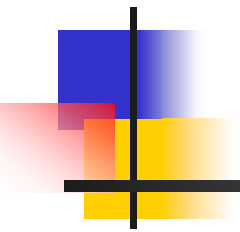
A Variational Approach to MR Bias Correction

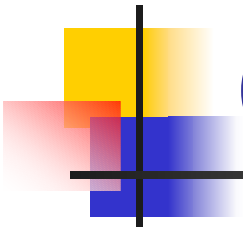


Ayres Fan

Stochastic Systems Group

Research Qualifying Exam

June 10, 2003

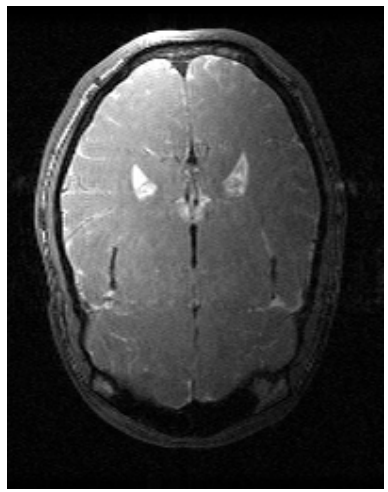


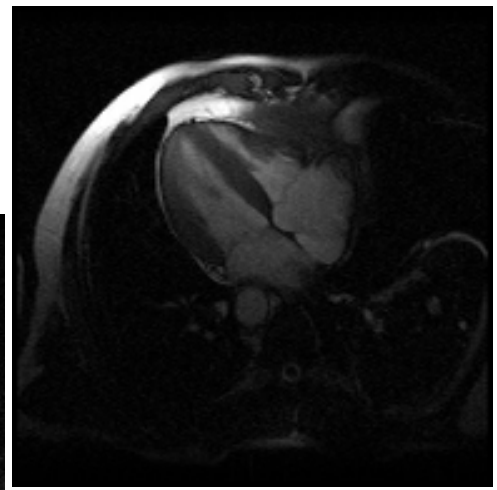
Outline

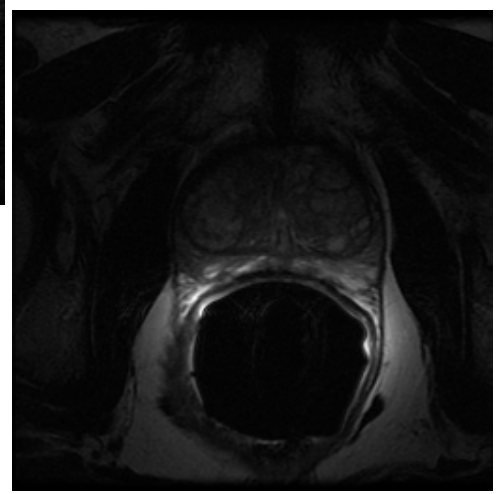
1. ***Introduction to bias correction***
2. Magnetic resonance imaging
3. Measurement model and problem formulation
4. Iterative Solver
5. Results
6. Conclusion

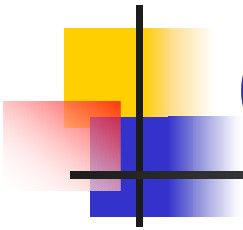
Problem Statement

- The bias field is a systematic intensity inhomogeneity that corrupts magnetic resonance (MR) images.
- Correcting for the bias field makes both human analysis (e.g., tumor detection, cartilage damage assessment) and computer analysis easier (e.g., segmentation, registration).
- General assumptions:
  - The bias field is slowly varying in space.
  - The bias field is tissue independent.
  - Tissue intensities are piecewise constant.









Outline

1. Introduction to bias correction
2. ***Magnetic resonance imaging***
3. Measurement model and problem formulation
4. Iterative Solver
5. Results
6. Conclusion

Magnetic Resonance

- Nuclear magnetic resonance (NMR) effect
  - Spins of nuclei become aligned either with or against an applied magnetic field. Nuclei precess around axis of rotation like a top.
- Larmour equation for the frequency of precession
  - $\omega_0 = \gamma ||\mathbf{B_0}||$
  - Resonant frequency proportional to the applied magnetic field strength.
  - γ is the magnetogyric ratio. Value varies depending upon element.
  - For hydrogen, γ = 42.57 MHz/T.
- Spatial Encoding
  - By varying the applied field strength in space, spatial location can be encoded in the frequency domain.

MR Imaging

- A pulse sequence is applied to tip the net magnetization vector into the transverse plane. The nuclei return to steady state with exponential decay while precessing at the Larmour frequency.
- The image generation is controlled by three intrinsic tissue properties and two user definable parameters.
  - e.g., for fast-spin echo (FSE) pulse sequences, the MR signal is given by:

$$\varphi(x) = \rho(x)\exp(-T_E/T_2(x))(1-\exp(-T_R/T_1(x)))$$

- We can target ρ, T1, and T2 measurements through appropriate selection of TE and TR.
  - e.g., by selecting a small TE and a large TR, we can minimize the effect of T1. This type of imaging is known as T2-weighted

T_E: time echo (time we measure signal)
T_R: time repeat (time between pulse sequences)
T_1: spin-lattice relaxation (recovery of z-magnetization)
T_2: spin-spin relaxation (loss of xy-magnetization)
ρ: proton density

Image Reconstruction

- A common approach is to encode location using gradient fields and convert the received MR signal from the frequency domain using the inverse Fourier transform.
  - Due to imperfections in the acquisition process (e.g., non-linear gradients, spatially varying $\mathbf{B_0}$), the result of the IFFT will be complex.
  - Generally, the final reconstructed image is taken to be the absolute value of the complex image.
  - The noise in the acquisition process is thermal and Gaussian, so reconstruction results in Rician noise. Rician noise is similar to Rayleigh at low SNR, Gaussian at high SNR.
- Maximizing SNR is a major goal
  - Signal level tends to be fairly low. Vast majority of atoms cancel each other out resulting in a small net magnetization. Increasing $\mathbf{B_0}$ increases net magnetization strength.
  - Increase SNR by spatial averaging, time averaging, filtering, increasing signal reception, and increasing $\mathbf{B_0}$.

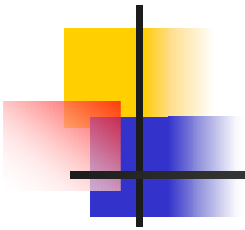
Bias Field

- The body coil is a large coil usually wrapped around the main cavity
  - Typically transmit RF pulse sequence with body coil due to good spatial homogeneity. Transmitted power density limited by FDA.
- Surface coils are coils placed near the object of interest.
  - In order to increase signal level, increase the induced magnetic field strength through coil design.
  - Surface coils have good signal strength near the coil, and the strength rapidly diminishes with distance. Thus we can get good SNR in the ROI.
- The signal observed at the receiver is then:

$$I(x) = \beta(x)\varphi(x) + n(x)$$

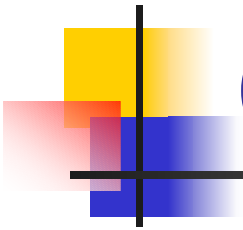
$\beta(x)$ is the magnetic field induced by the receiving coil.
$\varphi(x)$ is the ideal MR signal

- The severity of the bias field is determined by the spatial homogeneity of the surface coil reception profile.



Previous Work

- Earliest work in mid-80's: Haselgrove and Prammer (1986), Lufkin et al (1986) with homomorphic unsharp filtering, Axel et al (1987) with phantom correction.
- Window/level adequate for human vision tasks, but not for computer-based processing.
- More modern approaches include Dawant et al (1993) with spline fitting, Singh and NessAiver (1993) with embedded coil markers, Likar et al (2000) with entropy minimization.
- Various simultaneous bias correction and segmentation approaches such as Meyer et al (1995) using adaptive clustering and Wells et al (1996) using Expectation-Maximization.



Outline

1. Introduction to bias correction
2. Magnetic resonance imaging
3. ***Measurement model and problem formulation***
4. Iterative Solver
5. Results
6. Conclusion

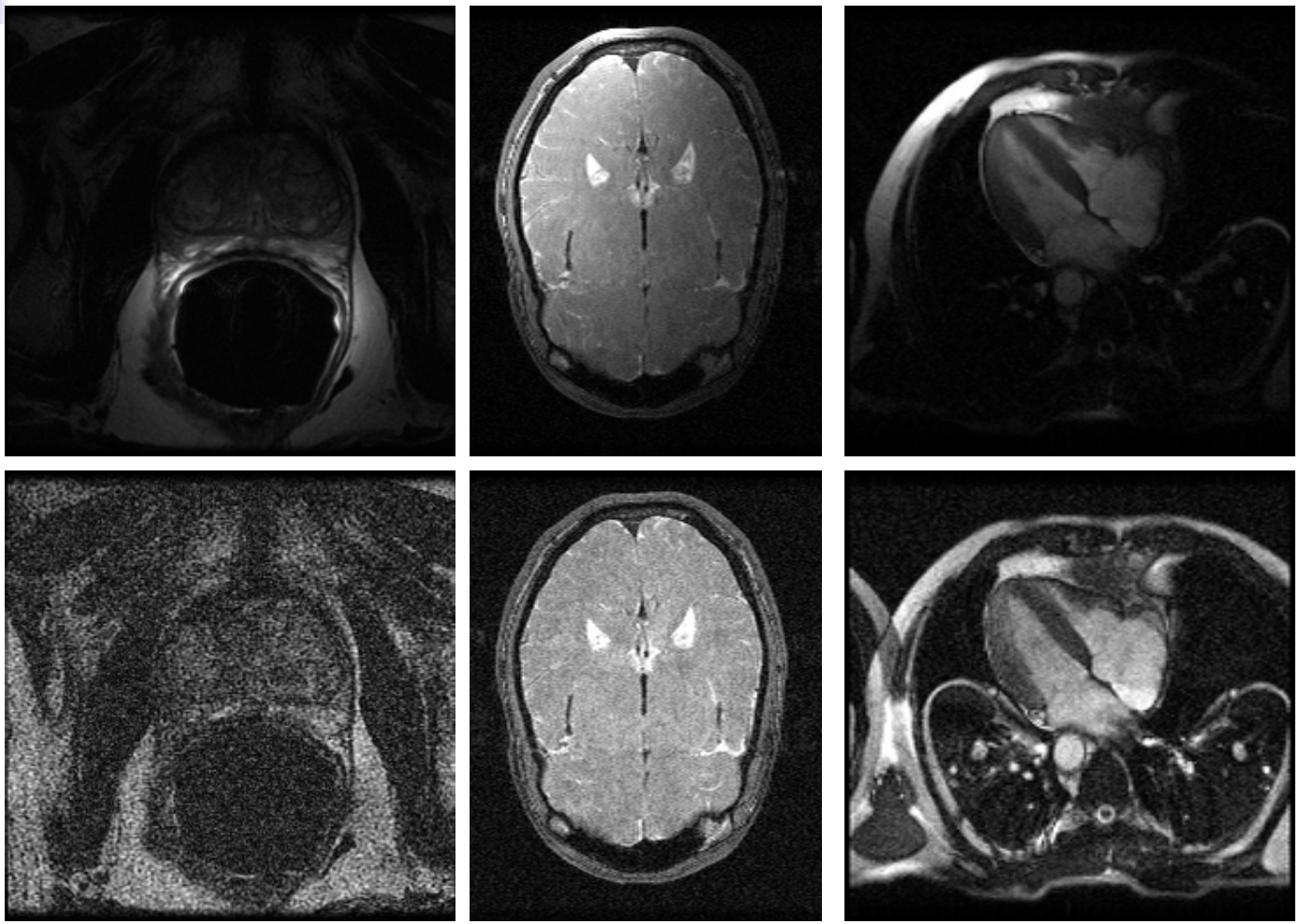
Measurement Model

- Brey and Narayana (1988) proposed capturing images from both the body coil and the surface coil. The measurement model is then:

$$I_B(x) = \varphi(x) + n_B(x)$$
$$I_S(x) = \beta(x)\varphi(x) + n_S(x)$$

- I_B is homogeneous but noisy. I_S has high SNR in the region of interest, but a potentially severe bias artifact.
- Note that gain in SNR from using a surface coil does not come from reduction of noise, but from increased signal gain from the bias field.

Example Data



Surface coil images (top) and body coil images (bottom)

Previous BC/SC Work

- Brey-Narayana filter the two observation images to denoise them and divide the results to estimate the bias field:

$$\begin{aligned} \hat{b} &= (h_1 * I_S)/(h_2 * I_B) \\ \hat{f} &= I_S/\hat{b} \end{aligned}$$

- Lai and Fang (1998) take I_S/I_B and select a sparse set of reliable control points and fit splines to the bias field.
- Pruessmann et al (2001) who fit local polynomials to the bias field.

ML Formulation

- We model the noise as Gaussian and IID.
  - This is a good approximation in medium-to-high SNR regions except the Rician noise adds a bias of 2-5%.
  - Generally low SNR regions correspond to air regions which we do not care about
- We stack the 2D or 3D images into vectors.
  - We now want to estimate the vector quantities **f** and **b**.
  - The discrete measurement model is:

$$\begin{aligned} \boldsymbol{y}_{\mathrm{B}} &= \boldsymbol{f} + \boldsymbol{n}_B \\ \boldsymbol{y}_{\mathrm{S}} &= \boldsymbol{b} \circ \boldsymbol{f} + \boldsymbol{n}_S \end{aligned}$$

- We can then write the log likelihood as (ignoring constant terms):

$$\ell(\boldsymbol{y}_{\mathrm{B}}, \boldsymbol{y}_{\mathrm{S}}; \boldsymbol{b}, \boldsymbol{f}) = -\frac{1}{2\sigma_B^2}||\boldsymbol{y}_{\mathrm{B}} - \boldsymbol{f}||^2 - \frac{1}{2\sigma_S^2}||\boldsymbol{y}_{\mathrm{S}} - \boldsymbol{b} \circ \boldsymbol{f}||^2$$

Regularization

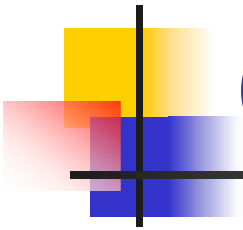
- Finding the ML estimate results in trivial estimates:

$$\hat{b} = y_S / y_B$$
$$\hat{f} = y_B$$

- We construct an augmented energy function that encourages smoothness in **b** and piecewise smoothness in **f**:

$$E(\boldsymbol{b}, \boldsymbol{f}) = \lambda_B ||\boldsymbol{y}_B - \boldsymbol{f}||^2 + \lambda_S ||\boldsymbol{y}_S - \boldsymbol{b} \circ \boldsymbol{f}||^2 + \alpha ||\boldsymbol{D}\boldsymbol{b}||^2 + \gamma ||\boldsymbol{L}\boldsymbol{f}||_p^p$$

  - We generally choose p <= 1 to help preserve edges
  - **D** and **L** are matrices chosen to implement differential operators
  - λ_B, λ_S, α, and γ are all positive constants
  - The λ's can be seen to be related to the inverse noise variances of the observed images.



Outline

1. Introduction to bias correction
2. Magnetic resonance imaging
3. Measurement model and problem formulation
4. ***Iterative Solver***
5. Results
6. Conclusion

Optimizing Energy Function

- Overall problem is non-convex.
- Use coordinate descent to alternately optimize **b** and **f**
  - Minimizing the energy simultaneously with respect to **b** and **f** is difficult. But given b, f is relatively easy to obtain, and vice versa.
  - A stationary point found using coordinate descent is also a stationary point of the overall energy functional.
- b-step
  - Minimize $\lambda_S \|\boldsymbol{y}_S - \boldsymbol{F}\boldsymbol{b}\|^2 + \alpha \|\boldsymbol{D}\boldsymbol{b}\|^2$

    F is a diagonal matrix with **f** along the diagonal
- f-step
  - Minimize $\lambda_B \|\boldsymbol{y}_B - \boldsymbol{f}\|^2 + \lambda_S \|\boldsymbol{y}_S - \boldsymbol{B}\boldsymbol{f}\|^2 + \gamma \|\boldsymbol{L}\boldsymbol{f}\|_p^p$

    B is a diagonal matrix with **b** along the diagonal

Solving for b

- With **f** fixed, the energy is quadratic in terms of **b**.
  - The quadratic matrix is positive definite which means that the energy function is strictly convex.
  - Hence there is only one local minimum, which is also the global minimum, for a given **f**.
- Setting the gradient to zero leads to a linear equation for the solution:

$$\left(\lambda_S \boldsymbol{F}^2 + \alpha \boldsymbol{D}^T \boldsymbol{D}\right) \boldsymbol{b} = \boldsymbol{F} \boldsymbol{y}_S$$

- We can solve by direct matrix inversion, but **b** is often very large (e.g., 65,536 elements for a 256x256 image).
- We use conjugate gradient to find a sub-optimal iterative solution.

Conjugate Gradient

- Conjugate gradient exhibits a superlinear convergence rate for convex quadratic optimization problems.
  - Want to minimize $\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{a}^T\boldsymbol{x}$
  - Updates of the form: $\boldsymbol{x}^{(i)} = \boldsymbol{x}^{(i-1)} + \delta^{(i)}\boldsymbol{d}^{(i)}$

    $$\boldsymbol{g}^{(i)} = \nabla E|_{\boldsymbol{x}^{(i-1)}} = \boldsymbol{Q}\boldsymbol{x}^{(i-1)} - \boldsymbol{a}$$

    $$\boldsymbol{d}^{(i)} = -\boldsymbol{g}^{(i)} + \frac{\|\boldsymbol{g}^{(i)}\|^2}{\|\boldsymbol{g}^{(i-1)}\|}\boldsymbol{d}^{(i-1)}$$
  - For quadratic problems, we can find the exact line minimization in closed form:

    $$\delta^{(i)} = -\frac{(\boldsymbol{g}^{(i)})^T\boldsymbol{d}^{(i)}}{(\boldsymbol{d}^{(i)})^T\boldsymbol{Q}\boldsymbol{d}^{(i)}}$$

Preconditioning

- Convergence rate of conjugate gradient depends on the condition number of the matrix.
- Do change of variables $\tilde{x} = Sx$
- Then the energy functional becomes

$$\frac{1}{2}\tilde{x}^T S^{-1} Q S^{-1} \tilde{x} - a^T S^{-1} \tilde{x}$$

  - In order to be effective, $\mathbf{S}^{-1}\mathbf{Q}\mathbf{S}^{-1}$ needs to have a smaller condition number than **Q**, and $\mathbf{S}^{-1}$ must be easy to apply.
  - Can write updates directly in terms of **x**.
- We find that using a tridiagonal preconditioner using the three main diagonals of **Q** increases convergence speed by 2x-3x, while increasing computation time by 10-15% per iteration.

Solution for f (γ=0)

- With no regularization on f, we can minimize the function pointwise:

$$\hat{\boldsymbol{f}}[n] = \frac{\lambda_B \boldsymbol{y}_B[n] + \lambda_S \boldsymbol{b}[n] \boldsymbol{y}_S[n]}{\lambda_B + \lambda_S \boldsymbol{b}^2[n]}$$

- This equation can be interpreted as a noise-weighted convex combination between y_B and y_S/b.
  - When **b**[n] is large, we mainly use the surface coil for the reconstruction
  - When **b**[n] is small, we mainly use the body coil
- Increases SNR by 0-3 dB over best image at each point.
  - In regions far from the surface coil, there will be a significant advantage to incorporating the body coil measurements.
- This is essentially the same fusion equation found by Roemer et al (1990). This is the canonical method of combining multiple surface coil images when the coil reception profiles are known.

Half-Quadratic Optimization

- To solve l_p regularization problems, we use half-quadratic optimization (Geman-Reynolds 1992).
  - Fixed-point iterative method where we form a succession of quadratic approximations $\mathbf{f}^{(0)}$, $\mathbf{f}^{(1)}$, …:

$$\|\boldsymbol{L}\boldsymbol{f}\|_p^p \approx \boldsymbol{f}^T \boldsymbol{L}^T \boldsymbol{W}^{(i)} \boldsymbol{L}\boldsymbol{f}$$

  - Choose $\mathbf{W}^{(i)}$ so that the relationship holds for $\mathbf{f}^{(i-1)}$:

$$\boldsymbol{W}^{(i)}[n,n] = \frac{p}{2}\left(((\boldsymbol{D}\hat{\boldsymbol{f}}^{(i-1)})[n])^2 + \xi\right)^{p/2-1}$$

- This makes each f-step quadratic, so we just need to solve this linear equation at each half-quadratic iteration:

$$\left(\lambda_B \boldsymbol{I} + \lambda_S \boldsymbol{B}^2 + \gamma \boldsymbol{D}^T \boldsymbol{W}^{(i)} \boldsymbol{D}\right) \hat{\boldsymbol{f}}^{(i)} = \lambda_B \boldsymbol{y}_{\mathrm{B}} + \lambda_S \boldsymbol{B}\boldsymbol{y}_{\mathrm{S}}$$

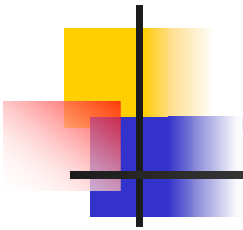
  - This is again a positive definite system, so we can use PCG to solve.
  - Results in three sets of nested iterations.

Multiple Coils & Pulse Sequences

- Many protocols use multiple surface coils to simultaneously receive the MR signal. This allows for better spatial coverage.
  - Generally combine into a single composite image using sum-of-squares.
- We can generalize our energy functional to find the bias field estimate for each coil and one composite true image estimate.
- When acquiring multiple pulse sequences, we only need one body coil image. The bias field is largely unchanged (though some minor effects may crop up due to, e.g., magnetic susceptibility of the tissue).
- We can create a generalized energy functional that handles a combination of multiple surface coils and pulse sequences:

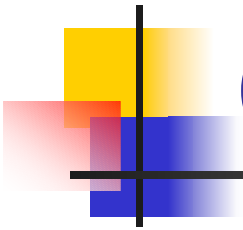
$$\sum_i \lambda_i \|\boldsymbol{y}_i - \boldsymbol{b}_{m_i} \circ \boldsymbol{f}_{n_i}\|^2 + \sum_m \alpha_m \|\boldsymbol{L}_m \boldsymbol{b}_m\|^2 + \sum_n \gamma_n \|\boldsymbol{D}_n \boldsymbol{f}_n\|_p^p$$

- We can use coordinate descent again, but on each b- or f-step we need to compute an estimate for each **b** and each **f**.



3D Processing

- Straightforward application of the exact same energy functional.
  - Slower than independent processing of slices.
  - Enables better coupling across slices than independent processing allows.
- Common clinical practice to capture volumes with different slice orientations (e.g., axial, sagittal).
  - Usually have higher in-plane resolution than inter-plane resolution.
  - Only need one body coil image to correct different orientation volumes because the bias field is slowly changing in space. Simply find the bias field for the volume we have a body coil image for, and interpolate onto other sampling grids.

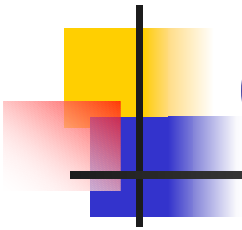


Optimizing Performance

- Multigrid
  - Multiresolution technique that can help avoid local minima, and increase convergence speed.
  - Simple coarse-to-fine implementation.
  - This allows us to find the low frequency components on coarser grids and propagate the results to the original scale.
- Many other parameters to choose (e.g., tolerances for each f-step and b-step). Need to optimize individually for each particular application.

Overview of Algorithm

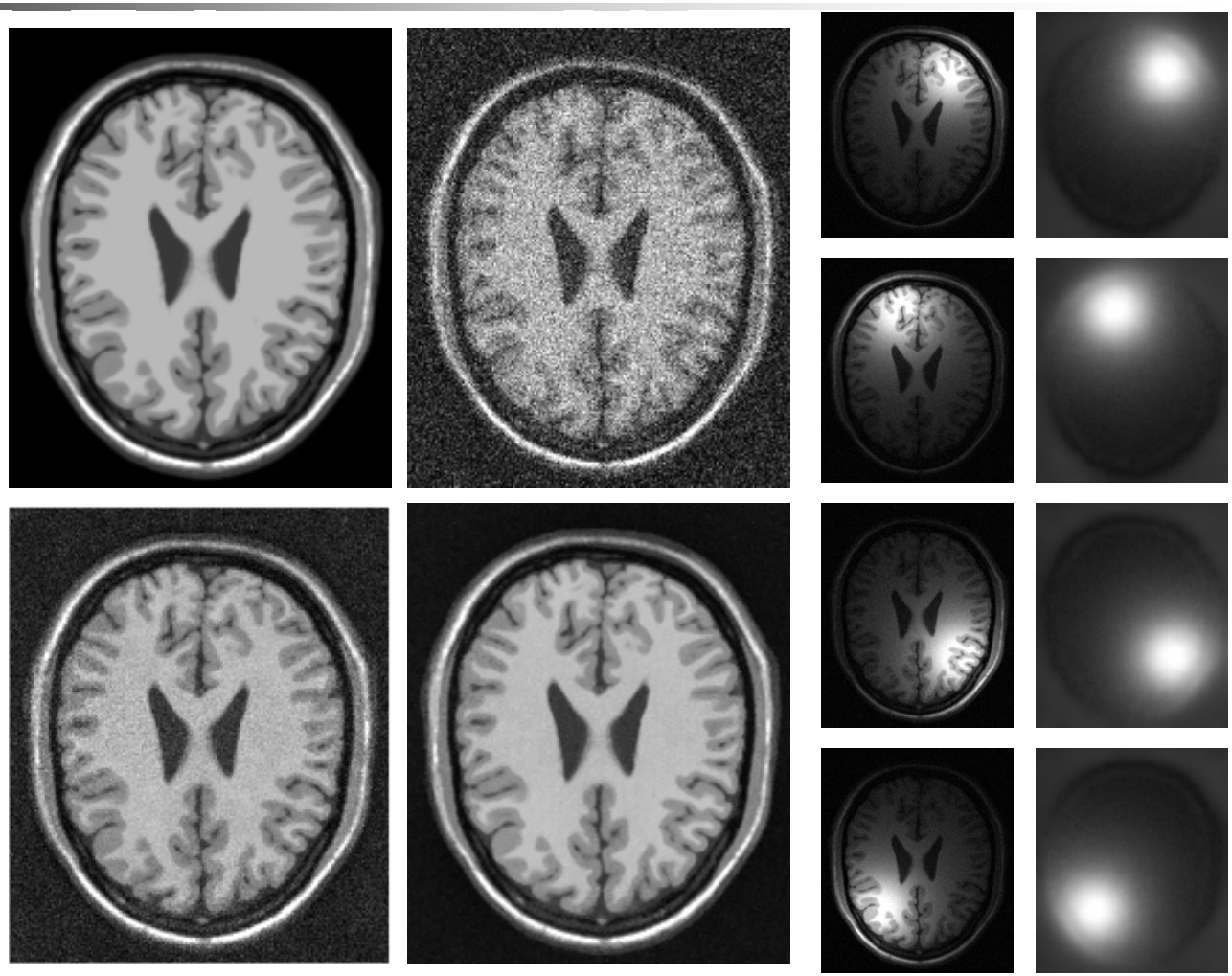
I. Downsample to finest desired grid.
II. Solve problem at scale s (using input from scale s+1)
    A. b-step: minimize energy with respect to b
        1. PCG iterations
    B. f-step: minimize energy with respect to f
        1. Half-quadratic iterations
            a. PCG iterations
    C. Repeat A and B until convergence.
III. Repeat II until we reach scale 0.



Outline

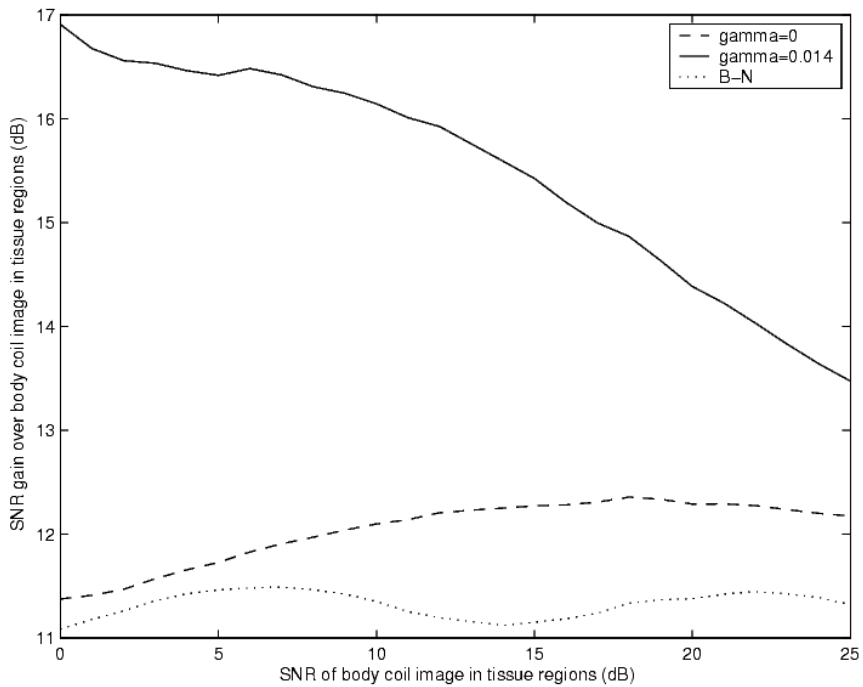
1. Introduction to bias correction
2. Magnetic resonance imaging
3. Measurement model and problem formulation
4. Iterative Solver
5. **Results**
6. Conclusion

MNI Example

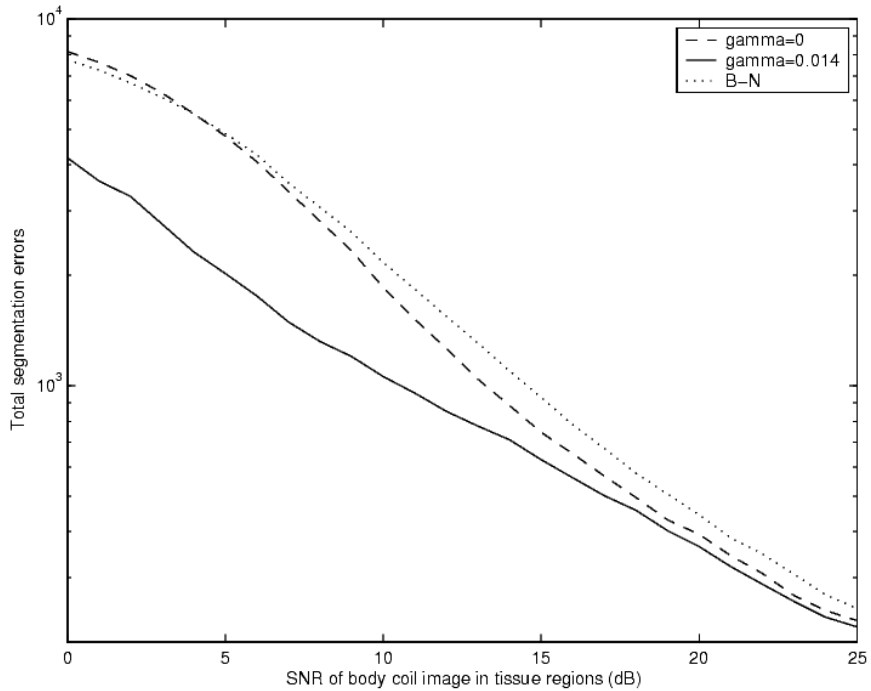


(left to right) True image (top), B-N estimate (bottom), body coil image (top), our f estimate (bottom), surface coil images, and bias field estimates.

MNI Scaling



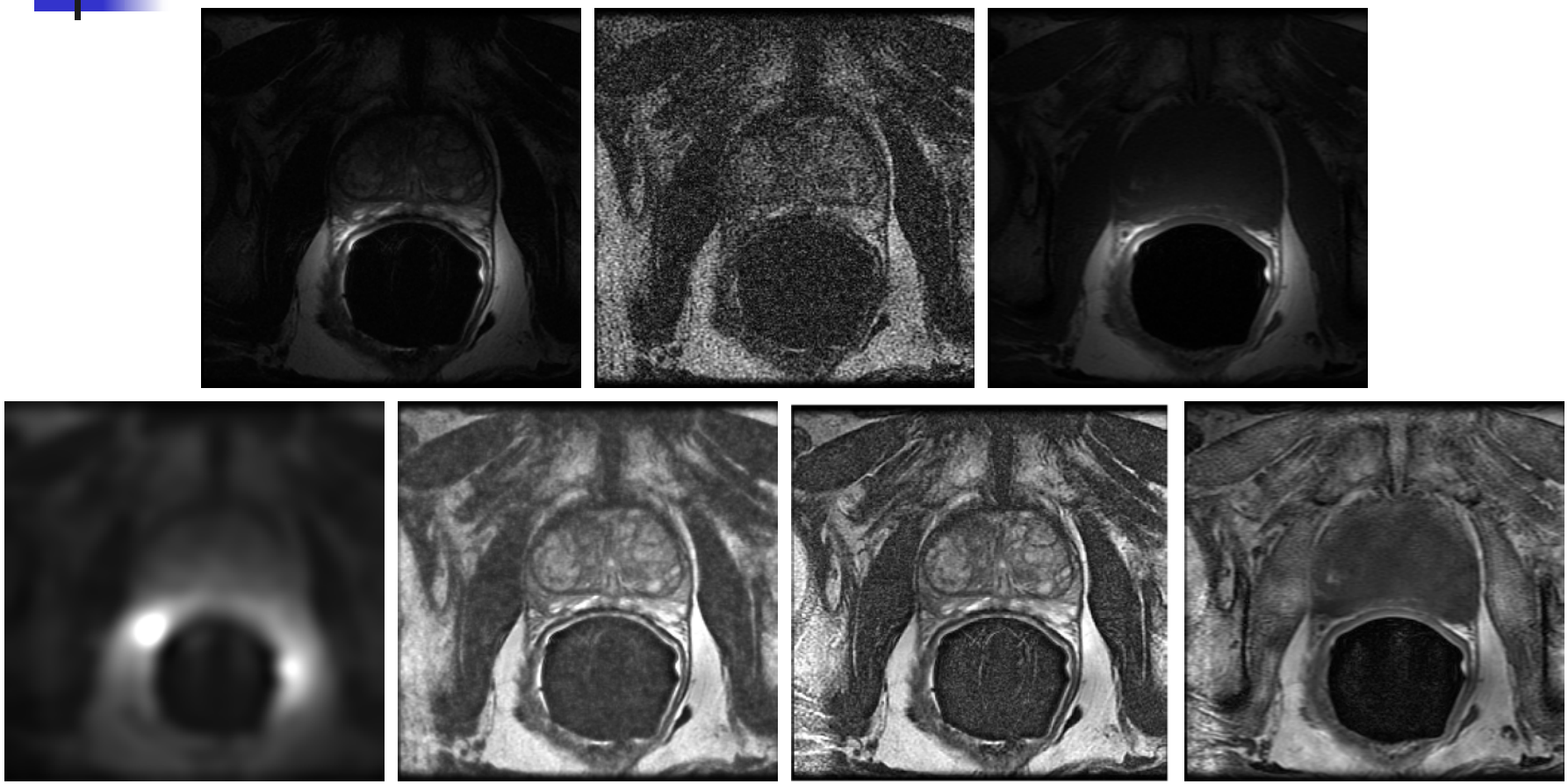
(a)



(b)

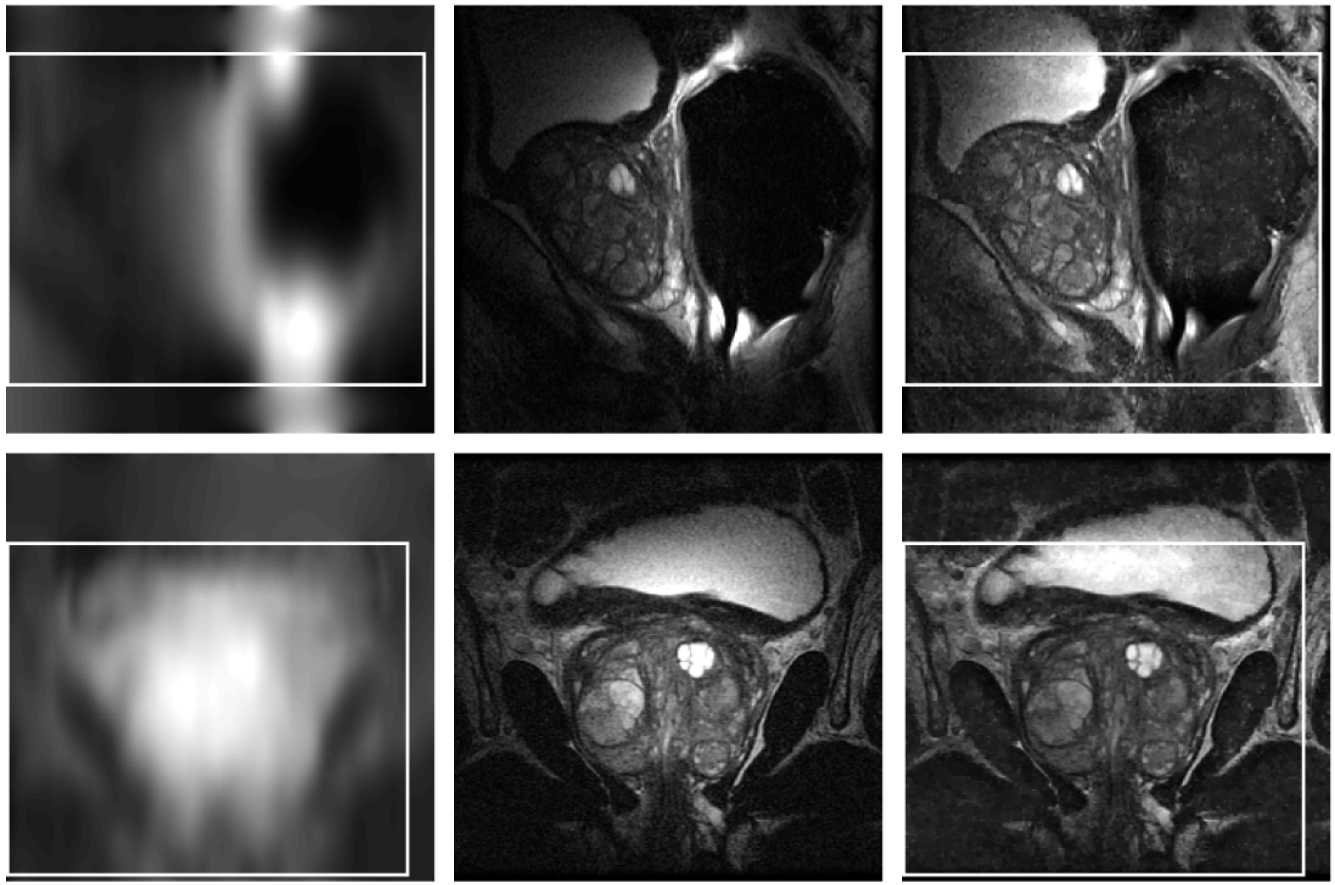
Performance of (a) SNR gain and (b) segmentation errors as a function of image acquisition SNR

Prostate Results



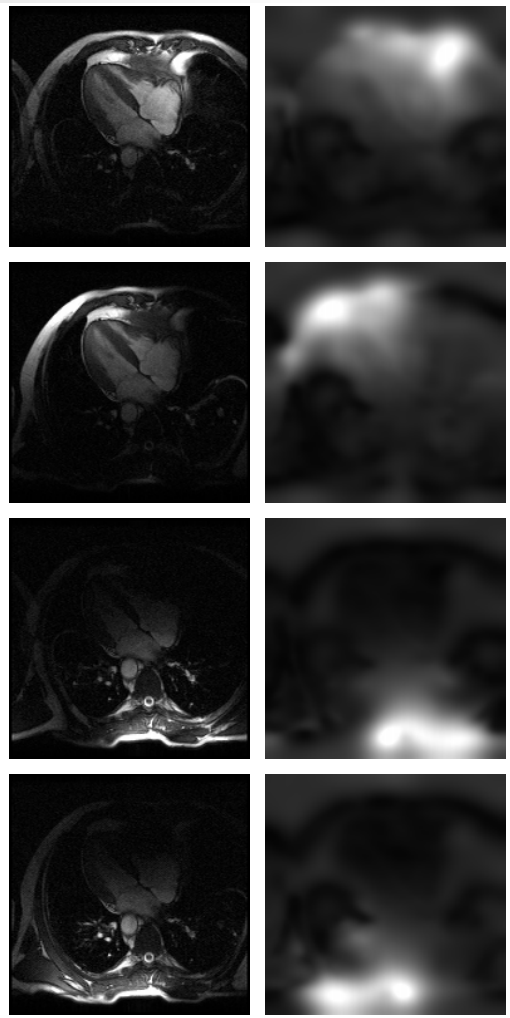
Top: T2W surface coil image, T2W body coil image, T1W surface coil image.
Bottom: Estimated bias field, true T2W image estimate, B-N T2W, true T1W image estimate.

Coronal and Sagittal Correction



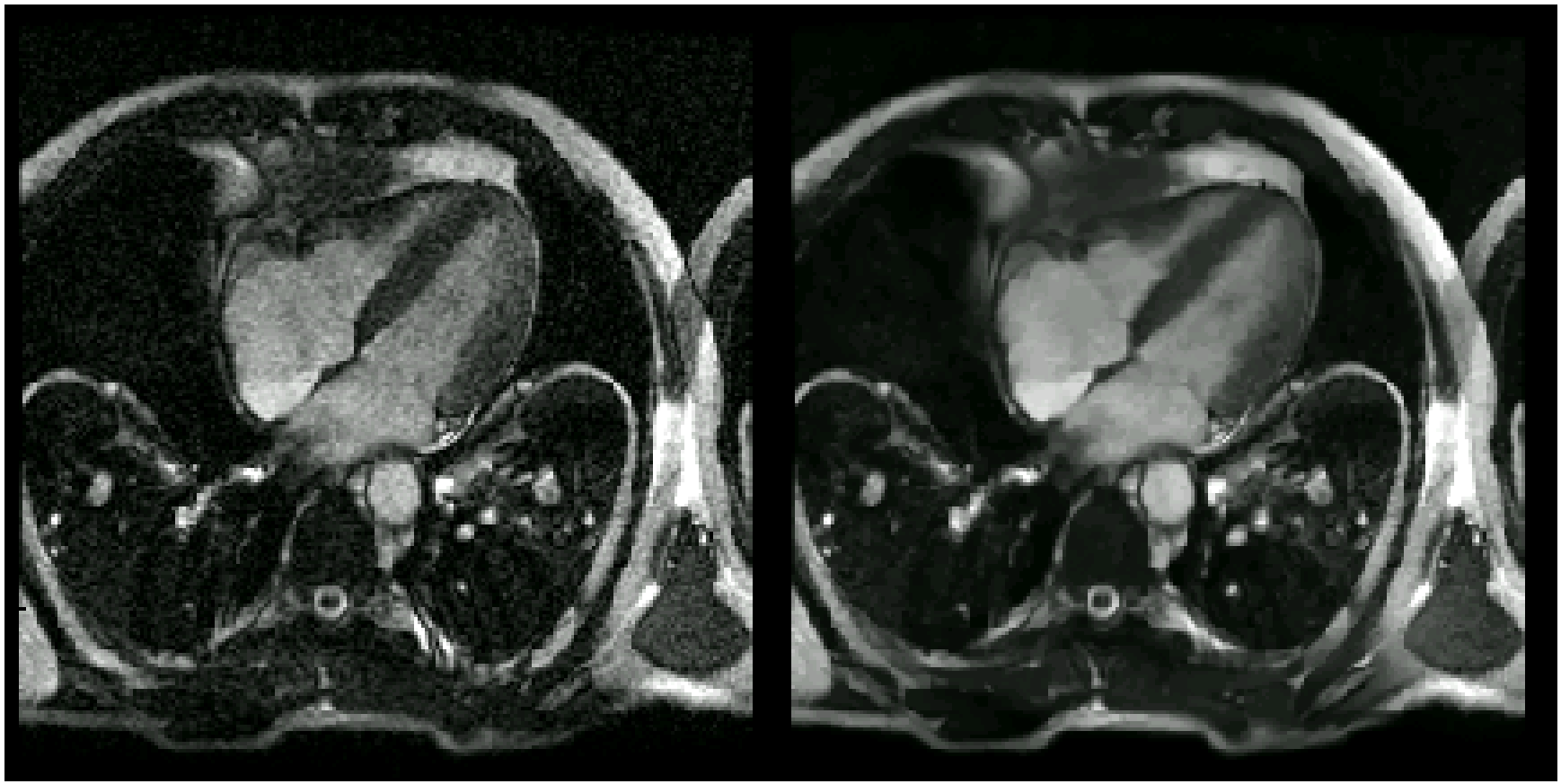
Sagittal (top), coronal (bottom). Bias field (left), surface coil image (middle), true image (right).

Heart Example

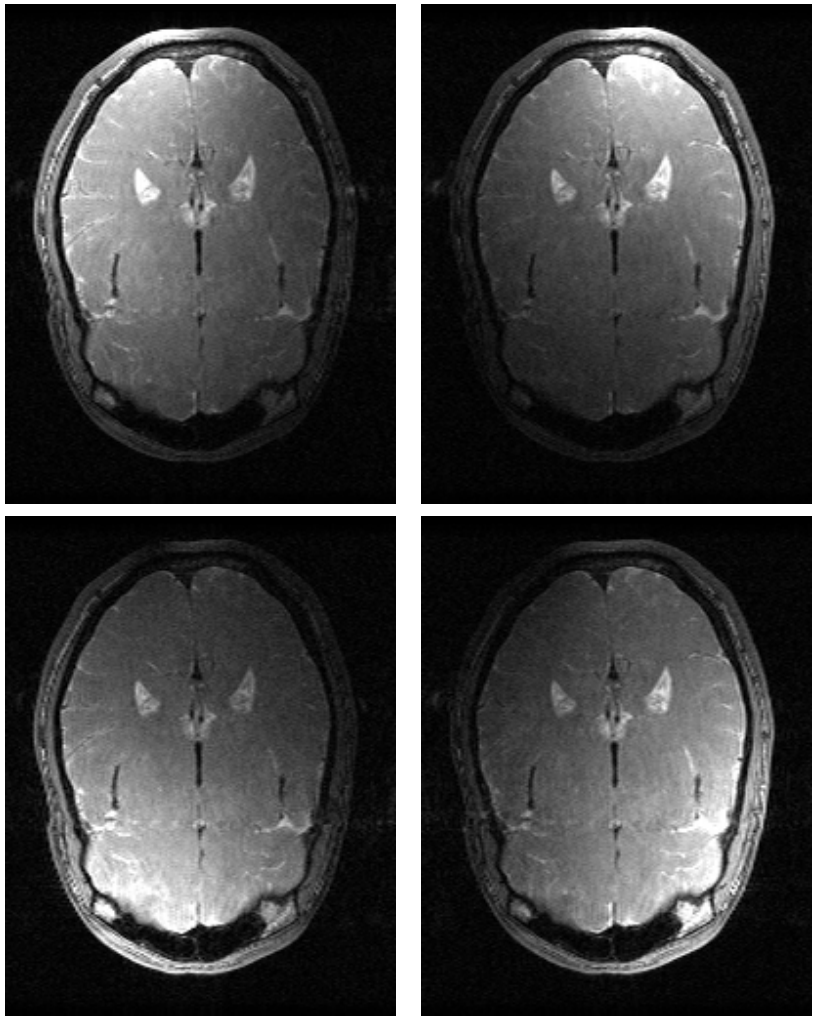


(left to right) Body coil image, B-N estimate, our estimate of f, surface coil images, and bias field estimates.

Heart Movie

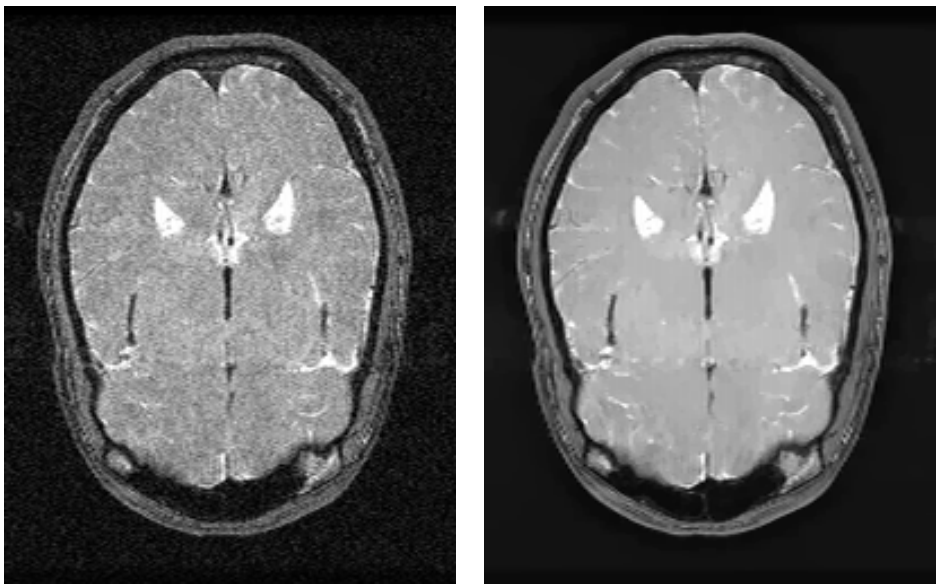


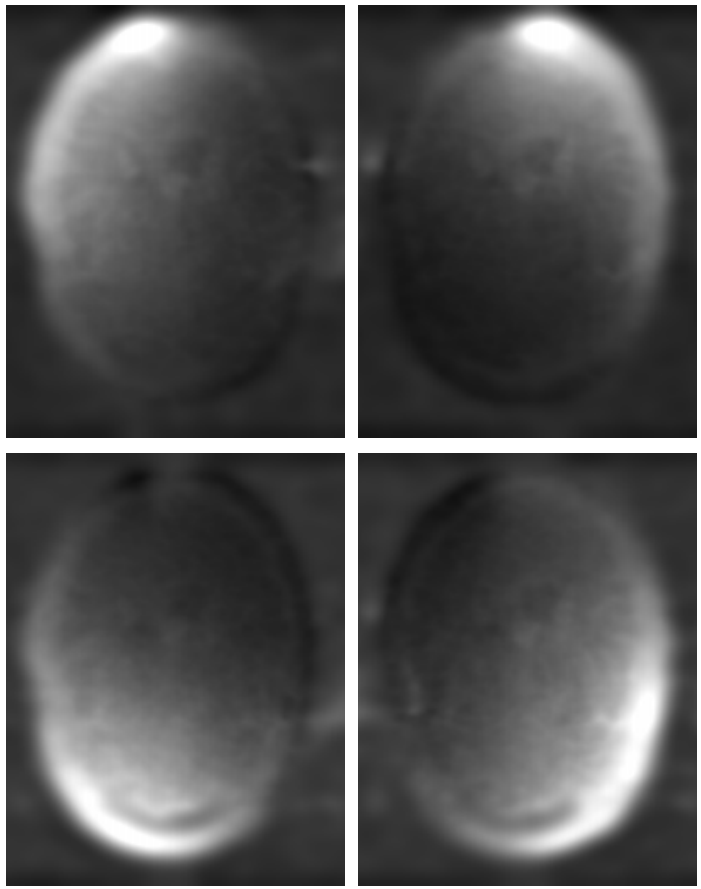
Brain Example



Body coil image (left) and surface coil images (right), gradient recalled echo (GRE)

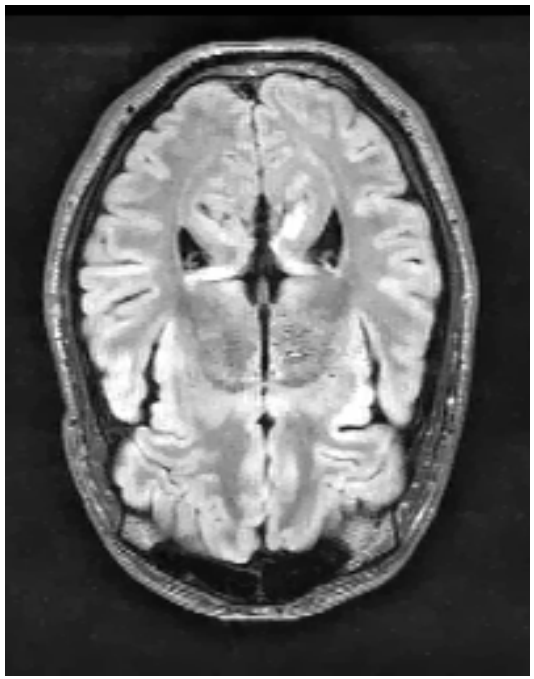
Brain Results

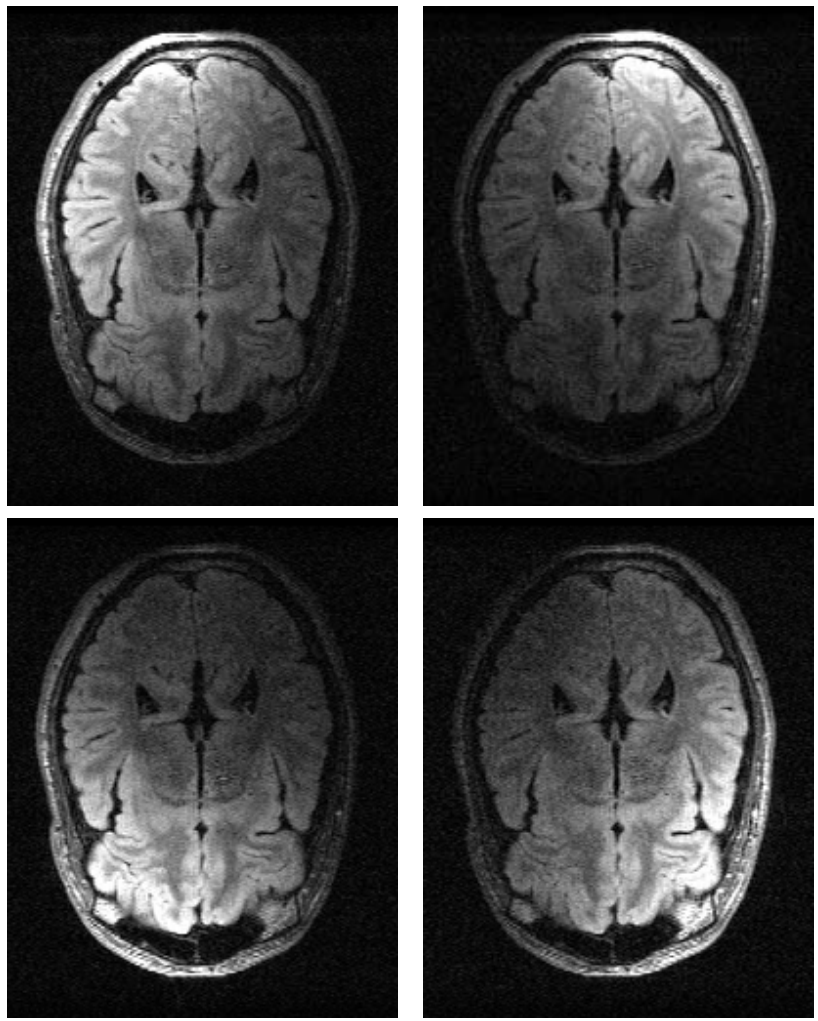




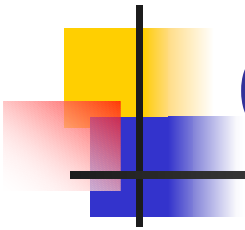
Body coil image (left),
estimated image (middle)
and bias fields (right)

Brain FLAIR Results



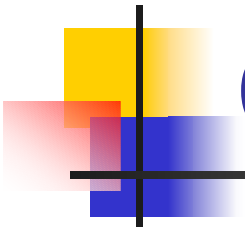


Estimated true image (left), surface coil images (right), FLAIR pulse sequence



Contributions

- Non-parametric variational formulation of image fusion problem with statistical estimation flavor.
- Demonstrably superior results on synthetic examples.
- Simultaneous bias correction and denoising.
- Seamless handling of multiple surface coils and multiple pulse sequences.
- Efficient solver using coordinate descent, preconditioned CG, multigrid.



Convergence

- In many ways, our coordinate descent approach can be viewed as being similar to Expectation-Maximization.
  - Can be viewed as an EM implementation if we believe that the regularization terms really are our statistical priors.
  - Have the same convergence properties as EM: every f-step and b-step is guaranteed to decrease the energy.
- In general, we can only hope to find a local minimum.
  - In practice, we have found very robust convergence properties, even without using multigrid.
  - Initialization with random noise will usually find a reasonable result, and the results seem to be the same local minimum.

Initialization/Parameter Choice

- Choose initial **b** based on method of Brey-Narayana.
- Then choose initial **f** by doing a f-step without regularization.

- Choose λ's based on measured noise variances in images.
- Choose α's and γ's based on empirical visual observations.
- Across scales, use simple multiplicative scaling for parameters:

$$\begin{aligned} \lambda^{[s]} &= 4\lambda^{[s-1]} \\ \alpha^{[s]} &= \epsilon_1 \alpha^{[s-1]} \\ \gamma^{[s]} &= \epsilon_2 \gamma^{[s-1]} \end{aligned}$$

PCG Update Equations

- Update equations directly in terms of **x**

$$\boldsymbol{H} = \boldsymbol{S}^2$$

$$\boldsymbol{x}^{(i)} = \boldsymbol{x}^{(i-1)} + \delta^{(i)}\boldsymbol{d}^{(i)}$$

$$\boldsymbol{g}^{(i)} = \nabla E|_{\boldsymbol{x}^{(i-1)}} = \boldsymbol{Q}\boldsymbol{x}^{(i-1)} - \boldsymbol{a}$$

$$\boldsymbol{d}^{(i)} = -\boldsymbol{H}^{-1}\boldsymbol{g}^{(i)} + \frac{(\boldsymbol{g}^{(i)})^T\boldsymbol{H}^{-1}\boldsymbol{g}^{(i)}}{(\boldsymbol{g}^{(i-1)})^T\boldsymbol{H}^{-1}\boldsymbol{g}^{(i-1)}}\boldsymbol{d}^{(i-1)}$$