KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING
School of Data Science and Analytics

One Family to Rule Robust Optimal Transport: The Weighted r-Power Framework

Rohan Jonnalagadda – May 2026, Advisor: Dr. Bin Luo

RESOURCES

INTRODUCTION

Optimal Transport (OT) provides a geometric framework for comparing probability distributions, yet it remains highly sensitive to outliers, which can severely distort transport plans. Recent advances such as ROBOT address this by hard-clipping excessive transport costs, improving robustness but limiting flexibility.

We extend this idea through WROT-r, a unified r-power formulation inspired by penalized weighted least squares. This framework generalizes multiple robust OT variants into a single mathematical family, where the parameter r acts as a "robustness knob." As $r \to 1^+$, the method enforces strict outlier rejection (ROBOT behavior); as r increases, it transitions into a soft-compression regime that gently damps large transport costs instead of truncating them. Our theoretical analysis and empirical results demonstrate that this tunable robustness achieves greater stability and accuracy, particularly under moderate noise and small-sample conditions, paving the way for adaptive and efficient

Mathematical Framework

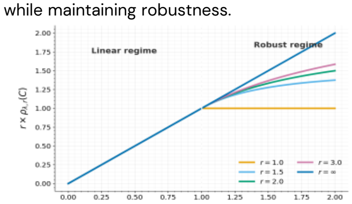
This unified cost function continuously links hard-clipping and soft-compression regimes through the parameter r, creating a smooth spectrum of robustness.

$$\min_{\pi \in \mathcal{F}(\mu,\nu)} \iint_{\mathcal{X}\times\mathcal{Y}} \rho_{\lambda,r}\big(c(x,y)\big)\, d\pi(x,y)$$

As $r \to 1^+$, it converges to the hard-clipped behavior of ROBOT, enforcing strict outlier rejection.

$$\rho_{\lambda,r}(C) = \begin{cases} \frac{1}{r} C, & C \le \lambda, \\ \lambda + \frac{1-r}{r}\left(\frac{\lambda}{C}\right)^{\frac{r}{r-1}}, & C > \lambda,\ r > 1, \\ \lambda, & C > \lambda,\ r = 1. \end{cases}$$

As r increases, it transitions into a soft-compression regime, where penalties decay gradually rather than abruptly, preserving stability while maintaining robustness.



Robustness Continuum

Robustness Spectrum (controlled by parameter r)

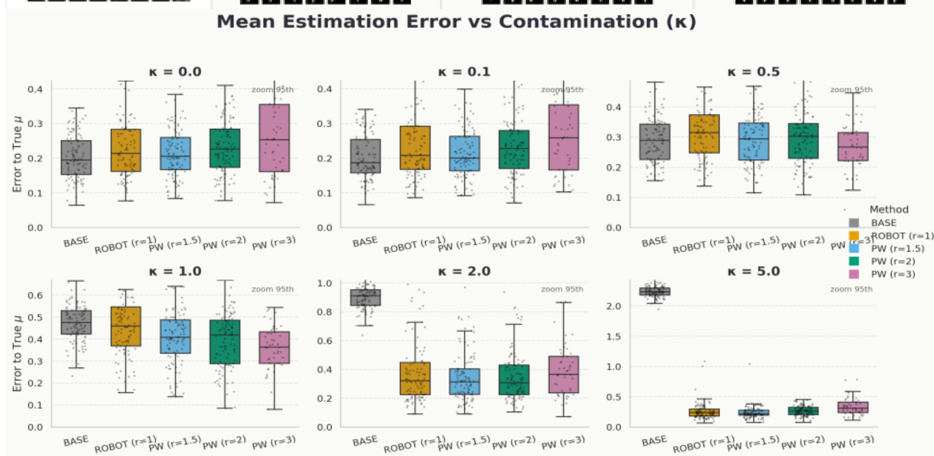
$r \to 1^+$
Hard-Clipping (ROBOT)
Cost: C' = min(C, 2λ)
Behavior: Strong rejection to outliers

r = 2
Soft-Compression
Cost: $\rho_{\lambda,2}$ (C)
Behavior: Gentle Damping of High Costs

$r \to \infty$
Smooth OT (less robust)
Cost: C (No clipping)
Behavior: Fully Sensitive to Outliers

ROBUST GAN PERFORMACE ACROSS THE ROBUSTNESS SPECTRUM

The top row shows digit 0s with white-image outliers, and the bottom row shows digit 1s with shoe outliers.



MEAN ESTIMATION ERROR ACROSS VARYING CONTAMINATION LEVELS (K)

This table reports the average estimation error (± standard deviation) of each method under increasing contamination ratios.

METHOD	K = 0.0	K = 0.1	K = 0.5	K = 1.0	K = 2.0	K = 5.0
BASE	0.204 ± 0.067	0.205 ± 0.068	0.288 ± 0.074	0.472 ± 0.082	0.899 ± 0.081	2.232 ± 0.085
ROBOT	0.371 ± 0.663	0.493 ± 0.942	0.605 ± 0.933	0.928 ± 1.249	0.495 ± 0.695	0.330 ± 0.525
PW(r =1.5)	0.220 ± 0.086	0.220 ± 0.090	0.303 ± 0.127	0.412 ± 0.141	0.343 ± 0.174	0.239 ± 0.107
PW (r=2)	0.232 ± 0.079	0.233 ± 0.083	0.297 ± 0.090	0.392 ± 0.123	0.349 ± 0.171	0.263 ± 0.082
PW (r=3)	0.265 ± 0.119	0.266 ± 0.113	0.267 ± 0.077	0.363 ± 0.099	0.386 ± 0.175	0.340 ± 0.146

EXPERIMENTAL SETUP

Robust GAN — MNIST (0) vs White Image:

- Inliers: digit "0"; Outliers: all-white images (pixel = 1); grayscale 28×28 flattened 784-D.
- DCGAN-style Generator (ConvTranspose2d + BN/ReLU + Sigmoid, z = 64) trained with robust OT loss (ROBOT/PW) using Adam ($\eta = 2\times10^{-3}$, $\varepsilon_{sh} = 0.2$, 10^4 steps); outliers from final good set.

Robust GAN — MNIST vs Fashion-MNIST

- Inliers: MNIST class 1; Outliers: Fashion-MNIST class 9 (n=6000,ε=0.2); 28×28 flattened pixels.
- Same generator and training setup (Adam 2×10^{-3}, $\varepsilon_{sh} = 0.2$, 10^4 steps, GPU/AMP enabled); λ∈{2, 4, …, 38}, metrics = masking & swamping vs λ + generated image grids.

Mean Estimation:

- Data: Synthetic Gaussian inliers with mean μtrue=[1,2,3,4,5], d=5, n=100. Outliers are shifted by k∈{0,0.1,0.5,1,2,5} with contamination ε=0.2.
- Estimator: Entropy-smoothed OT mean with robust costs (ROBOT, WROT(r=1.5), WROT(r=1), WROT(r=2), WROT(r=5)). Optimization via Adam, up to 100k outer steps; inner softmax updates on dual v every iteration.
- Errors were recorded for 100 simulations

OUTLIER DETECTION

- Synthetic samples $x_i \in \mathbb{R}^d$ are drawn from inliers $N(\mu_{\text{true}}, I_d)$ and outliers $N(\mu_{\text{true}}, -kI_d)$ with contamination ratio $\varepsilon = 0.2$.
- Points with max wij<1 are flagged as outliers, and detection performance is measured using masking (false negatives) and swamping (false positives).

$$w_i = \min\left(1,\ \left(\frac{\lambda}{C_i}\right)^{\frac{1}{r-1}}\right)$$

WROT–1.5

k	Accuracy	Precision	Recall	F1	Masking	Swamping
0	0.7695 ± 0.0362	0.1402 ± 0.1895	0.0495 ± 0.0614	0.0674 ± 0.0791	0.9505 ± 0.0614	0.0505 ± 0.0536
0.5	0.7778 ± 0.0348	0.2251 ± 0.2832	0.0700 ± 0.0880	0.0942 ± 0.1058	0.9300 ± 0.0880	0.0453 ± 0.0544
1	0.7980 ± 0.0301	0.4315 ± 0.3299	0.1935 ± 0.1933	0.2304 ± 0.1915	0.8065 ± 0.1933	0.0509 ± 0.0600
2	0.9119 ± 0.0435	0.8426 ± 0.1368	0.7470 ± 0.2666	0.7439 ± 0.1895	0.2530 ± 0.2666	0.0469 ± 0.0473
5	0.9708 ± 0.0351	0.8908 ± 0.1202	1.0000 ± 0.0000	0.9378 ± 0.0710	0.0000 ± 0.0000	0.0365 ± 0.0439

CONCLUSION

Robustness Behavior: For low contamination, smoother costs with higher r perform best, preserving more data while maintaining stability. Under moderate contamination , mid-range r achieves the best balance between rejecting outliers and retaining inliers. At high contamination (k ≥ 5), strict clipping with low $r \to 1$ (ROBOT) provides the most reliable estimates by fully suppressing extreme outliers.

Key Insight: The r-power formulation acts as a continuous robustness spectrum, enabling adaptive control between stability and resistance to contamination.

Applications: Reliable mean estimation in noisy data, Robust GANs that learn clean data manifolds under outlier corruption., Adaptive OT pipelines for generative modeling, clustering, and domain alignment.