On L-curve computation

Quanjun Lang, Fei Lu

October-December 2022

Abstract

L-curve method by Hansen computes the optimal hyper-parameter for regularization. Here we review the method and its use in the data-adaptive RKHS Tikhonov regularization (DARTR).

Hansen's L-curve method [Han00] selects the optimal hyper-parameter for Tikhonov regularization. The optimal parameter maximizes the curvature of the curve the norm of the regularized solution versus the residual norm. The advantages of the L-curve criterion are robustness and ability to treat perturbations consisting of correlated noise. The disadvantages are over-smoothing (the smoother the solution, the worse the optimal λ by the L-curve) and inconsistency when dimension increases. The analysis of L-curve regularization is complicated because it depends on the linear system and the way

"Every practical method has its advantages and disadvantages. "

1 L-curve for LS with l2 norm

Consider first the simplest l2 regularization for least squares (LS), for which we minimize the following loss,

$$L_\lambda(x) = \|Ax - b\|^2 + \lambda \|x\|^2 .$$

Here the LSE is $x_{reg} = (A^\top A + \lambda I)^{-1} A^\top b$. We consider the SVD of A such that $A = USV^\top$ with $U^\top U = I$ and $V^\top V = I$. Then

$$x_{reg} = V(S^2 + \lambda I)^{-1} V^\top \left(VSU^\top b\right) = V\left[(S^2 + \lambda I)^{-1} S\right] U^\top b$$

For given value of λ, the loss/residual of the LSE becomes

$$E(\lambda) = \|Ax_{reg} - b\| = \left\|U\left[(S^2 + \lambda I)^{-1} S^2\right] U^\top b - UU^\top b\right\| = \left\|\left[(S^2 + \lambda I)^{-1} \lambda I\right] U^\top b\right\|$$

and the norm of the regularized solution is

$$R(\lambda) = \|x_{reg}\| = \left\|\left[(S^2 + \lambda I)^{-1} S\right] U^\top b\right\| .$$

The L-curve is a log-log plot of the curve

$$\text{L-curve} : l(\lambda) = (y(\lambda), x(\lambda)) := (\log R(\lambda), \log E(\lambda)).$$

The L-curve method maximize the curvature of the L-curve to reach a balance between minimization of the residual and control of the regularization norm.

$$\lambda_0 = \mathrm{argmax}_{\lambda_{\min} \le \lambda \le \lambda_{\max}} \kappa(\mathrm{l}(\lambda)) = \mathrm{argmax}_{\lambda_{\min} \le \lambda \le \lambda_{\max}} \frac{\mathrm{x}'\mathrm{y}'' - \mathrm{x}'\mathrm{y}''}{(\mathrm{x}'^{\,2} + \mathrm{y}'^{\,2})^{3/2}},$$

where λ_{min} and λ_{max} are pre-assigned or computed from the smallest and the largest generalized eigenvalues of A.

Hansen's function **lcfun** computes the curvature explicitly using the parameters

$$\{\lambda_n\}_{n=1}^N,\ \beta = U^\top b,\ \xi = S^{-1}\beta,\ s = diag(S). \tag{1.1}$$

eq:para_lcfun

Notice that U and S are from the SVD of A, and the output is the curvature of the L-curve at $\{\lambda_n\}_{n=1}^N$.

Note that we used λ instead of using λ^2. But this will not change the position of the maximal curvature of the L-curve in the (E, R) coordinate since the curvature does not depend on the parametrization.

1.1 Quadratic form

In applications, we often have a loss function in a quadratic form:

$$L = c^\top \overline{A} c - 2c^\top \overline{b} + \overline{b}^\top \overline{A}^{-1} \overline{b} + \lambda c^\top c.$$

We need to transfer the quadratic form into Hansen's setting so that we can use the explicit computation of the curvature. Taking

$$A = \overline{A}^{1/2},\ b = \overline{A}^{-1/2}\overline{b},$$

(or equivalently, $\overline{A} = A^\top A$, $\overline{b} = A^\top b$), we have

$$L = c^\top \overline{A} c - 2c^\top \overline{b} + \overline{b}^\top \overline{A}^{-1} \overline{b} + \lambda c^\top c = \|Ax - b\|^2 + \lambda \|c\|^2.$$

To supply the inputs of the function **lcfun** in (1.1), we need to find the SVD of $A = \overline{A}^{1/2}$. Since $\overline{A}$ is symmetric, we have the SVD of $\overline{A}$ as $\overline{A} = USU^\top$, hence $A = US^{1/2}U^\top$ and $\overline{A}^{-1/2} = US^{-1/2}U^\top$. Then, the inputs of **lcfun** are

$$\beta = U^\top b = S^{-1/2}U^\top \overline{b}, \quad \xi = S^{-1}U^\top \overline{b}, \quad s = \mathrm{diag}(S^{1/2}) \tag{1.2}$$

eq:lcfun_qf

and $\{\lambda_n\}_{n=1}^N$ are taken to be the square root.

Then Hansen used optimization to find the minimal curvature around the argmin in the λ-grid. The above part is contained in the code **L_curve_standard_form**.

2 General regularization norms

The L-curve method applies to general regularization norms in the form $\|c\|_\Sigma^2 = c^\top \Sigma^{-1} c$. The corresponding loss function is

$$Loss = c^\top \overline{A} c - 2c^\top \overline{b} + \overline{b}^\top \overline{A}^{-1} \overline{b} + \lambda (c - x)^\top \Sigma^{-1} (c - x).$$

Here Σ^{-1} is the basis matrix of the basis functions in the Hilbert space defined by the norm. When $\Sigma^{-1} = Id$, we get the $l2$ regularization in the previous section. We use Σ^{-1} in the regularization norm since it corresponds to the covariance matrix of a Bayesian Gaussian prior. Suppose $L = \sqrt{\Sigma}$ is the square root of Σ computed from the SVD of Σ. Take $y = L^{-1}(c - x)$ and $c = Ly + x$, we have

$$\begin{aligned} Loss &= c^\top \overline{A} c - 2c^\top \overline{b} + \overline{b}^\top \overline{A}^{-1} \overline{b} + \lambda (c - x)^\top \Sigma^{-1} (c - x). && (2.1)\\ &= (y^\top L^\top + x^\top)\overline{A}(Ly + x) - 2\overline{b}^\top (Ly + x) + \overline{b}^\top \overline{A}^{-1} \overline{b} + \lambda y^\top y && (2.2)\\ &= y^\top (L^\top \overline{A} L) y - 2y^\top (L^\top \overline{b} - L^\top \overline{A} x) + \lambda y^\top y + Const. && (2.3) \end{aligned}$$

The problem can be reduced to the standard form using $\tilde{A} = L^\top \overline{A} L$ and $\tilde{b} = L^\top \overline{b} - L^\top \overline{A} x$. But the result is for y. Then use $c = Ly + x$ to get the estimation.

In short, the computation procedure is as follows. Given $\overline{A}, \overline{b}, \Sigma, x$:

1. SVD for Σ, $\Sigma = U_0 S_0 U_0^\top$ and set $L = U_0 S_0^{1/2} U_0^\top$.
2. SVD for $\tilde{A} = L^\top \overline{A} L = USV$, and compute $\tilde{b} = L^\top(\bar{b} - \overline{A}x)$.
3. Get λ_{opt} by using **lcfun** with inputs $\{\lambda_n\}$, $\beta = S^{-1/2}U^\top \tilde{b}$, $\quad \xi = S^{-1}U^\top \tilde{b}$, $\quad s = \mathrm{diag}(S^{1/2})$.
4. Get estimator $c = Ly_* + x$ where $y_* = (\tilde{A} + \lambda_{opt} I)\tilde{b}$.

2.1 DARTR

In DARTR, $\Sigma = B^{-1}\overline{A}B^{-1}$ with B being the basis matrix in $L^2(\rho)$, $x = 0$ and $L = B^{-1/2}\overline{A}^{1/2}B^{-1/2}$. Hence In particular, when $B = I$, we have $L = \overline{A}^{1/2}$, and $\tilde{A} = \overline{A}^2$ and $\tilde{b} = \overline{A}^{-1/2}\bar{b}$.

The computation procedure is as follows. Given $\overline{A}, \bar{b}, B, x$:

1. SVD for $\Sigma = B^{-1}\overline{A}B^{-1}$, $\Sigma = U_0 S_0 U_0^\top$ and set $L = U_0 S_0^{1/2} U_0^\top$.
2. SVD for $\tilde{A} = L^\top \overline{A} L = USV$, and compute $\tilde{b} = L^\top(\bar{b} - \overline{A}x)$.
3. Get λ_{opt} by using **lcfun** with inputs $\{\lambda_n\}$, $\beta = S^{-1/2}U^\top \tilde{b}$, $\quad \xi = S^{-1}U^\top \tilde{b}$, $\quad s = \mathrm{diag}(S^{1/2})$.
4. Get estimator $c = Ly_*$ where $y_* = (\tilde{A} + \lambda_{opt} I)^{-1}\tilde{b}$.

Remark1: a major issue is inversion of ill-conditioned matrix. SVD is in general better than $A\backslash b$, pinv or least squares. Thus, we compute the inversion B^{-1} by SVD to reduce numerical errors when B is ill-conditioned. A numerical stable way to implement Step 1 is: first, compute $L_0 = \mathrm{chol}(B)$ and do SVD for $L_0^{-1}\overline{A}L_0^{-1} = U_0 S_0 U_0^\top$. Then, get $\Sigma = B^{-1}\overline{A}B^{-1} = L_0^{-1}U_0 S U_0^\top L_0^{-\top}$.

Remark2: Our goal is solve $(\overline{A} + \lambda B\overline{A}^{-1}B)^{-1}\bar{b}$ with an optimal λ. Let $D = B^{-1}\overline{A}^{1/2}$. Then, $B\overline{A}^{-1}B = D^{-\top}D^{-1}$ and $(\overline{A} + \lambda B\overline{A}^{-1}B) = D^{-\top}(D^\top \overline{A} D + \lambda I)D^{-1}$. Here D is the same as the L in step 1.

References

sen_LcurveIts_a [Han00] Per Christian Hansen. The L-curve and its use in the numerical treatment of inverse problems. In *in Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering*, pages 119–142. WIT Press, 2000.