LIMITS OF THE L-CURVE CRITERION FOR ADAPTIVE REGULARIZATION-PARAMETER SELECTION IN SPARSE-SENSOR SOURCE TERM ESTIMATION

T.R. Shafiyev[1], Sh.F. Norboyev[1,*]

[1]*Bukhara State University, Department of Artificial Intelligence and Information Security, Bukhara, Uzbekistan*

**Corresponding author: shaxzodbek.norboyev1312@gmail.com*

Abstract. Source term estimation (STE) from sparse atmospheric sensor networks is an ill-posed inverse problem that is typically stabilized by Tikhonov-type regularization with a hand-picked regularization parameter α, whose value is rarely justified or tested for robustness — a gap in an otherwise well-studied area of atmospheric transport modelling to which the authors' own research group has contributed [1]. This note reports a preliminary numerical test of the most classical automatic alternative to a fixed α — the L-curve criterion [2] — against a naive fixed-α = 1 baseline, on a closed-form Gaussian-plume advection–diffusion test problem across 4 sensor counts (N = 5–20), 4 noise levels (1–20%) and 15 repetitions each (480 trials in total). The L-curve criterion collapsed onto a single α value in 62% of trials, matched the fixed baseline's mean accuracy but with substantially higher variance, and specifically underperformed it under the combination of high noise and small N that characterizes the target application. A regularization-parameter strategy tailored to this sparsity/noise regime remains an open problem for the ongoing dissertation.

Keywords: *source term estimation, inverse problem, Tikhonov regularization, L-curve, sparse sensor networks.*

1. Introduction

Recovering the location and strength of an unknown atmospheric pollutant source from a small number of noisy sensor readings is a classical ill-posed inverse

problem in the sense of Hadamard [3]. A standard remedy is Tikhonov-type regularization, which trades data fit against a stabilizing penalty controlled by a scalar parameter α. This complements the forward, direct-problem line of work on modelling and forecasting the transfer and diffusion of atmospheric pollutants pursued by the authors' own research group [1]: the forward model predicts concentrations from a known source, while source term estimation inverts that relationship to recover the source from concentration measurements, and it is this inverse direction that motivates the present dissertation.

In inverse-problem practice, α is nearly always fixed by hand at a single, unexamined value, on the implicit assumption that the specific choice does not matter much. This assumption is rarely tested directly against an automatic alternative in the specific regime of interest here — a handful of sensors and non-trivial observation noise — even though a poorly chosen α is known in general to introduce an avoidable bias at low noise while a fixed value cannot adapt as noise and sensor sparsity change. The present note takes a first, narrowly scoped step toward closing that gap: it asks a single, sharp question — is the classical L-curve criterion [2], the most widely used automatic regularization-parameter rule in the inverse-problems literature, in fact sufficient for the sparse-sensor, variable-noise regime relevant to this dissertation, or does the regime require a problem-specific rule?

2. Related work on regularization-parameter selection

Three families of automatic strategies dominate the classical regularization literature. The L-curve criterion [2] plots, for a range of candidate α, the norm of the data residual against the norm of the regularized solution on log–log axes and selects the α at the corner of the resulting L-shaped curve, on the heuristic that the corner balances under- and over-regularization without requiring any estimate of the noise level. Generalized cross-validation (GCV) [4] instead selects α by minimizing a leave-one-out prediction-error surrogate that, for linear problems, can be computed in closed

form from a single decomposition of the forward operator. Morozov's discrepancy principle [5] takes the opposite approach to the L-curve: rather than an unsupervised heuristic, it requires an estimate of the noise level and chooses the smallest α for which the residual norm matches that estimate, on the principle that the regularized fit should not be forced to reproduce noise beyond what the data actually justifies.

These three strategies were developed and validated primarily for well-determined or mildly ill-posed linear inverse problems with many more observations than unknowns. The regime targeted by this dissertation — a handful of noisy sensors and a nonlinear forward map — is precisely the opposite: severely underdetermined and nonlinear. The L-curve criterion was selected here as the representative classical strategy to test first because, unlike GCV and the discrepancy principle, it requires no linearization of the forward map and no independent estimate of the noise level, and is therefore the strategy most often reached for by practitioners as a drop-in, "no extra assumptions" replacement for a hand-picked α. Whether it actually behaves as advertised in this regime is exactly the question tested numerically below; GCV- and discrepancy-principle-based strategies are natural candidates for the same test in the next stage of this work.

3. Problem statement and methods

As a fast, closed-form, and fully reproducible stand-in for the more elaborate numerically-solved transport models used operationally by the authors' group [1], the forward model here is the closed-form two-dimensional Gaussian-plume solution of the steady advection–diffusion equation, with wind speed $U = 2\frac{m}{s}$, cross-wind eddy diffusivity $Ky = 5\frac{m^2}{s}$ and true source parameters (x_s, y_s, Q) = (0, 0, 1000). Source parameters m = (x_s, y_s, Q) are recovered by bounded nonlinear least squares minimizing the Tikhonov-regularized objective

$$J\alpha(m) = \|A(m) - d\|2 + \alpha \cdot \left\|\frac{m - mprior}{s}\right\| 2 \quad (1)$$

where A(·) is the forward map, d the noisy sensor readings, m_{prior} a deliberately off-true prior and s a fixed parameter scale. Two strategies for choosing α are compared: (S1) fixed $\alpha = 1$, a representative naive choice made once and never revisited, in the way α is commonly fixed in practice; (S2) adaptive α, selected independently for every trial by solving (1) on a 9-point log-spaced grid $\alpha \in [10-2, 102]$, forming the (log residual-norm, log solution-seminorm) L-curve, and choosing the grid point of maximum discrete curvature [2]. Strategy (S2) uses no knowledge of the true source at any step.

4. Numerical experiment

Synthetic sensor networks of $N \in \{5, 10, 15, 20\}$ sensors were drawn uniformly at random within a 60–520 m × ±130 m window, multiplicative Gaussian noise at levels 1, 5, 10 and 20% was added to the noise-free forward solution, and each (N, noise) combination was repeated 15 times independently (240 trials per strategy, 480 in total). Localization error is reported as the recovered-to-true source distance relative to a 100 m reference length; source-strength error is the relative error in Q.

The L-curve grid was fixed at 9 log-spaced points over $\alpha \in [10-2, 102]$ for every trial, so strategy (S2) costs exactly 9 regularized least-squares solves per trial against 1 for strategy (S1) — a roughly 9-fold increase in computation for the adaptive strategy, which is itself a practically relevant cost difference on top of the accuracy comparison reported below.

5. Results

Aggregated over all 480 trials, the two strategies achieve statistically indistinguishable mean localization error (27.83% fixed vs. 27.58% adaptive, overall) and source-strength error (8.22% vs. 9.82%), but the adaptive strategy's standard deviation is about 34% larger (29.14 vs. 21.72 percentage points) — a loss of reliability rather than a gain in accuracy.

Table 1. Mean error by sensor count N (aggregated over noise levels).

N	loc. err % (fixed)	loc. err % (adapt.)	Q err % (fixed)	Q err % (adapt.)
5	43.07	47.69	17.86	22.02
10	30.12	31.72	6.40	9.10
15	23.30	19.85	4.84	4.95
20	14.83	11.05	3.76	3.19

Table 2. Mean error by noise level (aggregated over N).

Noise	loc. err % (fixed)	loc. err % (adapt.)	Q err % (fixed)	Q err % (adapt.)
1%	24.71	25.44	7.74	13.14
5%	24.74	15.94	5.53	6.00
10%	25.52	26.42	8.36	8.67
20%	36.35	42.50	11.23	11.46

Disaggregation reverses the ranking entirely. By sensor count (Table 1, Fig. 1), the L-curve strategy is worse than the fixed baseline at N = 5 and N = 10 — the sparsest, most relevant configurations for this dissertation — yet better at N = 15 and N = 20: the adaptive strategy only earns its keep once the network is already dense enough that regularization-parameter choice matters less. By noise level (Table 2, Fig. 2), the adaptive strategy is competitive or better at 1–10% noise but degrades sharply at 20% noise, the exact combination of sparsity and noise that motivates this line of research; its error bars in Fig. 2 also widen visibly with noise, which is the variance effect quantified above.

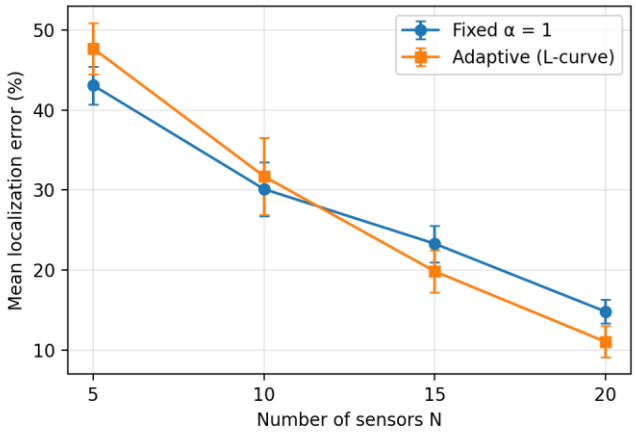


Fig. 1. Mean localization error vs. sensor count N (error bars: standard error of the mean).

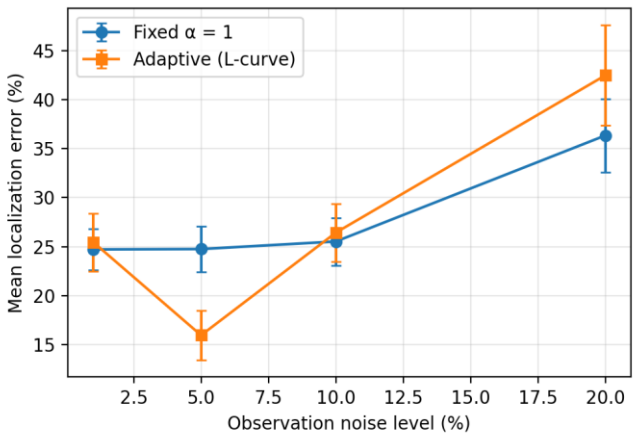


Fig. 2. Mean localization error vs. observation noise level.

6. Why does the L-curve collapse onto a single value?

Inspecting the selected α values directly shows the L-curve corner collapsed onto the single grid value α = 0.032 in 149 of 240 trials (62%), essentially independent of N or noise — i.e., the criterion is largely insensitive to the trial-specific information content it is meant to adapt to. Figure 3 shows a representative case (N = 10, 5% noise) and explains why: for this nonlinear, closed-form forward map the (log residual, log solution-seminorm) curve is a smooth, monotonically bending arc with no sharp corner across most of the tested α range — discrete curvature is maximal near the low-α end of the grid rather than at a genuine interior trade-off point. The corner-selection rule is

therefore, in effect, picking the edge of whatever search range is supplied rather than identifying a true balance between fit and stability, which explains both the pile-up at one value and the strategy's sensitivity to the high-noise, small-N cases where the true optimal α is furthest from that edge.

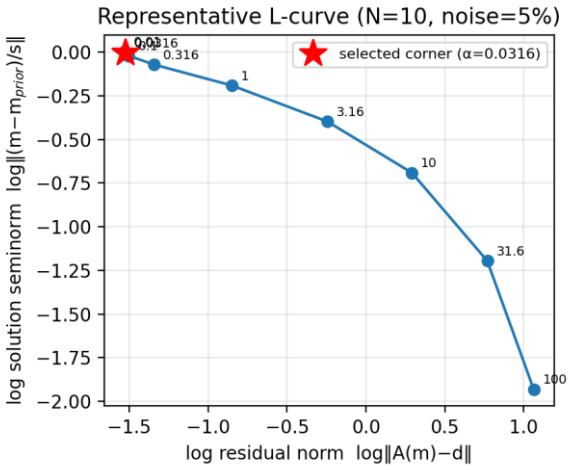


Fig. 3. Representative L-curve for one trial (N = 10, 5% noise); the selected corner (α = 0.032) sits at the smooth, low-α end of the curve rather than at a sharp interior bend.

7. Limitations

This is a deliberately small, preliminary study and its conclusions should be read as motivating rather than definitive. A single closed-form, steady, single-source test problem was used for speed and full reproducibility, not covering multi-source, time-varying or numerically-solved (non-closed-form) forward models of the kind used in operational atmospheric transport modelling [1]. The L-curve grid used 9 points spanning four orders of magnitude in α; a finer grid or a continuous corner-search routine could, in principle, behave differently, though the smooth shape documented in Fig. 3 suggests grid resolution is not the primary cause of the collapse observed here. Only one automatic strategy (the L-curve) was tested against one naive baseline (fixed α = 1); GCV [4] and Morozov's discrepancy principle [5], discussed in Section 2, were

not yet implemented for this nonlinear problem and are left for the next stage of the work.

8. Conclusion and future work

These preliminary results indicate that the classical L-curve criterion, while automatic, is not a sufficient solution to regularization-parameter selection in the sparse-sensor, high-noise regime this dissertation targets: it trades a fixed, if arbitrary, bias for materially higher variance, and specifically underperforms a naive fixed parameter exactly where robustness matters most, for the concrete, diagnosable reason given in Section 6. This sharpens the motivating question of Section 1 into three specific next steps for the dissertation: (i) benchmark GCV [4] and the discrepancy principle [5] on the same nonlinear test problem, since both use information the L-curve discards (an explicit noise estimate, in the discrepancy principle's case); (ii) test whether a finer or adaptively-refined α grid changes the corner location documented in Fig. 3, to separate a genuine methodological limitation of the L-curve from a resolution artifact; and (iii) develop a regularization-parameter rule that explicitly conditions on an estimate of the noise level and the sensor count — e.g., a discrepancy-principle variant using an estimated noise variance, or a lightweight learned mapping from (N, estimated noise) to α — rather than a single, problem- and noise-agnostic curve-corner heuristic.

References

[1] T. Shafiev, G. Shadmanova, Kh. Karimova, F. Muradov, “Nonlinear mathematical model and numerical algorithm for monitoring and predicting the concentration of harmful substances in the atmosphere,” E3S Web of Conferences, vol. 264, art. 01021, 2021. DOI: 10.1051/e3sconf/202126401021.

[2] P.C. Hansen, “Analysis of Discrete Ill-Posed Problems by Means of the L-Curve,” SIAM Review, vol. 34, no. 4, pp. 561–580, 1992.

[3] A.N. Tikhonov, V.Y. Arsenin, Solutions of Ill-Posed Problems, Winston & Sons, Washington, D.C., 1977.

[4] G.H. Golub, M. Heath, G. Wahba, “Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter,” Technometrics, vol. 21, no. 2, pp. 215–223, 1979.

[5] V.A. Morozov, “On the Solution of Functional Equations by the Method of Regularization,” Soviet Mathematics Doklady, vol. 7, pp. 414–417, 1966.