L-curve for Tikhonov Regularization

Keith Woodbury 11-06-2012

These notes are about generating an L-curve for optimal selection of the Tikhonov regularization parameter α_0. Reference Hansen pg 83 ff.

The "L-curve" is a log-log plot of the norm of the Tikhonov penalty term for the regularized solution on the vertical axis against the norm of the residuals. The Tikhonov objective function for zeroth order regularization is

$$S = (\mathbf{Y} - \mathbf{Xq})^T (\mathbf{Y} - \mathbf{Xq}) + \alpha_0 \mathbf{q}^T \mathbf{q}$$

and for first order regularization is

$$S = (\mathbf{Y} - \mathbf{Xq})^T (\mathbf{Y} - \mathbf{Xq}) + \alpha_1 (\mathbf{Hq})^T (\mathbf{Hq})$$

where **H** is the first derivative operator

$$\mathbf{H} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & 0 \\ 0 & \cdots & -1 & 1 & \vdots \\ \vdots & \cdots & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Zeroth Order L-curve

Hansen's book [1] is used as a reference for the L-curve.

The vertical axis is

$$\begin{aligned} \left\| \mathbf{L} x_{reg} \right\|_2 &= sqrt\left(\hat{\mathbf{q}}^T \hat{\mathbf{q}} \right) = sqrt\left((\mathbf{FY})^T (\mathbf{FY}) \right) = sqrt\left((\mathbf{F}[\mathbf{T} + \boldsymbol{\varepsilon}])^T (\mathbf{F}[\mathbf{T} + \boldsymbol{\varepsilon}]) \right) \\ &= sqrt\left((\mathbf{F}[\mathbf{Xq} + \boldsymbol{\varepsilon}])^T (\mathbf{F}[\mathbf{Xq} + \boldsymbol{\varepsilon}]) \right) \\ &= sqrt\left((\mathbf{Pq} + \mathbf{F}\boldsymbol{\varepsilon})^T (\mathbf{Pq} + \mathbf{F}\boldsymbol{\varepsilon}) \right) \end{aligned}$$

What is the expected value (under standard statistical assumptions zero mean, constant variance, uncorrelated)?

$$\begin{aligned} E\left((\mathbf{Pq} + \mathbf{F}\boldsymbol{\varepsilon})^T (\mathbf{Pq} + \mathbf{F}\boldsymbol{\varepsilon}) \right) &= E\left((\mathbf{Pq})^T (\mathbf{Pq}) + (\mathbf{F}\boldsymbol{\varepsilon})^T (\mathbf{Pq}) + (\mathbf{Pq})^T (\mathbf{F}\boldsymbol{\varepsilon}) + (\mathbf{F}\boldsymbol{\varepsilon})^T (\mathbf{F}\boldsymbol{\varepsilon}) \right) \\ &= E\left((\mathbf{Pq})^T (\mathbf{Pq}) \right) + E\left(\boldsymbol{\varepsilon}^T \mathbf{F}^T \mathbf{F} \boldsymbol{\varepsilon} \right) \\ &= \mathbf{q}^T \mathbf{P}^T \mathbf{Pq} + \sigma_Y^2 tr\left(\mathbf{F}^T \mathbf{F} \right) \end{aligned}$$

The horizontal axis is

$$\left\|\mathbf{A}x_{reg} - \mathbf{b}\right\|_2 = sqrt\left((\mathbf{X}\hat{\mathbf{q}} - \mathbf{Y})^T(\mathbf{X}\hat{\mathbf{q}} - \mathbf{Y})\right) = sqrt\left((\mathbf{XFY} - \mathbf{Y})^T(\mathbf{XFY} - \mathbf{Y})\right)$$
$$= sqrt\left(\mathbf{Y}^T(\mathbf{R} - \mathbf{I})^T(\mathbf{R} - \mathbf{I})\mathbf{Y}\right) = sqrt\left(\mathbf{Y}^T\mathbf{B}^T\mathbf{B}\mathbf{Y}\right) = sqrt\left(\mathbf{Y}^T\mathbf{\Phi}\mathbf{Y}\right)$$

The expected value of this is exactly $E(R_T)$

$$E(R_T) = E(\mathbf{Y}^T\mathbf{\Phi}_T\mathbf{Y}) = \mathbf{q}^T((\mathbf{I} - \mathbf{R})\mathbf{X})^T\left((\mathbf{I} - \mathbf{R})\mathbf{X}\right)\mathbf{q} + \sigma_Y^2 tr((\mathbf{I} - \mathbf{R})^T(\mathbf{I} - \mathbf{R}))$$

Note that both the vertical axis and the horizontal axis contain both a bias term and a random error term. This immediately suggests that the L-curve will not yield a precise estimate of the balance between bias and random errors.

Figure 1 shows the L-curve for Tikhonov regularization using step-interpolation. The "optimal" value of α_0 using the L-curve (visual inspection) is α_0 = 3.8E-4. Minimizing $E(R_T)$ yields α_0 = 2.7E-4, while Morozov yields α_0 = 4.3E-4. Hansen [1] suggests that the optimal value at the elbow of the L-curve can be found by maximizing the curvature of the L-curve; this value is 8.3E-5 (found by computing finite difference approximation to the second derivative and finding the maximum).

Hansen [1] says that the L-curve solutions tend to be oversmooth:

> *"The conclusion from these experiments is that the L-curve criterion for Tikhonov regularization gives a very robust estimation of the regularization parameter, while the GCV method occasionally fails to do so. On the other hand, when GCV works it usually gives a very good estimate of the optimal regularization parameter, while the L-curve criterion tends to produce a regularization parameter that slightly over-smooths, i.e., it is slightly too large."*

References

[1] P. C. Hansen, "The L-Curve and its Use in the Numerical Treatment of Inverse Problems," in *Computational Inverse Problems in Electrocardiology*, 2000.

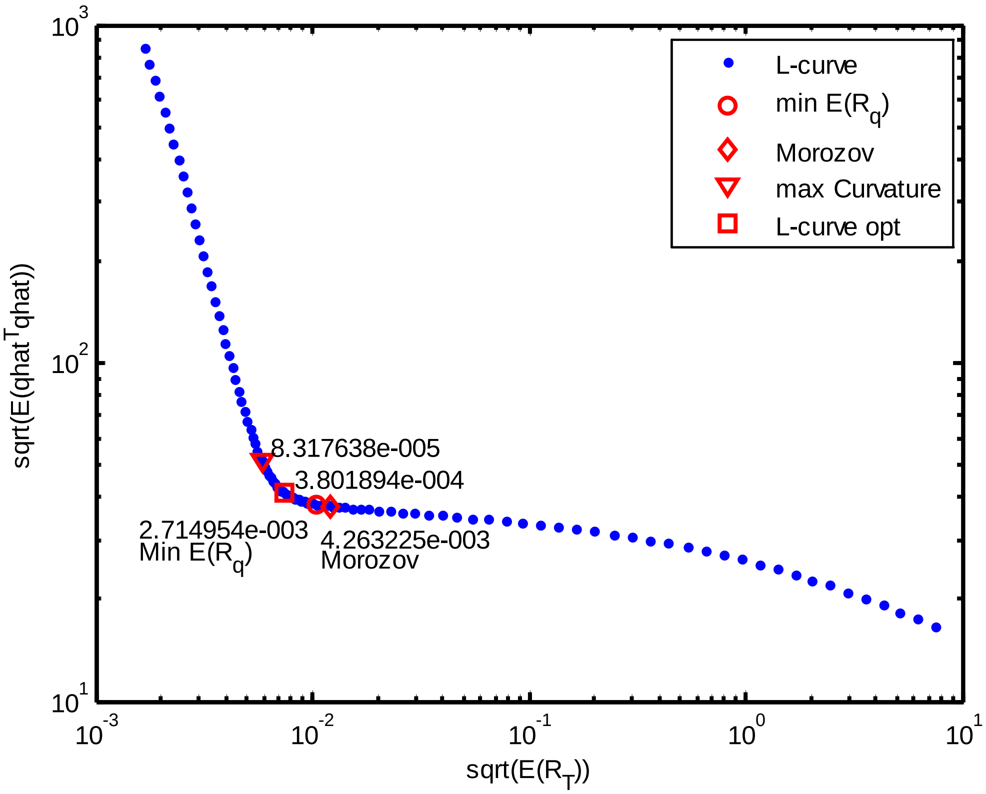

Figure 1. "L-curve" for 1e-6 < α_0< 0.9. Step interpolation, Δt = 0.05, x=L.