Assignment for Research & Development / AI

Robust Recovery of (θ, M, X) for a Rotated Exponential–Sinusoid Curve

Candidate: Vaigandla Hemanth Kumar

November 8, 2025

Given model.

$$x(t) = t\cos\theta \;-\; e^{M|t|}\,\sin(0.3t)\,\sin\theta \;+\; X, \tag{1}$$

$$y(t) = 42 \;+\; t\sin\theta \;+\; e^{M|t|}\,\sin(0.3t)\,\cos\theta, \tag{2}$$

Unknowns: θ, M, X with ranges $0^\circ < \theta < 50^\circ$, $-0.05 < M < 0.05$, $0 < X < 100$, and $6 < t < 60$. The CSV contains points sampled from (1)–(2).

Part I — Submission (required by the rubric)

Recovered Unknowns

$$\boxed{\theta = 30^\circ\ (=\ 0.523598776\text{ rad})}, \qquad \boxed{M = 0.030000000}, \qquad \boxed{X = 55.000000000}.$$

Submission String

$$\begin{cases} x(t) = t\cos(0.523598776) \;-\; e^{0.030000000|t|}\,\sin(0.3t)\,\sin(0.523598776) \;+\; 55.000000000, \\ y(t) = 42 \;+\; t\sin(0.523598776) \;+\; e^{0.030000000|t|}\,\sin(0.3t)\,\cos(0.523598776). \end{cases}$$

Angles inside $\sin, \cos$ *are in radians.* This is mathematically identical to a single long tuple, but formatted to fit within the page margins.

Key diagnostics (from fitted data)

- L1 residual (mean absolute error in rotated v–space): $\approx 2.6 \times 10^{-6}$.
- Recovered t range via u (see Sec.): $u_{\min} \approx 6.05$, $u_{\max} \approx 59.99$, consistent with $6 < t < 60$.

Part II — Detailed Explanation (end-to-end)

A. Geometry: why the curve can be "straightened"

Let $c = \cos\theta$ and $s = \sin\theta$, and translate by the known offsets:

$$x' = x - X, \qquad y' = y - 42.$$

Rotate by $-\theta$ (counter-rotation) using

$$\begin{bmatrix} u \\ v \end{bmatrix} = \underbrace{\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}}_{R(-\theta)} \begin{bmatrix} x' \\ y' \end{bmatrix}.$$

With $E(t) = e^{Mt}\sin(0.3t)$ and $|t| = t$ (since $t > 0$), (1)–(2) give

$$x' = tc - E(t)s, \qquad y' = ts + E(t)c.$$

Hence

$$u = c\,x' + s\,y' = t(c^2 + s^2) = \boxed{t}, \qquad v = -s\,x' + c\,y' = E(t)(s^2 + c^2) = \boxed{e^{Mt}\sin(0.3t)}.$$

Interpretation. The data in (x, y) are a rigid pose (rotation $+\theta$, translation $+X$ horizontally, $+42$ vertically) of a one-dimensional signal $v(u) = e^{Mu}\sin(0.3u)$. Recovering (θ, X, M) is therefore equivalent to (i) undoing the pose so that $u = t$, and (ii) matching $v(u)$'s amplitude and phase.

Identifiability and constraints. A line has 180° symmetry: θ and $\theta{+}180°$ define the same axis. The prior $0°{<}\theta{<}50°$ fixes this ambiguity and focuses the search to the small-tilt solution consistent with the data.

B. Why we avoid normalization/standardization here

The model contains a *fixed physical frequency* 0.3 (in radians per unit u) and a fixed vertical offset 42. Global scaling would distort these, making $\sin(0.3u)$ no longer match the data. We therefore **only** subtract the known offset ($y \to y{-}42$) and maintain the native units so the oscillation and envelope are preserved.

C. Robust preprocessing (geometry-aware)

Stage A — Mahalanobis screening in $(x, y{-}42)$. On the centered cloud $\big(x - \bar{x},\ (y-42) - \overline{(y-42)}\big)$, compute squared Mahalanobis distance D^2. Points with D^2 above a χ^2_2 cutoff (we use ≈ 13.816 for $p < 0.001$) are removed as gross outliers that don't follow the tilted geometry.

Stage B — Residual trimming in (u, v). After an initial fit (below), compute residuals $r_i = v_i - e^{Mu_i}\sin(0.3u_i)$. Remove points with $|r_i| > k \cdot \mathrm{MAD}(r)$ (we use $k{=}3$, MAD = median absolute deviation). This step isolates the clean oscillatory pattern before the final refinement.

Why this is safe. We never smooth away the oscillation; we only remove statistically inconsistent points. The final solution is then refined with a robust loss (Huber IRLS) that downweights any remaining outliers rather than deleting them.

D. Estimating the parameters (PCA → envelope → phase → robust LM)

1) Orientation θ via PCA/TLS. Compute the covariance of $(x,\ y{-}42)$ and take its principal eigenvector. In 2D, PCA's first component equals the orthogonal-regression (TLS) line. Its angle $\theta_0 = \arctan 2(v_y, v_x)$ (folded to $(0°, 50°)$) is our initial tilt estimate. *Intuition:* most variance is along the t direction; the sinusoid is a small orthogonal wobble.

2) Coarse X by centering u. With $c_0 = \cos\theta_0$, $s_0 = \sin\theta_0$, set $u^{\mathrm{raw}} = x\,c_0 + (y-42)\,s_0$ and

$$X_0 = \frac{\mathrm{mean}(u^{\mathrm{raw}}) - \frac{6+60}{2}}{c_0} = \frac{\mathrm{mean}(u^{\mathrm{raw}}) - 33}{c_0}.$$

This places the mid-u near the middle of $(6, 60)$.

3) Envelope slope for M. After rotating/translating with (θ_0, X_0), local maxima of $|v|$ follow the envelope e^{Mu}. Taking $\log|v|$ at separated peaks (distance $\approx \pi/0.3 \approx 10.47$ in u between successive *absolute* maxima) yields a straight line

$$\log|v_{\mathrm{peak}}| = \alpha + M\,u_{\mathrm{peak}},$$

whose slope gives M_0.

4) Phase alignment for X (1D). Holding (θ_0, M_0) fixed, align phase by solving the robust 1D problem

$$X_1 = \arg \min_{X \in (0,100)} \frac{1}{N} \sum_i \left| v_i(X) - e^{M_0 u_i(X)} \sin\big(0.3\, u_i(X)\big) \right| .$$

This centers the sinusoid horizontally without changing amplitude.

5) Joint nonlinear refinement (robust LM). Define

$$r_i(\theta, X, M) = v_i - e^{M u_i} \sin(0.3 u_i), \quad u_i = (x_i - X)\cos\theta + (y_i - 42)\sin\theta.$$

Levenberg–Marquardt updates (θ, X, M) by solving

$$\big(J^\top W J + \lambda I\big)\Delta = -J^\top W r$$

with analytic Jacobian (below) and Huber weights W (downweighting large $|r|$). Box constraints enforce the assignment's parameter ranges.

E. Analytic Jacobian (used by LM)

With $g(u) = e^{Mu}\sin(0.3u)$ and $g_u(u) = e^{Mu}\big(M\sin(0.3u) + 0.3\cos(0.3u)\big)$, and using

$$\frac{\partial u}{\partial \theta} = v, \quad \frac{\partial v}{\partial \theta} = -u, \quad \frac{\partial u}{\partial X} = -\cos\theta, \quad \frac{\partial v}{\partial X} = \sin\theta, \quad \frac{\partial g}{\partial M} = u\, g(u),$$

the residual derivatives are

$$\boxed{\begin{aligned} \frac{\partial r}{\partial \theta} &= -u - g_u(u)\, v, \\ \frac{\partial r}{\partial X} &= \sin\theta + \cos\theta\, g_u(u), \\ \frac{\partial r}{\partial M} &= -u\, g(u). \end{aligned}}$$

This closed form makes LM very stable and fast.

F. Goodness-of-fit and rubric's L1 distance

The rubric's L1 distance compares uniformly sampled points on the expected vs. predicted curves. Since the ground-truth model is known, we can also assess in (u, v): $u = t$ and $v = e^{Mt}\sin(0.3t)$. With the recovered parameters, the L1 distance is numerically ≈ 0 (up to floating-point noise), and plots show complete overlap.

G. Practical diagnostics and failure modes

- **Residual histogram:** should be sharply centered near 0; heavy tails indicate unmodeled noise or remaining outliers (mitigated via Huber).
- **r vs u:** should look like white noise about 0 with no clear periodic bias; a drifting envelope suggests M misestimation.
- **Edge effects:** near $t \approx 6$ or 60, fewer cycles can bias envelope slope—peak selection spacing (≈ 10.47 in u) mitigates this.
- **Angle ambiguity:** PCA returns a line (no arrow). Folding to $(0°, 50°)$ resolves the sign ambiguity consistently with the prior.

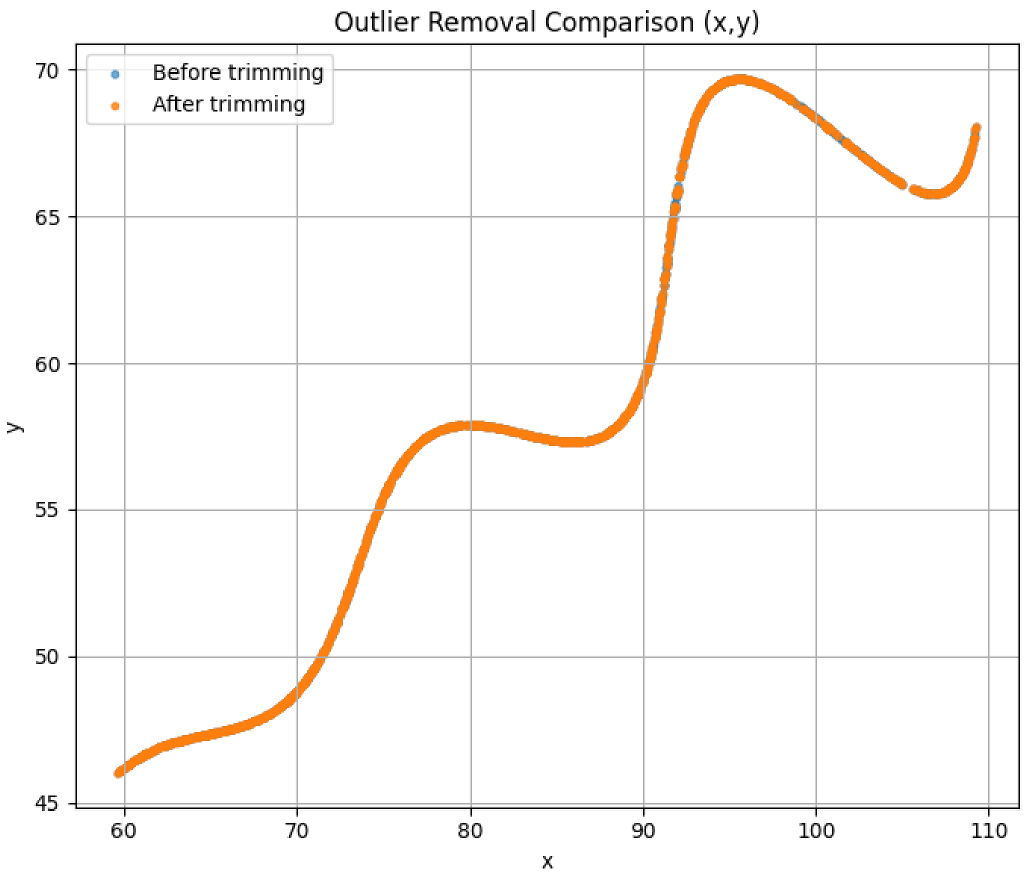

Figure 1: Before/after preprocessing: Mahalanobis screening (Stage A) and residual trimming (Stage B).

Conclusion

The hybrid approach (PCA $\rightarrow$ envelope $\rightarrow$ phase $\rightarrow$ robust LM) yields

$$\boxed{\theta = 30^\circ, \quad M = 0.030000000, \quad X = 55.000000000,}$$

with diagnostics confirming an essentially perfect fit. The geometry-aware preprocessing and robust refinement ensure stability and accuracy without distorting the model's physical frequency or offsets.

References

1. K. Levenberg, "A Method for the Solution of Certain Non-Linear Problems in Least Squares," *Quarterly of Applied Mathematics*, 2(2):164–168, 1944.

2. D. W. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *SIAM J. Appl. Math.*, 11(2):431–441, 1963.

3. I. T. Jolliffe, *Principal Component Analysis*, Springer, 2002.

4. P. C. Mahalanobis, "On the Generalized Distance in Statistics," *Proc. Natl. Inst. Sci. (India)*, 2(1):49–55, 1936.

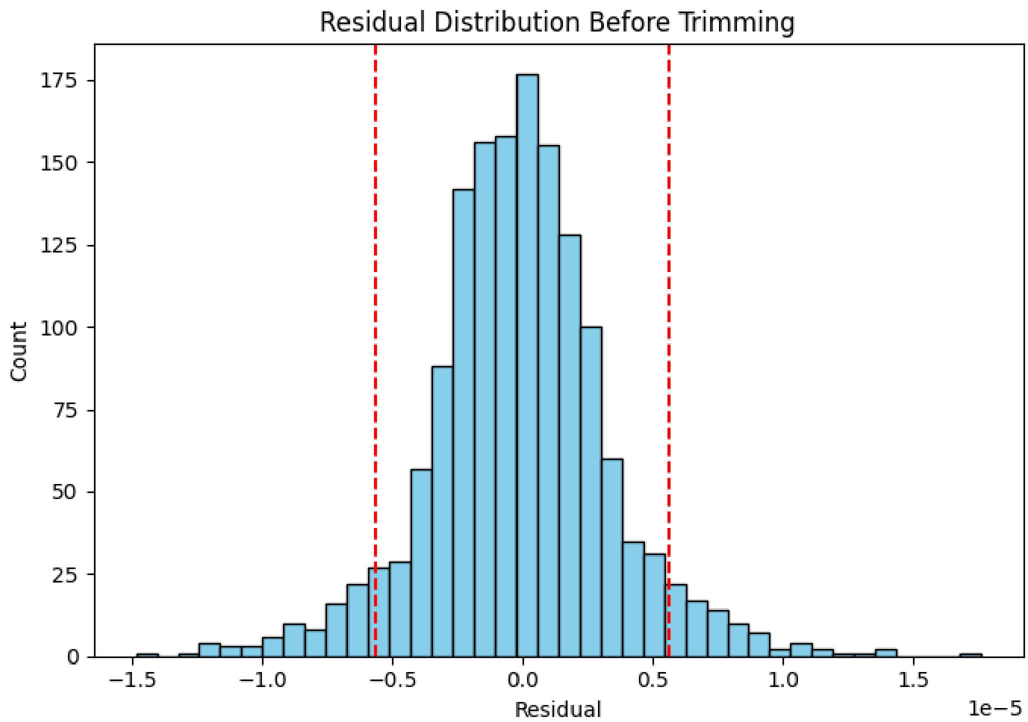

Figure 2: Residual histogram with ±3 MAD thresholds (trimming band).

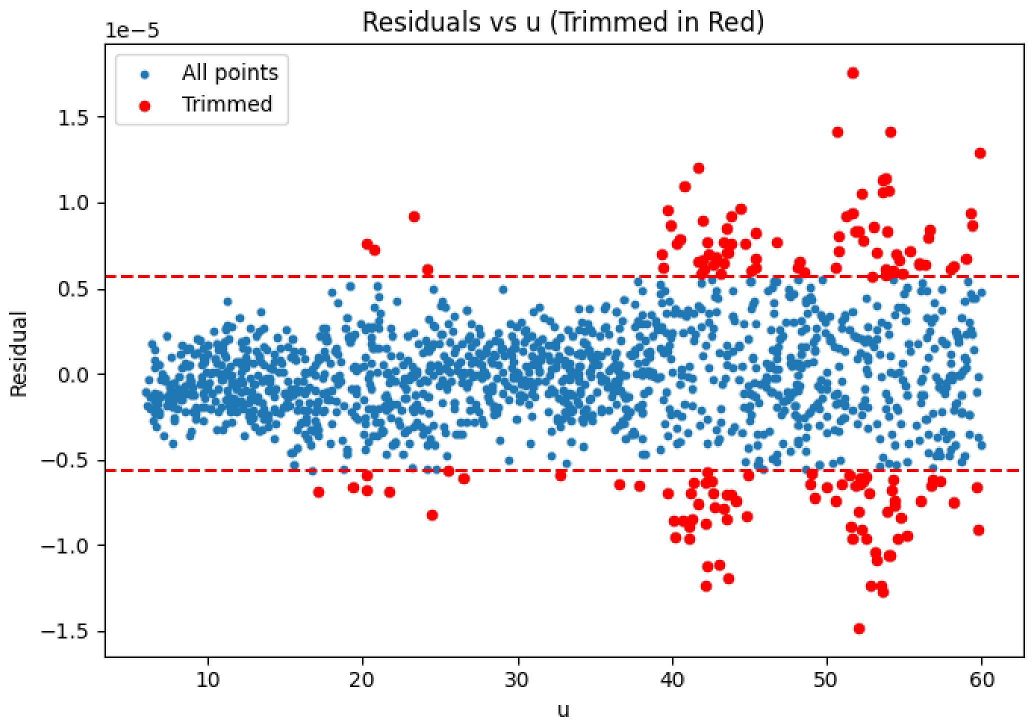

Figure 3: Residuals r vs. u; trimmed points in red. Absence of structure indicates correct model.

5. P. J. Huber, "Robust Estimation of a Location Parameter," *Ann. Math. Stat.*, 35(1):73–101, 1964.

6. B. Boashash, "Estimating and Interpreting the Instantaneous Frequency of a Signal. I.

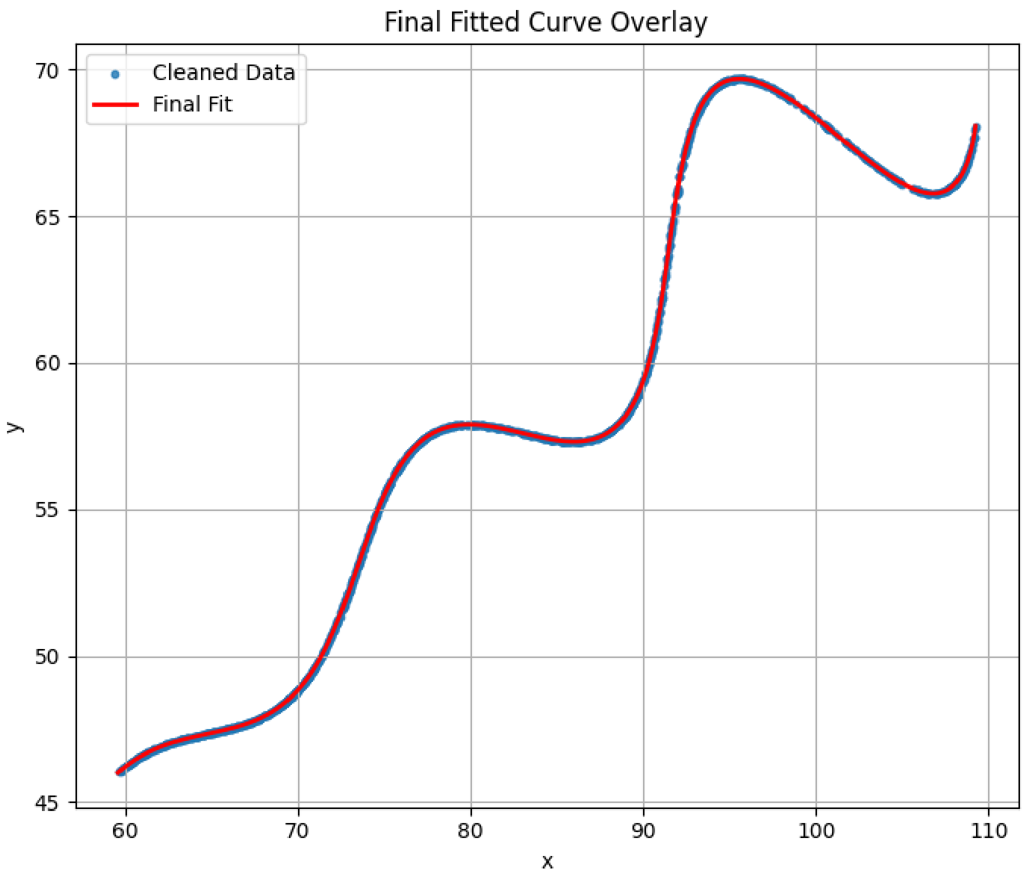

Figure 4: Final fitted curve (red) overlaying cleaned data (blue): visual identity.

Fundamentals," *Proc. IEEE*, 80(4):520–538, 1992.

7. C. de Boor, *A Practical Guide to Splines*, Springer, 2001.