Variational methods for image reconstruction in MR imaging

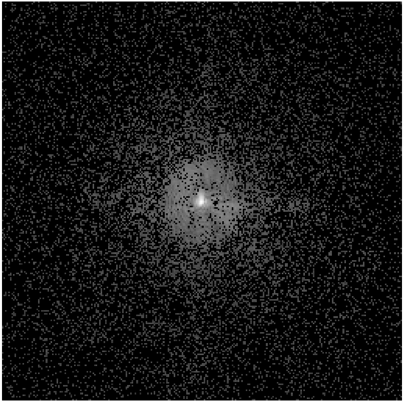
Christian Clason[1]

joint work with Kristian Bredies[1] and Florian Knoll[2]

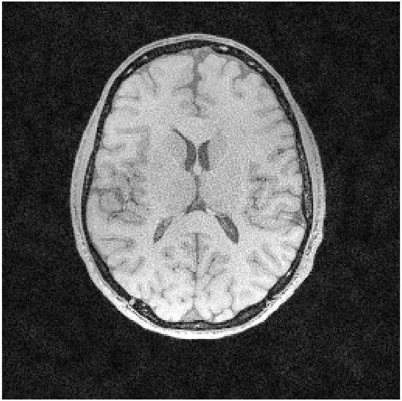
[1]Institute for Mathematics and Scientific Computing, Karl-Franzens-Universität Graz
[2]Institute for Biomedical Engineering, Technische Universität Graz

Dagstuhl Seminar “Efficient Algorithms for Global Optimisation Methods in Computer Vision”
Schloss Dagstuhl, November 21, 2011

Motivation

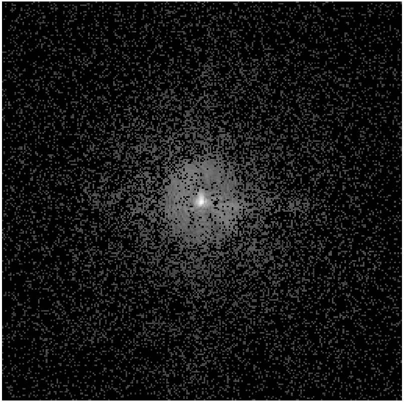


measured data

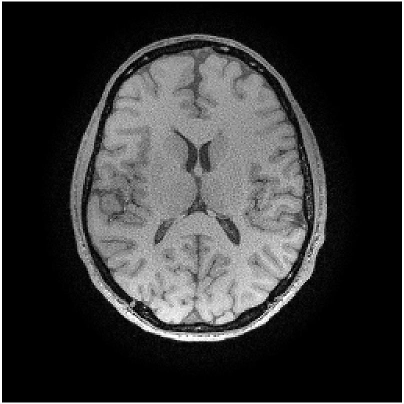


linear reconstruction

Motivation

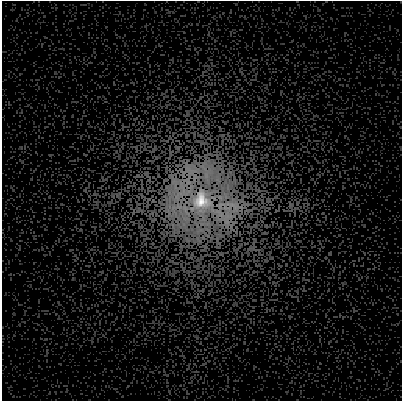


measured data

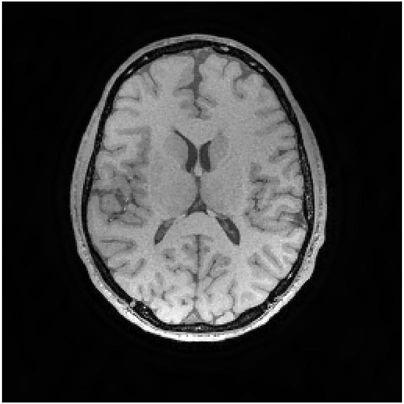


nonlinear reconstruction

Motivation



measured data



nonlinear TV-reconstruction

MR in a nutshell

Basic steps in MR scan:

1. Magnetic field is applied, aligns proton spins
2. Radio pulse at resonance frequency is absorbed by hydrogen nuclei, re-radiated over time at same frequency
3. Decaying time-dependent signal is measured by receiver coil

Fundamental principles of MR:

- Signal **amplitude** proportional to **hydrogen density**
- Signal **frequency** proportional to **magnetic field strength**

Spatial encoding

Problem:

- Measured time-dependent signal is **composite** over whole volume, no spatial information

Solution:

- Use **spatially dependent magnetic fields** to map resonance frequency to spatial location

But: linear superposition of (x, y, z) fields is not unique

$\rightsquigarrow$ **Sequential application** of fields to encode (x, y, z) separately

Spatial encoding

Slice selection (z):

- Use z-proportional magnetic field during RF **excitation**
- $\rightsquigarrow$ Only thin slice has resonance at RF pulse frequency, contributes to measured signal

Frequency encoding (x):

- Use x-proportional magnetic field **during measurement**
- $\rightsquigarrow$ Signal is superposition of resonance **frequencies**
- $\rightsquigarrow$ Each frequency can be identified with x-coordinate
- **Fourier transform** in time $\rightsquigarrow$ amplitude for each frequency $=$ spin density at each x-coordinate (integrated over y)

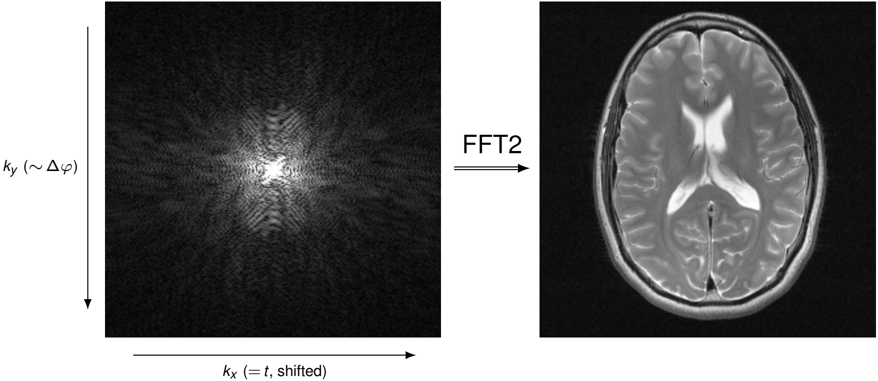
Spatial encoding

Phase encoding (y):

- Use y-proportional magnetic field **before measurement** to change **phase** of signal
- Phase cannot be measured directly, but **phase difference** can
- $\rightsquigarrow$ **Repeat measurements** with linear fields of increasing slope: $G_y = n\, y\, G_0$, $n = -N, \dots, 0, \dots N$
- $\rightsquigarrow$ Phase difference is y-specific: y larger $\Rightarrow$ gradient change, phase change with increasing n larger
- **Fourier transform** in $n \rightsquigarrow$ amplitude for each phase difference ($\sim y$-coordinate)

MRI: reconstruction

stack of measured data ("k-space") image ($\sim$ spin density)



Fourier transform of measured data gives image

Parallel imaging

Problem:

- Data needs to be acquired **sequentially** (for each k_y, z)
- Repetition cannot be faster than **physical limit**
- $\rightsquigarrow$ 4 minutes for single 256×256 slice

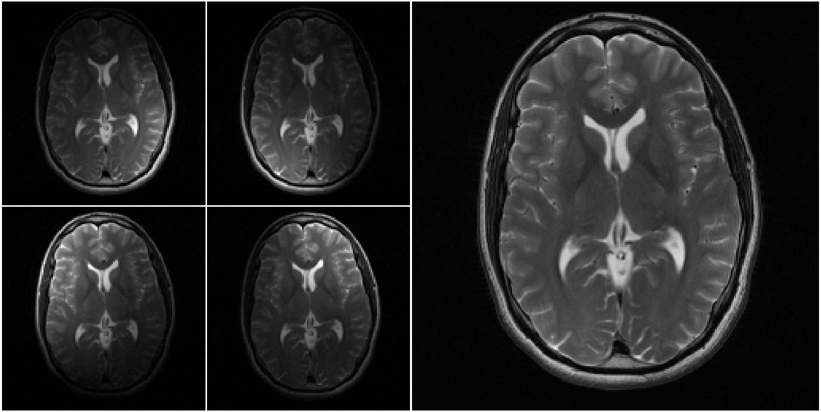
Solution:

- Acquire **fewer k_x lines**, e.g., skip every second line

But: subsampling in k_y $\rightsquigarrow$ **aliasing artifacts**

$\rightsquigarrow$ **Independent measurements** using multiple receiver coils with different **spatial sensitivities**

Example of coil sensitivities



Parallel imaging

Standard approach: if

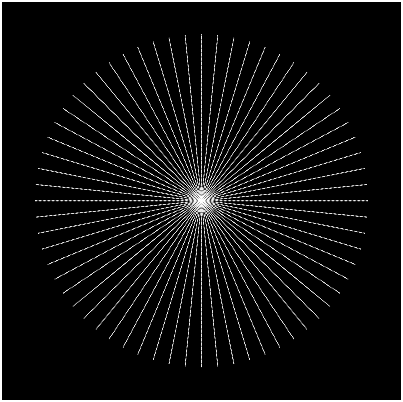
- sufficient number of coils
- coil sensitivities (approximately) known

$\rightsquigarrow$ Solve overdetermined linear system for each pixel

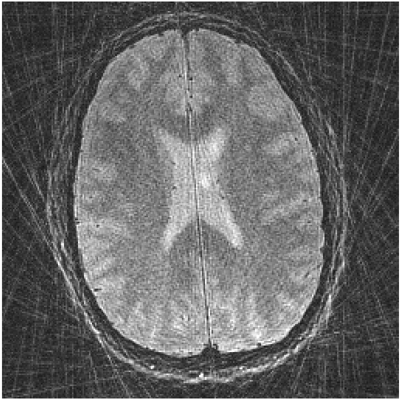
But: reconstruction quality depends on

- accurate estimation of sensitivities
- structure of subsampling artifacts

Subsampling example: Radial

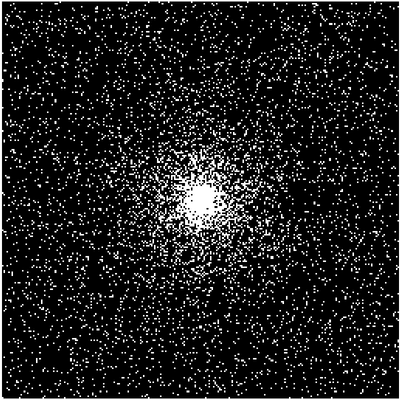


(a) pattern (acceleration factor $R \approx 12$)



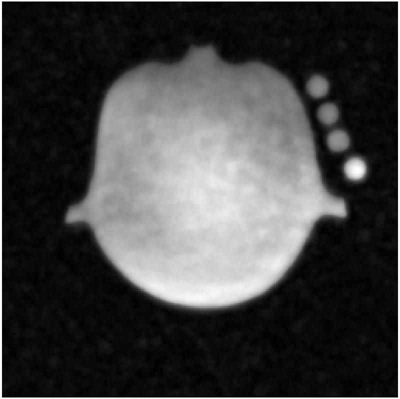
(b) aliasing artifacts

Subsampling example: Random Cartesian



(c) pattern (acceleration factor $R = 10$)

(k_x orthogonal to image plane, phase encoding in y, z)



(d) aliasing artifacts

Parallel imaging

Variational formulation allows including

- coil **sensitivity correction**
- **filtering** (artifact removal)
- arbitrary **sampling strategy**, penalty

in reconstruction

Parallel MRI as inverse problem

Given

- subsampling operator $\mathcal{F}_s$ (IFFT + mask, NFFT)
- acquired k-space coil data $g = (g_1, \dots, g_N)^T$

Find

- image u
- coil sensitivities $c = (c_1, \dots, c_N)^T$

such that

$$F(u, c) := (\mathcal{F}_s(u \cdot c_1), \dots, \mathcal{F}_s(u \cdot c_N))^T = g$$

Nonlinear inverse problem, ill-posed $\rightsquigarrow$ solve using IRGN method

Iteratively regularized Gauß–Newton method

1: choose $x^0 = (u^0, c^0)$, α_0, $q < 1$
2: **repeat**
3: solve for $\delta x = (\delta u, \delta c)$ (e.g., by CG on normal equations)

$$\min_{\delta x} \frac{1}{2}\|F'(x^k)\delta x + F(x^k) - g\|^2 + \frac{\alpha_k}{2}\|W(c^k + \delta c)\|^2 + \frac{\alpha_k}{2}\|u^k + \delta u\|^2$$

4: set $x^{k+1} = x^k + \delta x$, $\alpha_{k+1} = \alpha_k q$, $k = k + 1$
5: **until** $\|F(x^k) - g\| < tol$

W high-order differential operator (enforces smooth sensitivities)

[Uecker/Hohage/Block/Frahm '08]

IRGNTV

Replace L^2 penalty on u^{k+1} with **total variation** $TV(u^{k+1})$:

1: choose $x^0 = (u^0, c^0)$, α_0, β_0, $q < 1$
2: **repeat**
3: solve for $\delta x = (\delta u, \delta c)$

$$\min_{\delta x} \frac{1}{2}\|F'(x^k)\delta x + F(x^k) - g\|^2 + \frac{\alpha_k}{2}\|W(c^k + \delta c)\|^2 + \beta_k TV(u^k + \delta u)$$

4: set $x^{k+1} = x^k + \delta x$, $\alpha_{k+1} = \alpha_k q$, $\beta_{k+1} = \beta_k q$, $k = k + 1$
5: **until** $\|F(x^k) - g\| < tol$
6: **return** u, c

Solution of TV subproblems

Set $J(\delta x) := \frac{1}{2}\|F'(x^k)\delta x + F(x^k) - g\|^2 + \frac{\alpha_k}{2}\|W(c^k + \delta c)\|^2$

Step 3

$$\min_{\delta u, \delta c} J(\delta u, \delta c) + \beta_k TV(u^k + \delta u)$$

Non-smooth, convex optimization problem $\rightsquigarrow$ Use **convex duality**

$$\beta TV(u) = \sup_{\{|p(x)|_2 \leq \beta\}} \langle u, -\operatorname{div} p \rangle$$

Solution of TV subproblems

Saddle point problem

$$\min_{\delta u, \delta c} \max_{p \in C_{\beta_k}} J(\delta u, \delta c) + \langle u^k + \delta u, -\operatorname{div} p \rangle$$

with $C_\beta = \{p : \ |p(x)|_2 \leq \beta \text{ for all } x\}$ convex, J differentiable

$\rightsquigarrow$ Use **projected gradient descent/ascent method**:

- Requires only application of F', F'^* (i.e., $\mathcal{F}_s, \mathcal{F}_s^*$)
- Projection on C_β can be computed pointwise
- Straightforward parallelization (GPU)

Here: based on [Pock/Cremers/Bischof/Chambolle '09]

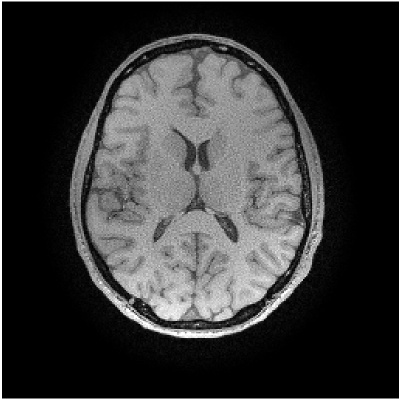
Primal-dual extragradient method

function TVSOLVE$(u, c, \alpha, \beta, \sigma_u, \sigma_c, \tau)$
  $\delta u, \bar{\delta u}, \delta c, \bar{\delta c}, p \leftarrow 0$
  repeat
    $p \leftarrow \text{proj}_\beta(p + \tau\nabla(u + \bar{\delta u}))$
    $\delta u_{old} \leftarrow \delta u, \; \delta c_{old} \leftarrow \delta c$
    $\delta u \leftarrow \delta u - \sigma_u(\partial_u J(u,c)(\bar{\delta u}, \bar{\delta c}) - \text{div}\, p)$
    $\delta c \leftarrow \delta c - \sigma_c(\partial_c J(u,c)(\bar{\delta u}, \bar{\delta c}))$
    $\bar{\delta u} \leftarrow 2\delta u - \delta u_{old}$
    $\bar{\delta c} \leftarrow 2\delta c - \delta c_{old}$
  until convergence
  return $\delta u, \delta c$
end function

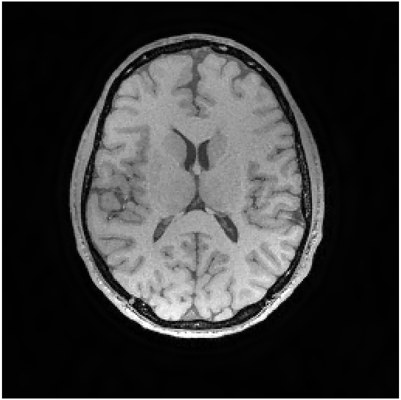
Examples: random sampling

- Raw data fom brain and phantom
- 3D gradient echo sequence, 3T system, 12 channel head coil
- 8 (phantom: 9) virtual channels (SVD) used for reconstruction
- Sequence modified using binary 2D mask for subsampling
- Subsampling $R = 4$ (10)
- Sequence parameters
  - repetition time TR $= 20\,\text{ms}$
  - echo time TE $= 5\,\text{ms}$
  - flip angle FA $= 18^\circ$
  - matrix size $(x, y, z) = 256 \times 256 \times 256$
  - FOV $= 250\,\text{mm}$
  - slice thickness brain 1 mm (phantom 5 mm)

Reconstructions: random ($R = 4$)

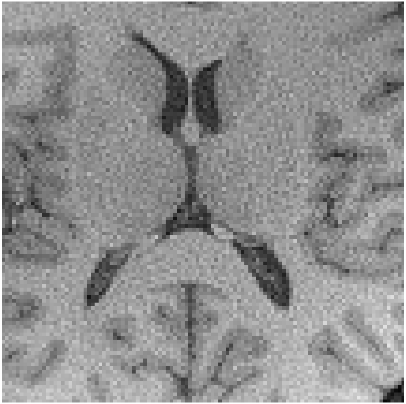


(a) IRGN



(b) IRGNTV

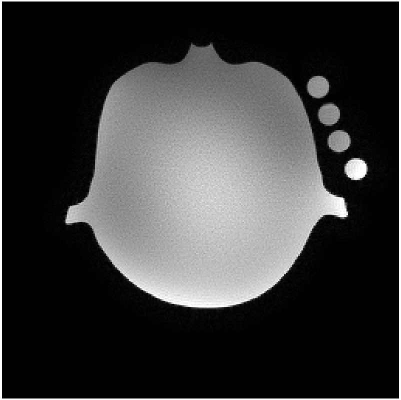
Reconstructions: random ($R = 4$)



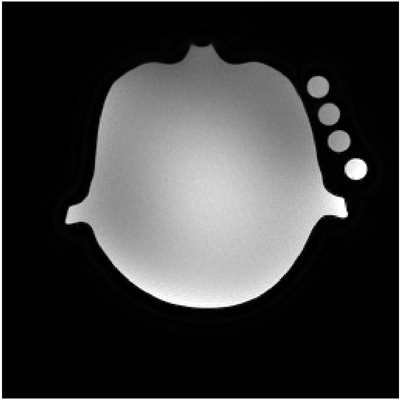
(a) IRGN (detail)

(b) IRGNTV (detail)

Reconstructions: random ($R = 4$)



(a) IRGN



(b) IRGNTV

Effect of TV

Since stopping $\beta_{k^*} > 0$ small, final TV effect is not very strong

Pro: No introduction of typical TV-artifacts (cartooning, stair-casing)

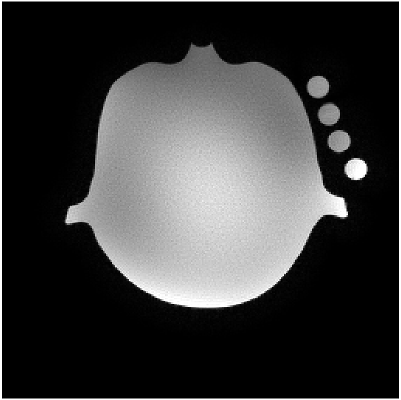
Con: More filtering can be desired for higher acceleration

$\rightsquigarrow$ Stop decreasing TV penalty parameter at desired value:

$$\alpha_{k+1} = \alpha_k q$$
$$\beta_{k+1} = \max(\beta_{\min}, \beta_k q)$$

For illustration: phantom with $\beta_{\min} = 5 \cdot 10^{-3}$

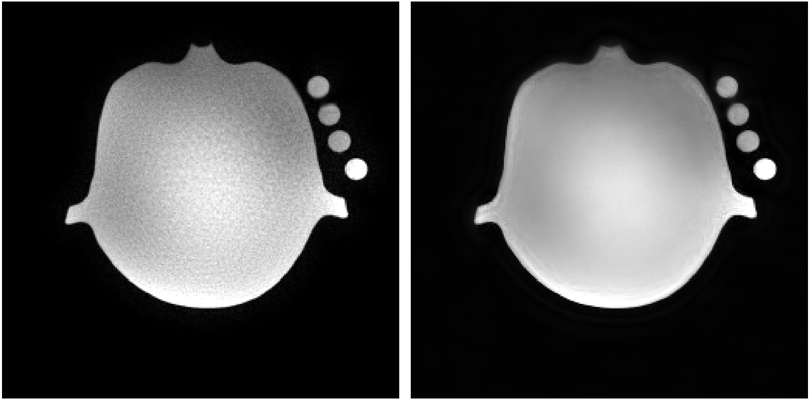
Effect of TV ($R = 4$)



(a) IRGN

(b) IRGNTV

Effect of TV ($R = 10$)



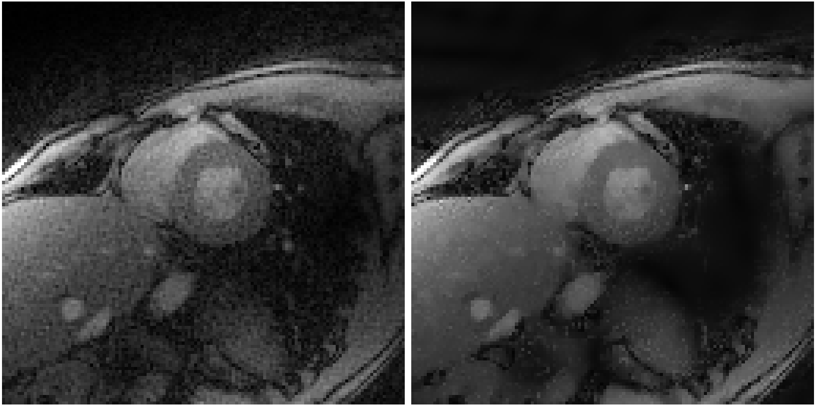
(a) IRGN (b) IRGNTV

Examples: radial sampling

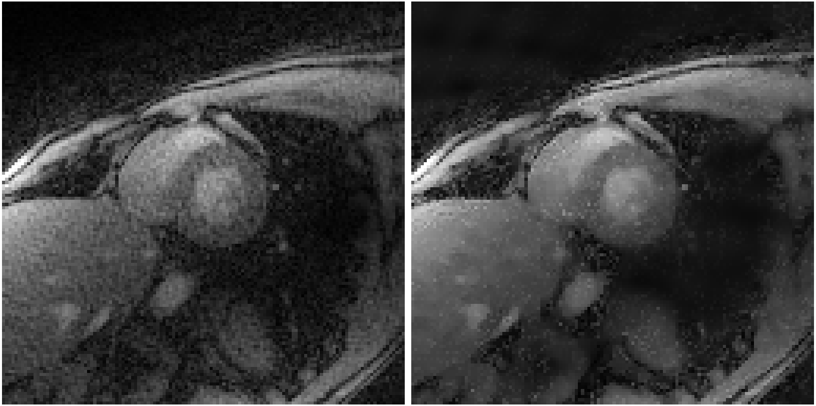
- Raw data of beating heart
- Radial FLASH sequence, 3T System, 32 channel coil
- 12 virtual channels (SVD) used for reconstruction
- 25 (19) projections, $R \approx 8$ (10.5)
- **No postprocessing, temporal view sharing**
- Sequence parameters
  - repetition time $\mathrm{TR} = 2.0\,\mathrm{ms}$
  - echo time $\mathrm{TE} = 1.3\,\mathrm{ms}$
  - flip angle $\mathrm{FA} = 8^\circ$
  - 256 points per proj. (2x oversampling) $\rightsquigarrow$ matrix 128×128
  - slice thickness 8 mm, in plane resolution $2\,\mathrm{mm} \times 2\,\mathrm{mm}$

(data courtesy of Martin Uecker, MPG Göttingen)

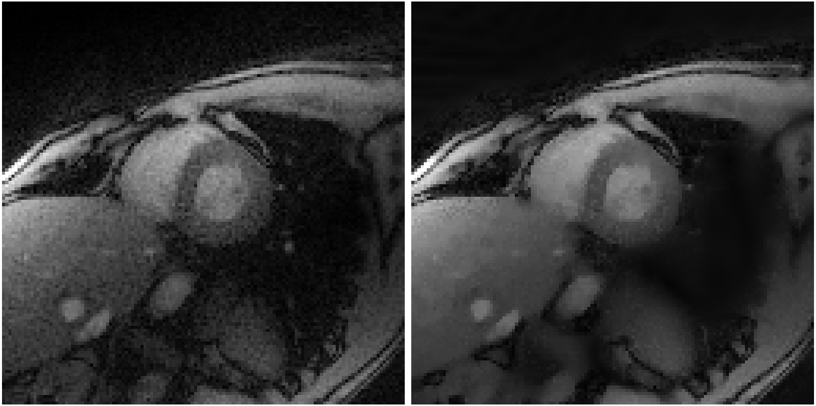
Radial sampling: cardiac (25 proj $\approx$ 20 fps)



Radial sampling: cardiac (19 proj $\approx$ 26 fps)



Radial sampling: temporal median filter



Total generalized variation (TGV)

Reduce stair-casing: use second order **total generalized variation**

$$\beta TGV^2(u) = \sup_{q\in C^2_\beta} \langle u, \operatorname{div}^2 q\rangle$$
$$= \inf_{v\in\mathcal{C}^1(\Omega,\mathbb{C}^d)} \beta\|\nabla u - v\| + 2\beta\|\mathcal{E}v\|$$

with

$$C^2_\beta = \{q \in \mathcal{C}^2_c(\Omega, \mathcal{S}^{d\times d}) : \ \|q\|_\infty \leq 2\beta, \|\operatorname{div} v\|_\infty \leq \beta\}$$
$$\mathcal{E}v = \frac{1}{2}(\nabla v + \nabla v^T) = (-\operatorname{div}^2)^* v$$

(see talk by **Kristian Bredies**)

IRGNTGV

Replace TV penalty on u^{k+1} with $TGV^2(u^{k+1})$:

1: choose $x^0 = (u^0, c^0)$, α_0, β_0, $q < 1$
2: **repeat**
3: solve for $\delta x = (\delta u, \delta c)$

$$\min_{\delta x} \frac{1}{2}\|F'(x^k)\delta x + F(x^k) - g\|^2 + \frac{\alpha_k}{2}\|W(c^k + \delta c)\|^2 + \beta_k TGV^2(u^k + \delta u)$$

4: set $x^{k+1} = x^k + \delta x$, $\alpha_{k+1} = \alpha_k q$, $\beta_{k+1} = \beta_k q$, $k = k+1$
5: **until** $\|F(x^k) - g\| < tol$
6: **return** u, c

Solution of IRGNTGV subproblems

Step 3:

Saddle point problem

$$\min_{\delta u, \delta c, v} \max_{\substack{p \in C_{\beta_k} \\ q \in C^2_{\beta_k}}} J(\delta u, \delta c) + \langle \nabla u^k + \delta u - v, p \rangle + \langle \mathcal{E} v, q \rangle$$

- Solution by projected primal-dual extragradient method
- Projection on C^2_β can be computed pointwise
- Cost dominated by (N)FFT $\rightsquigarrow$ comparable to IRGN(TV)

Primal-dual extragradient method

function TGVSOLVE$(u, c, \alpha, \beta, \sigma_u, \sigma_c, \sigma_v, \tau)$

  $\delta u, \bar{\delta u}, \delta c, \bar{\delta c}, v, \bar{v}, p, q \leftarrow 0$

  repeat

    $p \leftarrow \mathrm{proj}_\beta(p + \tau(\nabla(u + \bar{\delta u}) - v)$

    $q \leftarrow \mathrm{proj}^2_\beta(q + \tau(\mathcal{E} v))$

    $\delta u_{old} \leftarrow \delta u,\ \delta c_{old} \leftarrow \delta c,\ v_{old} \leftarrow v$

    $\delta u \leftarrow \delta u - \sigma_u(\partial_u J(u, c)(\bar{\delta u}, \bar{\delta c}) - \operatorname{div} p)$

    $\delta c \leftarrow \delta c - \sigma_c(\partial_c J(u, c)(\bar{\delta u}, \bar{\delta c}))$

    $v \leftarrow v - \sigma_v(-p - \operatorname{div}^2 q)$

    $\bar{\delta u} \leftarrow 2\delta u - \delta u_{old}$

    $\bar{\delta c} \leftarrow 2\delta c - \delta c_{old}$

    $\bar{v} \leftarrow 2v - v_{old}$

  until convergence

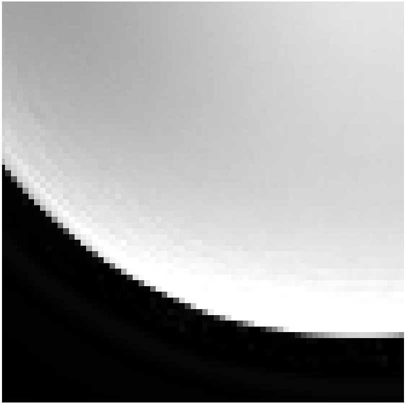
end function

Example: Random ($R = 4$)

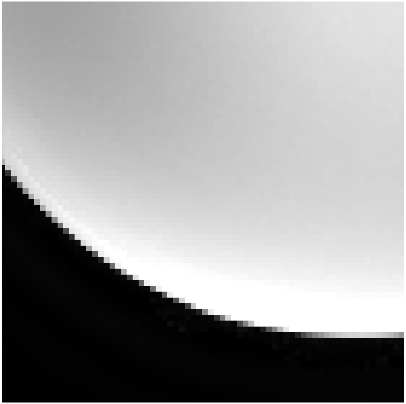
(a) IRGNTV

(b) IRGNTGV

Example: Random ($R = 4$)

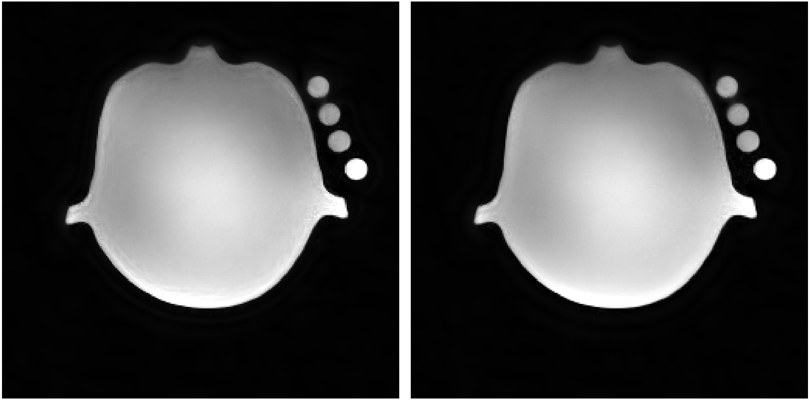


(a) IRGNTV (detail)



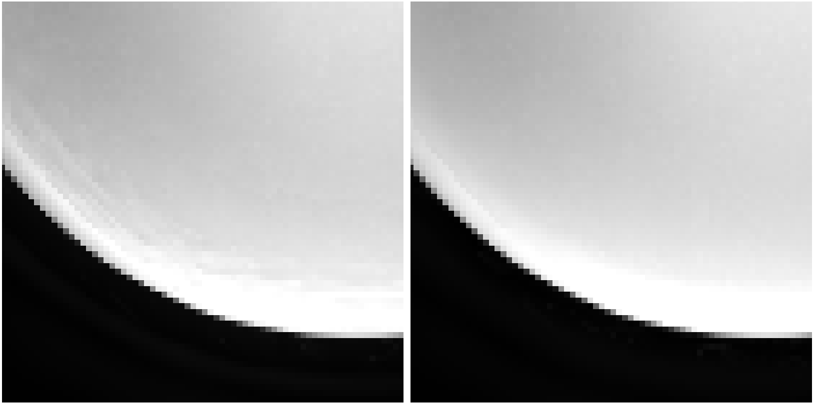
(b) IRGNTGV (detail)

Example: Random ($R = 10$)



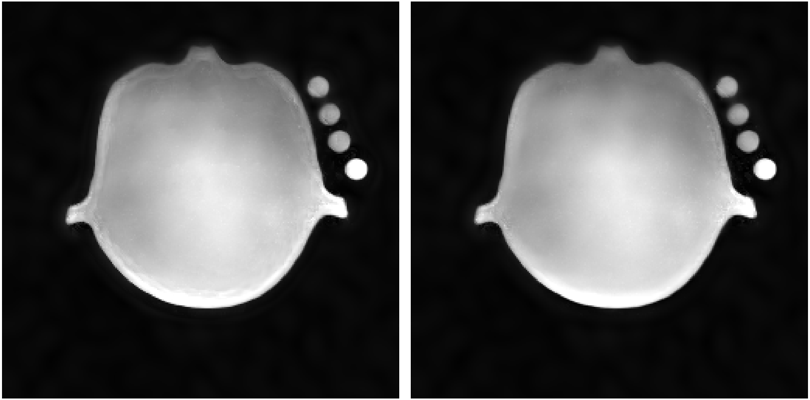
(a) IRGNTV (b) IRGNTGV

Example: Random ($R = 10$)



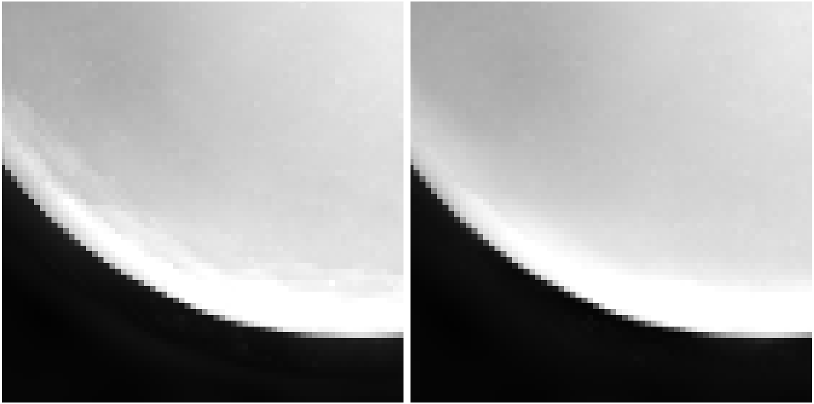
(a) IRGNTV (detail)   (b) IRGNTGV (detail)

Example: Random ($R = 18$)



(a) IRGNTV (b) IRGNTGV

Example: Random ($R = 18$)



(a) IRGNTV (detail) (b) IRGNTGV (detail)

Conclusion

Variational approach gives

- better reconstruction
- more flexibility

Outlook:

- Add constraint on slice/frame differences
- Full 3D T(G)V reconstruction
- Include parameter identification

Preprint, MATLAB code:

`http://www.uni-graz.at/~clason/publications.html`