Why Curvature in L-Curve: Combining Soft Constraints

Anibal Sosa, Martine Ceberio, and
Vladik Kreinovich

Cyber-ShARE Center
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
usosaaguirre@miners.utep.edu
mceberio@utep.edu, vladik@utep.edu

1. Inverse Problem: A Brief Reminder

- In science & engineering, we are interested in the state of the world, i.e., in the values of different quantities.
- Some of these quantities we can directly measure, but many quantities are difficult to measure directly.
- For example, in geophysics, we are interested in the density at different depths and different locations.
- In principle, we can drill a borehole and directly measure these properties, but this is very expensive.
- To find the values of such difficult-to-measure quantities $q = (q_1, \ldots, q_n)$, we:
  - measure auxiliary quantities $a = (a_1, \ldots, a_m)$ related to q_i by a known dependence $a_i = f_i(q_1, \ldots, q_n)$,
  - and then reconstruct the values q_j from these measurement results.

2. Enter Soft Constraints

- *Objective:* describe the constraint that the values q_j are consistent with the observations a_i.
- *Assumption:* measurement errors $a_i - f_i(q_1, \ldots, q_n)$ are indep. normal variables with 0 mean and same σ^2.
- *Resulting constraint:* $s \le s_0$, where

$$s \stackrel{\text{def}}{=} \sum_{i=1}^{m} (a_i - f_i(q_1, \ldots, q_n))^2.$$

- *Fact:* for each s_0, there is a certain probability that this constraint will be violated.
- *Soft constraints:* such constraints are called *soft*.
- *For convenience:* this constraint is sometimes described in a log scale, as $x \le x_0$, where $x \stackrel{\text{def}}{=} \ln(s)$.

3. Regularization

- Methods for taking additional regularity constraints into account are known as *regularization methods*.
- *Example:* In geophysics, the density values at nearby locations are usually close to each other: $q_j - q_{j'} \approx 0$.
- *Assumption:* differences $q_j - q_{j'}$ are indep. and normally distributed with 0 mean and the same σ_d^2.
- *Resulting constraint:* $t \le t_0$, where $t \stackrel{\text{def}}{=} \sum\limits_{(j,j')} (q_j - q_{j'})^2$.
- *In log scale:* $y \le y_0$, where $y \stackrel{\text{def}}{=} \ln(t)$.
- *How to combine constraints:* e.g., we can use the Maximum Likelihood method.
- *Result:* we find the values q_j that minimize the sum $s + \lambda \cdot t$, where λ depends on the variances σ^2 and σ_d^2.

4. How to Determine the Parameter λ

- *Fact:* for each λ, we can find $q_j(\lambda)$, and, based on this solution, compute $x(\lambda)$ and $y(\lambda)$.
- *Question:* what value λ shall we choose?
- *Often:* the curve $(x(\lambda), y(\lambda))$ has a turning point (is *L-shaped*).
- *In this case:* it is reasonable to select this turning point.
- *How to describe it:* it is a point where the absolute value $|C|$ of the curvature C is the largest:

$$C \stackrel{\text{def}}{=} \frac{x'' \cdot y' - y'' \cdot x'}{((x')^2 + (y')^2)^{3/2}}.$$

- *Fact:* this approach often works well.
- *Natural question:* explain why curvature works well.
- *What we do:* we show that reasonable properties select a class of functions that include curvature.

5. Analysis of the Problem: Scale-Invariance

- The numerical values of each quantity depend on the selection of a measuring unit a.
- If we change a to a new measuring unit c_a times smaller, then a_i and $a_i - f_i(q_1, \ldots, q_n)$ get multiplied by c_a.
- So, $s = \sum_{i=1}^{n} (a_i - f_i(q_1, \ldots, q_n))^2$ get multiplied by c_a^2, and $x = \ln(s)$ changes to $x + \Delta_x$, where $\Delta_x \stackrel{\text{def}}{=} \ln(c_a^2)$.
- If we change a measuring unit q by a new one c_q times smaller, then q_j and $q_j - q_{j'}$ get multiplied by c_q^2.
- Also $t = \sum (q_j - q_{j'})^2$, and $y = \ln(t)$ get multiplied by c_q^2, and $y = \ln(t)$ changes to $y + \Delta_y$
- Under these changes $x(\lambda) \to x(\lambda) + \Delta_x$ and $y(\lambda) \to y(\lambda) + \Delta_y$, and the curvature does not change.

6. Additional Invariance and Our Main Idea

- Instead of the original parameter λ, we can use a new parameter μ for which $\lambda = g(\mu)$.
- This re-scaling of a parameter does not change the curve itself and thus, does not change its curvature.
- *Our idea:* to describe all the functions which are invariant with respect to both types of re-scalings.
- *By a* parameter selection criterion *we mean a function* $F(x, y, x', y', x'', y'')$ *of six variables.*
- F *is* scale-invariant *if for all values* Δ_x *and* Δ_y,

$$F(x+\Delta_x, y+\Delta_y, x', y', x'', y'') = F(x, y, x', y', x'', y'');$$

- F *is* invariant w.r.t. parameter re-scaling *if for every function* $g(z)$*, for* $\widetilde{x}(\mu) = x(g(\mu))$, $\widetilde{y}(\mu) = y(g(\mu))$,

$$F(\widetilde{x}, \widetilde{y}, \widetilde{x}', \widetilde{y}', \widetilde{x}'', \widetilde{y}'') = F(x, y, x', y', x'', y'').$$

7. Main Result

Main Result. *A parameter selection criterion is scale-invariant and invariant w.r.t. parameter re-scaling if and only if it has the form*

$$F(x, y, x', y', x'', y'') = f\left(C(x, y, x', y', x'', y''), \frac{y'}{x'}\right)$$

for some function $f(C, z)$, where

$$C \stackrel{\text{def}}{=} \frac{x'' \cdot y' - y'' \cdot x'}{((x')^2 + (y')^2)^{3/2}}.$$

Comment. Once a criterion is selected, for each problem, we use the value λ for which the value

$$F(x(\lambda), y(\lambda), x'(\lambda), y'(\lambda), x''(\lambda), y''(\lambda))$$

is the largest.

8. Acknowledgments

This work was supported in part:

- by the National Science Foundation grants HRD-0734825 and DUE-0926721,
- by Grant 1 T36 GM078000-01 from the National Institutes of Health, and
- by UTEP's Computational Science program.

9. Proof

- For each tuple (x, y, x', y', x'', y''), by taking $\Delta_x = -x$ and $\Delta_y = -y$, we conclude that

$$F(x, y, x', y', x'', y'') = F(0, 0, x', y', x'', y'').$$

- Thus, we conclude that

$$F(x, y, x', y', x'', y'') = F_0(x', y', x'', y''),$$

where we denoted $F_0(x', y', x'', y'') \stackrel{\text{def}}{=} F(0, 0, x', y', x'', y'')$.

- So, we conclude that the value of the parameter selection criterion does not depend on x and y at all.
- In terms of the function F_0, invariance w.r.t. parameter re-scaling means that $F_0(\widetilde{x}', \widetilde{y}', \widetilde{x}'', \widetilde{y}'') = F_0(x', y', x'', y'')$.
- Re-scaling means that we go from the original function $x(\lambda)$ to the new function $\widetilde{x}(\mu) = x(g(\mu))$.

10. Proof (cont-d)

- In terms of the function F_0, invariance w.r.t. parameter re-scaling means that $F_0(\widetilde{x}', \widetilde{y}', \widetilde{x}'', \widetilde{y}'') = F_0(x', y', x'', y'')$.
- Re-scaling means that we go from the original function $x(\lambda)$ to the new function $\widetilde{x}(\mu) = x(g(\mu))$.
- The chain rule for differentiation leads to $\widetilde{x}' = x' \cdot g'$ and thus, $\widetilde{x}'' = x'' \cdot (g')^2 + x' \cdot g''$.
- Similarly, $\widetilde{y}' = y' \cdot g'$ and $\widetilde{y}'' = y'' \cdot (g')^2 + y' \cdot g''$.
- In particular, at the point where $g' = 1$, we have $\widetilde{x}' = x$, $\widetilde{x}'' = x'' + x' \cdot g''$, $\widetilde{y}' = y'$, and $\widetilde{y}'' = y'' + y' \cdot g''$.
- Thus, invariance w.r.t. parameter re-scaling means that $F_0(x', y', x'' + x' \cdot g'', y'' + y' \cdot g'') = F_0(x', y', x'', y'')$.
- In particular, for $g'' = -\dfrac{y''}{y'}$, we have $y'' + y' \cdot g'' = 0$ and thus, $F_0(x', y', x'', y'') = F_0\left(x', y', x'' - x' \cdot \dfrac{y''}{y'}, 0\right)$.

11. Proof (cont-d)

- We proved: $F_0(x', y', x'', y'') = F_0\left(x', y', x'' - x' \cdot \dfrac{y''}{y'}, 0\right)$.
- One can check that $x'' - x' \cdot \dfrac{y''}{y'} = C \cdot \dfrac{((x')^2 + (y')^2)^{3/2}}{y'}$.
- Thus, $F_0(x', y', x'', y'') = h(C, x', y')$, where

$$h(C, x', y') \stackrel{\text{def}}{=} F_0\left(x', y', C \cdot \frac{((x')^2 + (y')^2)^{3/2}}{y'}, 0\right).$$

- The curvature C is invariant w.r.t. parameter re-scaling.
- So, for $h(C, x', y')$, invariance means that $h(C, \widetilde{x}', \widetilde{y}') = h(C, x', y')$, i.e., $h(C, x', y') = h(C, x' \cdot g', y' \cdot g')$.
- In particular, for $g' = \dfrac{1}{x'}$, we have $x' \cdot g' = 1$ and thus,

$$h(C, x', y') = h\left(C, 1, \frac{y'}{x'}\right).$$

12. Proof (cont-d)

- We have proven:
  - that $F(x, y, x', y', x'', y'') = F_0(x', y', x'', y'')$,
  - that $F_0(x', y', x'', y'') = h(C, x', y')$, and
  - that $h(C, x', y') = h\left(C, 1, \frac{y'}{x'}\right)$.
- Thus, we conclude that $F(x, y, x', y', x'', y'') = h\left(C, 1, \frac{y'}{x'}\right)$.
- In other words, we get $F(x, y, x', y', x'', y'') = f\left(C, \frac{y'}{x'}\right)$

  for $f(C, z) \stackrel{\text{def}}{=} h(C, 1, z)$.
- The main result is thus proven.

References

[1] Hansen, P.C. Analysis of discrete ill-posed problems by means of the L-curve, SIAM Review 34(4), 561–580 (1992)

[2] Moorkamp, M., Jones, A.G., Fishwick, S.: Joint inversion of receiver functions, surface wave dispersion, and magnetotelluric data. Journal of Geophysical Research 115, B04318 (2010)

[3] Rabinovich, S.: Measurement Errors and Uncertainties: Theory and Practice. American Institute of Physics, New York (2005)

[4] Tikhonov, A.N., Arsenin, V.Y.: Solutions of Ill-Posed Problems, W.H. Whinston & Sons, Washington, D.C. (1977)