between subsets

…compared to sequential or random subset order

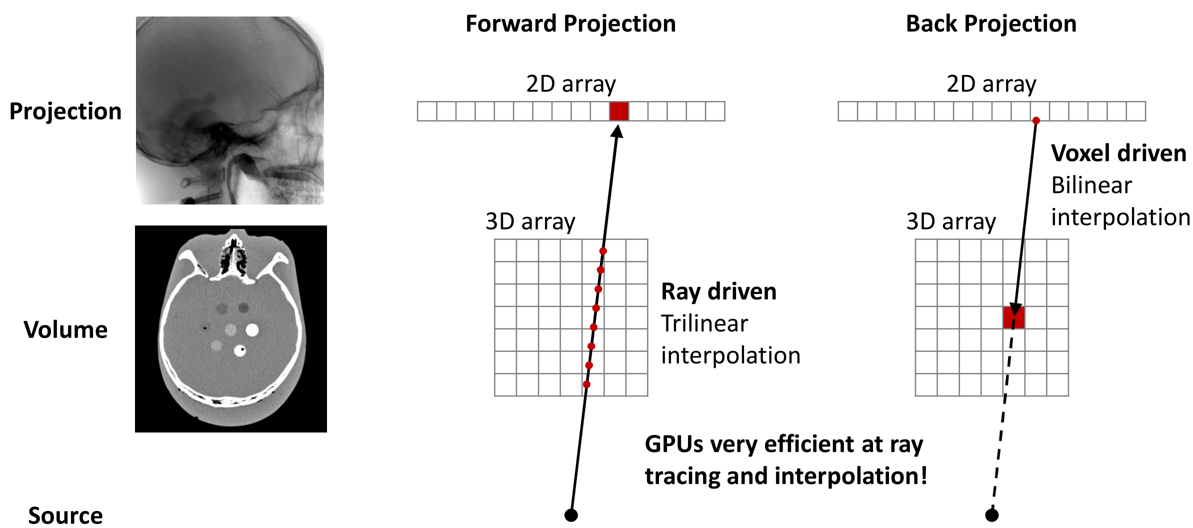
May exhibit limit-cycling behavior near converged solution

Convergence no longer guaranteed

…but rarely a problem in practice if M isn't too large

17

GPU-based forward and back projection



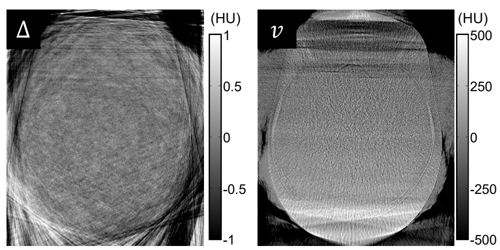
GPUs very efficient at ray tracing and interpolation!

18

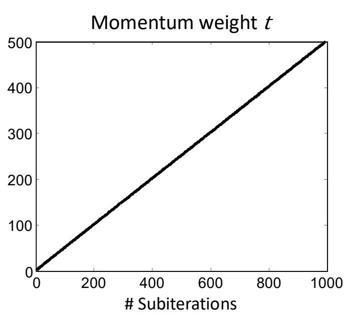
Accelerated Reconstruction

Conventional SQS-M
Initialize $\mu = \mu^{(0)}$
For $n = 1,2,3,\dots,N$
  For $m = 1,2,3,\dots,M$
    Compute Δ
    $\mu := [\mu + \Delta]_+$ update image

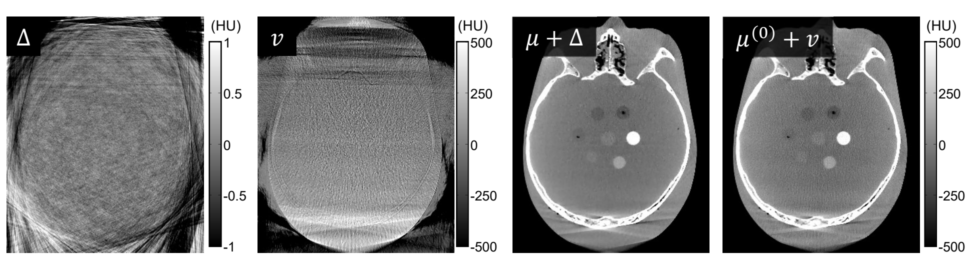
Nesterov Acceleration (Nes-M)
Initialize $\mu = \mu^{(0)}, v = 0, t = 1$
For $n = 1,2,3,\dots,N$ iterations
  For $m = 1,2,3,\dots,M$ subsets
    Compute Δ image update
    $v := v + t\Delta$ accumulated updates
    $t := (1 + \sqrt{1 + 4t^2})/2$ momentum weight
    $\mu := \left(1 - \frac{1}{t}\right)[\mu + \Delta]_+$ conventional update
      $+ \frac{1}{t}\left[\mu^{(0)} + v\right]_+$ momentum image



Accelerated Reconstruction



Nesterov Acceleration (Nes-M)
Initialize $\mu = \mu^{(0)}, v = 0, t = 1$
For $n = 1,2,3,\dots,N$ iterations
  For $m = 1,2,3,\dots,M$ subsets
    Compute Δ image update
    $v := v + t\Delta$ accumulated updates
    $t := (1 + \sqrt{1 + 4t^2})/2$ momentum weight
    $\mu := \left(1 - \frac{1}{t}\right)[\mu + \Delta]_+$ conventional update
      $+ \frac{1}{t}\left[\mu^{(0)} + v\right]_+$ momentum image



Acceleration Factor

Use objective function Φ to compare progress of algorithms A and B:

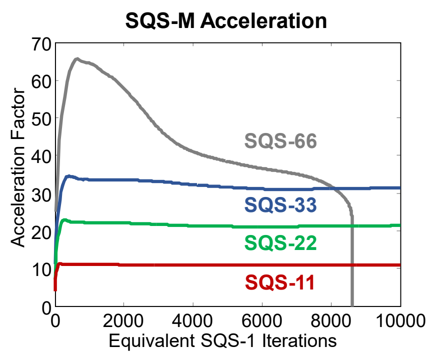
$$\min n_A \text{ s.t. } \Phi\left(\mu_A^{(n_A)}; y\right) \geq \Phi\left(\mu_B^{(n_B)}; y\right)$$

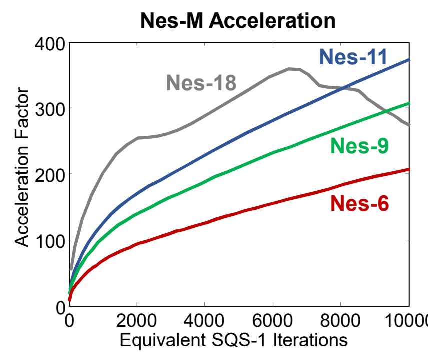
Acceleration Factor: $\mathrm{AF}(n_B) = n_B / n_A$

Baseline algorithm: SQS-1 (no subsets)

Limit-cycle / unstable when M too large

Nesterov AF increases with more iterations → Faster convergence rate than SQS





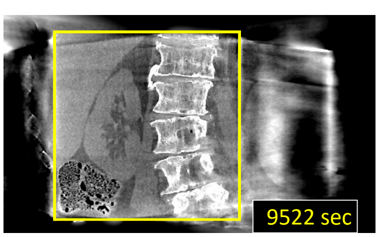
Nesterov acceleration

Momentum term accumulates all previous iterations to guide current iteration

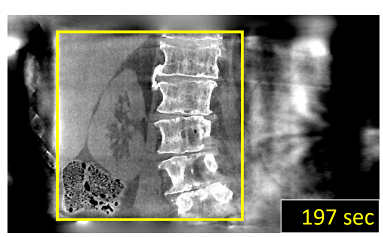
- Simple modification, very little computational/memory cost
- Accelerates convergence by >10x
- ~10 iterations sufficient

Full GPU implementation

Without acceleration



Nesterov acceleration



Wang et al, Med Phys **42** 2699 (2015)

Other optimization approaches

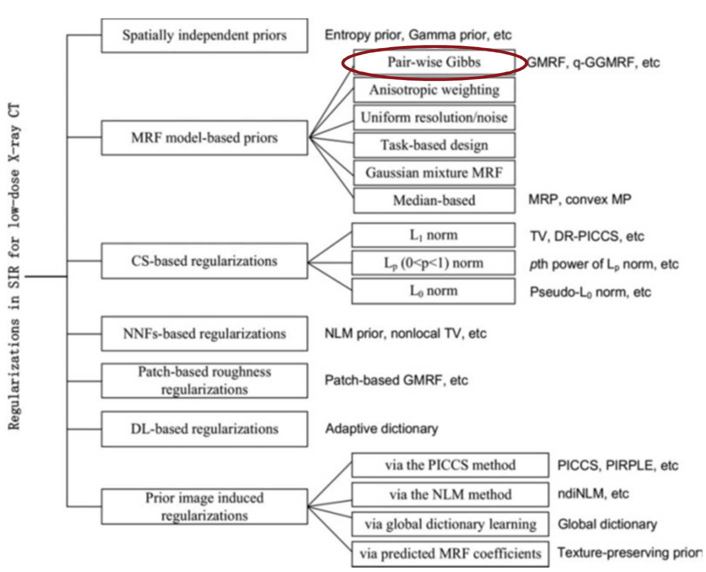
Algebraic Reconstruction Technique (ART) – the original (iterative) reconstruction!

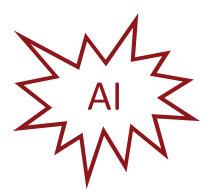
Parallel (SQS) vs serial (ICD)
Iterative coordinate descent (ICD)[1]
- Updates voxels serially
- Highly accurate updates, so only a few iterations required
- Slow on parallel hardware

Recent efforts to parallelize ICD on GPU[2]
- Voxels in z can be parallelized
- Distant voxels in axial plane are weakly coupled
- Memory ordering and access

[1] Bouman and Sauer, IEEE TIP 1996 [2] Hsieh et al, SPIE MI 2019

Other image regularization strategies





Mostly to enhance/denoise in image space …so far

Zhang et al: Regularization strategies in SIR of LDCT, Med Phys 2018

Other iterative methods

Other formulations

- ASD-POCS[1]

$$\text{argmin} \|f\|_{TV} \; s.t. |Mf - g| \leq \epsilon$$

Incorporating prior information

- PICCS[2]

$$\min_X [\alpha|\Psi_1(X - X_P)|_{\ell_1} + (1-\alpha)|\Psi_2 X|_{\ell_1}], \quad s.t. \; AX = Y$$

- dPIRPLE[3]

$$\underset{\mu,\lambda}{\text{argmax}} \log L(y;\mu) - \beta_R \|\Psi_R \mu\|_{p_R}^{p_R} - \beta_P \left\|\Psi_P(\mu - \mathbf{W}(\lambda)\mu_P)\right\|_{p_P}^{p_P}$$

[1] Sidky and Pan, Phys Med Biol **53** 4777 (2008)
[2] Chen et al, Med Phys **35** 660 (2008)
[3] Dang et al, Phys Med Biol **59** 4799 (2014)

Other iterative methods

Incorporating motion

- Motion-compensated recon[1]

$$f_{\text{MoCo}(i)} = \sum_{j=1}^{N} \left(\mathsf{X}^{-1}_{\text{PCF}(j)}\, p\right) \circ \mathrm{T}^i_j.$$

Incorporating energy information

- Spectral PICCS[2]

$$f_c(\underline{x}) = c\|\underline{x}\|_{\text{TV}} + (1-c)\left\|\underline{x} - \underline{x}_{\text{p}}\right\|_{\text{TV}}$$

Incorporating correlated measurements

- Gaussian Penalized-Likelihood with Blur and Correlations (GPL-BC)[3]

… and many more!

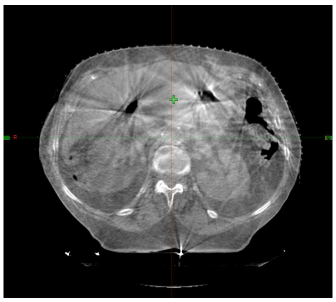
$$\psi = \frac{1}{2}\left(y - \mathbf{B}e^{-\mathbf{A}\mu}\right)^T \mathbf{W}\left(y - \mathbf{B}e^{-\mathbf{A}\mu}\right) + \beta\,\mathrm{R}(\mu)$$

[1] Brehm et al, Med Phys **39** 7603 (2012)
[2] Yu et al, Phys Med Biol **61** 6707 (2016)
[3] Tilley et al, IEE TMI **37** 988 (2017)

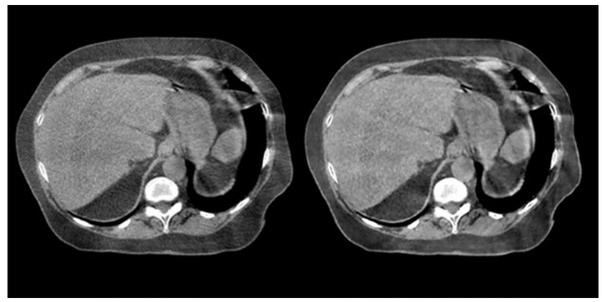
Iterative recon isn't a silver bullet!

Sometimes easier to address the underlying problem, e.g., motion

Free breathing



Breath hold



Standard CBCT   Iterative CBCT

Cai et al, Med Phys 46(3) 2019

Summary

Iterative CBCT benefits:

- Noise reduction
- Reduced cone-beam artifacts
- Improved HU accuracy

Computational methods include:

- GPU implementation
- Ordered subsets
- Nesterov momentum

Many different approaches/flavors, but generally model imaging system, x-ray physics, and prior knowledge

adamwang@stanford.edu