TV-Regularization in Tomography
Part 2: Algebraic Reconstruction Technique

Philipp Lamby

Interdisciplinary Mathematics Institute,
University of South Carolina

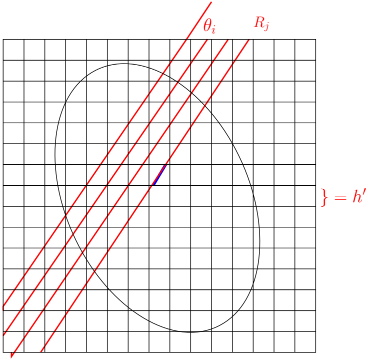
Peter Binev, Wolfgang Dahmen, Ronald DeVore,
Philipp Lamby, Andreas Platen, Dan Savu, Robert Sharpley

Summary of Part 1

- Limited Angle Tomography is a severly ill-posed problem.
- Reconstruction of edges parallel to missing views is unstable.
- Regularization is required.
- We use the TV-norm as regularizing term.
- Regularized equally sloped tomography is a possible approach.
  - But it requires an accurate data aquisition process,
  - and uses the projection-slice theorem,
  - i.e., the full Fourier-information is required.
- We want to use the algebraic reconstruction technique.

Outline

ART: Discretization



- Let $\alpha = (\alpha_1, \alpha_2)$ be the index of the pixel $\Box_\alpha$,
- $w_{(\theta_i,j),\alpha} = |R_j^{\theta_i} \cap \Box_\alpha|$ be the length of the intersection of ray $R_j^{\theta_i}$ with pixel α,
- and $g(\xi,\eta)$ be the distribution sought for:

$$g \approx \sum_\alpha c_\alpha \chi(\Box_\alpha)$$

Ray Integrals - Measurements

$$d_j^{\theta_i} = \int_{R_j^{\theta_i}} g(\xi,\eta) \approx \sum_\alpha w_{(\theta_i,j),\alpha} c_\alpha$$

ART Linear System

Linear System

- Discretization leads to a linear system

$$Wx \approx \mathbf{d}.$$

Properties of W

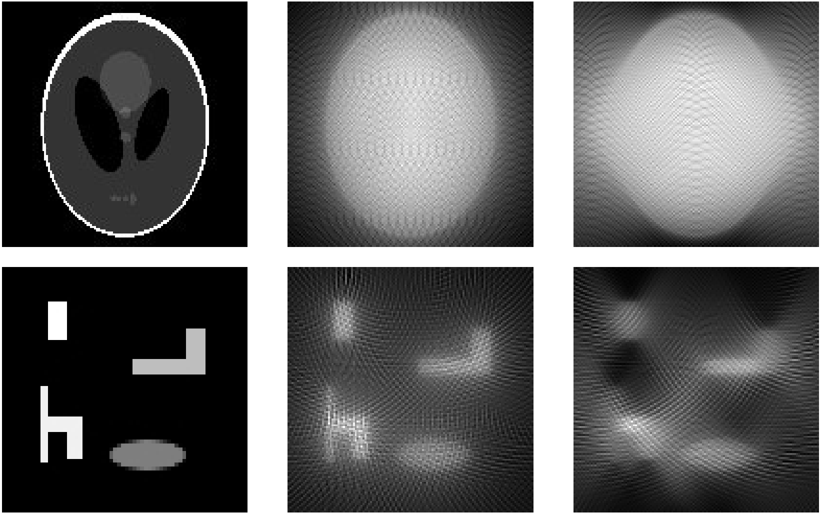
1. W is sparse: each row contains at most 2N non-zero entries.
2. In our case the system will be underdetermined:
   - To be reconstructed: a 1024×1024 image
   - Data: 149 projections, 1024 pixels each.
3. W will be ill-conditioned, rank deficient.

Singular Values of W in Limited Angle Tomography

#Views	h	θ_{max}	#Rays	#Pixels	$\#\sigma_k = 0$
30	0.02	90	1918	64^2	26
30	0.02	60	1962	64^2	38
60	0.02	90	3824	64^2	174
60	0.02	60	3844	64^2	176
90	0.02	60	5888	128^2	158

Discrete Backprojection

Reminder: The adjoint operator "almost" inverts the Radon transform. Let's try $x = W^T d$. (Original, DBP 90^o, 60^o)



Tradititional Solution Method: Kaczmarz' Algorithm

Basic Idea:

- Each row of the system defines a hyper-plane $\langle W_j, x\rangle = d_j$.
- Iteratively project the solution vector onto the hyperplanes:

$$P_j(x) = x - \frac{\langle x, W_j\rangle - d_j}{\langle W_j, W_j\rangle} W_j,$$

- $x^{k+1} = P(x^k) = P_n \circ P_{n-1} \circ \cdots \circ P_2 \circ P_1(x^k)$

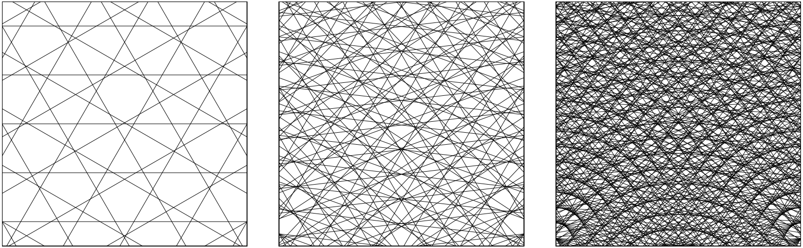
Theoretical Result:

- If $rank(W) > 2$, then iteration converges and

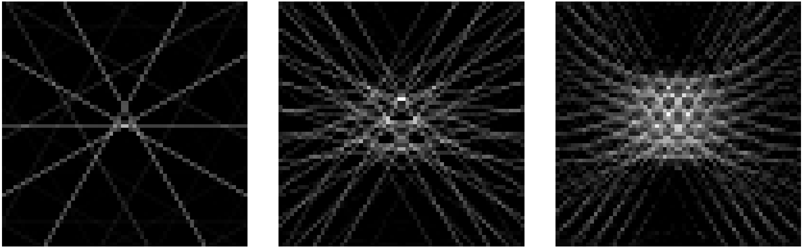
$$\lim_{k\to\infty} \mathbf{x}^k = \mathrm{Proj}_{N(W)}(\mathbf{x}^0) + G\mathbf{d}$$

where G is some kind of pseudo-inverse of W.

Least Squares Solutions, Phantom 1, $\theta_{max} = 60^\circ$



$n_a = 5,\ h = 1/5$ $n_a = 10,\ h = 1/10$ $n_a = 20,\ h = 1/16$



Regularization Approaches

- Unconstrained:

$$c_1^* = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{Wx} - \mathbf{d}\|_{l_2}^2 + \mu\|\mathbf{x}\|_{TV}, \qquad (1)$$

- Constrained:

$$c_2^* = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_{TV} \quad \text{s.t} \quad \|\mathbf{Wx} - \mathbf{d}\|_{l_2} \leq \varepsilon, \qquad (2)$$

- Selection of least squares solution:

$$c_3^* = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_{TV} \quad \text{s.t.} \quad \|\mathbf{Wx} - \mathbf{d}\|_{l_2} = \min \qquad (3)$$

- Selection for consistent systems:

$$c_4^* = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_{TV} \quad \text{s.t.} \quad \mathbf{Wx} = \mathbf{d} \qquad (4)$$

Kaczmarz with Trick

Theorem

Let $\sum_{k=1}^{\infty} \beta_k < \infty$ and $\|\mathbf{v}_k\| \leq 1$, $k = 1, 2, \ldots$. Then the sequence generated by the perturbed Kaczmarz iteration

$$x^{k+1} = P(x^k + \beta_k v^k)$$

still converges to a least squares solution, "possibly" to c_3^*.

Algorithm

- $\beta_1 = 1$, $\phi(x) = \|x\|_{TV}$.
- For $k = 1, 2, \ldots$ do
  1. Determine $v_k' \in \partial\phi(x^k)$. Set $v_k = v_k' / \|v_k'\|$.
  2. $x^{k+1} = P(x^k + \beta_k v^k)$.
  3. $\beta_{k+1} = c\beta_k$, $0 < c < 1$.

Remarks on Kaczmarz with Tricks

- Source: G.T. Herman, R. Davidi, Inverse Problems 2008
- "Tricks" are a standard technique in ART, usually used to enforce range constraints, e.g.

$$x^{k+1} = P(\text{compmin}(255, \text{compmax}(0, x^k)))$$

- The advantage of the method is, that one uses a reliable solver for the data term.
- The choice of β_k is critical:
  - If β_k decreases to fast, the result does not minimize ϕ.
  - But convergence is slow, if β_k decreases slowly.

SESOP

- SESOP = Sequential Subspace Optimization.
- Solves the unconstrained problem (1).

Algorithm

- Let $f(x)$ be a differentiable, convex functional.
- In our case: $f(x) = \frac{1}{2}\|Wx - d\|_2^2 + \psi\|Bx\|_1$.
- Choose an initial solution $x^0 \in \mathbb{R}^M$.
- For $k = 1, 2, \ldots$ do
  - Choose basis D of r-dimensional subspace of $\mathbb{R}^M$.
  - $\alpha^* = \arg\min_{\alpha \in \mathbb{R}^r} f(x_k + D\alpha)$.
  - $x^{k+1} = x^k + D\alpha^*$

SESOP Details

- $B \in \mathbb{R}^{N \times N}$ is defined by

$$\|Bx\|_1 = \|x\|_{TV} = \sum_{i,j} |x_{ij} - x_{i,j-1}| + |x_{ij} - x_{i-1,j}|.$$

- ψ is a smoothed absolute value, e.g. $(\varepsilon > 0)$

$$\psi(t) = \varepsilon \left(\left| \frac{t}{\varepsilon} \right| + \frac{1}{\left| \frac{t}{\varepsilon} \right| + 1} - 1 \right).$$

- The crucial optimization parameter is the choice of the subspaces.

Choice of the SESOP Subspaces

- Current gradient:

$$\nabla f(x_k)$$

- Nemirowski directions:

$$d_k^{(1)} = x^k - x^0, \quad d_k^{(2)} = \sum_{i=0}^{k} w_i \nabla f(x^i)$$

- Previous directions and gradients: $(i = 1, 2, \ldots, k)$

$$p_{k-1,i} = x^{k-i} - x^{k-i-1}, \quad \nabla f(x^{k-i})$$

- Conjugated gradients, etc.

SESOP Convergence

Theorem

Let $f(x)$ be a smooth, convex functional with L-Lipschitz continuous gradient. Let the subspaces in SESOP be spanned by the current gradient and the Nemirowski directions and possibly several other vectors. Then

$$\|f(x^{k+1}) - f(x^*)\| \leq \frac{LR^2}{k^2}$$

where $\|x^* - x_0\| \leq R$.

Remark

- Andreas currently prefers a choice of $7 - 11$-dim. subspaces.
- Method seems to work reasonably well with ART matrices.

Operator Splitting

- Solves the unconstrained problem (1).

Reminder: the Rudin-Osher-Fatemi Denosing Model

$$ROF_{\mu}(x) = \arg\min_{x} \frac{1}{2}\|Wx - d\|_2^2 + \mu\|x\|_{TV}$$

Algorithm

- Initialize $k = 0$, x^0.
- Choose τ, such that $\|I - \tau W^T W\| < 1$.
- For $k = 1, 2, \ldots$ do
  - $y_k = x_k - \tau W^T(Wx - d)$
  - $x_{k+1} = \mathrm{ROF}_{\mu/\tau}(y_k)$

Remarks on Operator Splitting

- Method is an alternating method, like Trick-Kaczmarz.
- Convergence can be proved under conditions (etc. RIP).
- As solver for the linear part, *Richardson iteration* is applied to the normal equations of the LS problem.

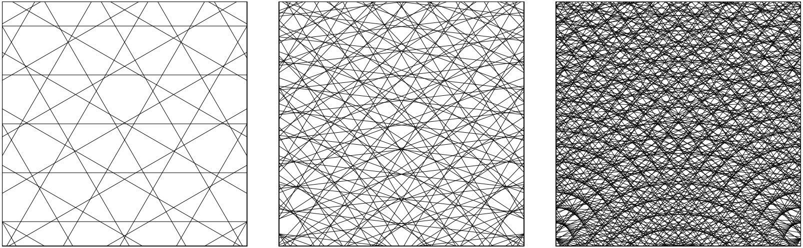
Richardson Iteration

- Solves $Ax - b$ with the iteration $x^{k+1} = x^k + \tau(b - Ax^k)$.
- The error after the k-th iteration is

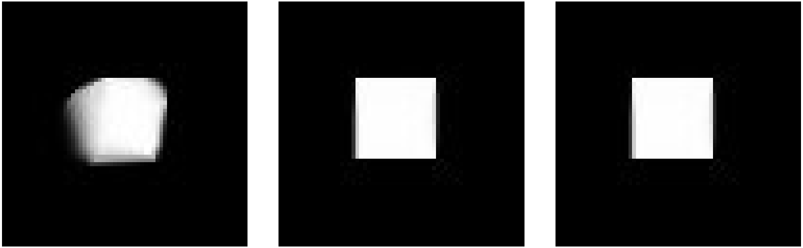
$$e^{k+1} = x^{k+1} - x^* = (I - \tau A)e^k$$

- The error converges to 0, if $\|I - \tau A\| < 1$.
- Since $A = W^T W$ is rank-deficient, no τ can achieve that.

Numerical Results - Phantom 1 - NESTA, $\theta_{max} = 60^\circ$



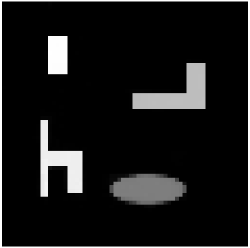
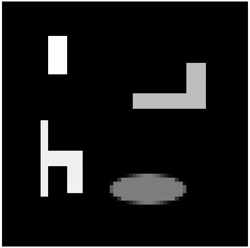
$n_a = 5,\ h = 1/5$ | $n_a = 10,\ h = 1/10$ | $n_a = 20,\ h = 1/16$



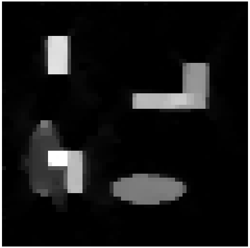
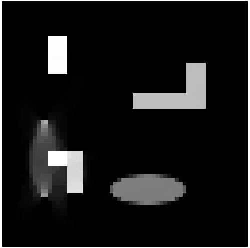
Phantom 3, TV-Kaczmarz

$TV(orig) = 44820$

$\theta = 90^\circ$
l: 510 Rays
r: 2046 Rays

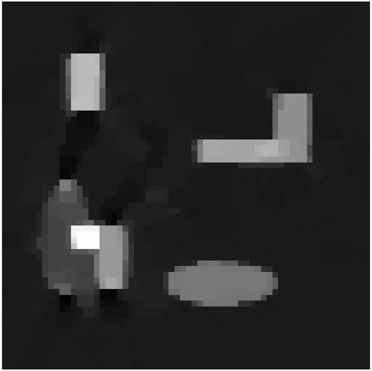



$n_a = 20,\ h = 1/20$ $n_a = 40,\ h = 1/40$

$\theta = 60^\circ$
l: 524 Rays
r: 2096 Rays
l: $TV = 42506$
r: $TV = 43562$

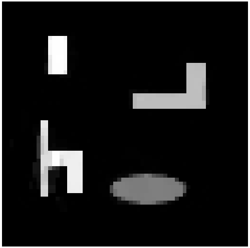
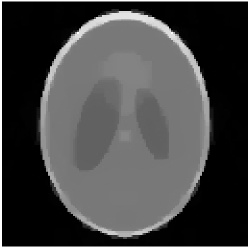
Range Constraints



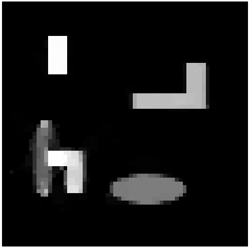
Left: Sesop-Reconstruction without constraints. Pixel Values vary in a range from -33 to 304. Therefore we add an additional penalty term to SESOP.

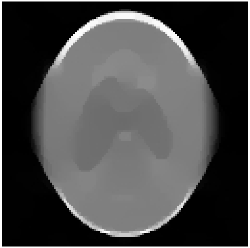
$$K(f) = \mu_2 \sum_{\text{Pixel } i} (\min(f(i), 0))^2 + (\max(f(i), 255) - 255)^2$$

Some SESOP Results


$n_a = 20, h = 1/40$ $n_a = 22, h = 1/40$ $n_a = 40, h = 1/40$





Tentative Conclusions

- TV-Kaczmarz sometimes reproduce phantoms exactly.
- But SESOP is the faster, and more reliable code.
- For limited tilt angles unconstrained optimization seems to work better than constrained optimization.
- TV-regularization should be supported by pixel-value constraints.

Current Questions

- Repeat the experiments with noise.
- Can one improve stability by choosing another basis?
- Do results improve with a non-convex regularization term?