Statistical Tests for Total Variation Regularization Parameter Selection

Jodi Mead
Department of Mathematics
Boise State University

BOISE STATE UNIVERSITY

This work is supported by NSF-DMS-1043107

Outline

- TV Regularization Parameter Choices
- Residuals as Risk Estimators
- Degrees of Freedom Estimates
- χ^2 Test for Regularization Parameter Selection
- Imaging Examples

Inverse Problem

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}$$

$$\mathbf{y} \in \mathbb{R}^m,\ \mathbf{A} \in \mathbb{R}^{m\times n},\ \mathbf{x} \in \mathbb{R}^n,\ \mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0},\ \mathsf{cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$$

Total Variation minimization

$$\hat{\mathbf{x}}_{tv} \in \arg\min_{\mathbf{x}} \frac{\lambda}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\nabla\mathbf{x}\|_1$$

Three TV algorithms

- **deconvtv** (S. Chan 2011) Augmented Lagrangian method
- **tvdeconv** (P. Getreuer 2010) Split Bregman method
- **FTVd** (J. Yang, Y. Zhang, W. Yin 2008)) Augmented Lagrangian method

Focus of talk is on regularization parameter (λ) selection
for any TV algorithm

Choice of λ example

$\lambda = 500$

$\lambda = 50000$

$\lambda = 500000$

TV regularization parameter selection

Approaches

- **L-curve**: no guarantee that norm of the data residual vs.the solution will be L-shaped.
- The following can be used when TV function is viewed or approximated with a quadratic functional:
  Discrepancy principle*, **Unbiased Predictive Risk Estimator (UPRE)****, **Generalized Cross Validation (GCV)*****.

*Wen et. al, 2012; **Lin et. al, 2012; ***Liao et. al, 2009

Residual properties

$$\mathbb{E}(\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2) = m\sigma^2$$

suggests solving nonlinear equation

$$f(\lambda) = \|\mathbf{y} - \mathbf{A}\mathbf{x}(\lambda)\|_2^2 - m\sigma^2 = 0$$

for λ. However

$$\hat{f}(\lambda) = \|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}(\lambda)\|_2^2 - m\sigma^2$$

is biased so choosing λ by solving $\hat{f}(\lambda) = 0$ leads to oversmoothing (Discrepancy principle).

Effective Degrees of Freedom (EDF)

Tikhonov regularization

$$\hat{\mathbf{x}}_{ls} \in \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{D}\mathbf{x}\|_2^2$$

gives ridge regression estimator

$$\hat{\mathbf{x}}_{ls} = (\mathbf{A}^T\mathbf{A} + \lambda\mathbf{D}^T\mathbf{D})^{-1}\mathbf{A}^T\mathbf{y}.$$

Predictors are

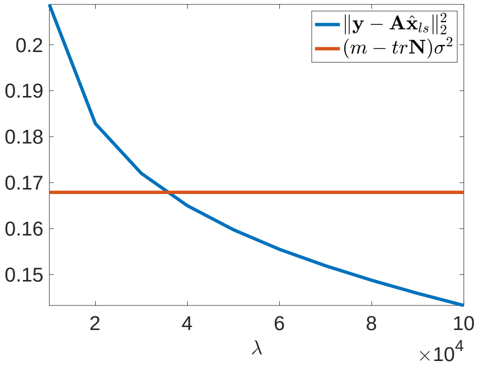
$$\mathbf{A}\hat{\mathbf{x}}_{ls} = \mathbf{N}(\lambda)\mathbf{y}$$

and*

$$\mathbb{E}(\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{ls}\|_2^2) = (m - \mathrm{tr}\mathbf{N}(\lambda))\sigma^2$$

*Hall et. al, 1987.

MRI example



$\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{ls}\|_2^2$
$(m - tr\mathbf{N})\sigma^2$
0.2
0.19
0.18
0.17
0.16
0.15
2
4
6
8
10
λ
$\times 10^4$

Degrees of Freedom in Nonlinear Regression

Tikhonov regularization term defines smoothing matrix $\mathbf{A}\hat{\mathbf{x}}_{ls} = \mathbf{N}\mathbf{y}$ while nonlinear smoothers (e.g. TV) have

$$\mathbf{A}\hat{\mathbf{x}}_{tv} = \delta(\mathbf{y}).$$

Degrees of freedom of δ are given by*

$$df(\mathbf{A}\hat{\mathbf{x}}_{tv}) = \sum_{i=1}^{m} \text{cov}((\hat{\mathbf{x}}_{tv})_i, \mathbf{y}_i)/\sigma^2$$

e. g. $\mathrm{df}(\mathbf{A}\hat{\mathbf{x}}_{ls}) = \mathrm{tr}(\mathbf{N})$.

*Efron, 2004.

Test Statistic

- Unbiased estimate of prediction risk

$$\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}\|_2^2 + (\omega_m df(\mathbf{A}\hat{\mathbf{x}}) - m)\sigma^2$$

AIC: $\omega_n = 2$, BIC: $\omega_m = \log(m)$

e.g. $\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{ls}\|_2^2 + (tr(\mathbf{N}) - m)\sigma^2$

Degrees of Freedom for TV (generalized Lasso)*,**

$$\hat{\mathbf{x}}_{tv} \in \arg\min_{\mathbf{x}} \frac{\lambda}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \|\mathbf{D}\mathbf{x}\|_1$$

No assumptions on $\mathbf{A}$ or $\mathbf{D}$:

$$df(\mathbf{A}\hat{\mathbf{x}}_{tv}) = \mathbb{E}[dim(\mathbf{A}(null(\mathbf{D}_{-\mathcal{A}})))], \quad \mathcal{A} = \{i : (\mathbf{D}\hat{\mathbf{x}}_{tv})_i \neq 0\}$$

If $\mathbf{A}$ is full rank, the number of non-zero predictors $\mathbf{D}\hat{\mathbf{x}}_{tv}$ is an unbaised estimate for $df(\mathbf{A}\hat{\mathbf{x}}_{tv})$.

---

*Tibshirani 2012;**Dossal 2013.

TV Degrees of Freedom Example - no groupings

- $nullity(\mathbf{D}_{-\mathcal{A}}) = n \rightarrow df(\mathbf{A}\hat{\mathbf{x}}_{tv}) = n$

$$\hat{\lambda} \in \arg\min_{\lambda} \|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{tv}\|_2^2 + (\omega_m n - m)\sigma^2$$

- Similar λ choice as χ^2 test for Tikhonov

$$\frac{\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{ls}\|_2^2}{\sigma^2} \sim \chi^2_{m-n}$$

Issue if $m \leq n$, instead use regularized residual*

$$\frac{\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{ls}\|_2^2}{\sigma^2} + \lambda\|\mathbf{D}\hat{\mathbf{x}}_{ls}\|_2^2 \sim \chi^2_m$$

*Mead et. al, 2008, 2009, 2013.

χ^2 Test for TV Functional

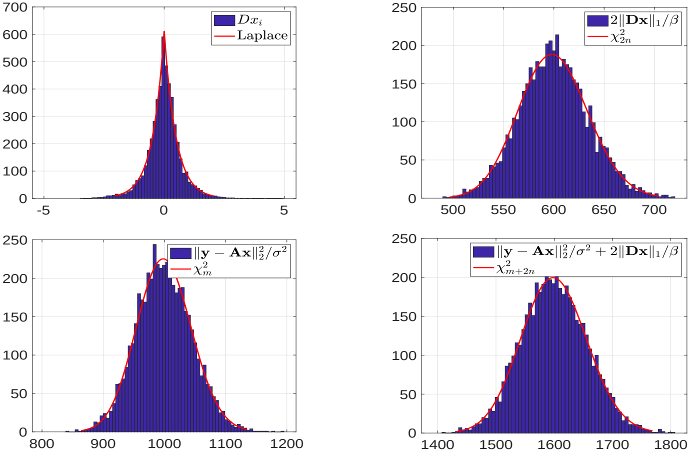
If $y_i \sim$ Laplace(θ, β) all independent, then

$$\sum_{i=1}^{n} \frac{2|y_i - \theta|}{\beta} \sim \chi^2_{2n}.$$

Since the TV functional is differentially Laplacian

$$\frac{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2}{\sigma^2} + \frac{2\|\mathbf{D}\mathbf{x}\|_1}{\beta} \sim \chi^2_{m+2n}$$

Histograms illustrating χ^2 Test for TV Functional



χ^2 Test for TV Estimate

$$\frac{\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{tv}\|_2^2}{\sigma^2} + \frac{2\|\mathbf{D}\hat{\mathbf{x}}_{tv}\|_1}{\beta} \sim \chi^2_{m - df(\mathbf{A}\hat{\mathbf{x}}_{tv}) + df(\mathbf{D}\hat{\mathbf{x}}_{tv})}$$

Since $\mathbf{D}$ is full rank

$$df(\mathbf{D}\hat{\mathbf{x}}_{tv}) = df(\hat{\mathbf{x}}_{tv}) = \mathbb{E}[nullity(\mathbf{D}_{-\mathcal{A}})]$$

e.g. if $\mathbf{A}$ is full rank ($\lambda = \frac{\beta}{\sigma^2}$)

$$\hat{\lambda} \in \arg\min_{\lambda} \|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}_{tv}\|_2^2 + \frac{2\|\mathbf{D}\hat{\mathbf{x}}_{tv}\|_1}{\lambda} - m\sigma^2$$

Numerical Tests - Evaluating Image Quality

MRI image filtered with a 15×15 uniform blur

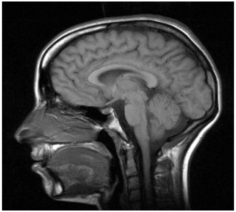
Input noise:

$$\mathbf{BSNR} = 20 \log_{10} \frac{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2}{m\sigma}$$

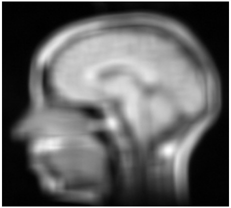
Recovered image quality:

$$\mathbf{ISNR} = 20 \log_{10} \frac{\|\mathbf{y} - \mathbf{x}\|_2}{\|\hat{\mathbf{x}} - \mathbf{x}\|_2}$$

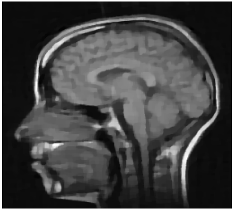
Exact



BSNR $= 40$

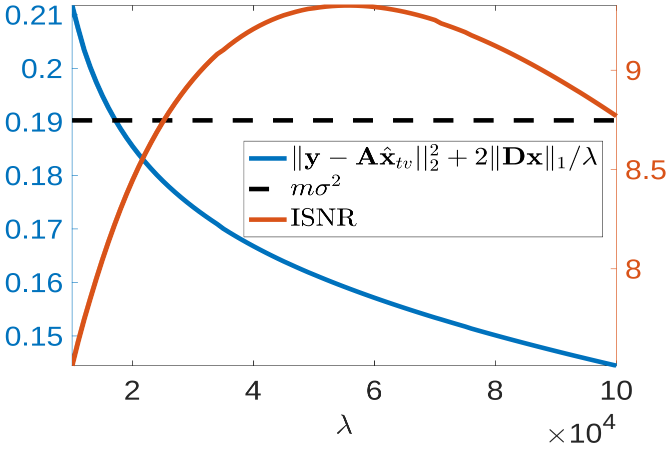


χ^2 ISNR $= 8.22$

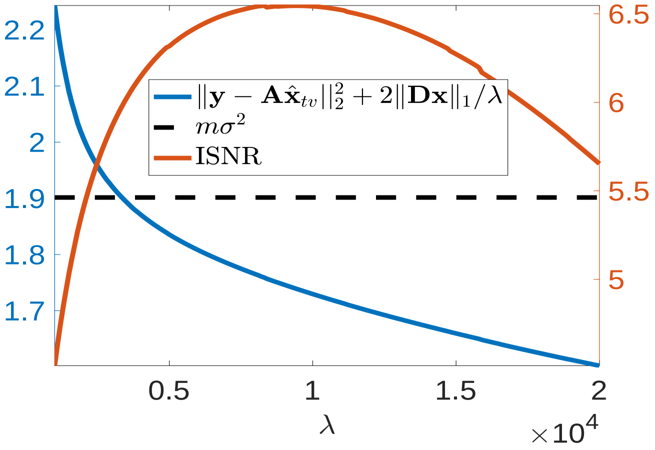


Maximum ISNR $= 9.33$

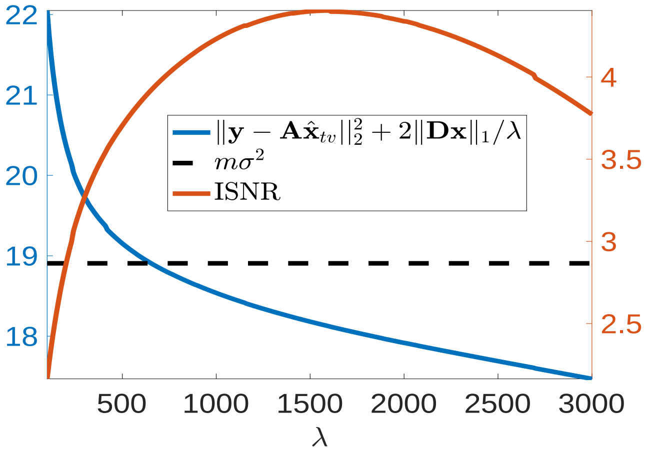
MRI BSNR = 40; χ^2 ISNR = 8.22; Max ISNR = 9.33



MRI BSNR = 30; χ^2 ISNR = 5.83; Max ISNR = 6.51



MRI BSNR = 20; χ^2 ISNR = 3.83; Max ISNR = 4.31



Summary and Conclusions

- We have developed a framework for automatic and efficient selection of TV regularization parameters. The approach extends results on residuals and risk estimators, in particular
  - The new measure of risk involves the regularized residual which follows a χ^2 distribution.
  - The degrees of freedom can be estimated from recent results on degrees of freedom for generalized Lasso.

- The proposed TV regularization parameter selection method* requires a data noise estimate and solves the TV problem multiple times during an optimization, rather than guess and checking.

*Mead, in revision, J. Inv. Imag.